

REPORTS AND STUDIES IN
**EDUCATION,
HUMANITIES,
AND THEOLOGY**

JANI-MATTI TIRKKONEN & ESA ANTTIKOSKI (EDS.)

*Proceedings of the 24th
Scandinavian Conference
of Linguistics*

PUBLICATIONS OF THE UNIVERSITY OF EASTERN FINLAND

Reports and Studies in Education, Humanities, and Theology No 5



UNIVERSITY OF
EASTERN FINLAND

EDITED BY JANI-MATTI TIRKKONEN & ESA ANTTIKOSKI

*Proceedings of the 24th
Scandinavian Conference of
Linguistics*

Publications of the University of Eastern Finland
Reports and Studies in Education, Humanities, and Theology

5

University of Eastern Finland
Joensuu
2013

Kopijyvä Oy
Joensuu, 2013
Series Editor: Maija Könönen

Distribution:
Eastern Finland University Library / Sales of publications
P.O. Box 107, FI-80101 Joensuu, Finland
tel. +358-50-3058396
<http://www.uef.fi/kirjasto>

ISBN: 978-952-61-0923-7 (nid.)
ISSNL: 1798-5641
ISSN: 1798-5641
ISBN: 978-952-61-0924-4 (PDF)
ISSN: 1798-565X (PDF)

Table of Contents

Contributors	6
Preface	9
1. SYNTAX	
Definite descriptions and Finnish complex demonstratives with ‘se’ <i>Erkki Ahlström</i>	11
The locality of expletive <i>der</i> in Danish embedded interrogative and relative clauses <i>Anne Bjerre</i>	23
Rethinking case assignment <i>Pauli Brattico</i>	37
Emphatic reflexives and logophoric marking in Modern Greek <i>Michael Chiou</i>	47
Gradient well-formedness of Finnish passive constructions <i>Fredrik Heinat and Satu Manninen</i>	59
Internal <i>wh</i> -movement in Finnish <i>Saara Huhmarniemi</i>	71
Observations on unique constituents in German <i>Juha Mulli</i>	80
Towards a phonological account of contrast in Italian: a pilot-study on contrastive topics <i>Jacopo Torregrossa</i>	90
Grammaticalization, productivity and analogy in Finnish converbs <i>Kendra Willson</i>	105
Against locative treatment of Polish object experiencers <i>Sylwiusz Żychliński</i>	115

2. LEXICOLOGY

- On the notion of “orthophemism” and other miscellanea on taboo and euphemism **130**
Michael Crombach
- The “small word” problem in medical dictation **147**
Michael Crombach
- Body parts and their names in Russian: the biological and semiotic pairs of body parts **155**
Grigory E. Kreydlin and Svetlana I. Pereverzeva
- Figurativeness of NPs containing lexemes ‘hand’, ‘head’ and ‘eye’ **165**
Jussi Niemi, Juha Mulli, Marja Nenonen, Sinikka Niemi, Alexandre Nikolaev and Esa Penttilä

3. SECOND LANGUAGE LEARNING

- Organising strategies: French and English in contrast **173**
Catherine Collin
- Second language acquisition: issues from Italian and Finnish at the syntax-discourse interface **187**
Lena Dal Pozzo
- The effect of social factors on the comprehension of a closely related language **201**
Charlotte Gooskens and Nanna Haug Hilton
- The influence of non-native morphosyntax and phonology on the intelligibility of a closely related language **211**
Nanna Haug Hilton, Charlotte Gooskens, Anja Schüppert, Renée van Bezooijen and Vincent van Heuven
- Teaching English tenses from a cognitive perspective **227**
Franka Kermer
- Finnish and English code-switching and borrowing in a Finland Swedish discussion program **237**
Anna-Kaisa Penttinen
- Lexical inhibition in trilingual speakers **249**
John W. Schwieter

4. DIACHRONIC LINGUISTICS

On the pronoun ego in the history of Russian <i>Elena Bratishenko</i>	262
The case of 'uninflected' infinitive in the Griko dialects of Sternatia and Calimera <i>Anna Frassanito</i>	274
Development of the relativization system in Scots and Scottish English <i>Sanna Hillberg</i>	289
Restructuring structures with modal and aspectual verbs in Archaic and Classical Latin <i>Rossella Iovino</i>	297
Archaisms in comparative forms of Modern Icelandic adjectives <i>Margrét Jónsdóttir</i>	311

5. LANGUAGE TECHNOLOGY

Automatic recognition of abbreviations and abbreviations' expansions in electronic texts – future development <i>Mojca Kompara</i>	321
Using a grammar checker and its error typology for annotation of statistical machine translation errors <i>Sara Stymne</i>	332

6. DISCOURSE

Language and social engagement in Asperger discourse <i>Jussi Niemi, Lidia Otsa, Aleksandra Evtjukova and John Niemi</i>	346
How to analyze the language of social networking sites – an analysis model <i>Michael Szurawitzki</i>	355
Conflict patterns in dialogues between wife and husband in Arthur Schnitzler's <i>Reigen</i> <i>Oliver Winkler</i>	364

Contributors

Erkki Ahlström

11 rue Albert Dupeyron
3150 Cenon, France
Erkkifi@yahoo.com

Anne Bjerre

Institut for Fagsprog, Kommunikation og
Informationsvidenskab
Syddansk Universitet, Kolding
Engstien 1
6000 Kolding
Danmark
bjerre@sitkom.sdu.dk
<http://www.sdu.dk/ansat/neville>

Elena Bratishenko

Department of Germanic, Slavic & East
Asian Studies
University of Calgary
Canada
bratish@ucalgary.ca

Pauli Brattico

Department of Computer Science and
Information Systems
University of Jyväskylä
FINLAND

Michael Chiou

Irinis Athineas 11,
11473, Athens
Greece
mchiou1234@gmail.com

Catherine Collin

University of Nantes (France)
Rue de la Censive du Tertre, BP 81227
F-44312 Nantes
Catherine.Collin@univ-nantes.fr

Michael Crombach

Healthcare Research and Development
Nuance Communications Austria
Triester Str. 64
1101 Wien
Austria
michael.crombach@gmx.at

Lena Dal Pozzo

CISCL - Centro Interdipartimentale di
Studi Cognitivi sul Linguaggio
Complesso S. Niccolò, Via Roma, 56, III
piano
I-53100 Siena, Italy

Anna Frassanito

Via delle Ghiacciaie, 39
50144 Florence
Italy
anna.frassanito@libero.it;
frassanito.anna@gmail.com

Charlotte Gooskens

Department of Scandinavian Languages
and Cultures
University of Groningen
P.O. Box 716
9700 AS GRONINGEN
The Netherlands
c.s.gooskens@rug.nl
<http://www.let.rug.nl/~gooskens/>

Fredrik Heinat & Satu Manninen

Stockholm University and Lund
University
Satu.Manninen@englund.lu.se
fredrik.heinat@ling.su.se

Sanna Hillberg

University of Eastern Finland
sanna.hillberg@uef.fi

Nanna Haug Hilton

Department of Frisian Language and
Culture

University of Groningen
P.O. Box 716
9700 AS GRONINGEN
The Netherlands
n.h.Hilton@rug.nl
<http://www.rug.nl/staff/n.h.hilton/>

Saara Huhmarniemi

Cognitive science
University of Helsinki
PL 9
00014 University of Helsinki
Finland
saara.huhmarniemi@helsinki.fi

Rossella Iovino

Department of Linguistic and Cultural
Studies
University Ca'Foscari of Venice
Ca' Bembo, Dorsoduro 1405
30123 Venice (VE)
Italy
rossella.iovino@unive.it

Margrét Jónsdóttir

School of Humanities
Faculty of Icelandic and Comparative
Cultural Studies
University of Iceland
IS-101 Reykjavík, Iceland
mjons@hi.is
<http://www.hi.is>

Franka Kermer

University of Eastern Finland at Joensuu
P.O. Box 111
80101 Joensuu
Finland
franka.kermer@uef.fi

Mojca Kompara

University of Primorska
Fakulteta za humanistične študije
Titov trg 5
6000 Koper
Slovenia
mokopty@yahoo.com

Grigory E. Kreydlin & Svetlana I.**Pereverzeva**

Institute of linguistics
Russian State University for the
Humanities
Miuskaya pl., 6
125993 Moscow
Russia
gekr@iitp.ru
P_Sveta@hotmail.com
<http://il.rsuh.ru>

Juha Mulli

German Language
University of Eastern Finland at Joensuu
P.O. Box 111
80101 Joensuu
Finland
juha.mulli@uef.fi

Jussi Niemi

Linguistics
University of Eastern Finland at Joensuu
P O B 111
FIN-80101 Joensuu, FINLAND
jussi.niemi@uef.fi
<http://www.uef.fi/filtdk/yleinen-kielitiede/henkilosto>

Anna-Kaisa Penttinen

University of Eastern Finland at Joensuu
Kaislakatu 10 N 124
80130 Joensuu
Finland
anpentti@student.uef.fi

Dr. John W. Schwieter

Department of Languages and Literatures
Wilfrid Laurier University
75 University Ave. West
Waterloo, Ontario
N2E 3Z6
Canada
jschwieter@wlu.ca
www.wlu.ca/homepage.php?grp_id=2267

Sara Stymne

Department of Computer and Information
Science
Linköping University
58183 Linköping
Sweden
sara.stymne@liu.se
<http://www.ida.liu.se/~sarst>

Michael Szurawitzki

Universität Regensburg
Germanistik
Deutsche Sprachwissenschaft
93040 Regensburg
Germany
Michael.Szurawitzki@sprachlit.uni-
regensburg.de
szurawitzki.blogspot.com

Jacopo Torregrossa

Department of Linguistics
University of Verona (Polo Zanotto)
Viale dell'Università, 4
37129 Verona (VR)
Italy
jacopo.torregrossa@univr.it
http://fermi.univr.it/live/people/Jacopo/people_jacopo.htm

Kendra Willson

Scandinavian Section
University of California, Los Angeles
212 Royce Hall
405 Hilgard Ave.
Los Angeles, CA 90095-1537 USA
willson@humnet.ucla.edu

Oliver Winkler

German Language and Literature
Åbo Akademi University
Fabriksgatan 2
20500 Åbo
Finland
owinkler@abo.fi
<http://www.abo.fi/institution/oliverwinklerty>

Sylwiusz Żychliński

Ul Kościelna 33/116
60-537 Poznań
sylwiusz@ifa.amu.edu.pl

Preface

The present collection of papers is based on the presentations given at the 24th Scandinavian Conference of Linguistics (SCL) held at the University of Joensuu (now: University of Eastern Finland) August 25 – 27, 2010. The listing of the preceding SCL meetings and their proceedings is to be found in the appendix of this volume.

Each paper was refereed by readers appointed by the Program Committee, and we as editors take this opportunity to thank each reader for their anonymous contribution to producing this review of linguistic studies in the Nordic countries. The organizers of the Joensuu meeting also wish to thank the University of Joensuu for its financial support to the present volume.

Jani-Matti Tirkkonen

Esa Anttikoski

1. *Syntax*

Erkki Ahlström

Definite descriptions and Finnish complex demonstratives with 'se'

This paper deals with the so-called *proto-article hypothesis* in Finnish. Some linguists, in particular Laury (1997), have suggested that in Finnish the demonstrative 'se' is, or is about to become, a definite article. Other linguists have taken a cautious stance towards this hypothesis (e.g. Larjavaara 2001a, 2001b, Juvonen 2000). Unfortunately, the debate on this question has not taken into account any standard semantic theories of definiteness or reference. The objective of this paper is to point out some relevant standard semantic phenomena which we need study in order to assess the validity of the proto-article hypothesis and to test some hypotheses derived from these phenomena.

1 BASIC DISTINCTION: TYPE AND TOKEN REFERENCE

Perhaps the most important distinction in modern semantics is that between *singular* and *general propositions*. Below we shall see that this distinction is essential in the explanation of the distinction between *definite descriptions* ('The F') and *complex demonstratives* ('This/that F'). In fact, it is the only basis which allows us to formulate precise hypotheses on the status of 'se' which can be corroborated (or falsified). Propositions refer to the truth-conditional content of utterances.¹ There are two (and only two) types of propositions, and the distinctive feature between them is that singular propositions are about particular individuals whereas general propositions are descriptive and do not involve directly any particular real-world entities. It is quite obvious that we can intend to refer to particular real-world objects with our utterances and we can make this intention manifest to our audience. Communication can thus be *object-dependent* or *object-independent*. Some determine phrases (henceforth DPs) in natural language are invariably used to refer to particular referents, while others can be simply descriptive. The former group consists of certain indexical pronouns ('I'), demonstratives ('that') and proper names ('Sarah Palin', 'The White House'), and the second group consists of *quantificational DPs* ('two

¹ We shall not discuss the thorny issues between literalism and contextualism which concern the question whether truth-conditions can be determined at the level of sentences or utterances. These questions are discussed in Récanati (1993).

Finns, 'most Finns'). The former DPs are *type-referential* whereas DPs of the second type can *token-referential*, i.e. token-referential DPs can use used to refer, but this is not an inherent part of their meaning. The former group imposes some cognitive conditions. For instance, the referent of the complex demonstrative '*this table*' has to be recognizable in the context (Récanati 1993). When we use a proper name, we are in a causal connection with the bearer of the name. The truth-conditional role of semantically referential DPs is limited to the introduction of the referent into the singular proposition, i.e. the referent itself is relevant from the point of view of what is said. In the case of descriptive DPs, the quantificational DP introduces a descriptive condition into the object-independent proposition of the utterance in which it is used. (Kripke 1972, Donnellan 1972).

1.1 Definite descriptions

According to the view we endorse, the semantic contribution of definite descriptions is quantificational. The semantic structure of the schematic sentence '*The F is G*' has the following semantic structure expressed used restricted quantification:

$$(1) \quad \text{'The } x: Fx \text{ } (Gx) \text{' is true iff } \int_{F-G} \neq 0 \text{ and } \int_F = 1$$

The contribution of the definite description in the proposition above is a *uniqueness condition*, which we can paraphrase in the following way: there is one and only x such that it is F, and there are no such Fs that they are not G, and the number (or the cardinality) of Fs is one. The essential point in the analysis is that the referent of the definite description is not directly relevant: the contribution of the definite description is descriptive (Neale 1991).

There is an obvious problem with this theory: most definite descriptions are not unique, and yet we can use very poor descriptions, like '*the table*' as if they were unique (if the analysis above is correct). However, it is obvious that quantifiers in general are used without any limitations. Consider the following case

$$(2) \quad \text{Everybody came to my party yesterday.}$$

To understand the precise meaning of this sentence in a context, one must be able to limit the relevant domain of interpretation of the quantificational pronoun '*everybody*'. In the same way a simple definite description is interpreted within a limited domain. Hawkins (1978) has presented the most relevant classification of these different domains:

- 1) Anaphoric use: '*A table ...the table*'; 2) Associative anaphoric use: '*A car...the engine/*the wheel*';
- 3) Immediate situation use (in a restaurant, speakers a waiter) '*The waiter*'; 4) Larger situation use (in the US, France...) '*the president*' (on Earth) '*the sun*' .

We shall not discuss the justification of this classification here in detail. It is our working hypothesis that definite DPs - like many other quantifiers - need to be localized within a limited domain and that this (or some similar) classification is relevant.

If definite descriptions are descriptive DPs, *i.e.* quantifiers, they must be sensitive to *scope*. This seems to be the case if we consider the following examples:

- (3) *Each guy admires the woman who brought him up.*
- (4) *Each guy had at least five beers last night.*

The DPs '*the woman who brought him up*' and '*at least five beers*' are affected by the quantificational subject DP. Furthermore, the quantificational DP in the adjunct can have the same effect:

- (5) *In most big companies, the CEO is given stock options.*

In this example, the uniqueness conditions of the DP '*the CEO*' is interpreted in the scope of the quantificational DP '*most big companies*', as our theory predicts.

However, in spite of the similarities between definite descriptions and other quantifiers (domain-sensitive interpretation and scope ambiguities), it is quite obvious that definite descriptions are not only quantificational; they are also used to refer to concrete entities, but this is a contingent fact belonging to the domain of pragmatics. The two dimensions of definite descriptions can easily be captured if we study the following example:

- (6) *The inventor of the light-bulb was a genius.*

This sentence has two interpretations, one which concerns Edison personally. The speaker wants to make his hearers understand that he wants to speak about the actual person who invented the light-bulb. This is the *referential reading*. On the other hand, the sentence can also be used to communicate a purely descriptive idea. The light-bulb is such a clever invention that whoever the inventor was, he was necessarily a genius, but this idea does not concern any particular individual. This latter interpretation is the *attributive reading*. (Donnellan (1966) 1999). There is a fairly wide consensus that the purely descriptive content is relevant in both these cases and that the referential interpretation has something extra besides the object-independent proposition.² The essential point is that in all uses of definite descriptions, the object-independent proposition has a role to play. In other words, the definite article (in its non-generic uses) introduces a uniqueness condition in the propositional content. How these two readings above are combined is more controversial, but we need not to go into details here.³

² This was Donnellan's ((1966) 1999) own view, and among others who agree we can mention Saul Kripke ((1979) 1991), Kent Bach (1987), Stephen Neale (1991) and François Récanati (1993).

³ Saul Kripke ((1979) 1991) uses the notion of speaker reference as opposed to semantic reference, Neale (1991) presents an analysis of this phenomenon in terms of conversation implicatures, and so forth.

1.2 Complex demonstratives

Simple and complex demonstratives are generally considered to be type-referential and they are used to express object- dependent or singular propositions. They belong to a limited class of indexical expressions which refer directly to the context of utterance. The reference of demonstratives is determined by the referential intentions of the speaker who uses them. When we use a demonstrative we need to look for the referent the speaker intends to refer to (Kaplan 1989).

The distinction between type and token reference can be illustrated using the notion of rigid designation. Type-referential DPs are rigid designators, and their reference is stable in all situations or circumstances (or possible worlds, to use a more technical term). On the other hand, the reference of descriptive DPs can vary from one situation to the other, *i.e.* they are not rigid designators. This difference is easy to illustrate with modal operators. Proper names and complex demonstratives are rigid terms, and their interpretation is not affected by the modal operator of possibility '*could*', whereas the interpretation of definite descriptions can change in different situations. We may speculate how things might have been by using the modal auxiliary of possibility '*could*':

(7) *That man/Barack Obama could have been a woman.*

In (7) we always talk about the same referent, and the proposition is patently false, whereas the definite description in (8) has different interpretations:

(8) *The President of the US could have been a woman.*

It is obvious that this sentence has a referential reading which is false if the referent is the present president of the US, but there is also a reading in which the modal operator affects the interpretation of the definite description; in this case, the speaker says that there might have been a situation in which a woman was elected the president. This reading is never possible with type-referential DPs. When we speak about modal scope with reference to DP, we refer to the fact that descriptive DPs *can* be affected modal operators.

From the point of view of the proposition expressed, the meaning of the demonstrative has no other role besides the identification of the referent. What is important is that the referent satisfies the predication of the rest of the utterance. Things are more complicated when we move to complex demonstratives. The role of the common noun phrase in the construction is controversial. Is it only part of the complex character the expression like the demonstrative element in it? Does it introduce descriptive condition like definite descriptions do? This question can be solved with a multipositional solution, but we shall not dwell on this issue here (on this topic see Dever (2001)).

Recently, Jeffrey King (2001) has drawn our attention to the fact that the interpretation which excludes the NP from the proposition cannot be true, at least not in all cases, because there are uses where the complex demonstrative behaves exactly like a definite description. Here are a couple of his examples:

- (9) *That hominid who discovered how to start fires was a genius*
(10) *Every father dreads that moment when his eldest child leaves home.*

It is obvious that in these cases the demonstratives are non-referring and can take narrow scope with respect to other operators in the sentence just like definite descriptions, too. However, we must bear in mind that in English these uses are possible only with the distal demonstrative 'that'.⁴ These cases imply that the picture of demonstratives we drew at the beginning of this paper is somewhat blurred, but these cases hardly falsifies the referential picture, because these uses are limited to certain syntactic structures.⁵

This presentation has not taken into account textual uses so far, but the distinction between the textual uses of definite descriptions and complex demonstratives is similar to the type vs. token reference distinction. A complex demonstrative needs an explicit antecedent in the text, whereas definite descriptions function on the basis of their descriptive uniqueness. Thus complex demonstratives are excluded from associative anaphora where there is no direct antecedent:

- (11) *I'd like to buy a Bentley, but I fear that the/*that engine would consume more gas than I can afford.*

The complex demonstrative is quite impossible in this use.⁶ We shall come back to this limitation below.

2 'SE' AND REFERENCE

The first part of this paper summarizes the essential points of the distinction between definite descriptions and complex demonstratives which we need to taken into account if we want to study the status of 'se' in Finnish. The second part applies into Finnish the insights presented briefly above. To assess the hypothesis that 'se' is a kind of a proto-article, I have taken examples from semantic literature on definiteness and I have first figured out if I myself would use 'se' in some critical uses. Then I have turned to people in my in-groups and peer groups (family and friends, living mainly, but not exclusively, in South-Western Finland), and then I have discussed these insights with some other informants. Unlike most previous work, my observations are thus NOT based on any corpus. My conclusions are only qualitative and synchronic. I concentrate exclusively on the critical uses which typically distinguish definite descriptions from complex demonstratives.

⁴ Larjavaara (1990: 158) discusses briefly similar uses in Finnish.

⁵ See Wolter (2006, chapter 4) for a discussion of these syntactic features, see Larjavaara (1990) for similar restrictions in Finnish.

⁶ See Kleiber (2001) for a particularly relevant discussion of associative anaphora in general this limitation in particular.

Before we start, we have to mention that in Finnish, there are some uses of 'se' which may blur the discussion of the proto-article hypothesis. All demonstratives are possible when the speaker tries to retrieve a particular lexical item in Finnish (Larjavaara 1990: 169):

- (12) *Tää, tää/toi, toi/se, se ...Pekkaki on kuollu.*
This/that... Pekka has died, too.

In this case the demonstrative makes no contribution to the proposition, it simply implies that the speaker is hesitating. Similarly, demonstratives are often used to indicate subjective attitudes,⁷ and here are some examples from French and Finnish:

- (13) *Ce Pierre, c'est un vrai con.*
Se Pekka on kyl täys nuija.
SE Pekka is a complete jerk.

Here again, the demonstrative makes no truth-conditional contribution. Some people one might be tempted to think that 'se' in these case makes no contribution and the cases below are similar. However, we shall see the use of 'se' is subject to conditions similar to those of genuine definite articles.

2.1 Non-referring demonstratives and 'se'

First all, we shall discuss some King's cases, in which the demonstrative is non-referring:

- (14) *(Colloquial Finnish) Joka äijä pelkää sitä/*Ø päivää, ku muija saa tarpeekses ja lähtee nostelee.*
(Formal Finnish) Jokainen mies pelkää sitä/Ø päivää, jolloin hänen vaimonsa saa tarpeekseen ja lähtee kotoa.
Every guy is afraid of SITÄ (partitive form of 'se') day when his wife has had enough and leaves.
Every guy is afraid of that/the day when his wife has had enough and leaves.
- (15) *(Colloquial Finnish) Sel/*Ø apinamies, ku tulen keksi, oli kyl oikee nero.*
(Formal Finnish) Sel/Ø apinamies, joka keksi tulen, oli nero.
SE ape-man who invented the fire was a genius.
That/the ape-man who invented the fire was a genius.

In these particular cases, 'se' is virtually obligatory in colloquial spoken Finnish; if we leave it out, the sentence is ungrammatical. These cases are interesting because the truth-conditional contribution of the English demonstrative in these uses is similar to that of a definite article and

⁷ These uses are well-known ever since Lakoff (1974).

the definite article is equally possible with no real difference in meaning, but in colloquial spoken Finnish, a bare noun phrase is quite impossible.

2.2 ‘Se’, scope and associative anaphora

In the literature on the use of proto-articles in Finnish, scope is never discussed because. There are two phenomena which we must take into account, as we noticed above. There are the cases where a quantifier takes wide scope over other descriptive DPs. These cases are not limited to the phenomena traditionally treated under the notion of *quantifier raising*; it is possible to find similar cases with quantifiers in adjuncts. Basically, the mechanism is the same in both these cases. The second aspect we include under the heading of scope concerns the interpretation of quantificational DPs in the scope of modal operators.

Associative anaphora is a particularly useful indicator of the eventual quantificational status of ‘se’. We shall discuss scope in Finnish together with associative anaphora. In formal Finnish, it is quite out of the question to use complex demonstratives in associative anaphora:

- (16) *Jos ostatte Lexus-katumaasturin, *se moottori ei ainakaan jätä toivomisen varaa.*
If you buy a Lexus SUV, SE engine should satisfy you.
If you buy a Lexus SUV, the engine should satisfy you.

On the other hand, in spoken Finnish, my informants and I myself were quite happy with the following cases:

- (17) *Jos sää joskus meinaat ostat Rutinoffin⁸, niin muist sit kans kattoo se moottori kunnolla!*
If ever you are planning to buy a Rutinoff, remember to check SE engine properly.
If ever you are planning to buy a Rutinoff, remember to check the engine properly.
- (18) *Aina ku mää ajan jonku kaveri Rutinoffil, nii se moottori pitää aina kyl niin maan kauhiaa äänt!*
Each time a drive a friend’s Rutinoff, SE engine always makes too much noise.
Each time a drive a friend’s Rutinoff, SE engine always makes too much noise.
- (19) *Jokases Rutinoffis kun oon ostanu, nii se moottori on aina levinny ekana.*
In each Rutinoff I have bought, it has always been SE engine that has broken down first.
In each Rutinoff I have bought, it has always been SE engine that has broken down first.

These examples have been selected on purpose so that the referential interpretation is unlikely or even impossible. In (17), we are talking about hypothetical future events, in (18) the quantificational adjunct quantifies (‘*aina*’) over points of time and takes wide scope over the DP ‘*se moottori*’, and in (19), there is also a quantificational adjunct (‘*Jokases Rutinoffis*’). Even if the

⁸ Because the examples are rather negative, we rather use an imaginary make of car, Rutinoff, invented by the late Finnish humorist Spede Pasanen.

demonstrative is impossible in any article language in these examples, the result sounds rather good and sometimes even better with 'se' in colloquial Finnish, even if the difference with the bare form is not as blatant as in King's cases we just saw. Here, we come close to article-like frequency, but it is evident that the register has to be very colloquial.

There is an important aspect concerning associative anaphora which we need to point out because it may cause confusion. We must not mix associative anaphora and *memorial deixis*. In the latter use, complex demonstratives are quite normal, and they compete with definite descriptions, as in the following where the speaker is evoking holiday memories:

(20) *Do you remember that/the bloke who went swimming with a bowler hat?*

Here the speaker recalls for instance holiday memories she shares with the hearer, and they both have a perception-based mental representation of the particular referent. We are thus dealing with object-dependent communication, even if the strict situational perception condition is not relevant any more. Demonstratives are quite common in these uses, even if this function in English only concerns the distal 'that'.

In Finnish, the demonstrative 'se' is quite common in this use both in formal and colloquial register:

(21) *Muistatko sen/Ømiehen, joka meni uimaan knalli päässä?
Mustaks sää sen/*Øäijä ku meni uimaa knalli päässä?
Do you remember that/the bloke who went swimming wearing a bowler hat?*

If the speaker and the hearer both have had a perception-based mental representation of the referent, demonstratives are possible both in Finnish and in English, but in Finnish these uses may be more natural with 'se', especially in colloquial spoken Finnish. This may be an indication of a closer article like status. This may be also the reason why 'se' can easily be used asymmetrically in memorial deixis in cases where only the speaker has a perception-based representation of the referent below:

(22) *Kuule kyl sulki olis ollu kattomist, Hartikaine, ku ne/Øhulahulatyöt oikee keikutti!
Listen Hartikainen, you shold have seen how NE (plural of 'se') hula-hula girls moved their hips!
Listen Hartikainen, you shold have seen how those/the hula-hula girls moved their hips!*

It is possible that in Finnish these latter uses are more natural than in some other languages; especially in French some of my informants found demonstratives rather unnatural in this use. If this is the case, it may be a sign that 'se' is indeed is close to the status of a genuine article.

De Mulder and Carlier (2006) consider these memorial uses as a bridge from the status of a genuine referential demonstrative to that of a definite article (they discuss some examples in Old French). This idea is plausible in the light of the following authentic example, taken from Sundbäck (1995: 2):

- (23) *Sittehä siinä oli tää oma, puutarha ja viinimarjapensaat ja yks omenapuu, joka oli semmonen oli semmone jokasyksyne murheenlapsi, ku, varkaat vei ne omenat aina, enne aikoja.*
Then there was this own, garden and currant shrubs and one apple tree, which was a problem each fall, because the thieves always took NE (plural of 'se') apples before they were ripe.⁹

The last DP *'ne omenat/those apples'* is interesting because the adjunct with the quantifier *'joka syksy /each fall'* takes wide scope over the DP *'ne omenat/those apples'*. In English, the demonstrative is quite impossible, but in Finnish the result is quite acceptable. The complex demonstrative *'ne omenat'* is naturally not a referential term because it has narrow scope, but the speaker has a perception based mental representation of the referents which were stolen and he is annoyed. So, we have an associative frame (*'apple tree – apples'*), a strong attitude (the speaker was annoyed) which is typical of the subjective uses of demonstratives, and perception-based representations of the apples each fall, even if the apples are different each year, and still the narrow scope complex demonstrative is possible. Because of the perception-based representation, this example illustrates nicely the eventual generalization of these demonstratives. In example (17) above, there is no possibility of any perception-based mental representation of the referent. The sentence is used to present a hypothetical state of affairs in the future, but in (23), there is an interesting combination of narrow scope and direct reference.

2.3 'Se' and modal scope

The second scope related feature concerns modal operators. The distal demonstrative, which in English can sometimes behave like a define article, does not take narrow scope with respect to the modal operator *'could'*:

- (24) *In England the PM could be a woman*
In England that PM could be a woman

The latter sentence has the impossible implication that the actual referent could be a woman, whereas in the first example the definite description can take narrow scope. However, in colloquial spoken Finnish, the demonstrative *'se'* can sometimes take narrow scope with respect to modals operators, *e.g.* in the case below:

- (25) *Suomes se pressa kyl vois yhtä hyvin olla naine.*
In Finland SE president could equally well be a woman.
In Finland the president could equally well be a woman.

This example is quite natural, but the demonstrative could also be left out without creating a stylistic anomaly. One of my informants found this use a bit oldish, and it is true that this kind

⁹ This is an authentic example with hesitation and pauses, and we shall edit it to give a perfect equivalent in English.

of a sentence would more likely be used by elderly people, perhaps more in rural areas, but these are just subjective comments.

2.4 'Se' in generic/habitual use

Generic uses are common with definite descriptions in article languages, but this is a contingent fact; the existence of a genuine definite article does not imply that it can be used in generic reference. Even if a given language uses systematically definite descriptions instead of bare noun phrases in generic reference (French: '*Le vin nest bon pour la santé*' versus English: '*Wine is good for health*'), these definite descriptions cannot be explained on the same basis as the cases discussed so far; uniqueness is not relevant in generic reference.

According to Vilkuna (1992: 135) 'se' is sometimes possible in generic use, as in the following case:

- (26) *Kyllä kai se majava siten on jyrksijä, kun kaikki niin sanoo.*
I suppose SE beaver is a rodent because everybody keeps saying so.
I suppose the beaver is a rodent/beavers are rodents because everybody keeps saying so.

Vilkuna limits the possibility of the generic 'se' to cases where there has already been a discussion about beavers and their classification, so that they have already been mentioned. It is unlikely that this use would appear 'out of the blue'. However, there are also cases where this does seem to be possible. For instance, this authentic example has been used by a foreman when a new employee starts work at a new job:

- (27) *Täällä ei sit juoda sitä viinaa!*
You don't drink SITÄ (partitive of 'se') booze here!
You don't drink alcohol here!

Alcohol had not been mentioned before in the discourse, but the 'se' is quite possible, and this is the case also in the example below:

- (28) *A: Mikäs Pekkaa vaivaa?*
B: No ku se polttaa sitä tupakkaa!
A: What's wrong with Pekka?
B: Well he smokes SITÄ (partitive of 'se') tobacco.
B. Well he smokes.

A is worried about Pekka's health, and B's answer seems to express frustration: Pekka smokes even if he knows quite well that it is bad for him, and B is annoyed. Both these examples were natural according to my informants, even if some of them indicated that they may be more typical of south-western Finland and that they would not use 'se'. In both (27) and (28), the demonstrative can be left out, but it weakens the subjective tone. The contribution of 'se' is thus subjective or affective.

3 CONCLUSIONS

The objective of this short paper is to give a more systematic explanation of some article-like uses of complex demonstratives with 'se' in Finnish. We started our discussion with a very brief analysis of the difference between definite descriptions and complex demonstratives. It was necessary, because the notions presented above are indispensable if we try to analyze the status of 'se' with any precision. It is quite obvious that there are syntactic phenomena which definite descriptions and complex demonstratives (as well as proper names) share (see *e.g.* Lyons 1999: 16- 27), but these facts are irrelevant for the truth-conditional contribution of these DPs. These similarities may be explained by cognitive notions like individual concepts. However, identification or individualization is too vague to give any solid starting point to discuss the problem at hand. Reference, rigid designation and quantification are notions which give precise arguments which we can test empirically. Even if the number of cases discussed here is limited, they do point out some interesting generalizations.

In the light of previous work and our observations, it is quite obvious that there is no definite article in Finnish. The use of a genuine definite article does not depend on any stylistic matters nor is it limited to any particular register or style. 'Se' is possible only in colloquial spoken Finnish. In formal language it is heavily stigmatized. However, the observations above do show that in colloquial spoken Finnish, 'se+NP' is frequently used in contexts where the definite article would appear in article languages, and in many cases, the use of 'se' is preferred or even practically obligatory. The cases discussed here are merely illustrative, but they should indicate how to proceed on this issue.

BIBLIOGRAPHY

- Bach, Kent. 1987. *Thought and Reference*. Oxford: Oxford University Press.
- Dever, Josh. 2001. "Complex Demonstratives". *Linguistics and Philosophy* 24, 771-330.
- Donnellan, Keith. (1966) 1999. "Reference and definite descriptions", in Ostertag, G. (ed.) *Definite Descriptions. A Reader*, Cambridge, Mass: MIT Press.
- Donnellan, Keith. 1972. "Proper Names and Identifying Descriptions", in D. Davidson and G. Harman (eds) *The Semantics of Natural Language*, Dordrecht: Reidel.
- Hawkins John. 1978. *Definiteness and indefiniteness. A study in Reference and Grammaticality Prediction*. London: Croom Helms.
- Juvonen, Päivi. 2000. *Grammaticalizing the Finnish Article*. Unpublished PhD-thesis. University of Stockholm.
- Kaplan, David. 1989. "Demonstratives", in Almog, J. et, al. (éd.). *Themes from Kaplan*, New York: Oxford University Press.
- King, Jeffrey 2001. *Complex Demonstratives: A Quantificational Account*. Cambridge, Mass: MIT Press.
- Kleiber, Georges. 2001. *Anaphore associative*. Paris : PUF.
- Kripke, Saul. 1972. *Naming and Necessity*, in D. Davidson and G. Harman (eds) *The Semantics of Natural Language*. Dordrecht: Reidel.

- Kripke, Saul. (1979) 1991. "Speaker's Reference and Semantic Reference", in Davis (ed.) *Pragmatics, A Reader*, 77-96, Oxford: Oxford University Press.
- Lakoff, Robin. 1974. "Remarks on This and That", in LaGaly, W., Fox, R. Et Bruck, A. (eds). 1974. *Papers from the Tenth Regional Meeting of the Chicago Linguistic Society*.
- Larjavaara, Matti. 1990. *Suomen deiksis*. Helsinki : SKS.
- Larjavaara, Matti. 2001a. "Määräinen artikkeli - suomessa? ". *Kielikello* 4, 25–27.
- Larjavaara, Matti. 2001b. "Suomen niin sanottu artikkeli". *Sananjalka* 43, 191–203.
- Laury, Ritva. 1997. *Demonstratives in Interaction*. Amsterdam: John Benjamins.
- Lyons, Christopher. 1999. *Definiteness*, Cambridge: CUP.
- de Mulder, Walter and Carlier, A. 2006. "Du démonstratif à l'article défini: le cas de ce en français moderne", in *Langue française* 152, 96-113.
- Neale, Stephen. 1990. *Descriptions*. Cambridge Mass: MIT Press.
- Récanati, François. 1993. *Direct Reference*. Oxford: Blackwell.
- Sundbäck, Irene. 1995. "Suomen kielessäkö artikkeli? ". *Kielikello* 2. 225 – 228.
- Vilkuna, Maria. 1992. *Referenssi ja määräisyys suomenkielisten tekstien tulkinnaissa*. Helsinki: SKS.
- Wolter, Lynsey. 2006. *That's that. The Semantics ad Pragmatics of Demonstrative Noun Phrases*. Unpublished PhD-thesis, University of California: Santa Cruz.

Anne Bjerre

The locality of expletive der in Danish embedded interrogative and relative clauses

1 INTRODUCTION

In Danish, *der* ('there') is used as a substitution for a missing or displaced subject in both impersonal passive, presentational, interrogative and bound and free relative clauses, cf. e.g. Erteschik-Shir (1984) and Vikner (1991). Examples are shown in (1).¹

(1) a. Nu kan der ringes og modtages SMS.

Now can there be called and received SMS

b. Her kan det faktum, at der løber en voksen
Here can the fact that there runs an adult

nogle meter bagved, være særdeles betryggende.
a few meters behind, be extremely reassuring

c. Jeg ved ikke hvem der synger den.

I know not who there sings it

d. Der er ingen signalement på den mand der førte bilen.

There is no description of the man, there drove the

e. Alt hvad der næres vokser.

All what there is nourished grows

Various analyses of this *der* have been proposed. The traditional grammarians have argued that *der* functions as an expletive subject in all these contexts, cf. e.g. Wiwel (1901), Diderichsen

¹ The data used in this paper are authentic examples from the Internet and examples from Vikner (1991) and Hansen (1974). Some of the uncontroversial examples are made up.

(1962), Hansen (1974) and Heltoft (1986). Within more formal analyses, discussions have focussed on whether the occurrences of *der* in (1c) through (1e) are indeed analyzed as occurrences of an expletive in subject position or perhaps better as occurrences of a complementizer in C⁰. Framed within the GB tradition, the analyses of e.g. Jacobsen and Jensen (1982) and Vikner (1991) treat *der* in embedded interrogative and relative clauses as a complementizer. Erteschik-Shir (1984) and Mikkelsen (2002), also within the GB tradition, propose analyses where *der* is an expletive in subject position. What the proposals have in common is a focus on the internal structure of the clause of which *der* is part.

In this paper a formal analysis of *der* in Danish embedded interrogative and relative clauses is presented which is formulated within an alternative formal framework, i.e. Head-Driven Phrase Structure Grammar, cf. e.g. Pollard and Sag (1994). The analysis differs from the GB analyses in not employing movement and traces to account for the constructions involving extraction, relying instead on surface-oriented structures and constraint satisfaction, cf. e.g. Sag and Fodor (1994), Sag (1997), Ginzburg and Sag (2000) and Sag (to appear). The presented analysis treats *der* in embedded interrogatives and relative clauses as an expletive subject substitute in subject position in line with traditional Danish grammar theory. In addition to the internal structure of the clauses with *der*, this analysis emphasizes the structural context in which the clauses appear. Taking the structural context into account will be shown to account for a distributional fact about *der* which so far has escaped explanation. In Danish the second conjunct in a coordinated relative clause cannot contain *der*.

2. MOVEMENT AND TRACE-BASED ANALYSES OF *DER*

Vikner (1991, p. 120) puts forward the constraints in (2) governing the distribution of *der*. The constraints are formulated in relation to the Empty Category Principle, cf. Chomsky (1981).

- (2) *der* may properly govern the spec of its complement iff this spec is coindexed with its own spec
der may only occur if the spec of its complement is coindexed with its own spec

The example in (3) is from Vikner (1991).

- (3) Vi kender de lingvister, der vil læse denne bog.
 We know the linguists, there will read this book

Vikner's analysis assumes the structure in (4) for the example in (3).

- (4) Vi kender de lingvister_i, [_{CP} OP_i der_i [_{IP} t_i vil læse denne bog]]

The empty operator moves from IP-spec to CP-spec, leaving a trace in IP-spec. *Der* is a complementizer in C⁰. The complement of *der* is the IP, and the specifier of *der* is the empty operator OP in CP-spec. The example is well-formed, as the constraints in (2) are met. The spec of the complement of *der* is coindexed with *der*'s own spec, and hence *der* properly governs the trace in IP-spec.

Mikkelsen (2002), also within the GB tradition, provides an analysis of the distribution of *der* in subject relative clauses. She proposes that the distribution of *der* is a consequence of its expletive status. *Der* only occurs in the position targeted by the Extended Projection Principle, cf. Chomsky (1981). The principle can be satisfied in two ways according to Mikkelsen (2002). When the subject in a subject relative clause moves directly from its thematic position to CP-spec, *der* is inserted in IP-spec to satisfy the principle. If the subject moves to CP-spec via IP-spec, it leaves a trace in IP-spec, and no expletive *der* is inserted. In this second case, only if the moved element is overt may its trace in IP-spec satisfy the Extended Projection Principle. Mikkelsen (2002) assumes the structure in (5) for the example in (3).

- (5) Vi kender de lingvister_i, [_{CP} OP_i [_{IP} *der* [_{VP} t_i vil læse denne bog]]]

The operator OP moves directly from its position in VP and expletive *der* is inserted to satisfy the EPP. In the example in (6), the EPP is satisfied differently.

- (6) Vi kender de lingvister, som vil læse denne bog.
We know the linguists, OP will read this book

For this example the structure in (7) is assumed.

- (7) Vi kender de lingvister_i, [_{CP} som_i [_{IP} t_i [_{VP} t_i vil læse denne bog]]]

The operator *som* moves via IP-spec and leaves a trace that satisfies the EPP.

The two analyses cover a wide range of Danish constructions involving extraction, and the short presentations here certainly do not do justice to them. However, what the two analyses have in common is that they do not refer outside the maximal projection of the CP containing the clause with the expletive *der*. Consequently, the analyses cannot predict the well/ill-formedness of the examples in (8), taken from Vikner (1991).

- (8) a Je kende mang lingviste der vil læs denn bo og so måsk vil synes den
I know many linguists ther wil read this boo an --- maybe wil like it
 b *Je kende mang lingvist der vil læs denn bo og der måsk vil synes den
I know many linguists ther wil read this boo an ther mayb wil like it

c Je kende mang lingviste so vil læs denn bo og so måsk vil synes den
I know many linguists --- wil read this boo an --- maybe wil like it

d *Je kende mang lingvist so vil læs denn bo og der måsk vil synes den
I know many linguists --- wil read this boo an ther mayb wil like it

The examples in (9) are also from Vikner (1991).

(9) a Je kend mang lingvist so der vil læs denn bo og so måsk vil synes de
I know many linguists --- ther wil rea this boo an --- mayb wil like it

b *Je kend man lingvist so der vil læs den bo og so der mås vil synes de
I know many linguist --- the wi rea this boo an --- the mayb wi like it

c Je kend man lingvist so vil læs den bog og so måsk vil synes de
I know many linguists --- wil rea this boo an --- mayb wil like it

d *Je kend mang lingvist so vil læs denn bo og so der måsk vil synes de
I know many linguists --- wil rea this boo an --- ther mayb wil like it

Neither of the analyses explains why the second conjunct in a coordinated relative clause cannot contain *der*. The two conjuncts have the same internal structure. In the next section, an analysis is proposed which accounts for this distribution of *der* by taking into account the different contexts in which the two conjuncts in the coordinated relative clauses occur.

3. THE DISTRIBUTION OF DANISH *DER* IN EMBEDDED CLAUSES

The distribution of *der* in embedded interrogative clauses and in relative clauses in standard Danish differs from its distribution in non-standard Danish. In non-standard Danish *der* insertion is more wide-spread, with varying degrees of acceptability, than in standard Danish. This paper will not have anything to say about why some non-standard occurrences of *der* are more acceptable than others, rather it will focus on what sets them apart as a group from standard occurrences. The examples in (10), repeated from (1), shows in which clauses *der* occurs in standard Danish, cf. Hansen (1974).

(10) a. Jeg ved ikke hvem der synger den.
I know not who there sings it

b. Der er intet signalement af den mand, der førte bilen.
There is no description of the man, who drove the car

- c. Alt hvad der næres vokser.
All what there is nourished grows

Der occurs in embedded interrogative clauses with local subject extraction, as in (10a), in bare subject relative clauses, as in (10b), and in free subject relative clauses, as in (10c).

On the other hand, we do not find *der* in subject relative clauses with the initial complementizer *som* or a relative pronoun, as in (11a) and (11b), and in non-local subject extractions as in (11c).

- (11) a. Hun mødte en mand, som gav hende en buket valmuer.
She met a man, --- gave her a bouquet of poppies
- b. Det handler om en mor, hvis søn er blevet voksen.
It is about a mother, whose son has become grown-up
- c. Tid til at ruste os til det næste opsving, som ingen ved, hvornår kommer.
Time to to prepare us for the next upturn, --- no one knows, when comes

The constraints which may be established for standard Danish on the basis of these examples are given in (12).

- (12) 1. If the expressed or unexpressed subject of a finite embedded clause does not appear in subject position, but inside its maximal projection, *der* obligatorily occurs in subject position.
2. If the expressed or unexpressed subject of a finite embedded clause does not appear in subject position, but outside its maximal projection, *der* is not permitted in subject position.

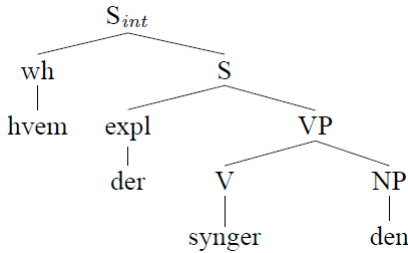
In non-standard Danish, the first constraint applies. However, the second constraint does not apply. Thus we find examples like (13a) with *der* in subject relative clauses with the initial complementizer *som*, or with a relative pronoun as in (13b), and in non-local subject extractions as in (13c).

- (13) a. ?Hun møder næste dag en mand som der giver hende en is.
She meets next day a man there gives her an ice cream
- b. ?Je fi ha af e venind hvis datter der desværre ikk kunn tåle ham
I go him fro a friend, whos daughte ther unfortunate not could tolerat him

c ?Je tra e fy so je bar ik ka huske hvo der boed (Hansen
I me a gu I jus no ca remembe wher ther lived

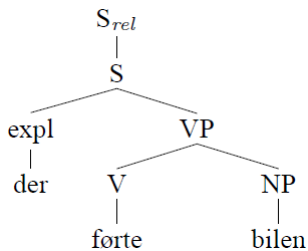
(14) through (16) illustrate the constraints in (12). In (14), an embedded interrogative clauses with local subject extraction, the expressed subject *hvem* is not in subject position, but still inside the maximal projection of the clause containing *der*.

(14) Jeg ved ikke



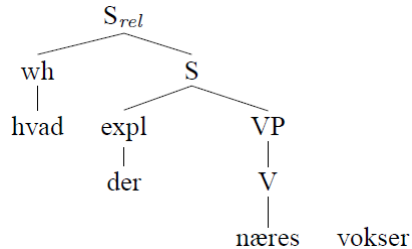
In (15), a bare subject relative clause, the unexpressed subject is again not in subject position. The unexpressed subject is a relative pronoun, the position of which, if expressed, would still be inside the maximal projection of the clause. In Danish subject relative pronouns are not expressed, cf. Hansen (1974).

(15) Der er intet signalement af den mand,



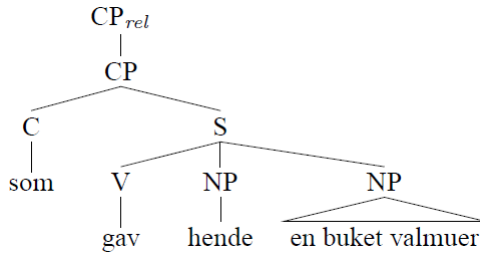
And finally, in (16), a free subject relative clause, the expressed subject *hvad* is not in subject position, but again, still inside the maximal projection of the clause.

(16) Alt



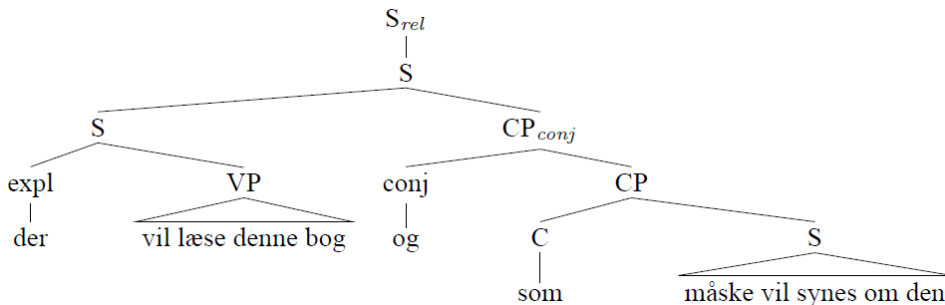
In contrast, if we look at a subject relative clause with an initial complementizer as in (17), we can see that the unexpressed subject is outside the maximal projection of the clause, and we do not get *der* insertion. Again subject relative pronouns are not expressed in Danish.

(17) Hun mødte en mand,



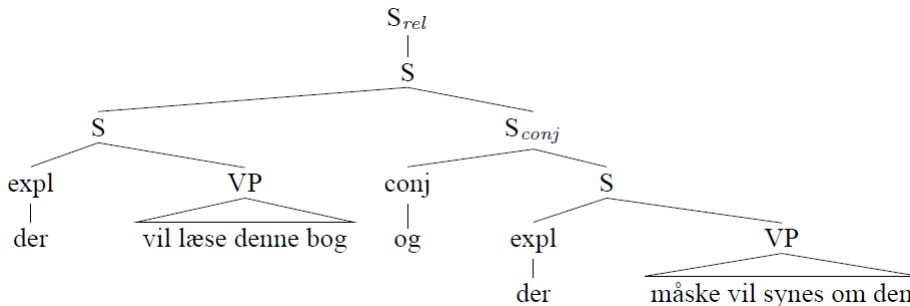
The constraints in (12) also apply to relative clauses in coordinations. In (18) the structure of the coordinated relative clause in (8a) is shown.

(18) Jeg kender mange lingvister



This example is fine, as we do not have *der* insertion. The coordinated relative clause in (8b), on the other hand, is not well-formed as can be seen in (19).

(19) *Jeg kender mange lingvister



The unexpressed subject is outside the maximal projection of the clause the second *der* is part of, S_{conj} , and we cannot have *der* insertion in this context.

Now, the present analysis predicts the judgements Vikner (1991) makes in (8) and (9). In our analysis, however, the occurrence of *der* in the second conjunct should be possible in non-standard Danish, since the second constraint above does not apply here. In fact, this also seems to be the case, as the examples in (20) show.

- (20) a ?I denn stor planlægningsproc skal der naturligv deltage studerend der har
I this big planning proces mus ther of course participat students, ther hav
 førstehåndskendsk til hva der ha funger i de nuværen studieordnin og der
firsthand knowledge of, what ther ha worked i the present curriculum, an ther
 har e fornemmels af hvilk ny undervisningstilta der ka funger i praksis
hav a feeling of, which ne teaching initiatives, ther can work i practice
- b ?Je vil gern anskaff mi såda e maskin evt. brugt e der noge der ved
I wil like acquire me such a machine, possibl used, is ther anyon ther know
 hvad den hedder og der har en liggende eller ved hvor man kan købe en?
what it is called and there has one lying or knows where one can buy one
- c. ?Med PrivatLeasing betaler du en fast månedlig ydelse, der ikke ændrer i
With PrivateLeasing pay you a fixed monthly sum, there not change during
 leasingperioden og som der dækker den reelle omkostning.
the leasing period and --- there covers the actual cost
- d. ?Vævet er opbygget af et svampagtigt netværk af bindevæv der har glatte
The is made up of a spongelike network of connective there has smooth
 muskelceller og som der har rigtig mange blodkar.
muscle cells and --- there has really many blood vessels

In (20) we find *der* in the second conjunct, both alone and following *som*, correctly predicted by the present analysis.

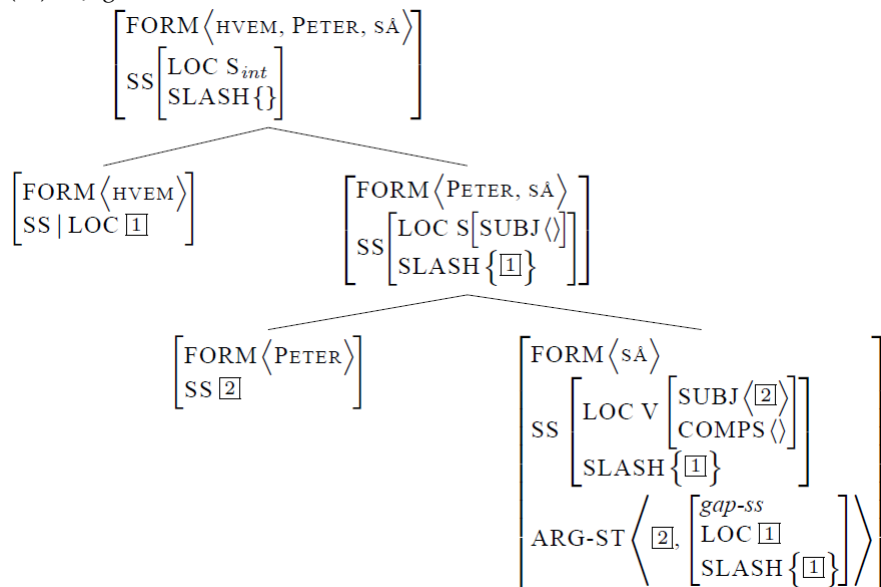
4. HPSG FORMALIZATION

Within the HPSG framework, constructions involving extractions are called filler-gap constructions. Filler-gap constructions are specified for the feature SLASH, and filler-gap

dependencies are established through the inheritance of SLASH specifications. A non-empty SLASH specification is introduced at the lowest level where the “gap” is introduced, then passed up through the structure, to be bound off by the filler.

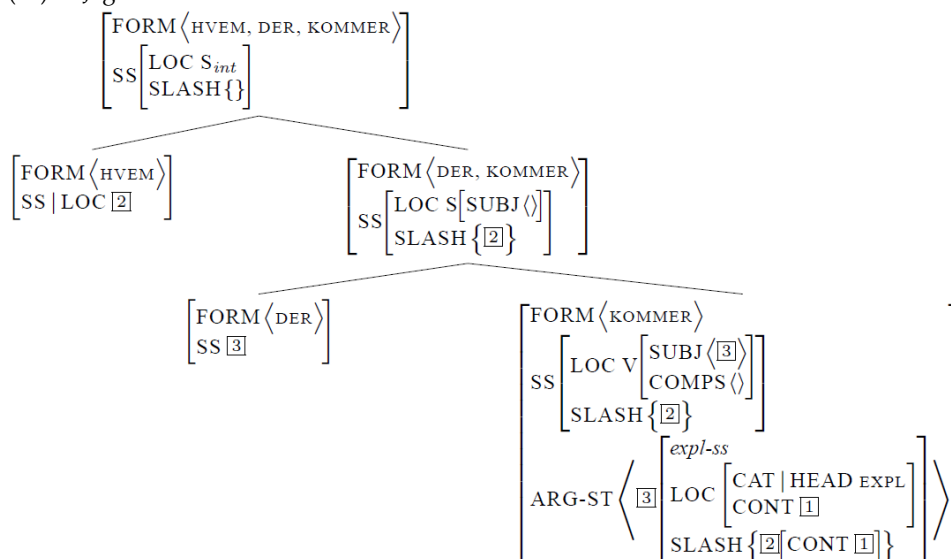
In (21) the specification for the SLASH feature in a filler-gap construction is illustrated.

(21) Jeg ved



Sometimes there is no overt filler. In the analysis presented here, *som* is treated as a complementizer, and consequently examples like (11a) does not involve an overt filler. In such cases the gap is bound off constructionally. This means that in certain construction types not involving a filler, the SLASH is nevertheless bound off. A relative clause is such a construction which may bind off a non-empty SLASH specification. Constructional gap-binding is illustrated in (22).

(23) Jeg ved



The constructions in (21), (22) and (23) are constrained by a set of principles and constraints. Firstly, the SLASH-Amalgamation Constraint, (Ginzburg and Sag, 2000, p. 169), which specifies the inheritance of SLASH values from the arguments of a word to the verb itself. Secondly, the Argument Realization Principle, (Ginzburg and Sag, 2000, p. 171), which maps the argument on the ARG-ST list to the valence lists. Thirdly, the Generalized Head Feature Principle, (Ginzburg and Sag, 2000, p. 33), which specifies inter alia the inheritance of the SLASH feature from the head-daughter to the mother in a construction. Finally, various constraints are responsible for binding off the SLASH value, either constraints involving a filler daughter or constraints involving constructional gap-binding.

To account for the locality of *der* we need to introduce a further constraint applying to Danish. To ensure that the SLASH value of an expletive does not “escape” the maximal projection of the clause it is introduced into so that it can be bound off outside this maximal projection, we formulate the constraint in (24).²

(24) The *Expletive SLASH* Constraint (Danish):

$$\neg \left[\begin{array}{l} \text{word} \\ \text{ARG-ST} \left\langle \left[\begin{array}{l} \text{LOC} | \text{CAT} | \text{HEAD} | \text{SUBJECT} \langle \text{expl-ss}_i \rangle \\ \text{SLASH} \{ \boxed{1}_i \} \uplus \boxed{\Sigma} \end{array} \right], \dots \right\rangle \end{array} \right]$$

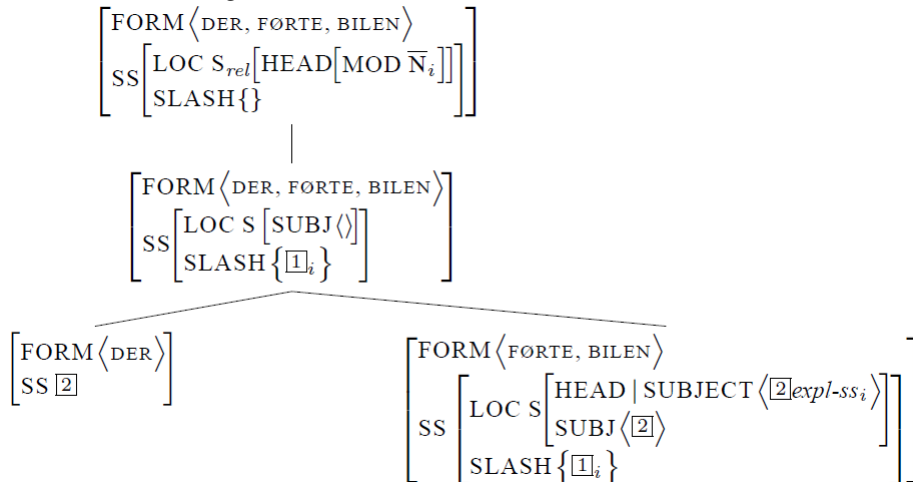
We mentioned earlier that the SLASH-Amalgamation Constraint specifies the inheritance of SLASH values from the arguments of a word to the verb itself. The *Expletive SLASH* Constraint makes

² Cf. Meurers (1999) for further arguments that we need a SUBJECT feature as part of the HEAD feature.

sure constructions with an expletive subject that is also in the SLASH set, is not allowed, thus preventing *der* in non-local dependencies.

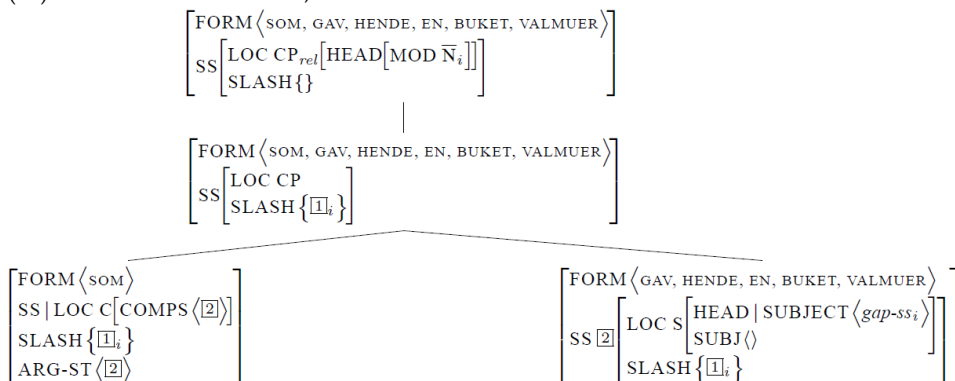
The bare relative clause in (25) meets the *Expletive SLASH Constraint*, as the SLASH value is bound off inside the maximal projection of the clause.

(25) Der er intet signalement af den mand,



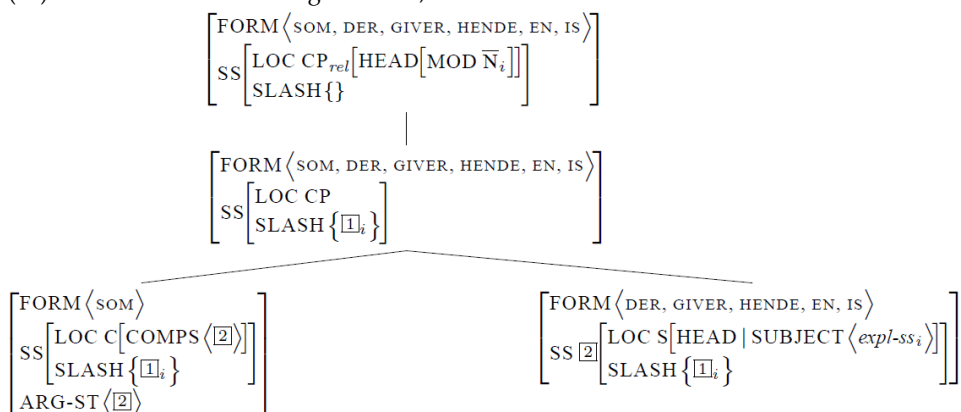
An example of a subject relative clause with an initial complementizer, as the example in (26), is not excluded by the *Expletive SLASH Constraint* either, as the complement of the complementizer does not have an expletive subject.

(26) Hun mødte en mand,



In non-standard Danish, the *Expletive SLASH Constraint* does not apply, and the a SLASH value corresponding to an expletive co-indexed with the SLASH value is not prevented from being inherited by the complementizer *som*. This is illustrated in (27).

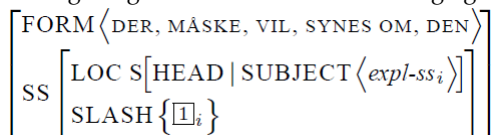
(27) Hun møder næste dag en mand,



In section 2 it was shown how the analyses in the GB tradition did not cover the distribution of *der* in coordinated relative clauses. Now this distribution is predicted by the *Expletive SLASH Constraint* and its prevention of the amalgamation by the conjunction of a SLASH value co-indexed with an expletive subject.

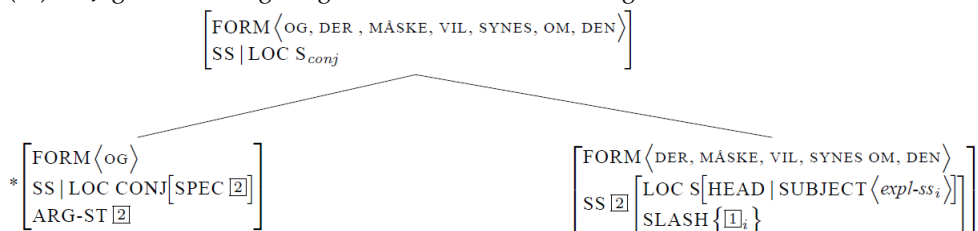
The unexpressed subject relative in the second conjunct in (8b) is outside the maximal projection of the clause containing *der*, i.e. the S_{conj} . (28) shows the representation of the S containing *der* in HPSG formalism.

(28) *Jeg kender mange lingvister der vil læse denne bog og



Because of the the *Expletive SLASH Constraint*, the conjunction cannot select this clause as its SLASH value is co-indexed with an expletive subject. The entire representation is shown in (29).

(29) *Jeg kender mange lingvister der vil læse denne bog



5 CONCLUSION

This paper presented a formal analysis of the distribution of the Danish *der* in embedded interrogatives and relative clauses within the framework of HPSG. The analysis relies on surface-oriented structures and constraint satisfaction rather than introducing movement and traces to account for extraction phenomena. The analysis is based on the SLASH feature, filler-gap and filler-expletive dependencies. The *Expletive SLASH Constraint* was established to account for the locality of *der* insertion. The analysis easily accounts for the different distributions in standard vs. non-standard Danish. The *Expletive SLASH Constraint* does not apply in non-standard Danish. It was shown that the present analysis can account for a distributional fact of *der* not previously accounted for, i.e. its exclusion in the second conjunct in relative clauses. The *Expletive SLASH Constraint* predicts this apparently exceptional non-occurrence of *der* in coordinated relative clauses in standard Danish.

REFERENCES

- Chomsky, Noam. 1981. Lectures on Government and Binding. Mouton de Gruyter.
- Diderichsen, Paul. 1962. Elementær Dansk Grammatik. København: Gyldendal.
- Erteschik-Shir, Nomi. 1985. Der. *Nordic Journal of Linguistics* 8(2), 131–148.
- Ginzburg, Jonathan and Sag, Ivan. 2000. *Interrogative Investigations: The Form, Meaning and Use of English Interrogatives*. Stanford: CSLI.
- Hansen, Erik. 1974. De nye *der*-konstruktioner. In *Festskrift til Kristian Hald*.
- Heltoft, Lars. 1986. Topologi og syntaks. En revision af Diderichsens sætningsskema. *Nydanske Studier og Almen Kommunikationsteori* 16-17, 105–130.
- Jacobsen, Bent and Jensen, Per-Anker. 1982. Some remarks on Danish weakly stressed *der*. Århus School of Business.
- Mikkelsen, Line. 2002. Expletive subjects in subject relative clauses. *Studies in Comparative Germanic Syntax. Proceedings from the 15th Workshop on Comparative Germanic Syntax*. John Benjamins Publishing Company. p. 69.
- Meurers, Detmar. 1999. Raising Spirits (and assigning them case). *Groninger Arbeiten zur Germanistischen Linguistik (GAGL)* 43, 173–226.
- Pollard, Carl and Sag, Ivan A. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.
- Sag, Ivan A. and Janet D. Fodor. 1994. Extraction Without Traces. In *Proceedings of the Thirteenth Annual Meeting of the West Coast Conference on Formal Linguistics*. Stanford: CSLI Publications. 365–384.
- Sag, Ivan A. 1997. English Relative Clause Constructions. *Journal of Linguistics* 33. 431–484.
- Sag, Ivan A. To appear. English Filler Gap Constructions. *Language*.
- Vikner, Sten. 1991. Relative *der* and other C0 elements in Danish. *Lingua* 84, 109–136.
- Wiwel, Hylling Georg. 1901. *Synspunkter for dansk sproglære*. E. Bojesen.

Pauli Brattico

Rethinking case assignment

ABSTRACT

It is frequently assumed that nominal case (nominative, accusative and so on) marks a local syntactic or semantic dependency between a predicate and its argument, while a separate case concord mechanism takes care of distributing the actual, attested morphological case suffixes among the adnominal elements and the noun head. This paper argues that these assumptions do not survive closer scrutiny. Case does not mark syntactic or semantic dependencies between predicates and their arguments, but is instead implied in agreement, aspect, polarity, transitivity, complementizers and tense; it is assigned to predicates as well; the dependency is neither local nor clause bound; and there being no concrete forms of abstract case, it is the theory of case concord which is responsible for almost all of the data concerning case. A hypothesis is put forward according to which there is no case assignment of abstract case; instead, case expresses a head-to-head relation, produces case concord as a side effect, and has no functional motivation.

1 CASE, ARGUMENTS AND PREDICATES

This article deals with nominal case. I begin by exploring the following hypothesis:

- (1) *Case and arguments*
Nominal case marks syntactic/semantic dependencies between thematic arguments and their predicates.

Generalization (1) can be found in various forms from a wide variety of grammatical theories. It constitutes the foundational block of theories which assume that case represents semantic notions such as agent/patient or figure/background, and theories which assume that case marks argument DPs (determiner phrases or noun phrases) for their syntactic dependencies. Some theories take (1) to constitute an axiom that characterizes the link between case forms and semantics. To illustrate, consider the following example:¹

¹The following abbreviations will be used in this article: A = the A-infinitival, one non-finite verb form in Finnish; ACC(0) = the nominative-looking zero-accusative case; ACC(n) = the genitive-looking n-accusative case; ACC = accusative case, any variant; ELA = elative case; INE = inessive; MA = MA-adjunct (one adjunct type in Finnish); PASS = impersonal passive; PRT = partitive case.

- (2) Pekka innostui sopimuksesta
 Pekka.NOM got.excited agreement.ELA
 'Pekka got excited about the agreement.'

The verb (or some structure therein) is responsible for assigning the elative case (glossed as ELA) to its thematic argument, in agreement with (1). Other cases are felt as ungrammatical (**Pekka innostui sopimuksen/sopimusta*, 'Pekka got.excited agreement.ACC/agreement.PRT'). Upon detecting the elative case a hearer can grasp the intended meaning. Without case, there would be much ambiguity. Now consider (3), which is a modification of (2).

- (3) Pekka innostui hyvästä sopimuksesta
 Pekka got.excited good.ELA agreement.ELA
 'Pekka got excited about the good agreement.'

What is marked by the elative is not the argument of the verb *innostui* 'got excited', but the individual words inside the argument. These words do not, as individual elements, constitute thematic arguments of the predicate. Depending on how (1) is interpreted, either the principle predicts this result wrongly (case marking on something that is not an argument, such as on the adjective) or it remains silent about (3)(argument-internal case marking being irrelevant to (1)). At any rate, we need to do something with (1) to get the facts right.

At first, it occurs that we could explain these facts away by requiring every word inside the case-marked argument to be marked by case (a case which would, in turn, be assigned to the argument by principle (1)). But this is too simple. Consider (4).

- (4) Pekka innostui hyvästä sopimuksesta myydä talo
 Pekka got.excited good.ELA agreement.ELA sell.A house.ACC(0)
 'Pekka got excited about the good offer to sell the house'.

It is not the case that all words inside the argument 'good agreement to buy the house' are marked by the elative case. Specifically, the non-finite verb *myydä* 'to sell' and its complement *talo* 'house' are not. In addition, the adjective marked by the elative in (4) is not just an adjective *word*, but a whole adjective *phrase* (bracketed in the example below):

- (5) Pekka innostui [tosi hyvästä] sopimuksesta myydä talo
 Pekka got.excited really good.ELA agreement.ELA buyl.A house.ACC(0)
 'Pekka got excited about the really good agreement to buy a house.'

If there is a rule which distributes the elative case to the elements inside the argument, that rule must be able to deal with whole phrases, such as the adjective phrase *tosi hyvästä* 'really good.ELA', not only with individual words. In many ways, the hypothetical case concord rule would then be like (1), which is also required to bestow case features from predicates to whole

phrases, namely to the arguments. Returning to the marking on the adjective phrase, what is case-marked by the matrix verb is the head of this phrase. Other arguments inside the adjective phrase (participle phrase in example (6)) are not case-marked by the matrix verb. In the following example, the adjective head case-marks its own argument by the accusative, while it itself receives the partitive from the matrix verb.

- (6) Pekka etsi leivän syönyttä koiraa
 Pekka searched bread.ACC eaten.PRT dog.PRT
 'Pekka searched the dog who ate the bread.'

There are constructions in which not even the elements of the projectional spine of the noun head get marked by the external case assigned by the matrix verb. In the following example, the demonstrative bears no overt trace of the object case, the adjective is marked by the accusative (by the transitive matrix verb), the numeral is in singular and without any case marking, while the noun head is in singular and is assigned the *partitive*:

- (7) Minä ostin ne hienot kolme taulua
 I bought those nice.ACC.PL three.SG painting.SG.PRT
 'I bought those nice three paintings.'

There is only *one* element which shows case marking that can be related to the matrix verb, the adjective, the rest shows something else, but the adjective is not an argument of anything that could be recognized as a predicate. The partitive case of the noun head is related to the numeral, but these do not form an argument-predicate pair either. This is all gibberish if we look at it through the prism of (1).

It is not true that only arguments are marked for case. Also predicates may be assigned case:

- (8) Me olemme kielitieteilijöitä
 We are linguist.PL.PRT
 'We are linguists.'

Hypothesis (1) forces us to say that the predicate 'linguists' is the thematic argument of the auxiliary copula.² This conclusion strikes me as dubious once we recognize that the same properties are attested if the predicate is an adjective, as in *Me olemme viisaita* 'we are wise.PRT'. And the whole point surely collapses because the partitive is actually a reflex of the plural

²This assertion depends on how we intend to read (1), more specifically, what we mean by the terms "predicate" and "argument". Assuming predicates must be saturated by arguments in order to create propositions, it follows that '(to be a) linguistic' is a predicate in (8). An alternative is to analyse 'linguistics' as the argument of the copula.

feature at the auxiliary. Thus we say *Minä olen kielitieteilijä* 'I am linguist.SG.NOM' and not **Minä olen kielitieteilijää* 'I am linguists.SG.PRT'.³

It is well-known that the Finnish accusative-partitive alteration is related, not to the dependency between an argument and its predicate, but to the aspectual properties of the whole event described by the verb phrase (Vainikka 1989, Kiparsky 1998). This is shown in (9).

- (9) Minä ammuin karhua/ karhun
 I shot bear.PRT bear.ACC
 'I shot at the bear/the bear (dead).'

Principle (1) remains silent concerning both aspect and verb phrases. It is also known that the partitive case is forced by a governing matrix negation:

- (10) Minä en ampunut karhua/ *karhun
 I not shoot bear.PRT bear.ACC
 'I did not shoot at the bear/the bear dead.'

Principle (1) remains silent concerning polarity. Then there are the two accusative variants, the n-accusative ACC(n) and the 0-accusative ACC(0), which are controlled by *agreement* on the matrix verb (Vainikka & Brattico 2009; see also Nelson 1998, Reime 1993, Toiconen 1995):

- (11)
 a. Me löysimme talon/ *talo
 we.NOM found.1PL house.ACC(n) house.ACC(0)
 'We found the house,'
 b. Me löydettiin *talon/ talo
 we.NOM found.PASS house.ACC house.ACC(0)
 'We found the house.'

Principle (1) has nothing to offer. The negation bears no argument-type relation with the direct object which it partitivizes. Agreement on a verb bears no argument-type relation to the direct object whose accusative form it controls; it is the verb which bears this relation, not *agreement on the verb*. But it is agreement that matters for case assignment.

Both the 'partitive of negation' and the 'accusative alteration of agreement' span over a longer distance than what is predicted by the hypothesis that the dependencies rely on argument-predicate pairs (Brattico 2009, 2010a, Vainikka & Brattico 2009, Toivonen 1995). In the following example, agreement on a matrix predicate affects the object case realization inside a remote adjunct phrase (an adverb phrase).

³A reviewer points out that the partitive form may be used in special contexts of 'pretension'. This refers to sentences such as *Sitä ollaan niin kielitieteilijää* 'EXPL be.PASS so linguist.SG.PRT', meaning perhaps something like 'you try so much to be/to present oneself as a linguist'.

(12)

- a. Lapset oppivat läksyt lukemalla
Children learned.1PL homework read.MA
*kirja/ kirjan
book.ACC(0) book.ACC(n)
'Children learned their homework by reading a/the book.'
- b. Me opittiin läksyt lukemalla
we.NOM learned.PASS homework read.MA
kirja/ *kirjan
book.ACC(0) book.ACC(n)
'We learned the homework by reading a/the book.'

The accusative forms in (12a-b) do not reflect a relation between an argument and its predicate, but a relation between *agreement* on a matrix predicate *oppia* 'learn' and an argument of another predicate *lukea* 'read' such that the latter is contained in an adjunct phrase. Similar facts can be demonstrated for other types of constructions, and for the partitive of negation (Brattico 2009, 2010b).

It is possible for one case assigner to assign case to several elements, to witness:

(13)

- a. Pekka näki Merjan syömässä leivän
Pekka saw Merja.ACC(n) eat.M bread.ACC(n)
'Pekka saw Merja eating the bread.'
- b. *Pekka ei nähnyt Merjan syömässä leivän
Pekka not saw Merja.ACC(n) eat.MA bread.ACC(n)
'Pekka did not see Merja eating the bread.'
- c. Pekka ei nähnyt Merjaa syömässä leipää
Pekka not saw Merja.PRT eat.MA bread.PRT
'Pekka did not saw Merja eating the bread.'
- d. Me nähtiin Merja syömässä leipä
We saw.PASS Merja.ACC(0) eat.MA bread.ACC(0)
'We saw Merja eating the bread.'
- e. *Me nähtiin Merjan syömässä leivän
We saw.PASS Merja.ACC(n) eat.MA bread.ACC(n)
'We saw Merja eating the bread.'

The predicate-argument pairs do not form a coherent system here: it is again the agreement and polarity properties of the matrix clause which affect case assignment of the direct object of an infinitival predicate. Besides, also whole adjuncts can be case-marked in Finnish:

- (14) Minä odotin koko päivän
I waited whole day.ACC(n)
'I waited the whole day.'

Finally, there are languages in which a given noun can show several case suffixes, some of them encoding aspect, tense and mood/force of the clause (Dench & Evans 1988, Nordlinger 1998).

These remarks provide a representative sample of data that I think are difficult to capture by (1). To delineate the argument a bit further, let's formalize (1). There are a number of ways to make it explicit. One particularly clear way is (15). This is what I think (1) 'says'.

(15) *Case and arguments (second version)*

An element X bears case F if and only if there is a syntactic or semantic relation (corresponding to F) between the argument X and its predicate P.

This condition is falsified into both directions. Consider the *only if*-part first. There were numerous examples of case marking that express something else than argumenthood: partitive by negation, accusative alteration, aspectual case, partitive by numeral, case at predicates and adjuncts, case at adjectives and other modifiers, and the long-distance case assignment which may mark even several assignees. Also the *if*-part fails. The *if*-part says that once we establish that X-P constitutes an argument(X)-predicate(P) pair by relation F, then there should be case marking F at X. But whole noun phrases/determiner phrases in Finnish are never case-marked.

2 CASE ASSIGNMENT AND CASE CONCORD

Linguists have visioned number of ways to sustain (1)/(15). A common strategy is to assume that although whole arguments are not always (or never) marked for case, in fact they *are* marked for case, namely, they are marked for *abstract case*. Abstract case is assigned to arguments. The reason why words inside the relevant noun phrases show case marking is because of *case concord*, or perhaps because 'morphological case' is dissociated from case assignment. Case concord of course does not need to fall under (1).

If case concord does not fall under (1), what *does* it fall under? There is surprisingly little explicit theorizing on this notion, which is quite striking in the light of the fact that it is assumed to handle complex data and that it is concerned with the overtly visible aspect of the phenomenon. Nevertheless, it often looks as if it is assumed that case concord were some sort of epiphenomenon of phi-concord (concord in number, gender and person). But it can't be a byproduct of phi-concord. In example (16), for instance, case concord is complete while the number feature changes in the middle of the noun phrase:

(16) niissä kalliissa kolmessa autossa
those.PL.INE expensive.PL.INE three.SG.INE car.SG.INE
'in those there expensive cars'

The opposite is observed in (17), where everything is marked for singular but case concord is not complete.

- (17) Minä odotin sen pitkästyttävän puoli minuuttia
 I waited that.SG.ACC boring.SG.ACC half.SG minute.SG.PRT
 'I waited that boring half a minute.'

Case concord and phi-concord are not the same thing. One possibility is to say that whereas abstract case marks argument phrases, case concord marks words/morphemes. This cannot be true, however, because it is mandatory for the putative case concord rule to handle also adjective *phrases*. It must also avoid marking several elements, such as complements of the noun head. It is the theory of case concord which is assumed to deduce almost *all of the data*.

The simplest remedy to these concerns is to take the data at its face value and assume that case assignment is a *head-to-head relation*. That is, instead of requiring predicates like verbs to assign case to their arguments, we claim that predicates and other case assigners directly case-mark words inside their arguments. Such case assignment represents a "head-to-head" relation, because it obtains between grammatical heads, not between heads and phrases (Brattico 2008, 2010b, Kayne 2005: Ch. 5). Taking this perspective, the data reviewed here shows that case assigners are functional-grammatical elements representing a variety of notions, such as phi-agreement, polarity, transitivity, aspect, prepositions, and semantic features of all sorts. The data also demonstrate that case assignment need not be clause-bound, but it is always downstream. Case assignment therefore resembles grammatical dependencies well-known from previous literature: it obtains between a functional element and another element it c-commands. It does not have any coherent "functional" or semantic motivation.

To see how the system works, let us conclude by revisiting some of the data. In example (2), repeated here as (2'), the predicate case marks the (sole) word inside its argument:

- (2') Pekka innostui^P sopimuksesta^G
 Pekka.NOM got.excited agreement.ELA
 'Pekka got excited about the agreement.'

Case assigners are marked by P, case assignees by G (from "probe" and "goal", respectively). If there are several words, then the predicate case-marks each of them, independently:

- (3) Pekka innostui^P hyvästä^G sopimuksesta^G
 Pekka got.excited good.ELA agreement.ELA
 'Pekka got excited about the good agreement.'

In example (4), repeated here as (4'), the main verb marks the adjective and the head noun, but the infinitival verb case marks its *own* object.

- (4') Pekka innostui^{P,1} hyvästä^{G,1} sopimuksesta^{G,1}
 Pekka got.excited good.ELA agreement.ELA
 myydä^{P,2} talo^{G,2}
 sell.A house.ACC(0)
 'Pekka got excited about the good offer to sell the house'.

This example illustrates another important principle regulating case assignment, that of *locality*. It is the most local case assigner which will case-mark a given set of assignees. Thus, the adjective and the noun head are case-marked by the finite matrix verb, while the direct object of the non-finite verb is case-marked by the non-finite verb. Therefore, it does not show elative case. Example (6) illustrates the same principle. The object *luun* 'bone' is case-marked by the participle head, which is the most local case assigner. The head itself, and the noun head, are case-marked by the finite matrix verb.

- (6) Pekka etsi^{P,1} luun^{G,2} syönyttä^{P,2/G,1} koiraa^{G,1}
 Pekka searched bone.ACC eaten.PRT dog.PRT
 'Pekka searched the dog who ate the bone.'

One of the most complex specimens documented so far is (7), but it, too, falls under the proposed locality system. The finite verb case-marks the two adnominals, the demonstrative and the adjective, while the numeral case-marks the noun head (7').

- (7') Minä ostin^{P,1} ne^{G,1} hienot^{G,1} kolme^{P,2} taulua^{G,2}
 I bought those nice.ACC.PL three.SG painting.SG.PRT
 'I bought those nice three paintings.'

In Finnish, the accusative/partitive alteration signals aspect. To handle this data it suffices to assume that the verb phrase contains a grammaticalized aspectual feature which functions as a case assigner.

- (9') Minä ammuin (Asp^P) karhua/ karhun^G
 I shot bear.PRT bear.ACC
 'I shot at the bear/the bear (dead).'

The same explanation applies to negation, which imposes the partitive to direct objects irrespective of the aspect feature (cf. 10). Therefore, the negation and other polarity features function as case assigners in Finnish. Example (11, 11') shows that agreement on finite verb changes the direct object case, and that the effect travels over a distance (12, 12'). The phenomenon remains a mystery under (1), but if case marks a relation between functional heads and nominal heads, there is nothing surprising about (11): agreement is a hallmark of functional structure.

- (11')
 a. Me löysimme^P talon^G / *talo
 we.NOM found.1PL house.ACC(n) house.ACC(0)
 'We found the house.'

b. Me löydettiin^P *talon/ talo^G
 we.NOM found.PASS house.ACC house.ACC(0)
 'We found the house.'

(12')

a. Lapset oppivat^P läksyt lukemalla
 Children learned.1PL homework read.MA
 *kirja/ kirjan^G
 book.ACC(0) book.ACC(n)

'Children learned their homework by reading a/the book.'

b. Me opittiin^P läksyt lukemalla
 we.NOM learned.PASS homework read.MA
 kirja^G/ *kirjan
 book.ACC(0) book.ACC(n)

'We learned the homework by reading a/the book.'

Arguably, it is more challenging to explain how long-distance case assignment is possible provided that most of the data follows a locality principle. On the other hand, by assuming that remote case assignment is possible it becomes easy to explain how one case assigner can mark several assignees over a distance, to witness:

(13')

a. Pekka näki^P Merjan^G syömässä leivän^G
 Pekka saw Merja.ACC(n) eat.MA bread.ACC(n)

'Pekka saw Merja eating the bread.'

b. *Pekka ei^P nähnyt Merjan^G syömässä leivän^G
 Pekka not saw Merja.ACC(n) eat.MA bread.ACC(n)

'Pekka did not see Merja eating the bread.'

c. Pekka ei^P nähnyt Merjaa^G syömässä leipää^G
 Pekka not saw Merja.PRT eat.MA bread.PRT

'Pekka did not see Merja eating the bread.'

d. Me nähtiin^P Merja^G syömässä leipä^G
 We saw.PASS Merja.ACC(0) eat.MA bread.ACC(0)

'We saw Merja eating the bread.'

This phenomenon is thus analogous to (2-2'), in which the matrix verb case-marks the two (or more) nominal elements: the adjective and the noun head. A useful working hypothesis is therefore one which says that a given case assigner will mark all potential case assignees until a more local and relevant case assigner intervenes. All in all, I propose the following locality-constrained head-to-head mechanism:

(18) *Case assignment*

Nominal case marks syntactic head-to-head dependencies between nominal heads and the closest c-commanding functional heads.

Principle (18) remains a guiding principle, requiring much further work that I cannot attempt here, but see Brattico (2011) for a possible continuation.

3 CONCLUSIONS

The belief that case marks arguments for their predicates remains a widely held assumption in linguistics. However, in Finnish all substantial facts concerning case involve "word marking", and virtually no facts involve abstract case and argument marking. Abstract case is nowhere to be seen. The data documented here requires us to submit two claims. First, instead of marking argument phrases, case assigners mark individual words inside those arguments and other phrases. Second, we need to posit a locality principle.

REFERENCES

- Brattico, P. 2008. Kayne's model of Case and Finnish nominal phrases. *Nordic Journal of Linguistics* 31: 135-160.
- Brattico, P. 2009. Long-Distance Case Assignment and the Theory of Phases. *Biolinguistica Fennica Working Papers* 1: 79-104.
- Brattico, P. 2010a. Long-distance case assignment in Finnish. University of Jyväskylä, ms.
- Brattico, P. 2010b. The Two-Part Models and One-Part Models of Nominal Case: Evidence from Case Distribution. *Journal of Linguistics* 45: 47-81.
- Brattico, P. 2011. Case assignment, case concord, and the quantificational case construction. *Lingua* 121: 1042-1066.
- Dench, A. & Evans, N. 1988. Multiple case-marking in Australian languages. *Australian Journal of Linguistics* 8: 1-47.
- Kayne, R. 2005. *Movement and Silence*. Oxford: Oxford University Press.
- Kiparsky, P. 1998. Partitive Case and Aspect. In Butt, M. & Geuder, W. (Eds.) *The Projection of arguments: Lexical and Compositional Factors*. Stanford: CSLI Publications, 265-307
- Nelson, D. 1998. *Grammatical case assignment in Finnish*. London: Garland.
- Nordlinger, R. 1998. *Constructive Case: Evidence from Australian Languages*. Stanford: CSLI Publications.
- Reime, H. 1993. Accusative marking in Finnish. In A. Holmberg & U. Nikanne (ed.), *Case and Other Functional Categories in Finnish Syntax*, 89-109, Berlin: Mouton.
- Toivonen, I. 1995. A study of Finnish infinitives. Brandeis University, unpublished Master's thesis.
- Vainikka, A. 1989. *Deriving syntactic representations in Finnish*. University of Massachusetts Amherst, PhD dissertation.
- Vainikka, A. & Brattico, P. 2009. The Finnish Accusative: Long Distance Case Assignment by phi-Agreement. University of Jyväskylä, ms.

Michael Chiou

Emphatic reflexives and logophoric marking in Modern Greek

*Evidence from parliamentary discourse. A pragmatic
analysis*

1 SETTING THE SCENE

Logophoric marking or logophoricity refers to the phenomenon whereby the 'perspective' of an internal protagonist of a sentence or discourse, as opposed to that of the current, external speaker, is being reported. Logophoricity and the use of logophoric pronouns were initially observed in a number of African languages such as Ewe, Dogon, Tuburi and so on. In these languages there is a separate class of pronouns dedicated to the encoding of logophoric interpretations. Nevertheless, apart from the purpose specific logophoric pronouns, other anaphoric expressions, such as long-distance anaphors for instance, can also be used in order to mark a logophoric interpretation under certain conditions. So far, there has not been any systematic account of logophoricity in Modern Greek. However, based on Kuno's (1987) theorizing that the contrast between an anaphor and a zero pronoun, where there is a free choice, is semantic/pragmatic in nature and it is associated with the notion of logophoricity, it will be argued that anaphor *o iðjos* (the same) when used instead of a zero pronoun encodes logophoricity. Data from parliamentary discourse indicate that *o iðjos*, along with its emphatic reading, is also used to encode logophoric interpretations when preferred over a zero pronoun. In this paper, I put forward that the logophoric reading of the emphatic anaphor *o iðjos* can be accounted for in terms of the systematic interaction of the neo-Gricean pragmatic principles (Huang 2000, Levinson 2000).

2 LOGOPHORICITY AND THE NOTION OF POINT OF VIEW

Originally, the notion of logophoricity was introduced by Hagège (1974) to refer to special pronominal forms found in West African languages which show that an argument of a subordinate verb is co-referential with the speaker or 'source' argument of a superordinate verb

of speech or thought. These African languages include Ewe, Dogon, Tuburi, Aghem among others (see Huang 2000 for a variety of examples). In these languages, which are called ‘full logophoric’, there is a separate paradigm of logophoric pronouns, i.e. a class of pronouns dedicated to the encoding of logophoric interpretations. Nevertheless, there is another way of logophoric marking apart from the purpose-specific logophoric pronouns. More precisely, some languages extend the use of long-distance anaphors so that they express logophoric contexts under certain conditions (see Clements 1975, Culy 1994, 1997, Huang 1991, 1994, 2000, Sells 1987, Zribi-Hertz 1989). For our purposes I am going to focus on the latter case.

In the literature there are various definitions and descriptions of the notion of logophoricity. According to Culy (1997: 845), “logophoric pronouns are usually described as pronouns that are used to refer to the person whose speech, thoughts, or feelings are reported or reflected in a given linguistic context”. This ‘person’ is also referred to as the ‘internal protagonist’ (Huang 2000) or the ‘minimal subject of consciousness’ (Zribi-Hertz 1989). In particular, Zribi-Hertz (1989) identifies the subject of consciousness with Kuno’s (1987) sense of logophoricity as “a semantic property assigned to a referent whose thoughts or feelings, optionally expressed in speech, are conveyed by a portion of the discourse” (Zribi-Hertz 1989: 711). Logophoricity is also related to the notion of ‘point of view’, yet Culy (1997) claims that logophoricity proper is rather distinct from point of view. More precisely, Culy points out that “morphologically distinct logophoric pronouns are grammatically licensed in indirect discourse...and only secondarily indicate point of view” (Culy 1997: 846). In a similar fashion, ‘indirect reflexives’ (reflexives which can be used logophorically) “can express point of view if they do not have grammatically determined antecedents” (Culy 1997: 856).

In a study on logophoricity based on long-distance reflexives in Icelandic and Japanese, Sells (1987) argues that there is no unified notion of logophoricity but logophoric phenomena are clusters of three primitive discourse-semantic notions, namely, source (the one who makes the report), self (one whose internal feeling is being reported) and pivot (point of view) Sells (1987: 445). Based on this tripartite division, Sells goes on to define four discourse environments which follow from the combination of these three semantic notions. Among these four environments there is one that also involves logophoric verbs. This, according to Sells, is the prototypical logophoric context. In that case, the internal protagonist carries the three semantic roles, namely the self, source and pivot. Let us illustrate with an example from Modern Greek:

- (1) O Janis lei oti o idjos ine eksipnos
 the John says that the same is clever
 ‘John says that he is clever.’

First of all, in (1) John is the source of the report. Moreover, he is the self since the fact that he is clever is established by him and finally, John is the pivot as the point of view of the sentence is the same with John’s point of view who is the internal protagonist.

What is more, there is cross-linguistic evidence (see Huang 2000, Stirling 1993) that certain verbs allow the logophoric interpretation of long-distance reflexives in a higher degree than others. Thus there is a kind of implicational universal for these types of verbs which is formulated as follows:

- (2) Universal for logocentric predicates, (Huang, 2000: 185)
 Speech predicates > epistemic predicates > psychological predicates > knowledge predicates > perceptive predicates.

The basic rationale behind this hierarchy is that each language allows a certain type of logophoric predicates and all the predicates higher in the hierarchy. So for instance, if psychological predicates are possible then both epistemic and speech predicates are equally allowed.

Coming now to Modern Greek, the study of notions such as point of view or logophoricity has not received a lot of attention in the literature. It is clear that Modern Greek does not possess separate classes of pronouns dedicated to the encoding of logophoric interpretations like the ones found in the West African languages mentioned above. Therefore, it remains to be shown whether long-distance anaphors in Modern Greek can be used in order to express logophoric aspects of meaning.

3 SOME BASIC FACTS

Modern Greek is a typical pro-drop language and as a result, it normally drops the overt subject¹ of finite clauses (pro-drop parameter)². Consider the following examples:

- (3) O Janisi ipe oti Øi tha djavazi perisotero
 the John said that (he) will study more
 'John said that he will study more.'
- (4) Ø pai kaθe mera sto sxolio
 (s/he) go-3sg every day to the school
 'She/he goes to school every day.'

Nevertheless, since pro-dropping is only a general tendency, it means that an overt phrase can equally occupy the relevant slot in the clause. Among all the potential candidates for filling in this slot I shall focus on the anaphor *o idjos* and more precisely on the distribution: {...zero / *o idjos*...}. By way of illustration, consider the following example:

- (5) Janisi nomizi oti Øi / o idjosi ine kalos maθitis
 the John thinks that (he) / the same is good student
 'John thinks that he/himself is a good student.'

O idjos is a long-distance anaphor and it can occur both in subject and object positions. In contrast to personal pronouns, anaphor *o idjos* cannot have independent reference but it always

¹ Modern Greek drops only subjects but not objects.

² The zero subject of finite clauses is known under the term *pro*, hence the pro-drop parameter.

needs to be bound by an antecedent in the sentence or in discourse. More importantly, *o idjos*, when in subject positions, systematically overlaps in reference with zero pronouns as example (3) illustrates. Finally, it is observed that it usually follows knowledge, psychological, epistemic and speech predicates (i.e. logocentric predicates) as it is illustrated in examples (6) – (17) respectively.

- (6) O Janis kseri oti o idjos ine o kaliteros maθitis
 the John knows that the same is the cleverest student
 ‘John knows that he is cleverest student.’
- (7) O Janis fovate oti o idjos θa plirosi ja oti eyine
 the John is afraid that the same will pay for what happened
 ‘John is afraid that he will pay for everything.’
- (8) O Janis nomizi oti o idjos ine o kaliteros maθitis
 the John thinks that the same is the cleverest student
 ‘John thinks that he is cleverest student.’
- (9) O Janis lei oti o idjos ine o kaliteros maθitis
 the John says that the same is the cleverest student
 ‘John says that he is cleverest student.’

On the basis of these facts, there is a twofold question which should be addressed: First, what could be the reason for using the anaphor *o idjos* where a non-morphologically expressed pronoun can be used, and second, given the option between a zero and the anaphor *o idjos*, what are the principles which condition their interpretation?

With regard to the first question, as Kuno (1987) and Kuno & Kaburaki (1977) note, the contrast between a pronoun and an anaphor, where there is a free choice, is semantic/pragmatic in nature and it is associated with the notion of ‘point of view’. The distribution {...zero / *o idjos* ...} seems to fit the above description in the sense that both the zero and *o idjos* are in free variation in these contexts. It remains to be examined whether there is also a difference in their intended use and interpretation related to the notion of ‘point of view’ in the sense of Kuno (1987) and Kuno & Kaburaki (1977). Consider the example:

- (10) O Janisi ipe oti Ø_i / o idjos_i den γnorize tipota ja tin ipoθesi
 the John said that (he) / the same not knew nothing about the case
 ‘John says that he knows nothing about the case.’

To start with, there is intuitive evidence that the use of *o idjos*, apart from emphasis or contrast, also encodes some kind of pragmatic meaning associated with the notion of ‘point of view’ as in Kuno (1987) and Kuno & Kaburaki (1977). In an example like (10), this pragmatic meaning can be spelled out as follows: When the null pronoun is used we have a mere report of what *John* said. By contrast, when *o idjos* is used, the current speaker is reporting on the internal protagonist’s ‘point of view’ without taking responsibility for what she/he is reporting. Therefore, as I understand, in the latter case there is a distinction between the perspective of the external speaker as opposed to that of the internal protagonist of the sentence.

4 THE DATA

So far, based on some preliminary data, there is evidence that the anaphor *o idios*, when used instead of a zero pronoun, can be associated with the notion of logophoricity or point of view in the sense of Kuno (1987) and Kuno & Kaburaki (1977). Nevertheless, there is need for some stronger evidence based on solid data in order to further support our claim. A way forward is to find data in which the speaker will tend to show logophoric intention in discourse. In other words, we need discourse data in which the difference in point of view between the speaker and the 'internal protagonist' would be as clear as possible. Given this, I am going to investigate parliamentary discourse data³. In this genre, a clearer distinction between the speaker and the 'internal protagonist' in discourse can be found since, in the majority of cases, the speaker and the internal protagonist belong to different political parties and as a result, they do not share beliefs and ideas, i.e. they have different points of view. Following this rationale, it is predicted that in certain cases the anaphor *o idios* will be preferred over the zero pronoun in order to promote a more logophoric reading. Moreover, due to the nature of parliamentary discourse, it may also be expected that the preference of *o idios* over zero will tend to be more frequent than in normal every day conversations. A first examination of a corpus of parliamentary discourse data seems to confirm our initial hypothesis. As it is illustrated in the examples below, the use of the anaphor *o idios* instead of a zero pronoun appears in several cases. For the purposes of this paper I am going to focus on some representative examples like the following:

- (11) Τα μέτρα τα οποία ελήφθησαν εχθές, μέτρα τα οποία είναι πρωτοφανή, βρίσκονται σε πλήρη αναντιστοιχία με αυτά που ο Πρωθυπουργός είχε επαγγελλθεί προεκλογικά, αλλά και σε πλήρη αναντιστοιχία με όσα έλεγε **ο ίδιος** μετά από αυτό εδώ το Βήμα...
- The measures that were taken yesterday, which are unprecedented, are not in accordance with what the Prime Minister had promised before the elections, and totally opposite to what **he himself** was saying here later on...
- (12) Τότε ο κ. Παπανδρέου, αφού δημιουργούσε τέτοιο κλίμα και αφού **ο ίδιος** ζητούσε εκλογές για να δημιουργήσει προεκλογικό κλίμα, είπε...
- Then Mr Papandreou, since he was generating this climate and since **he himself** was asking for elections in order to create a pre-elections atmosphere, he said...
- (13) Θέλω να πω ότι προσπαθείτε να κάνετε μια αύξηση εσόδων 1.000.000.000 ευρώ και **ο ίδιος** μάς ομολογήσατε ότι η έμμεση φορολογία επηρεάζει όλους.
- I want to say that you try to increased incomes by 1.000.000.000 Euros and **you** accepted that indirect tax affects everyone.
- (14) Όταν αναλάβαμε σαν χώρα τους Μεσογειακούς Αγώνες, κεφαλαιοποιώντας ουσιαστικά το κεφάλαιο που είχαμε δημιουργήσει με τους Ολυμπιακούς του 2004, **ο ίδιος** είχατε πει ότι αυτό είναι μια μεγάλη ευκαιρία ανάπτυξης για την Κεντρική Ελλάδα

³ All The data used for the purposes of this paper are from the Greek Parliament Minutes. (13th term of the Parliamentary Assembly, spring 2010. Online address: <http://www.parliament.gr/ergasies/>)

When we took over the Mediterranean Games, practically gaining from the assets we had created with the 2004 Olympic Games, **you** had said that this is a great development opportunity for central Greece.

In the case of (11), the speaker (who is an opposition MP) prefers the anaphor *o idjos* so as to expose contrast between the measures taken and what the PM was saying before that. Coming in (12), the intention of the speaker to distinguish the perspective of the internal protagonist from his/her own is more evident. In a sense the speaker implicitly states: 'you wanted the elections but we did not agree'. There is evidence to suggest that *o idjos* is marked for logophoricity or point of view. Finally, both in (13) and (14) the speaker by using *o idjos* brings to the surface the internal protagonist's perspective in order to juxtapose it with what the protagonist (or the protagonist's party in a broader sense) actually did. In these two last cases, this seems to be a kind of discourse technique especially when the speaker wants to reveal and stress the inadequacies between the protagonist's words and deeds. In general lines, it appears that the use of the anaphor *o idjos* instead of the zero pronoun carries a significant perspectival interpretation.

Based on our discussion so far, there are good reasons to believe that the preference of the anaphor *o idjos* over the zero pronoun is inherently pragmatic. In other words, it will be claimed that speakers generally tend to avoid using the anaphor *o idjos* without any particular reason for doing so. By contrast, when speakers prefer the use of the more marked anaphor they intend to convey more marked messages as well which cannot be inferred by the use of a zero pronoun. This brings us nicely to the second question addressed at in paragraph 3, namely, what are the principles which condition the choice and the interpretation, given the option {...zero/*o idjos*...}. As a way forward, I suggest that the interpretation of *o idjos* can follow from the systematic interaction of the neo-Gricean pragmatic principles of communication as introduced and developed by Levinson (1991, 2000) and Huang (1994, 2000, 2007).

5 THE NEO-GRICEAN PRAGMATIC PRINCIPLES

Levinson (1987, 1991, 2000) puts forth a neo-Gricean pragmatic model which reduces the original Gricean maxims to three inferential pragmatic principles, namely, the Q(uality), I(nformativeness) and M(anner)-principles. For our purposes, I will mainly focus on the I- and M-principles (for more on the Q-principle see Levinson 1987, 1991, 2000).

(15) The I-principle

Speaker's Maxim: the Maxim of Minimization

'Say as little as necessary', i.e. produce the minimal linguistic information sufficient to achieve your communicational ends (bearing the Q-principle in mind).

Recipient's Corollary: The Enrichment Rule.

Amplify the informational content of the speaker's utterance, by finding the most specific interpretation, up to what you judge to be the speaker's M-intended point.

Specifically:

- a) Assume that stereotypical relations obtain between referents or events, unless (i) this is inconsistent with what is taken for granted; (ii) the speaker has broken the Maxim of Minimization by choosing a prolix expression.
- b) Assume the existence of actuality of what a sentence is 'about' if that is consistent with what is taken for granted.
- c) Avoid interpretations that multiply entities referred to (assume referential parsimony); specifically: prefer co-referential readings of reduced NP's (pronouns or zeros)

(Levinson 2000: 114)

The I-principle is an upper bounding pragmatic principle, hence, when the speaker says '...p...' she/he conversationally implicates '... more than p...'. (Huang 2000: 209). In other words, according to the dictum of the I-principle, informationally weak expressions tend to be enriched informationally by the hearer.

(16) The M-Principle

Speaker's Maxim:

Indicate an abnormal, nonstereotypical situation by using marked expressions that contrast with those you would use to describe the corresponding normal, stereotypical situation.

Recipient's Corollary:

What is said in an abnormal way indicates an abnormal situation, or marked messages indicate marked situations, specifically: Where *S* has said "p" containing marked expression *M*, and there is an unmarked alternate expression *U* with the same denotation *D* which the speaker might have employed in the same sentence-frame instead, then where *U* would have I-implicated the stereotypical or more specific subset *d* of *D*, the marked expression *M* will implicate the complement of the denotation *d*, namely \bar{d} of *D*.

(Levinson 2000: 136-137)

Unlike the Q- and I-principles, which operate primarily in terms of semantic informativeness, the metalinguistic M-principle is operative primarily in terms of a set of alternates that contrast in form. The fundamental axiom upon which this principle rests is that the use of a marked or prolix expression M-implicates the negation of the interpretation associated with the use of an alternative, unmarked expression in the same set. This is illustrated in the example below.

- (17) a) The new manager is friendly.
 I +> The new manager is friendly in the stereotypical sense
 b) The new manager is not unfriendly.
 M +> The new manager is less friendly than the utterance in a) suggests.

It is tantamount to saying that genuine I-implicatures take precedence until the use of a marked linguistic expression triggers a complementary M-implicature to the negation of the applicability of the pertinent I-implicature (see e.g. Huang 2007 for further discussion).

5.1 A revised neo-Gricean Pragmatic Theory of Anaphora

Huang (1994, 2000, 2007) proposes a revised version of the neo-Gricean pragmatic theory of anaphora. More accurately, in cases where the use of a pronoun or anaphor instead of a zero does not M-implicate a contrast in reference, Huang argues that M-implicatures can be cancelled when they are inconsistent with contextual information. This explanation is in the spirit of the original Gricean theory given the distinctive properties of conversational implicatures, among which lies cancelability or defeasibility. Moreover, a way forward is to accept that M-implicatures do not operate only at the level of reference. In instances where the use of a more marked anaphoric expression instead of a less marked one does not give disjoint readings there is some sort of 'unexpectedness'. As Levinson (2000: 333) points out, M-implicatures naturally give rise to negative interpretations. It is left open on this account that the contrast intended may be either one of unexpectedness or point of view, on the one hand, or one of reference, on the other'. This prediction is spelled out in the revised neo-Gricean pragmatic apparatus in the following way (Huang 2000: 215):

- (18) Huang's revised neo-Gricean pragmatic apparatus for anaphora
- (a) Interpretation principles
- i) The use of an anaphoric expression x I-implicates a local co-referential interpretation unless (ii) or (iii):
 - ii) There is an anaphoric Q-scale $\langle x, y \rangle$, in which case the use of y Q-implicates the complement of the I-implicature associated with the use of x , in terms of reference.
 - iii) There is an anaphoric M-scale $\{x, y\}$, in which case the use of y M-implicates the complement of the I-implicature associated with the use of x , in terms of either reference or expectedness.
- (b) Consistency Constraints
- Any interpretation implicated by (a) is subject to the requirement of consistency with:
- i) the DRP.
 - ii) information saliency, so that
 - a) implicatures due to matrix constructions may take precedence over implicatures due to subordinate constructions, and

- b) implicatures of co-reference may be preferred according to the saliency of the antecedent, in line with the following hierarchy: topic > subject > object, etc.; and
- iii) general implicature constraints, namely,
 - a) background assumptions,
 - b) contextual factors,
 - c) meaning-*nn*, and
 - d) semantic entailments.

5.2 Analysis

Let us now return to our example cases (11-14) and examine in which way the anaphor *o idjos* introduces another layer of meaning in these utterances as well as how this meaning can be accounted for by the pragmatic apparatus described in (18). Consider again example (12) restated here for convenience in its original form in (19) and with the zero pronoun instead of *o idjos* in (20).

(19) Τότε ο κ. Παπανδρέου, αφού δημιουργούσε τέτοιο κλίμα και αφού **ο ίδιος** ζητούσε εκλογές για να δημιουργήσει προεκλογικό κλίμα, είπε...

(20) Τότε ο κ. Παπανδρέου, αφού δημιουργούσε τέτοιο κλίμα και αφού **Ø** ζητούσε εκλογές για να δημιουργήσει προεκλογικό κλίμα, είπε...

Then Mr Papandreou, since he was generating this climate and since **he himself /he** was asking for elections in order to create a pre-elections atmosphere, he said...

The reader is reminded at this point that (20) represents the unmarked way of saying things contrasting with (19) in which a more marked alternative is used. In (19) there is an opposition MP talking about the PM (*Mr Papandreou*) and the PM's demand for elections when he was in opposition. Therefore, there is a person who utters (19) (the speaker) and a person whose thoughts, words, acts and so on, are being reported by this utterance (the internal protagonist). It is clear that if the speaker wanted just to refer to the PM, the use of a zero pronoun, as in (20), would be rather unmarked and would not generate any ambiguous anaphoric interpretations. However, the speaker opts for the more marked *o idjos* intending thus to a more marked meaning. This meaning is marked in two ways: a) in terms of emphasis and contrast of the sort 'and not any other of the possible salient referents', and most importantly b) in terms of perspective. More precisely, the speaker by using *o idjos* wants the hearer to infer that he does not commit himself to the views (in this case demanding for elections) of the internal protagonist (i.e. *Mr. Papandreou*). This does not hold for (20) where the zero pronoun is used instead.

Coming to the examples (21) and (22) below, the choice of the anaphor *o idjos* gives an extra 'pragmatic weight'⁴ which is not conveyed by the use of the zero pronoun. In these examples the more marked anaphor shows the speaker's intention to put forward his/her own views and feelings in direct contrast with the views of the internal protagonist.

⁴ The notion of pragmatic weight was proposed by Davidson (1996) in a study on Spanish pronouns.

- (21) Τα μέτρα τα οποία ελήφθησαν εχθές, μέτρα τα οποία είναι πρωτοφανή, βρίσκονται σε πλήρη αναντιστοιχία με αυτά που ο Πρωθυπουργός είχε επαγγελλθεί προεκλογικά, αλλά και σε πλήρη αναντιστοιχία με όσα έλεγε **ο ίδιος** μετά από αυτό εδώ το Βήμα...

The measures that were taken yesterday, which are unprecedented, are not in accordance with what the Prime Minister had promised before the elections, and totally opposite to what **he himself** was saying here later on...

- (22) Όταν αναλάβαμε σαν χώρα τους Μεσογειακούς Αγώνες, κεφαλαιοποιώντας ουσιαστικά το κεφάλαιο που είχαμε δημιουργήσει με τους Ολυμπιακούς του 2004, **ο ίδιος** είχατε πει ότι αυτό είναι μια μεγάλη ευκαιρία ανάπτυξης για την Κεντρική Ελλάδα

When we took over the Mediterranean Games, practically gaining from the assets we had created with the 2004 Olympic Games, **you (yourself)** had said that this is a great development opportunity for central Greece.

What is more, it can be also claimed that in a discourse context like in the examples above, the intention of the speaker to express his/her point of view or perspective deals directly with the expression of evaluation/stance in discourse. Indeed, in most of the cases opposition MPs tend to make evaluative comments on the majority's views and deeds either directly or indirectly. The use of *ο ίδιος* in these contexts does not merely shows the speaker's preference for a specific antecedent but it is also a means of expressing point of view in an indirect way.

This evaluative stance is more evident in examples like (23).

- (23) Θέλω να πω ότι προσπαθείτε να κάνετε μια αύξηση εσόδων 1.000.000.000 ευρώ και **ο ίδιος** μάς ομολογήσατε ότι η έμμεση φορολογία επηρεάζει όλους.

I want to say that you try to increased incomes by 1.000.000.000 Euros and **you (yourself)** accepted that indirect tax affects everyone.

Here the opposition MP reports what the PM has said on indirect taxing. The speaker by using *ο ίδιος* in a context like this does not intend to oppose his/her perspective to the PM's one, since it is clear that he/she shares the belief that indirect taxing affects everyone. Therefore, the reason for choosing the anaphor *ο ίδιος* is because the speaker wants to make an evaluative comment highlighting the gap between what the PM was saying and what he actually did. As a result, the use of *ο ίδιος* intends an extra layer of pragmatic meaning which is not conveyed by the zero pronoun.

The logophoric interpretation of *ο ίδιος* can be claimed to be a non-stereotypical one compared the interpretation inferred by the zero pronoun. A step forward would be to suggest that the interpretation of *ο ίδιος* can be accounted for by the M-principle as described in (16). The M-principle predicts that "marked or prolix anaphoric expressions will tend to pick up the complement of the stereotypical extensions that would have been suggested by the use of the corresponding unmarked forms" (Levinson 2000: 38). The dictum of the M-principle has an intuitive basis since there must be a reason for choosing a marked expression where there is a choice for an unmarked one. Based on this rationale, the anaphor *ο ίδιος* and the zero pronoun

form an M-scale { \emptyset , *o idjos*} in which case, the use of the more marked anaphor, where the unmarked zero pronoun could have been used, M-implicates the complement of the interpretation associated with the use of the unmarked form. In this case the use of *o idjos* will tend to M-implicates two things: a) an emphatic/contrastive interpretation and b) a logophoric interpretation, i.e. the speaker intends to highlight the point of view of the internal protagonist of the sentence.

6 CONCLUSION

In this paper I examined the case of the Modern Greek anaphor *o idjos* and its relation with logophoric marking and the notion of 'point of view'. Our conclusion can be summarized in the following main points: First of all, Modern Greek is not a full logophoric language but it extends the use of a particular long-distance anaphor in logophoric contexts. More accurately, there is enough evidence to suggest that the anaphor *o idjos* can be used as a device for encoding logophoric or perspectival properties when in the subject position of a subordinate verb. Evidence from parliamentary discourse data suggests that in contexts where speakers want to highlight the point of view of the internal protagonist of the discourse tend to prefer *o idjos* over the unmarked zero pronoun. Therefore, in these contexts it appears that *o idjos* is marked for logophoricity. These logophoric interpretations are also reinforced by the presence of a logocentric verb. Moreover, the logophoric interpretation of *o idjos* seems to be pragmatically conditioned since it involves implicated (inferred) meaning. This interpretation can be accounted for by the systematic interaction of the neo-Gricean pragmatic principles of communication in the following way: given the distribution {... \emptyset /*o idjos*...}, there is an M-scale { \emptyset , *o idjos*} such that the use of the more marked anaphor will M-implicate the complement of the interpretation associated with the use of the zero pronoun i.e. a logophoric interpretation.

BIBLIOGRAPHY

- Chiou, Michael. 2010. NP-Anaphora in Modern Greek: A Partial Neo-Gricean Pragmatic Approach. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Clements, George. 1975. The pronoun in Ewe: Its role in discourse. *Journal of West African Languages* X, 2
- Culy, Christopher, 1994. Aspects of logophoric marking. *Linguistics* 32. 1055-1094.
- Culy, Christopher, 1997. Logophoric pronouns and point of view. *Linguistics* 35. 845-859.
- Hagège, Claude. 1974. Les pronoms logophoriques. *Bulletin de la Société de Linguistique de Paris* 69. 287-310.
- Huang, Yan, 1991. A neo-Gricean Pragmatic Theory of Anaphora. *Journal of Linguistics* 27. 301-333.
- Huang, Yan. 1994. *The Syntax and Pragmatics of Anaphora: A Study with special reference to Chinese*. Cambridge, Cambridge University Press
- Huang, Yan. 2000. *Anaphora: A Cross Linguistic Study*. Oxford, Oxford University Press.
- Huang, Yan. 2007. *Pragmatics*. Oxford, Oxford University Press.

- Kuno, Susumu. 1987. *Functional Syntax: Anaphora, Discourse and Empathy*. Chicago, University of Chicago Press.
- Kuno, Susumu. and Kaburaki, Etsuko. 1977. Empathy and syntax. *Linguistic Inquiry* 8. 627-672.
- Levinson, Stephen C. 1987. Pragmatics and the grammar of Anaphora: A partial pragmatic reduction of binding and control phenomena. *Journal of Linguistics* 23. 379-434.
- Levinson, Stephen C. 1991. Pragmatic Reduction of Pragmatic Conditions Revisited. *Journal of Linguistics* 27. 107-161.
- Levinson, Stephen C. 2000. *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. The MIT Press, Cambridge.
- Sells, Peter. 1987. Aspects of logophoricity. *Linguistic Inquiry* 18. 445-479.
- Stirling, Lesley. 1993. *Switch-reference and discourse representation*. Cambridge: Cambridge University Press.
- Zribi-Hertz, Ann. 1989. Anaphora binding and narrative point of view: English reflexive pronouns in sentence and discourse. *Language* 65. 695-727.

Fredrik Heinat and Satu Manninen

Gradient well-formedness of Finnish passive constructions¹

1 INTRODUCTION

In Finnish, a passive sentence is formed with the affixes *-ta-an* and *-ti-in* (present and past tense respectively) or with the auxiliary *olla* 'be' followed by the passive past participle of the lexical verb (present perfect and past perfect tense; see 1b-2b below):

	Active		Passive
1a.	(<i>Minä</i>) <i>ole-n</i> <i>avan-nut</i> <i>ove-n</i> .	b.	<i>Ovi</i> <i>o-n</i> <i>ava-ttu</i> .
	(I.nom) be-pres.1sg open-act.pcp door-acc		door.nom be-pres.3sg open-pass.pcp
	'I have opened the door'		'The door has been opened'
2a.	(<i>Minä</i>) <i>ol-i-n</i> <i>avan-nut</i> <i>ove-n</i> .	b.	<i>Ovi</i> <i>ol-i</i> <i>ava-ttu</i> .
	(I.nom) be-past.1sg open-act.pcp door-acc		door.nom be-past.3sg open-pass.pcp
	'I had opened the door'		'The door had been opened'

These are the only types of passive sentences recognized in most Finnish grammar books, including Setälä (1922), Penttilä (1963), Siro (1964), Hakulinen & Karlsson (1979) and Vilkuna (1996). Besides prototypical agent-patient predicates, these passives can be formed of even most intransitive and copular verbs. The preverbal position can remain empty, or it can be filled by various types of elements, including adverbials. This, together with the fact that the finite verb shows default agreement for third person singular and the verb is morphologically marked, in opposition to active verb forms, can be taken as evidence for an impersonal (i.e. non-agreeing) passive status of data like (1b)-(2b). For discussion about impersonal passives, see e.g. Siewierska (1984, 2005), Keenan (1985) and Shibatani (1985). For discussion about Finnish impersonal passives, see e.g. Manninen & Nelson (2004) and Helasvuo & Vilkuna (2008).

However, if (1b), (2b) and (3) are examples of a Finnish impersonal passive, a question arises about the status of sentences like (4) where the finite verb shows person and number agreement with the preverbal DP and the lexical verb is in the passive past participle form:

¹ We are grateful to Eva Klingvall for valuable comments on previous versions of this paper, and to Joost van de Weijer for the statistical analyses. The work is funded by the Swedish Research Council, grant number 2008-2081.

- | | | | |
|----|--------------------------------|-----------------------------|-------------------------------------|
| 3. | <i>Talot</i>
houses.nom | <i>o-n</i>
be-pres.3sg | <i>maala-ttu.</i>
paint-pass.pcp |
| | ‘The houses have been painted’ | | |
| 4. | <i>Talot</i>
houses.nom | <i>o-vat</i>
be-pres.3pl | <i>maala-ttu.</i>
paint-pass.pcp |

Although examples like (4) can be found in historical, dialectal and present-day sources, they are ignored in most literature on Finnish. The few sources that recognize their existence judge them prescriptively to be

- “transfer” from other languages (especially from Latin and Swedish; e.g. Häkkinen 1994, Löflund 1998, and the references cited therein)
- “mistakes” or “hypercorrect forms” when the writer has intended to use an impersonal passive (such as that in (3) above; e.g. Saarimaa 1947, 145f., Vilkkuna p.c)
- “mistakes” or “hypercorrect forms” when the writer has intended to use a copular sentence with nominative case (such as that in (5) below; e.g. Saarimaa 1947, 145f., Setälä & Nieminen 1957, 170, Ikola 2001, 149)

In Finnish copular sentences, the copula *olla* ‘be’ shows person and number agreement with the preverbal DP. It is followed by an adjectival phrase, which can be formed of a passive participle. The difference between the agreeing passive sentence in (4) and the copular sentence in (5) is that the participle in (5) shows number agreement with the preverbal DP and inflects for case (for nominative, partitive or essive, when the copula is *olla* ‘be’):

- | | | | | |
|----|---------------------------------|----------------------------|--|------------------------------------|
| 5. | <i>Talot</i>
houses.nom | <i>ovat</i>
be-pres.3pl | <i>maala-tu-t</i>
paint-pass.pcp-nom.pl | <i>/ sinise-t</i>
/ blue-nom.pl |
| | ‘The houses are painted / blue’ | | | |

The aim of this paper is to shed some light on agreeing passive sentences like (4). Although we can find examples of such sentences in different sources, there currently exists no description or analysis of the data. We do not know, for example, if examples like (4) really are “mistaken” or “hypercorrect” forms of either (3) or (5), or if we are instead dealing with variation (i.e. if (4) is an alternative form to either (3) or (5)). It is also possible that (4) is separate from (3) and (5) altogether, in which case we could be dealing with a new form (a “true” agreeing passive, perhaps). The paper is intended as a starting point for a description and analysis of data like (4). To be able to address questions such as those listed above, a first goal must be to gather information about how such sentences relate to both impersonal passives and copular sentences in native speakers’ minds.

2 OUTLINE OF A PILOT EXPERIMENT

As already noted, examples of sentences like (4) can be found in different sources; these include old Finnish texts (e.g. biblical texts from the 17th and 18th centuries, novels, short stories and plays from the 19th century), Finnish dialect archives (e.g. the *Muoto-opin arkisto*; we thank Mari Siironen and Maria Vilkuna for their valuable insights on the dialectal aspects of the data) and the internet (informal discussion forums, personal weblogs). Examples can also be found in many local newspapers and tabloid papers; however, because these sentences are prescriptively judged to be ungrammatical, it is often difficult to find examples in any “serious” newspapers or the existing corpora.²

In this paper, we focus entirely on present-day Finnish. Instead of investigating the frequency and uses of agreeing passive sentences like (4) in some specific source of data, our goal is to see how *acceptable* such sentences are perceived to be in native speakers’ minds, in relation to both impersonal passives (e.g. 3 above) and copular sentences (e.g. 5 above). To reach this goal, we have constructed a pilot experiment using the methodology of Magnitude Estimation (see e.g. Bard et al. 1996). In addition to agreeing passives like (4), impersonal passives like (3) and copular sentences like (5), we have included “incongruent” sentences like (6) in the experiment. In (6), which is an example of a form that is frequently found in informal spoken Finnish, the verb *olla* ‘be’ shows default agreement for third person singular while the passive participle shows number agreement with the preverbal DP and inflects for nominative case. Because both (4) and (6) are prescriptively judged to be “ungrammatical” it is interesting to see if there is a difference in acceptability between these sentence types in native speakers’ minds:

6.	<i>Talot</i>	<i>o-n</i>	<i>maala-tu-t.</i>
	houses.nom	be.pres.3sg	paint-pass.pcp-pl.nom

In section 2.1 we present the methodology of Magnitude Estimation, and in section 2.2 we outline the design of the experiment. In section 2.3 we present the results of the experiment. In sections 3 and 4 we interpret the results, discuss some of the questions that they raise and identify topics for future research.

2.1 What is Magnitude Estimation?

Magnitude Estimation is originally an experimental technique used in psychophysics; it is used to determine how much of a given sensation a person is experiencing of a given stimulus. Participants are first exposed to a standard stimulus (a modulus) which is assigned a fixed value by the experimenter, or to which the participants assign a value themselves. After this, the

² An anonymous reviewer suggests that the non-existence of examples like (4) in “serious” newspapers might be a result of such forms getting caught in the spell-check and grammar-check programs. A possible counter-argument is that examples like (4) can still be found in many local newspapers and the tabloids, many of which probably use the same programs.

participants are presented a series of stimuli which vary in intensity, and which they are asked to assess in relation to the modulus. The types of experiments where this methodology has been used include assessments of the brightness of a light source and the loudness of a sound source; both of these are situations where human perception and intensity of the physical stimuli are not linearly correlated. The methodology is also used in cases where it is difficult to measure the stimuli in absolute terms; one example is the participants' perception of pain.

Since the 1990s, Magnitude Estimation has been used successfully for testing linguistic acceptability (see e.g. Bard et al. 1996; Keller 2000; Featherston 2005; Sorace and Keller 2005). The methodology has several advantages over the traditional dichotomy grammatical-ungrammatical, or the use of a fixed scale such as 1 to 5. First, the participants can make as fine-grained distinctions as they want; this allows the researcher to capture the fact that linguistic constructions are hardly ever just grammatical or ungrammatical and that there is variation even within the two opposite poles. Second, there is no upper limit to the possible values that the participants can assign to the stimuli; the only restriction is that the numbers must be above zero (which usually means that the scale is from 1 to infinity; this is obviously what also enables the fine-grained distinctions). Third, Magnitude Estimation experiments yield results on an interval scale, instead of a nominal or ordinal scale, as is the case with categories or fixed numbers. An interval scale, in turn, allows the researcher to do more sophisticated and reliable statistical calculations.

2.2 The design of the experiment

The experiment was carried out over the internet using the WEBEXP2 free software package developed at the University of Edinburgh (see e.g. Keller et al. 2009). Although there are both advantages and disadvantages with conducting experiments over the internet, our experience is that the advantages by far outweigh the disadvantages; more discussion of this can be found in e.g. Reips (2007).

The participants in the experiment were 108 students at the University of Tampere and the University of Helsinki. They were contacted via their department mailing lists; the original message was sent to the Departments of Journalism and Mass Communication, Psychology, and Social Psychology in Tampere and the Department of Education in Helsinki, and it reached about 1000 students in total. Participation in the experiment was voluntary and the participants received no payment for their services. Out of the 108 participants, eight were excluded later: two because they were not native speakers of Finnish, and six because they had misunderstood the instructions and used either zero or negative numbers in their answers. 88 out of the 100 remaining participants were in their 20s, 11 in their 30s and one in her 40s. 74% of them were female, which probably reflects the gender division in the targeted departments.

The participants were first given detailed instructions on what they were expected to do. The instructions also contained several examples of estimations of both line lengths and linguistic acceptability. The participants were then asked to provide some personal details, such as their age, sex and if they were native speakers of Finnish. After this, they were asked to complete two short training sessions: one on assessing line lengths, the other one on assessing linguistic acceptability. After the training sessions, the actual experiment started.

The experiment consisted of 50 Finnish sentences which were presented to the participants in different randomized orders; this prevented the preceding sentences from affecting the values of the following sentences in any consistent way (because for each participant, these were different sentences). 40 of the 50 sentences were “real” test sentences, five were well-formed fillers and five were ill-formed fillers. All 50 sentences were assessed in relation to the well-formed modulus, (7), which was pre-assigned a fixed value of 100.

7. *Elokuvien ja muiden kuvaohjelmien nykymuotoisesta*
films.gen and other.gen picture.programs.gen current.shaped.elat
ennakkotarkastuksesta ollaan luopumassa.
pre-rating.elat be.passive give.up.iness
 ‘They are in the process of getting rid of the existing method of pre-rating
 films and other picture programs’
 (Helsingin Sanomat, 28th of January, 2010, www.hs.fi/kulttuuri)

The 40 test sentences contained five different passive participles. The participles were formed from the verbs *ostaa* ‘buy’, *syödä* ‘eat’, *avata* ‘open’, *sulkea* ‘close’ and *polttaa* ‘burn (down)’. The reason for selecting these participles/ verbs was that we found naturally occurring examples from the web of all of them, in all the four constructions. The naturally occurring examples were used as points of comparison when constructing the “real” test sentences. The participants were asked to judge two types of sentences for each participle: in type A, the preverbal position was filled by a plural DP (see 8a below) and in type B, by two coordinated singular DPs (8b):

- | | |
|---------------------------|---|
| Type A | Type B |
| 8a. <i>talot</i> ‘houses’ | b. <i>talo ja lato</i> ‘the house and the barn’ |

An overview of the constructions is given below. The numbers (1)-(4) correspond to the numbers above the circles in Figure 1, which can be found at the beginning of section 2.3. The letters A and B are explained above, in connection with (8a-b):

Label and sentence type	Form
A1 copular sentences with nominative case and plural DP (e.g. sentences like 5 above)	agreeing <i>olla</i> ‘be’ followed by passive participle in agreeing plural form
B1 copular sentences with nominative case and two coordinated singular DPs	same as A1
A2 agreeing passive sentences with plural DP (e.g. sentences like 4 above)	agreeing <i>olla</i> ‘be’ followed by passive participle in non-agreeing form
B2 agreeing passive sentences with plural DP	same as A2

A3	impersonal passives with plural DP (e.g. sentences like 3 above)	default <i>olla</i> 'be' followed by passive participle in non-agreeing form
B3	impersonal passives with two coordinated singular DPs	same as A3
A4	"incongruent" sentences with nominative case and plural DP (e.g. sentences like 6 above)	default <i>olla</i> 'be' followed by passive participle in agreeing plural form
B4	"incongruent" sentences with nominative case and two coordinated singular DPs	same as A4
5	well-formed filler sentences	varying types
6	ill-formed filler sentences	varying types

2.3 The results

The overall results of the experiment are presented in Figure 1. As already pointed out, the numbers above the circles correspond to the four sentence types under investigation. The first "dot" inside each circle is the A-sentence where the preverbal position is filled by a plural DP (e.g. 8a above) and the second "dot" is the B-sentence where the preverbal position is filled by two coordinated singular DPs (8b). We have done a repeated measures analysis of variance (ANOVA) and found a significant main effect of condition ($F(9, 945)=135.651$ $p=.000$); we have also performed pairwise comparisons of all ten sentence types using Sidak correction for multiple comparisons:

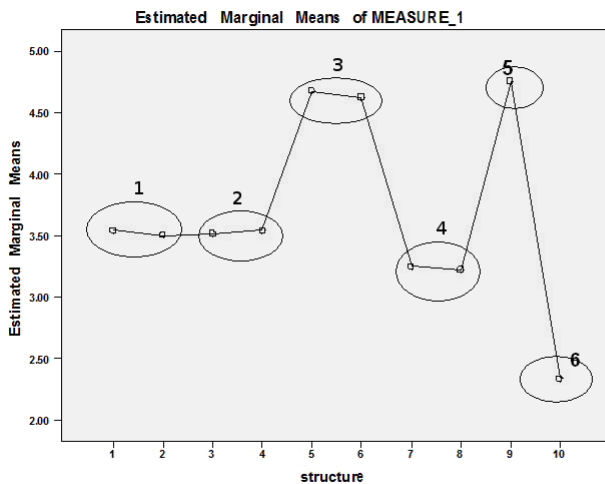


Figure 1

One of the first things that we notice is that the form of the preverbal DP does not have any effect on how acceptable a sentence is experienced to be: there is no statistically significant difference between e.g. Type A1 (copular sentences with plural DPs) and Type B1 (copular sentences with two coordinated singular DPs). The same holds for Types 2, 3, and 4. Secondly, the impersonal passives (Type 3 in Figure 1) are nearly as acceptable as the well-formed control sentences (Type 5 in Figure 1); there is no statistically significant difference between these two groups. The impersonal passives (Type 3) are judged to be considerably better than any of the other test sentences: there is a statistically significant difference between them and each of Type 1 (copular sentences), Type 2 (agreeing passive sentences) and Type 4 (“incongruent” sentences). There is also a small but significant difference between Type 4 (“incongruent” sentences) on the one hand and Types 1 (copular sentences) and 2 (agreeing passive sentences) on the other hand. Interestingly, there is no significant difference between sentences of Type 1 (copular sentences) and Type 2 (agreeing passive sentences).

3 DISCUSSION

The results of the pilot experiment are encouraging for several reasons, even if they at this stage seem to be raising more questions than they actually answer. This is, of course, a welcome result from something that was intended as a *starting point* for a description and analysis. In this section, we look at the results of the experiment, in light of what (little) has been said about agreeing passive sentences like (4) before.

3.1 Are we dealing with a pure “mistake” or “hypercorrect” form?

As pointed out in section 1, Finnish is often seen as a language which has only impersonal passives. One way to explain the existence of data like (4) would then be to treat them as pure “mistakes” or “hypercorrect” impersonal passives (i.e. as pure performance errors); this is indeed the explanation given in a lot of the literature. The reasoning behind this explanation is that speakers have simply been misled by the occurrence of the plural DP with nominative case in the preverbal position, as such DPs trigger agreement on the finite verb in Finnish active sentences. The fact that sentences like (4) are judged to be less acceptable than the impersonal passives could also be taken to support this line of reasoning: it is, after all, only natural to think that mistakes should be degraded relative to the forms that they are supposed to be mistakes of (i.e. that most speakers should be able to identify these sentences as mistakes). Secondly, the fact that sentences like (4) and (5) turn out to be equally acceptable in native speakers’ minds need not be anything more than a coincidence. However, if sentences like (4) really *are* mistaken impersonal passives, then the question immediately arises why they are not ranked even lower. Figure 1 shows that such sentences consistently fall within the medium range of acceptability: they are judged to be less acceptable than the impersonal passives and well-formed fillers, but they are still significantly better than the “incongruent” sentences and the ill-formed fillers.

The picture becomes even more unclear if we treat examples like (4) as pure “mistaken” or “hypercorrect” copular sentences with nominative case, which is the viewpoint taken in e.g. Saarimaa (1947), Setälä & Nieminen (1957) and Ikola (2001). Figure 1 shows that our participants

do not experience *any* difference in acceptability between these two sentence types; in other words, the mistakes are not degraded *at all* relative to the forms that they are supposedly mistakes of. This gives rise to (at least) two follow-up questions: first, why is there no difference in acceptability between sentences like (4) and (5) in native speakers' minds – can it *really* be just a coincidence? – and second, why do sentences like (5) fall within the medium range of acceptability in present-day Finnish? In view of what most grammar books say, we would have expected them to pattern with the impersonal passives and the well-formed fillers. Some answers to the second question can be found in the recent *Iso suomen verkkokielioppi* (henceforth VISK; see <http://kaino.kotus.fi/visk/>). According to VISK, copular sentences with nominative case are becoming less acceptable in sentences where case alternations (between nominative and partitive) are not directly linked to aspectual differences. As for the first question, most grammar books offer no explanation as to why sentences like (4) and (5) are equally acceptable (or equally degraded) in native speakers' minds. In our view, if sentences like (4) really are mistaken or hypercorrect copular sentences, which are themselves already degraded, then the important question is again why sentences like (4) are not ranked even lower.

It seems to us that neither of these pure “mistake” or “hypercorrect form” explanations can be on the right track. If agreeing passives like (4) were nothing more than mistakes or hypercorrect forms, we would have expected our participants – who are all relatively well-educated university students – to react consistently more strongly to such sentences, and to assign them lower values. This is, after all, what they have done with “incongruent” sentences like (6), or Type 4 in Figure 1 above, which are also prescriptively judged to be “mistakes” in most grammar books. The bottom line is that, while it is possible that the (medium range of) acceptability of data like (4) might be the result of a “mistake” or “hypercorrection” for *some* of the participants, it is unlikely to be the result of a “mistake” or “hypercorrection” for *most* of them. We therefore feel that the real answers to our questions lie elsewhere.

3.2 Are we dealing with variation?

An alternative account of data like (4) and of the results presented in Figure 1 would be to say that we are dealing with *variation*; in other words, instead of pure “mistakes” data like (4) are simply variants, in the sense of *alternative forms*, of “normal” impersonal passives or copular sentences, and they are assigned “normal” impersonal passive or copular readings. Under this line of reasoning, the next step should obviously be to determine *what kind of variation we are dealing with*; this would mean looking at both historical and dialectal data, and at the age-aspects of the phenomenon (as many older speakers we have discussed the data with have reported a strong dislike of sentences like (4)).

In our view, the line of reasoning pursued above is *partially* on the right track. The reason is that even copular sentences (e.g. 5) can actually be ambiguous between two different readings: a copular reading, which is stative, and a passive reading, which is eventive; see e.g. VISK (§1331-1333, §1336, §1432) for such claims. Crucially, according to VISK, the passive interpretations of

data like (5) are *not the same as the passive interpretations of data like (3)*. The examples, with the relevant interpretations, are repeated below for convenience:³

5. *Talot* *o-vat* *maala-tu-t.*
houses.nom be-pres.3pl paint-pass.pcp-nom.pl
‘The houses are painted’ (i.e. They are in a painted state)
‘The houses are painted’ (i.e. They are in a painted state as a result of someone having painted them’)
3. *Talot* *o-n* *maala-ttu.*
houses.nom be-pres.3sg paint-pass.pcp
‘The houses have been painted’ (i.e. Someone has painted the houses)

If we accept that sentences like (5) can be ambiguous between two different readings, then for sentences like (4) we have three different possibilities. These sentences could be variants, or alternative forms, of:

- “normal” impersonal passives like (3): they receive “normal” impersonal passive interpretations
- “normal” copular sentences like (5): they receive “normal” copular sentence interpretations
- passive readings of sentences like (5): they also receive the same kinds of passive interpretations

Below, we look briefly at the third possibility, which to us is the most interesting, and possibly also the most fruitful, alternative: sentences like (4) and (5) are examples of a “true” agreeing passive in Finnish, and this passive is separate from the “normal” impersonal passive (e.g. 3). At least at first sight, sentences like (4) and (5) meet (on their passive readings) the standard criteria attributed to agreeing (or, personal) passives cross-linguistically (see e.g. Siewierska 1984, 28ff., 2005, Keenan 1985, Shibatani 1985, and VISK §1331, 1332, 1336). First, they can have an active counterpart, such as that in (9) below:

³ The idea that Finnish has two different passive constructions could be motivated both on historical grounds, and in view of the fact that the closely related language Estonian has two different passive constructions. Estonian agreeing passives are formed with the auxiliary *olema* ‘be’ followed by the non-agreeing passive past participle of the lexical verb; in other words, the sentences look similar in form to the Finnish sentence in (4). The status of Estonian agreeing passives is subject to debate, however: Rajandi (1967) treats them as examples of a passive, while Erelt et al (1993) classify them as copular sentences, consisting of a copular verb and a deverbal adjective.

9. *Maalari* *o-n* *maala-nnut* *talo-t*.
 painter.nom be.pres.3sg paint-act.pcp houses-acc
 ‘The painter has painted the houses’

Second, their input verbs are two-place predicates (i.e. not intransitive or copular verbs, as in the case of impersonal passives; see e.g. VISK §1331, §1332, §1336). The logical subject is either demoted to an oblique phrase, or it disappears altogether (the more usual option, see e.g. 10-11 below). The logical object is promoted to passive subject; it controls agreement on the finite verb and undergoes alternation from accusative to nominative case. Finally, the verb is morphologically marked, in opposition to the corresponding active verb form:

10. *Talot* *o-vat* *maala-tut-t* (*maalarin toimesta*).
 Houses.nom be-pres.3pl paint-pass.pcp-nom.pl (painter.gen by)
 ‘The houses are painted (by a painter)’ (i.e. the houses are in a painted state because a painter has painted them’
11. *Talot* *o-vat* *maala-ttu* (*maalarin toimesta*).
 Houses.nom be-pres.3pl paint-pass.pcp (painter.gen by)
 ‘The houses are painted (by a painter)’ (i.e. the houses are in a painted state because a painter has painted them’

Of course, many of the criteria listed above also apply to Finnish impersonal passives: they can also have an active counterpart and show case alternation, and they are also morphologically marked, in opposition to active verb forms. Unless we can link the proposed agreeing passive vs. impersonal passive distinction to something more substantial, we risk classifying (4)-(5) as agreeing passives simply because their finite verb shows agreement with the preverbal DP. One key difference between the two types of passives is the interpretation assigned to the sentences: agreeing passive sentences seem to describe a state that results of or is brought about by a preceding event, whereas impersonal passives describe just the event itself; more discussion of the interpretive differences can be found in e.g. VISK (§1335-1339). Another difference seems (to us) to be the identity of the demoted/deleted agent. In “normal” impersonal passives like (3), the implied agent (logical subject) is typically interpreted as an unknown or generic human entity (see e.g. Manninen & Nelson 2004, Helasvuo & Vilkuna 2008). In agreeing passives like (4)-(5), on the other hand, both “specific” and “generic” human readings seem to be available. This is, at this stage, a topic that needs to be investigated more thoroughly, however.

3.3 A closer look at the interpretation: a mini-experiment

A property of passives cross-linguistically is that they are assigned eventive readings, while copular sentences are assigned stative readings; see e.g. Siewierska (1984). Under the variation view, an important goal should therefore be to examine if sentences like (4)-(5) are assigned eventive or stative readings by native speakers, or if they are ambiguous between these

readings. Furthermore, if they are assigned eventive readings, then are these readings the same as the ones assigned to impersonal passives like (3)?

As a first small step within the variation view, we have constructed a mini-experiment, to see which direction the data point. In this experiment, the participants were asked to judge the acceptability of agreeing passive sentences like (4) and copular sentences like (5) in contexts which were designed to favour either the stative (i.e. prototypical adjectival) or eventive (i.e. prototypical passive) reading. (12) is an example of a former type of context, (13) of a latter type of context:

12. *Millaisia talot ovat?* *Talot ovat _____ (maalattu / maalatut)*
'What are the houses like'
13. *Kyläläiset ovat maalanneet talot.* *Talot ovat _____ (maalattu / maalatut)*
'The villagers have painted the houses'

In both types of contexts, the participants reported that both sentence types could be used. However, many of them also indicated that in (12), they would have preferred the passive participle to have partitive case (i.e. *maalattuja*, instead of either *maalattu* or *maalatut*) while in (13) they would have preferred to use the impersonal passive (i.e. *on maalattu*, instead of either *ovat maalattu* or *ovat maalatut*). The number of participants and test sentences were too small to allow any statistical calculations or a proper analysis, but the results at least give an indication of what the future questions should be.

4 FUTURE TOPICS

The results from the pilot experiment described in section 2 and the mini-experiment described in section 3.3 suggest that data like (4) and (5) share the same interpretations; they can both have stative readings, and they can both have eventive readings in present-day Finnish. One factor that needs to be investigated in relation to this is the type of eventive reading. Another factor is the extent to which the judgments are dependent on the meanings of the passive participles/verbs; neither of our experiments showed any evidence of that, but the question is nevertheless worth exploring. Third, we need to investigate how much the choice of auxiliary and the context influence the participants' judgments of the sentences. The results of these investigations must also be complemented by a thorough investigation of both historical and dialectal data, to fully understand what is going on.

REFERENCES

Bard, Ellen, Dan Robertson, and Antonella Sorace. 1996. Magnitude estimation of linguistic acceptability. *Language* 72.32-68.

- Erelt, Mati, Reet Kasik, Helle Metslang, Henno Rajandi, Kristiina Ross, Henn Saari, Kaja Tael, and Silvi Vare. 1993. *Eesti keele grammatika. Sõnamoodustus, II Süntaks*. Lisa: Kiri. Tallinn: Eesti Teaduste Akadeemia Eesti Keele Instituut.
- Featherston, Sam. 2005. Universals and grammaticality: wh-constraints in German and English. *Linguistics* 43.667-711.
- Hakulinen, Auli, and Fred Karlsson. 1979. *Nykysuomen lauseoppia*. Helsinki: SKS.
- Helasvuo, Marja-Liisa, and Maria Vilkuna. 2008. Impersonal is personal: Finnish perspectives. *Transactions of the Philological Society*, 106.216-245.
- Häkkinen, Kaisa. 1994. *Agricolasta nykykieleeseen. Suomen kirjakielen historia*. Porvoo: WSOY.
- Ikola, Osmo. 2001. *Nykysuomen opas*. Turku: Publications of the Department of Finnish and General Linguistics of the University of Turku, 65.
- Keenan, Edward. 1985. Passive in the world's languages. *Language Typology and Syntactic Description*, ed. by Thomas Shopen, 243-281. Cambridge: Cambridge University Press.
- Keller, Frank. 2000. *Gradience in grammar*. Edinburgh: A University of Edinburgh dissertation.
- Keller, Frank, Subashini Gunasekharan, Neil Mayo, and Martin Corley. 2009. Timing accuracy of Web experiments: A case study using the WebExp software package. *Behavior Research Methods* 41.1-12
- Löflund, Juhani. 1998. *Suomen kirjoitetun yleiskielen passiivi*. Turku: A University of Turku dissertation.
- Manninen, Satu, and Diane Nelson. 2004. What is a passive? The case of Finnish. *Studia Linguistica* 58.212-251.
- Penttilä, Aarni. 1963. *Suomen kielioppi*. Porvoo: WSOY.
- Reips, Ulf-Dietrich. 2007. The methodology of Internet-based experiments. *The Oxford Handbook of Internet Psychology*, ed. by Adam Joinson, Katelyn McKenna, Tom Postmes, and Ulf-Dietrich Reips, 373-390. Oxford: Oxford University Press.
- Rajandi, Henno. 1967. Passiivne tud-adjektivisatsioon. *Keel ja Kirjandus* 4. 218-23.
- Saarimaa, E.A. 1947. *Kielenopas*. Porvoo: WSOY.
- Setälä, Emil Nestor. 1922. *Suomen kielioppi: äänne ja sanaoppi*. Helsinki: Otava.
- Setälä, Emil Nestor, and Kaarlo Nieminen. 1957. *Suomen kielen oppikirja*. Helsinki: Otava.
- Shibatani, Masatoshi. 1985. Passives and related constructions: A prototype analysis. *Language* 61.821-8.
- Siewierska, Anna. 1984. *The Passive: a Comparative Linguistic Analysis*. London: Croom Helm.
- Siewierska, Anna. 2005. Passive constructions. *World Atlas of Language Structure*, ed. by Martin Haspelmath, Matthew S. Dryer, David Gil, and Bernard Comrie, 434-437. Oxford: Oxford University Press.
- Siro, Paavo. 1964. *Suomen kielen lauseoppi*. Helsinki: Tietosanakirja Oy.
- Sorace, Antonella, and Frank Keller. 2005. Gradience in linguistic data. *Lingua* 115.1497-1524.
- Vilkuna, Maria. 1996. *Suomen lauseopin perusteet*. Helsinki: Edita.

VISK = Ison suomen kieliopin verkkoversio. [the web-edition of the Big Finnish Grammar] available at <http://kaino.kotus.fi/visk/etusivu.php>

Saara Huhmarniemi

Internal *wh*-movement in Finnish

1. INTRODUCTION

This paper examines *wh*-movement in Finnish *wh*-questions. It is often assumed that *wh*-movement is triggered by an interrogative feature residing in the left periphery of a finite clause. However, finite clauses are not the only domains that involve left-peripheral *wh*-movement. Finnish displays a general pattern of internal *wh*-movement inside pied-piped constituents (Huhmarniemi 2012).¹ The internal *wh*-movement is illustrated in (1b), in which the *wh*-phrase is contained within an adverbial clause that occupies a left peripheral position of the finite clause.² The *wh*-object *minkä* 'what' occupies the left peripheral position inside the fronted adverbial clause.

(1)

- a. Pekka tapasi Merjan [ostettuaan kirjan].
Pekka.NOM met Merja.ACC bought.INF book.ACC
'Pekka met Merja after he had bought a book.'
- b. [Minkä ostettuaan __] Pekka tapasi Merjan __?
what.ACC bought.INF Pekka.NOM met Merja.ACC
'What did Pekka buy before he met Merja?'

The question now arises, how the *wh*-phrase obtains the position at the edge of the adverbial clause in (1b). It is well established that the *wh*-phrase undergoes A'-movement inside finite clauses in Finnish (Vilkuna, 1989, 1995; Vainikka, 1989; Holmberg 2000; among others).³ The main objective of this study is to outline the evidence that the *wh*-phrase undergoes syntactic

¹ The term 'internal *wh*-movement' was introduced by Riemsdijk (1985).

² The non-finite clause type in (1) is called 'temporal construction' in Comprehensive Finnish Grammar (Hakulinen et al. 2004, §543). When the temporal construction does not contain overt subject, the verb receives possessive suffix according to the matrix subject. The possessive inflection is omitted in glosses. Syntactic properties of the temporal construction are examined for example in Vainikka (1989); Koskinen (1998); Huhmarniemi (2009).

³ The A'-movement inside non-finite clauses has been addressed by Koskinen 1998.

movement within pied-piped constituents such as the adverbial phrase in (1b). In addition, it is shown that this movement displays several properties typically associated to A'-movement.

This paper is organized as follows: Section 2 introduces some cross-linguistic properties of the target position of wh-movement, called here the 'edge' of a the relevant domain (Chomsky 2001, 2008). Section 3 outlines the syntactic properties of the internal wh-movement in Finnish that point toward A'-movement analysis. Section 4 contains the conclusions.

2. EDGE POSITION

The specifier (edge) position of the wh-phrase is often stated as a prerequisite for pied-piping in wh-movement languages (Webelhuth 1992, Cowper 1987, Grimshaw 2000). Examples (2a-b) show the relevance of the syntactic position of the wh-phrase in pied-piping constructions. In the relative clause (2a), the wh-phrase occupies the edge of the pied-piped DP. In (b), the wh-phrase occupies the complement position and the relative clause is not grammatical (Heck, 2008).

(2)

- a. a man [whose deck chair] you spilled coffee on __
- b. *a man [the deck chair of whom] you spilled coffee on __

Similarly, in Finnish, the wh-phrase has to occupy the edge of a pied-piped adposition phrase in a wh-question like (3a). The sentence (b), in which the wh-phrase occupies the complement position, qualifies only as an echo question (Manninen 2003).

(3)

- a. [Mitä kohti] Pekka juoksi __?
what.PAR towards Pekka.NOM ran
'What did Pekka run towards?'
- b. [Kohti mitä] Pekka juoksi __?
towards what.PAR Pekka ran
'Pekka ran towards what?'

Hungarian possessive construction provides another example of the importance of the edge position in pied-piped phrases. In Hungarian, the possessor argument in dative case may occupy the position at the edge of DP, as in (4a). However, the possessor in nominative case has to remain inside the NP (b-c) (Szabolcsi 1994, p. 98).

(4)

- a. Marinak a kalap-ja
Mary.DAT DET hat-POSS.3SG
'Mari's hat'

- b. (a) Mari kalap-ja
 DET Mari.NOM hat-POSS.3SG
 'Mari's hat'
- c. *Mari a kalap-ja
 Mari.NOM DET hat-POSS.3SG

The two possessors differ from each other in their ability to trigger pied-piping of the containing DP. DPs with a *wh*-possessor in the dative case may undergo *wh*-movement (5a), whereas this is not the case for DPs with nominative *wh*-possessor (b) (Szabolcsi 1994). Szabolcsi (1994) suggests that the *wh*-phrase has to move to the edge of DP in Hungarian possessive constructions for the pied-piping to be available.

- (5)
- a. [Kinek a [kalap-ja]]?
 who.DAT DET hat-POSS.3SG
 'Whose hat?'
- b. *[[Ki kalap-ja]]?
 who.NOM hat-POSS.3SG

Cross-linguistically, the pied-piping is associated with the edge position in a variety of languages. This finding has been formulated recently by Heck (2004, 2008) as edge generalization:

(6) **Edge generalization** (Heck, 2008, p. 88)

If a *wh*-phrase α pied-pipes a constituent β , then α has to be at the edge of β .

Finnish *wh*-questions, among others, provide support for the edge generalization: the *wh*-phrase has to occupy the edge position within pied-piped adposition phrases (PPs), DPs, APs, non-finite adjunct clauses (such as (2)), as well as certain non-finite complements (Hakulinen et. al. 2004, §1682, Huhmarniemi 2012). Nevertheless, the edge generalization (6) does not commit itself on the manner in which a *wh*-phrase obtains the edge position within the pied-piped phrase. The next section presents evidence that suggests that the edge position is in Finnish can be obtained via A'-movement.

3. INTERNAL WH-MOVEMENT TO THE EDGE

Structures that involve A'-movement share a group of characteristic properties that can be considered indicators of the presence of movement (Chomsky 1977). These properties include dislocation of an element from the position where its thematic roles and case are assigned, c-commanding between the landing site and the trace, reconstruction of reflexive anaphors, island

conditions, and pied-piping.⁴ This section examines Finnish internal wh-movement relative to these properties. Further examples on internal wh-movement inside different types of phrases in Finnish and their A'-movement properties can be found in Huhmarniemi (2012).

3.1 Thematic roles and case

The A'-movement preserves the thematic role and case of an argument, as illustrated with sentential wh-movement in (7a). Perhaps the most explicit examples of the same property in internal wh-movement come from the adverbial clauses such as temporal construction and MALLA-adjunct in (7b-c). The object wh-phrase bears the thematic patient role and accusative case. In both adverbial clauses, the wh-phrase retains its thematic role and case at the edge position.

- (7)
- a. Minkä Pekka osti __?
 what.ACC Pekka.NOM bought.INF
 'What did Pekka buy?'
 - b. [Minkä ostettuaan __] Pekka tapasi Merjan __?
 what.ACC bought.INF Pekka.NOM met Merja.ACC
 'After buying what did Pekka meet Merja?'
 - c. [Minkä ostamalla __] Pekka yllätti Merjan __?
 what.ACC buy.INF Pekka.NOM surprised Merja.ACC
 'By buying what did Pekka surprise Merja?'

The thematic role and case of the wh-phrase thus suggest that the wh-phrase has not been base-generated to the edge, but instead, dislocated from its case-assignment position.

3.2 Landing site and c-command

Example (2b), repeated here as (8a), shows that wh-phrase occupies the edge of a pied-piped phrase in a wh-question in Finnish. The same is true for wh-phrases in embedded questions (8b) as well as for relative pronouns in relative clauses (8c). Finally, there is a tendency for the so-called second position clitic particles to attach to the constituent at the edge. Example (8d) illustrates how the question particle -kO attaches to the constituent at the edge of an adverbial clause.⁵

⁴ Other tests for A'-movement include weak crossover (Postal, 1971; Wasow, 1979), quantifier scope changes, and the licensing of parasitic gaps (Chomsky, 1982). These properties are not addressed here.

⁵ The distribution of the -kO-particle is not as constrained as the distribution of wh-phrases. First, the -kO particle may in (8d) attach also to the whole fronted adverbial clause. In addition, there are some special cases, where the -kO particle does not have to occupy the edge position inside the pied-piped constituent (see Holmberg, 2008 for the distribution of the -kO-particle within DPs). However, the constituent that contains the -kO-particle occupies the edge of C.

(8)

- a. [Minkä ostettuaan __] Pekka tapasi Merjan __?
what.ACC bought.INF Pekka.NOM met Merja.ACC
'What did Pekka buy before he met Merja?'
- b. Juha ihmetteli, [minkä kirjan ostettuaan __] Pekka tapasi Merjan __?
Juha wondered which.ACC book.ACC bought.INF Pekka met Merja.ACC
'Juha wondered, which book Pekka bought before he met Merja.'
- c. kirja, [jonka ostettuaan __] Pekka tapasi Merjan __.
book which.ACC bought.INF Pekka.NOM met Merja.ACC
'the book that Pekka bought before he met Merja'
- d. [Kirjan-ko ostettuaan __] Pekka tapasi Merjan __?
book.ACC-kO bought.INF Pekka.NOM met Merja.ACC
'Was it after having bought a book that Pekka met Merja?'

In the examples above, the internal wh-movement targets the specifier of the relevant phrase and the moved wh-phrase thus c-commands the gap position. The c-commanding is a general property of syntactic relations and operations such as A'-movement (see e.g. Chomsky, 1995), whereas the lack of c-command between the moved wh-phrase and the gap would point towards an analysis in terms of sideward movement or scrambling.

3.3 Reflexive reconstruction

One of the diagnostic properties of A'-movement is that it preserves certain structural relations established at the base position, such as reflexive binding (see, e.g. Mahajan, 1990). This section investigates the binding properties of Finnish reflexive pronouns with regard to internal wh-movement. One of the available means to construct a reflexive anaphor in Finnish is to attach a 3rd person possessive suffix to a noun head, as in (9a) (Trosterud 1993; Vainikka 1989). The available correlates for the anaphoric expression *omistajalleen*, 'to its owner', are the sentence subject *Pekka* and the direct object *koiran* 'dog'. In sentence (9b), the anaphoric expression has moved to the edge of the finite clause, preserving its binding properties. We say that reflexives 'reconstruct' in A'-movement in Finnish.

(9)

- a. Pekka_k vei koiran_i omistajalleen_{i/k}.
Pekka.NOM took dog.ACC owner.to.Px
'Pekka took the dog to its owner.'
- b. Omistajalleen_{i/k}-ko Pekka_k vei koiran_i __?
owner.to.Px-kO Pekka.NOM took dog.ACC
'Was it its owner that Pekka took the dog to?'

The binding properties of reflexive anaphors are likewise structurally determined within Finnish adverbial phrases. Example (10a) shows that the direct object of the finite verb is an available correlate for the reflexive anaphor inside the PP in addition to the matrix subject. Example (10b) illustrates the situation after the internal *wh*-movement; the direct object remains to be an available correlate for the reflexive.

(10)

- a. Pekka_k palasi kotiin [vietyään koiran_i omistajalleen_{i/k}]
 Pekka.NOM returned home.to take.INF dog.ACC owner.to.Px
 'Pekka returned home after taking the dog to its owner.'
- b. [Omistajalleen_{i/k}-ko koiran_i vietyään __] Pekka_k palasi kotiin __?
 owner.to.Px-KO dog.ACC take.INF Pekka.NOM returned home.to
 'Was it its owner that Pekka took the dog to before returning home?'

Thus, the binding properties of reflexive anaphors reconstruct in internal *wh*-movement. This supports the contention that the edge position of the *wh*-phrase is achieved via A'-movement, and not via base-generation, for example.

3.4 Pied-piping

Sometimes more material moves to the edge position in addition to the *wh*-phrase itself. This phenomenon characteristic to movement operations is called *pied-piping*. For example, if the *wh*-phrase is contained within an adposition phrase in Finnish, the adposition is pied-piped along with the *wh*-phrase to the edge of C (11a). Another example of pied-piping is given in (11b), in which the *wh*-possessor pied-pipes the DP to the edge of C.

(11)

- a. [Mitä kaupunkia kohti] Pekka matkusti __?
 which.PAR city.PAR towards Pekka.NOM travelled
 'Which city did Pekka travel towards?'
- b. [Kenen kirjaa] Pekka luki __?
 whose book.PAR Pekka.NOM read?
 'Whose book did Pekka read?'

When the adposition phrase is contained within an adverbial clause, such as (12a), the adposition is pied-piped to the edge of the adverbial along with the *wh*-phrase. Similarly in (12b), the internal *wh*-movement involves pied-piping of the containing DP.

(12)

- a. [[Mitä kaupunkia kohti] matkustaessaan __] Pekka tapasi Merjan __?
 which.PAR city.PAR towards travel.INF Pekka met Merja.ACC
 'When traveling towards which city did Pekka meet Merja?'

- b. [[Kenen kirjaa] lukiessaan __] Pekka nukahti __?
 whose book.PAR read.INF Pekka.NOM fell.asleep
 'While reading whose book did Pekka fall asleep?'

Examples (12) thus illustrate that internal *wh*-movement involves pied-piping of the containing constituent in similar contexts as the sentence-level *wh*-movement

3.5 Islands for movement

The alternative *wh*-movement mechanism to pied-piping is extraction: in extraction, the *wh*-phrase leaves the containing constituent without pied-piping. However, certain phrases do not allow elements to be extracted from them; they are called islands (Ross, 1967). For example, it is not possible to extract a genitive argument out of a DP in Finnish (13a-b). In this case, the pied-piping is the only option for forming a *wh*-question (13c).

(13)

- a. Pekka lainasi [Merjan kirjaa]
 Pekka.NOM borrowed Merja's book.PAR
 'Pekka borrowed Merja's book.'
- b. *Kenen Pekka lainasi [__ kirjaa]?
 whose Pekka.NOM borrowed book.PAR
- c. [Kenen kirjaa] Pekka lainasi __?
 whose book.PAR Pekka.NOM borrowed
 'Whose book did Pekka borrow?'

Finnish internal *wh*-movement obeys island conditions, such as the restriction on possessor extraction, as illustrated in (14a-b). Similarly as in movement to the edge of the finite clause, the *wh*-possessor pied-pipes the containing DP to the edge of the adverbial.

(14)

- a. lainatessaan [Merjan kirjaa]
 borrow.INF Merja's book.PAR
 'while borrowing Merja's book'
- b. *kenen lainatessaan [__ kirjaa]
 whose borrow.INF book.PAR
- a. [kenen kirjaa] lainatessaan __
 whose book.PAR borrow.INF
 'while borrowing whose book?'

The *wh*-phrase thus obtains the edge position under conditions of pied-piping (section 3.4) and islands. These properties have been associated with A'-movement since they were first introduced by Ross (1967).

4. CONCLUSIONS

The aim of this study was to examine the syntactic properties of Finnish internal wh-movement. It was shown that some of the central syntactic properties associated with A'-movement within finite clauses are present in Finnish internal wh-movement contexts. These findings support a view in which the word order changes in different syntactic domains are similar to each other and possibly share a common footing in the linguistic computation.

REFERENCES

- Chomsky, Noam. 1977. On wh-movement. *Formal Syntax*, ed. by P. Culicover, T. Wasow, and A. Akmajian, 71–132. New York: Academic Press.
- Chomsky, Noam. 2001. Derivation by phase. *Ken Hale: A life in language*, ed. by Michael Kenstowicz, 1–52. Cambridge MA: MIT Press.
- Chomsky, Noam. 2008. On phases. *Foundational issues in linguistic theory: Essays in honor of Jean-Roger Vergnaud*, ed. by Robert Freidin, Carlos Otero, Maria-Luisa Zubizarreta, 133–166. Cambridge, MA: MIT Press.
- Cowper, Elizabeth. 1987. Pied piping, feature percolation, and the structure of the noun phrase. *Canadian Journal of Linguistics* 32. 321–338.
- Grimshaw, Jane. 2000. Locality and extended projections. *Lexical specification and insertion*, ed. by P. Coopmans, M. Everaert, and J. Grimshaw. Amsterdam: Benjamins.
- Hakulinen, Auli, Maria Vilkuna, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen, Irja Alho (ed). 2004. *Iso suomen kieliooppi*. Helsinki: SKS.
- Heck, Fabian. 2004. A theory of pied-piping. Ph.D. thesis, Universität Tübingen.
- Heck, Fabian. 2008. On pied-piping. Berlin: Mouton de Gruyter.
- Holmberg, Anders. 2001. The syntax of yes and no in Finnish. *Studia Linguistica* 55(2). 141–175.
- Holmberg, Anders. 2008. The syntax of the Finnish question particle. Manuscript.
- Holmberg, Anders and Urpo Nikanne. 2002. Expletives, subjects and topics in Finnish. *Subjects, expletives and the EPP*, ed. by Peter Svenonius, 71–105 Oxford University Press, New York.
- Huhmarniemi, Saara. 2009. Extraction islands in Finnish. *Biolinguistica Fennica Working Papers* 1. 21–78.
- Huhmarniemi, Saara. 2012. Finnish A'-movement: Edges and Islands. Ph.D. thesis. Institute of Behavioural Sciences, Studies in Cognitive Science 2. University of Helsinki, Helsinki. URL: <http://urn.fi/URN:ISBN:978-952-10-7712-8>.
- Koskinen, Päivi, 1998. Features and categories: Non-finite constructions in Finnish. Ph.D. thesis, University of Toronto. Distributed by Toronto Working Papers in Linguistics.
- Mahajan, Anoop Kumar, 1990. The A/A-bar distinction and movement theory. Ph.D. thesis, MIT. Cambridge, MA: Working Papers in Linguistics.

- Manninen, Satu. 2003. Finnish PPs and the Phase Impenetrability Condition. *Generative approaches to Finnic and Saami linguistics*, ed. by Diane Nelson and Satu Manninen, 295–320. Stanford, CA: CSLI Stanford University.
- Postal, Paul. M. 1971. *Cross-over phenomena*. New York: Holt, Rinehart, and Winston.
- Riemsdijk, Henk C. van. 1985. On pied piped infinitives in German relative clauses. *Studies in German grammar*, ed. by Jindrich Toman, 165–192. Dordrecht: Foris.
- Ross, John R. 1967. *Constraints on variables in syntax*. Ph.D. thesis, MIT.
- Szabolcsi, Anna, 1994. The noun phrase. *The syntactic structure of Hungarian*, ed. by Feren Kiefer and Katalin É Kiss. Vol. 27 of *Syntax and Semantics*, 179–274. London: Academic Press.
- Trosterud, Trond. 1993. Anaphors and binding domains in Finnish. *Case and other functional categories in Finnish syntax*, ed. by Anders Holmberg and Urpo Nikanne, 225–243. New York: Mouton de Gruyter.
- Vainikka, Anne. 1989. *Defining syntactic representations in Finnish*. Ph.D. thesis, University of Massachusetts Amherst.
- Vilkuna, Maria. 1989. *Free word order in Finnish: Its syntax and discourse functions*. Finnish Literature Society, Helsinki.
- Vilkuna, Maria. 1995. *Discourse configurationality in Finnish*. *Discourse Configurational Languages*, ed by Katalin É. Kiss, 244–268. Oxford University Press.
- Wasow, Thomas. 1979. *Anaphora in generative grammar*. Ghent: E. Story.
- Webelhuth, Gert, 1992. *Principles and parameters of syntactic saturation*. Oxford: Oxford University Press.

ACKNOWLEDGEMENTS

I would like to thank Pauli Brattico and Anne Vainikka for the insightful discussions about these topics. This work was supported by a grant from the Jenny and Antti Wihuri Foundation.

Juha Mulli

Observations on unique constituents in German

1 INTRODUCTION

There are many lexemes, for instance nouns, e.g. *Fersengeld*, that occur in a very limited number of linguistic contexts in modern German (see e.g. Fleischer 1997). In many Indo-European languages studied thus far, these constituents are prone not only to be lexically, but also morpho-syntactically tightly fixed. In English, for example, unique nouns do not undergo any syntactic transformations (cf. Moon 1998: 79, Langlotz 2006).

Unique constituents (henceforth UC) have been treated relatively extensively in many languages, for example in German, English and Finnish. According to Weinreich (1980: 252), these kinds of constituents are *unique*; Fillmore et al. (1988: 506) call them *unfamiliar pieces*. Nenonen and Niemi (1999) dub them *idiomatic isolates*. The most common term used in German is *unikale Komponente* (see Fleischer 1997: 37), for example *Fersengeld* in the phrase *Fersengeld geben* 'to turn tail and run' (lit. *Ferse* 'heel', *Geld* 'money'). The UCs are used exclusively in a specific idiomatic construction and these constituents belong to a special case of idioms consisting of a very frequent basic verb and a highly restricted noun. In the studies referred to above, the main thrust has been to describe the different groups of UCs and their use in language, mainly based on data retrieved from corpora.

However, there is not much research scrutinizing the use of UCs in German based on data retrieved from extensive corpora. In the present brief study, the use of two UCs will be examined based on a collection of texts in electronic form. In light of the limited scope, the focus will be on the unique nouns and idiomatic phrase constructions in contemporary German. The main aim of the current study is to examine whether they appear only in specific idiomatic phrases and only with specific verbs as is often taken for granted (cf. e.g. Fleischer 1997 and Dobrovolskij & Piirainen 1994). The data collected is based on an extensive corpus of written language (IDS-Korpus). The results of this study will provide more detailed information on the use of idioms and can be used as the basis for a more stringent treatment of this type of idioms in German lexicography. In addition, the results will challenge the traditional view on UCs by gathering insights into the actual use of these kinds of idioms and their components.

2 BACKGROUND: UCS IN GERMAN

As stated above, UCS in German have been treated quite extensively. Examples of these studies are – to name a few – Dobrovol'skij 1978, 1988; 1989; Feyaerts 1992, Fleischer 1982, Naumann 1987; Piirainen 1991, Dobrovol'skij and Piirainen 1994, Häcki Buhofer 2002. However, the overwhelming majority of these studies focus on data retrieved from dictionaries. Other kinds of corpora, e.g. the extensive corpus of contemporary German, IDS-Korpus, have – to the best of my knowledge – not been subjected to a detailed analysis.

In the present study, the notion of UC is used as follows: Unique constituents are those lexemes that appear only in a specific idiomatic construction, e.g. *Fersengeld geben*, while they do not appear at all as free lexemes outside these specific constructions. In this example, *Fersengeld* is the UC. In other words, traditionally, especially in the sphere of lexical semantics, the UCS do not have a meaning outside the idiomatic phrase at all (see Gehweiler et al. 2007). Nouns are said to account for the overwhelming majority of the UCS in German, but there also exist unique verbs and adjectives/adverbs in German (for more details, see Fleischer 1997: 37ff.).

Compared with Dutch and English, Dobrovol'skij (1989: 57) postulates that UCS are typical of German although not all cases are clear-cut. According to his study, there are 319 UCS in German, whereas a total of 251 UCS were found in Dutch and 52 UCS in English.

3 AIMS OF THE PRESENT STUDY AND DATA

Phraseology suffers from the problem that a dynamic phenomenon is represented as static. This is due to the fact that new editions of (commercial) dictionaries of idioms are published at long intervals and they in general suffer from restrictions of space. This may result in the fact that changes in meaning, for example, are often ignored although they have already entered the language.

The present study has two major aims. First, it will be examined whether typical UCS in German can be found in an extensive corpus of contemporary German, i.e. in the IDS-Korpus. Moreover, a number of dictionaries will be consulted. The followings questions will be addressed in depth: Do the UCS studied appear only in specific idiomatic constructions and only with a specific verb? How tightly fixed are these phrases in this sense?

The main thrust is to find out whether these UCS are really unique. The hypothesis is – based on previous studies in this sphere – that this is not necessarily the case (cf. Dobrovol'skij 1989; Dobrovol'skij and Piirainen 1994).

The present study is based on two prototypical instances of UCS: *Fettnäpfchen* (in the phrase *ins Fettnäpfchen treten* 'to blunder, to put one's foot in it', lit. 'to step into the grease pot') and the second one is *Fersengeld* (in the phrase *Fersengeld geben*, respectively, 'to turn tail and run', lit. 'to give heel money'). These cases are regarded as prototypical by Fleischer (1997). The first UC, *Fettnäpfchen*, goes back to the old days when it was common for farms to have a dish for boot/shoe grease close to the entrance so that people with wet boots/shoes coming in were able to grease their shoes. However, when somebody had inadvertently stepped in it, there were greasy spots on the floor boards. Not surprisingly, this resulted in displeasure. The second UC,

Fersengeld, dates back to the Middle Ages. In those days, it was quite common for travelers not to pay for their lodgings; instead of paying off old scores, they just turned on their heels and left very quickly. The only payment they made was “heel money” (for more details, see e.g. Duden 2008).

Why just these two UCs? The decision to study these UCs can be motivated as follows. First, previous studies have shown that not all idiomatic phrases comprising a UC can be attested in the IDS-Korpus. However, these two UCs tend to be quite prototypical in the sense that they are frequently used in the IDS-Korpus. In brief, these two seem also to be quite widespread in the German-speaking countries, i.e. their use is not limited to Germany. Second, contrary to expectations, based on data retrieved from the corpus, the use of the UCs will modify the results of previous studies (e.g. Fleischer 1997). One could also claim that these two UCs are by no means marginal (some other UCs can be regarded as *archaisms* since they do not appear in everyday German). For these reasons the present study is based on just these two UCs.

The IDS-Korpus on which the present study is based comprises altogether more than 3.7 billion running words and is mainly based on newspapers and some fiction. However, there is no comprehensive corpus of spoken German available. Due to these practical restrictions, the scope will be limited to written language. Hence, the subcorpus W (W – öffentlich – alle öffentlichen Korpora des Archivs W) consisting of 2.2 billion running words at the moment was chosen for the present study. It is the largest subcorpus available to all researchers and contains written language from the 18th century onwards. In addition, a number of dictionaries (of idioms) were consulted to provide information about the UCs in idiomatic phrases. The dictionaries on which the present study is also based are: Agricola (1992), Brockhaus-Wahrig (1980-1984), Duden (1999, 2002, 2003 and 2008), Friederich (1991), Grimm (1984), Paul (1992), Röhrich (2004), Schemann (1992, 1993 and 2000) and Wahrig (2007).

4 RESULTS

The citation form and a semantic paraphrase for each form will be provided. This information will be compared with those given in well-known and authoritative dictionaries of idioms and general dictionaries (see above). In addition, particular attention will be paid to occurrences retrieved from the IDS-Korpus.

4.1 Case study 1: *ins Fettnäpfchen treten*

The reference dictionaries examined give approximately the same semantic paraphrase ‘to blunder’; lit. ‘to step into the grease pot’ although the expression is treated somewhat differently with regard to the citation form. This is *ins Fettnäpfchen treten* but according to Duden (2008), *tappen* can also occupy the verb position. Most dictionaries remark on the restriction on the idiom-external, free use of the noun by adding *only in this expression*. However, Brockhaus-Wahrig (1980-1984) remarks *mostly in this expression*.

Based on the evaluation of more than 600 occurrences (i.e. *ins Fettnäpfchen* + verb) attested in the IDS-Korpus, a prototypical occurrence is as follows (1):

- (1) Der Tennis-Star Steffi Graf ist Sport-Geschichte. Darf ich Sie trotzdem noch Steffi nennen, oder trete ich dann ins Fettnäpfchen? STEFFI GRAF: Um Gottes Willen, nein. (M02/FEB.13880 Mannheimer Morgen, 21.02.2002, Ressort: Welt und Wissen; Auch wenn die Paparazzi das Anwesen belagern, genießt Steffi Graf das ungewohnte Familienglück)
'The tennis star Steffi Graf is part of the history of sports. May I still call you Steffi or do I put my foot in it? STEFFI GRAF: For heaven's sake, no.'

In general, the reference dictionaries list *treten* as a typical verb which is the most frequent one based on the IDS-Korpus. However, the verb *tappen* can also occupy the verb position but *ins Fettnäpfchen tappen* is not as often attested in the corpus (40 occurrences) as *treten* (420 occurrences). A typical occurrence with *tappen* is (2):

- (2) Im Geschäftsleben gelten oft andere Benimm-Regeln als im Privatleben. Diese müssen viele (Hoch-)Schulabgänger noch lernen – sonst tappen sie schnell ins Fettnäpfchen. (M05/NOV.93654 Mannheimer Morgen, 15.11.2005, Ressort: Wirtschaft und Mehr; Benimm-Kurse sollen "Klopse" am Arbeitsplatz verhindern)
'The etiquette for business differs from our private lives. Many graduates still must learn this or otherwise they will soon put their foot in it.'

In addition, there exist altogether 148 occurrences that are subject to modification. It is interesting to observe that the idiomatic phrase containing this UC can quite often be modified – however, the modifications found in the corpus in most cases do not trigger a change in meaning, i.e. the idiomatic phrase is modified as a whole and the basic meaning 'to blunder' remains the same. An example of this is (3). We can conclude that the modification in general provides more detailed information about the context, for example that somebody has made a number of mistakes.

- (3) In meiner Naivität trat ich in alle erdenklichen Fettnäpfchen. Uh, Sie meinen Darius, sagte ich. (M00/DEZ.81774 Mannheimer Morgen, 30.12.2000; 2000 by C. Bertelsmann Verlag, München)
'Due to my naivety, I made every imaginable mistake. Oh, you mean Darius, I said.'

In addition, a number of other verbs can occupy the verb position in this idiomatic phrase (46 occurrences found in the corpus) without a change of meaning. The verbs that occur in the corpus are (number of occurrences in parentheses): *greifen* 'to grip' (2); *hineintreten* 'to step into' (2); *setzen* 'to set/place' (6); *springen* 'to jump' (2); *steigen* 'to step' (8); *stolpern* 'to trip' (5); *tapsen* 'to fumble' (15); *trampeln* 'to stamp' (6).

Based on data retrieved from the IDS-Korpus, the UC *Fettnäpfchen* is by no means restricted to the phrase *ins Fettnäpfchen treten/tappen*, but it can also occur with a number of different verbs and in different syntactic patterns (cf. 4 and 5). Notwithstanding this kind of lexical and syntactic variation attested in the IDS-Korpus, *ins Fettnäpfchen treten* denotes more or less 'to screw up' or 'to say (or perhaps more basically to do) something inconsiderate' (cf. 4, number of

occurrences 27; cf. also 5, number of occurrences 18 including also the verbs *umgehen* and *ausweichen* 'to avoid').

- (4) Für manche Länder könnte ein interkulturelles Training vor der Abreise sinnvoll sein. Denn es lauern Fettnäpfchen. (C96/MAI.01836 Computer Zeitung, 23.05.1996, S.1)
'Intercultural training could be useful for some countries before departure since possible screw ups are lurking out there.'
- (5) Grundsätzlich sollte man natürlich immer freundlich und höflich sein. Fettnäpfchen vermeidet man, indem man das Verhalten der Kollegen beobachtet. (R99/NOV.88859 Frankfurter Rundschau, 03.11.1999, S. 5)
'Basically, one should be friendly and polite. In order to avoid screw ups, observe how your colleagues behave.'

To sum up, all occurrences attested in the IDS-Korpus cannot be assigned to the idiomatic phrase *ins Fettnäpfchen treten/tappen*. This means that the contemporary use of *Fettnäpfchen* is not limited to the idiomatic phrase found in dictionaries; instead, this UC appears outside the specific idiomatic phrases discussed hitherto and there are instances in which it occurs as a free lexeme with a more or less specific meaning irrespective of the verbs. Nonetheless, *Fettnäpfchen* can in many cases be paraphrased as 'screw up' or 'mistake'. Moreover, the UC tend to occur more frequently with *treten* than with *tappen*. As the present analysis shows, *Fettnäpfchen* is by no means marginally attested in the IDS-Korpus. To be more precise, *Fettnäpfchen* varies a lot concerning the phrases it can occur in, but the basic meaning of this noun remains in most cases more or less the same.

4.2 Case study 2: *Fersengeld geben*

According to the reference dictionaries, *Fersengeld geben* 'to flee'; lit. 'to give heel money' does not vary and the only verb that can occupy the verb position is *geben* 'to give'. Interestingly, neither this idiomatic phrase nor the lexeme *Fersengeld* itself is included in Schemann (2000). The UC with this specific idiom phrase is attested more than 450 times in the IDS-Korpus. A prototypical occurrence is (1).

- (1) Wie die Polizei berichtet, hatten Zeugen die Polizei gerufen. Als die Beamten dann am Tatort erschienen, hätten die Täter Fersengeld gegeben. Jeder rannte in eine andere Richtung. (R97/NOV.93222 Frankfurter Rundschau, 25.11.1997, S. 1, Ressort: LOKAL-RUNDSCHAU; 18jähriger nach Einbruch in Container festgenommen)
'According to the police, the witnesses had called the police. When the police officers arrived at the scene, the offenders had already fled. Everybody ran away, scattering in different directions.'

As the data retrieved from the corpus illustrates, the idiomatic phrase with this UC is not prone to syntactic variation. Nonetheless, an indirect object is common for phrases with this UC, see (2).

- (2) Die drei Insassen nahmen Reißaus und gaben den Polizisten Fersengeld. Die Beamten riefen Verstärkung und begannen, die Gegend abzusuchen. (R98/DEZ.102803 Frankfurter Rundschau, 21.12.1998, S. 3, Ressort: LOKAL-RUNDSCHAU; Polizeihund spürte Dieb im Gebüsch auf)

'The three car passengers made themselves scarce and fled from the police officers. The police officers called for back up and started to comb the area.'

Fersengeld also occurs with *zahlen* 'to pay' (this is attested seven times in the corpus), for example in (3), although *zahlen* is not listed in any of the reference dictionaries. In a number of occurrences *Fersengeld* can occur without a verb – especially in a headline (see 4).

- (3) Der Bote weigerte sich jedoch seine Barschaft herauszurücken, schrie um Hilfe und wurde darauf mit der Faust ins Gesicht geschlagen. Danach zahlten die Täter Fersengeld und tauchten in der Dunkelheit unter. (M05/MAR.20444 Mannheimer Morgen, 12.03.2005; Duo bedroht Pizzaboten)

'The courier refused to hand over the cash, he called for help and was struck into the face. After that, the delinquents fled and disappeared in the darkness.'

- (4) Vergebliches Fersengeld – Umsonst hat ein jugendlicher Dieb am Dienstag Fersengeld gegeben. Er hatte aus einem Kaufhaus in den K-Quadranten einen Kassettenrekorder und einen Rucksack geklaut. Ein Angestellter hatte den mutmaßlichen Dieb bei der Aktion beobachtet und verfolgte ihn zu Fuß. Als eine Polizeistreife vorbei kam, teilte der Verfolger den Beamten mit, der Dieb sitze an der Straßenbahnhaltestelle Luisenring. Als der Teenager die Polizisten entdeckte, ließ er seine Beute zurück und gab Fersengeld – aber umsonst, die Polizei war schneller. Der junge Mann wurde vorläufig festgenommen. dad/pol Polizeibericht (M04/NOV.78403 Mannheimer Morgen, 04.11.2004; 89-Jähriger mit Trick beklaut)

'The unsuccessful escape – A young thief fled in vain on Tuesday. He had shoplifted a cassette recorder and a backpack from a store. An employee had observed the suspect and followed him on foot. When a police patrol passed by, the pursuer informed the officers that the thief was sitting at a tram stop called Luisenring. When the teenager spotted the police officers, he left his booty behind and fled – but in vain, since the police officers were faster. The young man was arrested.'

In a number of cases the UC *Fersengeld* can occur with a number of different verbs – the verbs are by no means limited to *geben* or *zahlen*. However, not all occurrences of *Fersengeld* (+ verb) can be paraphrased as 'to flee'. The data contains 24 such instances – the following examples (5-7) illustrate the different homonymic meanings of the UC + verb phrase. Depending on the context, *Fersengeld* can also denote 'fund-raising campaign' (cf. 5) and 'compensation' (cf. 7). Provided that the occurrence is not compatible with the semantic paraphrase 'to flee', the context alludes to the respective meaning of the phrase.

- (5) Schüler kassierten viel Fersengeld – Kinder legten sich bei Ruandalauf schwer ins Zeug (RHZ08/JUN.18753 Rhein-Zeitung, 19.06.2008; Schüler kassierten viel Fersengeld)

'Pupils collected lots of money for charity – Children worked flat out for the Ruanda campaign'.

- (6) "Go for Ruanda" – so nennt sich eine Aktion, die von den Initiatoren der Afrika-Woche im Mai des kommenden Jahres laufen wird [sic!]. In einem Sternmarsch sollen sich die Kinder der Rhein-Hunsrücker Schulen ihr Fersengeld verdienen und die Einnahmen für das afrikanische Krisenland spenden. (RHZ98/NOV.02695 Rhein-Zeitung, 05.11.1998; "Fersengeld" für Hilfsprojekte in Ruanda)

"Go for Ruanda" is the name of the project that will take place in May next year. The pupils from the Rhein-Hunsrück schools will earn some money in a march from different starting points and donate the earnings to the developing country in Africa.'

- (7) Wenn das heute so weiter geht, beschwere ich mich noch an der Kasse. Wer weiß, vielleicht bekomme ich ja Fersengeld. (RHZ96/DEZ.10850 Rhein-Zeitung, 17.12.1996; Fersengeld)

'If this today goes on, I will complain at the checkout. Who knows, perhaps I will get compensation.'

As stated in chapter 1, changes in meaning or form are not always found in dictionaries – or at least they are noted quite long after they have entered the language – and case studies discussed above support this observation. On the other hand, the basic meanings of the UCs analysed above and prototypical properties of phrases containing the UCs *Fettnäpfchen* and *Fersengeld* can be gathered from the reference dictionaries consulted in the present study.

In actual language use ambiguity poses no problem at all since the meaning depends on the context (cf. Palm 1994: 435).

The analysis of more than 450 occurrences of *Fersengeld* + verb has yielded – basically – three different meanings. First, *Fersengeld geben* can in many cases be paraphrased as 'to flee'. As we have seen based on a small number of instances (5-7), the string may involve another reading: 'fund-raising campaign' and 'compensation'.

Depending on the context, the meaning of the lexeme *Fersengeld* can vary. Based on instances 5-6 (and other similar to these) we can conclude that *Fersengeld* basically stands for 'a fund-raising campaign' and in 7 for 'compensation/refund'. To be more precise, the UC *Fersengeld* is not as opaque/unmotivated as often postulated. This goes back to some instances found in the IDS-Korpus, for instance in (7). In this case, *Fersengeld* can be divided into two components: *Ferse(n)* + *Geld*; these components can be understood with reference to *subsequent* + *payment*. In fact, a compensation is a kind of subsequent payment. In other words, it is not surprising at all that *Fersengeld* can also denote 'compensation', since *Ferse(n)* can be interpreted in terms of 'posterior' and *Geld* 'money', respectively. Based on this interpretation, the meaning and the motivation of this lexeme might be interpreted as mentioned here¹.

¹ I would like to thank Professor Jussi Niemi for his perceptive remarks concerning the transparency of the components and for many other comments.

5 DISCUSSION

The present study has shown, based on two case studies, that the UCs are not necessarily as unique as often postulated. In short, they can occur with a number of different verbs. As idiomatic phrases in general are said to be fixed without much possibility for variation, the results discussed here are interesting – with the help the large corpus it was possible to document manifold variation in everyday German. A UC can appear outside a specific idiomatic phrase and it can have more than one meaning, depending on the context. Thus, the traditional view on *Fersengeld* was modified. There also seem to be a number of semantically unrelated verbs that occupy the verb slot. However, this is not necessarily systematically – if at all – attested in the contemporary dictionaries of German. Dictionaries provide in general basic information on frequencies, variation and syntax, which is not surprising at all, since they focus on crucial information. Due to the fact that none of the existing German dictionaries of idioms give comprehensive coverage of idiom-variation, the results gained in the present study are not surprising.

On the other hand, the notion of UC can be problematic. The idiomatic phrase *Fersengeld geben* has traditionally been regarded as a prototypical instance of containing a UC. However, as the occurrences retrieved from the IDS-Korpus show, *Fersengeld* is by no means limited to this idiomatic phrase. The question remains to be answered whether *Fersengeld* is a UC at all, although this is often implied in text books and dictionaries (cf. Fleischer 1997 and Duden 2008, for instance). In this respect the current study has provided new information about these lexemes and it has been able to gather insights into actual use of idioms.

In order to gain more detailed facts about unique constituents, questionnaire studies should also be conducted. As stated in the beginning, there is no comprehensive corpus of spoken language in German. Since there is a lack of research on spoken language, this could be the next area of research into the use of idioms. On the other hand, the notion of unique constituent and prototypical examples of it in German should be given more attention – synchronically *Fersengeld* is not as clear-cut case and prototypical of UCs as was in general postulated by many researchers (cf. Fleischer 1997, for instance).

To conclude, the IDS-Korpus can provide detailed information about these UCs. Nonetheless, the results retrieved from the corpus are restricted to written language. The internet can provide some information on the use of idioms, but questionnaires may well be the best way of getting information, especially on less frequent items in contemporary German. In addition, a multi-method approach could shed some light on the properties of UCs.

BIBLIOGRAPHY

Dictionaries and corpus used as reference material

Agricola, Erhard. 1992. Wörter und Wendungen: Wörterbuch zum deutschen Sprachgebrauch. Herausgegeben von Erhard Agricola unter Mitwirkung von Herbert Görner und Ruth Küfner. Überarbeitete Neufassung der 14. Auflage. Mannheim: Dudenverlag.

- Brockhaus-Wahrig. 1980-1984. Deutsches Wörterbuch in sechs Bänden. Herausgegeben von Gerhard Wahrig et al.. Wiesbaden: Brockhaus.
- Duden. 1999. Duden: Das große Wörterbuch der deutschen Sprache in zehn Bänden. Herausgegeben vom Wissenschaftlichen Rat der Dudenredaktion. 3., völlig neu bearbeitete und erweiterte Auflage. Mannheim: Dudenverlag.
- Duden. 2002. Duden: Das große Buch der Zitate und Redewendungen. Herausgegeben von der Dudenredaktion. Redaktionelle Bearbeitung: Brigitte Alsleben, Carolin Mülverstedt und Werner Scholze-Stubenrecht. Mannheim: Dudenverlag, cop.
- Duden. 2003. Duden: Deutsches Universalwörterbuch. Herausgegeben von der Dudenredaktion. 5., überarbeitete Auflage. Mannheim: Dudenverlag, cop.
- Duden. 2008. Duden: Redewendungen: Wörterbuch der deutschen Idiomatik. Herausgegeben von der Dudenredaktion. Redaktionelle Bearbeitung Werner Scholze-Stubenrecht und Wolfgang Worsch. 3., überarbeitete und aktualisierte Auflage. Mannheim: Dudenverlag.
- Friederich, Wolf. 1991. Moderne deutsche Idiomatik: Alphabetisches Wörterbuch mit Definitionen und Beispielen. 2. Auflage. Ismaning: Hueber.
- Grimm, Jacob und Wilhelm Grimm. 1984. Deutsches Wörterbuch. Fotomechanischer Nachdruck der Erstausgabe 1854-1971. München: Deutscher Taschenbuch Verlag.
- Paul, Hermann. 1992. Deutsches Wörterbuch. 9., vollständig neu bearbeitete Auflage. Von Helmut Henne und Georg Objartel unter Mitarbeit von Heidrun Kämper-Jensen. Tübingen: Niemeyer.
- Röhrich, Lutz. 2004. Lexikon der sprichwörtlichen Redensarten. Digitale Bibliothek; 42. Berlin: Directmedia.
- Schemann, Hans. 1992. PONS: Synonymwörterbuch der deutschen Redensarten. Stuttgart: Klett.
- Schemann, Hans. 1993. PONS: Deutsche Idiomatik: Die deutschen Redewendungen im Kontext. Stuttgart: Klett.
- Schemann, Hans. 2000. PONS: Deutsche Redensarten. Stuttgart: Klett.
- Wahrig. 2007. Wahrig digital: Deutsches Wörterbuch. United Soft Media.
<http://www.ids-mannheim.de/cosmas2/> (Viewed 01 March 2010)

References

- Dobrovol'skij, Dmitrij. 1978. Phraseologisch gebundene lexikalische Elemente der deutschen Gegenwartssprache. Leipzig: Universität Leipzig.
- Dobrovol'skij, Dmitrij. 1988. Phraseologie als Objekt der Universalienlinguistik. Leipzig: Enzyklopädie.
- Dobrovol'skij, Dmitrij. 1989. Formal gebundene phraseologische Konstituenten: Klassifikationsgrundlagen und typologische Analyse. Beiträge zur Erforschung der deutschen Sprache, eds. by Wolfgang Fleischer, Rudolf Große, Gotthard Lerchner, 57-78. 9. Band. Leipzig: Bibliographisches Institut.
- Dobrovol'skij, Dmitrij and Elisabeth Piirainen. 1994. Sprachliche Unikalia im Deutschen: Zum Phänomen phraseologisch gebundener Formative. *Folia Linguistica* XXVIII/3-4. 449-471.
- Feyaerts, Kurt 1992. Hebt u ook iets op uw kerfstok? Fraseologisch gebonden woorden in het Duits en het Nederlands. *Handelingen der Koninklijke Zuidnederlandse Maatschappij voor Taal - en Letterkunde en Geschiedenis* 44. 73-87.

- Fillmore, Charles J., Paul Kay and Mary Catherine O'Connor. 1988. Regularity and idiomaticity in grammatical constructions: the case of let alone. *Language* 64. 501-538.
- Fleischer, Wolfgang. 1982. *Phraseologie der deutschen Gegenwartssprache*. Leipzig: Bibliographisches Institut.
- Fleischer, Wolfgang. 1997. *Phraseologie der deutschen Gegenwartssprache. 2., durchgesehene und ergänzte Auflage*. Tübingen: Niemeyer.
- Gehweiler, Elke, Iris Höser and Undine Kramer. 2007. Types of changes in idioms – some surprising results of corpus research. *Idioms and collocations: Corpus-based linguistic and lexicographic studies*, ed. by Christiane Fellbaum, 109-137. London: Continuum.
- Häcki Buhofer, Annelies (unter Mitarbeit von Davide Giuriato). 2002. "Unikalia" im Sprachwandel: phraseologische gebundene Wörter und ihre lexikographische Erfassung. *Phraseologie in Raum und Zeit: Akten der 10. Tagung des Westfälischen Arbeitskreises „Phraseologie/Parömiologie“ (Münster, 2001)*, eds. by Elisabeth Piirainen and Ilpo Tapani Piirainen, 125-160. Hohengehren: Schneider Verlag.
- Langlotz, Andreas. 2006. *Idiomatic Creativity: A Cognitive-linguistic model of idiom-representation and idiom-variation in English*. Amsterdam: Benjamins.
- Moon, Rosamund. 1998. *Fixed expressions and idioms in English: a corpus-based approach*. Oxford: Clarendon Press.
- Naumann, Hans Peter. 1987. Unikale Komponenten im Schwedischen. In: Harald Burger & Rudolf Zett (Hrsg.). *Aktuelle Probleme der Phraseologie. Symposium 27.-29.9.1984 in Zürich*, 151-168. Bern, Frankfurt am Main, New York und Paris: Peter Lang.
- Nenonen, Marja and Jussi Niemi. 1999. Morphological isolates in idioms: cranberries or real words? *Brain and Language* 68. 158-164.
- Palm, Christine. 1994. *Habent sua fata idiomata. Beobachtungen zur Polysemie von Phraseologismen*. *Europhras* 92. *Tendenzen der Phraseologieforschung*, ed. by Barbara Sandig, 431-460. Bochum: Brockmeyer.
- Piirainen, Elisabeth. 1991. *Phraseologismen im Westmünsterländischen. Einige Charakteristika der westmünsterländischen Phraseologie (im Vergleich zum Hochdeutschen)*. *Niederdeutsches Wort* 31. 33-76.
- Weinreich, Uriel. 1980 [1969]. *Problems in the analysis of idioms. On semantics*, ed. by William Labov and Beatrice S. Weinreich. Pennsylvania: University of Pennsylvania Press.

Jacopo Torregrossa

Towards a phonological account of contrast in Italian: a pilot- study on contrastive topics

The idea that contrast is an autonomous informational component on par with topic and focus has been suggested in some recent contributions about information structure. This paper aims to support this idea with arguments of both theoretical and empirical nature. Focusing on the linguistic phenomenon of contrastive topics (CTs), I will show that the complex semantic/pragmatic effects they give rise to can be better understood disentangling CTs into their basic informational constituents, i.e., contrast and topic. In particular, I will analyze the semantic contribution of the former to the overall interpretation of CTs. This will represent the preliminary step for investigating the issue of how contrast is encoded linguistically. To this aim, I set up an experiment in which the phonology and the phonetics of both contrastive and non-contrastive constituents were examined. The preliminary data seem to suggest that contrastive interpretations are expressed by means of prosodic features, in the form of either phonetic or phonological features. From the general design of the experiment it emerges that certain linguistic phenomena could be better explained by relying on the notion of contrast; by this means contrast is shown to be a category of grammar. If the results of this pilot-study will be confirmed by extending the data-set, there will be several implications for the general architecture of grammar and for our understanding of the way in which the different components of language interact.

1 CONTRAST AS AN AUTONOMOUS INFORMATIONAL NOTION

The most striking fact when approaching the literature on information structure is that there is much debate on the linguistic encoding of topic and focus, while the notion of contrast is often derived as a side-effect of their use (see Ward and Prince 1986 as a main reference) or even considered as a category that is relevant only from a cognitive point of view, arising ‘from particular inferences which we draw on the basis of given conversational contexts’ (Lambrecht 1994:290). In opposition to this general tendency, I assume that contrast is an autonomous informational notion (on par with topic and focus) and a linguistically relevant category of grammar. Vallduvì and Vilkuna (1998) are the first to recognize that contrast has a specific

semantics, affecting the logical form of the sentence. In particular, they establish a clear-cut distinction between the notions of *rhematicity* and *thematicity* on the one hand (focus and topic, in my terms) and *kontrast* on the other. The former pertain to the information status of the constituents (they roughly correspond to new and old information, respectively) and to the structural organization of the sentence (they are associated with different instructions for context-updating). On the contrary, the latter has a quantificational (hence, semantic) import: it generates a set of objects of the same semantic type as the constituent with which *kontrast* is associated, such that each object represents an alternative to it. In the example given by the authors, the set associated with the *kontrastive* constituent Bill in (1) is $M = \{\text{Bill, Carl, Mark}\}$: the objects of M are all individuals and, apparently, the set is built up on the basis of contextual information.

- (1) John introduced BILL to Sue.

Italian data show that a contrastive interpretation could be associated with different positions in the sentence structure, e.g., in correspondence with constituents appearing in the left-periphery of the sentence (as in (2) and (3)) or with constituents which are in-situ (as in (4)).

- (2) A: Come hai conosciuto quelle ragazze? (How did you meet those girls?)
 B: *Michela*, l'ho conosciuta in ascensore, mentre *Gianna*, l'ho conosciuta al mare.
 (*Michela*, I met in the elevator, while *Gianna*, I met at the beach).
- (3) Mario ha incontrato Luca. Anzi no, *Gianni* ha incontrato. (Mario met Luca. No, it was *Gianni* that he met).
- (4) A: Hai invitato i tuoi fratelli alla festa? (Did you invite your siblings to the party?) B: Ho invitato *Marco*, ma ho deciso di non invitare *Gianni*. (I invited *Marco*, but I decided not to invite *Gianni*).

In all these cases, the constituents endowed with a contrastive interpretation evoke some alternatives, all belonging to a contextually given set, in line with Vallduvì and Vilkuna's generalization.

Once contrast has been identified as an autonomous category of grammar, one should investigate if it is marked by specific linguistic means, e.g., syntactic or phonological devices. Vilkuna (1995) claims that in Finnish contrast is encoded syntactically: contrastive constituents (both topic and focus) are assumed to move to a specific syntactic projection in the left-periphery of the sentence (ContrP). Italian seems to behave differently: contrastive constituents could appear in-situ as well, as (4) above shows: the movement to the left-periphery might, thus, be triggered by some other semantic feature (e.g., a topic feature expressing the familiarity in discourse of the denoted entity). Thus, the understanding of how contrast is linguistically encoded in Italian deserves further study. In this paper, I try to investigate the hypothesis that contrast is expressed through prosodic features, either in the form of phonological features (e.g., the use of specific accent-types) or of phonetic features (e.g., variation in the range or in the duration of the tonal vowels). The idea that contrast triggers specific prosodic effects is suggested by some investigations on the difference between informational foci and contrastive

foci in their phonological and phonetic marking. For example, Avesani and Vayra (2004) notice that contrastive foci are realized through complex accents (either H+H* or L+H*) which significantly differ from those associated with informational foci (i.e., H+L*). This generalization seems to apply cross-linguistically, as the contributions of Jun (1998) on Korean, Frota (2000) on Portuguese and Selkirk (2002) on English clearly show. This paper aims to investigate the issue of the linguistic encoding of contrast through a different way: it will compare the linguistic realization of contrastive and non-contrastive topics, thus taking into account the interaction between contrast and the informational notion of topic (which is still an unexplored field, at least as far as Italian is concerned. For a similar study on German, see Braun and Laid 2003) and it will be a preliminary step in analyzing which prosodic effects it triggers whenever it appears in isolation, i.e., whenever it is associated neither with a topic nor with a focus, which would provide crucial empirical evidence in favor of the assumption that contrast is an autonomous informational notion. The overall study design is built around the hypothesis that contrast is prosodically marked. In the next sections I will introduce the general experimental procedure which has been used to support this hypothesis with empirical evidence. Before doing this, however, it is necessary to spend some words on the linguistic nature of Italian contrastive topics (CTs, henceforth).

2 ITALIAN CONTRASTIVE TOPICS: A BRIEF INTRODUCTION

The aim of this section is threefold: a) to draw some generalizations on the linguistic nature of CTs; b) to show how this linguistic nature emerges from the interaction of the informational notions topic and contrast, considered independently from each other; c) to hint at the benefits of this analysis for the understanding of the phenomenon with respect to other analyses proposed in the last few years.

The italicized constituents in (2) above and in (5) and (6) here below are examples of CTs in Italian:

- (5) A: Mario ha invitato i suoi fratelli al matrimonio? (Did Mario invite his brothers to the wedding?)
 B: *Francesco*, l'ha invitato. (*Francesco*, he invited).
- (6) A: Abbiamo portato nostro figlio a vedere gli animali in fattoria. *I cavalli*, li adora, ma *i cani*, lo impauriscono. (We brought our son to the farm to see the animals there. *Horses*, he adores, but *dogs* scare him).

From a syntactic point of view, the constituents sit in the left-periphery of the sentence and are resumed by a clitic which is sentence-internal. As for their semantics, they are active in discourse in the sense of being 'either already evoked in the discourse or else in a salient set relation to something already evoked or inferable from the discourse' (Prince 1983:4): in the cases at stake here the relevant sets are *being Mario's brother* in (5) and that denoted by the property *being a farm*

animal in (6). It is crucial to notice that non-contrastive topics share the same linguistic behavior: the constituent in (7) is left-dislocated and it is active in discourse in Prince’s sense.

- (7) A: Mario ha invitato Francesco al matrimonio? (Did Mario invite his brothers to the wedding?)
 B: Sì, Francesco, l’ha invitato. (Yes, Francesco, he invited).

Returning to the examples (5) and (6), it is clear that CTs perform, furthermore, the function of evoking alternatives in discourse: in (6) the phrase *i cavalli* in the first sentence indicates the presence of the alternative topic in the second one, i.e., *i cani*. Crucially, the two alternatives are drawn from the contextually salient set and different properties are predicated of each of them in turn. The case in (5) is more complex in this regard, since no alternative to the topic *Francesco* is explicitly mentioned in the following discourse. However, one has still the impression that the property predicated of the CT does not hold for the other alternatives which are members of the relevant set.

The notion of CT could, thus, be disentangled into two notions, each with its own semantics, i.e., topic and contrast. Topichood is marked by means of syntactic movement to the left-periphery of the sentence. What follows aims to investigate the specific linguistic encoding of contrast. Before doing this, however, it is important to spend some words on the benefits that this analysis based on the disentanglement of the notion of CT could bring to the understanding of the phenomenon. Across literature on information structure, there have been several attempts to understand which components of language interact with each other for the expression of the complex nature of CTs described thus far. For example, some theoretical approaches developed in the last few years within the cartographic framework have proposed to identify CTs by means of positional parameters.

Benincà and Poletto (2004) and Frascarelli and Hinterhölzl (2007) assume that there is a specific functional projection in the left-periphery of the sentence which hosts CTs. More generally, they reject the idea that topic projections are freely recursive in the left-periphery, occurring both higher and lower with respect to a single focus projection (see Rizzi 1997 and the representation in a. below). Instead, every topic projection encodes a different kind of topic, each having a specific semantic interpretation (e.g., aboutness topic, contrastive topic, familiar topic; see b. below), in line with the cartographic assumption concerning the one-to-one correspondence between syntactic positions and semantic interpretations.

- | | |
|---|---|
| a. [TopP TOPIC* [FocP [TopP TOPIC* [IP | (Rizzi’s split-CP hypothesis) |
| b. [AboutnessP [ContrP [FocP [FamP* [IP | (Frascarelli and Hinterhölzl split-CP hypothesis) |

The authors’ generalizations are drawn from the analysis of a plethora of empirical observations: in particular, Frascarelli and Hinterhölzl work with large corpora of spoken data. However, there are some examples that do not fit well within their proposal of a rigid hierarchy of topic projections. The sentences (9) and (10) are both well-formed answers to the question in (8). The constituents *a Maria* (to Maria) and *le ortensie* (the hydrangea) alternate in their positions in the

CP, despite their being instances of different types of topic. The cartographic analysis would be taken to apply in this case only if the two constituents had a different semantic interpretation in the two sentences, which appears unlikely, since the presence of the context in (8) prevents any kind of interpretive ambiguity.

- (8) Quale tra questi fiori possiamo regalare a Maria? (Which of these flowers could be given to Maria as a present?).
- (9) A Maria le ortensie, gliele regalai l'anno scorso e le rose due anni fa. Quest'anno potremmo regalarle dei ciclamini.
(To Maria the hydrangea, I gave as a present last year and the roses two years ago. This year we can give her cyclamen flowers).
- (10) Le ortensie a Maria, gliele regalai l'anno scorso e le rose due anni fa. Quest'anno potremmo regalarle dei ciclamini.
(The hydrangea to Maria, I gave as a present last year and the roses two years ago. This year we can give her cyclamen flowers).

Examples like (9) and (10) seem to suggest that a syntactic approach alone cannot account for the complex nature of CTs. However, once the notions of topic and contrast are considered independently from each other, there is no need to assume the existence of a CT projection: topic projections can be seen as freely recursive in the left-periphery of the sentence (as in Rizzi's original proposal) and contrast could in principle be associated with each of them in turn, providing its semantic contribution and affecting their linguistic realization.

On the other hand, there are scholars who completely set aside syntactic considerations, relying on other criteria for the identification of CTs. For example, Büring (2003) focuses on their realization in English and proposes to identify them by means of prosodic cues: they bear a B-accent, i.e., accents realized as L+H* and followed by a default low tone and a high boundary tone (L H%). On the contrary, foci are realized by means of A-accent (plain high tones H*). According to Büring, the use of a CT marks a strategy of answering a super-question presupposed in the ongoing discourse. For example, (11) presupposes the super-question in (12) and could be uttered by means of two different prosodic patterns, i.e., either A+B or B+A. The choice of one or the other depends on which strategy is adopted to approach the issue raised by the super-question: in the former case, the speaker answers going by food and, consequently, *the beans* is promoted to CT status; in the latter, going by person; accordingly, *Fred* is marked as a CT. Phonology is the only way to express CTs. A consequence of this assumption is that there is no difference in the interpretation of (11), in the case it is realized with the accentual pattern A+B, and (13): the constituent *the beans* bears the B-accent in both cases. However, it is evident that the left-dislocation in (13) has the role of marking the constituent *the beans* as active in discourse, which in my system corresponds to the function of the topic. Therefore, I assume that B-accent does not mark CTs, but contrastive interpretations, both in-situ (as in (11)) and in left-dislocated position (as in (13)). This allows us to get the difference between (11) and (13).

- (11) Fred ate the beans.
(12) Who ate what?

(13) The beans, Fred ate.

It is important to notice that the association of a contrastive interpretation with an in-situ constituent or with a left-dislocated one will be crucial for the general design of the experiment presented in the next sections.

To conclude, the discussion of the two approaches to CTs discussed so far shows that the linguistic nature of CTs is too complex to be explained by means of univocal criteria, either syntactic (see cartography) or phonological (see Büring 2003). In the following I will focus on the informational notion of contrast investigating how it is encoded in Italian (at least in the variety I consider here).

3 METHODS

The following experiment was designed to investigate if it is prosodic features that differentiate between contrastive and non-contrastive constituents, turning out to represent the specific means of expressing the informational notion of contrast. I constructed five different paragraphs, each containing four sentences. The sentences involved minimal pairs differing with respect to two conditions, i.e., the position (in-situ or left-dislocated) and the interpretation (contrastive or non-contrastive) of the object constituent. On the one hand, the contrastive interpretation was obtained by introducing a relevant set into the discourse context and then creating an opposition between two of its members (one of which was the target object constituent) with respect to a certain property. In order to avoid any kind of ambiguity, both the objects were explicitly mentioned in discourse. On the other hand, to avoid triggering any kind of contrastive interpretation, I introduced some objects in the context and the object constituent in the target sentence was such that it denoted the corresponding superset. The following is one of the short paragraphs used for the experiment. Context A induces contrast and sentence (1) and (2) differ with respect to the position of the object constituent (left-dislocated and in-situ, respectively). On the contrary, context B does not induce any kind of contrast; sentences (3) and (4) differ again with respect to the position of the object constituent (left-dislocated and in-situ respectively):

Context A: Non amiamo tutte le razze di cani. (We do not appreciate every kind of dog).

- (1) *I maremmani* li adoriamo, mentre i carlini non ci piacciono affatto. (*Maremmas*, we adore, while pugs, we don't like them at all).
- (2) Adoriamo *i maremmani*, mentre i carlini non ci piacciono affatto. (We adore *maremmas*, while pugs, we don't like them at all).

Context B: Abbiamo adottato due maremmani. (We adopted two Maremmas).

- (3) *I maremmani*, li adoriamo. (*Maremmas*, we adore)
- (4) Adoriamo *i maremmani*. (We adore *maremmas*).

Ten Italian speakers (4 male and 6 female) between the ages of 21 and 35 took part in the experiment. They were all students or researchers at the University of Verona. The dialect origin

of the speakers was fully controlled, in order to exclude any potentially intruding variable that could distort the results of the experiment: they all spoke a Western Veneto Italian variety. The subjects were seated in a sound-proof room at the University of Verona and were asked to read the sentences in each paragraph aloud. I, thus, built up a corpus of two hundred sentences, which constitute my data-set. The PRAAT program was used to record directly into a computer and to conduct the analysis. After deriving the pitch-track of the target sentences, I labeled the object constituent according to the accent-type with which it was realized, following a TOBI annotation system: high (H) or low (L) tones were assigned both to syllables bearing the main stress and to those marking the boundaries of prosodic constituents. If it appeared that a single accent-type was associated with the same syntactic position (hence, with two different interpretations), I considered the sector's segmental level by referring to the spectrogram. In particular, the value of the pitch range (i.e., the difference between the maximal and minimal F0-value in the sector) and the length of the tonal vowel were measured and the results were analyzed statistically, applying a t-test.

4 RESULTS OF THE EXPERIMENT

In this section I will illustrate the results of the experiment. For the sake of clarity in the exposition, the section has been split into two main parts: in the former, I will show which accent-types are associated with specific syntactic positions (with focus on left-dislocated constituents) or with certain interpretations (with focus on contrastive interpretations). In the latter, I will deal with those subjects that tend to associate a single accent-type to the same syntactic position, in order to investigate whether contrast is marked by means of phonetic parameters in these cases.

4.1 Phonological choices

The purpose of the phonological analysis was to investigate whether the type of accent which was used is determined by the position of the object constituent (in-situ vs. left-dislocated) or by its interpretation (contrastive vs. non-contrastive). Here I will focus on the accent-type associated with left-dislocated constituents. It is well-known that left-dislocated items are prosodically set off in a separate intermediate phrase (see Avesani 1990 on this issue). Figure 1 lists the types of accents which are associated with left-dislocated constituents (both contrastive and non-contrastive) along with their frequency. There is a strong tendency to mark the end of the intermediate phrase by means of high boundary tones (following either L or H tonal accents). The use of one accent-type or the other is quite consistent among speakers, as Figure 2 shows.

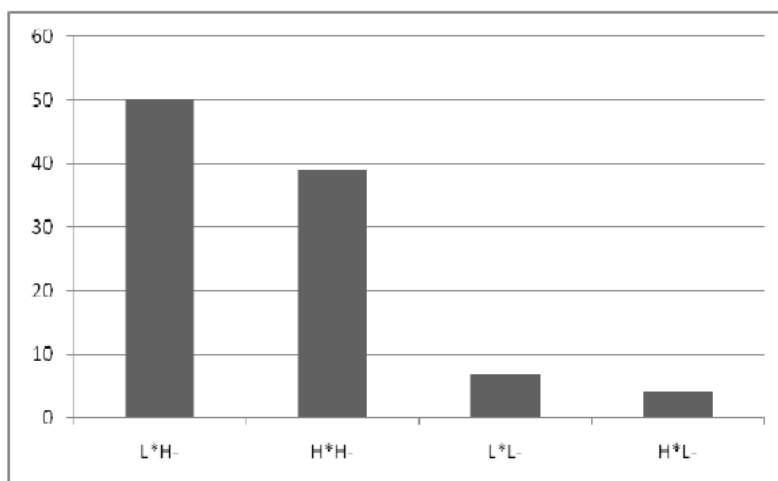


Figure 1: Frequency of accent-types used in association with left-dislocated constituents.

	F1	F2	F3	F4	F5	F6	M1	M2	M3	M4
L*H-	6	9	9	8	0	0	1	9	8	0
H*H-	4	1	1	1	10	10	9	0	1	2
L*L-	0	0	0	1	0	0	0	0	0	6
H*L-	0	0	0	0	0	0	0	1	1	2

Figure 2: Frequency of use of specific accent-types by the speakers.

The histogram in Figure 1 does not take into account the interpretation (contrastive vs. non-contrastive) of the left-dislocated constituents. It is remarkable, however, that speakers use the same accent-type in both contrastive and non-contrastive environments in 86 percent of cases. For example, Figure 3 and 4 show the prosodic contour of sentences containing a left-dislocated object with a contrastive and a non-contrastive interpretation, respectively. The two sentences were uttered by a female speaker (F6). In both cases, the phrase is constituted by a high tone on the accented syllable (H*) which in turn is followed by a high boundary tone (H-).

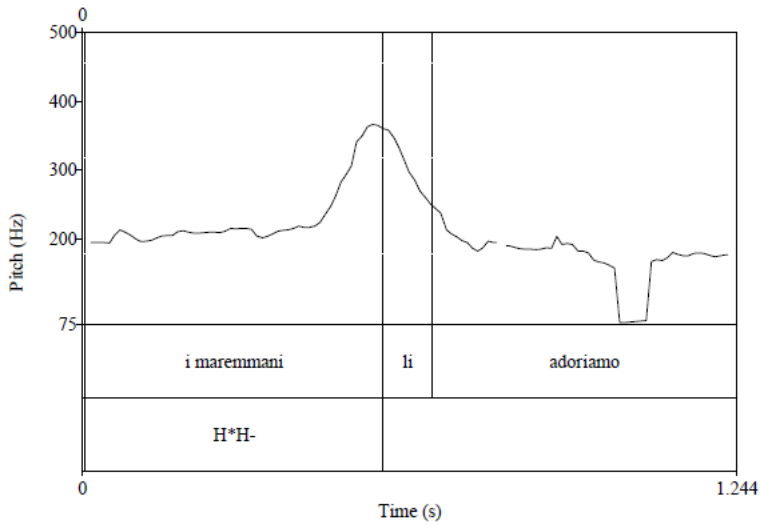


Figure 3: Left-dislocated constituent with a contrastive interpretation. (*I maremmani, li adoriamo*)

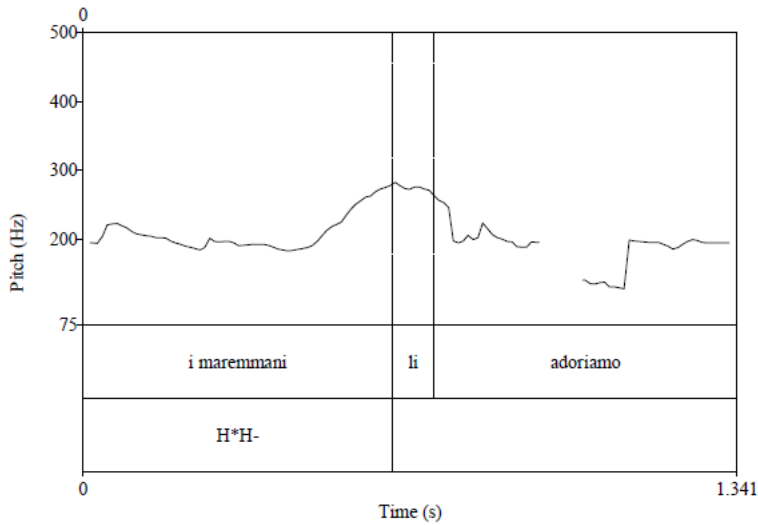


Figure 4: Left-dislocated constituent with a non-contrastive interpretation. (*I maremmani, li adoriamo*)

These data suggest that there is a systematic correlation between accent-types and syntactic positions, at least as far as left-dislocations are concerned: the type of accent is determined by the syntactic position of the constituent, while the interpretation seems to play no significant role. The analysis of the next section aims to investigate whether it is gradient phonetic features that differentiate between contrastive and non-contrastive interpretations in all these cases. Before

doing this, it is worth mentioning another interesting pattern which emerges from the analysis. There are some cases in which the same accent-type occurs both in-situ and in left-dislocated position, whenever the corresponding constituents are endowed with a contrastive interpretation. Crucially, the in-situ constituents with a non-contrastive interpretation are realized by means of different accent-types, suggesting that syntax does not play any role here (which follows an opposite trend with respect to what has been said so far). Figures 5 and 6 exemplify this pattern: the contrastive constituents (both in-situ and in left-dislocated position) bear a high peak on the stressed syllable (H*) followed by a high boundary tone (H-).

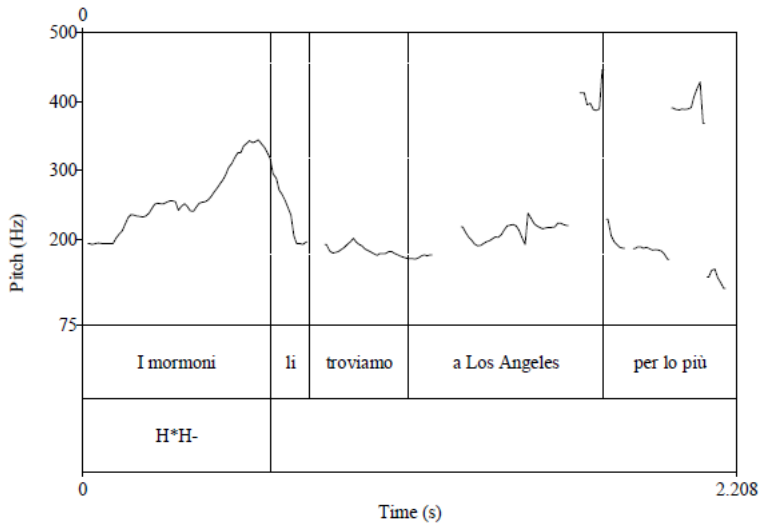


Figure 5: Left-dislocated constituent with a contrastive interpretation (F6). (*I mormoni, li troviamo a Los Angeles per lo più* – *The Mormons, we find in Los Angeles mostly*).

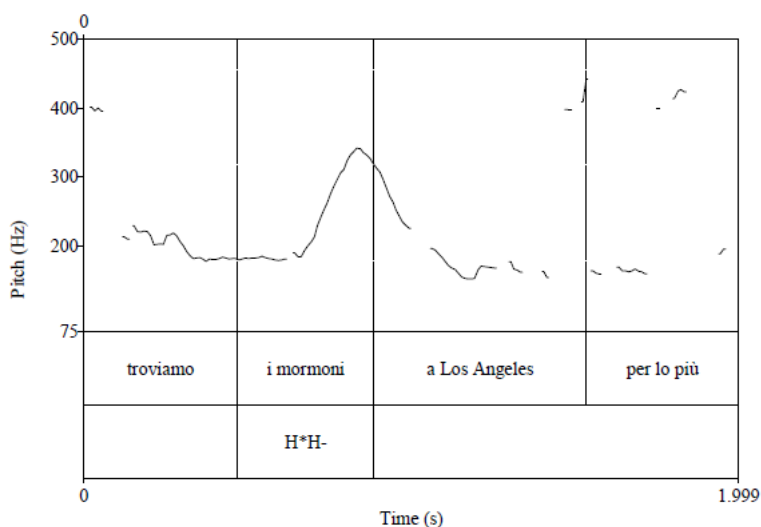


Figure 6: *In-situ constituent with a contrastive interpretation (F6). (Troviamo i mormoni a Los Angeles per lo più – We find the Mormons in Los Angeles mostly.)*

Though not statistically significant, examples like these suggest that speakers might resort to another strategy, involving some kind of correlation between phonological choices and semantic interpretations, instead of the afore-mentioned correspondence between phonological choices and syntactic positions. I leave the issue open for further research: in particular, once the data-set will have been extended, it will be interesting to investigate what conditions lead to exploit this second strategy.

4.2 Phonetic choices

The aim of the phonetic analysis was to determine whether the accents associated with the same syntactic position (in both contrastive and non-contrastive contexts) differ in some aspect of their prosodic realization which does not involve discrete phonological categories, but gradient phonetic parameters. Figure 7 shows that there is a significant tendency for the accents to be realized by means of higher ranges when they are associated with a contrastive interpretation (two sample t-test, $t=3.38$, $d.f.=84$, $p<0.05$).

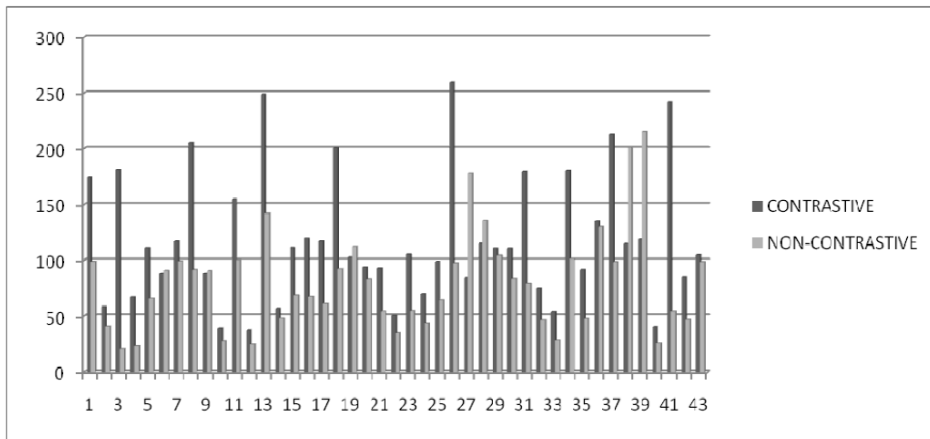


Figure 7: Range (Hz) of the left-dislocated constituents in contrastive and non-contrastive contexts.

As for the duration of the tonal vowel, the difference between contrastive and non-contrastive contexts turns out to be significant in seven speakers (two-sample t-test, $t=2.15$, $d.f.=60$, $p<0.05$). Interestingly, the other three speakers are those for which the values of the range are the highest. This parameter has been measured also with respect to the articulation rate of the sector to which the vowel belongs: the values of the ratio between the duration of the vowel and that of the sector differ significantly according to the interpretation – contrastive vs. non-contrastive (two sample t-test for means, $t=2.47$, $d.f.=30$, $p<0.05$).

On conclusion, it seems that phonetic variations play a role for the expression of contrast, especially in those cases where phonological choices do not seem to convey any difference in interpretation.

5 DISCUSSION AND CONCLUSIVE REMARKS

The present study was based on the assumption that contrast is an autonomous informational notion and a linguistically relevant category of grammar. In most literature concerning information structure, contrast has been considered in association with either topic or focus. In this paper, I decided to concentrate both on those cases in which contrast appears in isolation (i.e., in association with neither topic nor focus) and (mostly) on the phenomenon of CTs, showing how their linguistic nature emerges from the interaction of the informational notions of topic and contrast, considered independently from each other. By doing so, I also pointed out to the benefits of my analysis for the understanding of the phenomenon with respect to other analyses proposed in the last few years. In particular, I aimed to disentangle CTs into their two main informational constituents, i.e., topic and contrast, and, once contrast had been isolated as an autonomous informational component, I investigated the hypothesis that prosody is the linguistic means used to express it.

My exploratory data are consistent with this hypothesis: the results show that contrast is expressed by means of prosodic features, in the form of either phonetic or phonological features.

It has been shown that contrast is expressed by means of differences in the phonetic realization, whenever the left-dislocated constituents bear the same type of accent in both contrastive and non-contrastive contexts. This first result hints at the existence of a one-to-one correspondence between accent-types and syntactic positions. This is compatible with the hypothesis that the phonology of a sentence is determined by its syntactic representation (this corresponds to what Selkirk calls the *direct access hypothesis* - see Selkirk 2002 for a brief overview on different hypotheses on the syntax-phonology interface). However, the phonetic component responsible for the expression of contrast appears to be independent of syntax, as far as two different phonetic realizations correspond to the same syntactic position. This result would represent empirical evidence in favor of Selkirk's *prosodic structure hypothesis*, according to which some aspects of the phonological representation of a sentence stem from the direct interaction with the syntactic component (e.g., the organization into phonological words and phrases), while others reflect the way the prosodic constituent is organized internally. Among the latter, Selkirk includes the phonetic representation. The correlation between the contrastive interpretation and the presence of certain phonetic features indicates some kind of correlation between the semantic and the phonological component.

While discussing the results, I also hinted at a second marginal strategy, according to which speakers tend to express the same semantic interpretation by means of the same accent-type, which follows an opposite trend with respect to what has been discussed thus far. Syntax does not seem to play any role in this second case: both left-dislocated and in-situ constituents bear the same interpretation and accent-type, despite occupying different positions in the sentential structure. This result appears to especially challenge some recent theories formulated within the cartographic framework (see section 2 for further details). The methodological distinction between the informational notions of topic and contrast reveals useful to understand which linguistic means are made use of for expressing each of them: topic is marked through syntax, while contrast through phonology. Contrast does not seem to be encoded by means of positional parameters and, thus, there is no need to assume the existence of a CT projection (see section 2 for further details). What has been said so far is consistent with Rizzi's (1997) initial proposal on the recursivity of topic projections, each corresponding to the same semantic interpretation. CTs emerge from the interaction of contrast with one of these topic projections. More generally, it seems plausible that contrast associates with syntactic phrases in different positions of the sentential structure. Also the data relative to the second strategy clearly emphasize the need to explore the possibility that the semantic and the phonological components interface each other. The prosodic features mentioned thus far (in the form of both phonetic and phonological features) play a significant role for the expression of specific semantic interpretations, without any apparent mediation of the sentential syntactic structure. However, other alternative analyses which remain in line with the traditional model of grammar are worth exploring, e.g., the possibility that contrast is encoded by means of syntactic positions which occur recursively across the syntactic structure of the sentence. Throughout the discussion, I managed to provide some insights into the implications of the results of my analysis for general architectural issues. However, to support one or the other hypothesis, other empirical data need to be taken in consideration. In particular, I intend to extend my analysis in relation to the comparative dimension of language, each time isolating contrast as an autonomous informational notion and

investigating how it is expressed linguistically. Italian is likely to make use of prosodic features; Finnic and Hungarian seem to rely on syntax (see Vallduvi and Vilkuuna 1998). It could be the case that other languages resort to morphology. Both the structures and the informational categories seem to be universal. The way they pair-up gives rise to cross-linguistic variation.

BIBLIOGRAPHY

- Avesani, Cinzia. 1990. A contribution to the synthesis of Italian intonation. *Proceedings of the 1990 International Conference on Spoken Language Processing (ICSLP-90)*, vol. 1, 834-836. Kobe, Japan.
- Avesani, Cinzia, and Mario Vayra. 2004. Focus ristretto e Focus contrastivo in Italiano. *Atti del Convegno "Il parlato italiano"*, ed. by Federico Albano Leoni, Francesco Cutugno, Massimo Pettorino, Renata Savy, 1-20. M. D'Auria Editore.
- Beckman, Mary E., and Janet B. Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology* 3.255-309.
- Benincà, Paola, and Cecilia Poletto. 2004. Topic, Focus and V2: defining the CP sublayers. *The Structure of CP and IP. The Cartography of Syntactic Structures*, vol.2, ed. by Luigi Rizzi, 52-76. Oxford: Oxford University Press.
- Braun, Bettina, and Robert D. Ladd. 2003. Prosodic correlates of contrastive and non-contrastive themes in German. *Proceedings of the 8th European Conference on Speech Communication and Technology (EUROSPEECH-2003)*, 789-792. Geneva, Switzerland.
- Büring, Daniel. 2003. On D-trees, beans and B-accents. *Linguistics and Philosophy* 26.511-545.
- Frascarelli, Mara, and Roland Hinterhölzl. 2007. Types of topics in German and Italian. *On Information Structure Meaning and Form*, ed. by Susanne Winkler and Kerstin Schwabe, 87-116. Amsterdam/Philadelphia: John Benjamins.
- Frota, Sonia. 2000. *Prosody and focus in European Portuguese*. New York: Garland.
- Jun, Sun-Ah, and Hyuck-Joon Lee. 1998. Phonetic and phonological markers of contrastive focus in Korean. *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP-98)*, vol.4, 1295-1298. Sydney: ASSTA.
- Krifka, Manfred. 2001. Quantifying into question acts. *Natural Language Semantics* 9.1-40.
- Lambrecht, Knud. 1994. *Information Structure and Sentence Form. Topic, Focus and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press.
- Molnár, Valéria. 2002. Contrast – from a contrastive perspective. *Information Structure in a Cross-Linguistic Perspective*, ed. by H.Hallelgard, Stig Johansson, Bergljot Behrens and Cathrine Fabricius-Hansen, 147-161. Amsterdam/New York: Rodopi.
- Prince, F. Ellen. 1983. A comparison of Topicalization and Left-dislocation in discourse. Paper presented at the Annual Meeting of the LSA.
- Rizzi, Luigi. 1997. The fine structure of the left-periphery. *Elements of Grammar. Handbook in Generative Syntax*, ed. by Liliane Haegeman, 281-337. Dordrecht: Kluwer.
- Selkirk, Elisabeth. 2002. The syntax-phonology interface. *International Encyclopedia of the Social and Behavioral Sciences*, ed. by Neil J. Smelser and Paul B. Baltes, 15407-15412. Oxford: Pergamon.

- Vallduvì, Enric, and Maria Vilkuna. 1998. On rheme and kontrast. *Syntax and Semantics. The Limits of Syntax*, vol.29, ed. by Peter Culicover and Louise McNally, 79-108. San Diego, CA: Academic Press.
- Vilkuna, Maria. 1995. Discourse configurationality in Finnish. *Discourse Configurational languages*, ed. by Katalin É. Kiss, 244-268. New York: Oxford University Press.
- Ward, L. Gregory and Ellen F. Prince. 1991. On the topicalization of indefinite NPs. *Journal of Pragmatics* 16.167-177.

ACKNOWLEDGMENTS

Different versions of the material in this paper were presented at the Live Meeting (University of Verona), at the 24th Scandinavian Conference of Linguistics (University of Joensuu) and at the 4th Workshop on Prosody, Syntax and Information (University of Delaware). I thank all the participants of these events for their stimulating remarks. I am also deeply in debt to Birgit Alber, Barbara Gili Fivela and Stefan Rabanus for their helpful suggestions during the development of the analysis. Last (but not least), I am grateful to all the subjects that took part in the experiment for their patience and encouragement. Finally, I thank Erica Cei for reviewing a preliminary draft of the paper. Needless to say, I am responsible for all errors and shortcomings.

Kendra Willson

Grammaticalization, productivity and analogy in Finnish converbs

The term *converb* has been used especially since Haspelmath and König (1995) for a variety of adverbial adjuncts based on non-finite verb forms. Such constructions are also called adverbial participles, gerunds and absolute constructions. Languages vary typologically in the extent to which they make use of converb constructions as well as in the number of morphologically distinct converb constructions and their degree of specialization (cf. Nedjalkov 1995: 100, König 1995: 58; Ylikoski 2000: 228). Ample use of converbs is typologically characteristic of many Uralic languages (Comrie 1988: 475), and Finnish has a relatively rich converb system. Nikanne's (1997: 338) examples, shown in (1), present a "minimal sextuplet" showing the range of Finnish converb constructions:¹

- (1) a. Pekka tek-i rikokse-n **juo-tu-a=an olut-ta**.
Pekka do-PST crime-GEN drink-PART-PRTV=Px.3 beer-PRTV
'Pekka committed a crime having drunk beer'
- b. Pekka tek-i rikokse-n **juo-de-ssa=an olut-ta**.
Pekka do-PST crime-GEN drink-INF-INNESS=Px.3 beer-PRTV
'Pekka committed a crime while drinking beer'
- c. Pekka tek-i rikokse-n **juo-de-n olut-ta**.
Pekka do-PST crime-GEN drink-INF-INSTRTV beer-PRTV
'Pekka committed a crime, drinking beer'
- d. Pekka tek-i rikokse-n **juo-da-kse=en olut-ta**.
Pekka do-PST crime-GEN drink-INF-TRANSL=Px.3 beer-PRTV
'Pekka committed a crime in order to drink beer'
- e. Pekka tek-i rikokse-n **juo-ma-lla olut-ta**.
Pekka do-PST crime-GEN drink-INF-ADESS beer-PRTV
'Pekka committed a crime by drinking beer'

¹ I gloss the examples in order to show their morphological structure rather than treating converbs as single units. I choose to treat the person-number elements as clitics rather than suffixes, unlike *ISK*, but this is not important for the argument.

- f. Pekka tek-i rikokse-n **juo-ma-tta olut-ta**.
 Pekka do-PST crime-GEN drink-INF-ABESS beer-PRTV
 'Pekka committed a crime without drinking beer'

Finnish converbs form a subset of the non-finite constructions (complements, adjuncts and periphrastic constructions) collectively known as *lauseenvastikkeet*, "sentence equivalents" or "contracted sentences" (on which a classic work is Ikola 1974). Wiik (1981) analyzes overt and covert assumptions about "equivalence" in the way the term is used over the course of the 20th century. Nikanne (1997) criticizes the concept of "sentence equivalent" as being predicated on generative ideas of transformation. While the description of non-finite constructions as "transformed" finite clauses found support in generative linguistics (cf. Hakulinen and Karlsson 1979), its roots are already present in nineteenth-century grammars that shaped the modern written language. *ISK* (2004: section 876) regards *lauseenvastike* as a pedagogical but not a linguistic concept.

Lindén (1961, 1966) and Herlin and Kotilainen (2004, 2005) have discussed ways in which the development of written Finnish has led to a convergence in syntactic structure and pragmatic function between one converb type (the ESSA or first temporal construction, (1)b above) and a finite clause.

I have observed a similar transformation in the TUA or second temporal construction (1a) (Willson, forthcoming). In modern standard Finnish, the construction can admit an arbitrary number of modifiers, which appear in the same linear order as in a finite clause, as seen in (2). The genitive argument preceding the participle is identified with the subject of the verb represented by the participle, independent of its semantic role. A subject identical to that of the matrix clause must be expressed through a possessive suffix, as in (3). (Examples from Hakulinen and Karlsson 1979: 388.)

- (2) **Tuule-n revi-tty-ä kato-n** tuva-ssa palel-t-iin
 wind-GEN tear-PTCP-PRTV roof-GEN cottage-INESS freeze-PASS-PST
 'after the wind had torn the roof, people in the cottage were cold'
- (3) **Juo-tu-a=a-n maido-n** lapsi rauhoittu-i
 drink-PTCP-PRTV=P_x.3 milk-GEN child calm-PST
 'having drunk the milk, the child calmed down'

In traditional dialect usage, the TUA construction admits at most one argument, which precedes the participle (a small-clause like structure). In most cases this argument is the semantic theme; it may correspond to the subject of a finite clause that paraphrases the non-finite construction, as in (4), or to its object, as in (5). (Examples from the Morphological Archive of Finnish Dialects (MA).)

- (4) mi-tä si-tä **lehmä-n ehrü-ttü-ä** voi muu-ta teh-räk
 what-PRTV DEM.PRTV cow-GEN dry-PTCP-PRTV can other-PRTV do-INF
 kul laka-tar ruakki-ma-sta ja, anta-av vaiv vet-tä
 than stop-INF feed-INF-ELA and give-INF just water-PRTV

'What can one do when the cow has gone dry besides stop feeding and give just water?'
(Laihia)

- (5) hethäm minä **kirjee-n soa-tu-va=n** läh-i-n
immediately I letter-GEN get-PTCP-PRTV=Px.1SG leave-PST-1SG
'immediately I, having gotten the letter, left' (Saarijärvi)

The identification between the genitive modifier preceding the participle and the subject of the verb represented by the participle is part of the general pattern of genitive subjects for modal and non-finite constructions in Finnish, enhanced by standardization (cf. Jaakola 2004: 240); a change from semantic to syntactic distributional criteria is observed elsewhere in the standardization of Finnish (Laitinen 1997, 2004). This has also affected the distribution of possessive suffixes: the distributional criterion changed from discourse anaphora with some semantic restrictions to syntactic binding (cf. Lindén 1959).

Lindén (1961) and Herlin and Kotilainen (2004, 2005) also discuss functional changes in the ESSA construction. In pre-norm usage as seen in dialect archives the ESSA construction is used as a "temporal anchor" that does not introduce new information, and it generally has at most one argument, whereas in modern written Finnish converb constructions can take an arbitrary number of modifiers and regularly contain new information. Lindén (1961: 202) gives the example shown in (6), from Ilmari Kotimäki's *Ankara puutarhuri* (1956).

- (6) **Toimi-e-ssa=an 1840=luvu-n keski vaihee-lla Kuopio-n**
work-INF-INESS=Px.3 1840=period-GEN central period-ADDES Kuopio-GEN
maa=seurakunna-n kappalaise-n K.A. Granit-in perhee-ssä koti=opettaja-na
land-parish-GEN chaplain-GEN K.A. Granit-GEN family-INESS home-teacher-ESS
nuori ylioppilas August Ahlqvist merkits-i päivä=kirja-nsa
young high.school graduate August Ahlqvist note-PST day-book[GEN]=Px.3
lehd-i-lle miette-i-tä=än, vaikutelm-ia=an ja
page-PL-ALL thought-PL-PRTV=Px.3 impression-PL-PRTV=Px.3 and
tulevaisuu-den=suunnitelm-ia=an
future-GEN=plan-PL-PRTV=Px.3
'While working in the mid 1840s as a home tutor in the family of rural chaplain K.A. Granit, the young high school graduate August Ahlqvist noted in the pages of his diary his thoughts, impressions and future plans.'

Lindén (1966: 346) suggests that one catalyst to the loss of the functional distinction between finite and non-finite subordination may have been a second-language effect: many of the Fennomaniacs who contributed to the creation of modern Standard Finnish were native speakers of Swedish who lacked an intuition of the functional distinction between finite and non-finite subordination.

Ample use of non-finite constructions is a characteristic of Finnish in which it differs typologically from the Germanic languages. The creators of the modern standard language reveled in this difference. As Lindén (1966: 346) notes, Becker's (1824: 237) Finnish grammar

states that Swedish subordinate clauses "kan och bör ersättas" ("can and should be replaced") by non-finite constructions in Finnish. This nationalist aesthetic heightens the typological contrast between the languages, but at the same time tends to set present the functional boundaries between the environments where constructions are used are isomorphic even if the constructions themselves are different - hence a one-to-one mapping of translational equivalence. The cognitive mechanism here is a conscious mapping by adults of one construction onto another, a type of hypercorrection.²

In addition, extensive use of non-finite constructions is reminiscent of the Classical languages Latin and Greek which shaped written culture across Europe and thus may have been seen as a desideratum in establishing Finnish as a European language of culture, perhaps implicitly better suited to be one than the more analytic Swedish. Hence indirect contact in the form of both positive and negative models conspired to make the use of non-finite constructions desirable as part of national identity formation.

The influence of Latin and Greek non-finite constructions on other languages through Bible translations and learned culture has been discussed extensively. Such influence can appear in a variety of guises. Herlin and Kotilainen (2004) suggest that the result of Latin influence on the Finnish ESSA construction is a proliferation of directions - rather than following a single monodirectional grammaticalization path, the construction is extended in its semantic range by analogy or translational equivalence to the very general ablative absolute in the Latin language with its long literary tradition. This is an example of replication of grammatical use patterns (Heine and Kuteva 2005: 40-78). Grammaticalization thus interacts with borrowing and the development of written culture.

Synchronically in Modern Finnish, use of non-finite constructions is generally perceived as more "old-fashioned" and literary than use of finite subordination (Comrie 1988: 475), reflecting the overall "analytic tendency" in Uralic languages (cf. e.g. Tauli 1966). There is some irony here inasmuch as the literary usage in Finnish is young. The perceived equivalence between literary and archaic usage may be based on the model of languages with a long literary tradition, hence itself an effect of contact.

In addition to increasing the similarities between non-finite constructions and finite clauses, the notion of equivalence, transformation or paradigmaticity has also led to convergence in function and form among different non-finite constructions. The system of non-finite constructions seen in modern standard written Finnish is an emergent phenomenon in which constructions which are morphologically and historically unrelated come to be perceived as belonging to a larger metacategory or subsystem of the language. While some of this emergence occurred in a prehistoric "state of nature," the process has been catalyzed by standardization, the emergence of written culture, norm-makers' aesthetics and indirect language contacts.

This paradigmaticity has also led to increased productivity and generalization of some non-finite constructions. J. Leino (2005) demonstrates that the passive form of the ESSA construction

² Ehala (1998) describes a partly analogous case from Estonian. After the charismatic language planner Johannes Aavik described the predominant verb-final subordinate clause word order in Estonian as reflecting undesirable German influence, the word order of subordinate clauses in Estonian newspaper texts changed dramatically, presumably as a result of deliberate change on the part of adult speakers.

(shown in (7)), which is nearly absent from dialect usage (Lindén 1961: 196; Ikola, Palomäki and Koitto 1989: 331), develops in written Finnish to be used in a symmetric contrast with the active forms in a way that follows the diathesis contrast in finite clauses.

- (7) Tä-stä seteli-stä Suome-n Pankki maksa-a
 this-ELA banknote-ELAT Finland-GEN bank(NOM) pay-3SG
vaadi-tta-e-ssa sata mk.
 demand-PASS-INF-INESS 100 mk
 'For this note the Bank of Finland will pay 100 marks on demand ('when demanded').

Similarly, the TUA construction has increased in frequency because it has come to be perceived as the "past tense" or "perfect" form of the ESSA construction. Itkonen-Kaila (1991: 274) shows that the TUA construction is almost absent from Agricola's New Testament translation, appearing only in a very few "epäproduktiivisia kiteytymiä" ("non-productive frozen forms"), especially based on *tulla* 'to come'. Forsman Svensson (1992: 77) shows that the ratio of instances of the TUA construction to the ESSA construction in texts increases dramatically over the history of written Finnish. Some of this change may be due to the increased input of eastern dialects, where the construction is more robustly attested, in the 19th century, but the trend begins earlier. The cognitive process involved in this change would be the conscious substitution by adults of one converb construction for another, and/or of a converb construction for a finite clause - a type of hypercorrection.

This reflects the idea that the two constructions have the same function except that ESSA expresses events simultaneous with the matrix clause, TUA events temporally prior. The two are both traditionally called "temporal constructions," and are almost always treated together in both traditional and modern linguistic discussions, despite their morphological differences. However, as Lindén (1961: 204) points out, the TUA construction is lacking from many western Finnish dialects. Morphological variation in its formation (Itkonen et al. 1978: 44) also suggests it has a marginal status in the language.

In the northern part of the Kemi dialect area, the TUA construction is used in contexts where the standard language would require the ESSA construction (Cannelin 1888: 99, 105, cf. Lindén 1961: 205n). Itkonen-Kaila (1991: 274-275) observes that the very few examples of the TUA construction in Agricola's writings also include some where the modern language would require the ESSA construction (examples 8 and 9).

- (8) Nin iloit-zi-t sijs Opetuslapse-t **Herra-n nech-ty-e=ns**
 So rejoice-PST-3PL thus disciple-PL Lord-GEN see-PTCP-PRTV=Px.3
 'Thus the disciples rejoiced, having seen (i.e., seeing) the Lord' (Piina III 186)
- (9) Ja ette hen-en pit-i alasastu-ma-n wijmeis-ne peiue-ne **Jumala-n sana**
 and that he-GEN hold-PST descend-INF-INSTRUCT last-ESS day-ESS God-GEN word
ol-tu-a=nsa,
 be-PTCP-PRTV=Px.3
 Ise-ste Jumala-st totise-sta syndy-nyt
 father-ELA God-ELA true-ELA be.born-PTCP

'and that he should descend on the last day having been (i.e., being) the word of God, born of the true God the Father'

Although the paradigmatic contrast between the constructions is found in a few MA examples, such as (10), it may be largely a product of standardization (cf. Jaakola 2004: 239-240).

- (10) mi-tä tüö süü-ve-ssä juotte [maitoa], juokee
what-PRTV you eat-INF-INCESS drink-2PL [milk-PRTV] drink-IMPER.PL
süötü'ää=ne (Hirvensalmi)
eat-PTCP-PRTV=Px.2PL
'Why are you drinking [milk] while eating; drink after you have eaten'

This "paradigmaticization" reached its height in the 1970s, when it was suggested that the AKSE or final construction (Nikanne's 1d), which occasionally appears in non-purpose meanings (as in (11)) should be viewed as the future tense of the temporal construction, despite syntactic and semantic differences (such as the fact that the AKSE construction requires its subject to be identical to that of the matrix clause, while the ESSA and TUA constructions do not, as seen in (2) above) (cf. Vesikansa 1974, Bartens 1977, Hakulinen and Karlsson 1979, Vihonen 1991, P. Leino 2005).

- (11) hän horju-i ja kompastel-i **kaatu-a-kse=en** lopu-lta pitkin
s/he stagger-PST and stumble-PST fall-INF-TRANSL=Px.3 end-ABL along
pituu-tta=an kadu-lle
length-PRTV=Px.3 street-ALL
's/he staggered and stumbled (only/eventually) to fall finally flat onto the street' (A. Hakulinen and Karlsson 1979: 388).

The idea that morphologically unrelated converbs form a "tense system" (the present and perfect forms of the temporal construction) derives from the analogy between non-finite constructions and finite clauses - hence one perceived isomorphism catalyzes convergence along another dimension as well.

In general, the norms have aimed to define a clear division of labor among the converb types (as seen in the examples in 1) and limited semantics for each. This is often presented in terms of the type of finite subordinate clause (e.g. beginning with which conjunction) to which the construction is "equivalent" (cf. e.g. Kiuru 1981: 1). This again tends to increase symmetry between the finite and non-finite "systems" and establish the latter as an emergent system.

The TEN converb (a.k.a. *modaalirakenne*, second infinitive inessive, Nikanne's 1c) has been a particular target of normative attempts to limit its semantic range and to prevent individual forms from becoming grammaticalized. This normative aim, however, struggles against a natural tendency for grammaticalized converb forms to be created by analogy to others. In vernacular usage the TEN converb may be the most general of the Finnish converb types (cf. Ontermaa 2005).

Analogical influence among converb constructions also applies on the level of individual idiomatic forms. Valijärvi (2007) conducts an extensive corpus-based survey of converbs in Modern Finnish to assess the degree to which individual converb forms are fossilized and show signs of developing into other categories. According to Haspelmath (1995), converbs arise from participial forms or verbal nouns (1995: 17) and they "die" by becoming fossilized into adpositions, adverbs, adverb markers, conjunctions or connectors (1995: 37-42). Valijärvi (2007) defines "lexicalization" as development into lexical categories (mainly adverbs) and "grammaticalization" as category change toward a grammatical form but treats the processes in the same way in other respects. Valijärvi's detailed empirical study is a synchronic snapshot of written Finnish in the 1990s; she avoids speculation at greater time depth.

Examples of the TUA construction are underrepresented among Valijärvi's examples of grammaticalized converbs. She suggests that this may reflect the relative rarity of the TUA construction in general. The examples which she mentions are all unaccusative verbs referring to changes of location or state, referring to coming, going or ending. In the relevant meaning the verbs are non-durative. They show signs of developing into adpositions marking temporal sequences. These are listed in (12).

- (12) a. low degree of grammaticalization: GEN *tultua* 'after' - 'come:CVB',
 b. "more marginally grammaticalized converb forms" (178): GEN *lakattua* 'after' = 'end:CVB', GEN *loputtua* 'after' - 'end:CVB' GEN *mentyä* 'after' - 'go:CVB', ELA *päästyä* 'after, by the time of' - 'get:CVB', GEN *päätyttyä* 'after' - 'end:CVB' (2007: 178)
 c. "moderately" grammaticalized GEN *kuluttua* 'after, in'.

Valijärvi notes that grammaticalized converb forms may be created by analogy to others with similar meanings. This is likely to hold for the preceding examples. She also mentions the possibility that denominal adpositions with similar meanings could influence the semantic development of adpositions based on converb forms.

It also appears that in many cases morphologically different converbs based on the same stem will all show grammaticalization. Hence Valijärvi describes *tultua*, *tullessa*, and *tullen*, all based on the verb *tulla* 'to come', as showing grammaticalization toward adpositions.

- (13) Illa-n ~ tul-tu-a ~ tul-le-ssa ~ tul-le-n
 evening-GEN ~ come-PTCP-PRTV ~ come-INF-INESS ~ come-INF-INSTR
 'when evening had come ~ came'

The functional contrast associated with canonical meanings of these morphological converb types is minimal or absent in these grammaticalized forms. Lindén (1961: 207-208) likewise notes that idiomatic converb forms creating temporal expressions often appear in both TUA and ESSA forms based on the same verbs, with little semantic difference between the two constructions.

Hence the semi-grammaticalized forms are not each independently following a grammaticalization path; rather, analogy operates at the level of semi-grammaticalized forms to create other such forms.

Why would only these TUA forms show grammaticalization? The semantics of these verbs may be conducive to this type of reanalysis, but it also appears that the construction was not fully productive across the lexicon in most Finnish dialects. Examples of the TUA construction in dialect archives show a very limited range of lexical types - they appear with a small number of different verbs, most of which have similar meanings (cf. Kiuru 1981: 9). Almost all these verbs are intransitive. I found (Willson, forthcoming) that two-thirds of the 287 instances of the TUA construction in the Morphological Archive of Finnish Dialects represent only eight different verbs. These are listed in (14). All are statistically more frequent in this construction in the archive than in the vocabulary of Finnish overall.

- (14) *tulla* 'to come' (38 attestations), *kuolla* 'to die' (with 26), *lähteä* 'to leave, depart' (22), *mennä* 'to go' (20), *poikia* 'to calve' (17), *päästä* 'to get (from one place to another)' (22), *syntyä* 'to be born' (23), *syödä* 'to eat' (21).

All of these are intransitive except *syödä* 'to eat' (which often appears without an expressed object). They overlap considerably with the types identified by Valijärvi (2007) as showing a low degree of grammaticalization. These forms are idiomatic and conventional or "frozen" in some sense whether or not they have undergone a change in their syntactic category. It seems likely that there have been dialects and language stages in which the construction was restricted to verbs of the type statistically overrepresented in the MA.

It may not be the case that the grammaticalizing forms identified by Valijärvi (2007) are following a long-term grammaticalization trajectory from productive to frozen. Rather, they may be influencing each other and moving laterally along the margins of productivity. Many studies of gradient productivity treat productivity and analogy as differing in degree of abstraction or schematicity rather than in kind. Full syntactic productivity may in some cases be secondary, a new lease on life given the construction by standardization and conscious ideas of syntactic symmetry.

REFERENCES

- Bartens, Raija. 1977. Suomen temporaalirakenteen futuurista. *Virittäjä* 81.309-310.
- Becker, Reinhold von. 1824. *Finsk grammatik*. Åbo: Bibel-sällskapetets tryckeri.
- Cannelin, Knut. 1888. *Tutkimus Kemin kielenmurteesta*. Helsingissä: Suom. kirjallisuuden seuran kirjapainossa.
- Comrie, Bernard. 1988. General features of Uralic languages. In *The Uralic languages: description, history and foreign influences*, ed. by Denis Sinor, 451–477. Leiden: E. Brill.
- Ehala, Martin. 1998. How a man changed a parameter value. The loss of SOV in Estonian subclauses. In *Historical linguistics 1995: selected papers from the 12th International Conference on Historical Linguistics, Manchester, August 1995*, ed. by Richard M. Hogg and Linda van Bergen, 73-88. Amsterdam/Philadelphia: John Benjamins.

- Forsman Svensson, Pirkko. 1992. *Vanhan kirjasuomen nominaalirakenteista. I. 1600-luvun partisiippiattribuutteja. II. 1600-luvun lauseenvastikkeita.* Stockholm studies in Finnish language and literature, 8. Stockholm: Department of Finnish, Stockholm University.
- Hakulinen, Auli and Karlsson, Fred. 1979. *Nykysuomen lauseoppia.* Helsinki: SKS.
- Haspelmath, Martin. 1995. The converb as a cross-linguistically valid category. In *Converbs in cross-linguistic perspective*, ed. by Martin Haspelmath and Ekkehard König, 1-56. Berlin: Mouton de Gruyter.
- Haspelmath, Martin, and Ekkehard König (eds.) 1995. *Converbs in cross-linguistic perspective.* Berlin: Mouton de Gruyter.
- Heine, Bernd and Kuteva, Tania. 2005. *Language contact and grammatical change.* Cambridge: Cambridge University Press.
- Herlin, Ilona and Kotilainen, Lari. 2004. External factors behind cross-linguistic similarities. In *Up and down the cline. The nature of grammaticalization*, ed. by Olga Fischer, Muriel Norde, and Harry Perridon, 263-279. Amsterdam: John Benjamins.
- Herlin, Ilona and Kotilainen, Lari. 2005. Itsenäistyvä infinitiivi. 2. infinitiivin inessiivin kehitys kirjakielen aikana. In *Elävä kielioppi. Suomen infiniittisten rakenteiden dynamiikkaa*, ed. by Ilona Herlin and Laura Visapää, 258-291. Helsinki: SKS.
- Ikola, Osmo. 1974. *Lauseenvastikeoppia. Nykysuomen lauseenvastikkeiden ja niihin verrattavien rakenteiden selvittelyä.* Helsinki: SKS.
- Ikola, Osmo, Palomäki, Ulla and Koitto, Anna-Kaisa 1989. *Suomen murteiden lauseoppia ja tekstikielioppia.* Helsinki: SKS.
- ISK = *Iso Suomen Kielioppi.* 2004. Helsinki: SKS.
- Itkonen, Terho, Heikki Paunonen, Kaisu Juusela, Raija Lehtinen, Kerttu Mulari. 1978. *Muotoopin keruuoppaan 2. lisäviikko.* Helsinki: Helsingin yliopiston suomen kielen laitos / Kotimaisten kielten tutkimuskeskus.
- Itkonen-Kaila, Marja. 1991. Agricolan Uuden testamentin temporaalirakenteet. *Virittäjä* 95.255-280.
- Jaakola, Minna. 2004. *Suomen genetiivi.* Helsinki: SKS.
- Kiuru, Silva. 1981. Temporaalisen lauseenvastikkeen sanajärjestys. *Virittäjä* 85.1-20.
- König, Ekkehard. 2005. The meaning of converb constructions. In *Converbs in cross-linguistic perspective*, ed. by Martin Haspelmath and Ekkehard König, 57-95. Berlin: Mouton de Gruyter.
- Laitinen, Lea. 1997. Norms made easy: case marking with modal verbs in Finnish. In *Taming the vernacular: from dialect to written standard language*, ed. by Jenny Cheshire and Dieter Stein, 110-124. London and New York: Longman.
- Laitinen, Lea. 2004. Grammaticalization and standardization. In *Up and down the cline. The nature of grammaticalization*, ed. by Olga Fischer, Muriel Norde, and Harry Perridon, 247-262. Amsterdam: John Benjamins.
- Leino, Jaakko. 2005. Passiivin infinitiivimuodot suomen murteissa. In *Elävä kielioppi. Suomen infiniittisten rakenteiden dynamiikkaa*, ed. by Ilona Herlin and Laura Visapää, 231-257. Helsinki: SKS.

- Leino, Pentti. 2005. Kun tarkoitus häviää - finaalisesta infinitiivistä temporaaliseksi. In *Elävä kielioppi. Suomen infiniittisten rakenteiden dynamiikkaa*, ed. by Ilona Herlin and Laura Visapää, 194-230. Helsinki: SKS.
- Lindén, Eeva. 1959. Possessiivisuffiksien viittaussuhteista. In *Verba docent. Juhlakirja Lauri Hakulisen 60-vuotispäiväksi 6.10.1959*, ed. by Pertti Virtaranta, Terho Itkonen, and Paavo Pulkkinen, 294-308. Helsinki: SKS.
- Lindén, Eeva. 1961. Temporaalirakenne suomen murteissa. *Virittäjä* 65.194-210.
- Lindén, Eeva. 1966. Latinan vaikutus temporaalirakenteeseen suomen kirjakielissä. *Virittäjä* 70.1-6, 339-347.
- Nedjalkov, Vladimir P. 1995. Some typological parameters of converbs. In *Converbs in cross-linguistic perspective*, ed. by Martin Haspelmath and Ekkehard König, 97-136. Berlin: Mouton de Gruyter.
- Nikanne, Urpo. 1997. Suomen infiniittisten adjunktien temporaalinen tulkinta. *Virittäjä* 101.338-357.
- Ontermaa, Aki. 2005. Hilsettä ja herrastelua? Toisen infinitiivin instruktiivin ohjailu ja kielenhuollon ideologiat. In *Elävä kielioppi. Suomen infiniittisten rakenteiden dynamiikkaa*, ed. by Ilona Herlin and Laura Visapää, 318-343. Helsinki: SKS.
- Tauli, Valter. 1966. *Structural tendencies in Uralic languages*. Hague: Mouton.
- Vesikansa, Jouko. 1974. Temporaalirakenteen futuuri. *Virittäjä* 78.271-276.
- Vihonen, Sakari. 1991. Temporaalirakenteen nimeämisperusteista. *Virittäjä* 95.48-50.
- Valijärvi, Riitta-Liisa. 2007. A corpus-based study on the degrees of lexicalization and grammaticalization of Finnish converbs. Ph.D. dissertation, Uppsala University.
- Wiik, Kalevi. 1981. Mikä lauseenvastike on? *Virittäjä* 85.21-39.
- Willson, Kendra. forthcoming. Syntaktinen symmetria, epäsuorat kontaktit ja TUA-konstruktion standardisointi. In *Verbit ja konstruktioit*, ed. by Ilona Herlin and Lari Kotilainen. Helsinki: SKS.
- Ylikoski, Jussi. 2000. Konverbeistä ja konverbirakenteista. In *Näkökulmia kielitypologiaan*, ed. by Anneli Pajunen, 214-245. Helsinki: SKS.

Sylwiusz Żychliński

Against locative treatment of Polish object experiencers¹

1 INTRODUCTION

If there are a number of topics tackled regularly by generative linguists, psychological predicates are by far one of such evergreens. From very early on in the days of the generative enterprise, the linking/mapping conundrum associated with the distribution of psych verbs has presented a challenge in dire need of explanation. Such explanations have come, and they have abounded. However, whether we have reached the explanatory adequacy regarding the queer distribution of predicates with the Experiencer thematic role still remains to be checked against the rich cross-linguistic evidence. This paper focuses on a narrow selection of properties of psychological predicates, the status of the Experiencer argument being the most important among them. As more attention has been devoted to the question of unaccusativity, which is central to the proper treatment and representation of argument structure of psychological predicates, it is the class of object experiencer (or ObjExp) verbs that is the primary focus of investigation. The aim of this paper is twofold. On the one hand, it provides a brief overview and critique of different treatments of ObjExp verbs so far, mainly Belletti and Rizzi's (1988), Pesetsky's (1995) and Landau's (2010). On the other hand, it brings attention to some of the problems that a recent analysis of Object Experiencers, a locative approach (Landau 2010), faces.

2 SELECTED ASPECTS OF THE ANALYSIS OF PSYCHOLOGICAL PREDICATES

2.1 Belletti and Rizzi's seminal analysis

Over the years, psychological predicates have been the object of numerous analyses. Of the myriads of accounts, Belletti and Rizzi (1988), Grimshaw (1990), Pesetsky (1995) and Landau (2010) are all among recent authors of comprehensive theories striving to cast light on and explain the diversity of data.

The standard problem with psychological verbs is how to reconcile their thematic structure with UTAH². The canonical patterns manifested by psych verbs are presented in (1-3³) (Belletti and Rizzi⁴ 1988: 291):

¹ A version of this paper has been published as Żychliński (2011). This is a revised version with previously unpublished evidence presented in Section 6.

² It has been one of the challenges to prove UTAH right or wrong. In its original formulation, it is as follows:

- (1) Gianni teme questo.
Gianni.Nom fears this.Acc
- (2) Questo preoccupa Gianni.
this.Nom worries Gianni.Acc
- (3) a. A Gianni piace questo.
to Gianni.Dat piace questo.Nom
 b. Questo piace a Gianni.
this.Nom pleases to Gianni.Dat

At first blush the same θ -roles are distributed arbitrarily, with the Nominative experiencer in the subject position and the Accusative theme in the object position in (1) (frequently referred to as class I of psychological verbs), the Nominative theme in the subject position and the Accusative experiencer in the object position in (2) (class II), and the dative Experiencer and the nominative Theme in both orderings in (3) (class III). As such freedom of syntactic positioning is not otherwise widely attested, a convincing argument to explain this phenomenon was much needed. On the basis of a thorough investigation of Italian facts, Belletti and Rizzi (1988: 293) proposed one uniform underlying representation for sentences (2-3) (they treat (1) in an uncontroversial way, with the surface arguments marking their DP positions):

i) *Uniformity of Theta Assignment Hypothesis* (UTAH)

Identical thematic relationships between items are represented by identical structural relationships between those items at the level of D-Structure (Baker 1988: 45).

A weaker version of UTAH was proposed by Perlmutter and Postal (1984), which reads as UAH.

ii) *Universal Alignment Hypothesis* (UAH)

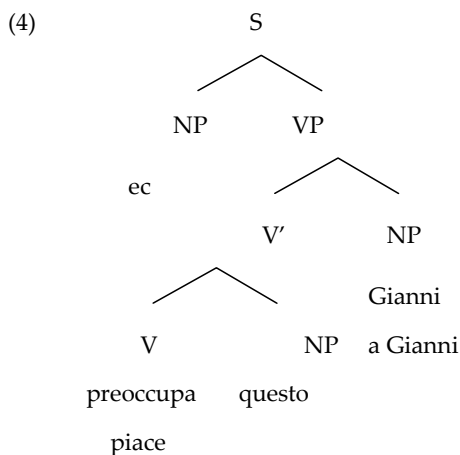
There exist principles of UG which predict the initial relation borne by *each* [argument] in a given clause from the meaning of the clause.

In *Zero Syntax* (1995), Pesetsky argues that it is actually possible to keep UTAH intact on the novel assumption that the repository of thematic roles is governed by finer-grained semantics, making it possible to distinguish among different types of the Theme role.

³ Only the Italian examples are cited, but English, Russian or Polish, to mention just a few languages, exemplify a similar pattern, e.g. in Polish:

- (1) Janek boi się tego.
 John fears this
- (2) To przeraża Janka.
 This frightens John
- (3) Jankowi podoba się to.
 To John (Dat) pleases REFL it
 To podoba się Jankowi.
 It pleases REFL to John

⁴ Henceforth B&R.



The diagram in (4) shows that psych verb sentences with different surface structures project the same configuration at the underlying level, with the Experiencer asymmetrically c-commanding the theme, formalized in (5) (B&R 1988: 344):

- (5) Given a θ -grid [Experiencer, Theme], the Experiencer is projected to a higher position than the Theme

The motivation for movement out of the VP to the subject position comes from B&R's interpretation of Burzio's Generalization (1988: 332):

- (6) V is a structural Case assigner iff it has an external argument

The practical implementation of (6) leads to the conclusion that verbs of the *preoccupare* and *piacere* type (2-3) do not have external arguments, with the subject position considered as athematic. To derive (2), then, we assume that the Experiencer gets the inherent Accusative⁵, and the Theme, by virtue of (6), moves to the subject position to get Case. In (3), since Dative is assigned by a preposition, the Experiencer is free to stay in the VP or move to the subject position.

2.2 Landau's reanalysis

In *The Locative Syntax of Experiencers*⁶ (Landau 2010), B&R's conclusions regarding the case of the experiencer are taken one step further. It is the author's original idea that all object experiencers are not only inherently case marked, but also governed by prepositions, thus Oblique. This applies equally well to Dative arguments and, albeit in a less obvious way, to Accusative ones. Such an assumption, to treat the source of Dative and (inherent) Accusative on a par, is not

⁵ This is possible if unaccusativity is restricted to the lack of structural case.

⁶ If not indicated otherwise, all references regarding Landau will consider this monograph.

necessarily novel as it basically follows the argumentation of Emonds (1985), who noticed that inherent case is assigned by either overt or null prepositions, depending on language-specific factors. But this is not where Landau's original contribution stops. The author goes on to present evidence that experiencers are in fact mental locations⁷ and, as such, locatives. To make it more plausible, periphrastic psych constructions are cited (Landau 2010: 11):

- (7) a. Nina is in love (with Paul).
- b. There is in me a great admiration for painters. (Arad 1998: 228)

It is easy to reproduce relevant examples in Polish:

- (8) Wzbiera we mnie gniew na myśl o spotkaniu.
 wells up in me anger at the thought of meeting.
 'I get angry at the thought of the meeting'

However, apart from conceptual reasoning Landau amasses more empirical evidence to corroborate the locative nature of experiencers, drawing heavily on the cross-linguistic adjunct control and Super-Equi facts. The gist of this argumentation will be presented in Section 6, where Landau's findings are confronted with the Polish data. For the time being, two crucial conclusions follow. One is the logical consequence of the claim that object experiencers are locatives, namely they must undergo (covert) locative inversion to a higher subject position. The other conclusion states that the nature of case on class II and class III object experiencers is basically the same⁸, i.e. both groups of verbs are governed and marked for case by prepositions.

2.3 Reanalysis of psych unaccusatives

One of the key diagnostics for unaccusativity comes from the domain of passivization. As observed by Marantz (1984: 144-149), unaccusative verbs are incompatible with passive morphology:

- (9) a. Passive morphology absorbs the external (underscored) Θ -role.
- b. Vacuous dethematization is impossible.

However, English ObjExp passives are relatively frequent (Pesetsky 1995: 22):

- (10) Bill was angered by Mary's conduct.
- (11) The paleontologist was pleased by the discovery of the fossil.
- (12) Bill was irritated by the loud noises coming from the next door.
- (13) Bill would not be satisfied by half measures.

⁷ Landau follows here the tradition of Jackendoff's decompositional analysis of mental states (1990), refined in the works of Bouchard and Arad, among others.

⁸ The distinction is finer-grained, though its details are not of primary significance in the context of this presentation. As a matter of fact, there is a structural difference between stative/eventive Object Experiencers as well as agentive/non-agentive ones. For further reference, see Landau (2005, 2010).

B&R go to great lengths showing that these passives are adjectival, not verbal, therefore not subject to facts noted by Marantz. Still, Pesetsky manages to show that some passives in English are unambiguously verbal⁹:

(14) Sue was continually being scared by odd noises.

As a result, the unaccusative nature of object Experiencers cannot be maintained.

3 AN ATTEMPT TO UNIFY THE FACTS

Landau (2010) pieces together all the facts compiled by his predecessors and takes his conclusions one step further, offering a cross-linguistic generalization on unaccusativity among class II psych predicates. A painstaking analysis of verbal/adjectival passives in a cross-linguistic perspective allows him to divide languages into two groups with respect to the status of their class II psych predicates:

(15) Psych Passives

Type A Languages: Only eventive (non-stative) Class II verbs have verbal passive.
(English, Dutch, Finnish)

Type B Languages: Class II verbs have no verbal passive.
(Italian, French, Hebrew) (Landau 2010: 47)

His predictions depart from the reasoning offered by Pesetsky, as Pesetsky assumed that languages are basically similar with respect to verbal/adjectival passives, and pointed out flaws in the selection of examples in B&R, whereas Landau partly agrees with both sides.

With the case on the object being determined as inherent, Landau starts out by explaining what two strategies can be employed to passivize a quirky object.

(16) Strategies for Passivization of Quirky Objects

a. P-stranding: The preposition that governs the object is stranded and reanalyzed with the verb.

Pseudopassive: [TP [DP Exp]₁ [T' Aux [VP [V VPASS + Ø₋][DP t₁]]]]]

b. Pied-Piping: The preposition that governs the object is carried along to the subject position.

Quirky passive: [TP [PP Ø₋ [DP Exp]]₁ [T' Aux [VP VPASS [PP t₁]]]]] (Landau 2010: 48)

English and Dutch are among languages showing the former strategy, whereas Finnish uses the latter one (Landau 2010: 48):

⁹ For Pesetsky, the presence of a progressive aspect and the choice of prepositions are some of the diagnostics for the verbal status of passives.

- (17) This bed was slept in.
 (18) Daar werd over gepraat. (Dutch)
 There was talked about
 (19) Sinu-sta pidetään. (Finnish)
 you.ELA like.PASS
 'You are liked'

This leads Landau to formulate a cross-linguistic generalization about the availability of verbal passives in ObjExp verbs (Landau 2010: 49):

- (20) Verbal passives of non-agentive ObjExp verbs will only be available in languages allowing either pseudopassives or (oblique) quirky passives.

Neither of the strategies described above for the passivization of quirky objects is attested in Italian, French and Hebrew (these three were tested by Landau), which is taken as a clear indication that they do not have verbal passives and, thus, belong to Type B Languages. Contrary to Pesetsky's criticism of B&R's treatment of Italian participles, Landau is able to show that even though their analysis is not adequate for English, it still passes muster in Italian.

4 POLISH OBJEXP VERBS AND THEIR UNSYSTEMATIC BEHAVIOR

The classification of participles put forth in the last section will only hold water on the assumption that object experiencers are PPs. What if they are not? All evidence leads to the inevitable conclusion that the nature of case on the object experiencer is different in Polish. It has been commonly assumed that the object experiencer bears inherent case. One of the universally accepted diagnostics for that is case suppression. Case suppression is clearly manifested in Russian, where Genitive of Negation rule is in full operation. This means that objects, which in positive sentences come in Accusative, switch to Genitive in negated sentences. The rule does not work with class II psych predicates, in which Genitive is not possible:

- (21) * Ètot šum ne pobespokoil ni odnoj devoc'ki.
 that noise.NOM not bothered not one girl.GEN
 'That noise did not bother a single girl'
 (22) * Ego neudacca ne ogorc'ila materi.
 his failure.NOM not upset mother.GEN
 'His failure did not upset mother'
 (Legendre & Akimova 1993, ex. 40)

This is to be expected given that "(a) standard account for this contrast exploits the fact that inherent case is fixed in the lexicon; GN, which is a syntactic rule, cannot override this case" (Landau 2010: 25). Polish, however, despite the fact that it is very similar to Russian in terms of the Genitive of Negation rule, surprisingly produces well-formed sentences:

- (23) Ten hałas nie zaniepokoił ani jednej dziewczyny.
 this noise-NOM not worried any single girl-GEN
 ‘This noise did not worry a single girl’
- (24) Jego porażka nie zdenerwowała matki.
 his failure-NOM not upset mother-GEN
 ‘His failure did not upset his mother’

Apart from case suppression, the non-inherent nature of case is also further substantiated by reflexivization facts. Neither in English-type languages nor in Italian-type languages can object experiencer verbs reflexivize:

- (25) a. Tedious talks irritate me.
 b. *I irritate myself.
- (26) *Gianni si preoccupa.
 Gianni *si* worries
 Gianni worries himself

A standard account for such a state of affairs involves the restriction on reflexivization (Landau 2010: 35):

- (27) Reflexive *si/se* may absorb accusative or dative case, but not oblique case.

Given that Accusative in Polish has all the hallmarks of structural case, it seems a correct prediction to assume that Polish reflexives should be well-formed¹⁰. This is, indeed, the case.

- (28) Złe wieści martwią Jana.
 bad news.NOM worry Jan.ACC
- (29) Jan martwi się (złymi wieściami).
 John worries CL.się (bad news.INST)
 John worries himself

As already stated, it is Landau’s underlying assumption that all object experiencers are introduced by prepositions and marked as Oblique. As prepositional phrases are islands, extraction out of them is predicted to yield ungrammatical results. The examples from Polish do not give straightforward results, the judgments being at least divided¹¹.

¹⁰ An anonymous abstract reviewer pointed out that in Russian similar reflexivization facts obtain:

- 1) Ivan volnuetsja.
 Ivan worries-himself.

This, in fact, may be a further fact obscuring Landau’s analysis. Alternatively, it may suggest that the clitic reflexivization is not a true instance of reflexivization.

¹¹ Similarly, different facts corroborating the locative nature of object experiencers are not unambiguous in Polish, e.g. object control into adjunct clauses and Super-Equi control facts.

- (30) ?Czyją irytowało to siostrę?
 whose irritated it sister
- (31) ?Czyją straszyłeś siostrę?
 whose frightened you sister
- (32) *Czyją irytują nocne telefony siostrę?
 whose irritate night calls sister?

Interestingly enough, the (at least partial) acceptability of (30) and (31) seems to be related to the (phonologically) light status of the subject more than to their agentive versus non-agentive subjects. In (32), whose subject is phonologically heavier, the judgment is straightforwardly ungrammatical.

5 MORE TROUBLE WITH POLISH PARTICIPLES

In the light of the discussion on participles in the preceding sections, Polish participles should be considered now in greater detail. Similarly as in the prior discussion, only class II ObjExp verbs are under observation, as they belong to the only problematic group as regards the question of unaccusativity. Verbs of this class include, among others, *przygnębiać* (depress), *irytować* (irritate), *straszyć* (frighten), *niepokoić* (worry), *kłopotać* (embarrass):

- (33) Krzyki za oknem irytują Piotra.
 shouts outside irritate Peter
- (34) Burza z piorunami przestraszyła dzieci.
 thunderstorm scared children
- (35) Wieści o kryzysie niepokoją obywateli.
 news about crisis worry citizens
- (36) Brak biletu zakłopotał pasażera.
 lack of ticket embarrassed passenger

Following Landau's line of reasoning, it should be possible to determine whether or not a language has verbal passive participles on the basis of (20), repeated below:

- (37) Verbal passives of non-agentive ObjExp verbs will only be available in languages allowing either pseudopassives or (oblique) quirky passives.

Crucially, Polish does not exhibit either of these two constructions. This follows if object experiencers are marked for structural Accusative. Therefore, a straightforward conclusion would require the classification of Polish as a Type B language (see (15)), which could be further supported by examples below:

- (38) ?Pogoda wciąż przygnębia wczasowiczów.
 weather.NOM still depresses holidaymakers.ACC
- (39) *Wczasowicze są bez przerwy przygnębiani (przez pogodę).
 holidaymakers.NOM are continuously being depressed (by the weather.OBL)

Two sentences below, however, show considerable improvement:

- (40) Krzyki na zewnątrz wciąż irytują Piotra.
 screams.NOM outside still irritate Peter.ACC
- (41) ?Piotr jest wciąż irytowany przez krzyki z zewnątrz.
 Peter.NOM is still being irritated by screams.OBL from outside

With more examples, the picture seems to be blurred even further. As has already been argued before, progressive aspect is said to be incompatible with passive morphology on ObjExp verbs. However, (42) shows that progressive is possible with a class II psych predicate in the active voice, and (43) proves that passive is also possible with the same verb. In the light of the discussion so far, this would indicate that *niepokoić* (worry) is an eventive predicate, thus not unaccusative, and it forms verbal passive participles.

- (42) Wieści o kryzysie właśnie teraz wyjątkowo niepokoją obywateli.
 news.NOM about crisis just now exceptionally worry citizens.ACC
- (43) Obywatele są teraz niepokojeni wieściami o nadciągającym kryzysie (choć ledwie co byli niepokojeni doniesieniami z rynku wschodniego).
 they have just recently been worried about news from the eastern market

It becomes apparent, then, that the classification under (37) is insufficient as it does not predict the behavior of Polish ObjExp verbs.

6 FURTHER IMPLICATIONS¹²

In the last section of this paper, a closer inspection will be given to a syntactic phenomenon possibly supporting the locative analysis of object experiencers, i.e. adjunct control. Drawing on the assumption that object experiencers move up to a subject position at LF, what should naturally follow is their capacity to control into adjunct clauses. Having analyzed languages such as French or Russian, this is precisely the claim that Landau (2010) makes. Below Russian data will be confronted with data from Polish to show that Landau's conclusions apply only to Russian.

¹² Another phenomenon, i.e. the Super-Equi facts in Polish, can also be shown to be unfavorable to Landau's locative hypothesis. For details, see Żychliński (2011).

6.1 Adjunct control in Russian

Let us analyze various patterns of control obtaining in Russian to see clearly how object experiencers stand out with respect to control.

The default type of control into adjunct clauses seems to be nominative subject control, as exemplified in (44-47)¹³:

- (44) *Vozvraščajas' domoj, Miša razgovarival s druž'jami.*
Going back home Misha-NOM talked with friends
'While going back home, Misha was talking to his friends.'
- (45) *Vernuvšis' domoj, Miša pozvonil na rabotu.*
Coming home, Misha-NOM called on office.
'Having come home, Misha called his office.'
- (46) *Tancuja s Olej, Miša vspominal Veru.*
Dancing with Olya Misha-NOM thought Vera
'Dancing with Olya, Misha was thinking about Vera.'
- (47) *Čitaja Kolino pis'mo, Miša emu zavidoval.*
Reading Kolya's letter Misha-NOM him envied
'While reading Kolya's letter, Misha envied him.'

The nature of nominative subject control is not affected by the placement of the adjunct to the right of the main clause, as the examples with right-adjunction in (48-50) illustrate:

- (48) *Miša govoril po telefonu leža.*
Misha-NOM spoke on phone lying.
'Misha was speaking on the phone lying down.'
- (49) *Saša perevodit stat'u ne pol'zujas' slovarem*
Sasha-NOM translates paper NEG using dictionary.
'Sasha is translating the paper without using a dictionary.'
- (50) *Kolja proiznosil reč' volnujas'*
Kolya-NOM made speech agitating.
'Kolya was making a speech agitatedly.'

That nominative subjects are unique as controllers is presented in (51), which is an illicit example of control into the adjunct by the accusative object:

- (51) **Tancuja s Olej, ego pozvali k telefonu.*
Dancing with Olya he-ACC asked to phone
'While dancing with Olya, (somebody) asked him to the phone.'

¹³ Examples (44-54) are taken after Legendre and Akimova (1993).

The real interesting facts, however, reach the surface of the analysis when we look at examples with accusative object experiencers. Unlike objects of non-psychological verbs, these arguments show the possibility of control into adjuncts in Russian, as seen in (52-54):

- (52) Čitaja Kolino pis'mo, Sašu vozmutilo ego legkomyslije.
 Reading K's letter Sasha-ACC irritated his lightheartedness
 'While reading Kolya's letter, his lightheartedness irritated Sasha.'
- (53) Vojdja v komnatu, Kolju porazil besporjadok.
 Entering in room Kolya-ACC impressed mess-NOM
 'Having entered the room, the mess impressed Kolya.'
- (54) ?Tancuja s Olej, ego rasstroiliee grustnye glaza.
 Dancing with Olya he-ACC upset her sad eyes-NOM
 'While dancing with Olya, her sad eyes upset him.'

Needless to say, the correlation between nominative subjects and accusative object experiencers in terms of their control characteristics strengthens the locative hypothesis advanced by Landau (2010). Given that the placement of adjuncts allows only elements situated in the subject position of the matrix clause to control them, the late locative movement of the PP containing the object experiencer to the subject position explains away this syntactic peculiarity.

6.2. Adjunct control in Polish

As mentioned above, the neatness of the locative hypothesis is weakened when the Polish data are considered. Initially, the grammaticality judgments concerning very similar sentences in Polish converge with the intuitions about Russian, with nominative subjects having the capacity to control into adjuncts both left- (examples 55-58) and right-adjoined (examples 59-61):

- (55) [eci wracając do domu], Piotr₁ rozmawiał z przyjaciółmi.
 coming back home Piotr-NOM talked with friends
 'Coming back home, Peter talked with his friends.'
- (56) [eci wróciwszy do domu], Piotr₁ zadzwonił do biura.
 having come back home, Piotr called the office
 'Having come back home, Piotr-NOM called the office'
- (57) [eci tańcząc z Ewą], Piotr₁ wspominał swą poprzednią dziewczynę.
 dancing with Ewa, Piotr-NOM remembered his previous girlfriend
 'Dancing with Ewa, Piotr remembered his previous girlfriend.'
- (58) [eci oglądając samochód Krzysztofa], Piotr₁ zazdrościł mu.
 watching car Krzysztof-GEN Piotr-NOM envied him
 'Watching Krzysztof's car, Peter envied him.'
- (59) Piotr₁ rozmawiał przez telefon [eci leżąc].
 Piotr talked on phone lying
 'Piotr was talking on the phone lying down.'
- (60) Piotr₁ tłumaczył artykuł [eci nie używając słownika].
 Piotr translated article not using dictionary

- 'Piotr was translating the article not using a dictionary.'
- (61) Piotr₁ wygłaszał przemowę [ec₁ gestykulując].
 Piotr made speech gesticulating
 'Piotr was making a speech gesticulating.'

Non-experiencer objects are excluded as controllers, which can be seen in (62):

- (62) ?*[ec₁ tańcząc z Ewą], Piotra₁ poproszono do telefonu.
 dancing with Ewa, Piotr-ACC was asked to telephone
 'Dancing with Ewa, Piotr was asked to the telephone.'

However, a crucial divergence in the data between Russian and Polish comes along with the analysis of the behavior of object experiencers in Polish. Although perhaps not completely out, the object experiencers as controllers in (63-65) are nonetheless strongly marked at best:

- (63) *[ec₁ oglądając wóz swoich marzeń], bałagan wewnątrz zadziwił Piotra₁.
 watching car his dreams-GEN mess inside surprised Piotr-ACC
 'Watching the car of his dreams, the mess inside surprised Piotr.'
- (64) *[ec₁ wchodząc do pokoju], woń papierosów zdenerwowała rodziców₁.
 entering to room smell cigarettes angered parents
 'Entering the room, the smell of cigarettes angered the parents.'
- (65) *[ec₁ tańcząc z Ewą], jej podkrążone oczy zasmuciły Piotra₁.
 dancing with Ewa, her tired eyes saddened Piotr-ACC
 'Dancing with Ewa, her tired eyes saddened Piotr.'

Placing the experiencer objects in the surface subject position, which a scrambling-prone language such as Polish allows for, does not ameliorate the situation, with the examples below clearly show:

- (66) ?*[ec₁ oglądając wóz swoich marzeń], Piotra₁ zadziwił bałagan wewnątrz.
 watching car his dreams-GEN Piotr-ACC surprised mess inside
 'Watching the car of his dreams, the mess inside surprised Piotr.'
- (67) ?*[ec₁ wchodząc do pokoju], rodziców₁ zdenerwowała woń papierosów.
 entering to room parents angered smell cigarettes
 'Entering the room, the smell of cigarettes angered the parents.'
- (68) ?*[ec₁ tańcząc z Ewą], Piotra₁ zasmuciły jej podkrążone oczy.
 dancing with Ewa Piotr-ACC saddened her tired eyes
 'Dancing with Ewa, her tired eyes saddened Piotr.'

On the basis of the control facts from Polish, it is hard to defend the locative treatment of object experiencers.

7 CONCLUSIONS

Landau's comprehensive treatment of object experiencers cannot accommodate Polish facts. The property of having verbal passives is not linked to the passivization options present in languages studied by Landau. Also, nothing seems to support the inherent nature of case on the experiencer, which ruins the locative hypothesis. Thus, the following two conclusions seem inescapable:

- a. Polish object experiencers do not show many of the syntactic characteristics typically associated with this group.
- b. Polish object experiencers are not locatives.

What is more, Polish potentially causes concerns for Burzio's Generalization, which relates the capacity to assign structural Accusative to the presence of the external argument, suggesting that more has to be said about this relation, especially given that in Polish there is at least one more known situation where Burzio's Generalization seems to run into trouble, namely impersonal passive constructions:

- (57) Napisano listy.
were written letters.ACC
The letters were written.
- (58) Nie napisano listów.
not were written letters.GEN
The letters were not written.

Not only is accusative preserved on the object, but also the operative Genitive of Negation confirms its structural status. It becomes necessary, then, to subject a greater number of languages to a painstaking analysis and find out if Polish stands alone as an exception to the locative hypothesis, or whether the locative hypothesis has to be reworked.

SELECTED LITERATURE

- Arad, M. 1998. *VP-structure and the syntax-lexicon interface*. Doctoral dissertation, University College London.
- Belletti, A., and L. Rizzi. 1988. Psych-Verbs and Theta-Theory. *Natural Language and Linguistic Theory* 6: 291-352.
- Bondaruk, A. 2004. *PRO and Control in English, Irish and Polish: A minimalist analysis*. Lublin: Wydawnictwo KUL.
- Bouchard, D. 1995. *The semantics of syntax: A minimalist approach to grammar*. Chicago: University of Chicago Press.
- Chomsky, Noam (1995) *The Minimalist Program*. Cambridge: MIT Press.
- Emonds, J. 1985. *A unified theory of syntactic categories*. Dordrecht: Foris.

- Grimshaw, J. 1990. *Argument Structure*. Cambridge: MIT Press.
- Grimshaw, J. 1990. Argument structure. Cambridge, Mass.: MIT Press.
- Grinder, J. T. 1970. Super Equi-NP deletion. *Chicago Linguistic Society* 6: 297-317.
- Higgins, R. F. 1973. *The Pseudo-cleft Construction in English*. PhD dissertation, MIT.
- Jackendoff, R. 1990. *Semantic structures*. Cambridge, Mass.: MIT Press.
- Landau, I. 2001. Control and Extraposition. The Case of Super-Equi. *Natural Language and Linguistic Theory* 19: 109-152.
- Landau, I. 2005. *The Locative Syntax of Experiencers*. Ms. Ben Gurion.
- Landau, I. 2010. *The Locative Syntax of Experiencers*. Cambridge: MIT Press.
- Legendre, G. and T. Akimova. 1993. Inversion and Antipassive in Russian. In: Avrutin, S., S. Franks & L. Progovac (eds.), *The 2nd Annual Workshop on Formal Approaches to Slavic Linguistics*, Ann Arbor, MI, 286-318, (Michigan Slavic Publication, 1994).
- Marantz, A. 1984. *On the Nature of Grammatical Relations*. Cambridge: MIT Press.
- Pesetsky, D. 1987. Binding Problems with Experiencer Verbs. *Linguistic Inquiry* 18: 126-140.
- Pesetsky, D. 1995. *Zero Syntax*. Cambridge: MIT Press.
- Żychliński, S. 2011. A Recalcitrant Nature of Object Experiencers. In: Pawlak M. and J. Bielak (eds.), *New Perspectives in Language, Discourse and Translation Studies*, Berlin: Springer, 97-90.

2. Lexicology

Michael Crombach

On the notion of “orthophemism” and other miscellanea on taboo and euphemism

This paper represents two presentations held at SCL24 that have been concatenated into a single paper. The first half of the paper investigates the value of the concept “orthophemism”, while the second half is dealing with the evolutionary emergence, the acquisition, and the prosody of taboo and euphemism.

THE NOTION OF “ORTHOPHEMISM”

Taboo once and taboo now

Taboo, the Polynesian souvenir imported by Captain James Cook, made a remarkable career for a word from a remote language. First recorded by Cook in 1777, it made its way right into the headlines of all major (Indo-European) languages. According to Betz 1978:141 (following an older interpretation by a physician named Shortland) the word *taboo* is a compound consisting of two syllables *ta* ‘intensive, strong’ and *pu* ‘marked’, thus giving the meaning ‘strongly marked’. After its introduction the word quickly received a mystification, mainly due to the works of Freud (1913) and Frazer (1894). This cumulated in the definition by the respectable and well-received sociologist Margret Mead 1934 in the Encyclopedia of Social Sciences:

Tabu [sic!] may be defined as a negative sanction, a prohibition whose infringement results in an automatic penalty without human or superhuman mediation.

In his 1956 book on taboo Franz Steiner ironically alludes to this definition, when he quotes an episode recorded by Cook 1784 (Steiner 1967:26):

the best sorts of food are *tabooed*, or forbidden to them [i.e. women] [...] and we were told that a poor little girl got a terrible beating, for having eaten, on board our ship, one of these

interdicted articles ... The islanders who tried to beat some respect for the laws and customs of her people into that foolish girl appear or have been quite unaware of Margret Mead's definition of taboo.

Definitions of this kind and the stress on the assumed ambivalence of taboo (purported mainly by Freud, who in his argumentation relies mainly on the Article *taboo* in the 11th Edition of the Britannica (Northcote 1911), a revision of the famous article by Frazer 1888 in the 9th Edition of the Britannica) have given taboo something puzzling that is not rooted in the fact of taboo itself.

But a recent definition by Keith Allan and Kate Burridge takes all the mysteries out of taboo:

"Taboo refers to a proscription of behaviour for a specifiable community of one or more persons, at a specifiable time, in specifiable contexts." (Allan and Burridge 2006:11).

I will refer to this definition throughout this paper. This relatively open and pragmatic definition would have been possible as early as the 1930s, when the German ethnologist Rudolf Lehmann (1930) in his diligent and thorough analysis of the Polynesian taboo customs clarified the central aspect of *tapu* to be nothing but 'forbidden'. In the following *taboo* refers to actions, objects, persons, words, ideas etc. that are for various reasons forbidden, or at least restricted in their use. But the first part of this paper is not primarily on taboo but deals with the linguistic possibilities to avoid taboo topics and words. Starting point is the distinction by Allan and Burridge 2006:34 that splits the linguistic handling of taboo into three possibilities: euphemism, dysphemism and orthophemism. Allan and Burridge illustrate this using a figure redrawn here, using a different example.

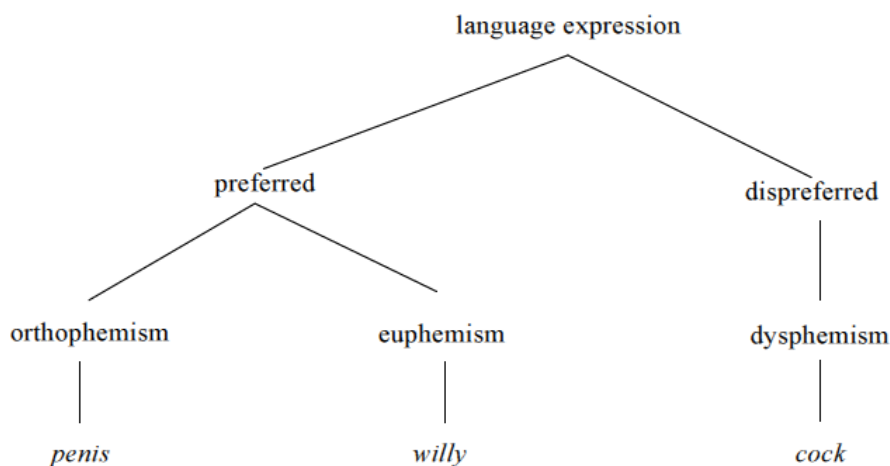


Figure 1: X-phemisms following Allan and Burridge 2006:34 (modified)

Euphemisms

Allan and Burridge 1991:11 suggest the following definition:

A euphemism is used as an alternative to a dispreferred expression, in order to avoid possible loss of face: either one's own face or, through giving offence, that of the audience, or of some third party.

But euphemisms are not simply substitutions of a word by another word (as already indicated by the vague use of "alternative"). I prefer to refer to euphemism as processes. I first found the term *process* associated with a definition of euphemism in Fuchs 2003, but she does not use it in plural (mostly there is more than one process involved at a given time) and she does not really explore the possibilities and implications of considering euphemism as *processes*. The most important implication is that most euphemisms can only be properly understood in the actual situation they are used in and, as a result, most euphemistic strategies may not find their way into a dictionary of euphemisms.

In the following I will give a short overview over the various processes that can be used as a euphemism. The most common strategy of avoidance is *silence*. As Stefan Rudas 1994:17 states, taboos are best handled by not speaking about them. When the topic/word etc. is completely avoided or ignored, then there is no need for a verbal handling of it. But when it becomes unavoidable to talk about the taboo, then a very common strategy is *vagueness*. Vagueness is possible in two ways, the first is still very close to silence, by means of an ellipsis: the tabooed word is simply left out, e.g. *holy* ... The second strategy is that the denotatum is either referred to by a something somehow related (e.g. some kind of "areal closeness" in German *Leistengegend* lit. 'iliac region' for 'private parts'), or more often referred to using *pronouns*, *interrogatives* etc. e.g. German *was*, English *it* (Crombach 2001), or local or deictic adverbs *there*, *here* (often combined with gesture) etc. *Gestures* are generally a way of avoidance, e.g. the gesture of the thumb crossing the throat indicating a non-verbal lethal threat, well known from Western movies. The most common euphemistic processes are of course words that replace other words. Especially in English there exist countless "dictionaries of euphemisms" that list taboo words and their corresponding euphemisms and/or slang terms. These processes of replacement of one lexeme by another lexeme could be called "euphemism in a narrow sense". Here it is remarkable to observe that euphemistic expressions always come in groups. In most cases there is not one word replaced by another word, but an ensemble of words replaces the word in question, which in turn belongs to a larger semantic field that is subject to taboo.

Another very important euphemistic strategy is the use of adjectives. By using the adjectival phrase *so-called* words invert their meaning to the opposite and can thereby have euphemistic character (see also e.g. Allan and Burridge 1991, 2006, Crombach 2001). If, for example, *UN soldiers* are labeled *so-called peace troops*, the speaker clearly indicates her disapproval of the UN presence in the area in question. *So-called* and similar expressions may be replaced by "*quotes*" in printed texts, so "*freedom-fighters*" is a euphemistic, or in this case more precisely dysphemistic, expression for 'rebel, revolter, insurgent'. These adjectives open the field of *phrases* as euphemistic strategies. The already mentioned *private parts* are a phrase that includes more than a euphemism for *penis*; they denote a body region otherwise unnamed. Another euphemistic strategy is *sentences* circumscribing the tabooed object. The most evil wizard of all times in J.K. Rowling's Harry Potter novels, *Lord Voldemort*, is referred to as *he-who-must-not-be-named* or *you-know-who*. The transitions from phrase to sentence to juxtaposition and thereby to *neologisms* are,

of course, gradual. *Neologisms* may also show up in other areas, e.g. in baby-talk, when the *vagina* is denoted as *cootchie* (Spencer 2005), or in hypocoristic talk between lovers, when the first name is used as a euphemism for the sex organs. Another escape strategy is the resort to foreign words, as shown by (Harris et al., 2003) taboo words are perceived less offensive in the second language; a French *merde* uttered in an entirely English setting is certainly less offensive (at least to some people) than a straight *shit*. The prestige of the foreign language used is, of course, also of importance in this context. In sum, these examples try to illustrate that euphemisms are not necessarily in dictionaries, but that they are creative processes, often resorting to ad hoc creations that are only decodable in their euphemistic setting in the given pragmatic situation.

Dysphemisms

Again I refer to Keith Allan and Kate Burridge (1991:26) for a definition:

A dysphemism is an expression with connotations that are offensive either about the denotatum or to the audience, or both, and it is substituted for a neutral or euphemistic expression for just that reason.

But I doubt the claim by Burridge and Allan (2007:6): *Dysphemism breaks social convention*. I think that dysphemisms can be the socially expected norm in certain communicative situations. If men slightly drunk in a sports bar somewhere in the middle of nowhere refer to their wives as *my squaw* (German *meine Alte*) instead of using her name, or simple *my wife* (German *meine Frau*), then they may not do so to insult their wives or to offend the audience, but they may simply follow the linguistic norm of this situation. It may be a taboo of the obvious display of sentiment that can be avoided by using a dysphemism. Marouzeau (1922, 1924), Looritz (1939), and Havers (1946) give examples where people give valuable things “bad names” to hide their material or emotional value. Therefore I see dysphemism as a different tool for the same purpose: avoiding a taboo (Crombach 2001:106f). And even more so when taking Holder 1995:vii seriously: “One man’s euphemism can be another woman’s dysphemism.”

Orthophemism

Allan and Burridge (2006) introduce “orthophemism” as the preferred expression according to the *middle class politeness criterion* (MCPC). The MCPC is defined by Burridge and Allan 2007:10 as “what would be considered the polite form when addressing a casual acquaintance of the opposite sex in a formal situation in a middle class environment.” In the following I want to show that this definition of “orthophemism” quickly blurs, because a) the MCPC is flawed, b) in a historical perspective things can change very quickly, and, finally, c) the taboo definition offered by Allan and Burridge excludes an idea like “orthophemism”.

MCPC

In my opinion the middle class politeness criterion (MCPC) is flawed: No matter how good one’s orthophemisms are: a taboo topic is not normally the type of conversation to get involved in with “a casual acquaintance of the opposite sex in a formal situation in a middle class environment.” It is strongly reminiscent of the absurd British humor most famously spread by

Monty Python to imagine someone in a situation as the one described touching on a taboo topic. Just imagine a typical coffee break at a conference, and right after you started conversation with a colleague (of the opposite sex) about the weather, her/his latest research interests, your recent grant application, all of a sudden she or he just starts talking about sex; of course using the orthophemisms required according to the MCPC: *vagina*, *penis*, *intercourse* etc. Wouldn't this be somehow pathological behavior? There are exceptions thinkable, for example, if the topic of the conference is somehow related to a taboo topic, then it is to be expected that the topic is discussed using the appropriate terminology, but certainly even under such circumstances the reference to the topic will be a professional one and not personal or autobiographic. Another situation where the MCPC may be expectable is when attending a doctor. In such a situation it may be unavoidable to speak about a taboo topic even with personal and autobiographic reference, but in such a case I doubt whether the orthophemisms are used by all people. It is equally possible in such a situation that people resort to euphemisms (and of course dysphemisms, as the distinction of these relies mainly on the personal point of view and social heritage). But even taking into account such situations, there is no really meaningful use of orthophemism under the MCPC, because, what ever conversation calls for an orthophemism is certainly not part of the kind of a conversation defined by the MCPC.

Historical perspective

It is of course always questionable if the historical perspective is a meaningful and valid approach in analyzing synchronic linguistic issues. There is always the danger or at least the scent of "language critique" (*Sprachkritik*), when comparing the current usage with the historical one. But I think as long as the results of the historical analyses are not used to judge the contemporary status of language use, it is certainly always worth analyzing the traces of development, as they can be a guide to possible future developments. I want to illustrate how the historical perspective blurs the definition of orthophemism by using two examples, the "orthophemisms" denoting "certain body parts": *vagīna* and *pēnis*. Both are of Latin origin with remarkable similarities and differences in their word biography. Their contemporary use in the Indo-European languages goes back to Latin terminology of the early anatomists in the 15th century. From this starting point these words spread into most European languages and they represent the standard expression for body parts otherwise nameless. Of course there are other words, but the other expressions referring to *pēnis* and *vagīna* are either euphemistic or dysphemistic. This namelessness is an interesting observation that is not restricted to sexual organs. The denotation of the various parts of the human body has some remarkable features in the Indo-European languages. There are body parts that have venerable old names that can be traced down for 4000 years; transformed by sound changes, but otherwise highly conservative, they are still used in everyday language, like e.g. English *foot* < Indo-European **pod-* (compare Greek *poús*, Latin *pēs*, Sanskrit *pad-* etc.) or English *heart* < Indo-European **kerd-* (compare Greek *kardía*, Latin *cor*, Sanskrit *hr̥d-* etc.) While the names of other parts are permanently changing, so that is impossible to reconstruct a common Indo-European precursor: e.g. English *hand* (compare Greek *kheír*, Latin *manus*, Sanskrit *hasta-*), English *head* (German *Kopf*, Greek *kephalē*, Latin *caput*, Sanskrit *çiras*). Havers 1946 and others tried to explain these changes with language taboo, but hardly any of these explanations is really satisfying. A very good overview of the relations is

provided by Buck 1949. Remarkably many of the frequently changing body-parts play an important role in figurative language as investigated by Niemi et al. (2010). An exception to this seems to be the *eye*, where Greek *ophthalmós*, Latin *oculus* and Sanskrit *akṣi-* point to an Indo-European **Hok^{w-}*. But this etymon also has a rather complicated kinship.

A quick biography of *vagīna* and *pēnis* reveals a very remarkable similarity of these two words: both words do not primarily designate the body parts in question in Latin. But there is also a substantial difference in this respect as the word *pēnis* is already in classical Latin considered offensive, and this pejorization has been observed by the Romans. Originally, it has obviously been a euphemism for *mentula*, but the euphemistic power has already faded in the classic period. A testimonial for this is Marcus Tullius Cicero in his letter to L. Papirius Paetus (*epistulae ad familiares* 9.22). In this letter Cicero discusses some of the “problematic” words in Latin, and whether the obscenity lies within the words or the objects they denote, without (of course) solving this problem. On the word *pēnis* Cicero notes: *caudam antique ‘penem’ vocabant [...] at hodie ‘penis’ est in obscenis* “the ancient called a tail ‘penis’ [...] but today *penis* is among the dirty words”. Already Cicero notes the inconsistency in linguistic behavior as the diminutive *penicellus* ‘brush, small tail’ (the word to become English *pencil*, German *Pinsel* ‘brush’) is not obscene. Even the medicine *Penicillin* derives from the diminutive *pēnicillium* ‘brush, bandage’ of the diminutive *penicellus*, as the mold *Penicillium chrysogenum* looks like a brush under the microscope. It is remarkable that the word *pēnis* ‘tail, penis’ can be traced down to Indo-European: **pes-nis* (Greek *péos*, Sanskrit *pásas*, Old High German *fasal*, Hittite *pesna* ‘man’). The word seems to always have had the meanings “tail” and “membrum virile” (cf. Buck 1949). And for some reason the obscenity was not felt as strong in the 15th century as the anatomists scientifically named the (human) body. This is different for *vagīna*. There is no recorded sexual meaning in classical Latin. Actually, there is one obscene wordplay in T. Maccius Plautus, Pseudolus 1181, that uses the metaphor of sword and scabbard, but this seems to remain the only record: (1180) *noctu in vigiliam quando ibat miles, quom tu ibas simul*, (1181) *conveniebatne in vaginam tuam machaera militis?* “At night, when the soldier was going on the watch, and you walked with him, didn’t the soldier’s sword match your scabbard?” But in general *vagīna* was only ‘scabbard’ to Latin speakers, sometimes used metonymic for ‘sheath, legume’. An argumentum ex silentio for this is the fact that Cicero does not mention this word. Cicero refers to the word *cunnus*. This is the taboo word in classical Latin that has to be avoided, even near homophone phrases are problematic, as documented by Cicero: *Quid, quod vulgo dicitur: ‘cum nos te voluimus convenire’, num obscenum est?* “What about, as commonly said ‘When we (cum nos) wanted to visit you’, is this now obscene?” Obviously the early anatomists were well aware of this fact and created a new term using a self-evident (Mitzka 1955:42) metaphor. The metaphor is then “reused” in the German since 1678 as loan translation *Scheide* (Mitzka 1955:42), but still *Vagina* is used as a scholarly expression like in all other European languages.

Concluding, it is worth mentioning that *vagīna* has some 1200 synonyms in English, while *pēnis* still has more than 1000 (Allan and Burridge 1991:96). And these are only the lexical replacements! There are certainly many euphemistic instances in referring to these body parts that are handled in a way that did not make it onto such a listing, e.g. *touch me here* together with a deictic gesture.

But now back to the analysis of the term “orthophemism”. The last argument that makes the definition of “orthophemism” problematic is the taboo definition provided by Allen and Burridge. They minimize the value of their concept “orthophemism”. Allen and Burridge stress the importance of the environmental setting (time and place, participants etc.) for the existence of a taboo. And this relativity may lead to a situation in which the dysphemism may seem the preferred expression (the “*my squaw*” situation mentioned above). As I tried to show above, the MCPC is not really meaningful when it comes to taboo topics, therefore it cannot function as a side condition of the definition of “orthophemism”. So an “orthophemism” is completely dependent on the situation and on the socialization and education of the speaker and hearer. They decide on their communicative experience and expectation which term is appropriate, “unmarked” in the given situation. An interesting undertaking would be an extensive data collection that records the different usages of different (potential) taboo terms in everyday language and the interdependence of the usage of these expressions with the socialization and education of the speaker and hearer. I can perfectly imagine that many people will resort to *penis* or *vagina*, when talking to their physician, but other people may still consider this “not their language” and use *cock* or *cootchie*. Some may even indicate that they do not know any better. I think that the difference between active and passive vocabulary may play an important role in such an investigation. It may not be enough to check whether people know a word, but to really verify if they actually use it. But as long as such a data collection is not available, I would suggest a different definition of orthophemism.

Redefinition of orthophemism

The term orthophemism is coined by Keith Allen and Kate Burridge using the Greek words *orthós* ‘straight, right seeming(ly), becomingly’ and *phemí* ‘say, utter, reply...’ (or the nominalization gr. *phēmē* ‘naming, language, rumor, utterance’). I suggest giving this term a more etymological reading: *seemingly naming*. The definition then would be: orthophemism is “the right, meaning successful, expression for the communicative purpose pursued by a speaker in a specifiable community of one or more persons, at a specifiable time, in specifiable contexts.”

MISCELLANEA ON TABOO AND EUPHEMISM

In the remainder of this paper I try to shed some light on three less often investigated aspects of language taboo and euphemism. First, I want to address the evolutionary emergence of language taboo and euphemism. The second issue is the acquisition of taboo, as it does not seem clear how something that is “not done” can be avoided, or even learned. And finally, I briefly touch on the prosody of taboo and euphemism.

Evolutionary emergence

A first step to prove if language taboo can be considered part of the evolution of language is to check whether taboo is found in all (or at least most of the) languages. There is evidence (especially in the older literature dating from the first half of the 20th century) that taboo can be found in virtually all languages investigated on this topic, for an overview see Havers 1946,

Meillet 1906. Is taboo then universal? It seems so, because even the language that in recent years has challenged many of the claimed language universals shows clear euphemisms. This in turn points to a taboo that makes speakers avoid certain words. The language that has challenged so many universals is Pirahã (for a starting point in the discussions about Pirahã see Nevins e.a. 2009, Everett 2009a). I am no expert for Pirahã but I checked Dan Everett's homepage (<http://llc.illinoisstate.edu/dlevere/>), as I knew he has an extensive collection of Pirahã data. There I started to read some of the available texts. In the text named "Kato's baby falls near the fire" I found the following passage (Everett 2009, quoted by kind permission of Dan Everett):

xí	oho	-ab	-a	-haí	pixái	-xíga [...]	
3person	eat	-remain	-vertical	-rel:cert	now	-emphatic [...]	
"(I) will eat you right now ('eat' = have sex). [...]"							
ig	-í	-a	ito	-ii	ib	i	-ab
com.	proximate	vertical	point	-thing	hit	-proximate	-remain
-aí	-haí						
-do	-rel:cert						
"With my pointy thing I will hit for a while."							

It was also Dan Everett who kindly informed me that *oho* 'eat' is the standard expression for 'having sex', meaning that there is no other expression. So it is doubtful whether it can be counted as a euphemism or not (but it clearly looks like a faded euphemism). But the expression *ito-ii* 'pointy thing' definitely represents a euphemism as there exists at least one different word for 'penis': *poobahai*.^{*} So it seems reasonable to assume that in this situation, the word *poobahai* is replaced by *ito-ii* because it would seem inappropriate or *taboo* to use *poobahai*. Panagl 1984:153f introduces 3 criteria to validate the assumption of taboo and euphemism. First, there have to be extra-linguistic indications that make the assumption of taboo plausible. The data from linguistic and cultural behavior around the world make 'penis' a very plausible subject to taboo and euphemism. Second, there has to be typological evidence that the term in question is subject to taboo: again this criterion is met as there is abundance of euphemistic strategies concerning 'penis' in languages of the world. Finally, the third criterion is the "euphemism tread-mill", euphemistic expressions turn into the standard expression and therefore tend to be replaced by a new euphemism, and again this criterion is met, as a) *poobahai* also means 'banana', and b) there is *xiitooi* 'stick thing' as an alternative euphemism to *itoii*. This final criterion blends into the assumption of Crombach 2001:179 that there is always an ensemble of co-existing or concurring euphemisms. So, I think it is valid to assume euphemisms in Pirahã, and this makes the assumption of taboo and euphemisms as linguistic universals a little bit more plausible.

Another very common test to see if a linguistic phenomenon is universal is the check for a counterpart in sign languages. Carol Padden and her interpreters informed me in an informal conversation during the EvoLang8 Conference in Utrecht, March 2010 that there are signs that

^{*} Thanks again to Dan Everett! In his emails from September 23, 2010 he writes: One word [for 'penis', CRM] is '*poobahai*'. *Poo* is a prefix, no longer productive, that means long, cylindrical object. But they also call it their 'stick instrument', *xiitooi* (and pencil is paper stick instrument - *kapiiga xiitooi*). banana is '*poogahai*'. Same tones."

should be avoided. Hearing sign language translators are e.g. advised to avoid the gesture *deaf* and use *hard of hearing* or *hearing impaired* instead. One interpreter explained to me, that she personally felt uncomfortable with a widespread “short hand” sign for ‘Chomsky’ as it is a combination of the letter ‘C’ and the sign for ‘god’. But she got used to this gesture, as the gesture is used frequently in linguistic contexts. These data give a clear indication that the linguistic behavior taboo and euphemism is a universal that can be meaningfully investigated in the context of language evolution.

Universals

But if taboo is considered a universal than the next thing to question is the term “universal”. It cannot be shown that all languages have a certain property, as there are thousands of extinct languages we have no knowledge of and therefore it is impossible to prove the universality of a universal. The claim of linguistic universals always lacks falsifiability. But universals can be proven probable. I think the distinction *biological universals* and *functional universals* is a reasonable starting point. *Biological universals* rest on the material foundations of our existence and will most likely be always present in a language, while *functional universals* are emergent properties of the working mind. They represent properties that seem important or fundamental (if not indispensable) for languages to work.

Biological universals

A very good indicator for a biological universal is if it can be shown that it is part of our (primate) heritage. If our close relatives share a trait that seems an essential feature in all languages of the world, we can assume that it is part of the biological basis. Language builds upon this foundation and the linguistic form is the manifestation of something biological. A candidate for such a biological universal is the distinction between proper and common names. To my knowledge it is Derek Bickerton (2009:80) who first pointed out that hardly ever anyone noticed that the language trained apes Washoe, Kanzi, and all the others, never had a problem to distinguish between the proper names of their trainers and the common names for objects. They never called an unknown woman *Sue*, they never wanted to know the name of their breakfast banana. It was clear to them when it makes sense to distinguish the individual object and when this is meaningless. Katz e.a. 1974 showed that the distinction between common and proper nouns emerges very early in language acquisition. Katz e.a. 1974:472 point out that, as they used English as the reference language, it is not absolutely clear whether the young children (17-22 months of age) actually made a distinction between proper names and common nouns or the usage of the article, as in English proper names are never accompanied by an article, while common nouns in most cases have definite or indefinite article, but I am not sure if this is actually an issue. The article could of course represent a clue, but as the development in (southern) German dialects indicates, the article can be used with proper names, so it seems very likely that the children actually distinguished the more basic difference than the grammatical refinements. Above all, it seems a reasonable biological universal to be able to distinguish individuals (at least of the own kind) and common objects (classes). This universal is reflected in language, as to my knowledge all languages have proper names for individuals in contrast to common names for classes of objects. Although the proper name of a person may for various

reasons not necessarily be revealed to everyone, as reported in e.g. Havers 1946, Crystal 2010: another remarkable aspect of language taboo and euphemism...

Functional universals

Opposed to biological universals functional universals may not have a biological basis, but rather they seem necessary for the proper functioning of language. A good candidate for a functional universal is the distinction of 1st and 2nd person singular (Comrie 1989:32). Lyons 1977:640ff presents a hypothetical language without *you* and *I*. But the efforts to have a proper conversation in such a language are considerable. Simple commands turn into complicated constructions, e.g. “*the person speaking wants the person next to him/her to open the window*” for a simple “*could you open the window*”. Therefore Lyons 1977:646 concludes: *deixis [you, I, here, there are always deictic as they can only be resolved semantically correct in the given situation, CRM] in any language ... is something that cannot be analysed away...* There may be of course a biological foundation for this universal, as the phenomenon of self-awareness is attested to the closest kin of humans, the great apes (de Waal 2008:1621). This conception of “I” may at least lead to a biological foundation of “me and the others”, which in turn may be reflected in language. But without a doubt the distinction of 1st and 2nd person singular is a very plausible functional universal.

As a matter of fact most of the biological universals will also be of a functional universal character, but I expect there to be very few biological universals. Figure 1 is a graphical illustration of the types of universals presented here that tries to integrate the distinctions as e.g. presented by Aitchison 1996:185.

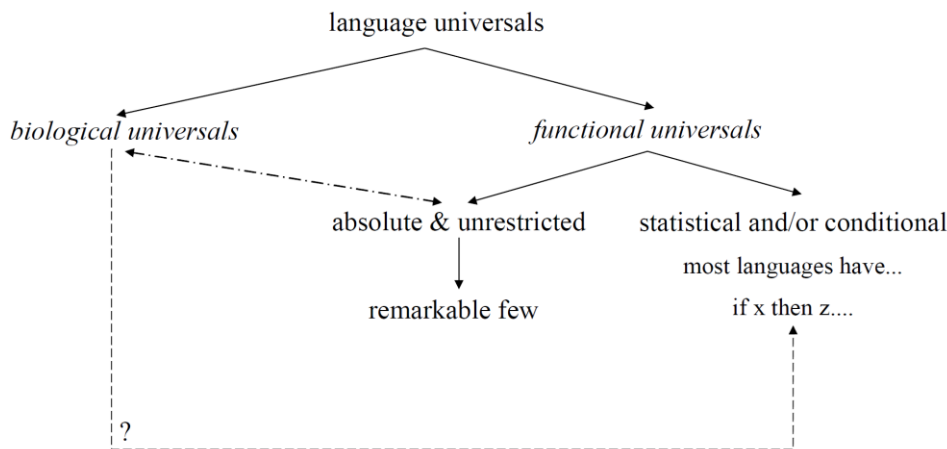


Figure 2: Types of universals

I am still not sure if languages can “choose” not to use a biological universal. But it seems quite possible. The usage of gender may be just such an example. Gender is based on the biological distinction of sex. Sometimes it is certainly of advantage to know if the person or animal referred to is female or male. Still, there are languages that do not use this biological distinction, e.g. Finnish, while others have a (sometimes amusing) mess in the mapping of sex and gender, e.g.

German *Frau Landeshauptmann* lit. approx. ‘Mrs. county-governor-man’. But even genderless languages get things somehow messy: in Finnish even a woman can be an *ovimies* lit. ‘door-man’. So these are very complicated matters, even when ignoring political correctness discussions. Therefore I want to be cautious about the relation between biological universals and their statistical or conditional status, and add a question mark in Figure 2.

Concluding these considerations on language universals I assume that taboo and euphemism are functional universals, necessary for successful communication, and implicative in nature: if there is a taboo, there will be a euphemism; if there is a euphemism detected/used (as in the Pirahã examples above) it is valid to assume a taboo in the given situation (as requested in the definition of taboo by Allan and Burridge) and therefore taboo exists in the language in question.

Words and meanings

In his book on the evolution of language W.T. Fitch 2010 assumes that the Ogden/Richards 1923 triangle (redrawn after Fitch 2010:123 in Figure 3) is one of the biological foundations of language acquisition. Children are born with the built-in assumption that *words mean things*. Bickerton 2009:81 notes that even apes *got it*, meaning that the language trained apes also finally had this insight. This connection between symbol and meaning is another promising candidate for a biological universal. The direct link of objects (or events) and the words (or symbols) is something that is rooted in the biological underpinning of language. (Although the ability for recursive application of this knowledge seems uniquely human; the learning of numbers seems a much more challenging task for animals than for children, cf. Carey 2001, Hauser e.a. 2002, Everett 2009a, Nevins e.a. 2009.) This link between word and objects leads to speech magic: it lets humans ascribe *magical powers to names and words* (Fitch 2010:125, Ogden and Richards 1923:43-65).

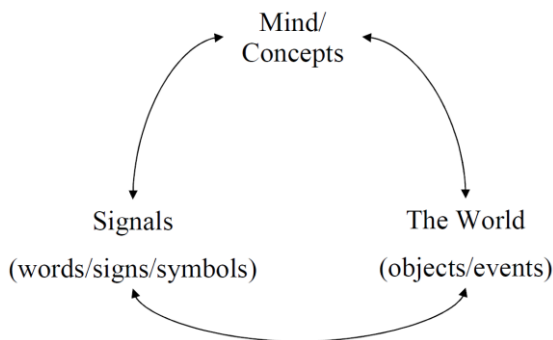


Figure 3: The Ogden/Richards triangle following Fitch 2010:123

The basis of taboo and euphemism are on the one hand the assumption that the name and the thing are strongly linked, and on the other hand the evocative powers. The observation that the right words can make things happen is a necessary insight in the (correct) usage of language. The right word is not only essential in animism, rites or religion. Here the speakers cling to languages long dead and unproductive, like Latin or Sanskrit. But it is also an everyday

experience. The right wording has significant impact on the communicative success. This is “orthophemism” as defined above. It can easily be observed in everyday situations: Let us suppose your mobile phone is for some reason not working as you expect. When you approach the service person in an aggressive manner, like “Your crappy phone is broken. I want my money back and I’ll buy one from your competitors”, the service will most likely not as good as when you approach the person with something like this: “I don’t know why, but every time I want to do this, the phone behaves in a very particular manner. Am I doing something wrong or is there something broken with my phone?” Another example is my son, 8 years of age, who is attending a French school. His mother tongue is Finnish, but the standard communication language at home is German. But if my son wants, for example, to play on the computer, and he estimates the odds to be allowed to do so rather low, then he will certainly address his mother in Finnish, and/or me in French, respectively, as these are the languages he regards (by experience) as more successful in such an impasse. Both examples have of course nothing to do with taboo and euphemism but they illustrate how important the proper usage of language and the usage of proper language are. An interesting question is how taboo and euphemism can come into such an important position within language as they contradict, or at least complicate an assumed primary function of language: information exchange. If the answer to “Where is your Grandfather?” is not a plain and straight “He is dead.” but an ambiguous “He went away.” this loss of information has to be made up for by other benefits. What is taboo and euphemism good for? Obviously, as mentioned before it helps handling the assumed magical powers of language. This very special application has been treated extensively by e.g. Havers 1946. But taboo and euphemism also contribute essentially to aspects of language (linguistic communication) that have been described by Michael Tomasello (2008): imitation, conformity, solidarity, affiliation, norms. It is *necessary* to be polite, cooperative and to avoid insult, in order to successfully use language (see also Knight 2006). These aspects can also be observed in the acquisition of taboo.

Acquisition of taboo words

The exact processes of taboo acquisition are hardly investigated, maybe because they are hard to investigate. Most of the research is done in regard to swear words (e.g. Jay 2000, Montagu 1967), and from this starting point it seems possible to interpolate the various other parts of tabooed language. An interesting aspect in this area is how to resolve the paradox that taboo (language) is forbidden, and therefore should not be used, and still most speakers of a language know these words. Of course there are tabooed words that get lost in the course of history, a well known example is the Indo-European word for ‘bear’ **h₂r̥k̥t-* (as in Greek *árktos*, Sanskrit *ṛk̥śas*, Hittite *hartagga*, Latin *ursus*) that is replaced in the Germanic languages by the euphemism “the brown one”, e.g. English *bear*, German *Bär*. But most of the words that “should not be used” are of venerable age, which implicitly means that they have been handed down quite as all the other words that are not considered taboo or offensive. The “vulgar” words German *Arsch* and English *arse*, both tabooed words in many situations, are of Indo-European heritage **h₃orso-*, (cf. Greek *órros*, Hittite *arra-* etc.), which in turn may represent an Indo-European euphemism for a more blunt **ǵ^hedos*, a nominalization to the verbal root **ǵ^hed-* ‘to shit’. Another remarkable aspect is how taboo the knowledge of taboo knowledge is. This is expressed by the promise

Fallwicl 2009:144 made to her informants on dirty jokes: they will stay anonymous. All this leads to the question: How is taboo (knowledge) acquired?

Following the general assumption taboo words have a low frequency. But in a quantitative approach it is not only necessary to assure that the counted items are what is intended to be counted (Baker 2010:44), but also that the corpus used is meaningful for the issue in question. Tables 1 and 2 illustrate the discrepancy in frequency determined by the corpus chosen. These two tables simply compare the frequencies of *fuck** in two different corpora. The first corpus is an American newspaper corpus compiled in the late 1990s; the second corpus is a collection of film scripts, also collected in the late 1990s. I chose the word *fuck** and its morphological derivatives for two reasons: the first being that this word is treated extensively by Fairmann 2009, and the second one being the finding by MacKay e.a. 2004 that the word *fuck* is on top of the list in all categories for taboo words: familiarity, obscenity, and even recall, when people are tested, how good they remembered taboo words. Table 1 clearly shows that taboo words are not acquired through reading newspapers. But taboo words are used in (admittedly simulated) spoken language as used in movies (table 2).

Table 1: The top 3 *fuck** in an American newspaper corpus (~8.1 x 10⁶ words)

Word	# of occurrences	Word with similar rank	Rank in corpus
fucking	0.37/mio	furlough	1692
fuck	0.37/mio	generically	1692
fuckers	0.25/mio	fruitcake	1693

Table 2: The top 3 *fuck** in an American film script corpus (~7.3 x 10⁶ words)

Word	# of occurrences	Word with similar rank	Rank in corpus
fucking	158,1/mio	deep	565
fuck	144.4/mio	bring	613
fuckers	4.5/mio	haired	1457

But the frequency of the words *fuck** in the actual movies will be even higher, as the screenplay corpus also includes all non-dialog information (e.g. setting, camera angle etc.). For a quick test I prepared a script that I was sure to contain taboo language. I deleted all the non-dialog text and found for the film *Die Hard* (1988, 20th Century Fox, Director: John McTiernan, Screenplay: Steven E. de Souza, Jeb Stuart) the following numbers:

Table 3: *fuck** in the dialogues of *Die Hard* (approx. 8500 words dialogue)

Word	# of occurrences	Word with similar rank	Rank in corpus
fucking	1293,6/mio	were	55
fuck	1058,4/mio	little	57
(mother)fuckers	470,4/mio	wrong	62

These figures should suffice to illustrate that the frequency of taboo words can and will be very high under certain circumstances. Another important factor in the acquisition of taboo words is their emotional value; frequency and feeling are important factors in all learning processes (for the importance of emotions in learning processes see e.g. Hu e.a. 2007). The emotional value points to another important feature of all taboo: only words and things of relevance for the speaker(s) and/or listener can be subject of taboo. To give a simple example, only a religious person can feel offended when someone “insults” her/his god(dess).

I assume five steps in the acquisition of language taboo and euphemism. These steps are not a procedural check-list, but they represent overlapping and recursive phases or a spiral in the development of taboo language use. I can offer here only a first overview on the processes that are involved in the acquisition of taboo. As mentioned before, Tables 1-3 illustrate that “forbidden words” may be forbidden, but that they are used anyway. The friction between facts (the actual use of taboo words) and norms (the interdiction of using tabooed words) adds an emotional value to these words. They seem something special. Taboo words have frequencies far above zero in everyday language use and are actually only banned/sanctioned in certain situations, so there are a lot of opportunities to make the first step (1): learning the words. It seems crucial that taboo words are learned in a very early stage of language acquisition, as Harris e.a. (2003) can show that taboo words are less offensive in the second language (L2) than in the mother tongue (L1). These findings are further supported by Kim e.a. 1997 who illustrate that L1 and L2 stored in distinct cortical areas, and above that taboo and course words are stored separately from other words in the brain (Burridge and Allan 2007 with further references). The last fact seems ultimately related to the high emotional value that taboo words have. Taboo words evoke reactions (see again MacKay e.a. 2004). In the course of acquisition the next step (2) is to learn that these words are not (always) appropriate. At first this is coupled only with a simple interdiction or contradiction “one must not say so”, or “how are you talking” but sooner or later sanctions will follow, e.g. parents quit playing with the “dirty talking kid”, or stop the conversation (Oberlechner 2010). At the same time there will be implicit and explicit instructions about the verbal avoidance strategies. The children acquire the euphemisms. This step (3) is accompanied by step (4): the abundance of use. As probably everybody with children has observed: when children pick up their first taboo words, maybe in kindergarten between 3 and 6 years, they make abundant, sometimes inappropriate, always annoying, use of taboo words. Fairman dedicates his book *Fuck* (Fairman 2009) to his daughter *in the hopes that she will say it less and understand it more*. But the abundance in use also leads to a loss of offensiveness, even for the afflicted parents. And as soon as the reactions are getting rarer the use of offensive languages loses its attraction. In the course of this process the ultimate result is the final step (5) adaptation to the norm. Of course there are tendencies of convergence and divergence (Crystal 2010:53). There are phases when the “forgotten forbidden words” suddenly gain a new attraction, e.g. in puberty. But the usage is also very much context-driven. This is even reflected in the ambiguous jurisdiction in the USA, as illustrated extensively by Fairman 2009. This is an area for future investigation and refinement, where appropriate test procedures have to be developed.

Prosody of taboo language

The last few lines in this paper will be dedicated to impressionistic observations on the prosody of taboo and euphemism. For people change the speed, pitch and volume of their speech when they are using words they consider taboo (for either themselves or their audience). I have not conducted any data analysis up to now. I only rely on observations easily made in every TV sitcom. One of the most obvious prosodic behaviors is yelling or hissing of taboo words, commonly referred to as cursing and swearing. The words are largely de-semanticized in these functions. Fairman 2009:44ff, for example, uses the notation *fuck*¹ and *fuck*², where *fuck*¹ means 'to copulate', while *fuck*² is not referential, and only conveys emotions like dismay, aggression, confusion, etc. Prosodic changes can also be overheard when two persons in a restaurant chat and utter sentences like *I hate to say it, but he is an asshole*. In such situations it is most likely that the speaker will lower the volume of his voice and maybe even check by a quick look if they are overheard, even with verbal euphemistic strategies like *praeterito* as in the example, or euphemistic gestures like e.g. quotation marks in a sentence like *Then she called him "sucker"*. This is another area where a lot of research has to be done.

REFERENCES

- Adams, James N. 1982. The Latin sexual vocabulary. Baltimore: John Hopkins University Press.
- Aitchison, Jean. 1996. The seeds of speech. Language origin and evolution. Cambridge: Cambridge University Press.
- Allan, Keith, and Kate Burridge. 1991. Euphemism and dysphemism. Language used as a shield and a weapon. Oxford: Oxford University Press.
- Allan, Keith, and Kate Burridge. 2006. Forbidden words. Cambridge: Cambridge University Press.
- Betz, Werner. 1978. Tabu – Wörter im Wandel. Meyers enzyklopädisches Lexikon in 25 Bänden. Vol. 23.141–144. Mannheim e.a.: Bibliographisches Institut.
- Bickerton, Derek. 2009. Adam's tongue. How humans made language, how language made humans. New York: Hill and Wang.
- Buck, Carl D. 1949. A dictionary of selected synonyms in the principal Indo-European languages. Chicago: Chicago University Press.
- Burridge, Kate, and Keith Allen. 2007. Taboo, censoring and the human brain. The Fine Print 4.6–19.
- Carey, Susan. 2001. Cognitive foundations of arithmetic: Evolution and ontogenesis [sic!]. *Mind&Language* 16.37-55.
- Comrie, Bernard. 1989. Language universals and linguistic typology. 2nd Ed. Chicago: University of Chicago Press.
- Crombach, Michael. 2001. Euphemismus und Tabu. Universität Salzburg: Department of Linguistics unpublished PhD thesis.
- Crystal, David. 2010. The Cambridge encyclopedia of language. 3rd Ed. Cambridge: Cambridge University Press.
- De Waal, Frans B.M. 2008. The thief in the mirror. *PLOS Biology* 6.1621-1622.

- Desalles, Jean-Louis. 2007. *Why we talk. The evolutionary origins of language*. Cambridge: Cambridge University Press.
- Everett, Daniel L. 2009. Kato's baby falls near the fire. <http://lc.illinoisstate.edu/dlevere/docs/Baby%20falls%20in%20fire.pdf>, retrieved July 22nd, 2010. Used with kind permission of Dan Everett.
- Everett, Daniel L. 2009a. Pirahã culture and grammar: A response to some criticisms. *Language* 85.405-442.
- Fairman, Christopher M. 2009. *Fuck. Word taboo and protecting our first amendment liberties*. Naperville: Sourcebooks.
- Fallwinkl, Mareike. 2009. *Der Witz als Tabubruch. Eine linguistische Untersuchung zum Witz im Sinnbezirk der Sexualität*. Saarbrücken: Verlag Dr. Müller.
- Fitch, W. Tecumseh. 2010. *The evolution of language*. Cambridge: Cambridge University Press
- Frazer, James G. 1888. *Taboo*. *Encyclopedia Britannica*. 9th Ed. Vol. 23.15-18. Edinburgh: Adam and Charles Black.
- Frazer, James G. 1894. *The golden bough. A study in comparative religion in two volumes*. New York: MacMillan.
- Freud, Sigmund. 1974. *Totem und Tabu*. [1913] Mitscherlich, Alexander. Angela Richards. James Strachey (eds). Sigmund Freud. Studienausgabe Vol. 9.287-444. Frankfurt am Main: Fischer.
- Fuchs, Johanna. 2003. *Tabu und Euphemismus eine sprachwissenschaftliche Betrachtung*. Universität Graz: Department of Romance Studies: unpublished MA thesis.
- Harris, Catherine L. Ayşe Ayçiçeği, Jean Berko Gleason. 2003. Taboo words and reprimands elicit greater autonomic reactivity in a first language than in a second language. *Applied Psycholinguistics* 24.561-579.
- Hauser, Marc D. Noam Chomsky. W. Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298.1569-1579.
- Havers, Wilhelm. 1946. *Neuere Literatur zum Sprachtabu*. Wien: Rudolf M. Rohrer.
- Holder, R. W. 1995. *A dictionary of euphemisms*. Oxford: Oxford University Press.
- Hu, Hailan. Eleonore Real. Kogo Takamiya. Myoung-Goo Kang. Joseph Ledoux. Richard L. Hukanir. Roberto Malinow. 2007. Emotion enhances learning via norepinephrine regulation of AMPA-receptor trafficking. *Cell* 131.160-173.
- Jay, Timothy. 2000. *Why we curse*. Amsterdam: John Benjamins.
- Katz, Nancy. Erica Baker. John Macnamara. 1974. What's in a name? *Child Development* 45.469-473.
- Kim, Karl H.S. Norman R. Relkin. Kyoung-Min Lee. Joy Hirsch. 1997. Distinct cortical areas associated with native and second languages. *Nature* 388.171-174.
- Knight Chris. 2006. Language co-evolved with the rule of law. Cangelosi, Angelo, Andrew D.M. Smith and Kenny Smith (eds.). *The evolution of language*. New Jersey: World Scientific. 168-175.
- Lehmann, Franz Rudolf. 1930. *Die polynesischen Tabusitten. Eine ethno-soziologische und religionswissenschaftliche Untersuchung*. Leipzig: Voigtländers.
- Looritz, Oskar. 1939. Gedanken-, Tat-, und Worttabu bei den estnischen Fischern. *Eesti Vabariigi Tartu ülikooli Toimetused* B 45.1-128.

- Lyons, John. 1977. *Semantics*. 2 Vols. Cambridge: Cambridge University Press.
- MacKay, Donald. Meredith Shafto. Jennifer K. Taylor. Diane E. Marian. Lise Abrams. Jennifer R. Dyer. 2004. Relations between emotion, memory, and attention: evidence from taboo stroop, lexical decision, and immediate memory tasks. *Memory&Cognition* 32.474-488.
- Marouzeau, Jules. 1922. Un trait du parler rustique. L'attenuation. *Bulletin de la Société de Linguistique de Paris* 23.28-31.
- Marouzeau, Jules. 1924. Le parler "paysan": Détour et formule. *Bulletin de la Société de Linguistique de Paris* 25.90-94.
- Mead, Margret. 1934. Taboo. Seligman, Edwin R.A. (ed.) *Encyclopedia of social sciences*. Vol. 14.502-505. New York: Macmillian.
- Meillet, Antoine. 1906. Quelques hypothèses sur des interdictions de vocabulaire dans les langues indo-européennes. Meillet, Antoine. 1948. *Linguistique historique et linguistique générale*. Paris: Edouard Champion.
- Mitzka, Walther. Eduard Brodführer. Alfred Schirmer. 1955. *Trübners deutsches Wörterbuch*. Begründet von Alfred Götze. Vol. 6.42. Berlin: de Gruyter.
- Montagu, Ashley. 1967. *The anatomy of swearing*. New York: Macmillian.
- Nevins, Andrew. David Pesetsky. Cilene Rodrigues. 2009. Pirahã exceptionality: A reassessment. *Language* 85.355-404.
- Niemi, Jussi. Juha Mulli. Marja Nenonen. Sinikka Niemi. Alexandre Nikolaev. Esa Penttilä. 2010. Degree of figurativeness of idioms containing 'hand/arm', 'head' and 'eye': A comparison of five languages. Paper presented at the 24th Scandinavian Conference of Linguistics, Joensuu, 25-27 August 2010.
- Northcote, Whitridge Th. 1911. *Encyclopedia Britannica*. 11th Ed. Vol. 26.337-341. Cambridge: Cambridge University Press.
- Oberlechner, Manuela. 2010. Das sagt man nicht. *Salzburger Nachrichten*, 28. September 2010.23.
- Ogden, Charles K. and Ivor A. Richards. 1923. *The meaning of meaning. A study of the influence of language upon thought and of the science of symbolism*. Ed. by John Constable. [2001] London: Routledge.
- Panagl, Oswald. 1984. Was man ungern ausspricht. *Altes und Neues vom Sprachtabu*. *Jahrbuch der Universität Salzburg* 1981-1983.147-158. Salzburg.
- Rudas, Stefan. 1994. Stichworte zur Sozialpsychologie der Tabus. Bettelheim, Peter, and Robert Streibel (eds.) 1994. *Tabu und Geschichte. Zur Kultur des kollektiven Erinnerns*.17-20. Wien: Picus.
- Spencer, Mimi. 2005. *The vagina dialogues*.
<http://www.guardian.co.uk/world/2005/mar/18/gender.uk> [retrieved 8.9.2010]
- Steiner, Franz. 1956. *Taboo*. [1967] Harmondsworth: Pelican.
- Tomasello, Michael. 2008. *Origins of human communication*. Cambridge: MIT Press.

Michael Crombach

The “small word” problem in medical dictation

INTRODUCTION

The recognition performance of “short” or “small” words is remarkably below the expected results in automatic speech recognition (ASR). There are different factors at play, each contributing to the phenomenon. This paper tries to briefly introduce them all and to draw some general conclusions from these observations. There is not the room to give an introduction into the technology that forms the basis of contemporary speech recognition. The basic ideas are that the speech signal is digitized and analyzed giving as a result a recognition tree with different probabilities for various recognition results. These possible acoustic results are compared to a statistical language model that has been derived from a huge amount of text (a very large corpus), and the results of the speech recognition process are the optimized probabilities of acoustic analyzes and word transition probabilities (for a more detailed review of the foundations of ASR see Jelinek 2009).

The data presented in this presentation refer to the usage of the Nuance SpeechMagic product, as the author happens to work in the development of this product. The main application area of SpeechMagic is medical dictation so the data analyzed are medical reports of all kind. The analyses are done by comparing verbatim transcripts of the actual dictation with the finalized and signed report, or by comparing the recognition result with the corrected and finalized text. Error-rates basically can be calculated in two ways: both by counting and comparing the words correctly recognized and the words not recognized correctly (word error-rate, WER), or by comparing the characters that appear in the recognized and the corrected texts (character error-rate, CER). Typically the character error-rate is lower than the word error-rate. Compare the following example: If *does he care* is recognized as **do she dare*, the word error-rate this example would be a 100%, but the character error-rate is only 25%. Because in aligning the sentences character by character then there are only 3 mismatches in the result: the missing *e*, the wrong position of the blank, and finally the substitution of *c* and *d*. The figures given in this paper are WER.

SPEECH RECOGNITION AND LINGUISTICS

Before going into the detailed analysis of the topic in question, it is worth mentioning that it is remarkable how little the influence of linguistics and its findings are on the technical

deployment of language (see also a short remark in Labov 1995). Essentially all “real” issues are solved using mathematical methods or models instead of linguistic findings or methods; some call this the “brute force” approach: processing power, vast amounts of data and memory produce statistics that are just as good as diligent language descriptions. This contradicts the beginnings of modern linguistics; some of Chomsky’s ground breaking works were financed by the US military, obviously in expectation of something military and (by implication) technically usable. Chomsky 1965, for example, has been explicitly founded by the US Army, US Air force and US Navy. How this relates to other activities by Noam Chomsky is presented by Knight 2004. As a matter of fact contemporary linguistics uses modern technologies (e.g. for corpus analyses or spell checking of scientific papers) but hardly contributes to the foundations and advancements of these technologies. This is in opposition to the basic analogies that linguists have drawn from the 19th century on: linguistics as a “hard science” with laws and principles like physics or chemistry. Givón 2008:xxviii argues to abandon these analogies for more appropriate and practicable parallels that can be found in biology. As in biology there are fewer laws and principles, but more trends, tendencies and statistics. There is actually only one linguist that comes to rescue the honor of the whole science: George Kingsley Zipf (1902-1950). His statistical approaches to explain linguistic phenomena are among the basic principles of language, and what’s more, Zipf’s methods form one of the bases of speech recognition. Although Zipf’s approach did not provide an explanation of sound-change, as he originally intended, his frequency tables and analysis revealed a basic property of all languages checked in this respect (Ferrer i Cancho - Solé 2003, Ferrer i Cancho 2006). There is a systematic correspondence between the absolute rank of a word in a corpus and its rank on a frequency table, the “Zipf curve”. This curve explains the phenomenon that only a few hundred, maybe some thousand, words make up 90% of all utterances, although many languages can be shown to have actual lexica of over 300k words (for more introductory information see Crystal 2010:91). Zipf also observed that there is a regular relation between the length and the frequency of words (see also table 1). The more frequent the words the shorter they tend to be. They have to be long enough to make distinctions possible, but not longer. Zipf called this the “principle of least effort”.

Table 1. Distribution short, medium and long words in a 129 k corpus

Grapheme	Distinct words		Whole text corpus	
	Number	Percentage	Number (= REF)	Percentage
All	3788	100.0	129278	100.0
Short	235	6.2	34104	26.4
Medium	1545	40.8	66106	51.1
Long	2008	53.0	29068	22.5

THE SHORT WORD PROBLEM

As can be seen from Table 1 the “short words” contribute a mere 6.2% of distinct units, but these units sum up to 26.4% of the tokens in the corpus. These figures become even more drastic when looking at Table 2 that illustrates the error rates of short words compared to other medium and long words.

Table 2: Performance comparison according to word length

Grapheme	REC	COR	DEL	INS	Recall	Precision	F-Measure
All	125495	110750	6127	2344	85.67	88.25	86.94
Short	32933	26844	3404	1270	78.71	81.51	80.09
Medium	65742	59101	2129	973	89.40	89.90	89.65
Long	26820	24805	594	101	85.33	92.49	88.77

The performance of distinct words or more generally groups of distinct words is measured using recall, precision and F-measure.

$$\text{Recall [\%]} = \text{COR/REF} * 100$$

represents the recognition rate, while

$$\text{Precision [\%]} = \text{COR/REC} * 100$$

determines the accuracy of the recognition.

COR gives the number of correct recognitions. *REF* is the number of occurrences in the (corrected) reference texts, the target. Finally, *REC* denotes the number of word occurrences in the recognition results.

F-measure is calculated as the harmonic mean of recall and precision:

$$\text{F-measure [\%]} = 2 * \text{Recall} * \text{Precision} / (\text{Recall} + \text{Precision}) * 100.$$

Table 2 illustrates the performance in relation to the length of the words. The numbers of REC and COR are given in Table 2. REF is given in the 4th column of Table 1. Short words clearly exhibit the weakest performance with recall and precision of ~80%. Their F-measure is almost 7% below the average for all words. Medium size and long words both perform above average. This means that the short words’ performance has significant influence on the whole system’s performance. Furthermore, Table 2 lists the number of erroneously inserted (INS) or deleted (DEL) words. Approximately 56% of all deletions are caused by short words. Likewise, about

54% of the insertions can be attributed to short words. When this is related to the fact that short words only make up 26.4% of the text corpus, then the ratio of deletions and insertions within the short words is significantly larger than for medium size and long words.

“SMALL” VS. “SHORT” WORDS

This introductory analysis of the recognition performance becomes even more interesting when taking a second look at the “short words”. It then shows that it is not only the shortness of the words that unites this group; their high frequency is related to their mainly grammatical function (see Table 3). These short words (as are endings for inflecting languages) are prone to be reduced in phonetic and prosodic realization (especially in English), while they form the basis of grammatically well-formed sentences. The words that cause the issues are these frequent function words; in the last years there was only one observed instance of a “short” word with recognition performance problems. The issue reported to the customer support and investigated by the R&D department was that in the UK English version of the ASR product the word *arm* [a:m] was “never recognized”. The analysis of the collected data revealed that this word was homophone with the “hesitation word”, a special “word” (actually an invisible recognition unit) that is used in speech recognition to prevent the *ahms*, *ehms*, *hmmms* from interfering with the recognition result. The “hesitation word” is by definition not output and had for UK English (amongst others) the transcriptions [a:m]. As the hesitation is (by its nature) much more frequent and thereby more probable than the word *arm* the recognition engine preferred on statistical reasons the more frequent representation of the phoneme sequence [a:m]. The problem could simply be resolved by removing the [a:m] transcription from the “hesitation word”. This example must suffice to illustrate that it is not the shortness alone that is the problem, but the combination of frequency, function and reduced/sloppy pronunciation. Therefore it makes sense to label these words “small words” as opposed to “short words”.

Table 3: the 15 most frequently inserted and/or deleted short words

Grapheme	REF
the	6531
is	3453
of	2940
and	2576
no	2530
are	1998
in	1501
two	1246
or	1131
a	1079
to	755
be	450
at	343
not	313

SMALL WORDS

So, the real issue are the “small words” that are remarkable in many ways. They are acquired late in the course of language acquisition (e.g. Pinker 1995), they are hard to learn in a second language (Ma 2009), and hardly ever handled absolutely correctly thereafter (e.g. Tetreault 2010). Finally, as Givón 2009 and Dessalles 2007 suggest, they may mark the final transition in language evolution from protolanguage to language. Actually, they mark the transition from pidgin to creoles. There is another fact that makes small words special: in a corpus of verbatim transcripts of dictated medical reports (~410 k spoken words) they expose the habit of obviously not being used at all. When the verbatim transcripts are compared to the actual finalized texts, it turns out that approx. 13% of the text appearing in the finalized reports cannot be found in the verbatim transcripts. The majority of these not-dictated items are punctuation marks and formatting commands, like “period” and “new line”. But about 2.5% of the words missing in the spoken corpus are words that are necessary to have well-formed sentences, or are otherwise linguistically and/or grammatically relevant: *small words*.

Table 4 gives the top 10 of the not-dictated items. As mentioned above the performance of small words is in itself below average, and this is enforced by this behavior.

Table 4: Not-dictated words in a ~410k verbatim transcribed medical report corpus

ND count	Item	% of ND	% Final	
2035	the	15,07	16,72	ND count: how often not dictated
1206	and	8,93	13,32	
798	a	5,91	15,41	% of ND: percentage within the not dictated items
444	of	3,29	5,38	
347	in	2,57	7,77	
286	to	2,12	4,39	
270	is	2,00	5,08	% Final: percentage of the ND items against the total occurrences
263	with	1,95	5,33	
228	he	1,69	8,86	
197	or	1,46	11,50	

In the following the contributing factors are briefly introduced. Each of these may not account for all instances, but contributes to this behavior.

Factor 1

Grammatical well-formedness is to some extent certainly an overrated idea; it is a necessary idealization that is not met by reality (Givón 2009). Utterances are not sentences, therefore speech is not necessarily grammatical; on the other hand grammaticality is not gradable (Haider 2007). So, in a strict interpretation the sentences dictated that, for example, lack the definite or indefinite article are grammatically incorrect. But there are different qualities of ungrammaticality. The omission of an article or a preposition does in itself not make an

utterance incomprehensible. Another factor is certainly that grammatical “well-forming” happens twice, first in speech production, second in speech perception. The expectation of grammaticality leads to interpolations of omitted, semantically empty but syntactically necessary parts of speech. It would be an interesting analysis to check how much of the verbatim transcripts are in itself influenced by interpolation by the (verbatim) transcribers. It is to be expected that there are even more not-dictated small words, but the efforts to conduct such an analysis are considerable and the data has to be prepared and selected very carefully, and even then such an investigation will be very error-prone.

Factor 2

This could be seen as a sociolinguistic approach. This aspect is linked to the findings of Tesak-Libben 1994, Tesak e.a. 1995, Tesak-Niemi 1997. These investigations show that there are registers of language that have their very own rules or grammar, e.g. the language of telegrams (telegraphese), SMSes etc., and obviously also spoken medical reports do not match the expected grammar of a written medical report. This puts emphasis on the highly specialized nature of this corpus. The language used in these dictations shows characteristics of a medical dictation jargon that licenses abbreviations within the syntactic structure, without being ungrammatical. This is a kind of economic behavior that leads to the omission of “redundant” words within the sentence. It is more a kind of “shopping list” dictation style that omits articles, conjunctions; list-like parts of information like e.g. ... *they relocated her knee cap dislocation [and] they relocated her patella ...* are put together. This is a different form of the “principle of least effort”: words that are not actually needed for the meaning can be omitted on economical grounds (Tesak e.a. 1995 with more references). It is not necessary to use the article when dictating the word *patient* that in turn is just a placeholder for a name, as a consequence *the patient* becomes *patient* (which in turn just serves as a “dummy” NP that sets medical reports apart from e.g. telegrams). The transcriptionists insert the article, to provide a proper sentence and certainly also because of factor 4 below: extra keystrokes.

Factor 3

Some of the missing small words can be explained by the work style. Dictation is not writing. Dictation is not speaking. It is a very own and special form of language use. Actually most of the doctors do not even care or know about the fact that they are talking to an ASR system (Leijten 2007 analyses the changes in behavior when users actually know they are talking to a machine, therefore her findings are not relevant for this use case). The recognition results reflect what Kleist around 1805 described in the sentence *l'idée vient en parlant*. The utterances are not preconfigured sentences but are formed while the speaker remembers the recent patients, procedures, diagnoses etc. This necessarily leads to unfinished sentences and fragments that are not in a traditional sense grammatical but a reflection of the process of thought.

Factor 4

Finally, there is the reformatting that takes place mainly to smoothen the differences between spoken and written language. It is the streamlining according to the rules of medical

transcriptions. A typical example is the replacement of contractions with the full forms, e.g. *that's* becomes *that is* in the finalized report, as required by Hughes 2002. But these undone contractions cannot account for most of the not dictated items as they have been tagged separately in the corpus. But as mentioned earlier, it cannot be excluded that the verbatim transcripts themselves contain corrections and conjectures that influence these cases. There is another important aspect that has to do with the transcription process: transcriptionists are paid by characters or lines. If there is an easy way to add keystrokes like an *and* between two sentences, or a *the* in front of every *patient*, the transcriptionist will certainly do so.

CONCLUSIONS

These four factors together may account for the *not-dictating behavior*, but the general weakness in the performance of small words in speech recognition may root in a mathematical phenomenon that has to do with Zipf curves. Actually, it is a psychological illusion. If over 25% percent of any given text are small words then it is very likely that within any given error-rate a fourth of all affected words is going to be one of the small words. And the small words in themselves again follow a Zipf curve (as does any subset of a Zipf curve; at least down to a reasonable number that may be something larger than 10) meaning that only very few of the small words account for most instances of small words. If then something goes wrong with small words it is very likely that only a very small number of words are again affected, and this in turn creates the psychological effect of *it never works...*

QUOTED REFERENCES

- Chomsky, Noam. 1965. Aspects of the theory of syntax. Cambridge: MIT Press.
- Crystal, David. 2010. The Cambridge encyclopedia of language. Cambridge: Cambridge University Press.
- Desalles, Jean-Louis. 2007. Why we talk. The evolutionary origins of language. Cambridge: Cambridge University Press.
- Ferrer i Cancho, Ramon. 2006. On the universality of Zipf's law for word frequencies. Grzybek, Peter and Köhler, Reinhard (eds.) Exact methods in the study of language and text. In honor of Gabriel Altmann. Berlin: de Gruyter. 131-140.
- Ferrer i Cancho, Ramon. and Richard V. Solé. 2003. Least effort and the origins of scaling in human language. PNAS 100.788-791.
- Givón, Talmy. 2009. The genesis of syntactic complexity. Amsterdam: Benjamins.
- Haider, Hubert. 2007. As a matter of facts. Theoretical Linguistics 33.381-394.
- Hughes, Peg (Ed.) 2002. The AAMT book of style for medical transcription. Modesto: AAMT.
- Jelinek, Frederick. 2009. The dawn of statistical ASR and MT. Computational Linguistics 35.483-494.
- Kleist, Heinrich von. 1984. Sämtliche Erzählungen und andere Prosa. Stuttgart: Reclam.
- Knight, Chris. 2004. Decoding Chomsky. European Review 12.581-603.

- Labov, William. 1995. The case of the missing copula. The interpretation of zeros in African-American English. Gleitman, Lila R. and Mark Lieberman (eds.). *Language. An invitation to cognitive science*. Second edition. Volume 1. Cambridge: MIT Press.25-54.
- Leijten, Mariëlle. 2007. *Writing and speech recognition*. Utrecht: LOT.
- Ma, Quing. 2009. *Second language vocabulary acquisition*. Bern: Lang.
- Pinker, Steven. 1995. Language acquisition. Gleitman, Lila R. and Mark Lieberman (eds.). *Language. An invitation to cognitive science*. Second edition. Volume 1. Cambridge: MIT Press.135-182.
- Tesak, Jürgen and Gary Libben. 1994. On economy in written language production: telegrams, short letters, and notes. *Grazer Linguistische Studien* 42.115-120.
- Tesak, Jürgen, Elisabeth Ahlsén, Gábor Györgi, Päivi Koivuselkä-Sallinen, Jussi Niemi & Livia Tonelli. 1995. Patterns of ellipsis in telegraphese: A study of six languages. *Folia Linguistica* 29. 297-316.
- Tesak, Jürgen and Jussi Niemi. 1997. Telegraphese and agrammatism: A cross-linguistic study. *Aphasiology* 11.145-155.
- Tetreault, Joel, and Jennifer Foster, and Martin Chodorow. 2010. Using parse features for preposition selection and error detection. *Proceedings of the ACL 2010 conference*. Stroudsburg: Association for Computational Linguistics. 353–358.

ACKNOWLEDGEMENTS

Tables 1 and 2 are modified versions of data collected and prepared by Peter Skala and Maria Hörndl for company internal analyses. “Referee C” made many valuable suggestions to improve this paper.

Grigory E. Kreydlin and Svetlana I. Pereverzeva

Body parts and their names in Russian: the biological and semiotic pairs of body parts

INTRODUCTION. THE SEMIOTIC CONCEPTUALIZATION OF THE HUMAN BODY

In studying human dialogs a linguist should take into consideration that in interaction people use not one but several semiotic codes.¹ One of them is the natural language (or languages); others are nonverbal semiotic codes, basically, the so-called body language. How people sit or stand, whether they look at each other or not, what distance they keep in different communicative acts, how they orient their bodies and parts of bodies, etc. – all these aspects of nonverbal behavior should be accounted for and described.

Gestures of hands, legs, head and shoulders, meaningful facial expressions, glances and touches, sign movements and facial orientations are usually considered separate conceptual and cognitive abilities. However, the evidence from communicative practices and theories of dialog, nonverbal semiotics and cognition suggests that the sets of nonverbal signs are not independent. Being formally and functionally different, they interact and make up an integral part of the speaking process. Nonverbal signs of each type combine with speech in production, arrangement and comprehension of meaning. The mechanisms, regularities and formal rules of interaction between verbal and nonverbal units, as well as the meanings and forms of co-speech gestures and paralinguistic units have been discussed in many articles and monographs; see, among others, McNeill 2000, Kendon 2004, Enfield 2009. These works demonstrate that nowadays linguists of different professional interests and specialists in nonverbal communication and other cognitive sciences join their efforts in adjusting conceptual tools and metalanguages.

The explorative fields to which our paper belongs are both semiotics and linguistics of the human body and corporeality. We study physical, structural and functional characteristics of the body as they are reflected in Russian verbal and nonverbal semiotic codes, mainly in the Russian body language. The insight into these characteristics helps to understand how Russian speakers talk about human bodies, parts of bodies, internal organs and other somatic objects, as well as to

¹ This work has been supported by the Russian Scientific Foundation for the Humanities (grant № 10-04-00125a).

find out how they present their bodies in communicative acts of various topics, styles and genres.

We have introduced elsewhere the notion of semiotic conceptualization (SC) of the human body. The SC is a formal model of the naïve semiotic world view and a generalization of the notion of linguistic conceptualization of the world. This means that the SC pertinent to the Russian culture and language provides theoretical and methodological tools and suggests some practical instruments for describing Russian bodily vocabulary and grammar.

The task of constructing the SC of the human body has been brought forward within the research project “Body parts in the Russian language and culture” we are working on. The project aims at solving three explorative problems: (1) constructing the SC of the body as it is represented in the Russian language and body language, (2) comparing the expressive possibilities that these semiotic codes possess, and (3) providing common ground for cross-linguistic, cognitive and cultural analyses of the SC.²

The kernel of the SC is formed by several sets of units. These are (1) the set of somatic objects, which includes the body itself, body parts, parts of body parts, organs, corporeal liquids, bones, etc., and the set of their names (e.g. *telo* ‘body’, *golova* ‘head’, *nos* ‘nose’); (2) the sets of features of somatic objects and their names, such as *forma* ‘form’, *razmer* ‘size’, *mestopolozhenije* ‘location’, *struktura* ‘structure’, *funkcija* ‘function’; (3) the sets of values of the features and their names (e.g. *dlinnyj* ‘long’, *prjamoj* ‘straight’, *vlazhnyj* ‘wet’). The SC also includes (4) the sets of typical gestures, facial expressions, postures and other nonverbal signs performed by/over somatic objects and the set of their names (*rukopozhatiije* ‘a handshake’, *ulybka* ‘a smile’, *sidetj na kortochkakh* ‘to squat’).

The objective of the paper is to discuss the conceptualization and linguistic representation of one feature of somatic objects, i.e. *the ability of certain somatic objects to group into pairs*³. This feature characterizes only several groups of somatic objects, i.e. body parts, parts of body parts and organs; it is not pertinent to liquids, muscles, corporal vessels and hair, horny or cutaneous coverings. In part 1 and 2 we will introduce the notions of *biological pairs* and *semiotic pairs of somatic objects* and consider some typical Russian lexical units which name somatic objects combined in pairs. In part 3 we will propose a generalization of the notion of *semiotic pairs* – the *unified somatic complexes*. Part 4 discusses the notion of salience of the member of a semiotic pair.

We base our investigation on linguistic data included into several electronic corpora, first of all, the Russian National Corpus (www.ruscorpora.ru). Sometimes we present the experimental data that display correct or incorrect usage of the linguistic names in different contexts. Besides, we analyze some of those Russian gestures that are performed by the members of a semiotic pair.

² The ideas, aims and methods of this project, as well as the notion of the SC of the body are discussed in many works written by the participants of the project, e.g. Kreydlin and Letuchiy 2006, Arkadyev *et al.* 2008, Kreydlin 2009.

³ We name this feature in such a complex manner because there is no one-word English equivalent for the Russian term *parnost'* (from Russian *para* ‘a pair’) denoting this feature.

1 BIOLOGICAL PAIRS OF SOMATIC OBJECTS

Two objects, as well as two properties, events, situations, etc., form a pair when they are acknowledged as a single whole. Arms, hands, legs, feet, shoulders, ears, nostrils are combined in pairs because they consist of two parts with the similar structure and function considered from anatomical and physiological points of view. These somatic objects are said *to form a natural biological pair* for we see them both (they are *perceived by sight*) and they *function* as if they were one object.

However, biological pairs may include invisible somatic objects as well. Although the heart, the brain, the kidneys and some other somatic objects are not visible (certainly we may see the *images* or videos of these organs, but not the organs themselves), they are regarded as united wholes. People come to know that these objects are combined in pairs through the process of education. For example, they learn that the heart contains the left and the right auricle, that the brain consists of the left and right hemispheres and that the kidneys are both on the left and on the right side of the body. The information about these somatic objects is not *perceptual*; it is *conceptual*, or *cognitive*.

We distinguish two types of biological pairs -- the *perceptual* and *cognitive biological pairs*. They are related to different types of world view: the first type correlates with the ordinary, or naïve semiotic world view and the second one – with the scientific world view. The fact that lungs or kidneys have two parts which are both important for the human life unsophisticated speakers get to know only in the process of education. This explains why in everyday conversations the Russian words *legkije* ('lungs') and *pochki* ('kidneys') are used in the plural form much more often than their singular counterparts *legkoje* ('lung') and *pochka* ('kidney'). The collocations *levoje/pravoje legkoje* ('left/right lung') or *levaja/pravaja pochka* ('left/right kidney') belong not to the nucleus of Russian but to its medical sublanguage. Ordinary people usually name the members of these pairs only in specific situations such as health diagnostic procedures, illnesses, physical damage and so on.

Another example of cognitive biological pairs is the two hemispheres of the human brain. Russian speakers refer to these somatic objects using the word *mozg* ('brain', Sg), despite human brain includes two main parts. These are called *polusharija* ('hemispheres', Pl). Each hemisphere has its own name – *pravoje/levoje polusharije* ('left/right hemisphere'). The word *mozg* denoting the whole is not used when we talk about its separate parts. In other words, Russians never use the word *mozg* implying *pravoje* or *levoje polusharije*.

But there is another Russian word – *mozgi* ('brains', Pl). Although being the plural of the word *mozg*, it does not mean 'two parts of the brain' – in contrast to the words *legkije* (the plural form of *legkoje* 'lung') and *pochki* (the plural form of *pochka* 'kidney'). One may point out one more important difference between the words *legkije* ('lungs') and *pochki* ('kidneys') regarded together and *mozgi* ('brains'): the first two words have the singular forms which both refer to the single part of the respective whole, and the last word in some of its meanings has a singular form too (*mozg*), but it never refers to a part of the whole.

The Russian word *mozgi* ('brains') is polysemic. Furthermore, it is linked with the word *mozg* in quite a peculiar manner. We will focus now on the structure of its polysemy.

The primary meanings of the word *mozgi* are:

- (a) 'two or more somatic objects each named *mozg'* (cf. *Medikov interesujet strojenije mozgov velikikh ljudej* 'the doctors are interested in the structure of brains of famous people'). This is the usual "arithmetic" meaning of the plural. The definition shows that if one wants to enumerate brains of different people s/he must use the word *mozgi*;
- (b) 'somatic object placed inside the head and used for thinking' (cf. *Mozgi plavjatsja ot zhary* lit. 'brains are melting in such heat');
- (c) 'intellectual power of a human being' (cf. *U nego khoroshije mozgi* lit. 'he has good brains');
- (d) 'intellectual power of people regarded as a single whole' (cf. *utechka mozgov* 'brain drain').

This definition is a correction of some dictionary entries for the word *mozgi*. We assert that the word has at least four separate meanings, or four separate lexemes⁴. Some of these lexemes can be used in texts metonymically, for example, the word *mozgi* can denote 'human being with outstanding intellectual power': *Eti deti – luchshije mozgi nashego klassa* lit. 'These children are the best brains in our class'.

2 SEMIOTIC PAIRS OF SOMATIC OBJECTS

In constructing the SC of the human body one has to distinguish between the properties that characterize somatic objects and those that characterize their linguistic names.

Thus, the notion of a biological pair introduced in part 1 applies to somatic objects, and the notion of *semiotic pair*⁵ that we will define below applies to the linguistic names.

Somatic objects that are biologically paired are called *semiotically paired* if they satisfy three conditions (so this definition is much more complex than that of the biological pair; besides, the notion of a semiotic pair is a derivative of the former notion):

- (1) the most frequently used name⁶ of the biological pair is in the plural (cf. *ruki* 'arms', *nogi* 'legs', *ushi* 'ears');
- (2) each member of the pair has a name, which is either (a) a word in the singular form that coincides with the name of the group (*ruka* 'arm', *noga* 'leg', *ucho* 'ear') or (b) a collocation of words, one of which is the noun in the singular form and the other is an adjective denoting spatial location of the referent of the noun. The location usually

⁴ We put aside the denotative and connotative meanings as well as metaphorical usages of the word *mozgi* 'brains' in order not to complicate the basic line of the composition of the paper.

⁵ The notion of a *semiotic pair* applied to shoulders was first introduced in the article Kreydlin and Letuchiy 2006.

⁶ In this case the Russian linguistic tradition prefers to use the term *standard name*.

relates to one of the two axes of the human body – the vertical one (more often) or the horizontal one (less often)⁷;

- (3) the name of the pair and the names of each member of the pair belong to the lexicon of the everyday language, i.e. they should not be regarded as terms. This implies that the native speakers know how to pronounce and spell these names, and that the names are pragmatically adapted, i.e. people know their morphological structure, their meaning and usage.

The examples of semiotic pairs, or semiotically paired somatic objects, are the right hand and the left hand that are named in Russian *pravaja ruka* and *levaja ruka*, respectively.

Indeed, the plural form *ruki* ‘hands’ is the name of a biological pair (condition 1). Both units refer to the members of this biological pair, and *levaja* and *pravaja* are the adjectives of location (condition 2). The collocations considered are certainly pragmatically adapted, because Russians understand such phrases as *On – moja pravaja ruka* ‘He is my right-hand man’, *Ja sdelažu eto odnoj levoj (rukoj)* ‘I’ll do it easily’ (lit. ‘I’ll do it with my left (hand) only’)⁸ and the like. They know idioms *idti ruka ob ruku* ‘to walk hand in hand’, *protjanutj ruku pomoschi* ‘to lend a hand’, *iz ruk v ruki* ‘from hand to hand’, and comprehend the cultural connotations of the words *levsha* ‘left-handed person’ or *levizna* ‘leftism’ (condition 3).

The feature that is attributed to the semiotic pair is usually identical to the one attributed to a member of the pair, but the values of the features may be different. For example, if a person attributes the feature of thickness to the object “legs”, s/he obligatory attributes the same feature to each leg; however, one leg of the person may certainly be a bit thicker than the other one.

Semiotically paired somatic objects are well suited to express actions or motions of a parallel or symmetrical character. Walking, sitting and standing, blinking (in contrast to winking!), rubbing hands, applauding, closing eyes when falling asleep and opening them awakening or the action of tug-of-war are carried out with both paired objects. If someone performs iconic (instrumental) gestures demonstrating the real objects that are parallel in form like rails or keys on a keyboard or metaphorically showing the abstract objects of geometrically parallel form like lines, s/he usually does it with both hands.

The semantic or communicative role of the members of semiotic pairs may be different. When a person is *sitting with his/her legs crossed*, one leg is active while the other one is passive, but when s/he is *standing* both legs are active. Sometimes, in their musical works, the composers assign the active role to one hand. That means that it plays the leading part in the musical composition. The composers reflect their intentions by naming their works like ‘Pjesa dlja levoj ruki’ (lit. ‘Piece for the left hand’).

Concluding the discussion of semiotically paired somatic objects, we want to stress one important distinction between biologic and semiotic pairs. Although the definition of the

⁷ For the ordinary Russian language the sagittal axis is much less significant than the two other axes, but in the medical sublanguage it can also play an important part, cf. collocations *perednjaja* resp. *zadnjaja stenka zheludka* ‘anterior resp. posterior wall of stomach’.

⁸ Cf. the English idiom *to work with the left hand* ‘to work negligently’ that corresponds to the Russian *rabotatj spust’ a rukava* (lit. ‘to work with the sleeves rolled down’).

semiotically paired somatic objects is based on the notion of biologically paired objects, the first one is linguistically and culturally specific and the second one is universal.

3 UNIFIED SOMATIC COMPLEXES

A semiotic pair consists of two objects. However, some somatic objects such as fingers, teeth, bones, the right hand's fingers, the left leg's bones, etc. exist in quantities more than two and make up natural groups because of their similar structures and functions. Fingers, the right hand's fingers, teeth, etc. form *unified somatic complexes* (USC-s).

The members of a USC, though called together unified, may have different functions and linguistic names.

A typical example of the USC is the fingers. The forefinger and the little finger both participate in pointing, so they have at least one common function, but the sets of their functions do not coincide. In particular, the forefinger, and not the little finger, is used in the gesture *to raise one's finger* (to attract people's attention). The nature of their names is different as well. The forefinger is called thus because of its location, and it has another name, *index finger*, which is given in accordance with its function. In contrast to the forefinger, the little finger is called only by its size. In Russian the word *mizinec* 'little finger' denotes the youngest son in a family; therefore, it is a metonymic name.

In spite of the obvious difference among functions, forms, names, etc. all fingers participate jointly in the most important actions of the hand like holding and grabbing some relatively small objects or clenching a fist. It is the fist that symbolizes the unity of the fingers of one hand and, more generally, the unity of many other sorts. The ability of the fingers to join in a fist is the feature that permits us to talk about the fingers as a USC.

In Russian the USC-s linguistically marked. We will mention only several linguistic markers and group them into categories. These are: (1) lexical markers, e.g. *pishchevaritel'naja sistema* 'alimentary system', *sistema dykhanija* 'respiratory system', *vnutrennosti* 'guts', *konechnosti* 'limbs', *cherty lica* 'facial features', etc., (2) morphological markers, e.g. the plurals of names of somatic complexes (cf. *nogi* 'legs', *palcy* 'fingers'), (3) syntactical markers: the USC-s can be codified by the coordinating construction (*On moi glaza i ushi* 'He is my eyes and ears'), by the binomial construction (*Ruki-nogi perelomaju!* lit. 'I'll break your arms and legs!') or with the aid of supra-phrasal units (cf. various descriptions of the body's appearance and behavior, e.g. the concepts of fighting styles or walking styles, table manners, etc.).

Each unified somatic complex can be treated as a natural analog and useful generalization of the semiotic pair.

4 THE SALIENCE OF A MEMBER OF THE SEMIOTIC PAIR

In the previous part we have suggested the natural generalization of the notion of semiotic pair. Now we will return to the basic notion and discuss some qualities of its members.

According to the definition of the semiotic pair the feature that differentiates the members of a pair is their spatial location. This is a structural feature of somatic objects. But from the semantic, pragmatic and cultural points of view some other features, structural, physical or functional, are more significant for the semiotic conceptualization of the human body.

In different contexts or semantic types of contexts one member of a semiotic pair appears to be more *semantically* or *culturally salient* than the other. We will give two examples to illuminate the difference between the *semantic* and *cultural salience*.

To express the meaning 'to do something easily, without any effort' Russians use the idiom *sdelatj chto-to odnoj levoj* lit. 'to do something with the left (hand) only' whereas the expression **sdelatj chto-to odnoj pravoj* (lit. 'to do something with the right (hand) only') does not exist despite most people being right-handed. The existence of the idiom *sdelatj chto-to odnoj levoj* is not accidental. In the context of physical actions performed with only one hand, the left hand is semantically more salient than the right one. On the contrary, if we want to express the complex idea that one person needs another to do something s/he wants to do and that this person cannot do it without another person's help, we use the idiom *moja pravaja ruka* 'my right hand'. In this context the right hand is marked, or salient.

The example considered demonstrates that both members of a semiotic pair can be semantically salient, but their salience manifests in different contexts.

The notion of cultural salience of a member of the semiotic pair embraces two cases: (1) the difference in the socially and ritually marked contexts within one culture and (2) the difference in cultures. Examples:

- (1a) In the Russian culture the married people wear the wedding ring on the right hand, whereas the divorced people wear it on the left hand.
- (2a) Left-hand gestures are often taboo in different West African cultures⁹, though in the Western cultures they are not tabooed.

The following example of the difference in cultures and the salience associated with this difference concerns not gestures but modes of nonverbal behavior in a dialog.

- (2b) Muslims consider the left hand dirty; they even often call it "toilet". This is why Muslims do not use the left hand in giving or accepting gifts.

The salience of a somatic object within the semiotic pair may have different reasons. We will enumerate only a few of them.

⁹ Cf. Ameka and Breedveld 2005.

(1) (biological reasons). One member of the pair becomes salient due to the biological specifics of a human being. Since our world is oriented towards the right-handed people the word *ruka* is interpreted as 'the right hand' in many linguistic contexts, e.g. *On privetstvenno protjanul ej ruku* 'He stretched out his hand to greet her'. Normally, in such contexts the word *pravij* 'right' is omitted, otherwise the sentences or texts are regarded as stylistically marked. The word *ruka*, of course, does not always mean 'the right hand', cf. *Cto u tebj s rukoj?* 'What's wrong with your hand?', *On nosit na ruke koljco* 'He wears a ring on his hand', etc. When uttering these phrases one does not presume that something is wrong with the right hand or that the person wears the ring on the right hand. Unfortunately, linguists do not know the rules that explain the omission of the word *pravij* so far; nor do they know how to distinguish the contexts in which *ruka* means 'the right hand' from the contexts in which *ruka* means 'hand'.

One of the biological factors that governs the human behavior in a real life and provokes the salience of the member of the pair is *the principle of physiological convenience*.¹⁰ It can be stated as follows: If a person wants to perform a particular action and has several somatic objects available as instruments for this action, then s/he chooses the *most convenient* object. This principle states that the salient member of the semiotic pair can be the most convenient one.

Sometimes the principle of physiological convenience comes into conflict with another principle of human behavior, that of *perception of spatial proximity of objects*. This principle asserts that a person who intends to perform a particular action on an object that is located in his/her real spatial domain using one of his/her body parts as an instrument for this action chooses the part that is the *closest* to the object. Thus, this principle states that the salient object is the closest one.

Let us consider the sentence *Malchik otodvinul stojashuju na stole tarelku* 'The boy moved the plate on the table away' as an example of the conflict of the two principles. Suppose the boy is right-handed. According to the principle of physiological convenience he employs his right hand for most everyday actions, and the action *to move away the plate* is not an exception. However, if the plate is on his left he is most likely to move it with his left hand. Here we ground our reasons on the principle of perception of spatial location, and this principle "wins" here over the first one.

(ii) (linguistic factors). One member of the semiotic pair may be linguistically marked as salient. For example, there is a Russian idiom *On – moja pravaja ruka* 'He is my right-hand man'. Here the adjective *pravaja* 'right' can be neither omitted nor replaced by another lexical unit; it is a constructively obligatory element of the idiom. In this case the semiotic salience of the right hand is caused by linguistic factors.

¹⁰ We maintain that the concept of convenience is a very important one. It should be related to different situations of usage, both linguistically and semiotically marked. One kind of convenience is a physiological one. When describing the gesture *to close the face with the hands* in the gesture dictionary one shouldn't burden the description with the propositions explaining that the right/left hand closes the right/left part of face. Instead, one may formulate the principle of physiological convenience that embraces this case and the similar ones. Another case of convenience is linguistically marked with the Russian word *udobno* 'convenient; proper'. We often ask *Udobno li zadatj takoj vopros?* 'Is it proper to ask such a question?'. This is an etiquette convenience. The elaboration of the concept of convenience and the taxonomy of types of convenience are stimulating topics for future investigations.

(iii) (cultural factors). Some somatic objects become salient when used as cultural signs or when included in culturally significant actions. Thus, there are rules and (not so strong) regularities that determine the culture-specific modes of carrying weapons, clothes or jewellery. For example, swords are carried on the left side of the body; the symbolic spitting is performed over the left shoulder¹¹. These rules of etiquette and social behavior as well as the regularities of semiotic and cultural activities mark the left side of the body and the left shoulder as salient members of the semiotic pairs.

CONCLUSION

Biological and semiotic pairs of somatic objects, unified somatic complexes and the salience of a member of the semiotic pair constitute the concepts and formal elements of the metalanguage that describes the human body and its semiotic conceptualization. If a linguist acknowledges that the somatic objects are semiotically paired, then s/he will obligatorily face some problems concerning the lexicographic interpretation of the words denoting these objects. What are the lexical entries of these words in an explanatory dictionary? Must they be in the singular or in the plural form? If the lexeme denoting the semiotic pair has a plural form should the singular form denoting the member of the pair be acknowledged as its morphological and semantic derivative? For example, if there is the dictionary entry *plechi* 'shoulders' should the word *plecho* 'shoulder' be mentioned somehow in the dictionary as well? If the answer to the question is positive, then how should the word *plecho* 'shoulder' be described in a dictionary? We state that if one member of the semiotic pair is salient, then its singular name has to be included into the same entry and its salience must be demonstrated in one way or another.¹²

We have had to leave aside many stimulating and perspective topics associated with the notion of semiotically paired objects. One of the topics concerns the characteristics of the possessors of these objects. We know perfectly well that such features as sex and gender should be taken into account when we deal with male and female semiotically paired objects. The social features of the possessors are significant as well. These features are reflected in the Russian texts. For example, only a man who does occupy (or just imagines that he occupies) a higher position on a social scale can say about another man *On – moja pravaja ruka* 'He is my right hand'.

Another topic is the difference of certain features of somatic objects. The physical and structural features of both members of the semiotic pair are usually the same. For example, the left and right nostril, the left and right ear, the left and right eye are not differentiated either linguistically or semiotically. But there are some rare exceptions, e.g. the right hand of a right-handed person is stronger than the left one. As for the functional features, they are mostly different. The physiological actions, the body signs, the modes and models of the behavior of one somatic object often differ from those of another.

The researches into these topics and problems will provide more detailed and thorough construction of the semiotic conceptualization of the human body.

¹¹ Cf. Shmel'ov 1999.

¹² This assertion has been thoroughly investigated and extensively illustrated in Kreydlin and Letuchiy 2006.

REFERENCES

- Ameka, Felix K., and Breedveld, Anneke. 2005. Areal cultural scripts for social interaction in West African communities. *Intercultural Pragmatics* 1/2. 167-187.
- Arkadyev, Peter M., Kreydlin, Grigory E., Letuchiy, Alexander B. 2008. Sravnitelnyy analiz verbalnykh i neverbalnykh znakovykh kodov (postanovka zadachi i sposobov yeye resheniya) [The comparative analysis of verbal and nonverbal semiotic codes]. *Dinamicheskiye modeli. Slovo. Predlozheniye. Tekst*, ed. by A.V. Bondarko, G.I. Kustova, R.I. Rozina, 439-449. Moscow: Yazyki slavyanskikh kultur.
- Enfield, Nick J. 2009. *Anatomy of meaning: speech, gesture, and composite utterances*. Cambridge: Cambridge University Press.
- Kendon, Adam. 2004. *Gesture: visible action as utterance*. Cambridge: Cambridge University Press.
- Kreydlin, Grigory E. 2009. Semioticheskaya kontseptualizatsiya tela i sopostavitelnyy analiz neverbalnykh sistem (leksikograficheskiy aspekt) [Semiotic conceptualization of body and comparative analysis of nonverbal systems]. *Materialy VIII Mezhdunarodnoy shkoly-seminara «Novoye v teorii i praktike leksikografii: sinkhronnyy i diakhronnyy podkhody»*, 53-54. Ivanovo.
- Kreydlin, Grigory E., Letuchiy, Alexander B. 2006. Kontseptualizatsiya chastey tela v rusском yazyke i v neverbalnykh semioticheskikh kodakh [Conceptualization of body parts in the Russian language and Russian nonverbal semiotic codes]. *Russkiy yazyk v nauchnom osveshchenii* №12 (2). 80-115.
- McNeill, David. 2000. *Language and gesture*. Cambridge: Cambridge University Press.
- Shmel'ov, Alexey D. 1999. Homo spuens: Simvolicheskiye zhesty i ikh otrazheniye v yazyke [Symbolic gestures and their reflection in the language]. *Logicheskiy analiz yazyka. Obraz cheloveka v kulture i yazyke: Sb. statey*, ed. by N.D. Arutyunova, I.B. Levontina, 186-193. Moscow: Indrik.

*Niemi, Jussi, Juha Mulli, Marja Nenonen, Sinikka Niemi,
Alexandre Nikolaev and Esa Penttilä*

Figurativeness of NPs containing lexemes ‘hand’, ‘head’ and ‘eye’

A comparison of frequency of use in five languages

1. INTRODUCTION

The aim of the present study is to compare figurativeness of meaning in the most frequent body-part idiom nouns in the five languages that have been for years under study in our joint project called *Crosslinguistic Studies of Phrasal Idioms* (see, e.g., Niemi, Mulli, Nenonen, Niemi, Nikolaev & Penttilä, 2010 and Niemi, Mulli, Nenonen, Niemi, Nikolaev & Penttilä, to appear). The five languages are: English, German, Swedish, Russian and Finnish.

According to our earlier lexical studies of four of the five target languages, the verb plus complement noun structure appears to allow a high degree of productive variation in its verb and noun (for English, see Penttilä 2006; for German, see Mulli 2008; for Swedish, see Niemi 2007; for Finnish, see Nenonen 2002, 2007; cf. radial categories of Lakoff 1987). This state of affairs most probably pertains also to our fifth language, i.e., Russian and, eventually, to all languages that carry this type of verb phrase. Thus, the common grammatical ground that was found the most suitable in the present context was that provided by the verb plus complement noun, in which the complement was to be a body-part noun (e.g., *E. kiss (some)one’s ass*). Another motivation behind this decision was that we have previously analyzed body-part idioms in these languages separately, and have there extended our research interests beyond the present contrastive analysis, for instance to psycholinguistic processing of phrasal idioms (see, e.g., Nenonen 2002, Nenonen & Niemi 1999, Nenonen, Niemi & Laine 2000, 2002 for Finnish, Niemi 1992, 2004a,b,c, 2007 for Swedish, Niemi & Niemi 2007 for Swedish and English, and Mulli 2007 for German). Moreover, we have also compared cultural versus biological motivation of V+N phrasal idioms across languages (English and Finnish) (Penttilä, Nenonen & Niemi 1998).

2. DETECTING THE MOST FREQUENT BODY-PART IDIOM NOUNS CROSSLINGUISTICALLY

We used the following general and idiom dictionaries to collect the idioms that were to be subsequently subjected to the qualitative and quantitative analyses: English: Collins (1995), Fowler (1982), Makkai et al. (1984), Makkai et al. (1995), Seidl & McMordie (1992), and Wallace (1981); German: Duden (2002); Swedish: Kari (1993b), Karlsson (1982–1987), *Mälände uttryck* (1990), and Romppanen et al. (1997); Russian: Евгеньева А.П. (1999); Finnish: *CD-perussanakirja* (1997); Kari (1993a), Kivimies (1964), Nurmi et al. (1991), and *Nykysuomen sanakirja I–VI* (1973).

A major result of the comparison was that three nouns, viz., ‘hand’, ‘head’ and ‘eye’, cover about one third or two fifths of the instances of body-part idioms in our samples. The relative (percent) shares (and absolute figures) are as follows: English 36.8% (157/427), German 35.5% (312/880), Swedish 33.1% (268/809), Russian 38.9% (306/787), and Finnish 32.3% (229/708) (for details, see Niemi et al., to appear). Since idioms are typically instantiations of non-literal language, it was thought plausible to test whether the idiom-proneness of these three nouns would entail that they are also typically used non-literally, even outside phrasal idioms.

3. OVERALL LITERAL AND NON-LITERAL USE OF IDIOM-PRONE BODY-PART NOUNS ‘HAND’, ‘HEAD’, AND ‘EYE’ IN TWO WRITTEN LANGUAGE GENRES

3.1 Hypothesis

Idiomaticity is highly associated (but not necessarily completely correlated) with non-literal or figurative meaning. For instance, bi-functional phrases like the English *break the ice* should be given a non-literal (figurative) reading (here: ‘relieve social tension’) for them to be idioms. Thus, they are figurative idioms. However, idioms like *take advantage of* or *take a look at* have no conceivable ingredient of figurativity. Thus, we thought that it would be of some theoretical interest to explore into the possible connection between figurative meaning and idiomaticity by analyzing the *textual* frequencies of the three most idiom-prone body-part nouns in the five languages. In addition, as text genres are expected to differ in their use of literal versus non-literal senses of words, we pitted against each other two text-types, viz., fiction and newspaper language. A most plausible hypothesis, in our opinion, was to expect that fiction would be more figurative than newspaper texts. This hypothesis was based on the assumption that in fiction (as the English term implies) one would have more degrees of freedom to use language abstractly and “creatively”, while in newspaper language, which usually narrates about facts, one would be more constrained by the so-called concrete aspects of narrative language.

3.2 Data and Procedures

For the present purposes, we used the open-access web research resources described in Table 1 as the corpus sources for the frequency analyses of the literality of ‘hand’, ‘head’, and ‘eye’ in the five languages.

Language	Source	No. words (mill.)	Web address
1. Fiction			
English	British National Corpus – Domain: Imaginative	19.7	www.natcorp.ox.ac.uk/
German	IDS-Korpus, div-pub – Belletristik des 20. und 21. Jahrhunderts: Diverse Schriftsteller	3.6	ids-mannheim.de/cosmas2/
Swedish	Språkbanken, Bonniersromaner I	5.6	spraakbanken.gu.se
Russian	Russian National Corpus	70.4	ruscorpora.ru/mycorpora-main.html
Finnish	Kielipankki, Otava-a	2.8	csc.fi/kielipankki
2. Newspapers			
English ¹	British National Corpus – Medium: Periodical	27.9	www.natcorp.ox.ac.uk/
German	IDS-Korpus, <i>mm</i> – <i>Mannheimer Morgen</i> : Januar 1995 – Dezember 2008	248.5	ids-mannheim.de/cosmas2/
Swedish	Språkbanken, <i>Göteborgs-Posten</i> 2004	19.4	spraakbanken.gu.se
Russian	Russian National Corpus	113.3	ruscorpora.ru/search-paper.html
Finnish	Kielipankki, <i>Karjalainen</i>	35.7	csc.fi/kielipankki

Table 1. Data sources for literacy – non-literality analyses of ‘hand’, ‘head’, and ‘eye’ in English, German, Swedish, Russian and Finnish.

¹ It must be pointed out that the subsection investigated in the BNC for newspaper texts is not totally comparable with other searched corpora, since in addition to newspaper texts it also contains texts from periodicals. However, in terms of textual genre, there should be no difference between newspapers and periodicals, which both contain journalistic texts. Another problem, due to the internal division of the BNC, is that there is a theoretical chance of the same texts occurring in both subcorpora, i.e. BNC/imaginative and BNC/periodical. Fortunately enough, this theoretical possibility did not materialize itself in the current searches, since the text samples randomly chosen for the analysis were different in each genre.

The fiction and newspaper corpora were exhaustively searched using all inflectional forms of the target nouns. Subsequently, all the target items, with their textual contexts, were randomized. The first 100 instances of these random-order listings of the target noun (plus context) were analyzed for their literality. Thus, the total set of instances in the analysis was 3,000 (5 languages, each with 100 instances per 3 body-parts in 2 text-types, i.e., 600 instances per each language).

3.3 Results

The percentages of literal use of the target nouns in fiction are as follows: English 73.2, German 59.3, Swedish 56.0, Russian 55.5, and Finnish 60.8, and the corresponding data for the newspaper corpora are: 40.6, 32.2, 35.0, 35.2, and 26.1 (see Niemi et al., to appear, for details). Thus, contrary to our hypothesis, the majority of the meanings of the triad of 'hand', 'head' and 'eye' in fiction are literal, while in the newspapers the reverse is the case. Thus, if we take the literality versus non-literality of these three nouns in the five languages as reflections of the two genres in general, we are in the position to claim that these linguistic expressions tend to be *more* literal in fiction than in newspapers.

4. DISCUSSION AND CONCLUSION

Our analysis shows that – somewhat counter-intuitively – fiction contains relatively more instances of literal meanings of the target nouns 'hand', 'head' and 'eye' in the five languages, English, German, Swedish, Russian and Finnish (cf. similar observations on the somewhat lower proportion of metaphoric words in fiction, when compared to news corpora, Steen et al., 2010). This relatively higher degree of literal meanings in the present fiction corpora, compared to newspapers, can be understood through the worlds that the two genres prototypically convey in their texts. In fiction the authors create a world to their readers, and this requires that they describe the concrete, physical reality of this creation for the readers in order to enable them to conceptualize the state of affairs in their minds. After all, the description takes place through language without visual aids or stimuli. Newspapers, on the other hand, describe and explain the world which already exists and which is more or less shared by their writers and their readers. In newspapers we can even see photographs of the reality, and therefore the text itself can concentrate on a more abstract explication rather than on a description of it. Of course it should also be mentioned that news often concern politics, society, business and administration, many aspects of which are often abstract rather than concrete. Thus, this predicts more instances of figurative meanings of the target nouns mentioned above. We call this major principle of ours the Principle of *Shared World of Discourse*. In addition, there is another principle that leads towards the same end-result, viz., to the relative literality of fiction, or, seen from the opposite perspective, to the relatively high use of non-literal expressions in newspapers. The second of our principles is that of *Timeliness*. The old dictum says that there is nothing as old as yesterday's

newspaper, while a book of fiction should find its readers at least for a few years. Thus, cotemporality, or timeliness, while contributing to the increasing degree of the shared world and experiences, also requires that the writers use current expressions. By fulfilling this requirement, the writers of newspaper texts can be relatively sure that their “vivid expressions” will be understood by their readers and the (quite) frequent use of these kinds of expressions can be regarded as motivated.

BIBLIOGRAPHY

- CD-perussanakirja* 1997. Kotimaisten kielten tutkimuskeskuksen julkaisuja 94. Helsinki: Kotus.
- Collins Cobuild dictionary of idioms* 1995. London: HarperCollins.
- Duden* 2002. *Redewendungen: Wörterbuch der deutschen Idiomatik*. Editors: Matthias Wermke, Kathrin Kunkel-Razum & Werner Scholze-Stubenrecht. (Der Duden in 12 Bänden; 11). 2nd ed. Mannheim: Dudenverlag.
- Fowler, William S. 1982. *Dictionary of idioms*. New edition. Walton-on-Thames: Nelson.
- Kari, Erkki 1993a. *Naulan kantaan: nykysuomen idiomisanakirja*. Helsinki: Otava.
- Kari, Erkki 1993b. *Svenska här och nu*. Helsinki: Otava.
- Karlsson, Göran, ed. 1982–1987. *Stora svensk-finska ordboken*. Helsinki: SKS.
- Kivimies, Yrjö 1964. *Näinkin voi sanoa: suomen kielen fraseologiaa*. Helsinki: Tammi
- Lakoff, George 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
- Makkai, Adam, Maxine Tull Boatner & John Edward Gates 1984. *Handbook of commonly used American idioms*. Woodbury, NY: Barron's.
- Makkai, Adam, Maxine Tull Boatner & John Edward Gates 1995. *A dictionary of American idioms*. 3rd ed. Hauppauge, NY: Barron's.
- Målande uttryck: en liten bok med svenska idiom* 1990. Göteborg: Språkdata.
- Mulli, Juha 2007. Lexical observations on German idioms. In Marja Nenonen & Sinikka Niemi, eds., 193–199.
- Mulli, Juha 2008. Zur Variation deutscher Somatismen. In Christopher Hall und Sebastian Seyferth, eds. *Finnisch-deutsche Begegnungen in Sprache, Literatur und Kultur. Ausgewählte Beiträge der Finnischen Germanistentagung 2007*. Berlin: SAXA Verlag, 181–195.
- Nenonen, Marja, ed. 2004. *Papers from the 30th Finnish Conference of Linguistics, Joensuu, May 15–16, 2003*. Studies in Languages, 39. Joensuu: University of Joensuu.
- Nenonen, Marja 2002. *Idiomit ja leksikko: suomen kielen lausekeidiomien, syntaktisia, semanttisia ja morfologisia piirteitä*. University of Joensuu, Publications of the Humanities 29. Joensuu: University of Joensuu.
- Nenonen, Marja 2007. Prototypical idioms: Evidence from Finnish. *SKY Journal of Linguistics* 20: 309–330.
- Nenonen, Marja & Jussi Niemi 1999. Morphological isolates in idioms: Cranberries or real words? *Brain and Language* 68: 158–164.

- Nenonen, Marja & Sinikka Niemi, eds. 2007. *Collocations and idioms 1: Papers from the First Nordic Conference on Syntactic Freezes, Joensuu, May 19–20, 2006*. Studies in Languages, 41. Joensuu: University of Joensuu.
- Nenonen, Marja, Jussi Niemi & Matti Laine 2000. Morphologically isolated words in idioms: A case study with an aphasic. *Journal of Language and Verbal Behaviour* 3: 72–79.
- Nenonen, Marja, Jussi Niemi & Matti Laine 2002. Representation and processing of idioms: Evidence from aphasia. *Journal of Neurolinguistics* 15: 43–58.
- Niemi, Jussi, Juha Mulli, Marja Nenonen, Sinikka Niemi, Alexandre Nikolaev & Esa Penttilä 2010. Body-part idioms across languages: Lexical analyses of VP body-part idioms in English, German, Swedish, Russian and Finnish. In Stefaniya Ptashnyk, Erla Hallsteinsdóttir and Noah Bubenhofer, eds. *Korpora, Web und Datenbanken. Corpora, Web and Databases. Computergestützte Methoden in der modernen Phraseologie und Lexikographie. Computer-based methods in modern phraseology and lexicography*. Baltmannsweiler: Schneider Verlag Hohengehren, 67–76.
- Niemi, Jussi, Juha Mulli, Marja Nenonen, Sinikka Niemi, Alexandre Nikolaev & Esa Penttilä, to appear. Idiomatic proclivity and literality of meaning in body-part nouns: Corpus studies of English, German, Swedish, Russian and Finnish. To appear in *Folia Linguistica*.
- Niemi, Sinikka 1992. Idiomaticity and context effects in word recognition: A study in crosslanguage psycholinguistics. In Jussi Niemi, ed. *Studia linguistica careliana: A festschrift for Kalevi Wiik on his 60th birthday*. Studies in Languages, 26. Joensuu: University of Joensuu.
- Niemi, Sinikka 2004a. Kielitytologia ja kognitiivisia havaintoja ruuminosaidiomeista. In Marja Nenonen, ed. pp. 144–150.
- Niemi, Sinikka 2004b. Ruotsin ja suomen idiomatiikka yhteisen kulttuuritaustan indikaattorina. In Marja Nenonen, ed. 151–165.
- Niemi, Sinikka 2004c. Svenskans kroppsdelidiom ur ett språktypologiskt perspektiv. In Melander, Björn, Melander Marttala, Ulla, Nyström, Catharina, Thelander, Mats & Östman, Carin, eds. *Svenskans beskrivning 26, Förhandlingar vid Tjugosjätte sammankomsten för svenskans beskrivning, Uppsala den 25–26 oktober 2002*. Uppsala: Hallgren & Fallgren, 246–254.
- Niemi, Sinikka 2007. Idiomatic and structurally related VPs in Swedish: A corpus analysis. In Marja Nenonen & Sinikka Niemi, eds. *Collocations and idioms 1: Papers from the First Nordic Conference on Syntactic Freezes, Joensuu, May 19–20, 2006*. Studies in Languages, 41. Joensuu: University of Joensuu, 235–242.
- Niemi, Sinikka & Jussi Niemi 2007. Verb + bare singular constructions in Swedish and English. In Randi Alice Nilsen, Nana Aba Appiah Amfo and Kaja Borthen, eds. *Interpreting utterances: Pragmatics and its interfaces: Essays in honour of Thorstein Fretheim*. Oslo: Novus. 231–246.
- Nurmi, Timo, Ilkka Rekiaro & Päivi Rekiaro 1991. *Suomen kielen sanakirja*. Jyväskylä: Gummerus.
- Nyky-suomen sanakirja I–VI* 1973. Porvoo: WSOY.
- Penttilä, Esa 2006. *It takes an age to do a Chomsky: Idiomaticity and verb phrase constructions in English*. Unpublished doctoral dissertation, University of Joensuu.

- Penttilä, Esa, Marja Nenonen & Jussi Niemi 1998. Cultural and biological bases of idioms: A crosslinguistic study. In Jussi Niemi, Terence Odlin & Janne Heikkinen, eds. *Language contact, variation and change*. Studies in Languages, 32. Joensuu: University of Joensuu, 234–245.
- Romppanen, Birgitta, Ilse Cantell & Mats-Peter Sundström, eds. 1997. *Stora finsk-svenska ordboken*. Porvoo: WSOY.
- Seidl, Jennifer & William McMordie, eds. 1992. *Oxford pocket English idioms*. Oxford: Oxford University Press.
- Словарь русского языка (МАС)* 1999. В 4-х т. / РАН, Ин-т лингвистич. исследований;
Под ред. А. П. Евгеньевой. 4-е изд., стер. М.: Рус. яз.; Полиграфресурсы.
- Steen, Gerard J., Aletta G. Dorst, J. Berenike Herrmann, Anna A. Kaal & Tina Krennmayr 2010. Metaphor in usage. *Cognitive Linguistics* 21: 757-788.
- Wallace, Michael J. 1981 *Dictionary of English idioms*. Glasgow and London: Collins.

3. Second language learning

Catherine Collin

Organising strategies: French and English in contrast¹

1 OUTLINE

1.1 Introduction

Linguistic analysis aims at understanding the functioning of discourse. Confronting languages shows the underlying mechanisms at work in both languages. Infinitive and imperative modes are both used in French and English to express instructions and directives. The aim of the present paper is to show that although imperative and infinitive modes may both be used in directions and orders, the judgment they imply is based on different modal operators, and the organisation they involve is motivated by different factors. Recipes are characterised by a certain number of linguistic features. Our database consists of about 900 translated recipes. But contrasting authentic texts with their translation sheds new light on differences that tend to be overlooked in monolingual treatments. When a choice of translation is seen to recur in different texts, this frequency can be analysed in terms of a tendency that needs to be explained.

1.2 Organisation of the processes in French and in English

Imperative in French and English refers to deontic and injunctive values in the vast majority of cases in recipes. In French, both the infinitive and the imperative modes are used to express injunction. But a quantitative analysis would be of little help considering the risk of reducing the marker to its use without taking into account the specificity of its construction in both languages. Moreover, the reason why the injunctive value is translated by the infinitive or the imperative mode would remain unexplained. Indeed, what is noticeable is the fact that infinitive and imperative are rarely translated as such in French and English, and that the organisation of the processes requires another marker. Nominalisation can be considered as one of those means, as in the following examples.

- (1) Préparer du riz ou des pommes de terre *comme garniture* au coq au vin.
Prepare rice or potatoes *to serve with* Coq au vin. (auf²)
- (2) *Macération* : 2 fois 2 heures.
Allow marinating time of four hours, in two-hour blocks. (sof)

¹ I would like to express my gratitude to the anonymous reviewers, and to Jemma Buck for their advice and valuable comments on an earlier version of this article.

² For the complete references of examples used here see the bibliography. The original language is given first.

- (3) Répéter l'opération jusqu'à l'épuisement des ingrédients.
Repeat, until *you have used up* all ingredients. (auf)

In each case, the verb phrase is translated into French by a nominal phrase (*garniture, maceration, opération, épuisement*). This overwhelming tendency can also be seen in the fact that the nominal phrase is added in the French extract while no corresponding verb phrase is present in English, as in example (3) *Répéter l'opération/ Repeat*. The nominalisation process encountered in recipes enables the action to be described without the presence of a human agent. The action is then understood as an autonomous process in French but remains under the control of a human agent in English ((3) *you have used up all ingredients*). Two types of nominalisation can be encountered. Some of them indicate that an action is in progress ((2) *macération*), while others do not interfere in the chronology of the recipe ((3) *opération*). Nominalisation offers the possibility of referring to another process in the making ((4) *coloration*) that has to be taken into account in the description of the actions. *Coloration* is translated by a dynamic verb phrase in English:

- (4) Découvrir le carré de porc pour lui permettre de prendre une belle *coloration*, 20 minutes avant la fin de la cuisson.
Uncover the rack of pork and *let brown* 20 minutes before the end of cooking. (yop)

Nominalisation in French enables an action to be described in its development (*découvrir*) while another process is presented in its telic aspect (*coloration*). As P. Cappeau (2003: 90) puts it: "La nominalisation permet une sorte de bifurcation pendant qu'un processus est engagé puis laissé en suspens, un autre procès peut être présenté".

The presence of a personal pronoun in (3) *you have used up* is not compulsory in the English verbal phrase, but it anchors the direct resort to an agent to perform the action, as in (5):

- (5) Ajoutez la margarine et travaillez avec vos doigts jusqu'à *obtention* d'une préparation granuleuse.
Add margarine and knead together with your fingers until *you have* a crumbly mixture.
(auf)

Contrary to the preceding examples, the French imperative (5) *ajoutez* is used, and the *-ez* ending testifies for the address to a second person marker. In this example, the French imperatives *ajoutez* and *travaillez* act as a frame ruling the rest of the sentence, so that the nominalisation *obtention* is connected with the imperative verbs. In English, the sentence *you have a crumbly texture* is in the indicative mode, and its realisation depends on the co-speaker *you*. The presence of the personal pronoun implies the construction of a speaker. The representation of the intersubjective construction is different in French since the *-ez* verbal ending indicates that the speaker commits himself/herself to the value of the prospective validation of the predicative relation. The nominalisation of the process is thus a means of avoiding subordination, while another process presented is in the making. Nominalisation leads to erase any agentivity of the action itself. Each time the process is translated by a verb phrase in English, even though the personal pronoun is not necessarily mentioned as in the following example (6):

- (6) Couvrez et laissez à feu doux jusqu'à *cuisson* des légumes et *évaporation* de l'eau.
Cover and leave to simmer at a low heat until vegetables *are cooked* and liquid *has reduced*. (auf)

The use of the passive voice enables the agent to be avoided, and to hierarchise the processes. What can be deduced from the analysis of the preceding examples is the fact that the English translation tends to concentrate on the realisation of the action (hence the presence of a verbal phrase) where the order of the actions reflects the order of the 'events'. In other words each verb corresponds to a gradual step in the process, so that the succession of the actions depends on the verbs. The French text very often breaks the chronology of the events, as in (7):

- (7) Plongez-les dans une marmite d'eau salée dans laquelle vous aurez ajouté la farine *délayée* dans un peu d'eau.
Put a big pan of salt water on to boil. *Sieve* flour into a small amount of water, and add to pot. (pti)
- (8) Faire revenir l'oignon dans un peu de beurre ajouter les tomates sans la peau *coupées en dés*.
Brown the onion in butter, *dice* and peel the tomatoes and add to pan. (auf)
- (9) *Peel and slice* the apples and arrange on the pastry.
Disposez sur la pate les pommes *épluchées et coupées* en fines lamelles. (sof)

The chronological order of the successive actions is disrupted in French with the use of the past participles (*délayée, coupées en dés, épluchées et coupées*). They refer to prior processes, which have not even been mentioned in the preceding context of the extract; nonetheless they are considered as being accomplished. The past participle corresponds to a time omission of the process which is only referred to in its resultative or telic step. This trend can also be seen with the following examples:

- (10) *Avec la main*, former des boulettes de la taille d'une noix avec le mélange au poulet.
Using damp hands, form mixture into evenly shaped balls that are small enough to eat with your fingers. (eur)
- (11) Testez la cuisson à l'aide de la pointe d'un couteau qui doit ressortir propre.
Bake until sides pull away from the pan and skewer *inserted in middle comes out clean*. (eur)
- (12) Lorsque le sabayon fait un ruban, retirer du feu et poursuivre *au fouet* jusqu'à refroidissement.
Whip the mousse while continuing to beat with a whisk until thickened. Remove from heat and continue *whisking* until cool. (ces)

In French the mention of the tool (*pointe d'un couteau, fouet*) equates with the action itself while in English a verbal participle is needed (*inserted, using damp hands*). In (12), the prepositional phrase *au fouet* refers directly to the instrument, and metonymically to the action. Such a reference to the

noun is absent in English, leaving room for the movement of the action, which is even repeated (*beat with a whisk, continue whisking*).

- (13) Verser l'eau sucrée, bien mélanger *après avoir rebouché* la bouteille.
Add the "syrupy" water. *Carefully close the bottle with its lid*, and mix well. (auf)
- (14) Cover with cling film and put in fridge.
Mettre dans le réfrigérateur *après avoir recouvert* le bol du mixeur d'un film alimentaire.
(auf)
- (15) Pelez l'ananas. Supprimez les "yeux" incrustés dans la chair. Taillez le fruit en rondelles de 5 mm d'épaisseur. Recoupez-les en petits morceaux *après avoir ôté* la partie centrale.
Peel the pineapple and remove the eyes. Slice into rounds approximately 1/2 cm thick and then cut into quarters, *removing* the core. (auf)
- (16) *Après avoir fait tremper* les brochettes de bambou dans l'eau pendant 15 à 20 minutes, enfilez les cubes d'agneau et badigeonnez la viande du mélange huile aux herbes de Provence et romarin.
After dipping the bamboo brochettes in water for 15 to 20 minutes, spear the lamb cubes and baste the meat with the herbes de Provence and rosemary oil mix. (ces)

In (16), the French past participle is translated by a present participle (*dipping*). The structure *after -ing* is rare in our corpus and only 18 occurrences can be found. More generally, the presence of the *-ing* participle of the most frequent verb in our corpus is very low.

Table 1: Five frequent English verbs in the bilingual corpus in the *-ing* participle form.

Verb	Number of imperative form	Number of <i>-ing</i> participle	% of <i>-ing</i> participle
Add	953	59	6.19
Pour	302	8	2.64
Mix	335	107	31.94
Remove	317	17	5.36
Put	261	6	2.29

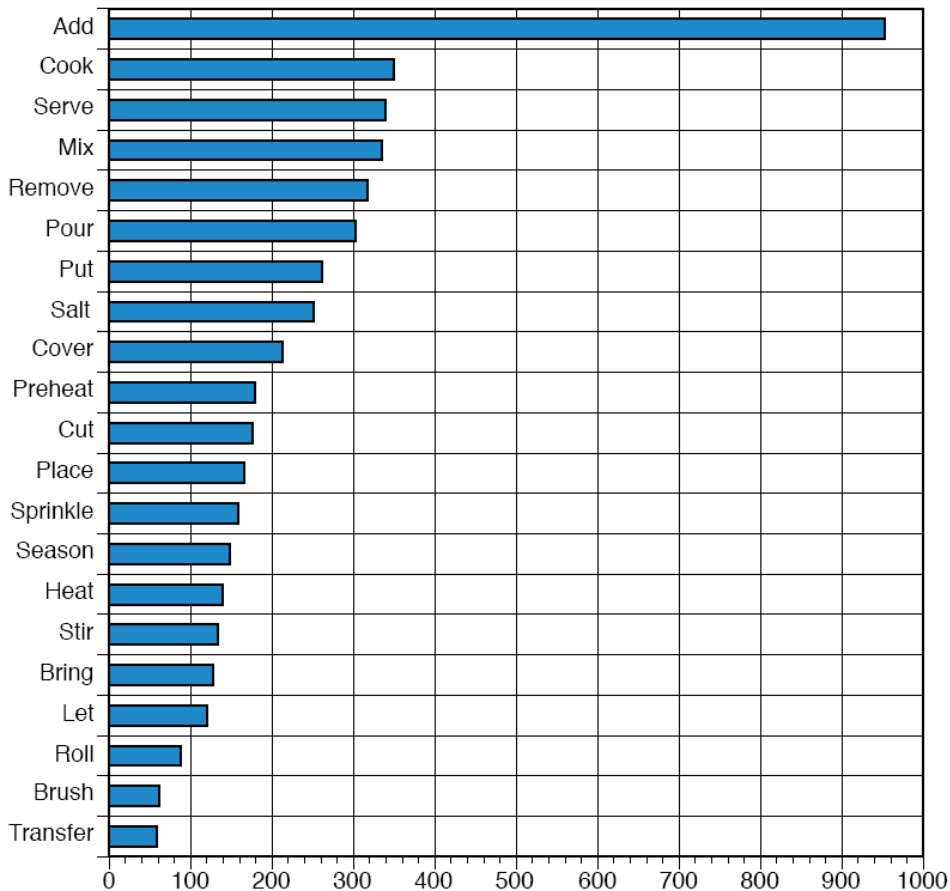


Figure 1: The most frequent English verbs in the bilingual corpus.

The use of participles is different in the two languages, since the French past participle is an historical means to build injunction in recipes. Indeed, a closer analysis of old French texts³ shows that the past participle is very often used in recipe instructions.

A. Englebert (2009: 141) points out that this structure may represent the only verbal form of some recipes: “Le participe apparaît dans les trois mêmes textes que les premiers infinitifs injonctions, le *Manuscrit de Sion*, le *Viandier de la Mazarine* et le *Mesnagier* – trois textes où il peut être la seule forme verbale de la totalité de certaines recettes⁴.” In the following extract all the different instructions are presented in the past participle tense.

³ *Manuscrit de Sion* (1320-1325); le *Viandier* (c.1350); *Mesnagier* (1393).

⁴ “The participle form appears in the same texts than the first infinitive and injunctive forms, the *Manuscrit de Sion*, the *Viandier de la Mazarine*, and the *Mesnagier* – in those three texts it can be the unique verbal form in the whole recipe”.

- (17) Ales *rosties* en fillopant, *mengees* a la moustarde; ou *pelees*, puis *cuictes* en l'eaue ung trespetit, puis *enfarinees*, frictes a l'uile, et *mengees* a la jance ou aux ailletz. (*Manuscrit*, rubrique 73, A. Englebert)

Even though the modern use of the past participle has changed in French, up to the point that no such instructions could now be entirely given in this verbal form, its injunctive meaning in context might have kept at least part of its heritage.

Considering the series of examples we have seen, the English version turns out to be respectful of the chronology of the recipe and delineates each step of the process. In the case of French, validation is anticipated (*mélanger après avoir rebouché, après avoir recouvert*), and a hierarchical distinction is established between the actions, so that the use of the participle or the nominalisation is a means of referring to actions without affecting the course of the main process. To sum up, the participle is used in French texts to underline an action that refers to a prior process whose result is the most distinctive quality.

2 LOCALISATION AND DETERMINATION

The question of localisation and determination is a key issue because it provides arguments in favour of discriminating features of recipes in French and in English. It also supports the view held in this paper that in French, the organisation of the recipe promotes the intersubjective orientation and the English texts are organised relative to the actions and processes. Of particular interest to the present subpart of the study is the analysis of the link between the processes, and the different constructions in both languages of a viewpoint.

2.1 Syntactic arrangement of the processes

The difference in terms of arrangement lies in the fact that the English recipe rests upon the implicit succession of the actions in such a way that each imperative verb describes a distinct step of the activity to be carried out. This prominent trait can be particularly illustrated in the following examples:

- (18) Mettez tout cela dans de l'eau à laquelle vous avez ajouté une partie du jus de limon.
Add lemon juice to a pan of water, *then* soak artichoke. (ces)
- (19) Dressez les filets de dorade rotis sur lesquels vous verserez le caramel au balsamique, accompagnez-les de crudités assaisonnées.
Drizzle fish with sauce *and* serve with vegetables or crudits. (cui)
- (20) Incorporez peu à peu la crème parfumée, après avoir retiré la gousse de vanille fendue.
Add the cream little by little, *but* remove the vanilla pod. (yop)

In the English version, the segmentation of the processes into separable units tends to divide the actions to be performed. This division is very often associated with the use of coordinators *then*, *and*, *but* that mark and define explicitly the link between the different steps of the action. In (20), the chronology of the action is reversed because the action of removing the vanilla pod is prior

to the introduction of the cream in the preparation, and the English text marks this disruption in the succession with the *but* coordinator. The relation between the actions is explained, as the following example shows:

- (21) Retirez du four, faites glisser la pizza sur une planche à découper ou une grande assiette de service. Attendez 5 à 6 minutes avant de découper.

Take the pizza out of the oven and transfer it to a cutting board or your plate. *In order to allow the cheese to set a little*, wait 3-5 minutes before slicing or serving. (pti)

In the English version of (21), the constituent *In order to allow the cheese to set a little* accounts for the process *wait*. In French, such justification is absent, and the different actions remain autonomous. In this extract, another trend made explicit is the fact that the propositional content of the English text is always integrated to the main clause and to the verb, and even time references are set relative to the preparation itself, while in French, the reference is satellised in anteposition or postposition relative to the sentence:

- (22) 5 minutes avant la fin de la cuisson, versez le jus d'1 citron vert et mélangez.

Mix well, allow to *simmer* for a final 5 minutes. (sof)

- (23) Dès l'obtention d'une purée lisse, ajouter, en mixant toujours, le vinaigre, le Tabasco, le sel et le poivre.

Mix with crushed garlic, liquidise for a couple of minutes *until you have a smooth puree*. (auf)

In English, the time reference *final 5 minutes* or *until you...* is established relative to a verb form (*simmer, have*) that belongs to the process, so that a thematic coherence can be followed within the recipe. The reference is salient in the French extracts and the anteposition of the constituent illustrates the fact that it does not pertain to the same localisation process as the rest of the recipe.

The crucial question of the reference is also brought up by the following extracts:

- (24) Plusieurs heures plus tard faites une chantilly avec la crème, le sucre et la vanille à l'aide d'un batteur et déposez-la sur la tarte ; déposez le tout à nouveau dans le congélateur.

10 minutes before serving, make the topping. Whisk the cream, sugar and vanilla and put on top of flan filling. Leave to chill for 10 minutes in the fridge. (eur)

- (25) On évitera de faire cuire les lentilles trop longtemps ce qui transformerait la légumineuse en purée.

Do not cook *the lentils* too long or *they* will turn into a puree. (auf)

The first remarkable point is the mention of the time reference *plusieurs heures plus tard* in French which involves an external existence of the event outside the situation of utterance. On the contrary in English, *10 minutes before serving*, the localisation of the reference is relative to the serving, i.e. the action in progress. In French, the *-la* (feminine) determination (in *déposez-la*) refers back to the *Chantilly* (feminine, singular), but once the topping is put on the flan, its

reference is modified (*déposez le tout*). In the second example, the plural reference *les lentilles*, is changed to the singular *la légumineuse*. This specific use of the pronoun corresponds more directly to the textual and structural organisation of the context than to the ontological properties of the referent. This distinction has been first mentioned by G. Kleiber (1994), and defined as follows in Kleiber (1997: 19): “Le sens du pronom comporte en fait deux parties: une composante vériconditionnelle formée par les propriétés lexicales ou descriptives intrinsèques (...) et une composante instructionnelle ou procédurale qui regroupe les indications référentielles sur la façon spécifique de retrouver le référent⁵”.

2.2 Construction of a viewpoint

The recipe in French seems to obey a global and structural localisation of the reference, which explains the presence of different levels of organisation, and the creation of distinct viewpoints.

(26) Coupez les 3 citrons confits restants en quartiers (*et retirez les pépins*).

Quarter and *unseed* the 3 remaining preserved lemons. (auf)

(27) Puis, avec précaution et en vous protégeant les mains (*attention aux éclaboussures*), versez les 25cl d'eau.

For safety reasons, *have ready* a bowl of ice water to plunge your hands into *if any caramel should land* on your skin. (auf)

In both examples, the French version introduces parentheses, which traditionally correspond to secondary processes or even afterthoughts that could be paraphrased by *don't forget to unseed* in the first example. In the second example, the content of the parenthesis *attention aux éclaboussures* contrasts with the rest of the text because of the exclamative-like tone of such a unit. The injunctive meaning of the noun *attention* is partly translated by the imperative *have ready* and the construction of a hypothetical situation *if any*. In any case, the French extract represents a change of point of view, while the English one has a unique viewpoint on the situation. The longer extract (28) is also illustrative of such a difference between both languages:

(28) *Disposez un grand morceau de papier sulfurisé dans le moule à tarte. Je vous conseille d'étaler la pâte de suite sur un plan fariné puis de l'abaisser dans le moule à tarte. Et de la mettre au frais pendant une ou deux heures. Ainsi, elle ne se rétractera pas à la cuisson.*

Line a pie plate with parchment paper. *Allow* the dough to warm slightly to room temperature and *roll* the dough to a thickness of 1/8 inch, on a floured board. Cut a circle about 2 inches larger than the plate by rolling it onto the rolling pin. *Refrigerate* for 1 to 2 hours. (pti)

⁵ The meaning of a pronoun is divided into two constitutive parts: vericonditional component made of lexical and descriptive properties, and instruction or processing component including the specific way to trace back to the source of the reference.

In this example, the point of view changes in the French extract with the mention of the reference to the speaker⁶ Je who commits himself/herself as to the necessity of the validation of the process, and signals the speaker's stance as regards the value of the utterance which is vouched for.

3 VALIDATION AND PROSPECTIVE VALIDATION

In French, the assertive modality is able to refer to a prospective event in the same way as when injunction is at stake. In English the use of the hypothesis is the invariable answer to this trend.

3.1 Assertive modality versus hypothesis

The assertive modality is another means used in French instructions. Assertion is combined with the use of the indicative present or future in the following examples:

- (29) Ce poulet est tout autant un délice *lorsque l'on remplace* le curry par du safran.
This dish is even better *if you replace* the curry powder with saffron, for a more delicate taste. (auf)
- (30) When mixture is cold, make ice-cream according to ice-cream maker instructions. If you don't own an ice-cream maker, *make* this mascarpone ice-cream and *add* orgeat syrup.
Si vous n'avez pas de sorbetière, *je vous conseille* ma glace au mascarpone sans sorbetière, à laquelle *vous rajouterez* le sirop d'orgeat. (pti)
- (31) Selon le contenu de vos placards, *vous pourrez* bien entendu les remplacer par d'autres farines ou flocons de céréales, avec ou sans gluten. *Pour les personnes sensibles* aux flocons d'avoine, *je conseillerais* de prendre des flocons de millet.
If you prefer, choose other flours or rolled grains, with or without gluten. *If you are sensitive* to oats, prefer rolled millet. (pti)
- (32) *Vous pouvez* également verser sur la tarte chaude un petit verre de rhum et faire flamber.
Try pouring a second shot of rum over the pie while hot, then flambing at the table. (auf)

The indicative marks a break in the recipe, splitting the text so that a comment can be inserted introducing a change in the endorsement of validation. The indicative construction locates the propositional content of the sentence in an external moment relative to the situation of the instruction. This construction can represent a distant moment relative to the recipe where the speaker assumes an extended position or remark on the validation as in *lorsque l'on remplace, je vous conseille, vous pouvez*. But it can also represent an iterative moment each time the

⁶The presence of the 1st person imitates the traditional way French Chefs write or tell their recipes, launched by Raymond Oliver. This corresponds to another injunctive means as argued by A. Englebert (2009: 125) "Par cette 1^è personne, on pense moins à une injonction qu'à l'exposé par le chef de ce qu'il fait ; mais cette 1^è personne peut également constituer une invitation ("faites ce que je fais"), et par là une injonction." (The 1st person can be used in reference to the Chef in his presentation of the recipe, or invitation to "to the way I do", and thus constitutes an injunctive form).

prescriptive actions are performed, as in *vous rajouterez, vous pourrez*. In the latter case, the future tense appears. The future tense basically predicates the existence of an event, and the speaker in French texts has recourse to this tense to comment on events, including conjecture on events, by fictitiously displacing them⁷. In English, the imperative⁸ form (*add, choose*) is the expected form in such a context. But a striking element is the presence of *If* in each extract to introduce this external situation. In English, any comment that does not belong to the proper situation of the recipe needs to be marked by distinctive features, so as to delineate with separate markers the different levels of enunciation. Consequently, the instruction cannot be attributed to a precise origin of commitment in English. The processes *if you replace, if you prefer, if you are sensitive*, are orientated relative to the situation of utterance.

3.2 Prospective validation and the construction of intersubjective relations

The use of the periphrastic forms in French in injunctions creates a modal distance between the situation of utterance and the source of the injunction. The English imperative form corresponds to 86.6% of translations of the periphrastic forms.

(33) Il faut laisser reposer la viande en veillant à ne pas les laisser baigner dans son jus de cuisson.

Let meat rest, making sure not to let them soak in their cooking juices. (auf)

(34) Il suffit de râper les betteraves, courgettes, céleri, carottes... en fines lamelles et les déposer sur une plaque dans le four, chauffé à 90°C (T 2/3).

Try making 'vegetable crisps' for a fun starter or snack: Thinly slice some beetroot, courgette, celery, carrots or any other vegetables you fancy, then arrange on an oven tray. (auf)

The periphrastic *il faut/il suffit* illustrates moments in the recipe where the prospective validation is oriented relative to the situation of utterance. In other words, the periphrastic form marks injunctive forms whose origin is different from the speaker's, introducing athetic predication, and posits the existence of a particular state of affairs which characterises the situation. In English, the distance between the injunctive form and the origin is also created with the use of *try making, let meat rest*. The speaker's stance is clearly enhanced, and the introduction of a modal use of *try* or *let* allows the content of the propositional clause to be distinguished from the imperative form. The content is modalised by the periphrastic forms in French. In such a context, the corpus data show that the conditions in which modalisation occurs is different in both languages. In English, modalisation concerns the speaker's stance towards the validation of the situation. In the same contextual environment in French, modalisation takes place through the use of a third person subject in assertive statements. This difference between French and English in terms of modalisation greatly influences the way the two languages organise the construction of intersubjective relations. In English, the intersubjective relation is created within the situation

⁷For further analysis of the future tense, see A. Celle (2007).

⁸The translation of a French future tense by an imperative form in English corresponds to the majority of cases (53.6%), the translation by a modal *will/would* (23.6%), aspectual change (10.7%) and past participle (7.1%) in our corpus.

of utterance, which accounts for the massive presence of verb phrases and imperative forms to give instructions. In the same contextual conditions in French, the intersubjective relation frames the instructions which can be illustrated by the comments and interpretations breaking the chronology of the process. Likewise, the choice of the infinitive in French recipes is definitely not bound to erase any commitment. Contrary to the assumption that the imperative mode is supplanted by infinitive forms in French recipes⁹ the two modes share specific functions in the recipes. A closer analysis of our corpus¹⁰ shows that the infinitive mode marks specific rather than temporal disconnection:

(35) *Retirez* la tarte du four et *laisser* reposer.

Remove the dough from the oven and let cool. (pti)

In this example, the presence of both the infinitive and the imperative mode illustrates the fact that the intersubjective relation serves as a general frame for the instructions, and within this structure, there is still room for other syntactic forms with a specific purpose that gives its orientation relative to the process. Furthermore, a lexical analysis of the most frequent verbs in our corpus shows that each verb tends to specialise in a particular mode. For instance, *laver* (to wash) is massively used in its imperative form, while *couvrir* (to cover) is used in the infinitive form in the majority of cases. The following figures give details of this analysis.

⁹ See A. Englebert (2009: 126) "L'infinitif se présente comme le moyen d'expression de l'injonction le plus remarquable des traités actuels". (The infinitive is the most important injunctive means in the contemporary instruction recipe texts).

¹⁰ Other illustrations of this phenomenon in P. Cappeau (2003) and C. Collin (2006).

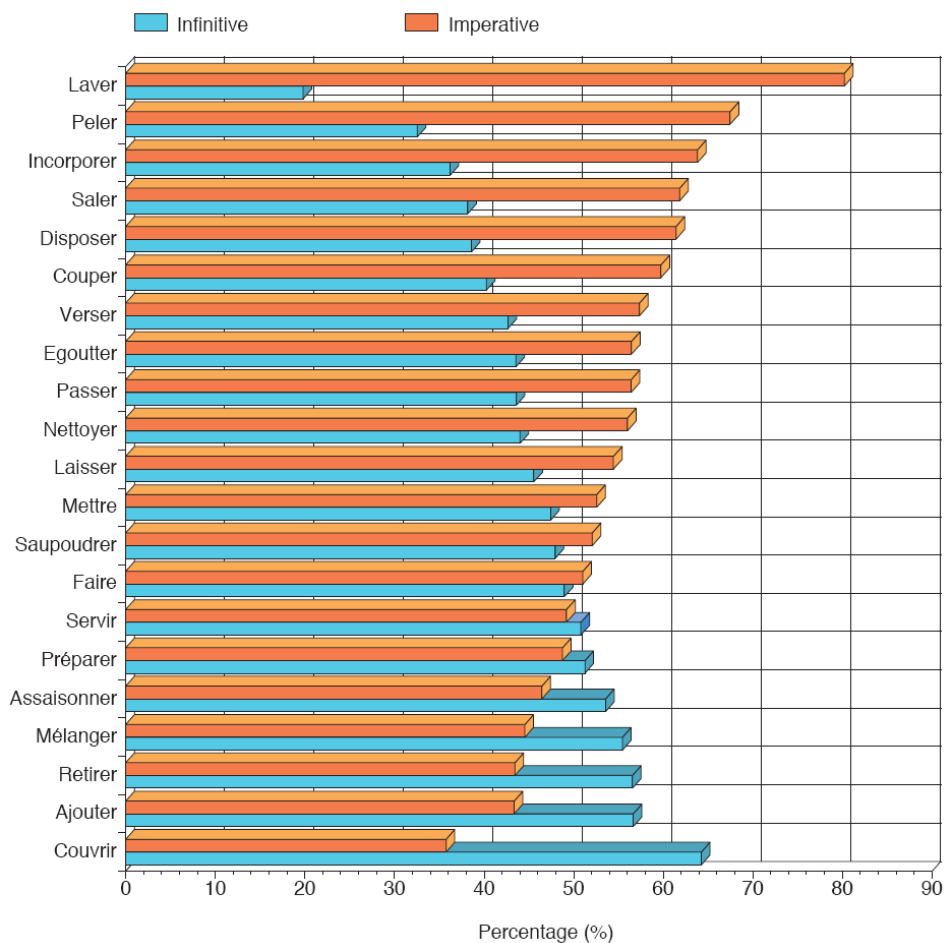


Figure 2: Infinitive versus imperative use in the bilingual corpus concerning the most frequent French verbs

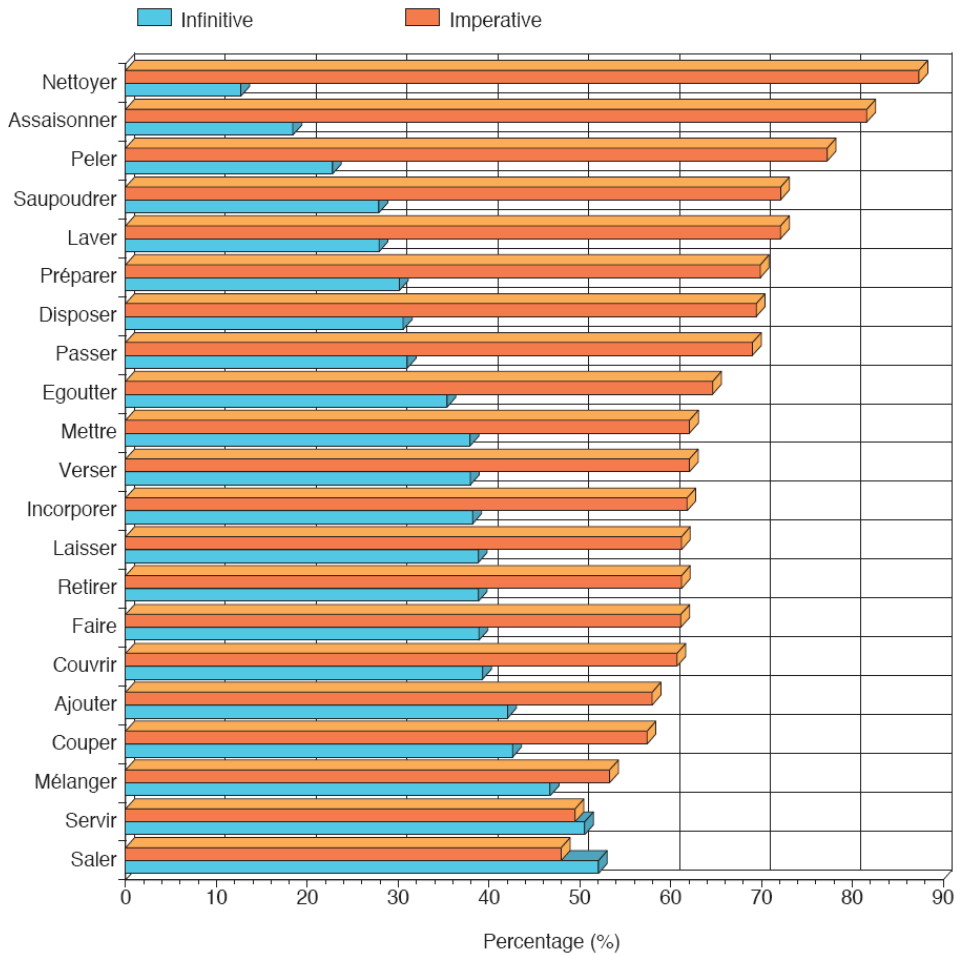


Figure 3: Infinitive versus imperative use in the monolingual corpus concerning the most frequent French verbs

4 CONCLUSION

Though this study should be extended to other instruction types of texts, several remarks and working hypotheses can be made about the organising strategies in both languages. The investigation of authentic occurrences drawn from bilingual instruction environments shows that French and English behave differently as far as instruction is concerned. The purpose of this paper was to stress that regular choices in terms of translation could be found, providing evidence of a tendency. In English, the propositional content of the injunctive form cannot be attributed to a precise origin of commitment, so that the syntactic arrangement of the predicates generally follows the succession of represented instructions. The representation of intersubjective relations are created within the situation of utterance. Crucially, modalisation is

carried out in an independent way in French texts where the instruction construction is built on an intersubjective framework localising the description of the processes. Different injunctive means are used which mark transitions and trigger subjective assessment in the choice of a specific form.

BIBLIOGRAPHY

- Cappeau, Paul. 2003. Vous mélangez tout ? De l'ordre dans les recettes. *Ordre et distinction dans la langue et le discours*, ed. by Bernard Combettes, Catherine Schnedecker, and Anne Theissen, 87-99. Paris: H. Champion.
- Celle, Agnès. 2007. Temps et modalité l'anglais, le français et l'allemand en contraste. *Études contrastives*. Volume 7. Bern: P. Lang.
- Collin, Catherine. 2006. *Le sujet dans l'injonction*. Rennes: Presses Universitaires de Rennes.
- Englebert, Annick. 2009. *Mange! L'impératif français, du mythe à la réalité*. Bruxelles: Editions de l'université de Bruxelles.
- Kleiber, Georges. 1994. *Anaphores et pronoms*. Louvain-la-Neuve: Duculot.
- Kleiber, Georges. 1997. Anaphore pronominale et référents évolutifs ou Comment faire recette avec un pronom. *Relations anaphoriques et (in)coherences*, ed. by Walter De Mulder, Liliane Tasmowski-De Ryck and Carl Vetters, 1-29. Amsterdam: Rodopi.
- Kleiber, Georges. 2001. *L'anaphore associative*. Paris: Presses Universitaires de France.
- Siewierska, Anna. 2004. *Person*, Cambridge: Cambridge University Press.
- Van der Wurff, W. (ed.) 2007. *Imperative Clauses in Generative Grammar, Studies in Honour of Frits Beukema*. Amsterdam: John Benjamins.
- Xrakovskij, Victor. 2001. *Typology of Imperative Constructions*. Munich: Lincom Europa Studies in Theoretical Linguistics.

CORPUS

Monolingual corpus

<http://www.bbc.co.uk>

<http://www.cuisine-de-saison.ch/fr>

<http://www.cuisinez.free.fr>

Bilingual corpus

<http://www.aufeminin.com> (auf)

<http://www.sofeminine.co.uk> (sof)

<http://www.cestmamanquilafait.com> (ces)

<http://www.europesbest.ca/fr> (eur)

<http://www.ptitchef.com> (pti)

<http://www.yoplait.ca> (yop)

<http://www.cuisine-france.com> (cui)

Lena Dal Pozzo

Second language acquisition: issues from Italian and Finnish at the syntax-discourse interface

1 INTRODUCTION

The study concerns the interlanguage grammar of adult second language (L2) speakers of Italian with Finnish as their first language (L1) at the subtle syntax-discourse interface. It is part of a larger research project started by Belletti and Leonini (2004) and successively Belletti et al. (2007) with respect to the *answering strategies* (Belletti 2006) adopted by non-native speakers of Italian in discourse-pragmatic contexts in which the subject is new information (henceforth SNI). The aim of the present study is to provide further data and observations in the domain of language acquisition, in particular of Italian L2 in two populations which differ for the language attainment in their L2. The differences between the L1 and the target language of the participants, (namely between Finnish and Italian), such as the properties of the subject and the different answering strategies adopted in the two languages in SNI contexts, make the study innovative. Since VS inversion structures in Italian with transitive verbs involve the use of clitic pronouns, which are absent in the L1 of the participants, the use of clitic pronouns in the L2 interlanguage was also observed. Data was collected in two countries through an elicitation task in which the contextual situations were presented in a video test: 22 short videos were used to elicit the relevant structures in SNI contexts. It is traditionally assumed within the Principle and Parameters Theory that the availability of postverbal subjects is a consequence of the positive setting of the null subject parameter (Rizzi 1982, Burzio 1986 and related works). Interestingly, rather recent studies on language acquisition (Belletti and Bennati 2004, Belletti et al. 2007) have shown that although subject inversion correlates to the availability of a referential *pro*, typical of null subject languages (NSL), it is not sufficient to instantiate VS inversion, where the subject is new information. Crucially, the availability of this kind of subject inversion structures is also related to some discourse-pragmatic factors. Thus, in looking at second language acquisition, the availability of subject inversion structures in the interlanguage grammar of L2 learners is not straightforward due to the complex interplay between syntax and discourse.

At the syntactic level the results are interesting for the status of Finnish as a partial null subject language (PNSL) (cf. Holmberg et al. 2009). The results show that despite the very high L2 competence the VS order is not used to the same extent as in the native speaker's control

group. Finally, the results provide further support to the assumption that an interlanguage grammar allows for a wider range of grammatical options than the L1.

From a theoretical point of view, the research is based on relevant recent literature within the cartographic approach (cf. Cinque 2002, Rizzi 2004, Belletti 2004) and assumes, in particular, the analyses proposed by Rizzi (1997 and related works) for a fine-grained left-periphery (CP) and those assumed by Belletti (2001, 2004, 2005) for the vP-periphery, which parallels the clause external CP periphery in sentences with new information focus.

2 BACKGROUND

As hinted above, the collected data is discussed in light of some of the main contributions provided under the cartographic approach, which will be briefly introduced in this section. Rizzi (1997, 2004) has proposed the structure in (1) for the complementizer system mainly based on the interaction of different elements in the left periphery of Italian. The left periphery of the clause is analyzed as an articulated area composed by distinct functional heads and their corresponding projections. Belletti (2001, 2004) postulates an articulated vP-periphery as represented in (2), along the lines of the rich left-periphery postulated by Rizzi (1997, 2004).

(1) [ForceP ... [TopP ... [IntP ... [TopP ... [FocP ... [TopP ... [FinP ... [IP ...]]]]]]]]]

(2) [CP ... [TP [TopP ... [FocP **Foc** [TopP vP]]]]]]

The left-periphery proposed by Rizzi is formed by a fixed and a non fixed component. The first one is constituted by the Force and Fin heads located at the top and at the bottom of the CP system, respectively. ForceP is the part of the clause where the so called Specification of Force (Rizzi 1997, following Chomsky 1995) is expressed. In other words, Force is where the information about the sentence type (declarative, interrogative, exclamative, relative, and comparative) is given. The Fin head provides the information about the finiteness of the following IP. The second component is formed by those categories that appear between ForceP and FinP and which are activated when necessary, namely whenever a constituent has topic, focus or WH features are to be satisfied by a Spec-head criterion (Rizzi 1997:288).

One of the baselines of the cartographic approach is that pragmatic-discursive interpretations like those of topic and focus are univocally associated with dedicated functional heads. Hence, according to this framework, the interpretation of a constituent as a new information focus is a consequence of its filling the Spec position of a Focus head. Observing the syntax of VS structures with S as new information focus in Italian, Belletti (2004) shows that postverbal subject occurs very low in the linear order of the clause. Thus, in order to account for the new information focus interpretation, the author proposes the existence of a vP periphery along the lines of (1), which contains one focus position surrounded by recursive topic positions, as shown in (2).

Assuming the vP periphery in (2), Belletti (2001, 2004, 2005) proposes that in Italian a sentence with SNI, as in (3), has the structure in (4). The subject is in Spec,FocP in the vP

periphery, the verb moves to a head higher than FocP and a *pro* satisfies the EPP requirement in the canonical preverbal subject position.

- (3) (Chi ha parlato?) Ha parlato Gianni
 (who has spoken?) has spoken John
- (4) [CP ... [TP *pro*... ha parlato ... [Top [FocP Gianni [TopP [VP...]]]]]]

The movement of the postverbal (new information) subject to the Focus position in the vP periphery is assumed to be related to the availability of a referential *pro* (Belletti 2005), which however has been shown to be a necessary but not sufficient condition for licensing VS order because of the discourse-related constraints which play a crucial role in the distribution of postverbal subjects (Belletti, Bennati & Sorace 2007).

Consequently, the type of *free inversion* observed in Italian is assumed not to take place in non null subject languages. As discussed in (Belletti 2009), NNSL typically focalize the subject as new information focus mainly through two different strategies: (i) SV structures with a particular intonation on the subject, namely *in situ* focalization, as is the case of English, (5), and (ii) cleft strategies, as is the case of French, (6)

- (5) a. Who came?
 b. John came
- (6) a. Qui a parlé?
 who has spoken
 b. C'est Jean.
 ce is Jean
 'It's Jean'

2.1 Previous studies on the acquisition of postverbal subjects

The pilot study conducted by Belletti and Leonini (2004) investigated the use of null subjects and of postverbal subjects in free inversion structures (cf. example (4)) in the L2 Italian of 26 speakers with different L1s (mainly NNSL). The results showed that null subjects were correctly used to a wider extent than postverbal new information subjects, meaning that a referential *pro* was licensed in the interlanguage grammar of the L2 learners. However, the vP peripheral focus position, which is assumed to host the new information subject in null subject languages was not (extensively) activated. Thus, the authors suggest that the two investigated phenomena are not correlated in the interlanguage grammar of the participants and there is a difficulty for the interlanguage grammar at the interface between the computational system and discourse. A later study by Belletti et al. (2007) investigated the same phenomena in the interlanguage grammar of near native speakers of Italian. This strengthened the previous results and showed that the unbalanced correlation between the use of postverbal new information subjects and the use of null subjects cannot be considered a developmental effect, since even at a near-native level null subjects are correctly used at a significantly higher rate with respect to postverbal subjects.

Finally, let us mention the adaptations of the video task to Finnish (Dal Pozzo 2010) and to Brazilian Portuguese (Gessner 2007) which have provided interesting data for further discussion on the syntax of subject in those two languages.

3 THE EXPERIMENTAL DESIGN

Two groups of participants took part in the test. The first one (Group 1) consists of 15 native Finnish speakers who speak Italian as their L2. Testing took place in Italy and at that time, they had all lived in Italy from a minimum of 9 months to a maximum of 40 years (mean 20, 5). All the participants had moved to Italy as adults and have been exposed to Italian in a linguistically natural environment (implicit learning). They have reached a very high level of attainment in the L2. Also the subject with the shortest stay (9 months) has a good competence in Italian, due to her previous stays in Italy and studies of Italian and other romance languages. The participants used the target language at home and in daily life at work or for studying. The mean age is 43,5, with a range between 22 and 65. All the subjects have an academic education and 11 of them have attended Italian language classes mainly during the first period of their stay in Italy, 4 of them have not attended any Italian classes.

The second group (Group 2) consists of 10 adult speakers of Italian L2 with an intermediate L2 competence. Testing took place in Finland and at that time all the subjects lived in Finland and were students in the intermediate class at the Turun Työväenopisto (Centre for Workers' Education of Turku). The participants had received only formal instruction of Italian in classroom environment (explicit learning) and the time of their exposure to Italian range between 1 and 10 years (mean 4,35), which in some cases was not continuous. The mean age is 55,2 with a range between 25 and 66.

Each participant was tested separately and the answers were recorded. The task took about 20-30 minutes per participant but it was untimed for the purpose of the study. The same collection methods were adopted for the two groups. Relevant information on Group 1 and Group 2 is provided in Appendix 1.

The participants were shown 22 short videos. After each video a question concerning what they had just seen was asked by one of the characters, followed by 1-3 questions in the presentation. The tested subject was previously instructed to answer in the most spontaneous way and to use a verb in the answer. Filler questions were used as distractors. Verbs were classified in transitives (n=20), unergatives (n=10) and unaccusatives (n=4), some examples directly drawn from the test are given in (7)-(9).

- (7) a. Chi ha mangiato la mela?
who has eaten the apple
- b. L'ha mangiata la signora.
 CL has eaten the woman
 'The woman ate it'
- (8) a. Chi ha parlato nel video?
 who has spoken in the video

- b. Ha parlato l'uomo.
has spoken the man
'The man spoke'
- (9) a. Chi è partito?
who is left?
- b. È partito il ragazzo.
is left the guy
'The guy left'

The procedure adopted in the task confirmed to be effective in providing the correct pragmatic conditions for eliciting spontaneous but controlled answers with the subject as new information focus. Only sentences containing a verb were considered valid answers for the analysis.

4 RELEVANT DIFFERENCES BETWEEN FINNISH AND ITALIAN

Finnish, the L1 of the subjects, and Italian, the L2 of the subjects differ substantially in many aspects. Finnish is a Finno-Ugric language with a rich morphology, it does not have articles nor clitic pronouns, and it is a partial null subject language, namely null subjects are allowed for first and second person singular and plural but not for third, as exemplified in (10) and (11):

- (10) a. __ tulen huomenna
come-PRES1sg tomorrow
- b. __ tulette huomenna
come-PRES2pl tomorrow
- (11) a. *(hän) tulee huomenna
(s/he) comes-PRES3sg tomorrow
- b. *(he) tulevat huomenna
(they) come-PRES3pl tomorrow

Nonetheless, 3rd person null subjects are allowed under special circumstances: (i) in subordinate clauses when the subject is co-referential with the subject of the main clause, observe the different interpretation between the null subject in (12)a and the overt one in (12)b:

- (12) a. Mattii sanoi, että __i/*k soittaa myöhemmin
Matti-NOMsg say-PAST3sg that __ call-PRES3sg later
'Matti said that he will call later'
- b. Mattii sanoi, että häni/k soittaa myöhemmin.
Matti-NOMsg say-PAST3sg that (s)he-NOMsg call-PRES3sg later
'Matti said that (s)he will call later'

(ii) in generic sentences when the 3rd person null subject is referring to a generic 'one':

- (13) Jos __ syö terveellisesti __ voi paremmin.
 If __ eat-PRES3sg healthy __ feel-PRES3sg better-NOM
 'If one eats healthy one feels better'

Finally, a null expletive pronoun is found in extraposed clauses, (14)a and with weather verbs, (14)b. Notice that in colloquial Finnish the expletive pronoun can be optionally overt in both cases. In the extraposed sentence it is considered referential whereas with weather verbs it is often referred to as a quasi-argumental or quasi-referential expletive¹ (Holmberg & Nikanne 2002).

- (14) a. __ oli kiva, että soitit.
 be-PAST3sg nice that call-PAST2sg
 'It was nice that you called'
 b. __ sataa.
 rain-PRES3sg
 'It rains'

Finnish has a basic SV (O) order in declarative clauses. From a previous adaptation of the same video test to Finnish (Dal Pozzo 2010) it emerged that the preferred answering strategy in SNI contexts was overwhelmingly of the SV(O) type even though other strategies, such as O/AdvVS, O= topic/known information, clefts, reduced clefts and locative clefts were not excluded, examples are given in (15)-(18).

- (15) a. – Kuka söi omenan? – Sen söi vaalea nainen.
 who eat-PAST3sg apple-ACC? It-ACC eat-PAST3sg blond-NOMsg woman-NOMsg
 b. – Kuka puhui videossa? – Videossa puhui se poika
 who speak-PAST3sg video-INEsg? Video-INEsg speak-PAST3sg that-NOMsg
 boy-NOMsg
 '- Who spoke in the video? - In the video spoke that boy'
- (16) a. Kuka vastasi?
 who-NOMsg answer-PAST3sg

¹ See Holmberg and Nikanne (2002, 2008), Holmberg (2005) on the possible use of the expletive pronouns *se/sitä* in impersonal constructions in colloquial Finnish, (i) and (ii) are the counterparts of (14)a-b, respectively:

- (i) Se oli kiva, että soitit.
 expl was nice that you called
 (ii) Nyt se taas sataa.
 now expl again rains
 'Now it is raining again'

b. Se oli tuo tyttö, joka vastasi
it was that girl-NOMsg who-NOMsg answered

(17) a. Kuka soitti?
who-NOMsg call-PAST3sg

b. Se oli Kaisa
it was Kaisa-NOM

(18) a. Kuka on lakaissut?
who-NOM has swept

b. Siinä oli yksi tyttö, joka lakaisi
this-INE was one girl-NOM who swept
'There was a girl who swept'

Italian is a romance language without a nominal inflectional morphology (only in the pronominal system some forms are reminiscent of cases). It has articles to mark definite and indefinite noun phrases, it has clitic pronouns and it is a pro-drop language, as shown in (19):

(19) Parto/parti/parte/partiamo/partite/partono domani.
leave-PRES1sg/2sg/3sg/1pl/2pl/3pl tomorrow

Italian has a basic SVO order and it is a null subject language. Postverbal subjects are allowed under various discourse-pragmatic circumstances and the postverbal subject in (3) can be interpreted in different ways depending on the context: as new information focus, (20), as contrastive focus, (21), or as topic, (22). Moreover, a postverbal subject is pragmatically correct also in all-new contexts, (23). As previously introduced, in the present study we follow Belletti's work (2001, 2004, 2005) on postverbal subjects, where S=new information focus, and the empirical studies on the topic.

(20) a. Chi è partito / ha parlato ?
who has left / has spoken
b. E' partito / ha parlato Gianni
has left / has spoken Gianni

(21) a. Sara ha mangiato la mela.
Sara has eaten the apple
b. No, l'ha mangiata Lucia.
No CL has eaten Lucia

(22) a. Che cosa ha poi fatto Gianni?
What has then done Gianni
b. Ha (poi) parlato, Gianni
has (then) spoken Gianni

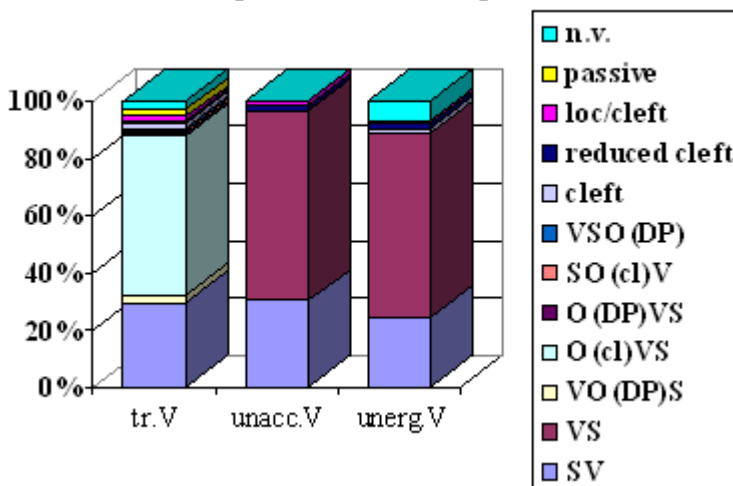
- (23) a. Che cosa è successo?
 what has happened
 b. Ha telefonato Piero
 has telephoned Piero

In sum, Finnish and Italian differ for their nature as partial null subject language and null subject language, respectively. On the one hand in Italian there is a null referential *pro* which allows for postverbal subjects, whereas in Finnish postverbal subjects are allowed only if there is an overt topic in sentence initial position (cf Holmberg & Nikanne 2002). Moreover, Italian has object clitic pronouns, which may further complicate the target use of VS structures with transitive verbs.

5 RESULTS

The L2 speakers with high L2 attainment show a good rate of postverbal subjects across all verb classes. Nevertheless, the SV order is still widely adopted independently of the verb class. Other answering strategies emerge at a very low rate. The results of Group 1 are summarized in Table 1, detailed data is given below it.

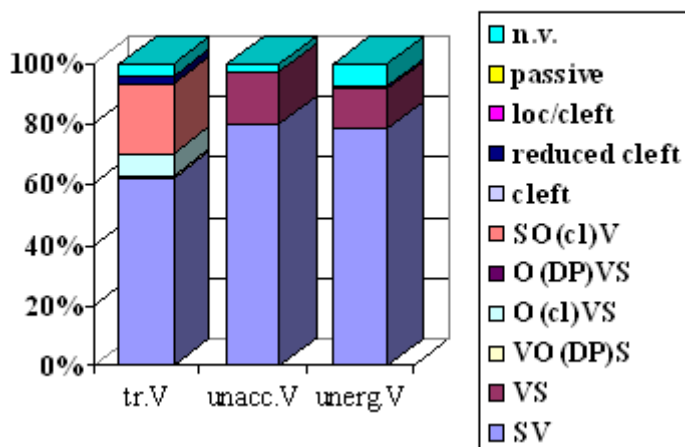
Graph 1: Results Group 1



	tr V	unacc V	unerg V
SV(O)	29,7%(89/300)	30% (18/60)	24,7%(37/150)
VS	--	66,7% (40/60)	64,7% (97/150)
VO(DP)S	2,3% (7/300)	--	--
O(cl)VS	56,3%(169/300)	--	--
O(DP)VS	1% (3/300)	--	--
SO(cl)V	0,7%(2/300)	--	--
VSoDP	1% (3/300)	--	--
cleft	1,7%(5/300)	--	2%(3/150)
Red.Cleft	1%(3/300)	1,7%(1/60)	1,3%(2/150)

The intermediate group shows a very poor use of postverbal subjects across all verb classes. The preferred answering strategy is overwhelmingly SV(O), which is the most widely adopted strategy in their L1 too. The results of Group 2 are resumed in Table 2; detailed data is given below it.

Graph 2: Results Group 2



	tr. V	unacc. V	unerg. V
SV(O)	62,1%(118/190)	80% (32/40)	79%(79/100)
VS	--	17,5% (7/40)	13% (13/100)
VO(DP)S	1%(2/190)	--	--
O(cl)VS	6,8% (13/190)	--	--
O(DP)VS	0,5%(1/190)	--	--
SO(cl)V	22,6%(43/190)	--	--
cleft	0,5%(1/190)	--	--
Red.Cleft	2,1% (4/190)	--	1% (1/100)

Observing the L2 data it seems plausible to explain the wide use of the SV strategy as a phenomenon of *transfer*, in other words the informational value provided by the Italian free inversion in SNI contexts is obtained through the preferred L1 strategy, namely SV, which is grammatical, but not pragmatically correct, in Italian.

The most relevant non target answering strategy is SO(cl)V. A clear shift is observed from the Intermediate L2 group (22,6%) to the advanced one (0,7%). The SO(cl)V strategy seems to be a transitional phase between the L1's SV(O) and the L2's O(cl)VS. Thus, I suggest that following the pattern typical to Finnish, the new information subject appears in the sentence-initial position and that the object, represented as a clitic pronoun, is correctly placed in preverbal position.

A related phenomenon which emerged from the data is that L2 speakers produce overt subjects in contexts where native speakers of Italian would not use them, as deemed unnecessary, (24)a. Interestingly, the use of overt subjects are attested also for 1st person, (4)b, which is apparently unexpected.

- (24) a. – Dov'era seduto l'uomo? – L'uomo era seduto sul divano.
where was sitting the man? The man was sitting on the couch.
b. Quante macchinette del caffè hai visto? – Io ho visto tre ...
how many coffee machines did (you) see? I saw three ...

Overt pronominal subjects can be licensed in Italian but their distribution is governed by discourse-pragmatic factors. I suggest that the use of overt subjects under these discourse-pragmatic circumstances is an instance of transfer from the L1 of the participants, a PNSL.

5.1 Use of clitic pronouns

Clitic pronouns were elicited through questions with transitive verbs with either a full DP or a clitic pronoun, respectively, as exemplified in (25)a-b. Italian realizes the direct object as a clitic pronoun in both cases and this is the most natural way to answer to the type of questions in the task, as was evident from the control group that had an above chance performance.

- (25) a. Chi ha mangiato la mela?
who has eaten the apple
b. Chi l'ha spenta?
who CL has turned off
'Who has turned it off?'

As expected, Group 1 perform in a more 'native like' fashion with respect to Group 2. However, for both groups the presence of a clitic pronoun in the question influences the use of the clitic pronoun in the answer: if the object is realized as a full DP in the question it is more easily realized as a full DP in the answer, too, (25)a and (26)a. In a parallel way, when there is a clitic pronoun in the question then it is more easily used in the answer, too, (25)b and (26)b.

- (26) a. La signora ha mangiato la mela
 the woman has eaten the apple
 b. L'ha spenta Maria
 CL has turned off Maria
 'Maria has turned it off'

Clitic pronouns are never misplaced and whenever used they appear in the correct proclitic position with tensed verbs. When the clitic is omitted it is always replaced by a full lexical noun phrase, strong pronouns are never found. We follow Belletti and Leonini (2004) in considering this replacement as an avoidance strategy in the use of a clitic pronoun. The observation is strengthened by the present data which shows that the use of clitic pronouns clearly increases from the Intermediate level (Group 2) to the Advanced level (Group 1).

General results of both groups are summarized in Appendix II, the results of the two groups were analyzed separately and taking into account the presence of the clitic (Q/cl) or of a full DP (Q/full DP) in the question and in the answer (A/cl and A/full DP, respectively).

6 CONCLUSION

The highly advanced L2 group perform in a more 'native-like' fashion with respect to the intermediate L2 group. However, at the discourse-syntax interface as for the use of postverbal subjects in context of new information and for the use of clitic pronouns, even the highly advanced (native-like) group, with daily contact to the L2 shows a relatively poor performance. It follows that the low-peripheral focus position is not extensively activated in the interlanguage of the highly advanced L2 group and it is very poorly activated in the interlanguage of the intermediate L2 speakers. This is coherent with the non target-like use of overt subjects in the L2. Overall, the results are consistent with previous findings and show that the interface level is prone to more variability than narrow syntax, which is an autonomous and more rigid system (see Sorace et al. 2009 for discussion).

Finally, I assume the non canonical orders attested in the data to be typical instances of the interlanguage grammar: L2 speakers have more available strategies than L1 speakers and, as expected, the emerging orders do not show violation of any formal conditions.

REFERENCES

- Ayoun, D., 2003. Parameter setting in language acquisition. London: Continuum.
- Belletti, A., 2001. Inversion as focalization. In A. Hulk and J.Y. Pollock (eds.) *Subject Inversion in Romance and the theory of Universal Grammar*. 60-90. New York: Oxford University Press.
- 2004. Aspects of the low IP area.. In L. Rizzi (ed.) *The Structure of CP and IP*, L.Rizzi. New York: Oxford University Press.
- 2005. Extended Doubling and the VP periphery. In *Probus*, 17,1:1-35.
- 2006. Answering strategies. A view from acquisition. In *Proceedings of Going Romance 2005*.

- 2008. Answering strategies and the nature of clefts. In Belletti, A. *Structures and Strategies*. London: Routledge.
- Belletti, A. and Leonini C., 2004. Subject inversion in L2 Italian. In S. Foster-Cohen, M. Sharwood Smith, A. Sorace, M. Ota (eds.) *Eurosla Yearbook 4*. Amsterdam/Philadelphia: John Benjamins.
- Belletti, A., Bennati, E. and Sorace A., 2007. Theoretical and developmental issues in the syntax of subjects: Evidence from near-native Italian. *NLLT* 25.657-689.
- Cinque, G. (ed) 2002. *Functional Structure in DP and IP. The Cartography of Syntactic Structures*. vol. 1. New York: Oxford University Press.
- Dal Pozzo, L., 2010. Evidence from experimental results and some questions about the syntax-discourse interface. Paper presented at the XI Finnougric Conference, Pielisaba.
- Guesser, S., 2007. *Soggetto nullo e focalizzazione del soggetto in portoghese-brasiliano*. Master Thesis. University of Siena.
- Holmberg, A., 2002. Deriving OV order in Finnish. In Svenonius, P. (ed.). *The Derivation of VO and OV*. Amsterdam: Benjamins.
- 2005. Is there a little pro? Evidence from Finnish. *Linguistic Inquiry* 36:4. 533-564.
- Holmberg, A., Nayudu, A., Sheehan, M. 2009. "Three partial null-subject languages: a comparison of Brazilian Portuguese, Finnish and Marathi". *Studia Linguistica* 63:1. Wiley-Blackwell Publishing.
- Holmberg, A. and Nikanne, U., 2002. Expletives, Subjects and Topics in Finnish. In Svenonius, P. (ed.) *Subjects, expletives and the EPP*. New York: Oxford University Press.
- 2008. Subject doubling in Finnish: the role of deficient pronouns. In Barbiers, S; Koenenman, O, Lekakou, M, (eds.) *Microvariations in Syntactic Doubling*. Bingley: Emerald. 325-349.
- Modesto, Marcello, 2008. Topic prominence and null subjects. In (ed.) Biberauer, Teresa, *The limits of syntactic variation*. 375–409. Amsterdam: John Benjamins.
- Rizzi, L., 1997. *The Fine Structure of the Left Periphery*. In L. Haegeman (ed.) *Elements of Grammar*, Dordrecht: Kluwer Publications.
- 2004. *The Structure of CP and IP. The cartography of syntactic structures*. Vol. 2. New York: Oxford University Press.
- Sorace, A., Serratrice, L., Filiaci, F., Baldo, M., 2009. Discourse conditions on subject pronoun realization: Testing the linguistic intuitions of older bilingual children. In *Lingua* 119:460-477.
- White, Lydia, 2003. *Second Language Acquisition and Universal Grammar*, Cambridge: Cambridge University Press.

I thank the members of the Firenzen Suomi-Seura (The Finnish-Italian Association) and the Italian students and their teacher, Maria Erämaja of the Turun Työväenopisto (Centre for Workers' Education of Turku) who took part in the test.

APPENDIX I: PARTICIPANTS

Table 1: Group 1- High Italian L2 competence

L2 learner	age	education in Italy	permanence in Italy	Use of Italian at home	Use Italian for working/ studying
S1	65	no	40 years	yes	yes
S2	48	yes (univ. degree)	18 years	yes	yes
S3	45	yes (univ.degree)	25 years	yes	yes
S4	26	yes (as an Erasmus student)	1,6 years	yes	yes
S5	43	yes (univ.degree)	23	yes	yes
S6	23	yes (univ.lessons – Erasmus student)	9 months	yes	yes
S7	29	yes (univ.degree)	5 years	yes	yes
S8	65	no	40 years	yes	yes
S9	45	yes (univ.degree)	24 years	yes	yes
S10	40	no	9 years	yes	yes
S11	62	no	39 years	yes	yes
S12	40	yes (univ.degree)	20 years	yes	yes
S13	30	yes (as an Erasmus student)	2 years	yes	yes
S14	54	yes (degree at the music academy)	30 years	yes	yes
S15	55	yes	30 years	yes	yes

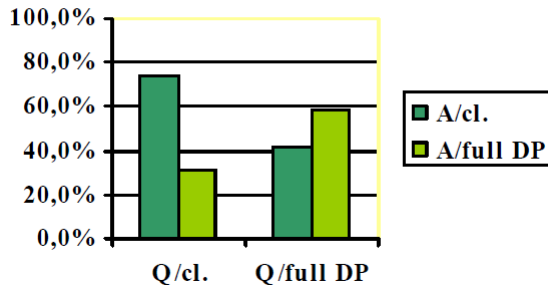
Table 2: Group 2 - Intermediate Italian L2 competence

L2 learner	Age	Exposure to Italian	do you use Italian at home?	do you use Italian for working/ studying?
S1	63	7 years	no	no
S2	60	3 years	no	no
S3	65	5 years	no	no
S4	60	3 years	no	no
S5	62	5 years	no	no
S6	66	5 years	no	no
S7	65	2,5 years	no	no
S8	25	1 year	no	no
S9	26	2 years	no	no
S10	60	10 years	no	no

APPENDIX II

Highly advanced L2 speakers:

Use of clitic pronouns - Group 1



Intermediate L2 speakers:

Use of clitic pronouns - Group 2



Charlotte Gooskens and Nanna Haug Hilton

The effect of social factors on the comprehension of a closely related language

1 INTRODUCTION

The three mainland Scandinavian languages, i.e. Danish, Swedish and Norwegian, have a reputation of being mutually intelligible, which means that the speakers are able to communicate each using their own languages. However, in daily practice inter-Scandinavian communication sometimes fails. The results of a number of studies have shown that especially Danes and Swedes have difficulties understanding each other's language, while Norwegians are generally quite good at understanding the two neighbouring languages (see e.g. Delsing and Lundin-Åkesson 2005, Maurud 1976).

The level of inter-Scandinavian intelligibility may depend upon two kinds of factors: (1) linguistic distance between the languages at different linguistic levels (phonetics, vocabulary, morphology, and syntax), (2) non-linguistic factors such as attitude towards the languages of the other Scandinavian countries and experience with the other languages. In this paper we will focus on the role of non-linguistic factors for the intelligibility of Danish among Norwegians. Previous studies suggest that non-linguistic factors may contribute to the successful communication between speakers of closely related languages (Maurud 1976, Bø 1978, Börestam 1987, and Wolff 1959). The existence of negative attitudes or social stigmas attached to languages is often seen as a potential obstruction for successful intergroup communication. Ultimately, of course, the level of intelligibility also depends on the amount of experience with the other language(s), including formal instruction.

In the previous investigations of mutual intelligibility within Scandinavia, mentioned above, only listeners from the southern part of Norway close to the neighbouring countries were tested. It could be that informants who live further away, with fewer opportunities to visit neighbouring countries, have a lower level of intelligibility of the neighbouring languages. To investigate this we test the intelligibility of Danish among two groups of Norwegians, one from the southern part and one from the far north. These two groups have approximately the same linguistic prerequisites for understanding the neighbouring languages. Therefore, if a difference in intelligibility is found we have a strong indication that non-linguistic factors play a role for intelligibility of a neighbouring language. Our expectation is that the northern group has had less contact with Danish than the southern group because they live so far away from speakers of

the language. It is not clear beforehand whether there is a difference in attitudes towards Danish between the two groups of listeners, however.

2 METHOD

To test the intelligibility of spoken Danish among Norwegians we conducted a word intelligibility experiment via the Internet among a northern group and a southern group of subjects. To investigate the attitudes towards the Danish language and its speakers and the amount of contact with the language we also had the two groups of pupils fill in an in-depth questionnaire.

Subjects

Subjects were recruited through secondary schools and were pupils who attended university preparatory courses. The northern group consisted of 12 pupils from the places Honningsvåg, Tana and Alta in Finnmark which is the northern-most province in Norway, more than 2000 kilometres from Denmark. The southern group was 13 pupils from Hønefoss in the province of Buskerud. The distance to Denmark for these speakers is approximately 300 km (see Figure 1). The dialects spoken in the northern and the southern areas are linguistically very similar (see Gooskens and Heeringa 2005). Both varieties can be said to be heavily influenced by the written standard Bokmål, and employ forms that are predominantly found in this codified variety.

The subjects were asked a number of questions about their background before they participated in the experiment. Part of the information is summarized in Table 1. We see that the mean age of the two groups of subjects is almost the same (16.4 versus 17.1 years). Both groups consist of approximately the same number of boys and girls. All of the subjects indicated that they spoke Norwegian with both their mother and their father so that it can be assumed that Norwegian is their first language.

Table 1: Number of subjects, gender and mean age per group of listeners.

	Subjects	
	North	South
Number	12	13
Gender	F 8, M 4	F 9, M 4
Age	16-18 (mean 16.4)	16-19 (17.1)



Figure 1. Map of Scandinavia showing where the test subjects lived.

Intelligibility experiment

To test intelligibility, an Internet-based experiment was conducted.¹ In this experiment, Norwegian subjects were confronted with 384 Danish isolated singular nouns in spoken form. The nouns were randomly selected from a list of 2575 highly frequent Dutch words and translated into Danish. We assume that this random selection of words is representative for the Danish language as far as their linguistic properties (for example the distribution of phonemes and percentages of non-cognates) are concerned. We preferred to test isolated words rather than a whole text because this gave us the possibility to control the input completely. The experiment that was used was developed to also be applicable for other studies, among others a study of which specific word characteristics influence intelligibility of a closely related language. Furthermore, the decoding of content words is central to understanding speech. As long as the subject correctly recognizes content words, he will be able to piece the speaker's message together. In a pre-test to the word comprehension experiment, we assured that all the nouns in the stimuli were known to subjects from the test group, i.e. high school pupils aged 15-18 years. The test words were read aloud by a male native speaker of Danish (from Frederiksberg near Copenhagen) and recorded in a professional sound recording studio.

¹ The experiment also included other Germanic language pairs. It may be found on the Internet at <http://www.let.rug.nl/lrs>. It is possible to participate in the test with a guest account (login: germanic, password: guest).

The experiment would have been too arduous if all subjects had been given all test words. Therefore, each subject heard only one word block consisting of one quarter of the 384 words. The choice of the words and the order of presentation were randomized so that possible tiredness effects were neutralized. The subjects listened to the test words via headphones. They were requested to write the Norwegian translation into a text field within ten seconds. Prizes were offered to the highest-scoring participants in an attempt to encourage them to complete the tasks to the best of their ability.

The results from the word comprehension experiment were automatically categorized as right or wrong through a pattern match with expected answers. The answers that were categorized as wrong were subsequently checked manually. Responses that deviated from the expected responses due to a mere spelling error were counted as correct identifications. Spelling errors were objectively defined as instances where only one letter had been spelt wrongly without resulting in another existing word. So, for example the mistake in *terrase* (correct *terrasse*) 'terrace' is considered a spelling mistake and therefore counted as correct (only one wrong letter without resulting in another existing word), while *hånd* (correct *ånd* 'spirit') was not counted as correct because the spelling mistake results in an existing word meaning 'hand'. Some words have more than one possible translation. For example, the Danish word *arbejde* was sometimes translated into Norwegian *arbeide* and sometimes into *jobb* both meaning 'job'. Both translations were counted as correct. In the case of homonyms, both possible translations were accepted as correct. For example, Danish *hær* can be translated correctly into Norwegian *her* 'here' or *hær* 'army'. After this procedure, we had obtained a score of zero (word not identified) or one (word identified) per word for each subject. We then also calculated the percentage of correct translations per subject, obtaining the intelligibility score per subject.

Contact questionnaire

After the intelligibility experiment had taken place, the subjects were asked to fill in the questionnaire with attitude and contact questions at home on a voluntary basis.

The subjects were asked questions about their frequency of visits to Sweden and Denmark and the frequency of Swedish and Danish linguistic input. The questions were the following:

1. 'How often have you been to Denmark?'
2. 'When did you last visit Denmark?'
3. 'How often do you hear Danish?'
4. 'How often do you talk to a Danish -speaking person?'
5. 'How often do you read Danish?'

As answer to questions 1, the subjects could choose between the following: 'never', '1 or 2 times', '3 to 5 times' or 'more than 5 times'. Question 2 could be answered with 'last year or more recently', '1 to 2 years ago', 'more than 3 years ago' or 'never'. Questions 3, 4 and 5 could be answered with 'never', 'less than once a year', 'a few times a year', 'once a month', 'once a

week', and 'every day'. 'In order to be able to calculate the mean contact scores, the answers were recalculated to an ordinal scale from 0-3 and 0-5 respectively.

Attitude questionnaire

To elicit the attitude of the subjects they were asked to characterise the neighbouring languages and the speakers. First, the subjects were asked directly about their attitude towards Danish: 'What do you think of the Danish language (please select the alternatives that you think describes Danish?)'. The alternatives are listed in Table 2. They are the negative and positive characteristics on the dimensions of dynamism, superiority and attractiveness as defined by Zahn and Hopper (1985). Some of the adjectives listed in Table 2 are antonyms, e.g. 'correct' - 'incorrect', while others do not have a positive or negative antonym.

The subjects were asked to give their opinion about the speakers: 'What do you think of Danes?' (see Table 2). The alternatives were presented in random order and the subjects could select as many characteristics as they liked.

Table 2. Overview of the positive and negative adjectives that the subjects could chose to characterise the Danish language and its speakers (the adjectives have been translated from Norwegian)

Danish language		Danes	
negative	positive	negative	positive
Incorrect	Correct	Reserved	talkative
Unattractive	attractive	poor	rich
Unalluring	alluring	self-centred	considerate
Repulsive	beautiful	insecure	self-confident
Un-cool	cool	unintelligent	intelligent
uncultivated	cultivated	bad-tempered	calm
boring	amusing	cross	happy
ugly	pretty	unserious	serious
unsexy	sexy	pushy	friendly
dull	interesting/fun	slow	cosy
tedious	cosy	dishonest	honest
		snobbish	not snobbish
		unsympathetic	sympathetic

Next the subjects were asked 'Would you like to live or study in Denmark?' They could answer 'yes', 'no' or 'maybe'. We assume for the sake of the analysis that a positive answer may signal a positive attitude towards Danish and Danes.

Finally, the subjects were asked how much of a Danish news item on the radio they think they would understand. They could choose between the following answers: '1-25%', '26-50%', '51-75%' and '76-100%'. In the strict sense this question may not measure the subjects' attitudes towards Danish, but still the answers express the expectations that they have when being

confronted with Danish. These expectations do not necessarily reflect how well the subjects are in fact able to understand Danish, however.

3 RESULTS

3.1 Mean results

In the first row of Table 3 the intelligibility expressed as the mean percentages of correct translations of the spoken Danish words is presented for the two groups of subjects. Furthermore, the attitudes expressed as the mean number of positive and negative characterisations of the Danish language and Danes are presented. Finally, the amount of contact with the Danish language and Danes expressed as percentages is presented. In the last column of the table, the results of independent samples t-tests are given.

Table 3. Results of the intelligibility tests and the attitude and contact scores for the northern and the southern group of subjects. The intelligibility scores are expressed as percentages, the contact scores as points (see above), the attitude scores are the mean number of points given by the subjects (see explanation in Section 2).

	Northern group	Southern group	sign. (df=23)
intelligibility	76.3	79.6	$t = -1.015, p = 0.32$
contact:			
hear Danish (max 3)	2.5	2.85	$t = -1.082, p = 0.29$
talk to Danes (max 3)	1.27	1.42	$t = -0.382, p = 0.71$
read Danish (max 5)	0.83	1.85	$t = -1.727, p = 0.10$
frequency: visits to Denmark (max 5)	0.67	2.0	$t = -4.891, p = 0.00$
recentness: visits to Denmark (max 5)	0.67	2.0	$t = -3.913, p = 0.01$
attitude:			
positive towards Danish (max. = 11)	1.0	2.5	$t = -2.376, p = 0.03$
negative towards Danish (max. = 11)	2.5	1.7	$t = 0.843, p = 0.41$
understand Danish (max. = 4)	3.1	2.3	$t = 1.705, p = 0.10$
positive towards Danes (max. = 13)	1.8	4.5	$t = -2.947, p = 0.07$
negative towards Danes (max. = 13)	1.6	0.3	$t = 1.915, p = 0.07$
live or study in Denmark (max. = 2)	0.9	0.8	$t = 0.582, p = 0.57$

Intelligibility

The southern group translated a little more words correctly (79.6%) than the northern group (76.3%), but the difference is not significant ($p = 0.32$). Since the data presented here are part of an investigation where the intelligibility of a larger number of subjects were tested (see note 1) we can compare the results of this small group of subjects to that of a larger group. The data set

used in the present analysis was smaller because only for this dataset also attitude and contact scores were available (see Section 2). In the large survey the intelligibility of 39 subjects from the north and 118 from the south were tested. These subjects translated 65.9% (northerners) and 65.0% (southerners) of the Danish words correctly. This percentage is lower than for the restricted dataset with 25 subjects, probably due to the fact that participation was obligatory in the large investigation. The subjects in the restricted set also volunteered to participate in the questionnaire and are probably more eager students. Most importantly, we find that for neither the large data set nor for the restricted data set the difference between the northern and the southern group is significant. This means that the restricted data set can be considered representative as far as the difference between the two groups is concerned. Both groups translated Danish words equally well.

Contact

In the second part of Table 3 it is shown how often the subjects have contact with the Danish language and its speakers. We see that on all contact scales southerners have higher scores than the northerners. The differences for the two first contact scales that are concerned with the spoken language are not significant. The difference in reading patterns is closer to being significant, (0.83 versus 1.85) but the difference is not as large as between the groups' frequencies of visits to Denmark (northerner's score is 0.67, and the southerners' 2.0). The southerners have also visited Denmark more recently (northerners score is 0.67 versus southerners' 2.0). Of the twelve northerners, five have never been to Denmark and only one has been 3-5 times. Only one of the thirteen southerners has never been to Denmark, while the rest of them have either been between 3 and 5 times (10) or more often, i.e. 5 or more times (2).

Attitude

We calculated the mean number of positive and negative characteristics chosen for Danish in the two groups of listeners (bottom part of Table 3). In general, the northern group is more negative about the Danish language than the southern group, while the northern group is more positive about their ability to understand Danish. The difference between number of positive characteristics of Danish is significant between the two groups. Many southerners find the Danish language beautiful, attractive and interesting, while the northerners more often characterise it as boring and unattractive. When asked to characterise the speakers rather than the language the southern group often chose the positive characteristics 'happy', 'jovial' and 'talkative' while the northern group often chose negative adjectives such as 'pushy', 'sluggish', 'snobbish' and 'unintelligent' to characterise the Danes.

The attitude data show that the southerners are more positive towards Danish and Danes than the northerners. It is possible that northerners in general tend to evaluate languages and their speakers more negatively than the southerners. In order to check whether this is the case, we compared the attitude results to attitude scores for Swedish and Swedes. These results are gained from a questionnaire with the same set-up about Swedes and the Swedish language as the questionnaire about Danish and Danes. The listeners came from the same schools and classes in the two geographic areas as the listeners who participated in the experiment testing the attitudes towards Danish and Danes. The results show that the northern listeners were not

significantly more negative towards Swedish and Swedes than the southerners on any of the scales. This shows that the northerners are not in general more negative in their judgments of languages and their speakers than the southerners. The negative judgments of Danish cannot be explained by a general negative attitude.

3.2 Correlations

Relationship between contact and attitude

The results presented in Table 3 show that in general southerners have more contact with and are more positive towards the Danish language than the northerners. To investigate whether there is a relationship between these two factors we correlated the ordinal contact scores with the attitude scores using a Spearman rho correlation. The results show that on most scales no significant correlations are found. However positive correlations are found at the 1% level for contact scores expressing numbers of visits to Denmark and the number of positive characteristics of the Danes (correlation coefficient .51). A positive correlation at the 5% level is found between the number of visits to Denmark and positive attitudes towards the Danish language (correlation coefficient .42). There is also a positive correlation at the 5% level between recentness of visits to Denmark and the number of positive evaluations given of Danes (coefficient .43) It is tempting to conclude from this that visits to Denmark result in positive attitudes, but it should be kept in mind that such correlations do not say anything about the causal relations. It is also possible that subjects who have a positive attitude go to Denmark more often. However, this possibility is smaller seen in the light of the fact that the listeners are so young that a positive attitude could not really have led to an increased contact pattern.

Relationship between intelligibility and extra-linguistic factors

It is generally assumed that a positive attitude and previous experience with a language will make it easier for a listener to understand it than when the listener has a negative attitude or little experience. The fact that the southerners in our investigation generally have more contact with, and are more positive towards, the language and the people in Denmark may therefore give rise to the expectation that the southerners are also better at understanding Danish. However, as became clear from Table 3 the northerners translated just as many Danish words correctly as the southerners. This gives rise to the expectation that there is no relationship between intelligibility and extra-linguistic factors in our investigation. We correlated the intelligibility scores with the attitude and contact scores as found in Table 3. The results show that there is no correlation between the mean intelligibility scores and the contact or attitude scores in the data.

4 CONCLUSIONS

Some previous studies have established a link between language attitudes, contact and language comprehension abilities (e.g. Delsing and Lundin-Åkesson 2005). In the current study no link

has been established between the intelligibility scores from 25 informants and the attitudes as expressed by the same informants. Nor has a correlation been found between intelligibility scores and the informants' history of contact with the language in question. Informants in Northern Norway have very different contact patterns with the Danish language and people from informants in Southern Norway. These different contact patterns do not correlate significantly with different comprehension scores obtained in the two locations, however. The contact scores calculated in this study are based on day to day contact with the neighbouring language, but also on the frequency of visits to the neighbouring country. This contact was reported as relatively infrequent by both groups of informants, after all 3-5 visits to a country during a lifetime do not constitute frequent contact. However, the reading and listening frequencies that informants reported were rather high, and no meaningful relationship can be found between these scores and comprehension either. The link between dispersed contact frequency on the one hand and comprehension scores on the other therefore seems contestable. Probably, more frequent or intensive contact is necessary to result in improved intelligibility. Future studies should consider the role of much more frequent contact on comprehension. Informants who work in the neighbouring countries for a longer period of time every year or people who go regularly for meetings could be interviewed in more detail for this type of study. If these informants score radically better in comprehension tests than informants who work elsewhere for longer periods of time or commute to elsewhere, one could argue a link between contact and comprehension more successfully. As of now, however, no indication of a relationship between the two variables has been found.

There is, however, a significant relationship between language attitudes and language contact patterns in the current data set. It appears that Norwegian pupils who have had more contact with the Danish language are also substantially more positive towards the variety. The two populations in this study live at opposite sides of Norway and the northernmost informants are very restricted in their possibilities of visiting Denmark. Our results give an indication that this geographical difference is meaningful for attitudes towards the neighbouring country and its language. The dynamics of this relationship must be explored further in the future, however. It is unclear from our data set whether Southern informants hear and read Danish more than the northern informants because they are more positive towards the language, or whether they are more positive towards the language because they have been exposed to it on a more regular basis. Future research must also focus more in depth on the whole contact and attitude profiles of the informants. The pupils in the current study differ in their attitudes towards Danish only, while they have comparable attitudes towards the Swedish language and people. This result shows that the negative attitudes towards Danish cannot be explained by a general negative attitude among the northerners.

Finally, this current study has also shown that language attitudes are not good predictors of comprehension. This is a finding that somewhat contradicts previous results obtained by other researchers (e.g. Delsing and Lundin-Åkesson, 2005). It could be that the overt attitudes elicited in the current study are not particularly good predictors of linguistic behaviour, while more subconscious attitudes could help us explore the link between extra-linguistic factors and comprehension more in depth. Future research should focus on obtaining informants'

subconscious attitudes towards neighbouring languages and cultures and explore a potential link between them and comprehension scores.

BIBLIOGRAPHY

- Bø, I. 1978. *Ungdom og naboland. En undersøkelse av skolens og fjernsynets betydning for nabospråksforståelsen* [Youth and neighboring country. A survey of the importance of school and television for understanding neighbouring languages] (Rogalandforskning 4). Rogalandforskning, Stavanger.
- Börestam, U. 1987. *Dansk-svensk språkgemenskap på undantag*. Uppsala: Uppsala Universitet.
- Delsing, L.-O. and K. Lundin-Åkesson. 2005. *Håller språket ihop Norden? En forskningsrapport om ungdomars förståelse av danska, svenska och norska* [Does the language keep the Nordic countries together? A research report on how well young people understand Danish, Swedish and Norwegian]. Copenhagen: Nordisk ministerråd.
- Gooskens, C. and W. Heeringa. 2005. De moderne norske dialekters stilling indenfor den nordiske sprogfamilie [The position of the modern Norwegian dialects within the Nordic language family]. In: S. Lie, G. Nedreliid and H. Omdal (eds.), *MONS 10. Utvalde artiklar frå det tiande Møte om norsk språk i Kristiansand 2003*. Kristiansand: Høyskoleforlaget. 133-146.
- Maurud, Ø. 1976. *Nabospråksforståelse i Skandinavia. En undersøkelse om gjensidig forståelse av tale- og skriftspråk i Danmark, Norge og Sverige* [Mutual intelligibility of languages in Scandinavia. A study of the mutual understanding of written and spoken language in Denmark, Norway and Sweden]. Stockholm: Nordiska Rådet.
- Wolff, H. 1959. Intelligibility and inter-ethnic attitudes. *Anthropological Linguistics* 1. 34-41.
- Zahn, C and R. Hopper. 1985. Measuring Language Attitudes: The Speech Evaluation Instrument. *Journal of Language and Social Psychology* 4 (2), 113-123.

ACKNOWLEDGMENTS

We thank pupils at Tana VGS, Alta VGS, Nordkapp VGS, Hønefoss VGS, Ringerike VGS and Kongsberg VGS for participating in the intelligibility experiment and for filling in the questionnaires.

*Nanna Haug Hilton, Charlotte Gooskens, Anja Schüppert,
Renée van Bezooijen and Vincent van Heuven*

The influence of non-native morphosyntax and phonology on the intelligibility of a closely related language

1 INTRODUCTION

This article discusses the role that a foreign morphology and syntax might play for the comprehension of a closely related language. We investigate this with an experiment testing Danish listeners' comprehension of Norwegian morphology, syntax and phonology. The claim that differences in morpho-syntax have an influence on our comprehension abilities when listening to a closely related language is previously untested. This study aims to rectify this and add to our understanding of how individual linguistic factors contribute to comprehension of speech. Before turning to the details of the experiment we present some background literature about mutual intelligibility in Scandinavia alongside general literature about the role of syntax for intelligibility.

2 BACKGROUND

The Scandinavian languages Swedish, Danish and Norwegian are closely related. Within the Scandinavian Germanic speech community, communication mostly happens in speakers' native varieties rather than in a lingua franca. This situation differs from inter-dialectal communication where speakers mostly have a standardised variety to refer to if miscommunication occurs. The situation in Scandinavia has been referred to as one of 'receptive bilingualism' (Braunmüller and Zeevaert, 2001) or 'semi-communication' (Haugen, 1966). The intelligibility between speakers of Swedish, Danish and Norwegian is far from perfect, however.

Extra-linguistic factors like attitudes and contact frequency have been identified as influential for intelligibility of closely related varieties within Scandinavia (Delsing and Lundin-Åkesson, 2005) but a causal relationship between the extra-linguistic factor attitude and

comprehension is difficult to establish (van Bezooijen and Gooskens, 2007). Previous research has also shown that intrinsic linguistic factors play a role for the comprehension of a closely related language. In addition to lexical differences, phonetic distances between the languages as well as differences in word length, word frequencies and the occurrence of foreign sounds were found to affect intelligibility negatively between speakers of Scandinavian languages (Gooskens, 2007; Kürschner et. al 2008). There has been very limited research as of yet, however, of the role that syntactic differences might play for the comprehension of closely related languages.

This article focuses on that topic within the Scandinavian linguistic community, an ideal situation to explore the role of syntactic differences on comprehension. The Scandinavian languages Norwegian, Swedish and Danish have a large number of cognate words in common. They differ rather substantially with respect to their phonologies, but only minor morphological and syntactic differences exist between them. The languages are all verb-second languages, for example, all have expletive subjects, and suffixes to express definiteness on nouns (for a discussion in English of more syntactic similarities between Scandinavian languages see Holmberg and Platzack, 2005). These three national languages are thus structurally similar to a large degree but with certain differences. Investigations of the relative influences of these differences on comprehension help create a clearer view of which linguistic factors are important to aid multilingual communication.

Some empirical studies of differences in morphology or word order have previously been conducted in the Scandinavian speech community, but these are concerned with differences in production rather than the potential problems differences might cause for comprehension. Ridell (2008) studies speech accommodation in Swedes and Danes who work together, and focuses partly on syntactic differences. She notes that word-order differences between Danish and Swedish are rare and that particle placement in phrasal verb constructions is the only difference frequent enough to study quantitatively. Her results show that Swedish speakers accommodate the particle placement pattern in their speech towards that of Danes (Ridell, 2008:65,203) but it is unknown whether these results mean that particle placement differences play a role for Danes' comprehension of Swedish in the first place.

When it comes to the role that syntax plays for comprehension of a variety that is linguistically similar to one's own, studies have focused largely on comprehension of or by second language learners. Van Heuven and de Vries (1981), for example, investigated the impact of foreign phonic features versus foreign non-phonic features word order, lexical choice and morphological features on the comprehension of spoken Dutch. The non-phonic factors were influential for the native listeners' comprehension of the speech material but phonic factors contributed more to listeners' comprehension than did the non-phonic ones (see also van Heuven and de Vries, 1981:316). Blau (1990) found that although syntactic modifications to English affect second language learners' comprehension when reading, the effect of such modifications was not significant for the comprehension of spoken language.

The role of morphosyntactic differences for comprehension of a different dialect or, specifically, a standard variety, has also received some attention in previous studies. Speidel *et al.* (1985), investigated to what degree a simplified syntax or a native phonology might improve Hawaiian English school children's comprehension of Standard English. A simplified syntax,

attained by reducing morphological complexity, had no significant effect on children's comprehension of a different dialect.

The question of how humans process and understand sentences has received attention in different linguistic disciplines. What most accounts have in common is an agreement that grammatical constructions are created online by the listener during the comprehension process (e.g. Frazier 2002 [1987]). We assume that the comprehension of a foreign word order can be compared to the comprehension of scrambled sentences. When confronted with sentences where a word is not in its canonical order it is believed that a listener relies somewhat on the distance between the canonical gap and the word that has been moved. Nearby words are more likely to be associated with one another (Fodor and Frazier, 1980:442) and it is therefore assumed that the greater the linear distance of the displacement is between the actual (incorrect) and the canonical sentence position of a word (or larger constituent), the longer comprehension will take (cf. Sekerina, 2003). A question that needs more investigation, however, is whether syntactic anomalies in foreign linguistic varieties are processed in the same manner. This article discusses this issue in greater detail and explores whether the processing of Norwegian syntax and morphology is problematic for Danes in comparison with the processing of Danish syntax and morphology by the same listeners.

The focus of this article is to explore the relative influence morphosyntactic differences might have on comprehension of aural input as opposed to the influence of phonological differences. It is hypothesised that aural comprehension of sentences is the most efficient in one's native phonology and morphosyntax. Furthermore we predict that input with native phonology but foreign morphosyntax will make comprehension more difficult but less difficult again compared to input with foreign phonology but native morphosyntax. The prediction is also that sentences where both phonology and morphosyntax is foreign are the most difficult to comprehend for listeners. To investigate the issues above, an experiment was designed to assess to what degree aural comprehension is compromised by differences in phonology and morphosyntax between Danish and Norwegian for Danish listeners.

3 METHODOLOGY

3.1 Stimuli

Identifying syntactic differences between varieties is not straightforward. It is difficult to create an exhaustive overview of all syntactic differences that exist between two languages, even if one were to consider written language only. There is little data available on syntactic constructions that are possible in spoken varieties of Scandinavian languages (although these are currently being mapped out in the projects attached to the ScanDiaSyn network (ScanDiaSyn, 2010)). Constructions that are available in the standard written languages were therefore chosen for the current investigation. The syntactic constructions were identified by consulting literature for Norwegian, Swedish and Danish learners of the neighbouring languages. A number of syntactic constructions that cause problems for Scandinavian learners of another Scandinavian language are described in the literature about inter-Scandinavian language learning (e.g. Brøndsted, 1967; Christensen, 2007; Teleman, 2008). The accounts assess which linguistic factors cause problems

for Scandinavian learners of other Scandinavian languages and hence also give an indication of which linguistic factors are different between the languages.

According to the literature written on inter-Scandinavian language learning (Brøndsted, 1967; Christensen, 2007; Teleman, 2008) the biggest discrepancy in syntactic constructions identified is between Danish and the other two Scandinavian national varieties. We therefore decided to test the influence on comprehension of differences between Norwegian and Danish, as the number of cognate words has been shown in an earlier study (Gooskens, 2007) to be somewhat larger between these two varieties than between Swedish and Danish. An aim for the investigation design was to create stimuli that consisted of cognate words only so as to enable testing of the particular influence of phonology and morphosyntax (as opposed to the influence of lexical differences).

Testing whether constructions in Norwegian are problematic for Danes (rather than whether Danish constructions are problematic for Norwegians) was believed the most likely to yield results in this study. This is because the number of differences between Norwegian and Danish are not symmetric. Because syntactic variability is wide-spread in the languages, and perhaps especially in Norwegian, the problematic syntactic constructions for Danes in Norwegian do not necessarily have corresponding problematic constructions for a Norwegian in Danish. A number of constructions available in Norwegian do not exist in Danish, e.g. double marking of demonstratives, or placement of phrasal verb particles before, rather than after, an object. Although phrasal verb particles are always placed after the object in Danish sentences, however, this is a variable feature in Norwegian, where particles can either be placed before or after the object, as is also the case for English in the following Norwegian example: *Jeg skrudde lyset av* 'I switched the lights off' and *Jeg skrudde av lyset* 'I switched off the lights'. The syntactic constructions used for this type of experimental study are therefore only problematic to listeners from one language group (e.g. Danes listening to Norwegian), but not the other (e.g. Norwegians listening to Danish) due to more syntactic variants being available in the latter's language.

Five syntactic or morphosyntactic constructions in Norwegian that Danes struggle with when learning the language were chosen for our experimental study. All of the Danish construction variants are also possible (although somewhat archaic) constructions in Norwegian.

Firstly, a number of Germanic languages, including Norwegian, have variable particle placement in clauses consisting of a transitive particle verb (a phrasal verb) and its object, but Danish does not (cf. Holmberg and Platzack, 2005:427). This means that in Norwegian both construction 1a and 1b exists, whereas in Danish, 1b is ungrammatical. In phrases consisting of a transitive particle verb and a direct object, the particle always succeeds the object in Danish. In Norwegian, this is variable, and the particle can be placed before the object. Construction 1b, ungrammatical in Danish, but grammatical in Norwegian, may therefore present a comprehension problem for a Dane listening to Norwegian.

1a	<i>En</i>	<i>mann</i>	<i>spiser</i>	<i>maten</i>	<i>opp</i>
	A	man	eats	food+THE	up
	'A	man	eats	the food'	

1b	<i>En</i>	<i>mann</i>	<i>spiser</i>	<i>opp</i>	<i>maten</i>
	A	man	eats	up	food+THE
	'A	man	eats		the food'

The second syntactic difference tested is the placement of infinitive markers and negators, also a word-order difference between the two varieties. In Danish the infinitive marker cannot precede negation, whereas in Norwegian, the infinitive marker can both precede and succeed negation before a verb phrase. In Norwegian this variability could reflect a semantic difference, cf. example 2a and 2b. Note that sentences with this semantic difference are not used as experimental stimuli for the current investigation.

2a	<i>Jeg</i>	<i>lover</i>	<i>å</i>	<i>ikke</i>	<i>vanne</i>	<i>blomstene</i>
	I	promise	to	not	water	flowers+THE
	'I	promise	to	not	water	the flowers'
2b	<i>Jeg</i>	<i>lover</i>	<i>ikke</i>	<i>å</i>	<i>vanne</i>	<i>blomstene</i>
	I	promise	not	to	water	flowers+THE
	'I do not	promise	to		water	the flowers'

It is not possible to create this distinction in Danish, however, where the infinitive marker always follows the negator (cf. Christensen, 1997). This means that a sentence like 2d will be ungrammatical to a Dane, whereas it will be one of two possible constructions in Norwegian. 2c is grammatical in Danish.

2c	<i>Fotomodeller</i>	<i>prøver</i>	<i>ikke</i>	<i>å</i>	<i>bli</i>	<i>tykke</i>
	models	try	not	to	become	fat
	'Models	try	not	to	become	fat'
2d	<i>Fotomodeller</i>	<i>prøver</i>	<i>å</i>	<i>ikke</i>	<i>bli</i>	<i>tykke</i>
	models	try	to	not	become	fat
	'Models	try	to	not	become	fat'

For stimuli belonging to this category, only sentences that are semantically near-identical (in Norwegian) were chosen, as in 2c and 2d, instead of constructions that result in entirely dissimilar meanings like in 2a and 2b.

Thirdly, there exists a difference between Norwegian and Danish in the usage of the reflexive possessive pronouns. In Norwegian, the reflexive possessive pronoun *sin(e)* can refer back to both plural and singular subjects, whereas the form can only refer back to a singular subject in Danish. The Danish plural reflexive pronoun is *deres*, which is the same form as the 3pl possessive pronoun, also used as such in Norwegian. The semantic scope of the reflexive marker *sine* is thus that it can express either a singular or a plural subject's possession in Norwegian, while it can only express a singular subject's possession in Danish. It is believed, therefore, that a

sentence like the one presented in 3b may cause some comprehension difficulty for a Danish listener of Norwegian. Only 3a is a grammatical Danish construction, whereas both constructions are grammatical in Norwegian. In Norwegian, the first sentence could be read as referring to another proprietor of the flowers than ‘the gardeners’, however. 3b is ungrammatical in Danish. The gloss in 3a represents the meaning of the sentences in both Danish and Norwegian, while the gloss in 3b represents the meaning of the sentence in Norwegian.

3a	<i>Gartnerne</i>	<i>vanner</i>	<i>deres</i>	<i>blomster</i>
	Gardeners	water	their (?own)	flowers
	‘Gardeners	water	their	flowers’
3b	<i>Gartnerne</i>	<i>vanner</i>	<i>sine</i>	<i>blomster</i>
	Gardeners	water	their own	flowers
	‘Gardeners	water	their own	flowers’

The fourth construction type is a morphological difference that exists between the varieties: the marking of definiteness in nouns. In Norwegian (and Swedish) demonstrative noun phrases the marking of definiteness occurs twice, once with an overt definite article (*den* ‘the’ or ‘that’ or *denne* ‘this’, for example) and with a definite article suffix. In Danish noun phrases, the definiteness is only marked once. A definite phrase is marked with a definite article suffix on the noun. If the phrase is demonstrative, a definite phrase is marked with a definite article before the noun phrase. The morpheme used in Norwegian for definiteness in a demonstrative phrase is therefore not foreign to Danes. The same definite-article morpheme is found in their native language. It is the double marking of definiteness that is the anomaly.

In rare instances, spoken Norwegian also shows the single marking of definiteness in these types of sentences. This is the variable in certain expressions, like *den amerikanske president* or *den amerikanske presidenten* ‘the American president’. To Danes, the extra marking of definiteness on the Norwegian nouns may cause some comprehension problems. 4a is grammatical in Danish (but archaic in Norwegian). 4b is ungrammatical in Danish and the unmarked variant in Norwegian.

4a	<i>Februar</i>	<i>er</i>	<i>den</i>	<i>korteste</i>	<i>måned</i>
	February	is	the	shortest	month
	‘February	is	the	shortest	month’
4b	<i>Februar</i>	<i>er</i>	<i>den</i>	<i>korteste</i>	<i>måneden</i>
	February	is	the	shortest	month+THE
	‘February	is	the	shortest	month’

The final and fifth construction that may prove difficult for Danes in Norwegian is both morphological and word-order based: the placement of possessive pronouns and the addition of a definite article suffix in possessive noun phrases. In Danish (and Swedish) the possessive

pronouns are placed before the noun. The pre-posed possessive marker also marks definiteness (cf. Holmberg and Platzack, 2005:440). In Bokmål Norwegian, however, the possessive pronoun can either be pre-posed, as in Danish, or post-posed. When post-posed, the possessive pronoun does not work as a definite marker and a definite article suffix must be added to the noun. Norwegians thus have available sentence type 5a as well as 5b, whereas the 5b is ungrammatical in Danish.

5a	<i>En</i>	<i>bonde</i>	<i>arbeider</i>	<i>på</i>	<i>sin</i>	<i>traktor</i>
	A	farmer	works	on	his	tractor
	'A	farmer	works	on	his	tractor'
5b	<i>En</i>	<i>bonde</i>	<i>arbeider</i>	<i>på</i>	<i>traktoren</i>	<i>sin</i>
	A	farmer	works	on	tractor+THE	his
	'A	farmer	works	on	his	tractor'

3.2 The Experiment

To test the effect of the Norwegian (morpho-)syntactic constructions described above on comprehension by Danish listeners, a sentence plausibility experiment was designed where informants had to judge the content of aural input as 'plausible' or 'implausible'. This approach was chosen over a more traditional sentence verification test (where the content would be deemed as either 'true' or 'untrue') as it was considered impossible to design enough sentence types for the experiment that are universally true at the same time as being formulated within the syntactic frameworks discussed above.

3.3 The Participants

The participants in this investigation were all native Danish speakers from the Jylland (N=45) and Copenhagen (N=8) areas. In the chi-square analysis of correct answers there were no significant differences between speakers from these two regions in the sample. The informants are therefore all included together in the analysis. The subjects ranged between 21 and 50 years old. Twenty-five informants were female and 28 were male. For the analysis of reaction times (discussed in 3.2), a subsample of 41 right-handed informants was used.

3.4 The Stimulus Material

Forty sentences consisting of cognate words only were constructed for the experiment: eight sentences for each syntactic construction type. Four of these eight had a semantic content that could be deemed plausible while the other four were semantically implausible. The distribution of the sentences across construction types and plausibility are given in table 1.

Table 1 The distribution of test sentences across construction type and plausibility conditions

	Particle Placement	Infinitival and negator placement	Possessive Type	Double definiteness	Possessive Placement
Plausible	4	4	4	4	4
Implausible	4	4	4	4	4
Total	8	8	8	8	8

Certain sentences in the corpus consisted of semantically related words, as ‘only child’ and ‘sister’ as above or ‘doctors’ and ‘patients’ in the sentence ‘doctors treat their patients’. Semantically related concepts like these were found both in implausible as well as plausible conditions. Some sentences did not have related concepts, but could still be deemed plausible or implausible from the context. An example is the sentence ‘A man prints out an article’, a sentence with plausible content. Only sentences that were deemed semantically unambiguous in a pre-test with a small group of native speakers of Danish and Norwegian were included in the experiment. All 40 statements in the experiment were articulated in the present tense. None of the content words in the stimuli are used more than twice. See appendix A for all test sentences used in the experiment.

Stimulus recording

The 40 sentences described above were recorded in four different conditions: with Danish phonology and Danish morphosyntax (condition A), with Danish phonology but Norwegian morphosyntax (condition B), with Norwegian phonology but Danish morphosyntax (condition C), and with Norwegian phonology and Norwegian morphosyntax (condition D). A bilingual speaker was chosen to produce the stimuli to control for differences in personal voice characteristics between the conditions. All sentences consisted of morphemes that are available in Danish as well as Norwegian. The bilingual speaker was therefore presented with an unnatural, yet fairly unproblematic task when producing the sentences in different conditions.

The bilingual speaker was born in the Aarhus area in Denmark but had lived in Oslo, Norway, from the age of 8. Two ‘voice parades’ were designed to test how native the bilingual speaker sounded to Norwegian and Danish listeners. For both voice parades text passages were recorded and presented to listeners through a web site. Listeners were asked to state to which degree the recordings they heard sounded foreign. In the Danish voice parade, four recordings of the same text passage were presented to listeners. One of these recordings was made by the bilingual speaker. The three others were produced by native Danish speakers from Aarhus and Aalborg, i.e. the same geographical area that the bilingual hailed from. The Norwegian voice parade consisted of five recordings of the same text passage. The distracter recordings in the Norwegian version were all recorded by speakers from Oslo. In neither test was the bilingual speaker rated significantly less native sounding than the other speakers. In both tests, one, or more, recordings of native speakers scored worse on a perceived native-ness scale than the

bilingual speaker. We claim therefore that all stimuli recorded for the experiment sound natively Norwegian and Danish, phonetically speaking.

Running the Experiment

Our experiment was designed and conducted using E-prime, version 2.08.22. Informants were first presented with a short demonstration task where they heard two sentences with Danish phonology and two sentences with Norwegian phonology. The syntactic structure of these test sentences were both grammatical in Norwegian and Danish, and sentences consisted of cognate words only.

In both the demonstration task, as well as the actual experimental setting, listeners were instructed to listen to sentences and evaluate their content as either ‘plausible’ or ‘implausible’. Subjects wore headphones and heard the stimuli while an hour glass and the two alternative answers ‘implausible’ and ‘plausible’ were displayed on the computer screen in front of them. Two keys on the keyboard, M and Z on a Standard Dutch Qwerty Keyboard, were marked with a green and a red sticker, respectively. The green key (to the right) represented a plausible reply, and the red (to the left) represented an implausible reply. All informants were informed explicitly, by the experimenter, that they should respond as quickly and correctly to the stimuli as possible by pressing the appropriate key on the keyboard.

The recorded sentences (160 in total) were not all presented to the same informants. A crossed design was used, such that each informant heard the 20 plausible and 20 implausible sentences divided equally over the four phonology-by-morphosyntax conditions. Listeners heard the same sentence in just one version and heard an equal number of sentences (10) from every morphosyntax-by-phonology condition. Correct responses as well as decision times were collected from the informants. All informants participated in the experiment on the same laptop computer.

3 RESULTS

A number of sentences with Danish phonology and Danish morphosyntax were judged incorrectly by the Danish listeners. No listener replied correctly to all stimuli. The errors the listeners made in each condition are presented in table 2.

Table 2 N of correct replies to stimuli in the four conditions, total N responses: 2160

Response	DaPhon DaSyn (A)	DaPhon NoSyn (B)	NoPhon DaSyn (C)	NoPhon NoSyn (D)
Error	40 (7%)	49 (9%)	72 (13%)	104 (19%)
Correct	500 (93%)	491 (91%)	468 (87%)	436 (81%)
Total	540	540	540	540

As expected, the Danish listeners had fewest difficulties to comprehend sentences with Danish phonology and morphosyntax (condition A) and most difficulties to comprehend sentences with Norwegian phonology and morphosyntax (condition D). They responded correctly to 93% of the sentences in condition A, but only on 81% of the sentences in condition D. The data also suggests that sentences with native phonology and foreign morphosyntax (condition B) are easier to comprehend than sentences with foreign phonology and native morphosyntax (condition C). Ninety-one percent of the sentences in condition B are decoded, but only 87% of the sentences in condition C.

Certain sentence types were judged incorrectly in condition A by a number of informants. Some sentences were judged incorrectly only by one informant in this condition, and these are subsequently kept in the analysis. The remaining eight sentences that were judged incorrectly at least twice in the native phonology and native morphosyntax condition are excluded from the remainder of the analysis as we believe that these sentences were too problematic, semantically speaking, for the experimental design. For an overview of all the sentence types that were judged incorrectly, see appendix B. The percentages of incorrect answers for the remaining 32 sentences are presented in figure 1.

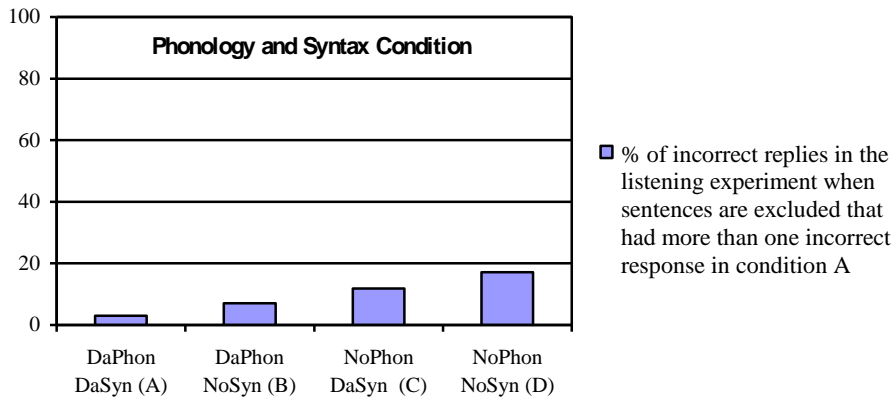


Figure 1 Percent incorrect judgments of sentences in four different phonology -by- morphosyntax conditions.

As the percentages of correct replies were not distributed normally, but all ranged from 41.3-100% they were transformed using a rationalised arcsine transform (Studebaker, 1985) to make them more suitable for statistical purposes. A 2 x 2 repeated measures ANOVA was conducted on the transformed data. The effect of phonology and morphosyntax were both reported as significant at $p < .05$. There was a significant effect of phonology on the correctness results $F(1) = 44.203$ and of morphosyntax on the correctness results $F(1) = 10.203$.

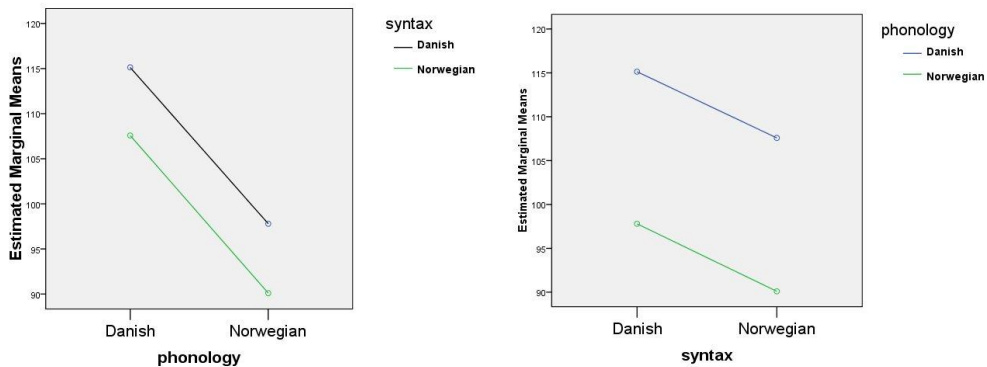


Figure 2 The effect of phonology (left) and syntax (right) on correctness results for sentences of the two morphosyntax types. As can be seen from figure 2 the introduction of a foreign phonology results in a more dramatic decline in comprehension scores than the introduction of a foreign morphosyntax does (note that the y-axis reflect the transformed correctness results).

4 CONCLUSION

The results of the current study have shown there is an effect of Norwegian morphosyntax and phonology on Danes' comprehension of sentences. When listening to sentences made up solely of cognate words and native morphemes Danish listeners do not comprehend sentences with Norwegian morphosyntactic constructions as well as sentences with Danish morphosyntactic constructions. Likewise, a Norwegian phonology impedes sentence comprehension for Danish listeners. Although they both play a role, the influence of a foreign phonology on comprehension appears to have a more dramatic effect on comprehension than the influence of foreign morphosyntax. It was hypothesised that comprehension would be slowest for Danish listeners when hearing sentences with phonology and morphosyntax from the closely related language Norwegian.

The findings of this study are in line with previous literature on sentence comprehension in a closely related variety. The finding that phonological differences play a larger role for comprehension than syntactic differences do was indicated by previous studies such as Van Heuven and de Vries (1981), who found that word order, morphological features as well as lexical choice affect comprehension of spoken Dutch. They also concluded that the role of phonic factors was much larger as that of non-phonic factors. One of the reasons why the phonological differences have a larger effect on intelligibility in our study could be the fact that there are simply more phonological differences in the stimuli than there are morpho-syntactic differences. Future studies should investigate this in more detail and determine whether it is also the case that phonological differences play a large role if there is only a smaller number of them. It could also be the case that certain phonological differences are more difficult to process than others. If

one were to control for prosodic and segmental features individually, one could investigate whether one of the two factors have a larger effect on intelligibility. This is also a question that researchers can investigate in more detail in the future.

This current study has not investigated the influence of lexical choice on comprehension, as was done by van Heuven and de Vries (1981), but this factor could be another important constraint on sentence comprehension and can be investigated more in depth in future studies, especially within Scandinavia. It could be that the usage of a word that is a cognate, but that does not entirely overlap semantically in the two languages, also provides difficulties for comprehension of a closely related language. Word frequency and complete semantic overlap was not controlled for in the stimuli for the current experiment. Neither has the current study looked at the relative influence on comprehension of different types of lexical content in the stimuli. It could be that a significant difference in comprehension can be found between sentences that have semantically related concepts, such as 'doctors treat their patients' and sentences with content words that are less closely connected semantically, like 'a man prints an article out'. It could be that morphosyntax plays a larger role for sentence comprehension when the lexical content of sentences is not semantically related, and future research could focus more on this possibility.

Overall, the results presented in this article indicate that there is an effect of morphosyntactic differences on the mutual intelligibility within the Scandinavian linguistic area. The finding that phonology plays a large role for comprehension is in line with previous studies of speech intelligibility in Scandinavia (e.g. Gooskens, 2007; Kürschner *et al.* 2008). The effect of morphosyntactic differences must not be disregarded in future studies, however. The effect of these types of differences is likely to be found also for the other language pairs in the area (Swedish and Norwegian, and Swedish and Danish). Syntactic differences also, presumably, play a role for Norwegians comprehension of Danish, although this was not investigated empirically in this study. What is more one could deduce that syntactic differences also play a role for intelligibility of a different *dialect*. The fact that syntactic differences influence our ability to comprehend sentences should have implications for teaching in schools where children are expected to understand a standard language but have a non-standard native syntax themselves. Awareness of which individual linguistic factors that determine intelligibility is important then not only to improve communication between speakers of closely related languages, but also in other situations of receptive bilingualism and in situations of (language) learning and teaching.

REFERENCES

- Bezooijen, R, van and C. Gooskens. 2007. Interlingual text comprehension: linguistic and extralinguistic determinants. In: J. D. ten Thije and L. Zeevaert (eds.). Receptive Multilingualism and intercultural communication: Linguistic analyses, language policies and didactic concepts. Hamburger Studies in Multilingualism. Amsterdam: Benjamins, 249-264.
- Blau, Eileen K. 1990. The Effect of Syntax, Speed, and Pauses on Listening Comprehension TESOL Quarterly, 24, 4, 746-753

- Braunmüller, K. and L. Zeevaert. 2001. Semikommunikation, rezeptive Mehrsprachigkeit und verwandte Phänomene. Eine bibliographische Bestandsaufnahme. *Arbeiten zur Mehrsprachigkeit – Folge B*, nr. 19. Hamburg University.
- Brøndsted, M. 1967. *Dansk for nordmænd*. H. Aschehoug & Co.
- Christensen, R. Z. 2007. *Dansk for svensktalende*. Lund: Studentlitteratur.
- Delsing, L.-O. and K. Lundin Åkesson. 2005. Håller språket ihop Norden? En forskningsrapport om ungdomars förståelse av danska, svenska och norska (TemaNord 573). Copenhagen: Nordisk ministerråd.
- Fodor, J. D. and L. Frazier. 1980. Is the human sentence parsing mechanism an ATN? *Cognition* 8. 417-459.
- Frazier, L. 2002 [1987]. Sentence processing. In: G. Altmann, *Psycholinguistics: critical concepts in psychology*, Volume 3, Routledge 2002.
- Gooskens, C. 2007. The contribution of linguistic factors to the intelligibility of closely related languages. *Journal of Multilingual and Multicultural Development*, 28 (6). 445-467.
- Haugen, E. 1966. Semicommunication: The language gap in Scandinavia. *Sociological Inquiry* 36. 280-297
- Heuven, V.J.J.P. van and J.W. de Vries. 1981. Begrijpelijkheid van buitenlanders: de rol van fonische versus niet fonische factoren. *Forum der Letteren*, 22. 309 - 320.
- Holmberg, A. and C. Platzack. 2005. The Scandinavian languages. In: *The Oxford Handbook of Comparative Syntax*, edited by G. Cinque and R. Kayne. Oxford University Press, Oxford. 420–458.
- Kürschner, S. C. Gooskens and R. van Bezooijen. 2008. Linguistic determinants of the intelligibility of Swedish words among Danes. *International Journal of Humanities and Arts Computing*, 2. 83–100.
- Ridell, K. 2008. *Dansk-svenska samtal i praktiken*. Uppsala University.
- ScanDiaSyn. 2010. *Scandinavian Dialect Network*. Website <http://www.tekstlab.uio.no/nota/scandiasyn> accessed 07.05.2010
- Sekerina, I. A. 2003. Scrambling and processing: Dependencies, complexity, and constraints. In: S. Karimi (Ed.), *Word order and scrambling*. Malden, MA: Blackwell Publishing. 301–324.
- Speidel, G.E., R.G. Tharp, and L. Kobayashi. 1985. Is there a comprehension problem for children who speak nonstandard English? A study of children with Hawaiian-English backgrounds. *Applied Psycholinguistics*, 6. 83-96.
- Teleman, U. 2008. *Svenska för danskar*. Roskilde: Roskilde universitetsforlag.

APPENDIX A: ALL SENTENCES

Training sentences

Plausible

Drager finnes i eventyr

'Dragons exist in fairy tales'

Et hus er større enn en bil

'A house is larger than a car'

Implausible

Russland er Europas minste land
'Russia is the smallest country in Europe'
En fisk leser sin avis
'A fish reads his newspaper'

Test sentences

Plausible

En mann skriver ut en artikkel
'A man prints out an article'
En rådgiver ramser opp en liste
'A councillor regurgitates (*up) a list'
En politibetjent legger ned en penn
'A police officer puts down a pen'
En nabo finner frem en sykkel
'A neighbour takes out a bike'
Farmor er i familie med faren min
'Grandma (lit. Father's mum) is related to my father'
Du ser med øynene dine
'You see with your eyes'
En statsminister leder regjeringen sin
'A prime minister leads his government'
Ministerens sekretær skriver brevet hans
The minister's secretary writes his letter'
Lærere underviser sine elever
'Teachers teach their pupils'
Leger behandler sine pasienter
'Doctors treat their patients'
Gartnere vannet sine blomster
'Gardeners water their flowers'
Hunder følger gjerne sine eiere
'Dogs follow their owners gladly'
De fleste mennesker håper å ikke bli syke
'Most people hope to not become ill'
Konkurransedeltagere ønsker å ikke tape
'Competitors wish to not lose'
Flaggermus pleier å ikke pusse vinduer
'Bats usually do not clean windows'
Fotomodeller prøver å ikke bli tyke

'Photo models try to not become fat'
I kristendommen er bibelen den viktigste boken
'In Christianity the Bible is the most important book'
I Danmark er København den største byen
'In Denmark, Copenhagen is the largest city'
Februar er den korteste måneden
'February is the shortest month'
I England er engelsk det viktigste språket
'In England English is the most important language'

Implausible

Et bibliotek låner ut kaffe
'A library lends out coffee'
Et insekt leser av temperaturen
'An insect reads off the temperature'
En gris legger ut penger
'A pig advances money'
En elefant slår opp et ord
'An elephant looks up a word'
Et enebarn leker med søsteren sin
'An only child plays with its sister'
Du lukter med hendene dine
'You smell with your hands'
En katt snakker med moren din
'A cat speaks with your mother'
Jeg er eldre enn faren min
'I am older than my father'
Hester spiser sine hover
'Horses eat their hooves'
Kaniner selger sine haler
'Rabbits sell their tails'
Elektrikere drikker sine ledninger
'Electricians drink their cables'
Mekanikere koker sine biler
'Mechanics boil their cars'
Piloter håper å ikke lande trygt
'Pilots hope to not land safely'
Forfattere ønsker å ikke skrive
'Writers wish to not write'
Fugler pleier å ikke kunne fly

'Birds are usually not able to fly'
 Vektløftere prøver å ikke bli sterke
 'Weight lifters try to not become strong'
 Sukker er den sunneste maten
 'Sugar is the healthiest food'
 I Sverige er gress den sjeldneste planten
 'In Sweden grass is the rarest plant'
 Murstein er den mest dyrebare edelstenen
 'Bricks are the most valuable jewel'

APPENDIX B

Construction type	Sentence	Plausible?	N error responses
Particle placement	A police officer puts a pen down	Yes	6
	A councillor reads a list out loud	Yes	5
	A neighbour takes the bike out	Yes	4
	An insect reads (out) the temperature	No	2
	A pig makes an advance payment	No	1
Possessive placement	You smell with your hands	No	3
	I am older than my father	No	2
	An only child plays with its sister	No	1
	The minister's secretary writes his letters	Yes	1
Possessive type	Dogs like to follow their owners	Yes	1
	Doctors treat their patients	Yes	1
	Mechanics boil their cars	No	1
Negator and infinitival placement	Bats usually do not clean windows	Yes	3
	Photo models try not to become fat	Yes	2
	Participants in contests do not wish to lose	Yes	1
	Most people hope not to become ill	Yes	1
	Pilots hope not to land safely	No	1
	Weight lifters try not to become strong	No	1
	Birds usually do not fly	No	1
Demonstrative marking	February is the shortest month	Yes	1
	Bricks are the most valuable precious stone	No	1

Franka Kermer

Teaching English tenses from a cognitive perspective

1 INTRODUCTION

An aspect of language acquisition that has been found problematic in numerous English as a Second Language (ESL) studies (e.g. Brown 1973; Close 1962; Hakuta 1974; Kennedy 2002) is the learners' correct application of the English tense-aspect system to the corresponding context. These studies show that non-native English speakers from diverse language backgrounds face different challenges when acquiring the English tenses, particularly if the tense-aspect system in their mother tongue shows significant contrarities to the English one (Kennedy 2002: 4). According to Niemeier and Reif (2008: 328), German-speaking learners of English suffer from a lack of accurate use of the target language, with 'the English tense-aspect system [as] a major source of error.' Additionally, they tend to use wrongly, or rather overuse, the English polysemous progressive form (Rohde 2000). The misuse of English tenses can result from the differences between the German and English tense-aspect systems; in contrast to English, German has no grammaticalized progressive aspect, and the German *Perfekt* differs in meaning and form from the English *present perfect* (Engel 2004; Klein 1995). On the other hand, however, the difficulties may be ascribed to the kinds of materials, exercises, and examples that teachers currently use in the German ESL classroom. Therefore, traditional English tense teaching calls for a critical review and deserves to be viewed from new perspectives of language education.

The Cognitive Grammar (CG) theory with its inherently symbolic function provides major advantages for a meaningful understanding of the English tense-aspect system. The CG theory offers help in showing the conceptualisation and reasoning behind the linguistic realisation of the tense-aspect system in English, which can ensue in a productive grammar instruction in the ESL classroom (Taylor 2002).

In a broad sense, Langacker's (1982, 1987) CG theory is concerned with how language functions without the constraint of syntax and morphology, which means that it does not resort to the rules of sentence constructions or the analysis of word structures. Langacker's central idea of the CG theory is that nothing else than symbolic units are required to describe grammatical structures (Langacker 1991). A symbolic unit is defined as a bipolar unit, which means it consists of both a semantic pole (its meaning), and a phonological pole (its form) and thus serves as a conventional element linking meaning and form ([SEM] / [PHON]). Therefore, grammatical structures can be seen as inherently symbolic rather than being an abstract and autonomous part of a language. According to Gibbs (2005), humans selectively perceive happenings, states, and emotions of their environment and conceptualise them as the so-called concepts in their mind. A

concept is thus the most central and essential unit of the human mind (Langacker 2008). Human cognition and perception of the world are triggered by the dichotomy of significant versus trivial information, which means that we incline to select important events according to their vantage point (known as figure/ground organisation in psychology) (Zimbardo 1992).

In addition to its function of selecting information at a cognitive level, grounding is a technique that is used at sentence level. The process of grounding includes certain grammatical components such as articles, quantifiers, possessive pronouns, demonstratives, morphemes as well as modal and tense verbs (Radden and Dirven 2007). According to Langacker (2002), tense is a case of viewing arrangement, which means it is a tool used by the speaker to locate a situation in time from his/her viewpoint. Langacker's (1991) 'model of evolving reality' portrays the classification of a situation's reality status in terms of two main features. First, the application of modal and finite verb forms to a sentence (in a finite clause) primarily reveals information about the temporal status of an event, which can be either bounded or continuous. Second, the grammatical class TENSE marks the temporal location of an event, which can be located in the past, the present, or the future time (Evans & Green 2006: 617). Future situations, however, can be expressed by means of the modal verbs such as 'will', 'can', 'may', or 'wish', the semi-modal going-to, and the simple present tense, and hence, encompass events that are potential, unforeseeable or (un)likely to happen in the future. In addition to the CG theory, Fauconnier (1994, 1997) also provides an approach to grounding situations in time from a cognitive perspective. According to Fauconnier (1997: 72) and his theory about mental spaces, 'in moving (mentally) from one mental space to another, we try to keep track of the time shifts and epistemic shifts between the spaces in focus.' By means of tense and mood, we are able to keep track of time through discourse. To create a time space, we can use time adjuncts, or space builders, as yesterday, tomorrow, last year or in 1999, and morphological categories added to the verb (-s or -ed). A space builder can set up past, present or future Event spaces, and tense generally indicates the relation between times of events in different spaces (Fauconnier 1997: 73).

A similar approach was discussed by the German logician Hans Reichenbach (1947) in his 'theory of tense' in 1947, who argued that the meaning of tenses could be mentally grasped if they are described as a sequence of three time points. He developed, in contrast to traditional grammar, a tense system that involves three points of times: first, the present moment of speaking (S); second, the event time (E), which is the interval where an event occurs, and third, what Reichenbach termed, the reference time (R). The speaker can shift his/her viewpoint to a time that either precedes or follows the time at which an event occurs. Thus, anterior or posterior times, denoted by complex tenses, can be expressed (see Binnick 1991; Davidsen-Nielsen 1990; Hornstein 1990; Klein 1992; Michaelis 2006). In the simple tenses (e.g. simple present or simple past), the time points E and R are simultaneous, which means they coincide. For instance, the simple past is characterised by an event (E) that happens before the present moment of speaking (S), whereupon R is shifted to the past, simultaneously to the event's occurrence (E).

Langacker's CG theory and Fauconnier's mental space theory can be ideally transferred to Reichenbach's classification of English tenses. The speech time is part of the base space and the event time is part of the event space. The reference time allows the speaker to shift his/her

viewpoint to the past or future time, which means that the point from which a situation is viewed is not restricted to the present time anymore (see Table 1).

Table 1: Perfect and Prospective Tenses involving the Location of S (speech time), E (event time) and R (reference time) (adapted from Davidsen-Nielsen, 1990: 61)

Present Perfect	E	R, S
Past Perfect	E – R	S
Future Perfect		S – E – R
Present Prospective		S, R – E
Past Prospective	R – E	S
Future Prospective		S – R – E

One of the major advantages of Reichenbach’s system is the representation of the simple past versus the present perfect, which according to Dürich (2005) is a significant source of problems of German ESL learners. Based on Reichenbach’s approach, the speaker is capable of shifting his/her reference point to the past (simple past) or let it remain at the present moment of speaking (present perfect). If learners understand how a situation is mentally represented as a space-building construction by means of tense-aspect marking, they may be able to conceptualise and comprehend the temporal relation of situations (Archard & Niemeier 2004). Accordingly, the joining of Reichenbach’s theory with the tenets of a cognitive perspective of tenses seems worthwhile and helpful for ESL students.

2 COGNITIVE GRAMMAR IN ESL TEACHING

I have presented the basic theoretical framework of my study, which now ties in with the question: How can we adapt the tenets of Langacker’s CG theory, Fauconnier’s mental space theory, and Reichenbach’s logical approach to the ESL classroom in order to tackle the learner’s obstacles? Niemeier and Reif (2008), who tested the CG approach in an empirical study at a university in Germany (120 students), have conducted one of the remarkable (ongoing) investigations in this field. One group of English students were presented with traditional teaching materials, whereas the other group were taught with a cognitively oriented programme. My new teaching design avails of Langacker’s, Fauconnier’s and Reichenbach’s approaches in terms of the meaningful understanding of time, and, beyond that, combines the selected information with the methodological standards of ESL teaching.

The national core curriculum in Germany does not explicitly apply or mention a cognitive approach to grammar instruction in English lessons and studies on this realm are surprisingly rare. Nevertheless, the German curriculum endows ESL teachers with aims, or rather prerequisites, regarding the contents and methods that they are demanded to use in the classroom. My study investigates the teaching of English tenses from a cognitive perspective in secondary schools, which is in line with the formal requirements postulated in the curriculum. The preliminary results discussed in my MA thesis (Kermer 2009) show that pupils are able to comprehend the cognitive perspective of the English tenses in a teaching and learning environment that meets the formal standards. I presented materials and activities for practising

and broaden eighth-graders' knowledge of the past perfect and the present perfect. Based on my observation during pair work exercises, dialogues, classroom interactions as well as the evaluation of short written texts and interviews (teacher and pupils), it becomes clear that the ESL learners can benefit from the cognitive grammar instruction. The pupils found not only the mental representation and conceptualisation of the tenses useful, but also the activities and materials I put into practice (see 3). In order to specify and understand the new teaching design, however, we first have to glance at the formal requirements stated in the German core curriculum.

2.1 The formal expectations of German ESL learners' merits

Since the German curriculum for English in Lower Saxony (KC 2006) requires specific standards of competences, language teachers are challenged to afford a learning environment that contains communicative activities, continuous use of the target language and the acquisition of correct grammatical structures. The overall intent of the KC is a successful acquisition of English grammar, described as follows: 'Die Schülerinnen und Schüler verfügen in zunehmendem Maße über häufig verwendete grammatische Strukturen, lernen diese intentions- und situationsangemessen anzuwenden [*The pupils have an increasing command of commonly used grammatical structures and learn to use these appropriately in regard to their intentions and the situation.*]' (KC 2006: 20). At the end of class ten, the pupils are expected to understand and express activities and events located in the past, present and future time (KC 2006: 21). These statements clearly indicate that pupils predominantly need to understand and comprehend the English tense system to fulfil the formal requirements. The KC (2006: 8) ensures the successful realisation and application of these competences, if a lesson complies with the following prerequisites:

1. Der handlungs-, prozess- und ergebnisorientierte Unterricht erfolgt in der Zielsprache (Prinzip der funktionalen Einsprachigkeit). [*The action, process and outcome oriented teaching takes place in the target language (informed monolingualism).*]
2. Das Sprachhandeln im Englischunterricht findet in authentischen, bedeutungsvollen und herausfordernden Situationen statt. [*The use of language in English lessons takes place in authentic, meaningful and challenging situations.*]

As can be noticed, the formal requirements of ESL grammar teaching contain an authentic and meaningful learning environment, in which the learners achieve an intentional, almost native-like, use of grammatical structures. These grammatical structures include, among others, the expression of continuous, repetitive, instantaneous, past, future or timeless situations (KC 2006). In order to see whether the requirements are fulfilled, we should have a critical look at one of the devices used in the classroom: the schoolbook.

2.1 Past perfect and present perfect in German schoolbooks

I will now refer to English G 2000, one of the schoolbooks used in German schools, and discuss how the book fulfills the standards. On page 118–119, the book presents a revision of the present perfect, the simple past and the past perfect. The explanations entail rules and instructions in

order to repeat existing background knowledge. Example 1 (English G 2000: 118) includes an instruction on how to use the present perfect correctly.

Wie du weißt, benutzt man das present perfect, wenn man sagen will, dass etwas passiert ist, nicht wann. So kommt es auch häufig mit Adverbien der unbestimmten Zeit (*ever, just, never, already*) vor. [*As you know, the present perfect is used to express that something has happened, not when it happened. Thus, it is often used with indefinite time adverbials (ever, just, never, already).*]

As one can see, Example 1 is solely expressed in German and only the time adverbials are given in English. Furthermore, the pupils read the instruction out of context, and there is no reference to an authentic or challenging situation. The requirements stated in the KC (2006: 8–9) are not achieved and, consequently, Example 1 does not provide an appropriate platform for ESL learners to acquire an intentional and automatic use of proper grammatical structures. The similar set of problems applies to Example 2 (English G 2000: 118). Although the given explanation is written in English, it is lacking a larger or an authentic context. Additionally, the sentence refers to a person (Glen) with whom the pupils most likely do not have any association or relationship and thus the realisation of a comprehensible learning environment appears to be hindered.

Example 1

In der Vergangenheit [<i>In the past; in German</i>]	geschah noch davor: past perfect [<i>happened before the past; in German</i>]
Glen ate the dinner	he had bought on the way home.

The above description for the past perfect can be ascribed to the traditional grammarian Otto Jespersen (1927). Jespersen renders useful, applicable aids for studying the English language and, beyond that, files dozens of practical and adaptable uses and examples for grammatical constructions. Jespersen (1927: 262) declares that the past perfect indicates a time which occurs ‘before [the] past’; an account that is displayed in Example 1. Jespersen attempts to provide one theory for different grammatical cases and does not undertake, like Reichenbach, investigations where he thoroughly characterises single explanations for single grammatical structures. Nevertheless, it becomes apparent that Jespersen’s treatment of time largely coincides with Reichenbach’s time system (e.g. the representation of the past perfect; see 1). Hence, the basic intentions of Example 1 may be applicable at an initial stage in ESL teaching, but for an adequate and a sufficient use it would not seem appropriate.

This brief illustration leads to the assumption that hitherto methods of grammar explanations in that particular schoolbook may not account for an intentional, meaningful and challenging comprehension or use of the English tense system. An alternative concept of teaching the English tenses, specifically the present perfect and the past perfect, is presented in the proceeding chapter.

3 APPLYING COGNITIVE GRAMMAR IN ESL TEACHING

On the basis of the theoretical background specified in the previous chapters, the following text illustrates intentions and ideas for implementing CG based teaching facilities in English classrooms. The chapter is founded on my investigation, which I conducted within the framework of my MA thesis in a German secondary school in 2009. The investigation includes a practical application spectrum that consists of materials, activities and methods for teaching the simple past, present perfect and past perfect in an eighth grade. In order to organise lessons which are in line with the formal KC requirements, the lessons are based on ‘informed monolingualism’, the communicative approach and the idea of an ‘authentic, meaningful and challenging learning environment’. I will suggest possible teaching steps, whereupon I highlight the embedment of the cognitive concept of time in applicable examples.

The core for understanding sentences that include the perfect tenses is the comprehension of the three time points: the speech time (S), the event time (E), and the reference time (R). Therefore, Figure 2 is employed as a supporting illustration to present the interaction of S, E, and R when using the past perfect, the present perfect or the simple past. Teachers can pin, for example, Figure 2 to the blackboard or to a different, visible position in the classroom. However, it is important to design individual gaps for E, S and R in the figure and to emphasise their relation by means of, for instance, arrows.

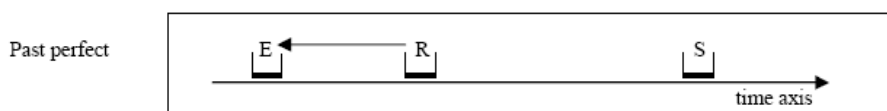


Figure 2: Representation of Past Perfect

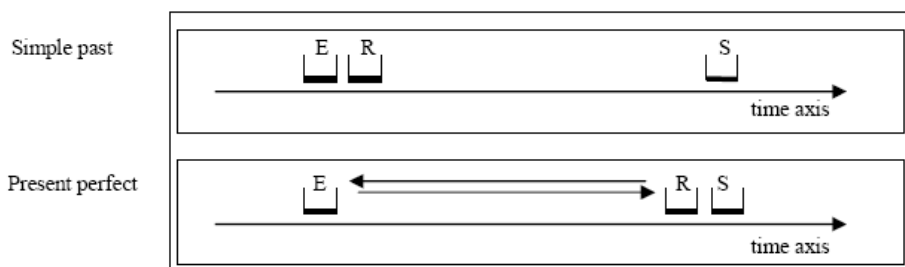


Figure 3: Representation of Simple Past and Present Perfect

To clarify the terms S, E, R, and their purpose accordingly, the teacher can prepare large, coloured posters as visual aids; marking each of them with the term and an example such as shown in Figure 4:

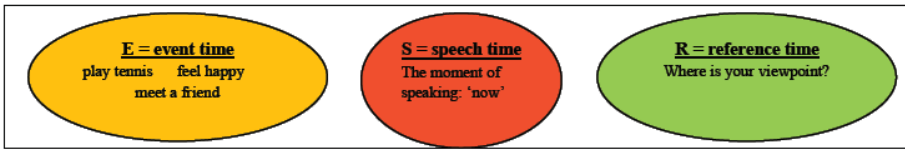


Figure 4: E (event time), S (speech time) and R (reference time) as visual aids for the classroom

In order to realise the notion of the three reference points in time in a meaningful and exemplary way, the teacher could ask two pupils to come in front of the classroom. The teacher has to decide what tense should be presented and discussed first. As the past perfect contains an explicit distinction between E and R, it seems appropriate, from my point of view, to begin with this tense. While S is marked with a red cross on the floor in the middle of the classroom, each of the pupils obtains a tag (poster) with either E or R on it. A sentence that includes the past perfect is written on the blackboard (e.g. 'My friend had left the classroom when I arrived.'). The assignment to the whole class could be as follows: Look at the sentence on the blackboard. Arrange both classmates (E and R) relative to S and discuss your decision in the class.

As the assignment entails the class' active participation, an action-oriented and communicative introduction phase can occur. After the pupils achieved this assignment, they should consider own past situations and experiences, which happened before R and thus require the past perfect. To bestow a correct application, the teacher might provide (R) for the pupils on the time line. During this activity the pupils deepen, apply, and practise the new information in a meaningful discourse. Furthermore, talking about their own experiences advances the pupils' motivation and leaves space for individual development and integration into the learning process. To ensure that the pupils understand and initialise the basic and fundamental notions of E and R, the teacher should place emphasis on this activity and allow extra time for it in the lesson.

As pointed out in the first chapter, the correct application of the English present perfect presents an elusive feature for German ESL learners. This can be attributed, on the one hand, to the distinctive meaning and use of the perfect tense in the German language. On the other hand, it can be ascribed to the misinterpretation of how to express a past situation appropriately with its actual meaning. To expose the distinguishing features between the simple past and the present perfect, the teacher can refer to Figure 3. The pupils should study the graphic cautiously and develop an idea or explanation for questions such as: What difference can you see between the two graphics? What do you think does the different location of R mean? How does the meaning of a sentence change if you locate R a) next to S or b) next to E?

The purpose of this exercise is to foster the pupils' ability of shifting their viewpoint according to a certain situation and, beyond that, understanding the consequences of it. The significant component here is that the teacher should underscore the simultaneous occurrence of E and R when using the simple past, and the simultaneous occurrence of S and R when using the present perfect. The former characterises past events (E) which are viewed (R) from the past, which means the speaker's focus is on the past event and time, detached from the present. In the latter case, the speaker's viewpoint (R) remains in the present moment of speaking and the past situation (E) is observed by a backward-looking stance. Using the present perfect in this case

enables the speaker to highlight his/her focus on the present time as well as on the current significance of the past event. This activity should preferably be performed in the whole class and it is advised to use the target language at any time and at any stage. In order to practise the pupils' correct understanding and application of the present perfect and the simple past, the teacher can conduct an activity that is based on pair work or small groups. The pupils receive a worksheet with sentences containing both tenses as well as different coloured cards for S, E and R. The assignment for the class could be as follows: Arrange the cards according to the given sentences and discuss your decision with your partner! The teacher may perform as a supervisor or assistant and fosters the pupils' learning process and comprehension. Although the pupils are expected to communicate in the target language, the focus should be on the accurate application of the tenses to the exercise.

Another activity the teacher can conduct in the classroom represents a resuming, summarising and repetitive exercise. The teacher projects sentences that contain the simple past, past perfect and present perfect onto the blackboard. If applicable, the teacher uses sentences which were used by the pupils in previous exercises. Thus, the pupils' personality is involved in the learning process which consequently leads to an authentic learning environment. S, the speech time, is still marked with a red cross in the classroom. The pupils try to locate (E) as well as (R) to the presented sentences accordingly. Thereby, they are obliged to analyse and discuss their suggestions and, additionally, justify their decisions in a collective. The teacher should guide the class and support their critical analysis whenever this is appropriate and useful.

The introduced examples of classroom activities and exercises were emerged at the early stage of my investigation. Applied linguists and teachers may consider the proposed materials or, at least, reflect on the cognitively based concept for teaching the English tenses. Additionally, it has to be mentioned that I developed most activities due to the formal requirements postulated in the KC as well as to the social environment of that particular school class. However, the new teaching concept will be amplified during the next stages of my PhD research and its profitableness will be tested in other experiments.

4 CONCLUSION

Langacker's CG theory is the basis for understanding the abstract ideas and assumptions of grounding a situation in time. Grounding a situation in time is an essential part for a successful realisation of any communication (either written or spoken). Langacker (2008) focuses on the relation of CG and social interaction; moreover, he outlines that language should be analysed as a cognitive process. Since the connection of these two phenomena is established and presented in many books, the analysis of grammatical structures and their impact on human interaction can be examined to a great extent. As successful communications form a central quality of humans' life, certain elements, such as the right use of time, need to be acquired. English language learners are possibly presented with rather uncommunicative and meaningless discourse materials and activities in which they are expected to learn the English tense system. The CG model and the involving assumptions on grounding a situation in time offer new perspectives in the field of ESL teaching. The teaching of tenses should neither involve simple

and unchallenging tasks nor meaningless learning situations. New analysis and teaching methods could be anchored within the field of Second Language acquisition and Langacker's CG theory might be the basis for a deeper understanding how grammar and social interaction are related.

My forthcoming PhD research includes different steps. The first part of the study is the observation and critical analysis of traditional English grammar teaching in German secondary schools. During this period, the central view is on the entire English tense/aspect system and I will attend and observe different ESL classrooms over a longer period. The gathered data will contribute to an elaborated understanding of traditional grammar instruction in German ESL classrooms. Based on these data as well as on the prerequisites postulated in the KC, I design a CG based concept which consists of materials, exercises, classroom activities and instructions for teaching certain English tenses (e.g. present perfect, simple past). The materials that I developed during my MA provide the basis for the concept. The experiment is planned to be executed in spring 2012. One group will be taught with the traditional concept of teaching English tenses, whereas the other group is presented with the CG based approach. Based on written tests, I will assess and evaluate the pupils' performance and critically reflect and compare both teaching designs.

I aim to offer a new syllabus of ESL grammar teaching in which the instructions lead to a successful and an accurate comprehension of the English tense system. If this attempt is successful, meaningful, effective and motivating ESL grammar teaching may possibly result.

REFERENCES

- Achard, M. & Niemeier, S. (eds). (2004). *Cognitive Linguistics, Second language Acquisition, and Foreign Language Teaching*. Berlin: Walter de Gruyter.
- Brown, R. (1973). *A first language*. Cambridge, Mass.: Harvard University Press.
- Binnick, R. I. (1991). *Time and the Verb. A Guide to Tense and Aspect*. New York: Oxford University Press.
- Close, R.A. (1962). *English as a Foreign Language. Grammar and Syntax for Teachers and advanced Students*. London: Allen & Unwin.
- Davidsen-Nielsen, N. (1990). *Tense and Mood in English. A Comparison with Danish*. Berlin & New York: Mouton de Gruyter.
- Dürich, K. (2005). *The Acquisition of the English tense and Aspect Aystem by German Adult Learners*. Master thesis (unpublished). Technische Universität Chemnitz.
- Engel, U. (2004). *Deutsche Grammatik*. München: Iudicium.
- Evans, V. & Green, M. (2006). *Cognitive Linguistics. An Introduction*. Edinburgh: Edinburgh University Press.
- Fauconnier, G. (1994). *Mental Spaces: Aspects of Meaning Construction in Natural Language*. Cambridge: Cambridge University Press.
- Fauconnier, G. (1997). *Mappings in Thought and Language*. Cambridge: Cambridge University Press.

- Fauconnier, G. & Sweetser, E. (1996). *Spaces, Worlds, and Grammar*. Chicago: University of Chicago Press.
- Gibbs, Jr., W. (2005). *Embodiment and Cognitive Science*. United States of America: Cambridge University Press.
- Hornstein, N. (1990). *As Time Goes by: tense and universal grammar*. Massachusetts: MIT Press.
- Jespersen, O. (1927). *A Modern English Grammar on Historical Principles*. Heidelberg : Carl Winter.
- Kennedy, G. (2002). *Structure and Meaning in English: A guide for Teachers*. Harlow: Pearson Longman.
- Kermer, F. (2009). *Cognitive Grammar: Grounding Situations in Time*. Master thesis (unpublished). University of Hildesheim.
- Klein, W. (1992). *Tempus, Aspekt und Zeitadverbien*. [Tempus, Aspekt and temporal adverbs; in German]. In *Kognitionsforschung*, 2, 107–118.
- Klein, W. (1995). 'The Acquisition of German'. In Dietrich, Klein, Noyau (1995), *The Acquisition of Temporality in a Second Language*. Amsterdam: John Benjamins, 71–115.
- Langacker, R. W. (1982). *Remarks on English Aspect*. In Paul J. Hopper (ed.), *Tense-Aspect: Between Semantics & Pragmatics*, 265–304. Amsterdam & Philadelphia: John Benjamins.
- Langacker, R. W. (1987). *Foundations of Cognitive Grammar. Volume I. Theoretical Prerequisites*. Stanford: Stanford University Press.
- Langacker, R. W. (1991). *The Cognitive Basis of Grammar*. Berlin & New York: Mouton de Gruyter.
- Langacker, R. W. (2000). *Grammar and Conceptualization*. Berlin: Mouton de Gruyter.
- Langacker, R. W. (2002). *The Cognitive Basis of Grammar (2nd ed.)*. Berlin: Mouton de Gruyter.
- Langacker, R. W. (2008). *Cognitive Grammar as a Basis for Language Instruction*. In Peter Robinson & Nick C. Ellis (eds), *Handbook of Cognitive Linguistics and Second Language Acquisition*, 66–88. New York & London: Routledge.
- Michaelis, L. A. (2006). *Time and Tense*. In Aarts, B. & McMahon, A. (eds), *The Handbook of English Linguistics*. 220–244. Oxford: Blackwell.
- Niedersächsisches Kultusministerium. (2006). *Kerncurriculum für die Realschule Schuljahrgänge 5–10. Englisch. Niedersachsen*. Unidruck Hannover.
- Niemeier, S. & Reif, M. (2008). *Making progress simpler? Applying cognitive grammar to tense-aspect teaching in the German EFL classroom*. In de Knop, S. & de Rycker, T. (eds), *Cognitive Approaches to Pedagogical Grammar: A Volume in Honour of Rene Dirven*. 335–365. Berlin: Walter de Gruyter.
- Pütz, M., Niemeier, S. & Dirven, R. (2001). *Applied Cognitive Linguistics: Theory and language acquisition*. Berlin: Mouton de Gruyter.
- Radden, G. & Dirven, R. (2007). *Cognitive English Grammar*. Amsterdam & Philadelphia: Benjamins.
- Reichenbach, H. (1947). *Elements of Symbolic Logic*. New York: Macmillan.
- Taylor, J. R. (2002). *Cognitive Grammar*. Oxford: Oxford University Press.
- Zimbardo, P. G. (1992). *Psychology and life*. New York: Harper Collins Publishing.

Anna-Kaisa Penttinen

Finnish and English code-switching and borrowing in a Finland Swedish discussion program

1 INTRODUCTION

Nowadays, code-switching and borrowing have become important research areas, as the world is becoming more global all the time. When different cultures interact with each other, new phenomena and new words and expressions spread constantly. The constant spread of English has also caused that most research about code-switching and borrowing has concentrated on English. For example, Hansen & Lund (1994) and Jarvad (1995) have studied English influence on Danish; Ljung (1985), Chrystal (1988) and Sharp (2001) on Swedish; and Sandøy (2000) and Johansson & Graedler (2002) on Norwegian. Most of them have observed written language. The only exception is Sharp's (2001) study, which concentrates on English in spoken Swedish. Also a larger project called *Moderne importord i språka i Norden* (Modern loanwords in the languages of the Nordic countries) has been led by Helge Sandøy at the University of Bergen in Norway. It seems to be one of the few studies about loanwords in Finland Swedish context. As Swedish functions as a minority language in Finland in contact with Finnish, it creates an interesting situation, where Swedish sometimes borrows words from Finnish. Furthermore, English functions as the third language in the Finland Swedish environment.

This article examines Finnish and English code-switching and borrowing in Finland Swedish speech. It depicts if Finland Swedish is endangered by the loan expressions or if it can retain its original expressions as a minority language in Finland despite the influence of constantly spreading English. The first aim of this study is to show that Finnish and English expressions appear in Swedish-spoken discussion. The second aim is to categorize the Finnish and English expressions into code-switching and borrowing. The third aim is to examine how the speakers mark the expressions by metalinguistic and paralinguistic means.

2 MATERIAL AND METHODOLOGY

The research material consists of eight episodes of a Finland Swedish discussion TV-program called *Hetimiten* (2005). The results have been taken from my master's thesis *Om inlåning av finska och engelska ord samt kodväxling i samtalsprogrammet Ögonaböj* (Penttinen 2008). The database consists of eight episodes of the discussion program *Hetimiten*. There are four participants in the program: Stan Saanila, André Wickström, Heli Roiha and Thomas Perret. Saanila and Wickström are Swedish-speaking Finns, Roiha is originally Finnish-speaking but has lived in Sweden, and Perret is Swedish. Conversation in *Hetimiten* is very casual and everyday-like, and the topics concern occasions in the discussants' personal lives or current events in the world. The eight episodes were broadcasted on TV 10.10.2005, 17.10.2005, 24.10.2005, 31.10.2005, 7.11.2005, 14.11.2005, 21.11.2005 and 28.11.2005. Each episode lasts about 35 minutes. The analysis began by watching the episodes and simultaneously gathering all the Finnish and English expressions that arose in the Finland Swedish discussion. The phrases that included Finnish or English were transcribed into written form. After that, the functions of the expressions and how the speakers marked them e.g. through hesitations, laughter or gestures, were also documented.

After the data was compiled in written form, the Finnish and English expressions in the Finland Swedish speech were categorized using the framework of Lauttamus' (1990) and Knookala's (1982) studies. The benefit of Lauttamus' model is the four partitions: code-changes, code-mixes, nonce loans and integrated loans, which are one way to look at what different kinds of Finnish and English expressions appear in the data. The fifth category for quotations was formed with help of Knookala's research. Furthermore, the functions of the expressions and how the speakers mark them in their speech were analyzed using the frameworks of Löfström's (1995), Sharp's (2001), Lampinen's (2002a & 2002b) and Hiidenmaa's (2003) studies. This was done by looking at the metalinguistic and paralinguistic means which the speakers used in their speech.

3 ASPECTS OF FINNISH AND ENGLISH INFLUENCE ON FINLAND SWEDISH

About 5.5 % of the Finnish population is Swedish-speaking Finns. Swedish is one of the two national languages in Finland. It means that Finland is a bilingual country, and the Constitution of Finland guarantees that it is possible to communicate with the state authorities either by Finnish or Swedish. Also the social services, schooling and information as well as laws, regulations and important documents have to be equally available on both Finnish and Swedish. Swedish is a compulsory subject for Finnish-speaking as well as Finnish is compulsory for Swedish-speaking pupils in the last three years of the primary education. The situation in Finland leads to an aspect which makes Swedish spoken in Finland and Swedish spoken in Sweden clearly different; that is the loanwords from the Finnish language. The loans can be divided into direct loans and translation loans. As Beijar et al. (1998: 61-62) exemplify, a direct loan is e.g. Finnish *juttu* 'thing' instead of Swedish *grej* 'thing'. Finnish functions sometimes as slang in Finland Swedish. Nyholm (1978: 49) remarks that e.g. Swedish-speaking Finnish students use Finnish in such situations where they want to express toughness.

As Hiidenmaa (2003) states, today English language is a major lender all over the world, and in some branches terminology is mostly or only in English. The English language is being used more and more also in Finland as the time goes by. Battarbee (2002: 262) states that during the 1980s and 1990s Finnish society underwent an extremely rapid change into an urban, technological, postmodern society, which supported the spread of English loanwords. Although English is not a compulsory subject in Finnish schools, practically all the students learn English either as a compulsory or as an optional in Finland. Hiidenmaa (2003: 93-96) presents us with five different situations where English loanwords can occur in the Finnish language: 1) on their own (e.g. in employment announcements) 2) as a slogan in a text (e.g. in advertisements where the authentic information is in Finnish, but where there is a loose expression at the end of it, e.g.: “– For the way you are. ”), 3) in product names, 4) as quotations in Finnish phrase structure (e.g. *controlling-tehtävät* ‘controlling-tasks’), and 5) integrated into Finnish (e.g. *nämä ovat tyypillisii spinoffeja* ‘these are typical spin-offs’).

4 ABOUT CODE-SWITCHING AND BORROWING

In this article *code-switching* includes *code-changes*, *code-mixes* and *quotations*. The first two of the categories have been taken from Lauttamus (1990) study, and the third has been formed with help of Knookala’s (1982) study. Lauttamus (1990: 49-50) states that the guest language grammar operates on code-changes and code-mixes, and furthermore on quotations. Code-changes, code-mixes and quotations are non-smooth transitions from the host language to the guest language of the switched elements. According to Lauttamus (1990: 15-16) study, code-changes are typically multi-word fragments (usually sentences or whole phrases), which function lexically, syntactically, morphologically and phonologically according to the guest language. An example of a code-change from the data is: “*Nu sitter de där och tittar på Ögonaböj och äter. ‘Antaa olla! Choko går dit!’*” (‘Now they are sitting there watching Hetimiten and eating. ‘Let it be! Chocolate goes there!’’) Lauttamus (1990: 15-16) states that contrary to code-changes, code-mixes appear typically as one word or compound word and are usually nouns. A code-mix has not adapted syntactically or morphologically into the host language. An example of a code-mix is: “*Och sen är det den advanced, vilka är att man sku bara äta saker som är blå.*” (‘And then there’s that advanced, which means that we should eat only blue things.’) The category of quotations includes such code-switches which appear when the speaker quotes something that someone else has earlier said in different language. An example of a quotation from the data is: “*För vad som hon frågade, sa de: ‘Hilleri!’*” (‘Whatever she asked, they said: ‘Fitch!’’)

According to Lauttamus categorization (1990: 20) *borrowing* is divided into two sub-categories: *nonce loans* and *integrated loans*. Lauttamus (1990: 20) describes nonce loans as morphologically and syntactically, but not phonologically, adapted into the host language. In addition, integrated loans are also phonologically adapted. Borrowed elements function fluently in the speech. *Finlandssvensk ordbok* (2008) and *Svenska Akademiens ordlista över svenska språket* (2006) have been used in order to find out which loans are already integrated and which are not. In this study nonce loans can be such Finnish or English expressions that are e.g. conjugated according to Swedish syntax. Integrated loans are such expressions that are adapted into

Swedish at all the levels mentioned above and are also listed in the dictionaries. Sometimes the adaptation is not clear at all levels, e.g. the phonology might be undeniably English, but the word has been listed in the dictionary and thus has to be categorized as an integrated loan. An example where a nonce loan and an integrated loan appear in a same sentence is: "*Riksdagen borde koncentrera sig på att stifta lagar mot hackare och krackare i stället för att...*" ('The government should concentrate on passing laws against hackers and crackers instead of...'). The integrated loan *hackare* is already listed in the dictionary, but the nonce loan *krackare* has not yet been integrated.

5 HOW THE SPEAKERS MARK CODE-SWITCHING AND BORROWING

Sharp (2001: 157-158) has studied how speakers mark their code-switching, and categorizes these means into metalinguistic and paralinguistic means. Metalinguistic comments are e.g. translations, synonyms, clarifications, pauses and repetitions. Paralinguistic comments are laugh, giggling, changes in voice, gestures, pauses and hesitating. The metalinguistic markers highlight the foreign expressions, and thus the item becomes marked. In this article the metalinguistic markers include hesitation words, translations and clarifications. Here is an example where a hesitation word *liksom* ('like') and a Swedish translation *tokig* ('crazy') appear in a same sentence: "*Terry Gilliam som ska vara liksom crazy och tokig...*" ('Terry Gilliam who should be like crazy and mad...') An example of a clarification after a loanword is: "*Du har alltså inga clown-auditions? Alltså, nå visa vad du kan...*". ('So you don't have any clown-auditions? Like, show me now what you can do...')

Paralinguistic markers include changes in voice, laughter and gestures. An example where all these means are used together is: "*Och sen sprang han runtom lägenheten och skrek: 'Mä oon Heli-man! Mä oon Heli-man!'*" ('And then he ran around the apartment and screamed: 'I am Heli-man! I am Heli-man!') The speaker imitates voice of a small boy, sways her hands like she would be running and laughs when the quotation comes up. Also pauses build up a category. Pauses can be divided into filled and silent pauses. An example of pauses is: "*...jag tänker att dom borde börja köra med såndär telefon... [pause] vad heter det... [pause] ööö... [pause] marketing...*" ('...I think that they should begin to promote with that kind of telephone... [pause] what is it called... [pause] erm... [pause] marketing...') A filled pause has been blended in with the hesitation words *vad heter det* ('what is it called'), and also a filled pause *ööö* ('erm') comes up in the same sentence.

6 CODE-SWITCHING AND BORROWING IN HETIMITEN

The amount and percentage of the Finnish and English expressions in the research material is presented in the Table 1 below. Altogether, 223 Finnish and English items appeared in the data. English expressions account for 165 and Finnish expressions for 58 items, which means that 74.0 % of the material are English and 26.0 % are Finnish expressions. Consequently, the results show that English clearly dominates over Finnish. The amounts of items in the different categories will

be presented more distinctively below. Also the different categories show that it is more common to use English than Finnish in Finland Swedish speech. The speakers are marked with capital letters in the beginning of every example. S is for Stan Saanila, A for André Wickström, H for Heli Roiha and T for Thomas Perret. The date after every example refers to the broadcasting date, when the expression appeared in the program.

Code-changes account for 12 Finnish and 69 English expressions. It means that Finnish code-changes cover 20.7 % of all the Finnish-spoken material, and the English code-changes cover 41.8 % of all the English-spoken material.

Table 1: The absolute amounts and percentages of code-switching and borrowing in the data.

	CODE-SWITCHING						BORROWING				total	
	code-changes		code-mixes		quotations		nonce loans		integrated loans		amount	%
	amount	%	amount	%	amount	%	amount	%	amount	%		
Finnish expressions	12	20.7	5	8.6	15	25.9	16	27.6	10	17.2	58	100
English expressions	69	41.8	29	17.6	10	6.1	19	11.5	38	23.0	165	100
total	81	36.3	34	15.2	25	11.2	35	15.7	48	21.5	223	100

In the example (1) the speaker changes his voice when he imitates how a foreigner would speak Finnish, and in the example (2) there is a longer code-change sequence, where the speaker switches completely into English. There appears an attempt to create comic associations through using Finnish or English in the examples (1) and (2), and a lot of the expressions are idiomatic which would be hard to translate:

- (1) S: Nån intresserad?! Chockoplator? *Halu ostaa? Onko?* ('Someone interested?! Chocolate bars? Want to buy? Anyone?') (10.10.2005)
- (2) (2) A: Man vill inte själv gå till den här stor svart liksom gangster: "Excuse me, do you call me shitty-looking white boy? You know what, you take it back now." ('You do not want to go in front of that big black, like, gangster and say: --')
 S: "You yourself have very large pants."
 A: "You are shitty-looking. So, now you say you pardon me and then you go on with your day and I go to Macy's. Want to play one-on-one on basketball, I show some tricks. I'm a mean shooter."
 S: "Back in Paippis I was a legend."
 A: "I was a legend. A basketball legend. You know Harlem Globetrotters was actually my idea. At the time." (7.11.2005)

Only five Finnish items can be categorized as code-mixes. On contrary, there are 29 English code-mixes. In example (3) the speaker clearly uses Finnish grammar, because the inflection is Finnish, and in example (4) English plural marker -s indicates that it is a code-mix:

- (3) H: Det finns såna new age-affärer där finns såna där valsångsskivor. Tänk om en platta skulle heta Mussång. *Hiirilaulua*. ('There are such new age shops where you can find whale song albums. Think about if an album had been named Mousesongs. Mousesongs.')(14.11.2005)
- (4) S: Du har alltså inga clown-*auditions*? Alltså, nå visa vad du kan... ('So, you don't have any clown-*auditions*? Like, show me now what you can do...')(17.10.2005)

Quotations make up the only category in which there appear fewer English items than Finnish ones. Finnish quotations are used in 15 items and English in 10. In example (5) the speaker tries first to translate the quote into Swedish, but changes into Finnish and formulates the whole sentence again in Swedish. In example (6) the speaker quotes something that has been earlier said in English:

- (5) H: När Rasmus hade kommit hem han sa att: "Joo, det här luktar liksom Heli, *tää tuoksuu ihan Heli-tädiltä*, det här luktar liksom moster Heli och luktar vanilj då." ('When Rasmus had returned home, he said that: 'Yeah, this smells like Heli, this smells just like aunt Heli, this smells like aunt Heli and it smells like vanilla.')(10.10.2005)
- (6) A: Ett gäng av stora svarta killar, så sa en efter jag gick förbi, så en sa: "*Look at that shitty-looking white boy.*" ('A gang of big black guys, so when I went by, one of them said: --')(7.11.2005)

There are 18 Finnish and 19 English nonce loans in the material. In example (7) the pronunciation is clearly Finnish and the word goes fluently in the sentence, and in example (8) the loan word has been conjugated according to Swedish grammar but uses English phonology. In both cases the speaker brings up a new word through a phenomenon, that he or she cannot describe in Swedish as well as it goes in Finnish or in English:

- (7) H: Jag tycker att vad som är sorligast för jag som är en gammal *mummo* är det att numera kan man varken virka eller sticka i flygplanet. ('I think that the saddest thing for me as a grandma is that nowadays we are not allowed to do crochet-work or knit in on the plane.')(28.11.2005)
- (8) S: Vi ska slåss så där *faggigt* så där, vet du, så blir det inte så farligt. ('We are going to fight like fags like that, you know, so it doesn't become so dangerous.')(10.10.2005)

There are 10 Finnish integrated loans, where as there are 38 English. In example (9) the Finnish word *juttu* belong grammatically, morphologically and syntactically so fluently into Swedish that it nearly appears more Swedish than Finnish. In example (10) the English word *blog* has got one more *g* in order to look more Swedish:

- (9) H: Ni förstör alltid mina *juttun*. ('You always ruin my stories.')(10.10.2005)
- (10) S: ... som är en *blogg* som skrivs alltså på finska och svenska... ('...which is a blog that is being written in Finnish and Swedish.')(10.10.2005)

7 MARKING THE CODE-SWITCHING AND BORROWING IN *HETIMITEN*

Altogether there appear 223 Finnish and English loan expressions in the data. 144 of these items are marked with various means and 79 of them are left unmarked. Thus 64.6 % of all the expressions are marked and 35.4 % are used without marking them. The speakers mark both Finnish and English expressions in relatively same amount: 65.5 % of Finnish and 64.5 % of English expressions are marked. Table 2 below illustrates the amount of marked and unmarked code-switching and borrowing. Table 3 below illustrates the amount of different metalinguistic and paralinguistic markers used in the examples. The markers used with code-changes, code-mixes, quotations, nonce loans and integrated loans have been categorized separately, because it is easier to compare different types of expressions with markers. The figures are higher in Table 3 than in Table 2 because sometimes one item is marked with several means.

Table 2: The absolute amounts of marked and unmarked code-switching and borrowing in the data.

		code- changes	code- mixes	quota- tions	nonce loans	integrated loans	total
		amount	amount	amount	amount	amount	amount
marked	Finnish	10	4	15	7	2	38
	English	66	24	6	6	4	106
<i>total</i>		76	28	21	13	6	144
unmarked	Finnish	2	1	0	9	8	20
	English	3	5	4	13	34	59
<i>total</i>		5	6	4	22	42	79
<i>total of Finnish and English</i>		81	34	25	35	48	223

Table 3: The absolute amounts of metalinguistic and paralinguistic markers in the data.

	code- changes	code- mixes	quota- tions	nonce loans	integrated loans	total	%
hesitation words	4	8	0	4	3	19	8.7
translations and clarifications	6	5	8	1	1	21	9.6
changes in voice	59	12	15	8	2	96	44.0
laughter	13	2	2	0	0	17	7.8
gestures	30	10	7	6	2	55	25.2
filled and silent pauses	4	4	0	1	1	10	4.6
<i>total</i>	116	41	32	20	9	218	100
<i>%</i>	53.2	18.8	14.7	9.2	4.1	100	

The expressions are rather marked than left unmarked, but there are some differences between the categories. The speakers mark the expressions more when it comes to code-changes, code-mixes and quotations, but leave them unmarked in cases of nonce loans and integrated loans. The speakers use metalinguistic markers 40 times in order to mark Finnish and English expressions. That makes 18.3 % of all the marking means, which indicates that it is more unusual to mark the expressions with metalinguistic means than with paralinguistic. Paralinguistic means appear 168 times with the expressions, which makes 77.1 % of all the marking means. Pauses mark 10 expressions, which is 4.6 % of all the marking means.

The most common marker is a change in voice, which is used with 96 expressions. It is 44.0 % of all the marking means. The changes in voice that came up in the data were emphasizing, imitation, lowering or arousing the voice and singing. The change in voice is most used with code-

changes, 59 code-changes are marked with change in voice. Example (11) is about code-change, which is marked through imitation:

- (11) A: Man vill inte själv gå till den här stor svart liksom gangster: "Excuse me, do you call me shitty-looking white boy? You know what, you take it back now." ('You do not want to go in front of that big black, like, gangster and say: --') (7.11.2005)

The next common marker is gestures, which are used to mark 55 items. It is 25.2 % of all the marking means. Gestures can be divided into two groups: to gestures that are clearly bound to the word and to gestures that ease up the atmosphere. In example (12) the speaker switches into Finnish and simultaneously shows his muscles in order to present strength:

- (12) S: Där var en bild av en tjur och sån där bodybuilderkille och sen sjöng de sån där sång om buljong: "Buljong! Siitä tulee vahva!" ('There was a picture of an ox and a bodybuilder and they sang a song about consommé: 'Consommé! It makes you strong!') (24.10.2005)

The third common class of markers is translations and clarifications. 21 expressions are marked through these means, which makes 9.6 % of all the marking means. The speakers can translate or clarify Finnish and English expressions into Swedish. Sometimes it also happens the other way round, when the speaker first uses a Swedish word but then translates it into Finnish or English, like in example (13):

- (13) S: Byoriginal. Ja, respekt åt originalen. *Respect the original.* ('Village original. Yes, respect the original. Respect the original.') (24.10.2005)

19 items are marked by hesitation words, which make 8.7 % of all the marking means. Hesitation words appear sometimes simultaneously with pauses, and it is hard to differentiate them. There are three hesitation words, *såna här* 'such', *liksom* 'like' and *vad heter det* 'what are they called', in example (14):

- (14) S: Ibland misslyckas också såna här liksom *practical jokes* och skämt, vad heter det... ('Sometimes also such, like, practical jokes and jests fail, what are they called...') (21.11.2005)

Laughter is the least used marker in the data. 17 expressions are marked with laughter, which makes 7.8 % of all the marking means. The speakers laugh almost all the time in *Hetimiten*, which makes it hard to distinguish if the laughter has something to do with the expression or not. There are some cases where the speakers mark their nervousness through laughter when they e.g. cannot remember an appropriate word or are insecure to use a foreign word, like in example (15):

- (15) H: Ja, den heter på, det här är också från USA, *the tingler, the ultimate head massager*. ('Yes, this is called, this is also from the USA, --') (31.10.2005)

Pauses are used as markers with 10 expressions only. It makes 4.6 % of all the marking means. The filled and silent pauses come often together, and the filled pauses are sometimes blended with hesitation words. In example (16) the filled pause has been blended with hesitation word *den där* 'that':

- (16) "Jag blir här och väntar den där [paus] *brainytåget*. Är det okej?" ('I stay here and wait for that brainy train. Is it okay?') (17.10.2005)

8 DISCUSSION

This study has shown that the Swedish-speaking Finns in the discussion program *Hetimiten* use both Finnish and English expressions during their Swedish speech. Although, the amount of Finnish and English is not equal. English expressions cover for 74.0 % of the data whereas the percentage for Finnish expressions is 26.0 %. This means that in *Hetimiten* the speakers use rather English than Finnish expressions, if they want to code-switch or borrow.

The speakers use relatively many English code-changes in the data in comparison to Finnish. Many English code-changes are quotes from movies, TV-programs, computer games, songs etc. They can also be idiomatic expressions, which have been repeated e.g. in school again and again. That kind of expressions do not require a lot of thinking, so they can easily slip into the Swedish discussion without that the speaker even notices having switched the language. Some of the expressions are also used as loose idiomatic phrases in the discussion in order to comment on something, usually they express the speakers' emotions or attitudes. Some longer code-change segments also arise when the speakers switch their language into English for a longer period of time. Code-mixes in the data are often compound words or pieces of combinations, when usually one part of the word is Swedish and the other Finnish or English, e.g. *Marianne-karkkeja* 'Marianne candies', *hiirilaulua* 'mousesong', *foxy ladies* and *practical jokes*. The speakers use also quotations in order to quote something that someone else has earlier said: *För vad som hon frågade, sa de: "Hilleri!"* ('Whatever she asked, they said: 'Fitch!').

Code-switching is most commonly used in order to express idiomatic phrases and bring up comic associations. This may be due to the construction of code-changes, code-mixes and quotations. As the code-switches use morphology, phonology and syntax of the guest language, they are more apt to be idiomatic expressions and create comic associations. Also the foreign sound itself can bring up comic associations when the utterances become highlighted just because they are not Swedish. On the contrary, loanwords most typically express lexical variation or refer to new phenomena. It is obvious that just loanwords are used to give variety, because they can also have Swedish synonyms which could be used as well in the discourse. When the loanwords are being used to present new phenomena and thus new words, they first become nonce loans and then integrated loans. Consequently, it is evident that loanwords, not code-switches, should bring up the new words.

The speakers mark 64.6 % of the expressions and leave 35.4 % unmarked. Both Finnish and English expressions are marked evenly. In this article the metalinguistic markers include hesitation words, translations and clarifications. Paralinguistic markers include changes in voice, laughter and gestures. Pauses are observed as a category of their own, because it is hard to distinguish if they belong to metalinguistic or paralinguistic markers. Filled pauses can be hesitation words which are uttered slowly or hesitation sounds which indicate that the speaker is insecure. Silent pauses are clear breaks between the words. Naturally, the speakers tend to mark the expressions more in cases of code-changes, code-mixes and quotations, but leave them unmarked when it comes to nonce loans and integrated loans. The difference between the marking of the code-switching and borrowing depends again on the different structure of the code-switches and loanwords as already mentioned above.

This article has presented what kind of Finnish and English expressions comes up in a Finland Swedish discussion program. One could think that Swedish spoken in Finland would have its external sources or influence mostly from the Finnish language, which is spoken in the same environment. However, the results show that English expressions dominate over Finnish in what comes to the data from a Finland Swedish discussion program. English has gained a strong position as a world language, and it seems likely that the English expressions will increase also in Swedish spoken in Finland. Nevertheless, it is unlikely that the Finnish expressions would completely disappear from Finland Swedish because of the shared environment and the long common history.

RESEARCH MATERIAL

Hetimiten (Ögonaböj). Episodes broadcasted in Finlands Svenska Television: 10.10.2005, 17.10.2005, 24.10.2005, 31.10.2005, 7.11.2005, 14.11.2005, 21.11.2005 and 28.11.2005.

REFERENCES

- Battarbee, K. 2002. "Finnish." In: Görlach M. (ed) *English in Europe*. New York: Oxford University Press Inc. 261-276.
- Beijar K., Ekberg H., Eriksson S. & Tandefelt M. 1998. *Ett land, två språk – den finländska modellen*. Espoo: Schildts.
- Finlandssvensk ordbok. 2008. af Hällström-Reijonen C. & Reuter M. (eds.) Espoo: Schildts.
- Hiidenmaa, P. 2003. *Suomen kieli – who cares?* Keuruu: Otavan kirjapaino Oy.
- Knookala T. 1982. *En beskrivande studie över tvåspråkighet bland finska invandrare – kodväxling och interferens*. FUMS Rapport Nr 105. Uppsala.
- Lampinen, K. 2002a. "Om finlandssvenskt radioprat." In Alanen K., Johansson C.-E., Nikula K. (eds.), *Från ord till handling - språkliga, kulturella och historiska artiklar*. Tampere 2003: Juvenes Print. 119-133.
- Lampinen K. 2002b. "Om ordvalet i finlandssvenskt radioprat." In Sorvali I. & Rossi P. (eds.), *Svenskan i Finland 7*. Oulu 2003: Oulu University Press. 83-89.

- Lauttamus, T. 1990. Code-Switching and Borrowing in the English of Finnish Americans in an Interview Setting. Studies in languages Nr 20. University of Joensuu. Faculty of Arts. Joensuu: Joensuun yliopiston monistuskeskus.
- Löfström, J. 1995. "Språk och attityder bland helsingforssvenska ungdomar. En projektpresentation." In Saari M. & Lindström J. (eds), *Svenskan i Finland 3. Meddelanden från institutionen för nordiska språk och nordisk litteratur vid Helsingfors universitet*. Helsingfors: Universitetstryckeriet.
- Nyholm, L. 1978. "Svenskan i Helsingfors." In: Laurén C. (ed), *Finlandssvenskan - fakta och debatt*. Porvoo: Söderströms. 39-56.
- Penttinen, A.-K. 2008. Om inlåning av finska och engelska ord samt kodväxling i samtalsprogrammet Ögonaböj. Master's thesis.
- Sharp, H. 2001. *English in Spoken Swedish. A Corpus Study of Two Discourse Domains*. Acta universitatis Stockholmiensis. Stockholm Studies in English. Göteborg: Elanders Novum.
- Svenska Akademiens ordlista över svenska språket. 2006. Utgiven av Svenska Akademien. Stockholm: NordBook.

John W. Schwieter

Lexical inhibition in trilingual speakers

Speaking is an intricate part of our everyday lives. Nonetheless, it is a procedure that perhaps is taken for granted in terms of its cognitive complexity. Although speaking comes so natural to us—most would consider it to be something that requires little conscious “thinking” at all—when all is said and done, it is a complicated procedure that manifests itself in the mind at multifaceted, lightening speeds. Putting aside the morpho-syntactic construction of words and phrases, just take, for instance, the retrieval of words from the mental lexicon. With everything we say, there is a communicative intent which underpins the retrieval of the words that best represent such intention. The situation becomes much more complicated in the case of the bilingual because they many times have two (or more) words for each object. For an English-Spanish bilingual, just saying the word “dog” in English may activate the words “dog” and “cat” in both languages. How is the right word in the right language ultimately chosen among all the activated lexical candidates? Some researchers view lexical access as a tri-level procedure in which the semantic/conceptual level sends activation to the lexical level which in turn sends activation the phonological level (see Costa, 2005, La Heij, 2005, and Schwieter, 2010 for overviews). Because studies have overwhelmingly demonstrated cross-linguistic competition at the semantic level (Jescheniak & Schriefers, 1998; Peterson & Savoy, 1998), lexical level (Finkbeiner, Almeida, Janssen, & Caramazza, 2006; Schwieter & Sunderman, 2009), and phonological level (Costa, Miozzo, & Caramazza, 1999; Hermans, Bongaerts, De Bot & Schreuder, 1998), the idea of inhibitory control, or the cognitive suppression of competitors, as specified in Green’s (1986, 1998) Inhibitory Control Model (ICM) has proven to be one of the most prominent theories in bilingual speech production to help explain lexical access. Nonetheless, many studies in bilingualism have shown that its predictions may be sensitive to second language (L2) proficiency (Costa & Santesteban, 2004; Costa, Santesteban, & Ivanova, 2006; Schwieter, 2008, 2010; Schwieter & Sunderman, 2008, 2009). The present study tests whether or not lexical access entails inhibitory control among speakers of three languages. In the next sections, the ICM and the studies that have supported it are presented along with the studies that have recently argued for an alternative account. Finally, the current study is presented including detailed accounts of the participants, experimental procedures, results of the statistical analyses, and implications for future research.

1 INHIBITORY CONTROL AND LEXICAL ACCESS

To begin our discussion of inhibitory control, we must first consider how the process of word retrieval functions. As briefly mentioned above, lexical access involves the activation of concepts, words, and sounds. An example is illustrated in Figure 1 in which an English-Spanish bilingual is asked to name a picture of a chair. First, the object is identified at the conceptual level. Many times, concepts overlap with other semantically related items that share features and because of this, other concepts such as *table* may also be activated. The activation at the conceptual level feeds down to the lexical level activating both the words *chair* and *table* in English and their translation equivalents in Spanish (*silla* and *mesa*, respectively). Clearly, when several lexical items become activated, there must be a cognitive mechanism that helps bilinguals select the right word in the correct language. Focusing on the lexical level, which is the focal point of the present study, Green's (1986, 1998) ICM posits that active inhibition is applied to the non-target words. In particular, the model supports the notion of language tags, an idea that words in both languages have a built-in label specifying to which language they belong. These tags help inhibitory control distribute inhibition accordingly (i.e., more inhibition is applied to words in the non-target language due to the fact that they do not carry the appropriate language tags). Furthermore, the ICM hypothesizes that inhibition is proportionate to the amount of activation sent from the conceptual level. As such, more inhibition is applied to non-target words with higher activation levels. Some empirical tests of the ICM have included natural speech situations (Grosjean, 1988, 1997; Grosjean & Miller, 1994; Li, 1996, 1998), brain activity (Abutalebi & Green, 2008; Hernandez, Dapretto, Mazziotta, & Bookheimer, 2001; Price, Green, & von Studnitz, 1999), and Stroop translation tasks (Schwieter, 2008; Schwieter & Sunderman 2009). However, without a doubt, switching tasks in which bilinguals are forced to switch back and forth between their languages have been the most elaborated empirical test of the ICM. For instance, the picture- or numeral-naming task with language switches has been employed by a number of researchers in bilingualism (Abutalebi & Green, 2007; Costa & Santesteban, 2004; Costa et al., 2006; Hernandez & Kohnert, 1999; Finkbeiner, Almeida, Janssen, & Caramazza, 2006; Gollan & Ferreira, 2007; Hernandez et al., 2001; Hernandez, Martinez, & Kohnert, 2000; Jackson, Swainson, Cunningham, & Jackson, 2001; Linck, Hoshino, & Kroll, 2008; Meuter, 1994; Meuter & Allport, 1999; Schwieter, 2010; Schwieter & Sunderman, 2008; Wodniecka, Bobb, Kroll, & Green, 2005).

Meuter (1994) and Meuter and Allport (1999) originally put forth the numeral-naming task as a means of estimating reliance on inhibitory control by measuring the time it takes to switch into the first language (L1) vs. the L2. The idea behind this was that because the ICM argues that more inhibition would be required for words in the more dominant language (L1) as compared to the less dominant language (L2), it should take longer (in ms) to reactive the L1 (e.g., an L2-L1 switch takes longer than an L1-L2 switch). Meuter and Allport asked the bilingual participants to name lists of numbers on a computer screen in the appropriate language according to the coloured box around which each target was displayed. The lists were designed so that participants were switching back and forth between their two languages. As anticipated, there were asymmetrical switch costs between the L1 and L2 with the L1 taking a substantially larger

hit when being reactivated from suppression. Their study provided the initial empirical evidence that bilingual lexical access during speech production entails inhibitory control.

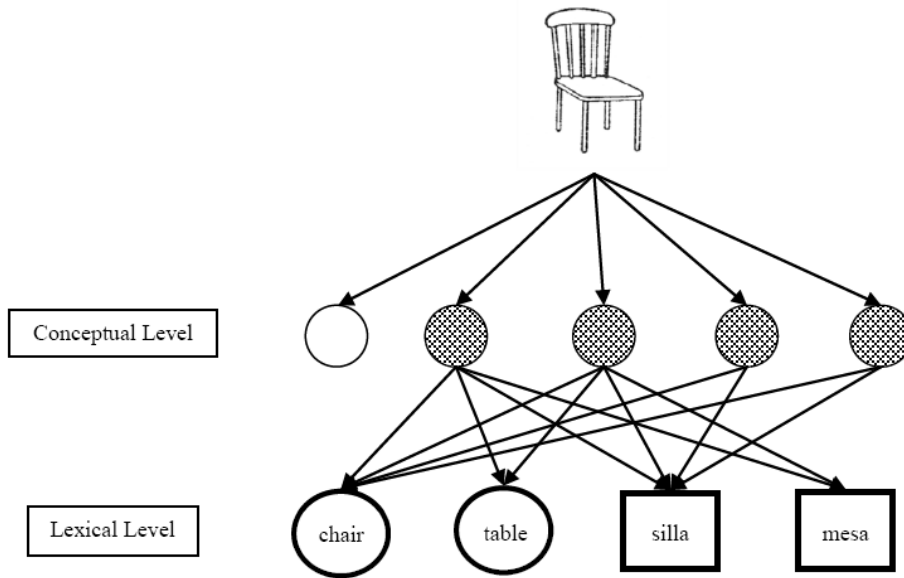


Figure 1. Bilingual lexical access.

2 L2 PROFICIENCY AND LEXICAL ACCESS

One major development in bilingual lexical access has been the argument that the ICM may be sensitive to proficiency level in the L2. This first arose from Costa and Santesteban's (2004) study in which highly-proficient bilinguals showed evidence of symmetrical switch costs. Because *asymmetrical* switch costs had been widely accepted as empirical support for the ICM, Costa and Santesteban's study demonstrated that not all bilinguals will need to rely on mechanisms of inhibitory control. In their study, they also compared the language switching performance of less-proficient L2 learners to highly proficient bilinguals. Only the less-proficient L2 learners suffered asymmetrical switch costs, thus supporting reliance on inhibitory control only at lower levels of L2 proficiency. Costa and Santesteban argued that with increases in L2 proficiency, bilinguals shift away from reliance on inhibitory control. One alternative they have considered is that at higher proficiency levels, bilinguals are able to make use of a language cue that automatically points them in the direction of the target word without having to rely on inhibitory control. Additional support for this account was replicated in Costa et al. (2006).

La Heij and colleagues have been the leading voices for the widely-accepted language-selective account of lexical access which is similar to what Costa and Santesteban (2004) identified as the alternative to inhibitory control—and perhaps what bilinguals shift towards

when moving away from inhibitory control (Bloem & La Heij, 2003; Bloem, van den Boogaard, & La Heij, 2004; Costa & Caramazza, 1999; Costa et al., 1999; La Heij, 2005; Roelofs, 1998). Under the scenario, lexical access is a procedure that need not rely on mechanisms of inhibitory control due to the fact that bilinguals are able to rely on a language cue that is present at the beginning stages of the speech planning procedure (i.e., at least by the conceptual level). As such, the language cue would help ensure that the activation sent from the conceptual level to the lexical level is greater than words competing in the non-target language. Because the non-target words are not considered in the lexical selection procedure, inhibitory control will simply not be required.

Acknowledging that both inhibitory control and language-selective mechanisms could both be plausible explanations for the cognitive processes of bilingual lexical access, Schwieter and Sunderman (2008) recently tested the notion of shifting away from reliance on inhibitory control towards a language-selective mechanism in a picture-naming experiment. In this study, less- and more-proficient English language learners of Spanish participated in a picture-naming experiment in accordance with Costa and Santesteban (2004). Once again, switching into the L1 was more costly than switching into the L2 suggesting reliance on inhibitory control. However, this effect was indeed restricted to those with less proficiency; for the highly-proficient language learners, there were symmetrical switch costs obtained for the L1 and L2. Further regression analyses were able to pinpoint a specific verbal fluency (proficiency) level at which the difference in switch costs became null. The researchers interpreted this as an important threshold which language learners must surpass in order to be able to access words without inhibitory control. Essentially, the threshold was a theoretical point of proficiency at which the ability to go about lexical access without the assistance of inhibitory control surfaced. Further support for Schwieter and Sunderman was extended using the Stroop interference paradigm in Schwieter and Sunderman (2009). The results from Schwieter and Sunderman (2008, 2009) lend themselves to interpret bilingual lexical access as a dynamic procedure which may not always be underpinned by inhibitory control as had been postulated by Costa and Santesteban. The researchers put forth the Selection by Proficiency (SbP) Model (see Figure 2) which demonstrated the variability of inhibitory control as modulated by L2 proficiency. Under this view, proficiency is interpreted as L2 lexical robustness, the familiarity with and frequency of word retrieval that leads to greater automaticity of access of lexical items.

As shown in Figure 2, the bilingual lexical access procedure is demonstrated for two language learners. The one on the left represents a bilingual who has very low L2 proficiency while the one on the right depicts a bilingual who has a very high L2 proficiency. The main difference between the procedures is that the less-proficient bilingual must rely on inhibitory control mechanisms due to the lack of conceptualization of a language cue. The opposite holds true for the more-proficient bilingual on the right: because this bilingual can make use of the language cue at the conceptual level, having to inhibit non-target words at the lexical level may no longer be necessary. Nonetheless, the SbP Model includes a diagonal line that allows for more-proficient bilingual to revert to mechanisms of inhibitory control when the language cue fails or is not strong enough to assist lexical access.

1. Does lexical access during trilingual speech production entail inhibitory control?
2. Will the magnitude of inhibitory control for each language be proportionate to the proficiency level?

As for the first question, it is anticipated that lexical access for trilinguals will show evidence of inhibitory control as interpreted through asymmetrical language switch costs. This is expected because the participants in the present study were neither highly proficient in their L2 nor third language (L3). Therefore, regarding the second research question, it is expected that in terms of the magnitude of the switch costs, the following order will be observed (from largest asymmetrical switch costs to least): $L1 > L2 > L3$, mirroring the proficiency distribution of each language. The next section presents the current study that test these research questions. First, information about the participants and experimental procedures is outlined. This is followed by sections on the results of the statistical analyses, the theoretical implications drawn from the results, and suggestions for future research.

3 PRESENT STUDY

3.1 Participants

English (L1) speakers of French (L2) were recruited from intermediate and advanced French courses at a university in an English-speaking region of Canada. These participants were also currently enrolled in introductory or intermediate Spanish (L3) courses at the same university. In all, 38 participants completed a language history questionnaire and took part in a picture-naming task individually in a psycholinguistics laboratory. The descriptive statistics from the language history questionnaire revealed an average age of 20.4 years with ages of language acquisition beginning at < 1 year for the L1, at 6.8 years for the L2, and at 17.2 years for the L3. The questionnaire also gathered participants' self-ratings of their overall language abilities (i.e., including reading, writing, speaking, and listening) in all three of their languages on a ten-point scale where 1 = not fluent; 10 = very fluent. Participants rated their abilities as follows: L1: 9.9; L2: 7.4; and L3: 4.6. *T*-tests were conducted and ensure an even spread between language proficiencies showing significant differences between each of the languages: L1 and L2, $t(37) = 9.60$; $p < .0001$; L2 and L3, $t(37) = 7.37$; $p < .0001$; and L1 and L3, $t(37) = 17.59$; $p < .0001$. This suggests that participants were aware that their L2 abilities were significantly weaker than the L1 but stronger than the L3 (i.e., $L1 > L2 > L3$). This is an important point to highlight because it is expected that not only will these participants rely on inhibitory control to retrieve words when speaking, but that the magnitude of such inhibitory control for each language will be proportionate to the proficiencies.

3.2 Picture-naming task

The picture-naming task employed in the present study investigates the extent to which IC underpins lexical access when trilingual language learners speak. The experimental design follows Costa and Santesteban (2004) and Schwieter and Sunderman (2008) with an added response language (i.e., an L3). The experiment contained ten line drawings from the Snodgrass

and Vanderwart (1980) standardized picture list. These included pictures of a pencil, house, car, dog, book, cat, chair, table, bear, and a heart. Sixty randomly sequenced lists of 5-14 pictures were created. For lists with 5-10 pictures in length, no pictures were duplicated. However, for those with 11-14 pictures, repeated pictures were placed at least three pictures away from their first presentation. Each list included anywhere from 0-4 switch trials. A switch trial was identified as a picture in which the immediate preceding picture was named in a different language. A nonswitch trial was a picture which was named in the same language as its preceding trial. The total number of trials named included 600 pictures where 70% were nonswitch trials and 30% were switch trials (420 and 180 trials, respectively). A colour code was used as in previous studies (e.g., Costa & Santesteban, 2004; Schwieter & Sunderman, 2008) to establish the language of production for each trial presented on the computer screen. In the present study, participants were told to name pictures in English (L1) if they appeared on blue background screens, in French (L2) on red background screens, and in Spanish (L3) on yellow background screens. Of the 600 pictures, 200 were named in each of the three languages and each of the ten pictures appeared 60 times during the entire experiment.

During the experimental session, each participant was given verbal instructions and advised that they would be completing a series of five practice lists of pictures before beginning sixty experimental lists. The experimental procedure of each list of pictures (both practice and experimental) consisted of the following format: 1) a black fixation point (+) was presented in the centre of the screen on the white background for 500 ms; 2) the first picture of the list appeared on one of three coloured backgrounds for 2000 ms or until the participant responded into a microphone; 3) a black fixation point (+) was presented; 4) the next picture was shown (either a switch or nonswitch trial) and the cycle was repeated until the end of the list. The reaction times (RTs) of the participants were measured by the computer and the naming accuracy was coded by the researcher as correct or incorrect on a master key. The participants had the option of taking a short break after every 10 lists (approximately every 6 minutes) to avoid fatigue.

3.3 Results

A 2 × 3 ANOVA was carried out using participant means as random factors with trial type (nonswitch or switch) and response language (L1, L2, or L3) as critical variables. Only correct responses were used in the RT analyses and both correct and incorrect responses were used in the error analyses. A response was considered to be correct if the participant correctly named the picture in the appropriate language and was deemed incorrect in all other instances (e.g., naming the picture in the incorrect language, stuttering, microphone malfunction, etc). RTs faster than 300 ms and slower than 2000 ms were removed from the analyses and treated as outliers. Additionally, RTs that were 2.5 SDs above or below the participant's mean RT for each condition were excluded from the analyses and also treated as outliers. The total amount of data removed was roughly 4.3% of the data.

The descriptive statistics for RTs, accuracy, and switch costs can be seen in Table 1. For the RT analyses, there was a main effect for trial type, $F(1, 37) = 174.97$, $MSE = 1,123,103$, $p < .0001$, suggesting that switch trials (1074 ms) were slower than nonswitch trials (935 ms). There was also a main effect for response language, $F(2, 36) = 43.14$, $MSE = 389,370$, $p < .0001$. Further analyses revealed differences between each of the three languages: the L1 (925 ms) was faster

than the L2 (1060 ms), $F(1, 37) = 82.08$, $MSE = 710,447$, $p < .0001$; the L1 (925 ms) was faster than the L3 (1029 ms), $F(1, 37) = 46.27$, $MSE = 419,590$, $p < .0001$; and the L3 (1029 ms) was faster than the L2 (1060 ms), $F(1, 37) = 4.07$, $MSE = 38,073$, $p < .02$. There was also an important significant interaction for trial type * response language, $F(2, 36) = 32.55$, $MSE = 63,092$, $p < .0001$. Post-hoc analyses revealed that the L1 switch cost (184 ms) was significantly greater than the L2 switch cost (156 ms), $F(1, 37) = 3.65$, $MSE = 7,837$, $p < .03$; the L2 switch cost (156 ms) was larger than the L3 switch cost (75 ms), $F(1, 37) = 29.71$, $MSE = 64,345$, $p < .0001$; and the L1 switch cost (184 ms) was larger than the L3 switch cost (75 ms), $F(1, 37) = 77.92$, $MSE = 117,093$, $p < .0001$.

For the analyses of accuracy, there was a main effect for trial type, $F(1, 37) = 10.75$, $MSE = .009$, $p < .0001$, suggesting that switch trials (95.1%) were less accurate than nonswitch trials (96.4%). There was also a main effect for response language, $F(2, 36) = 13.61$, $MSE = .011$, $p < .0001$, demonstrating differential accuracy between all three languages. Post-hoc analyses revealed that L2 (94.6%) was less accurate than L1 (97%), $F(1, 37) = 23.30$, $MSE = .022$, $p < .0001$; L2 (94.6%) was less accurate than L3 (95.5%), $F(1, 37) = 3.43$, $MSE = .003$, $p = .03$; and L3 (95.5%) was less accurate than L1 (97%), $F(1, 37) = 13.66$, $MSE = .008$, $p < .0001$.

Table 1. Mean Reaction Times (ms) and Accuracy (%).

	<u>L1</u>		<u>L2</u>		<u>L3</u>	
	RT	Acc	RT	Acc	RT	Acc
Switch	1017	98.8	1138	95.3	1066	96.0
Nonswitch	833	96.3	982	94.0	991	95.1
Switch cost	184		156		75	

To explore whether or not there was a relationship between proficiency (as operationalized by participant self-ratings) and language switching performance, correlation analyses were conducted using the Pearson product-moment correlation coefficient (Cohen, 1988). Although there were no correlations found for L1 ratings and L1 switch costs or for L2 ratings and L2 switch costs ($r = 0$), there was a significant (large) correlation for L3 ratings ($M = 4.64$, $SD = 1.89$) and L3 switch costs, ($M = 93.80$, $SD = 45.45$), $r(37) = .617$, $p < .0001$. To further verify and explore this effect, an additional 2×3 ANOVA was carried out using the same participant means as reported above (with trial type (nonswitch or switch) and response language (L1, L2 or L3) as critical variables) but with the addition of L3 proficiency rating as a between-subjects factor. By using a mean split ($M = 4.6$) of the L3 proficiency ratings, two groups were created (lower L3 proficiency group vs. higher L3 proficiency group). As can be seen in Table 2, fairly similar switch costs were elicited for the L1 and L2 but not so much for the L3. When looking at the results of the second ANOVA, as expected, the same main effects for trial type and response language were identified in addition to the interaction for trial type * response language.

However, by splitting the data to compare two groups of L3 proficiency levels, differences were identified in terms of the magnitude of the L3 switch cost. There was a significant three-way interaction for trial type * response language * L3 proficiency rating interaction, $F(2, 35) = 5.60$, $MSE = 4,530$, $p < .01$. Post-hoc analyses demonstrated a significant difference in L3 switch costs for the less proficient L3 group (60 ms) and more proficient L3 group (114 ms), $F(1, 36) = 17.71$, $MSE = 12,606$, $p < .0001$. The correlation analysis and the ANOVA L3 proficient split analysis demonstrate a relationship between proficiency of the least dominant language and the magnitude of switch costs. This is, however, confined only to the L3 switch costs and was not observed with the L1 or L2 proficiency ratings.

Table 2. Mean Reaction Times (ms) by L3 proficiency.

	<u>Less- L3 proficient</u>			<u>More- L3 proficient</u>		
	<u>L1</u>	<u>L2</u>	<u>L3</u>	<u>L1</u>	<u>L2</u>	<u>L3</u>
Switch	938	1014	943	1014	1161	1088
Nonswitch	744	841	883	828	979	974
Switch cost	194	173	60	186	182	114

4 DISCUSSION AND CONCLUSIONS

Returning to the two research questions presented above, the results of the ANOVA suggest that trilinguals—at least those having L2s and L3s significantly weaker than their L1s—will rely on inhibitory control to assist word retrieval while speaking. Indeed, the interaction for trial type * response language pointed out massive asymmetries in language switch costs (L1 = 184 ms; L2 = 156 ms; L3 = 75 ms). *T*-tests confirmed statistically significant differences between all three languages supporting reliance on inhibitory control for all three languages. Given the imbalance of language proficiency between all three languages of the participants in the present study and given that previous research (Costa & Santesteban, 2008; Costa et al., 2006; Schwieter & Sunderman, 2008) with less-proficient bilinguals has uncovered similar asymmetrical switch costs with less dominant languages, it may not be surprising that inhibitory control is needed to support lexical access for these trilingual speakers. In fact, even though the SbP Model (Schwieter & Sunderman, 2008, 2009) is a bilingual model, the trilinguals in the present study could be placed in the left portion of the model (refer back to Figure 2). Under this assumption, they have not obtained strong enough lexical robustness in the less-dominant language to be able to rely on the language-selective mechanism seen in the right portion of the model and must undertake lexical access with the help of inhibitory control.

The second research question in this study was whether or not the magnitude of inhibitory control for each language would be proportionate to the proficiency level. It was expected that because the ICM (Green, 1986, 1998) argues more inhibition for more dominant (proficient) languages, there would have been larger switch costs for the L1 than for L2s and L3s. Although this was the case and even though the proficiency ratings were similarly distributed and significantly different from one another (L1: 9.9; L2: 7.4; and L3: 4.6), there was only a correlation between the L3 proficiency ratings and the L3 switch costs. Furthermore, when the data was split into two groups of lower and higher L3 proficiency, a sharper image of the effects of L3 proficiency emerged and were clearly confined to performance in the L3—participants with less L3 proficiency suffered significantly less switch costs in their L3 (60 ms) than participants with more L3 proficiency (114 ms). Their language switch costs for L1 and L2, however, were virtually the same. Schwieter and Sunderman (2008) suggest that in bilingual speech production, proficiency in the less dominant language (L2) will modulate reliance on inhibitory control. Although there were no universal findings across all three languages as there were in studies such as Costa and Santesteban (2004) and Schwieter and Sunderman, the results from the present study looking at these issues in trilingualism shed light on the familiar story that the proficiency level of the least dominant language is a determining factor in whether or not inhibitory control underpins lexical access.

Future studies will need to further examine how lexical access functions in trilingual speech production while taking note of the potential effects that the proficiency level may have. While the present study has taken a first look at inhibitory control in trilingual speech production through participants with substantially less L2 and L3 proficiency than the L1, more is yet to be uncovered. It will be beneficial to research in lexical access if future studies incorporate participants with very high L2 and L3 proficiency (ideally, balanced trilinguals). Only with more comparative data can we begin to better understand the cognitive processes that help facilitate how words are retrieved when multilinguals speak.

BIBLIOGRAPHY

- Abutalebi, Jubin, and David Green. 2007. Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics* 20.242–275.
- Abutalebi, Jubin, and David Green. 2008. Control mechanisms in bilingual language production: Neural evidence from language switching studies. *Language and Cognitive Processes* 23.557-582.
- Bloem, Ineke, and Wido La Heij. 2003. Semantic facilitation and semantic interference in word translation: Implications for models of lexical access in language production. *Journal of Memory and Language* 48.468-488.
- Bloem, Ineke, Sylvia van den Boogaard, and Wido La Heij. 2004. Semantic facilitation and semantic interference in language production: Further evidence for the conceptual selection model of lexical access. *Journal of Memory and Language* 51.307-323.
- Cohen, Jacob. 1988. *Statistical power analysis for the behavioral sciences* (2nd ed.).

- Costa, Albert. 2005. Lexical access in bilingual production. *Handbook of bilingualism: Psycholinguistic approaches*, ed. by Judith Kroll and Annette de Groot, 308-325. New York: Oxford.
- Costa, Albert, and Alfonso Caramazza. 1999. Is lexical selection in bilingual speech production language-specific? Further evidence from Spanish-English and English-Spanish bilinguals. *Bilingualism: Language and Cognition* 2.231-244.
- Costa, Albert, Michele Miozzo, and Alfonso Caramazza. 1999. Lexical selection in bilinguals: Do words in bilingual's lexicon compete for selection? *Journal of Memory and Language* 41.365-397.
- Costa, Albert, and Mikel Santesteban. 2004. Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language* 50.491-511.
- Costa, Albert, Mikel Santesteban, and Iva Ivanova. 2006. How do highly proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32.1057-1074.
- Finkbeiner, Matthew, Jorge Almeida, Niels Janssen, and Alfonso Caramazza. 2006. Lexical selection in bilingual speech production does not involve language suppression. *Journal of Experimental Psychology: Learning, Memory and Cognition* 32.1075-1089.
- Green, David. 1986. Control, activation, and resource: A framework and a model for the control of speech in bilinguals. *Brain and Language* 27.210-223.
- Green, David. 1998. Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition* 1.67-81.
- Grosjean, François. 1988. Exploring the recognition of guest words in bilingual speech. *Language and Cognitive Processes* 3.233-274.
- Grosjean, François. 1997. Processing mixed languages: Issues, findings and models. *Tutorials in bilingualism: Psycholinguistic perspectives*, ed. by Annette de Groot and Judith Kroll, 225-254. Mahwah, NJ: Erlbaum.
- Grosjean, François, and Joanne Miller. 1994. Going in and out of languages: An example of bilingual flexibility. *Psychological Science* 5.201-206.
- Hermans, Daan, Theo Bongaerts, Kees De Bot, and Robert Schreuder. 1998. Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition* 1.213-229.
- Hernandez, Arturo, Mirella Daprettob, John Mazziottab, and Susan Bookheimer. 2001. Language switching and language representation in Spanish-English bilinguals: An fMRI study. *NeuroImage* 14.510-520.
- Hernandez, Arturo, Antígona Martínez, and Kathryn Kohnert. 2000. In search of the language switch: An fMRI study of picture naming in Spanish-English bilinguals. *Brain and Language* 73.421-431.
- Jackson, Georgina, Rachel Swainson, Ross Cunnington, and Stephen Jackson. 2001. ERP correlates of executive control during repeated language switching. *Bilingualism: Language and Cognition* 4.169-178.

- Jescheniak, Jörg, and Herbert Schriefers. 1998. Discrete serial versus cascading processing in lexical access in speech production: Further evidence from the coactivation of near-synonyms. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24.1256-1274.
- La Heij, Wido. 2005. Selection processes in monolingual and bilingual lexical access. *Handbook of bilingualism: Psycholinguistic approaches*, ed. by Judith Kroll and Annette de Groot, 289-307. New York: Oxford.
- Li, Ping. 1996. Spoken word recognition of code-switched words by Chinese-English bilinguals. *Journal of Memory and Language* 35.757-774.
- Li, Ping. 1998. Mental control, language tags, and language nodes in bilingual lexical processing. *Bilingualism: Language and Cognition* 1.92-93.
- Linck, Jared, Noriko Hoshino, and Judith Kroll. 2008. Cross-language lexical processes and inhibitory control. *The Mental Lexicon* 3.349-374.
- Meuter, Renata. 1994. Language switching in naming tasks. Oxford: Oxford University dissertation.
- Meuter, Renata, and Alan Allport. 1999. Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language* 40.25-40.
- Peterson, Robert, and Pamela Savoy. 1998. Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24.539-557.
- Price, Cathy, David Green, and Roswitha von Studnitz. 1999. A functional imaging study of translation and language switching. *Brain* 122.2221-2235.
- Roelofs, Ardi. 1998. Lemma selection without inhibition of languages in bilingual speakers. *Bilingualism: Language and Cognition* 1.94-95.
- Schwieter, John. 2008. The effects of bilingual type on language selectivity. *Readings in language studies: Language across disciplinary boundaries*, ed. by Miguel Montero, Paul Chamness Miller, and John Watze, 417-431. New York: Information Age Publishing.
- Schwieter, John. 2010. *Cognition and bilingual speech: Psycholinguistic aspects of language production, processing, and inhibitory control*. Saarbrücken, Germany: Lambert Academic Publishing.
- Schwieter, John, and Gretchen Sunderman. 2008. Language switching in bilingual speech production: In search of the language-specific selection mechanism. *The Mental Lexicon* 3.214-238.
- Schwieter, John, and Gretchen Sunderman. 2009. Concept selection and developmental effects in bilingual speech production. *Language Learning* 59.897-927
- Wodniecka, Zofia, Susan Bobb, Judith Kroll, and David Green. 2005, August. Is the first language inhibited when speaking the second language? Evidence from a competitor priming paradigm. Paper presented at the 14th conference of the European Society for Cognitive Psychology, Leiden, The Netherlands.

4. *Diachronic linguistics*

Elena Bratishenko

On the pronoun *ego* in the history of Russian

This paper examines the variant spelling *ego/evo* – the Gen. and Acc. (syncretic with Gen.) cases of the 3rd pers. masc. sg. pronoun **jī* – in the well-known 17th century *Life of Archpriest Avvakum, written by himself*. The text, due to the unprecedented amount of colloquial features permeating its hagiographic style, contains a great number of the pronunciation spellings *evo* alongside the orthographic *ego*. The data come from copy *A* of the text, the first out of the three redactions (corresponding to the three first letters of the Cyrillic alphabet *a, b, v*) which, along with copy *V*, is believed to have survived in the original. The consonant phoneme /g/ comes from the proto-Indo-European emphatic particle *-go*. This particle was reanalyzed in Slavic as part of the Gen.-Abl. ending (Meillet 1956:434). It can be assumed that, by the time the text was written, /v/ had long become the pronunciation norm, the actual change beginning some time in the late 14th century (Bratishenko 2010). The spelling *evo* vs. the original *ego* reflects the change that affected the Gen. sg. masc./neut. ending in the declension of Russian adjectives and non-personal pronouns in the history of Russian whereby *-ogo/-ego* > *-ovo/-evo*¹; although Russian orthography has preserved the etymological spelling to this day, as illustrated in (1a-c).

- (1) a. *dorogogo* [dərəgónə] ‘dear’
b. *ego* [(j)ivó] ‘his’
c. *segodnja* [sivódn’ə] ‘today’ (lit. ‘of this day’)

As determined by Černyx (1927:68), Avvakum’s native dialect (*starogrigorovskij* ‘old Grigorovo’, after a village near Nižnij Novgorod) is close to the Moscow dialect which was the central variety of Russian during the Middle Russian period. Based on this, Avvakum’s usage, his personal idiosyncrasies notwithstanding, may be considered a representative sample of the 17th century Russian allowing one to hypothesize about the overall state of this pronoun at the time.

The discussion focuses on two syntactic usages of the anaphoric pronoun with the highest number of attestations, namely the Acc. case as direct object of transitive verbs, and the adnominal possessive Gen. case. From the analysis of the data, it emerges that, despite the fact that Acc. and Gen. of this pronoun are syncretic, the relative frequency of *ego* and *evo* in the *Life of Avvakum* differs between the two syntactic functions. While the attestations of the new *evo* by far prevail over *ego* in both usages, the higher number of *evo* attestations for Acc. suggests that,

¹ The two subtypes of the ending (*-ogo/-ego*) correspond to the two types of stems within the same declensional pattern. For the sake of economy, a single notation *-ogo* will be used hereafter for both variants.

even with the case syncretism, the two aren't treated as identical by the author. The uneven distribution of the two spelling variants in the two morphosyntactic constructions thus raises an interesting question regarding the homonymy of the two forms. It also has possible implications for the spread of the new ending from adjectives, where it arose, to pronouns. Acc. direct object of the anaphoric pronoun appears to have adopted the new form before the possessive Gen. The reason for this may be that the 3rd pers. anaphoric pronoun is in the process of developing into two different pronoun types: a personal 'him/it' and a possessive 'his/its'.

1 THE HISTORY OF THE ANAPHORIC 3RD PERS. PRONOUN *JĪ

The Common Slavic anaphoric pronoun *jĭ has been subject to many changes both historically and pre-historically. It is the result of the merger of two distinct proto-Indo-European pronouns: the relative *yo- and the anaphoric *i- (Vaillant 1958:423), and this inherent duality might be partly responsible for the proliferation of reflexes of *jĭ in Russian morphology. The following table contains the Old Church Slavonic (OCS), and thus late Common Slavic, declension of the anaphoric *jĭ in the singular, as it is the one relevant for the discussion below.²

Table 1. The anaphoric pronoun *jĭ.

Sg.	Masc.	Neut.	Fem.
Nom.	*jĭ	*je	*ja
Acc.	jĭ	je	jǫ
Gen.	jego	jeje	jeje
Loc.	jemĭ	jeji	jeji
Dat.	jemu	jeji	jeji
Instr.	jimĭ	jeje	jeje

There are no attestations of *jĭ in the Nom.³ In general, no Nom. of personal pronouns was normally used for subjects in early Slavic – only when the subject needed to be emphasized – because the person was expressed analytically by the verbal ending. Only with the restructuring of the Russian past tense system, such as the disappearance of the Aorist and the Imperfect, and the increase in the Perfect without an auxiliary (and thus uninflected for person), the personal pronoun came into regular use. In fact, all personal pronouns have suppletive Nom. case (Kuznecov 1959:113). The Nom. for the 3rd pers. anaphoric pronoun is borrowed from demonstrative pronouns. In OCS it is most often *tŭ* 'that one', sometimes *onŭ* (Lunt 1974:53, Vaillant 1958:433-34);⁴ in Russian *on* (<*onŭ*) has become standard. Here is an example from the *Life of Avvakum*:

² For a complete paradigm refer, for example, to Vaillant (1958:421-22).

³ This is also the case in the plural; while the dual has Nom. attestations.

⁴ The three types of demonstrative pronouns in early Slavic were *sĭ* 'this here', *tŭ* 'that here', and *onŭ* 'that there' – distinguishing not only between an object near by and an object at a distance, but also between an

- (2) *desetī lětū onū menja mučilū ili ja evo – ne znaju*
 ten years he-NOM me tortured or I him-ACC not know
 ‘for ten years either he tortured me or I him – I do not know’ (38)⁵

Cocron (1962:158-159) says that by incorporating the pronoun *onū* into the category of personal pronouns Russian has isolated it almost entirely from its original demonstrative function. Yet, the fact that this pronoun alternated with the demonstrative *tū* in the Nom. of the anaphoric paradigm, as well as its own demonstrative origin made it a link between personal and demonstrative pronouns. Avvakum uses it both as a personal and as a demonstrative pronoun employing it before surnames and first names of people (3a). Avvakum also uses two other demonstratives *sej* (< *sī-jī*) ‘this one’ (3b), *toj* (< *tū-jī*) ‘that one’ (3c) in this way:

- (3) a. *onū že-emph. Dionisij* ‘he/that one, Dionysius’ (5)
 b. *sej Dionisij* (3)
 c. *toj že Dionisij* (3)

Cocron notes that starting in the 16th century the pronoun *onū* is similarly associated with animate objects in many texts of administrative nature.

Incidentally, there is a long prehistory of suppletion in the Nom. of the Indo-European masc. and fem. sg. forms, although not neut., of demonstrative pronouns themselves. Slavic and Baltic, however, underwent an innovation not shared by their sister languages: they eliminated the suppletive Nom. forms. The new Slavic Nom. of the demonstrative pronoun (e.g. *tū* and *ta*, masc. and fem. sg. respectively) is based on the stem of the rest of the paradigm. This innovation brought Nom. and Acc. closer, which, in Meillet’s (1956:433) opinion, may have contributed to the merger of these two cases in **o*-stem nouns, which in turn triggered Gen.-Acc. syncretism in personal/animate masc. nouns of this declension. (Gen.-Acc. syncretism will be returned to shortly.)

Although no Nom. **jī* is attested in Slavic, its prehistoric existence is transparent in several etymologically related items. Thus it constitutes the OCS relative pronoun *jīže*, as in the citation from the *Life of Avvakum* below, *-že* being another particle akin to *-go* (Meillet 1956:434). (Note also the anaphoric pronoun *jemu* in the Dat.).

- (4) *iže xoščet spastisja, preže vsexū podobaetū*
 Who-rel.NOM wishes to be saved before everyone should
jemu deržati kafoličeskaja věra
 he-anaph.DAT hold catholic faith
 ‘the one who wishes to be saved, should before everyone else hold the catholic faith’ (6)

object related to the person speaking, an object related to the person one is speaking with, and an object without any relation to those speaking (Meillet 1965:439).

⁵ The reference is to the page number in the edition cited.

The Nom. is also discernible in other formations that can be traced back to the pronoun *jĭ. This pronoun has grammaticalized into the ending of various forms of the so-called pronominal (as opposed to nominal) declension that, after several stages in its development, eventually transformed into the compound adjectival declension in Russian (Cocron 1962:116). The pronominal type was mostly represented by the long-form (definite) adjectives and participles, as shown in (5), formed by an addition of *jĭ in a corresponding case and number to the short form.

(5) Long-form adjectives:

	<i>starŭ</i>								
	<i>starŭ</i>	+	<i>jĭ</i>	>		<i>starŭjĭ</i>	'the old one'		
	<i>stara</i>	+	<i>jego</i>	>		OCS <i>stara- jego</i>	(> <i>stara-ego</i> > <i>stara-ago</i> > <i>starago</i>)		
	<i>staru</i>	+	<i>jemu</i>	>		OCS <i>staru-jemu</i>	(> <i>staru-emu</i> > <i>staru-umu</i> > <i>starumu</i>), etc.		

According to Vaillant (1958:465), *-jĭ* (<*yo-) was commonly added to pronouns with very short stems for extension, producing such demonstratives as *sĭ-jĭ*, *tŭ-jĭ* already mentioned, or the interrogative *kakŭ-jĭ* > *kakŭjĭ* (> *kakoj*) 'which one'. The possessive sg. masc. 1st pers. *mo-jĭ* 'my', 2nd pers. *tvo-jĭ* 'your', reflexive *svo-jĭ* 'one's own', and interrogative *čĭ-jĭ* 'whose' too were formed this way. The plural possessive pronouns are younger and follow the model of possessive adjectives. The addition of the possessive adjective forming suffix *jĭ to the Gen. pl. of the personal pronouns produced the possessive pronouns *našĭ* (< *našĭjĭ*) 'our' and *vašĭ* (< *vas-jĭ*) 'your'. Prehistorically, the possessive suffix *jĭ- (perhaps also related to the anaphoric *jĭ?) combined with the stems of masc. personal, especially proper, nouns (*čĕsar-* 'ceasar', *čĕsarjĭ* 'ceasar's') to form possessive adjectives, the main means of expressing individual personal reference. This ancient suffix, already unproductive at the time of the earliest East Slavic records, was replaced by a younger suffix *-ov-* in possessive adjectives formed from the same masc. *o-stems, especially of foreign origin – the source of Christian patronymics and surnames. (See more on these in section 2.) The old East Slavic possessive adjectives (both those formed by *-jĭ-* and by *-ov-*) did not decline in the compound manner of the long-form adjectives, they declined like nouns. In the course of history the morphosyntactic category of possessive adjectives disappeared from Russian. Practically the only sphere where they have survived, at least morphologically, is the surnames in *-ov* and *-in*.⁶ Otherwise the adnominal Gen. case has taken their place in possessive constructions. The 3rd pers. indeclinable Gen. of the anaphoric pronoun (masc./neut. sg. *ego*, fem. sg. *jeje* (OCS)/*jeje* (old East Slavic); pl. (*j*)*ixŭ*) stands out both morphologically and syntactically from the rest of the possessive pronouns that belong to a class of their own. Originally, the most common alternative to Slavic possessive pronouns was enclitic Dat. of personal pronouns.

⁶ Possessive adjectives can still be formed in Modern Russian but only with the help of the suffix *-in* and are limited to hypocoristic names, such as *Saša*, hence *Sašin*, and kinship terms, such as *mama* 'mom', hence *mamin* 'mom's', of the feminine declension – former *a stems.

Regarding the compound declension, the situation in East Slavic is further complicated in that, in contrast to OCS, the Common Slavic endings changed yet again through analogical substitution based on the declension of the demonstrative pronoun *tū* 'that'. This produced the East Slavic Gen. *starogo* and Dat. *staromu* (after Gen. *togo*, Dat. *tomu*) corresponding to OCS *starago* and *starumu* – the latter two resulting from the loss of intervocalic /j/, assimilation and contraction, as explained in (5) above. Possessive pronouns weren't subject to this analogy: Cf. Gen. *mojego*, Dat. *mojemu* that preserve the integrity of **ǰi*.

There is still another change that affected the paradigm of the anaphoric pronoun, namely the rise of the Gen.-Acc. syncretism mentioned earlier. As Meillet (1956:434-35) notes, the Acc. **yon* produced, on the one hand, the anaphoric enclitic *ǰi* and, on the other, the emphatic relative *jego* used with reference to animate beings. Recall that, as a result of the loss of the Indo-European suppletive Nom. masc. and fem. forms, the Nom. and Acc. of demonstrative pronouns in Slavic became quite similar. On the other hand, in the major masc./neut. declension (**o*-stems) of nouns the two cases coalesced completely due to an apocope. During the historical period, as a consequence of the Nom.-Acc. syncretism, the Acc. of masc. personal, and eventually, of all masc. animate nouns was replaced by the Gen.-Acc. form. Gen.-Acc. also arose in personal pronouns, and in the anaphoric pronoun (*jego*).⁷ The process started when the anaphoric pronoun served as direct object of long active participles, presumably to prevent the confusion of the subject participle's Nom. ending in *-ǰi* with the Acc. direct object *ǰi* of the anaphoric pronoun. The degree of grammaticalization of the pronoun into an adjectival ending must have been not as high then for its confusion with the anaphoric pronoun to be possible.

It is noteworthy that, in contrast to nouns distinguishing between animate and inanimate masc. direct objects by means of either Gen.- or Nom.-Acc., the Russian 3rd pers. pronoun has a single Acc. form *ego*, syncretic with Gen., for both, despite the fact that it too first appeared when the pronoun referred to a male human being. Moreover, Gen.-Acc. also spread to the neuter (*ego*) and to the feminine (*ee*) of this pronoun. Interestingly, in OCS, early in the rise of Gen.-Acc., the relative frequency of the old Nom.-Acc. *ǰi*, regardless of personal/animate reference, is much higher than that of the new Gen.-Acc. *jego* (David Huntley – personal communication). In other words, this is a pronoun where Gen.-Acc. syncretism does not seem to be so sensitive to animacy distinctions, otherwise at the centre of this phenomenon. Perhaps this is due to the fact that most anaphoric pronoun attestations have human antecedents (Kuznecov 1959:131, Klenin 1983:26). This is certainly the case in the *Life of Avvakum*. In connection to this, Kuznecov (1959:131) points out the likely influence of the interrogative *kogo* on *jego* becoming the only Acc. form, since *kogo* always refers to a person and the Gen.-Acc. *kogo* is the only known Acc. of this pronoun.

2 -OGO > -OVO IN RUSSIAN ADJECTIVAL DECLENSION

The question of the substitution of the Gen. sg. ending *-ogo* by *-ovo* in Russian adjectives and pronouns has been addressed repeatedly from the 19th century on. The most prevalent, phonetic,

⁷ Klenin (1983:32), following other scholars, believes that the much earlier and predominant Gen.-Acc. of the relative *jegože* was a major influence on the rise of Gen.-Acc. of the anaphoric *jego*.

approach considered the substitution a gradual process: *ogo* > *oγo* > *oo* > *owo* > *ovo*. Its critics pointed out that a purely phonetic interpretation failed to account for the fact that the change occurred only in one particular morphological category, without affecting other identical phonological sequences, as is evident from (1a).⁸ This approach was superseded by Flier's (1983) widely-accepted hypothesis of direct phonemic reidentification (*-ogo* > *-ovo*) which considered morphological factors as well. An alternative explanation first proposed in general terms by Malinowski (1875)⁹ and recently taken up anew by Bratishenko (2010) is based on analogy. The idea of analogy is suggested by the existence of conspicuous non-Christian Russian surnames based on qualitative adjectives where the Nom. *-ov* did not arise etymologically, such as *Tolstov* (< *Tolstoj*) 'Fat', *Gorbatov* (< *Gorbatoj*) 'Hunchbacked'. In view of Bratishenko (2010), *-ovo* first arose in non-Christian patronymics – qualitative long-form adjectives by origin (such as *Tolstoj*), as a by-product of their becoming fixed surnames, a means of personal identification. The model for analogy was provided by Christian patronymics – possessive adjectives by origin (such as *Petrov*) formed by the suffix *-ov*. The earliest attestation of the ending *-ovo* is in a 14th century charter, in a surname *Xorobrov* based on the byname *Xorobroj* 'the Brave one' (Bratishenko 2010:82). The formant *-ov-*, abstracted from the **u*-stem declension, has long been associated with personal, especially proper, **o*-stem masc. nouns in the history of Slavic. Besides possessive adjectives (*Petr-ov* 'Peter's'), as already mentioned, it is found in the Dat. sg. (*Petr-ov-i* 'Peter'), in the Nom. pl. (*syn-ov-e* 'son'),¹⁰ and the Gen. pl. (*otc-ov* 'fathers') of the **o*-stem masc. nouns. In sum, a derivational morpheme – the possessive suffix *-ov-*, in combination with the Gen. ending *-a* of possessive adjectives contaminated an inflectional morpheme – the Gen. ending *-ogo* in qualitative adjectives.

2.1 Distribution of *ego* vs. *evo* in *The Life of Avvakum*

As is well known, the *Life of Avvakum* stands out from the preceding hagiographic tradition in many ways (Cocron 1962:23). One of the most written about aspects of the text is its blend of high and low registers. Some have theorized that the duality in the Archpriest Avvakum's switching between lofty and archaic Church Slavonic style and extremely colloquial "blabber" parallels the ideological opposition of the so-called old believers and the followers of Patriarch Nikon's reforms at the heart of the author's writing during the period of schism in the Russian Orthodox church. Indeed, there are plenty of dichotomies that have been commented on.¹¹ On the linguistic side, they include variation between the forms of the past tense in verbs – Aorist vs. Perfect (Timberlake 1995); and between the forms of the 1st pers. sg. pronoun *azv(že)* vs. *ja* 'I' (Worth 2000).

According to Černyx (1927:43), the *Life of Avvakum* is perhaps the first of the northern Russian monuments reflecting the ending *-ovo*, although, as mentioned at the outset, by the 17th century /v/ would have been standard in speech. Many instances of *ego* and *evo* reflect "register harmony" (Timberlake 1995:26), as citations in (6)-(11) show. The Church Slavonic variant of the

⁸ See Tolkačev (1960) for an extended overview.

⁹ Černyx (1927:44) is another proponent of this otherwise unpopular idea.

¹⁰ The noun *synŭ* is, of course, an original **u*-stem, but it joined the **o*-stem declension very early.

¹¹ See, for example, Vinogradov (1923/1980).

1st pers. pronoun *azŭ* ‘I’ and the archaic Aorist *molixŭ* ‘implored’ match the form *ego* in (6). The contrast between *tělo evo* ‘his body’ and *sogrešenie ego* ‘his sin’ (8) is especially striking. (Note also the Church Slavonic Gen. *-ago* in the adjectives *Svjatago* ‘Holy’ and *istinnago* ‘true’, and the participles *životvorjaščago* ‘life-creating’ and *poklonjaemago* ‘worshipped’ in (7).)

- (6) *azŭ molixŭ ego*
I implored him-ACC
‘I implored him’ (10)
- (7) *i v Duxa Svjatago, Gospoda, istinnago i životvorjaščago, světa*
and in Spirit Holy Lord true and life-creating light
*našego, věruemŭ, so Otcemŭ i Synomŭ poklonjaemago, za nego*¹²
our believe with Father and Son worshipped for him-ACC
že stražemŭ i umiraemŭ, pomoščiju Ego Vladyčneju
part. suffer and die help His-GEN Master’s
‘and we believe in the Holy Spirit, the Lord, the true and life-creating one, our light, worshipped with the Father and the Son; for Him we suffer and die, with his, our Master’s’, help’ (1)
- (8) *i my so vладыкожу prikazali tělo evo sredi ulicy*
and we with His Grace instructed body his-GEN in the middle street
sobakamŭ brositŭ, da že graždane opláčjutŭ sogrešenie ego
dogs throw so that part. townsfolk mourn sin his-GEN
‘and His Grace and I ordered to throw his body to the dogs, in the middle of the street, so that the townsfolk would mourn his sin’ (20)

Emotion, pity in particular, plays a role in Avvakum’s adherence to the colloquial register ((7)-(9)), and correspondingly, to the variant *evo*.

- (9) *i mne evo stalo žalŭ*
and I him became pity
‘And I felt pity for him’ (35)
- (10) *Ja, zaplakavŭ, blagoslovilŭ evo gorjuna*
I having cried blessed him poor thing
‘I, having burst into tears, blessed him, the poor thing’ (51)
- (11) *Milenŭkojmoj! boitsja Boga, sirotinŭka Xristova: ne pokinetŭ evo Xristosŭ!*
Darling my fears God orphan Christ’s not abandon him Christ
‘My darling! fears God, Christ’s little orphan: let Christ not abandon him!’ (53)

However, register harmony is clearly not the only factor in the spelling variation. Contrast (12) with (6)-(11) above. (Note also the substantivized adjective *mertvova* ‘dead’ with *-ova*.)

¹² Already in the first Slavic records, such as OCS, the epenthetic *n-* is added to the anaphoric pronoun when preceded by prepositions. See Ferrell (1958) for a discussion of the 3rd pers. pronoun in Slavic in relation to prepositions.

- (12) *a ja nočtju, vstavŭ, pomolja Boga, blagoslovja*
 and I at night having risen having prayed God having blessed
evo mertvoova, i s nimŭ pocelovaosja, opjatŭ podle
 him-ACC dead and with him having kissed again next
ego spatŭ ljaŭu
 him-GEN sleep lie
 ‘and at night I, having risen, having prayed to God, having blessed him, the dead one,
 and having kissed him, would again go to bed next to him.’ (73)

2.2 Analysis of the data

From the pattern of distribution of *evo* vs. *ego* in the *Life of Avvakum*, it emerges that Acc. attestations (most of which are direct objects) exhibit the highest frequency of the non-orthographic form *evo* (93%). Attestations of the Gen. as direct object of certain verbs (1 *ego* vs. 1 *evo*), and in negated constructions (0 *ego*, 7 *evo*) are too small a number to be taken into account, except that perhaps the predominance of *evo* in the Gen. direct object is influenced by the variant’s predominance in the Acc. with the same syntactic function. The prevalent usage of the Gen. is in possessive constructions. Interestingly, the ratio of possessive Gen. attestations of *evo*, although still in the majority (65%), is almost 30% lower than Acc. direct object in this text, although the total number of Gen. attestations (31) is also considerably lower in comparison to the Acc. direct objects (77), and thus may not be as indicative of a trend.¹³

Table 2. Distribution of *ego* vs. *evo* in the *Life of Avvakum*

	<i>ego</i>		<i>evo</i>	
Acc. (Direct object)	5	7%	66	93%
Gen. (Possessive)	11	35%	20	65%
Total	16	16%	86	84%

The question arises why the distribution of *evo* in the Acc. and in the Gen. is uneven and why the ratio is considerably lower in the Gen. First of all, it is, of course, conceivable that Avvakum actually pronounced the ending with [g] at times, because, as a priest, he would have been accustomed to reciting religious texts and to sermonizing, which he most certainly did with archaic Church Slavonic pronunciation. Secondly, *ego* may also sometimes be a purely graphic phenomenon. In fact, this emerges quite clearly from the two variant spellings of the Gen. and Gen.-Acc. ending in adjectives and participles in the same text. With absolute consistency Avvakum spells the older variant as the non-native OCS reflex *-ago* (as in (6) above), never as *-ogo* (which would be the expected East Slavic form, the result of the analogy to *togo*), and the newer one with as the native reflex *-ovo/-ova*¹⁴ (11).

¹³ As objects of prepositions, both Acc. and Gen. have still fewer attestations of *evo*, and the ratio is approximately half for Gen. (9 *ego*, 10 *evo*) and 1/3 for Acc. (6 *nego*, 3 *nevo*). Again, the total attestations are not many.

¹⁴ The choice between these two seems to be dependent on the stress.

Yet, why is Avvakum freer in employing the non-orthographic version in the Acc. than he is in the Gen., in both the adjectival forms and in the anaphoric pronoun? Perhaps the Gen. of the anaphoric pronoun used for expressing possession, an indeclinable form, by his time disconnected in the minds of the speakers from the paradigm of the anaphoric pronoun, began drifting away from its original paradigm towards that of the possessive pronouns before the change *-ogo* > *-ovo* was over.

Regarding the possible mechanism of the substitution of *ego* by *evo*, it may have occurred due to the association of the personal pronoun with the demonstrative *togo*, on the one hand, and the role that the latter played in the restructuring of the compound adjectival declension in East Slavic on the other. Recall that the juxtaposition of the personal pronoun *onŭ* and a proper name or a personal noun was a typical usage for the period, including in Avvakum. The suppletion in the Nom. of the personal pronoun by the demonstratives *onŭ* or *toj* imported a certain demonstrative value to it, as already mentioned. This, of course, concerns the whole paradigm. The following is an example. (See also *evo gorjuna* 'him, the poor thing' (10), or *evo mertvova* 'him, the dead one' (12) already cited – the latter being a substantivized adjective, all with personal reference.)

- (13) *udavili evo Feodora na Mezeni*
 strangled him Feodor by the Mezen'
 they strangled this Feodor by the (river) Mezen' (50)

Thus the Acc. and the Gen. *evo* may be due to the influence of *tovo* which, based on its morphological connection with the adjectives, may have undergone the change from *togo* soon after they did. Note in this regard that the attestations of *togo* in this text are infrequent (5, 11%), *tovo* being the overwhelming majority (40, 89%).

One of the most peculiar features of the *Life of Avvakum* is the declinable post-posed demonstrative *tŭ*, akin to a definite article, that can be added to nouns, adjectives and, on occasion, even infinitives of verbs (Cocron 1962:141-142). For example:

- (14) *Kakŭ doščeníkŭ-otŭ v vodu-tu ne pogrjazŭ so mnoju?*
 how raft-that-NOM into water-that-ACC not sunk with me
Stalo u menja v tě pory kosti-te ščemitŭ
 began at me at those times bones-those-ACC to squeeze
i žily-tě tjanutŭ, i serdce zašlosŭ, da i umiratŭ
 and sinews-those-ACC to pull and heart fluttered well and to die
stalŭ.
 began

'How did that raft not sink in that water together with me? At that time those bones of mine began aching and those sinews of mine began contracting and my heart fluttered and, well, I began to die.' (23)

According to Cocron (1962:144), attestations of this post-posed demonstrative are particularly frequent in the Acc. and the Gen. cases. The post-posed demonstrative *tŭ* had the same Gen. and Acc. as the demonstrative *toj* (3c) related to it. In (15) it appears with a variant of the same proper name as in (13) above.

- (15) *u* *Fedora* *tovo*
 at Fedor-GEN that
 ‘that Fedor had’ (56)

All the factors just described may have contributed to the substitution of *evo* for *ego* in the Gen. and Acc. of the anaphoric pronoun. However, if, as has been suggested, the possessive Gen. was at this time disconnecting from its source anaphoric paradigm, it may have been slower at adopting the new ending. This gradual division of a single anaphoric pronoun into two distinct pronouns, personal on the one hand and possessive on the other, is indicated by several developments observed in the history of Russian. The syntactic difference between the two may be motivating the separation. The personal pronoun is a substantive, and the possessive pronoun an attributive modifier.

Like all personal pronouns, the 3rd pers. pronoun exhibits the suppletion in the Nom. (masc./neut. sg. *on/ono*, fem. sg. *ona* & pl. *oni*) completing its declensional paradigm. Another typical characteristic of personal pronouns is Gen.-Acc. syncretism. This is true of *ego* as well, despite the fact that, in contrast to the 1st & 2nd pers., the 3rd pers. is the only one that may refer to both animate and inanimate objects. As a possessive pronoun, the Gen. *ego* continues to exhibit signs of innovation in the direction of a further morphological adaptation to the possessive pronoun paradigm. The pressure to become an attributive form conforming to the other possessive pronouns is evident in dialectal and children’s speech: sg. [(j)ivónyj] for *ego* ‘his’, [jéjnyj] for *ee* ‘her’, pl. [ixniij] for *ix* ‘their’, etc. (Šaxmatov 1941:495), modelled on the sg. *moj*, *tvoj*, *svoj* (the latter in complementary distribution with the 3rd pers. possessive pronoun). The less integrated pl. are under similar pressure, e.g.: [nášij] for *naš*, [vášij] for *vaš*. Moreover, the fact that, in contrast to the Gen. and Acc. of the homonymous personal pronoun, the epenthetic *n-* is not added to the possessive pronoun after prepositions in Russian, further speaks to the lack of identity between the personal and the possessive 3rd pers. pronouns. Contrast *u nego* ‘he has’ and *u ego sestry* ‘his sister has’. Previously this distinction between the personal and the possessive usage of *ego* was made by means of word order. The possessive Gen. was originally post-posed to the head noun it modified, and was thus separated from the preposition (Ferrell 1958:80). This is indeed the most typical usage in the Life of Avvakum, for example:

- (16) *gortani ego* ‘his throat’ (10)

Avvakum places *ego* pre-positively only once out of 11 total attestations, and *evo* 4 times out of 20. This correlation of the older form with post-position and the younger one with pre-position indicates a growing trend towards pre-positive usage of possessive forms – which is, of course, the situation in modern Russian.

3 CONCLUSION

Putting stylistic and orthographic factors aside, it appears that Avvakum is more liberal with the spelling *evo* in the Acc. case (the overwhelming majority of which serve as direct objects), but is more conservative with the same graphic variant for the Gen. (the majority of which are possessive). The difference in the extent of the spelling *evo* suggests that Gen. of possession and Gen.-Acc. of the anaphoric pronoun, although belonging to the same paradigm, aren't treated as the same form by Avvakum. It may be concluded that different relative frequency of *ego* and *evo* in the Acc. and in the Gen. has to do with the divergence of a single pronoun into two separate ones. The history of the Common Slavic pronoun *jego* reflects a shift from the ubiquitous anaphoric pronoun **ji* at the advanced stages of grammaticalization already in the earliest historical records, to a full-fledged 3rd pers. personal pronoun on the one hand, and a homonymous possessive pronoun on the other. The two alternative spellings (*ego* & *evo*) in the *Life of Avvakum* manifest the emerging difference in two pronoun class allegiances of the Russian *ego*.

REFERENCES

- Zhitie protopopa Avvakuma, im samim napisannoe [The Life of Archpriest Avvakum, written by himself]. Izdanie Imperatorskoj Arxeografičeskoj Kommissii. Petrograd, 1916.
- Bratishenko, Elena. 2010. Tracing the Genealogy of the Gen. -ovo in Russian Adjectival Declension, *Russian Linguistics* 34/1.67-86.
- Černyx P. Ja. 1927. Žitie protopopa Avvakuma, im samim napisannoe, kak pamjatnik severnovelikoruskoj reči XVII-go stoletija, Očerki po istorii i dialektologii severnovelikoruskogo narečija, 1. Irkutsk: Izdatel'stvo Irkutskogo gosudastvennogo universiteta.
- Cocron, Friedrich. 1962. La langue russe dans la seconde moitié du XVII-e siècle (Morphologie). Paris: Institut d'études slaves de l'université de Paris.
- Ferrell, James. 1958. The systems of the third person pronoun with particular reference to the category of the preposition in Slavic Languages (American Contributions to the Fourth International Congress of Slavists, Moscow, September 1958), 77-100. 's-Gravenhage: Mouton & Co.
- Flier, Michael. F.: 1983. The origin of the desinence -ovo in Russian, in Markov, Vladimir and Dean S. Worth (eds.), *From Los Angeles to Kiev* (Papers on the occasion of the Ninth International Congress of Slavists), 85-104. Columbus, Ohio.
- Klenin, Emily. 1983. Animacy in Russian: A New Interpretation. Columbus, Ohio: Slavica Publishers.
- Kuznecov, P. S. 1959. Očerki istoričeskoj morfologii ruskogo jazyka. Moscow: Izdatel'stvo Akademii nauk SSSR.
- Lunt, Horace. 1974. *Old Church Slavonic Grammar*. 6th ed. (revised). The Hague-Paris: Mouton.

- Malinowski, L. 1875. Über die endung des genitiv sing. masc.-neutr. der pronominalen und zusammengesetzten declination im russischen und kaschubischen. Beiträge zur vergleichenden Sprachforschung auf dem Gebiete der indogermanischen Sprachen, 8/3.356-362.
- Meillet, A. 1965. Le slave commun. 2nd ed. (revised and enlarged). Paris: Champion.
- Timberlake, Alan. 1995. Avvakum's aorists, *Russian Linguistics*, 19.25-43.
- Tolkačev, A. I. 1960. Ob izmenenii -ogo > -ovo v roditel'nom padeže edinstvennogo čisla mužskogo i srednego roda člennyx prilagatel'nyx i mestoimenij russkogo jazyka, *Materialy i issledovanija po istorii russkogo jazyka*, Avanesov R. I. (ed.), 235-267. Moscow: Izdatel'stvo Akademii nauk SSSR.
- Šaxmatov, A. A. 1941. Sintaksis russkogo jazyka. 2nd ed. Leningrad: Učpedgiz.
- Vaillant, André. 1958. Grammaire comparée des langues slaves. Vol. 2. Part 2: Flexion pronominale. Lyon: IAC.
- Vinogradov, V.V. 1923/1980. O zadačax stilistiki: Nabljudenija nad stilem Žitija protopopa Avvakuma, *Izbrannye trudy: O jazyke xudožestvennoj prozy*, 3-41. Moscow.
- Worth, Dean. 2000. Avvakum as az and ja, *Južnoslovenskij filolog* 56/1-2.227-233.

Anna Frassanito

The case of ‘uninflected’ infinitive in the Griko dialects of Sternatia and Calimera

1. FINITE COMPLEMENTATION IN GRIKO DIALECTS: AN OVERVIEW

Aim of this paper is to investigate the categorial status and the main properties displayed by the so called ‘fossilized’ (uninflected) infinitive in the Griko varieties of Sternatia and Calimera. The first section of this contribution is devoted to the morphological analysis of the Griko uninflected form along the lines of the unified model of morphosyntax as illustrated in Manzini-Savoia (2005 ff.). The second and final section focuses on the syncretism of Griko ‘fossilized’ infinitive with other forms of the paradigm (third person singular present perfect, second person singular negative imperative). The salience of the current study lies in a different approach to phenomena of syncretisms: the ability of the Griko infinitive to serve as a number of different forms of the paradigm ultimately depends on interpretive processes taking place at the LF interface, to the exclusion of unnecessary machinery postulated by DM and the OT.

As it is widely known from the literature (Rohlf's 1969; Joseph 1993), both M(odern) G(reek) (henceforward MG) and Griko dialects of the Salento region lack infinitive. In control contexts, Romance languages like English or Italian (1a-b) have recourse to a morphological infinitive; this is not true of both MG and Griko. In contexts such as the control ones MG and Griko dialects lack infinitival clauses as in (1c-d); instead, a finite form of the verb shows up preceded by the particle *na*¹. The occurrence of the ‘particle+finite verb’ sequence where an infinitive is needed can also be found in Balkan languages (Arbëresh here) and in some Southern Italian Romance varieties (Calabrian/Salentino). The particles vary according to the different languages; they surface, respectively, as *të* for Albanian and Arbëresh (1e), as *mu/ma/mi* for Calabrian (1f) and as *ku/ka* in Salentino (1g):

- | | | | |
|-----|---|-----------------------------|------------------|
| (1) | a | I left not to see him | English |
| | b | Me ne andai per non vederlo | Italian |
| | | I.left for not to see-him | |
| | c | efija na mi(n) ‘t:ɔ blepɔ | M(odern) G(reek) |

¹ Due to space limitations, I am not dealing with the categorial status of *na* here. For further discussion on the position occupied by the ‘subjunctive’ particle, see Roberts and Roussou (2003) and Frassanito (2008).

	I.left	Prt	not	him	I.see				
d	ε	'pirta	na	mi(n)	't:ɔ	't:ɔ:risɔ	Griko: <i>Calimera</i>		
	I.left	Prt	not	him	I.see				
e	ika	mɔs	t	ε	`fikja		Arbëresh: <i>Firmo</i>		
	I.left	not	Prt	him	I.saw				
f	'mi	ndi	'jivi	(pa)	nɔ	mm(i)	u	'viju	Calabrian: <i>S. Agata del Bianco</i>
	I.left		(for)	not	Prt	him	I.see		
g	mε	nda	ji:	ku	nu	lu	'viju		Salentino: <i>Sogliano Cavour</i>
	I.left		Prt	not	him	I.see			

A complete list of the control contexts requiring *na* followed by a finite VP is given in (2a-e); the *na*-inflected verb construction shows up in combination with modals as in (2a), with subject control verbs as in (2b) and with aspectuals as in (2c), as well as with complements to causatives as in (2d) and in adverbial clauses as in (2e).

- (2) a ε 't:ɛlɔ na paɔ
not I.want Prt I.go
'I do not want to go' Griko: *Sternatia/Calimera*
- b Mu 'pε n 'ɔ 'v:risɔ
to.me he.said Prt him I.look for
'he said to me to look for him'
- c ɛntʃ iɲasa n i milisɔ
I.began Prt to.her I.talk
'I began to talk to her'
- d ε 't:ɔ(n) ɛfika na plɔsi
not him I.let Prt he.sleeps
'I did not let him sleep' Griko: *Sternatia/Calimera*
- e ɛpir'tanɛ na 'mi(n) 't:ɔ 't:ɔ:risɔnɛ
they.left Prt not him they.see
'they left not to see him'

1.1 The 'infinitive'

The examples in (1c) for MG as well as the examples in (1d) and (2a-e) for Griko show that a morphological infinitive is ruled out in control contexts. Yet Griko dialects of Sternatia and Calimera attest to the existence of an apparent counterexample; in combination with *potere* (English *can*; *sɔzɔ* in Griko) a 'fossilized' infinitive (Rohlf's 1969; Joseph 1993; Italia-Lambroyorgu 2001) shows up. Note that the subset of predicates involved (verbs of ability) undergo 'restructuring' phenomena in the sense of Rizzi (1982). Evidence in favour of the complex predicate formation comes from inflectional-level material: clitics lexicalizing argument of the embedded verb surface as complement to the matrix verb, giving rise to 'clitic-climbing' (3a). In contexts of the type in (3a) a lexicalization of the lower verb by means of an inflected predicate yields the ungrammatical configuration in (3b).

- (3) a ε 't:ɔ sɔzɔ rɔdisi/arɔtisi Griko: *Sternatia/Calimera*
 not it I.can ask-INF
 'I can not ask it'
 b *ε sɔzɔ na 't:ɔ rɔ'tɔ
 not I.can Prt it I.ask
 'I can not ask it'

While the inflectional morphology of the finite verb in (1)-(2) varies according to the embedded EPP, the specialized inflection *-i* of the 'fossilized' infinitive is invariable. The infinitival form is always associated with a *-i* ending whatever the agreement properties of the embedded EPP (4a-b).

- (4) a '(T):ɔ sɔzɔ 'fruntɛtsi/ankuntrɛtsi av:ri Griko: *Sternatia/Calimera*
 him I.can meet-INF tomorrow
 'I can meet him tomorrow'
 a' * '(T):ɔ sɔzɔ 'fruntɛtsɔ/ankuntrɛtsɔ av:ri
 him I.can I.meet tomorrow
 b 'ε sɔzɔne plɔsi
 not they.can sleep-INF
 'they can not sleep'
 b' * 'ε sɔzɔne plɔsɔne
 not they.can they.sleep

1.2 Cross-linguistic evidence: the case of Calabrian and of the Geg Albanian *paskajore*

Invariable verb forms of the type illustrated for Griko can be found in Calabrian as well as in a number of Geg varieties of Albanian. The Calabrian varieties of Seminara show the same distributional patterns as their Griko counterparts: a morphological infinitive is selected in combination with restructuring modals and causatives (see 5a-b)².

- (5) a lu 'puɛttsu 'harɛ Calabrian: *Seminara*
 it I.can do-INF
 'I can do it'
 b u 'fattsu 'dɔrmiri
 him I.make sleep-INF
 'I make him sleep'

In control contexts other than restructuring ones, a '*mmi/mi*+finite verb' sequence is attested. The relevant contexts for Seminara are the ones exemplified for Griko in (2a-e), i.e. modals (6a), aspectuals (6b), subject and object control verbs (6c), causatives (6d) and adverbial sentences (6e).

² All relevant data and assumptions are from Manzini and Savoia (2005).

- (6) a 'vɔʃʃu pɛ mmi 'vɛni dɔ'mani Calabrian: *Seminara*
 I.want for Prt he.comes tomorrow
 'I want him to come tomorrow'
- b 'idi fi'niʃʃunu pɛ mmi 'maɲdʒanu
 they they.finish for Prt they.eat
 'they finish eating'
- c mi 'ðissuru mi 'vɛɲɲu dɔ'mani
 to.me they.told Prt I.come tomorrow
 'they told me to come tomorrow'
- d u fitʃi pɛ mmi/mi ðɔrmi
 him I.made for Prt he.sleeps
 'I made him sleep'
- e 'vinni pɛ mmi ti viði
 I.came for Prt you I.see
 'I came to see you'

An invariable verb form is displayed in Geg Albanian control contexts too. Where standard Albanian selects for a fully inflected embedded verb with *të* (1e), Geg dialects have recourse to the preposition *mɛ* (with) followed by an invariable verb form; such an uninflected participle is known as *paskajore*, 'infinitive' (Demiraj 1985, 1997; Banfi 1985; Joseph 1993; Pellegrini 1995; Manzini-Savoia 2002). However, unlike Griko and Calabrian, the syntactic construct of Geg Albanian is not sensitive to restructuring phenomena; the *paskajore* occurs in all control contexts, irrespectively of the semantic/syntactic properties of the matrix verb. Thus the *mɛ*-participle structure is found with modals (8a), with aspectuals (8b), with subject/object control verbs (8c-d) as well as with adverbial sentences (8e).

- (8) a s 'munɛm mɛ ɛ 'bɔ̃ Geg Albanian: *Shkodër*
 not I.can to it do
 'I can't do it'
- b kam fi'tu mɛ 'hɔɲɛr
 I.have begun to eat
 'I have begun to eat'
- c i kam 'θɔ̃:n m u m'lu
 to.him I.have said to M/R³ cover
 'I told him to cover up'
- d i kan ' θɔ̃:n (tʃi) (mas) mɛ ɛ 'bɔ̃
 to.him they.have said that not to it do
 'they told him not to do it'
- e kam 'v:rð mɛ t a 'ðɔ̃:n
 I.have come to to.you it give

³ Read middle-reflexive.

‘ I came to give it to you’
 f ε kam ‘bũ m u ‘tʃu
 him I.have made to M/R wake.up
 ‘ I have made him wake up’

The mε-uninflected verb sequence equally occurs in causative contexts too (8f); in a nutshell, the contexts which require the *paskajore* are the core contexts for infinitives in Romance languages. A schematic summary of the distributional pattern of the ‘infinitive’ in the three dialects of *Sternatia/Calimera*, *Seminara* and *Shkodër* is given in (9).

(9) Occurrences of the ‘infinitive’ in the dialects of *Sternatia/Calimera*, *Seminara* and *Shkodër*.

	Control contexts	Verbal classes	Prepositions
<i>Sternatia/Calimera</i>	Restructuring	Modals(ability)/causatives	None
<i>Seminara</i>	Restructuring	Modals/Causatives	None
<i>Shkodër</i>	Restructuring/ non-restructuring	All	mε/pv/tu(i)

2 SYNCRETISMS: GRIKO AND MODERN GREEK

It is also a property of both a number of Calabrian and Geg Albanian varieties that they show a non-finite form in control contexts. The relevance of this cross-linguistic comparison lies in the morphological similarities between the verb forms involved: in all the languages examined a lexical (verb) root selects for an invariable ending which varies according to the different dialects. This invariable ending coincides, in turn, with a vowel in the case of *paskajore* (see 8a-f), with the familiar *-re* inflection for Calabrian (5a-b) and with *-si* in Griko varieties (4a-b)⁴.

The lexical entry found as Griko ‘fossilized’ infinitive (3a-b) coincides with both the present perfect (10a) and the perfect participle (10b) in MG; the present perfective forms in (10a) show up in embedded sentences only, where they follow *na*.

(10)	a (na)	di-s-i	Modern Greek
	(Prt)	he.dresses (perfect)	
	b exo	di-s-i	
	I.have	dressed	

⁴ This schema is to be intended as an oversimplification: the formation of the invariable participle varies according to the consonantal/vocalic ending of the verb stem in Geg Albanian. Verbal bases ending in a vowel will select for a thematic vowel-like element; instead, the participle formation simply involves a lengthening of the stressed nucleus in the case of consonantal bases: see, for example, ve: • -participle, long nucleus- as opposed to the bare stem ve •. The formation of Griko ‘fossilized’ infinitive also proves more complex than above; further details on the morphological structure of both Griko/MG and Geg Albanian will be given in the following sections.

A syncretic pattern of the type found for MG in (10a) shows up in Griko dialects too. Like MG, the ‘fossilized’ infinitive of Griko does not extend to all persons of the present perfect; though, the two Greek varieties differ as to the forms which admit of a syncretic lexicalization. These include third person singular only in MG, both second and third person singular of the present perfect in Griko. The relevant paradigm is given in (11a) below.

- (11) a (na) rɔd-i-s-o, -s-i,-s-i,-s-ome,-s-ete,-s-one Griko: *Calimera/Sternatia*
 (Prt) ‘I ask, you ask, he/she asks, we ask, you ask, they ask’
 b (na) t:ɔrisi
 (Prt) you.see/he.sees

As for MG, Griko present perfective forms are characterized as ‘dependent’ ones.

With the second person singular negative imperative (12a-c), Griko ‘fossilized’ infinitive shows up as a suppletive form for the negative imperative as illustrated in (12a’-c’)⁵.

- (12) a min kli’si ti ‘ppɔrta Griko: *Sternatia*
 not close the door
 ‘Do not close the door!’
 a’ kli’sɛ ti ‘ppɔrta
 close the door
 ‘Close the door!’
 b min arɔ’tisi/rɔ’disi
 not ask
 ‘Do not ask!’
 b’ a’rɔtiso
 ‘Ask!’
 c min tɔ’risi
 not watch
 ‘Do not watch!’
 c’ ‘tɔriso
 ‘Watch!’

⁵ Suppletion in the negative imperative is not an isolated property of Griko dialects; independent evidence comes from Romance languages like Italian and also some Southern Italian varieties. Unlike positive imperative, which is endowed with dedicated forms, negative imperative has recourse to suppletive forms. The suppletive form which is typically used to negate a ‘true’ imperative coincides with an infinitive in both Italian and Southern Italian dialects, as it is the case with Griko. The relevant examples for Italian are given in (13a-b) below.

- (13) a Canta!
 ‘Sing!’
 b Non cantare!
 Not sing-INF
 ‘Don’t sing!’

nominal or verbal. If so, the lexical item *l-* in (18) is endowed with predicative content only. What is, then, the role played by the thematic vowel *-v* here? The model proposed by Manzini-Savoia (2005 ff.) allows us to easily account for the participial structure in (18). The insertion of the thematic vowel *-v* under a N node, where the former lexicalizes nominal class specifications, derives directly from their unified model of morphology and syntax. According to Manzini and Savoia, the morphological component strictly mirrors the categories and hierarchies holding for the syntax: the morphological structure of the verb within which the inflection inserts is a word-internal reduplication of the syntactic structure hosting a subject⁸. Manzini and Savoia will take the inflection inside the verb to have the same status as the clitic subject within the sentence; as to the relevant status, Manzini and Savoia label it as a D category according to the common assumption (Chomsky 1995) that the sentential subject is an instantiation of denotational (D-type) properties. What is relevant for the analysis of the Geg Albanian participle in (18), though, is the role performed by the vowel *-v* in N. As it is widely known, the predicative content of a sentential VP includes at least two arguments: the external argument (sentential subject) and an internal argument corresponding to the object in the sentence structure. Manzini and Savoia assume the sentential object to lexicalize N(ominal) class properties (gender specifications); hence they label it as N. Based on the parallelism between the predicative content of the finite verb and of the perfect participle and due to the agreement pattern of the perfect participle in Romance languages like Italian⁹, Manzini-Savoia treat the *-v* morpheme in (18) as a word-level lexicalization of the internal argument of the Geg Albanian participle. If so, the categorial status of the thematic vowel in (18) will be the same as the one of the sentential object, i.e. N. The specialized *-v* inflection in N is, then, said to provide a low level lexicalization of the internal argument of the predicate. The ability of the thematic vowel to insert under a N node allows for a nominal characterization of the invariable participle in (18); hence, the availability of the participle *lv* in ‘infinitival’ contexts.

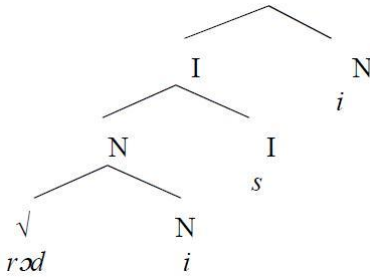
3.1 The MG perfect participle/Griko ‘fossilized’ infinitive

MG/Griko invariable bases are superficially more complex than their Geg Albanian counterparts. The comparison between the Albanian participle in (18) and the Griko invariable form *rɔdisi* in (19) shows that the ‘verb root-thematic vowel’ sequence combines with a *-s* ending and an extra *-i* inflection in Griko.

⁸ The set of arguments projected by the sentential VP includes two slots: one for the subject as the external argument and the second one for the object as a saturation of the internal argument of the verb.

⁹ As it is widely known from the literature (Burzio 1986, Kayne 1989a) the perfect participle of Italian shows the same gender/number specifications as its sentential object as sentences like ‘*la hanno vista*’ exemplify. Note that the thematic vowel of Geg Albanian in (18) is not exactly the same as the agreement morphemes of fully inflected participles; vocalic elements like *-v* in (18) rather lexicalize reference to an indefinite variable whose value is contributed by the sentential internal argument.

(19) Calimera



As we saw in section 3, the vocalic bases of Geg Albanian form the participle by combining a category neutral verb root with a thematic vowel as in (18). Following Manzini-Savoia (2007), we associated the thematic vowel with a word-internal N position, where it lexicalizes the internal argument of the participle. If we extend this analysis to *rɔdisi*, we can easily account for the status of the *rɔdi* subconstituent in (19). Based on the parallel with Geg Albanian in (18), the lexical base *rɔd* is associated with predicative properties only; the nominal character of the lexical base stems from the insertion of *-i* under a lower N node. As to the Griko *-s* morphology; we can analyze the specialized *-s* ending in (19) as a bearer of perfective properties. It is not an isolated property of the perfect participle of Griko/MG to overtly lexicalize aspectual specifications; the treatment of the *-s* morphology as an overt marker of perfectivity is independently motivated for the inflected participles of both Italian and Albanian¹⁰. According to the analysis provided by Manzini-Savoia (2005, 2007) for both Romance and Geg Albanian participles, I will assume the *-s* element to project a perfective I head position; as such it takes as its complement the lexical base inclusive of the thematic vowel in N. The latter will, then, insert to the immediate left of the aspectual I head (*-s* here) as a result of the right-branching nature of the morphological component. The word-final *-i* ending becomes associated with a higher N node internal to the participle as in (19). The Griko/MG higher *-i* is not an agreement inflection: the absence in *-i* of gender/plurality specifications allows for the participial form of Greek-based varieties to occur in all participial/infinitival contexts; this is not the case with the inflected participles of both Italian and Geg Albanian. Indeed, the inflected participles are excluded from the *paskajore*; the lexicalization of the infinitive is restricted solely to the uninflected participial forms in (18) in Geg Albanian (see 8a-f). As we have seen before, the invariable verb form in (19) serves both as the perfect participle in MG and it surfaces as an infinitival form in control contexts of Griko. The compatibility of (19) with two different interpretations, namely participial/infinitive, will depend on the semantics of the vowel *-i*. Following Manzini and Savoia's assumptions on the thematic vowel of (Geg) Albanian invariable participles in (18), we will identify the *-i* of Griko/MG with a N element introducing an indefinite (free) variable. As such, the nominal form of Griko/MG in (19) will surface, alternatively, as either a participle or an infinitive according to the way we bind the indefinite variable at LF. When the variable introduced by the thematic vowel is bound by an antecedent (the sentential object here) it

¹⁰ I refer the reader to Manzini-Savoia (2007) for a detailed morphological analysis of the participial formations above.

introduces the N sentential argument as a variable, thus allowing for the emergence of (19) in MG participial contexts.

4. SYNCRETISMS: THEORETICAL BACKGROUND

Syncretic phenomena are treated in a number of different morphosyntactic approaches; these include the D(istributed) M(orphology) (henceforward DM; see Halle & Marantz 1993,1994) and Optimality Theory (OT here; see Grimshaw 2001) among others. According to these approaches, the exponent of a morpheme and its feature specifications are not bundled together in the lexicon; as a result, terminal nodes are associated with bundles of features deprived of any phonological content in the syntax. Vocabulary insertion only takes place at the end of the morphosyntactic derivation, hence its characterizations as 'Late'. In accordance with the above discussion, syncretic phenomena simply involve a manipulation of the terminal strings in DM and the OT; terminal elements undergo a modification of their feature specifications as a result of special processes operating at the M(orphological) S(tructure). The latter qualify as specialized MS rules for DM; instead, the OT representation will follow from a marked hierarchy of constraints. Against both DM and OT and in accordance with Manzini-Savoia (2005 ff.), I will argue for a early insertion and an *interpretive* characterization of syncretism. A syntactic approach of this type is perfectly in line with minimalist postulates; following Chomsky (1995), syntactic structures are projected directly from lexical material. Lexical items are, then, overtly and entirely instantiated in the syntax (what is referred to as 'early' insertion) in this lexicalist model. This has important consequences for the analysis of syncretisms; lexical elements are merged because of their positive specifications in the current approach. The latter allow for interpretive processes at LF to associate the syncretic element with the relevant range of interpretations. An argument in favour of this theory comes from the distribution of the syncretic patterns in both Griko/MG and Albanian; as we have seen above, the ability of the uninflected form in (18) and (19) to emerge is restricted to more or less the same contexts in both Greek and (Geg) Albanian varieties. The relevant contexts include both the participle/infinitive and the perfect either in the present or in the past in all the varieties. Furthermore, the ability to select for the syncretic form is restricted to the second and the third person of the singular only. The DM and the OT frameworks would posit special principles whereby some of the input properties get lost in the derivation; an underspecified lexical item will, then, insert under terminal nodes accounting for syncretic phenomena. The recourse to abstract notions of 'underspecification' and extrinsic re-ordering is perfectly adequate to describe the facts. Nonetheless, there is no obvious reason why certain tense/aspect/person specifications should be preferred to the exclusion of the others if lexical insertion ultimately depends on language-specific rules. On the other hand, the restrictions observed in (14) and (17), far from being accidental, follow from the aspectual and modal properties of the uninflected verb bases in our model.

4.1 The syncretisms of Geg Albanian *paskajore*

The (Geg) Albanian *paskajore* (vocalic bases) constitutes a lexicalization of both the invariable perfect participle and the infinitive, as well as serving as the 3rd person singular of the past perfective middle-passive. When embedded under the auxiliaries *jam/kam* it is a perfect participle, while when it is preceded by prepositions like *mε/pv/tu* it serves as the *paskajore* proper. When it associates with a specialized *u* morphology in (15c) it provides a lexicalization of the middle-reflexive in the perfective 3rd person. According to Manzini-Savoia (2007), the ability of the verb stem in (18) to associate with a perfective interpretation depends on the nominal character of the verb base. Recall from above that the thematic vowel-like element in (18) constitutes lexicalization of a N denotation internal to the participle. They assume this low-level N specification to correspond to a word-internal direct object. It is precisely this nominal character of the verb stem that allows for a perfective reading (nominal predicates are typically endowed with stative properties). As to the temporal reference, the past here, it can be inferred on the basis of the perfective aspect. We still have to account for the compatibility of the verb stem with the third person denotation in (15c). The impossibility of the vocalic ending in (18) to lexicalize D-type properties accounts for its restriction to third person only; the availability of bare verb stems in (18) to third person singular interpretation and only to it directly follows if discourse-linked referents (1st/2nd person) as well as plural (cf. 3rd plural) require a D inflection. Let us now turn to the syncretisms in Greek-based varieties.

4.2 The MG perfect participle and the Griko ‘fossilized’ infinitive: syncretisms with finite forms

The invariable verb form in (19) serves as either the perfect participle in MG or the ‘infinitive’ in Griko. It also lexicalizes the second/third person singular (Griko) or the third person singular only (MG) of the present perfect in both dialects and it finally occurs as a suppletive form for the positive imperative in Griko; as such, it lexicalizes the second person singular negative imperative (see 12a-c). Despite the striking regularities, Griko/MG invariable base in (19) and its Albanian counterpart in (18) differ as to the lexicalization of aspectual properties. The association of the bare verb base with a perfective interpretation is inferential in Albanian. On the contrary, aspectual specifications are overtly lexicalized in the verb bases of both Greek varieties as a specialized *-s* morphology. Independent evidence in favour of a treatment of the *-s* ending as an aspectual marker comes from the inflected participles of both Italian and Albanian. The ‘verb base-thematic vowel’ sequence combines with a language specific aspectual marker followed by participial agreement morphology in both Italian and Albanian; the aspectual ending surfaces as a *-t* element in Italian regular participle (see *mangia-t-o* ‘eaten’) and as either a *-m* (in case of vocalic bases) or a *-un* morphology (for consonantal bases) in Albanian (see *mlu-m-ε* ‘covered’/ *vef-un-a* ‘dressed’).

The ability of the invariable verb in (19) to serve as a perfect participle in MG can be derived on the same basis as its Geg Albanian counterpart in 4.1. As for the infinitival reading associated with the ‘fossilized’ form of Griko in (4a-b), we will assume that the *irrealis* reading is inferred on

the basis of a scopal mechanism introducing a generic quantification over possible worlds at LF¹¹ (recall from Albanian that the thematic vowel simply lexicalizes the internal argument of the participle as a free variable). We still have to account for the occurrence of (19) as a lexicalization of either second/third or third person singular of the present perfective, respectively in Griko and MG. The perfectivity is contributed by the specialized –s morphology here, as it was for the perfect participle of MG. With respect to the restriction of the verb base to third person singular in both dialects, it can be accounted for in exactly the same way as we did for the perfective past of Geg Albanian. The emergence of the verb stem in (19) as a syncretic form for third person singular (present perfective) depends on the lack of subject agreement inflection in the latter. Nonetheless, unlike Albanian and MG, Griko ‘fossilized’ infinitive is also found as an instantiation of the second person singular (of the present perfect). This syncretism between second/third person singular is not an isolated property of Griko dialects; a person split opposing second/third person singular to the other forms is independently exemplified by Manzini and Savoia for the auxiliary selection processes in Central/Southern Italian Romance varieties. A coincidence between second and third person singular in contrast with the other forms also characterizes the present of consonantal bases in Geg Albanian (see 17b). Following Manzini and Savoia, we assume that second person singular is the core referent for the discourse-anchored set in exactly the same way as third person is for the complementary event-anchored set. Their salience within their own interpretive domain allows for their availability in the absence of any subject agreement inflection.

The set of interpretations covered by the pure verbal base extends to include second person singular negative imperative in Griko. The compatibility with second person singular has been already accounted for. As to the imperative interpretation, it is closest to the infinitival one in that they both imply *irreality*.

CONCLUSIONS

In the present paper we have considered the distribution and the properties of the uninflected infinitive of Griko dialects. Based on the parallelism with Albanian varieties (Geg Albanian and Arbëresh), we associated the thematic vowel of both Griko and MG with a word-internal lexicalization of N properties. We argued that it is precisely this nominal character of the Greek verbal stem that allows for the latter to serve as equally the perfect participle (in MG) and the infinitive (in Griko).

In the last section we focused on the syncretisms of the Griko ‘fossilized’ infinitive with finite forms of the verb. It was argued that we get a better understanding of the regularities in both Griko/MG and (Geg) Albanian syncretic phenomena by adopting Manzini and Savoia’s view of syncretism as ambiguity. Syncretic elements are merged because of their fully specified properties in our model; their aspectual, modal and temporal specifications follow from interpretive processes at the LF interface. An approach of this type provided us with a clearer

¹¹ For a detailed analysis of the connection between aspect/mood as a result of scopal mechanisms we refer the reader to Iatridou (2000) as to Modern Greek and to Manzini -Savoia (2007) for Geg Albanian.

picture of the data without postulating the existence of unnecessary machinery (MS rules in DM; language-specific constraints in the OT).

REFERENCES

- Banfi, E. 1985. *Linguistica Balcanica*. Bologna: Zanichelli.
- Burzio, L. 1986. *Italian syntax*. Dordrecht: Kluwer.
- Calabrese, A. 1993. The lack of Infinitival Clauses in Salentino: A synchronic analysis. *Syntactic Theory and the Dialects of Italy*, ed. by Adriana Belletti, 28-98. Turin: Rosenberg & Sellier.
- Chomsky, N. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Demiraj, Sh. 1985. *Gramatike Historike e Gjuhës Shipe*. Tirana: 8 Nëntori.
- Demiraj, Sh. 1997. *La lingua albanese. Origine, storia, strutture*. Rende: Centro Editoriale Librario dell' Università della Calabria.
- Frassanito, A. 2007-2008. *Costrutti dipendenti: l'assenza di 'infinito' nelle varietà grecofone di Sternatia e Calimera*. University of Florence: MA dissertation.
- Grimshaw, J. 2001. Optimal clitic positions and the lexicon in Romance clitic systems. *Optimality-theoretic syntax*, ed. by Jane Grimshaw and Sten Vikner. Cambridge, MA: MIT Press.
- Halle, M.; Marantz, A. 1993. Distributed Morphology and the pieces of inflection. The view from Building 20, ed. by Ken Hale and Samuel Jay Keyser, 111-176. Cambridge, MA: MIT Press.
- Halle, M.; Marantz, A. 1994. Some Key Features of Distributed Morphology. *Papers on Phonology and Morphology*, ed. by Andrew Carnie, Heidi Harley and Tony Bures. MIT Working Papers in Linguistics 21.275-288.
- Iatridou, S. 2000. The grammatical ingredients of counterfactuality. *Linguistic Inquiry* 31.231-270.
- Italia, G. G.; Lambroyorgu G. 2001. *Grammatica del dialetto greco di Sternatia (Grecia Salentina)*. Galatina: Congedo Editore.
- Joseph, B.D. 1993. *The Synchrony and Diachrony of the Balkan Infinitive*. Cambridge: Cambridge University Press.
- Kayne, R. S. 1989a. Facets of Romance past participle agreement. *Dialect Variation and the Theory of Grammar*, ed. by Paola Benincà. Dordrecht: Foris.
- Kayne, R. S. 2005. On the parameters and the Principles of Pronunciation. Online article, New York University.
- Kayne, R. S. 2006a. Expletives, Datives and the Tension between Morphology and Syntax. Online article, New York University.
- Kayne, R. S. 2006b. Silent Clitics. Paper presented at the Fourth University of Cambridge postgraduate conference in language research, Cambridge.
- Ledgeway, A. 2007. Osservatorio linguistico: la sintassi dei dialetti meridionali. *La lingua italiana. Storia, strutture, testi* 3.161-170.

- Manzini, M. R.; Savoia, L. M. 2003. Morpho-sintaksa e pjesores dhe e paskajores në të folmen e Shkodrës. *Studime Filologjike*, 3-4.
- Manzini, M. R.; Savoia L. M. 2003. Participio e infinito nella varietà di Scutari. Cinque secoli di cultura albanese in Sicilia. *Atti del XXVIII Convegno Internazionale di Studi Albanesi*, ed. by Matteo Mandalà, 401-432. Palermo: A. C. Mirror.
- Manzini, M. R.; Savoia L. M. 2005. *I dialetti italiani e romanci. Morfosintassi Generativa*. Alessandria: Edizioni dell'Orso, 3 vols.
- Manzini, M. R.; Savoia, L. M. 2007. A Unification of morphology and syntax. *Studies in Romance and Albanian dialects*. London: Routledge.
- Manzini, M. R.; Savoia L. M. 2008. *Work Notes on Romance morphosyntax/Appunti di morfosintassi romanza*. Alessandria: Edizioni dell'Orso.
- Manzini, M. R.; Savoia L. M. 2008. 'Non-active voice in Albanian: Implications for the theory of movement. *Studi sulle varietà arbëresh*, ed. by Leonardo Maria Savoia. Rende: Università degli studi della Calabria.
- Manzini, M. R.; Roussou, A.; Savoia L. M. In press. The morpho-syntax of non-active voice in Greek and Albanian. *Proceedings of Mediterranean Syntax meeting 2, Istanbul*.
- Marantz, A. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *Proceedings of the 21st Annual Penn Linguistics Colloquium*, ed. by Alexis Dimitriadis, Laura Siegel, Clarissa Surek-Clark and Alexander Williams. U. Penn Working Papers in Linguistics, 4.2.201-225.
- Pellegrini, G. B. 1995. *Avviamento alla linguistica albanese*. Università di Palermo.
- Rizzi, L. 1992. *Issues in Italian syntax*. Dordrecht: Foris.
- Roberts, I.; Roussou, A. 2003. *Syntactic change: a Minimalist Approach for Grammaticalization*. Cambridge: Cambridge University Press.
- Rohlf, G. 1969. *Grammatica storica della lingua italiana e dei suoi dialetti. Sintassi e formazione delle parole*. Turin: Einaudi.
- Trumper, J.; Rizzi, L. 1985. Il problema sintattico di ca/mu nei dialetti calabresi mediani. *Quaderni del Dipartimento di Linguistica*. 63-76. Università della Calabria.
- Zanuttini, R. 1997. *Negation and clausal structure*. New York: Oxford University Press.

Sanna Hillberg

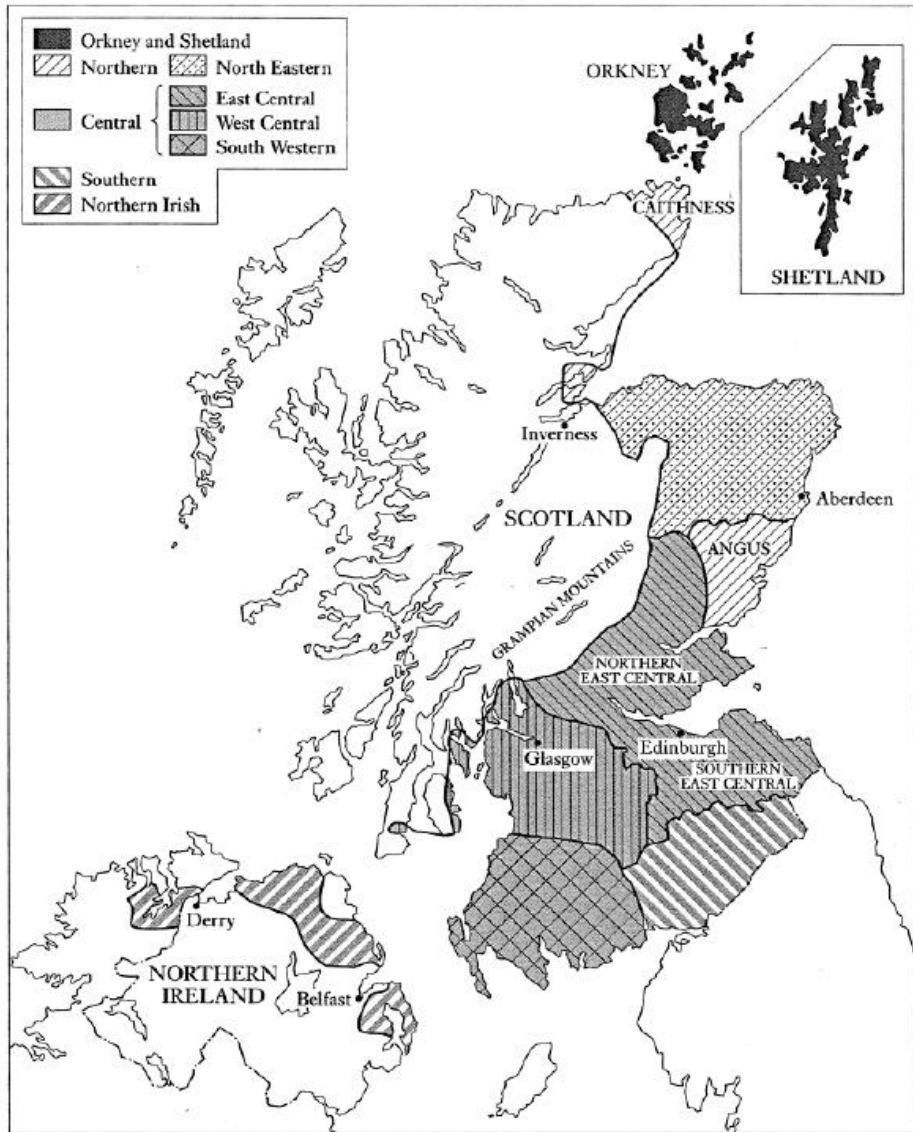
Development of the relativization system in Scots and Scottish English

1 INTRODUCTION

This paper discusses two major changes that have taken place in the written Scots/Scottish English (ScE) relativizer system between the Older and Modern Scots periods. First, I will discuss the development of the relativizer *that* becoming a restrictive relativizer in formal written discourse. Secondly, I will deal with the incorporation of the English *wh*-relativizer *which* into written Scots. The findings of Caldwell (1974) and Romaine (1982) on Early and Middle Scots are contrasted with the preliminary findings from my current study, which focuses on the relativization features of the present-day educated written ScE.

For my data I have compiled a corpus that I will call *The Corpus of Scottish English On-line Press News (CSEOPN)*, and as a comparison material I use the sections of press news reportage from International Corpora of English (ICE), i.e. *ICE-Ireland* and *ICE-Great Britain Release 2*. Since the latter, however, contains samples from two Scottish newspapers that are included in the CSEOPN, *The Herald* and *The Scotsman*, these samples are extracted from the ICE-GB material and replaced by a sample that I have collected from *The Guardian* and *The Times* on-line newspapers: this sample amounts to the same number of words as the material that has been extracted from ICE-GB. Hence, the Standard English figures presented in this paper comprise ICE-GB, *The Guardian* and *The Times* figures.

CSEOPN is divided into two sub-corpora: Highland and Island ScE (HIScE) and Lowland ScE (LScE), under which the regional dialect areas are placed. This division is based on historical-linguistic grounds, i.e. the established dialect border between the Scottish Highlands and Lowlands. In the Lowland areas English was an alternative to Scots from the 15th century and the English norms in writing were fully accepted by 1700. The Highlands were Gaelic-speaking relatively late, and the gradual adoption of the English language did not start until the 17th century (see e.g. Görlach 2002: 168–175).



Map 1. Dialect map of Scotland (in Smith 2000: 161).

The history of the Scots language is divided into two main periods: Older Scots (OS), until 1700, and Modern Scots (ModS), from 1700 onwards. The OS period is grouped into the following subcategories: Pre-literary Scots (PIS) (until 1375); Early Scots (ES) (1375–1450) and Middle Scots (MS) (1450–1700). Detailed studies on the OS morphology and syntax including relative pronouns and clauses have been carried out, for example, by King (1997) and Moessner (1997). Rather surprisingly, the relativization strategies of Late Middle Scots (LMS) (1550–1700) and the first centuries of the ModS period have not been investigated. In addition, the syntax of HIScE

has received very little scholarly attention so far. All published studies on Scots relativization concern Lowland Scots/LScE, while relativization in HScE has not been investigated until very recently.

Relativization strategies in Scots and ScE have developed somewhat independently in comparison to StE, in which the adoption of relativization strategies may have followed Latin and French models. However, this foreign influence has been seen as strengthening rather than initiating the trend (Fischer 1992: 299). During the OS period relative pronouns in Scots had different spellings from their English counterparts. Some peculiarities in the relativizer use still exist in ScE, just as in any non-standard variety of English. It has to be noted that the earliest texts investigated with respect to the relativizer use in Scots date back to 1375¹, because earlier written records are not available.

2 THE RELATIVIZER *THAT/AT*

The indeclinable OS relativizer had two forms: *that* and *at*. *That/at* occurs with personal, non-personal and indefinite subjects. Contrary to the Modern StE system, in which the use of the relativizer *that* is mainly confined to restrictive relative clauses, in ES *that* functions freely in both restrictive and non-restrictive relative clauses. In ES *that/at* functions as subject, direct object, prepositional object, copulative and adverbial complement. There are some differences in *that/at* usage: the instances of the relative *that* clearly outnumber *at* in ES writing. *At* occurs mainly in restrictive relative clauses, while *that* is common in both types of adnominal relative clauses (see e.g. Cadwell 1974, King 1999). Use of *at* declines after the sixteenth century in written Scots, but it is still heard in speech (Caldwell 1974: 31).

An interesting distinction arises in the uses of the restrictive and non-restrictive *that* in the ES writings: in literary texts *that* is frequent in both types of adnominal relative clauses, while in record and official prose it is common in restrictive relative clauses (Caldwell 1974: 27–28). This finding suggests that non-restrictive *that* might have been considered somewhat informal already during the OS period. In MS *that* typically functions in restrictive relative clauses, which is similar to its modern StE and ScE usage. The preliminary results of the study show that today, non-restrictive *that* is rather uncommon in educated written ScE, but some examples occur in my data (1). According to Biber et al. (1999: 610–1), non-restrictive *that* is frequent in conversation, but rarely occurs in news writing, and, for example, in fiction it is ‘used for special stylistic effect’.

- (1) The Moffat branch of the charity, *that* supports the serving and ex-service community and their families, is to have its first standard dedicated. (LScE, *Dumfries and Galloway Standard*)

¹ The Bruce, compiled by Master John Barbour, Archdeacon of Aberdeen, which is the oldest extant book written in Scots (Caldwell 1974: 8).

Unfortunately, it has not been recorded when non-restrictive *that* became less common in written ScE. However, non-restrictive *that* occurs in present-day spoken Scots, see example (2).

(2) Jock Allan, *that* has done so well in Embro, was a herd. (T25/14, 15, Macafee 1983: 52)

One obvious reason for *that* becoming rare in non-restrictive relative clauses is the incorporation of the personal relativizer *who* into Scots in the 16th century. *Who* often modifies, for example, proper nouns, which are always followed by a non-restrictive relative clause. According to Ryden (1979: 14), *that* was functionally overloaded and therefore *who* was a natural alternative with personal antecedents.

Overall, the relativizer *that* is rather common in educated written ScE: the mean frequency is 15.0 per 10,000 words of running text (16.6 in LScE and 13.3 in HScE), while the corresponding figure for StE is 14.2 and for IrE 7.4. In comparison to the personal relativizer *who*, *that* has become overall less common over the centuries, and today the majority of personal antecedents are relativized by *who*. For comparison, the mean frequencies of the relativizers *that* and *who* per 10,000 words of running text are provided in figure 1.

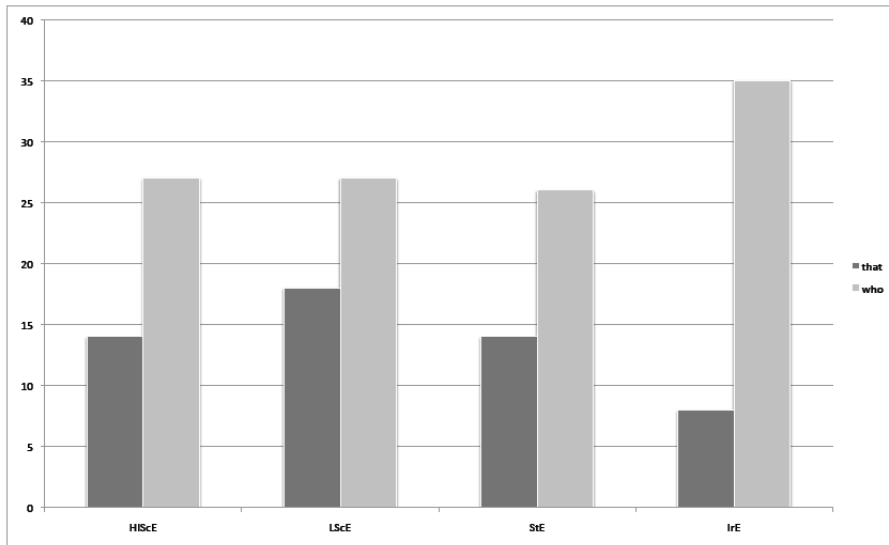


Figure 1. The mean frequencies of the relativizers *that* and *who* in H15cE, L15cE, StE and IrE / 10,000 words of running text.

3 THE *QUH*- AND *WH*-RELATIVES

The so-called *quh*-relatives, which correspond to ModE *wh*-relativizers, occur in OS. Numerous variants exist for each relativizer, e.g. the relativizer *quhilk* (*which*) was also attested, for example, in the following forms *quhich*, *quilk*, *qlk* and *qwelk* (Meurman-Solin 2003: 185). The origin of these northern *quh*-forms is somewhat unclear, but it has been suggested that *quhilk*, which also has a definite form *the quhilk* and an inflected plural form (*the*) *quhilkis*, derives either from OE or ON (King 1997: 172).

The plural form (*the*) *quhilkis* occurs only in Scots. It is either of French, Latin or native Scottish origin, and although it is a plural form it does not necessarily follow a plural antecedent (see King 1997: 172, Caldwell 1974: 36). However, it typically occurs with plural head nouns (3):

- (3) Herculis and ... Alexandere *the quhilkis* ware baith borne Grekis (*Asl. MS.* 1. 186/19, c. 1515, Caldwell 1974: 37)

(*The*) *quhilkis* disappears from Scots around the 18th century, and the reason for this is unclear. However, it is possible that the establishment of the English norms concerning the use of the *wh*-forms simply made (*the*) *quhilkis* useless.

The indefinite form, *quhilk*, which is established in Scots after the incorporation of the definite form, becomes more frequent from 1450 onwards and competes to some extent with *that/at* (McClure 1994: 54). Since then *quhilk* has slowly worked its way from the most formal types of writing and the most complex types of relative clause structures to the modern spoken Scots taking forms *whilk* and *which*. *Whilk* is still occasionally heard in Scots:

- (4) Guid fowk, we are convent here neist the auld brugh o Stirling *whilk* the auld makar Lindsay, in his “Papyngo” has screevit about.

(<http://www.scottishcorpus.ac.uk/corpus/search/>, Borrowman, A.S. 1976. Sermon at Stirling, audio transcription)

It is likely that the relativizers (*the*) *quhilk* and its plural form (*the*) *quhilkis* were a convention of the written discourse, and not used in ES speech (Caldwell 1974: 33). Today, also the definite form has become extinct.

The English *wh*-forms appear in Scottish texts around 1540. They are first used with the *quh*-forms, which, however, become increasingly uncommon towards the end of the 17th century, and ultimately they disappear altogether.

In the present data of educated written English *which* is the most common relativizer. In the overall ScE data the mean frequency for the relativizer *which* is 32.0 per 10,000 words of running text, in StE 33.4 and in IrE 30.4. The correlations between the frequencies of *which* and *that* are shown in figure 2. In HIScE, which is more recently influenced by StE, the “formal” relativizer *which* is more salient than in LScE, and the mean frequencies per 10,000 words of running text are 34.4 and 29.5, respectively. On the other hand, the mean frequencies for the relativizer *that*, which is a dominant relativizer in LScE speech, indicates that *that* is more frequent in LScE than in HIScE on-line news writing.

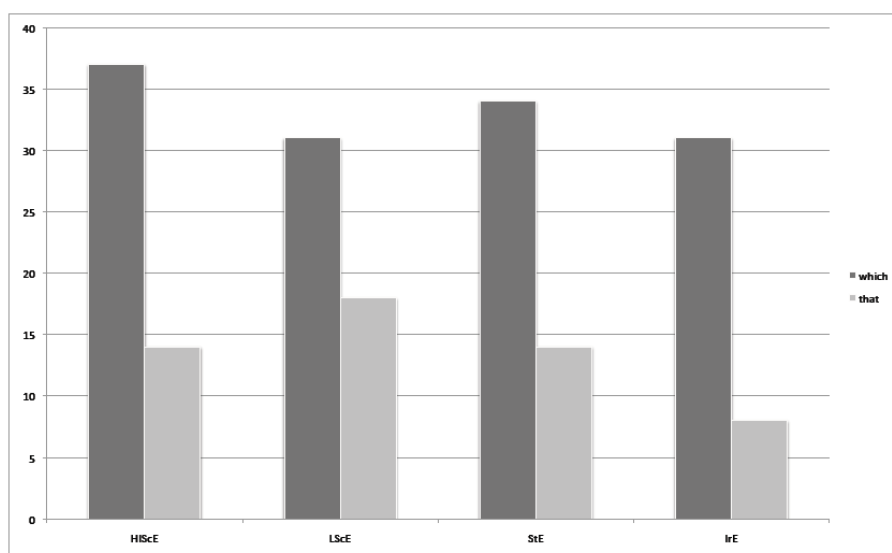


Figure 2. The mean frequencies of the relativizers *which* and *that* in HIScE, LScE, StE and IrE / 10,000 words of running text.

During the ES period *quhilk* is already emerging as a non-restrictive relativizer. This is another reason for *that* starting to lose ground as a non-restrictive relativizer and acquiring a position as a restrictive relativizer in written language. Therefore, this finding suggests that the distinction

between *that* and *quhilk* with respect to restrictiveness developed during the ES period. Overall, *quhilk* is more common in non-restrictive than restrictive context in ES writing (Caldwell 1974: 36–7). The initial findings of the present online news data suggest that all varieties prefer non-restrictive *which*. These results contradict those of Biber et al. (1999: 610), who claim that restrictive *which* is preferred in news writing, while non-restrictive *which* is preferred in spoken discourse. Studies show that relativizer *which* is rare in LScE speech, especially in restrictive relative clauses, and in non-restrictive functions it is the most frequent in sentential relative clauses (Tagliamonte et al. 2005: 86, King 1997: 173).

With respect to the animacy of the antecedent, before the personal nominative relativizer *who* was incorporated into Scots, *quhilk* referred to both animate and inanimate antecedents, unlike the ModE *which*, which refers to inanimate and nonhuman antecedents only (with some exceptions, e.g. the word *baby*).

- (5) Beatrice, *quhilk* wes gevin to Cithrik in mariage² (BELL 86.30f. Moessner 1997: 144)

The current data shows that personal *which* has become nonexistent in educated written language. However, it is common in ScE with collective antecedents (6), which are nouns that refer to groups of people (for more detailed definition see e.g. Quirk et al. 1985: 316–317).

- (6) Su Daintith Watch Manager for Liverpool Coastguard *which* were called to the scene said: "Our priority at the moment is the safety of life, we have search and rescue units on scene, who are working closely with the barge support vessels to recover all the workers to a place of safety. (LScE, *Dumfries and Galloway Standard*)

4 CONCLUDING REMARKS

To conclude, remarkable changes have taken place in the Scots/ScE relativizer system over a period of centuries. The relativizer *that* was formerly common in both types of adnominal relative clauses with personal and non-personal antecedents. However, in today's ScE on-line news writing *that* has become infrequent in non-restrictive relative clauses and with personal antecedents. The northern *quh*-relativizers were replaced by the English *wh*-forms in the 16th century and the native Scottish relativizer (*the*) *quhilkis* became extinct. With regard to restrictiveness, contrary to the findings of Biber et al. (1999), the preliminary results of the study suggest that non-restrictive *which* is preferred in news writing.

Unfortunately, the timeline of relativization in Scots/ScE cannot for the time being be completed because its relativization system in the 18th and 19th century has not been investigated. For the same reason it remains unclear when and why the relativizer (*the*) *quhilkis* disappears from Scots and *that* becomes rather informal non-restrictive relativizer. However, these will be interesting topics for future research.

² BELL 86.30f. Moessner (1997: 144).

CORPORA

International Corpus of English (ICE) – Great Britain

International Corpus of English (ICE) – Ireland

The Corpus of Scottish English On-line Press News (CSEOPN)

BIBLIOGRAPHY

- Aitken, A.J. 1984a. 'Scottish accents and dialects'. In Trudgill, P. (ed.), *Language in the British Isles*. Cambridge: Cambridge University Press, 94–114.
- Beal, J. 1997. 'Syntax and Morphology'. In Jones, C. (ed.), *The Edinburgh History of the Scots Language*. Edinburgh: Edinburgh University Press, 335–377.
- Caldwell, S.J.G. 1974. *The Relative Pronoun in Early Scots*. Helsinki: Société Néophilologique.
- Fischer, O. 1992. 'Syntax'. In Blake, N (ed.), *The Cambridge History of the English Language*. Vol. 2, 1066 - 1476. Cambridge: Cambridge University Press, 207–408.
- Görlach, M. 2002. *A Textual History of Scots*. Heidelberg: Winter.
- Herrman, T. 2005. 'Relative clauses in English dialects of the British Isles'. In Kortmann B., T. Herrmann, L. Pietsch and S. Wagner, *A Comparative Grammar of British English Dialects*. Berlin, New York: Mouton de Gruyter, 21–123.
- Hillberg, S. 2008. *Relativization in Scottish and Standard English On-line Newspapers*. Unpublished MA Thesis. Department of English, University of Joensuu.
- King, A. 1997. 'The Inflectional Morphology of Older Scots'. In Jones, C. (ed), *The Edinburgh History of the Scots Language*. Edinburgh: Edinburgh University Press, 156–181.
- Meurman-Solin, A. 1997. 'Differentiation and Standardisation in Early Scots'. In Jones, C. (ed), *The Edinburgh History of the Scots Language*. Edinburgh: Edinburgh University Press, 3–23.
- Meurman-Solin, A. 2003. 'Corpus-based Study of Older Scots Grammar and Lexis'. In Corbett, J., J. D. McClure and J. Stuart-Smith (eds.), *The Edinburgh Companion to Scots*. Edinburgh: Edinburgh University Press.
- Moessner, L. 1997. 'The Syntax of Older Scots'. In Jones, C. (ed), *The Edinburgh History of the Scots Language*. Edinburgh: Edinburgh University Press, 112–155.
- Quirk, R., S. Greenbaum, G. Leech and J. Svartvik. 1985. *A Comprehensive Grammar of the English Language*. New York: Longman.
- Romaine, S. 1982. *Socio-historical Linguistics: Its Status and Methodology*. Cambridge: Cambridge University Press.
- Ryden, M. 1979. *An Introduction to the Historical Study of English Syntax*. Stockholm: Almqvist & Wiksell International.

ELECTRONIC SOURCE

Scottish Corpus, <http://www.scottishcorpus.ac.uk>

Rossella Iovino

Restructuring structures with modal and aspectual verbs in Archaic and Classical Latin

This paper presents an analysis of restructuring structures (henceforth RSs) with modal and aspectual verbs in archaic and classical Latin. In particular, extending previous analyses (Zennaro (2004) and Costantini and Zennaro (2010)), I will provide various examples in order to demonstrate that all Latin modal (*audeo*, I dare; *debeo*, I must; *exopto*, I desire; *malo*, I prefer; *nequeo*, I cannot; *nolo*, I do not want; *possum*, I can; *queo*, I am able to; *volo*, I want) and aspectual verbs (*aggredior*, to be about to; *cesso*, I stop; *coepi/incipio*, I start; *conor*, I try; *cunctor*, I wait; *curo*, I make effort to; *desino*, I leave out; *festino*, I speed up; *maturo*, I accelerate; *persevero*, I continue; *soleo* I usually) enter RSs. Furthermore, I will discuss some of their properties in relation to Cinque's (1999; 2004) theory on the hierarchy of functional projections. I will show that the syntactic behaviour of each of the modal and aspectual verbs depends on the kind of modality or aspectuality they express, that is, on the functional projection they occupy with respect to the others, and to the Past Tense in particular.

As for data, important writings of authors active from earliest times to the 2nd century A.D. was considered and gathered using the *B(ibliotheca) T(eubneriana) L(atina)*. During this period, in fact, Latin shows an interesting syntactic homogeneity.

In view of the goals it pursues, the paper is organized in the following way: the first section is devoted to a brief description of the framework which forms the basis of the present approach to the syntax of the Latin RSs (§1), while the second gives a brief overview of the main properties of Latin RSs with modal and aspectual verbs (§§ 2-3). In particular, two tests which permit the individuation of monoclausal structures will be illustrated (position of the weak pronouns and marked order between the finite verb and the infinitive); finally, the last section proposes a syntactic analysis of the RSs with modal and aspectual verbs (§§ 4-4.2). This theoretical contribution will be inferred by the direct analysis of significant Latin texts, used as examples (cf. Bortolussi forthcoming).

1 THE FRAMEWORK

As argued in Rizzi (1976), Burzio (1986) and Cinque (2004), the term ‘restructuring’ is used to refer to a syntactic structure in which an infinitive is contained within a monoclausal structure with the main verb of particular verb classes, namely, modal (i.e. *to want, can, to seem*), aspectual (i.e. *to finish, to stop, to continue*) and motion verbs (i.e. *to come*), as in the following Italian and English sentences:

- (1) (a) Gianni [*vuole rientrare tardi la sera*]_{VP}
(b) John [*wants to come back late in the evening*]_{VP}

Cinque (2004) analyses RSs within his theory on the hierarchy of the functional projections, proposed in Cinque (1999), and claims that they are rigidly ordered in the following hierarchy:

Cinque’s (2004) hierarchy

MoodP_{speech act} > MoodP_{evaluative} > MoodP_{evidential} > ModP_{epistemic} > TP (Past) > TP (Future) > MoodP_{irrealis} > Mod_{alethic necessity} > AspP_{habitual} > AspP_{repetitive(I)} > TP (Anterior) > AspP_{terminative} > AspP_{continuative} > AspP_{retrospective} > AspP_{proximative} > AspP_{durative} > AspeP_{generative/progressive} > AspP_{prospective} > ModP_{obligation} > ModP_{permission/ability} > AspP_{completive} > VoiceP > AspP_{celerative(II)} > AspP_{repetitive(II)} > AspP_{frequentative(II)}

In particular, Cinque (1999) suggests that the adverbs are generated in the specifier of each functional projection. This theory of adverbs is also useful to understand the restructuring phenomenon: Cinque (2004), in fact, proposes that the verbs which give rise to a monoclausal structure are generated in the head of a functional projection depending on their semantics. The fact of being generated in functional projections justifies their definition as ‘functional verbs’. In this perspective, a Latin modal verb like *queo* (I can) is merged in the head of the functional projection ModP_{permission/ability}, while an aspectual verb like *soleo* (I am used to) is merged in the head of the functional projection in AspP_{habitual}.

2 THE LATIN RSS WITH MODAL VERBS

In order to individuate RSs in Latin, Salvi (2004) proposes two tests. First of all he shows that Latin has strong and weak pronouns, as put forward by Cardinaletti and Starke (1999). In particular, Salvi (2004: chap. IV and V) argues that, although there is only one pronominal morphological form in Latin, in some cases strong pronouns can be distinguished from weak ones on the basis of syntactic and pragmatic evidence. Strong pronouns can occupy any position in the sentence and can be substituted by nouns, unlike weak pronouns, which only have anaphoric reference and occupy the position after the first element of the sentence (‘Wackernagel’s position’), independently of their syntactic function. This is shown in the following example (from Salvi (2004: 126), in which the pronoun *me* is placed in Wackernagel’s position:

- (2) *Caninius noster me tuis verbis admonuit* (Cic., *epist.*, 9,6,1)
Our friend Caninius gave me your message.

Salvi (2004) uses pronoun collocation in Wackernagel's position as a test of monoclausality. Following Costantini and Zennaro (2010), who demonstrate that the presence of a weak pronoun in Wackernagel's position involves the verb *possum*, and the *verba voluntatis*, I will show that RSs can be found with all modal verbs, as can be seen in the following examples regarding *queo*, *nequeo*, *debeo* and *audeo*:

- (3) *Neque eam queo locare cuiquam* (Varro., *ling.*, 5,2)
I cannot give her in marriage to anyone.
(4) *Ego hoc nequeo mirari satis* (Plaut., *Trin.*, 1132)
But I can't marvel enough at this.
(5) *Quae ipsi debeo huic libro libenter inserui* (Stat., *Silv.*, 2,21)
For him I gladly have to insert these things in this book.
(6) *Nondum id quidem audeo dicere* (Cic., *fin.*, 4,1,1)
On that point I do not yet dare to speak

The above examples demonstrate that not only *possum*, *debeo* and the *verba voluntatis* can enter RSs, but other modal verbs can do too, constituting an homogeneous class from a syntactic point of view.

As regards the second test, Salvi (2004: 43) affirms that the unmarked word order in Latin is that with the verb in the final position in the clause, as in the following example:

- (7) *Neque ego ad te his duobus mensibus scripseram* (Cic., *epist.*, 8,9,1) (from Salvi 2004: 43)
And I have not written to you for the last two months.

Salvi (2004: 47-48) also notes that in sentences with two predicates, namely finite verb plus infinitive, the former follows the latter, at least in the unmarked order:

- (8) *Hoc tempore Catilinam, competitorem nostrum, defendere cogitamus* (Cic., *Att.*, 1,2,1) (from Salvi 2004: 47)
At the present time I am thinking about defending my fellow candidate Catiline.

However, Salvi (2004: 47) adds that the unmarked SOXV (S-O-infinitive-finite verb) order above can change into SOVX (S-O-finite verb-infinitive), if the verb (V) and the infinitive (X) form a unit:

- (9) *Si haec mala fixa sunt, ego vero te quam primum, mea vita, cupio videre* (Cic., *epist.*, 14,4,1) (from Salvi 2004: 48)
If these ills can never be removed, I assure you, my dears, that my desire is to see you as soon as possible.

According to Salvi (2004), the verb position at the end of the sentence, the infinitive-finite verb order with lexical verbs and the inversion of that order in the case of functional verbs all constitute an important tendency which can be found in archaic and classical Latin. However, as noted in Devine-Stephens (2004: chap. 5), in addition to this there is a wide range of options in the placement of the verb (or of the verbal group) and of the other elements of the sentence. Let us consider the following examples:

- (10) *Audeo dicere hoc* (Liv., 2,34,11)
I dare to say that.
- (11) *Is tamen hanc causam ab illa debet seiungere* (Cic., *Cluent.*, 96)
He ought none the less to make a distinction between that case and this.
- (12) *Te ipsum mihi iam dudum exoptabam dari* (Ter., *Heaut.*, 758)
I've been hoping to run into you for a while.
- (13) *Proditorem nolo dicere, certe speculatorem habemus in sinu* (Liv., 40,5,12)
I do not want to say a traitor, but we have in our bosom at least a spy.
- (14) *Mavult dicere voluptatem quam vacuitatem doloris* (Cic., *fin.*, 2,16)
He prefers to talk of the pleasure than of the absence of the pain.
- (15) *Quid istuc sit [...] nequeo noscere* (Plaut., *Asin.*, 36)
I cannot know that's the place you mention.
- (16) *Sed tamen hoc queo dicere* (Cic., *Cato.*, 32)
But still I can say this much.

Sentences (10) to (16) confirm that in the case of a complex predicate, the infinitive generally follows the finite verb. As noted in Costantini and Zennaro (2010) for *debeo*, finding this property in predicates, especially those involving modal verbs, constitutes an important piece of evidence for the hypothesis that modal verbs form a single predicate with the infinitive. Nevertheless, the above examples show that, even if the verbal group is generally placed in the final position in the sentence (11), (12), (15), (16), this is not the only possibility. In (10) and in (14), for instance, the predicate precedes the object, while in (13) the verbal group *nolo dicere* is contained within a complex object (*proditorem nolo dicere, certe speculatorem*), which regularly precedes the main verb.

Although in archaic and classical Latin there is a great variability in the word order, the sentences above show that it is possible to find some regularities and that the disposition of the elements in the sentences is changeable, but it is not unrestricted. This is confirmed on considering a larger number of occurrences. Taking into account the verbal structures with a modal verb plus an infinitive found in Latin texts available in the BTL, the situation is the following in archaic and classical Latin:

- (17) (a) *Audeo dicere* 24 vs *dicere audeo* 3
(b) *Debet seiungere* 1 vs **seiungere debet*
(c) *Exoptabam dari* 1 vs **dari exoptabam*
(d) *Nolo dicere* 4 vs **dicere nolo*
(e) *Mavult dicere* 3 vs **dicere mavult*

- (f) *Nequeo noscere 2 vs *noscere nequeo*
 (g) *Queo dicere 5 vs *dicere queo*

From these data it is evident that, independently of the collocation of the verbal group in the sentence, in the presence of RSs, the most attested word order in many cases is that in which the infinitive follows the finite verb to which it is adjacent. The only three cases of classical text in which the infinitive-finite verb order is attested can be accounted for when considering their information structure. This pattern is very marked, but it is still grammatical:

- (18) [TOP *Rem a nostris positam* | [FOC *nec dicere* [IP *audeo* [tTOP [tFOC]]]], *quia infirma uidetur, nec praeterire* (Sen., *nat.*, 4b,5,1)
 I am afraid to mention a theory established by our (Stoic friends), because they seem to be weak. I am also afraid to be silent about them.
- (19) *Tum ille vere vertens annus appellari potest; in quo* [FOC *vix dicere* | [IP *audeo* [*quam multa hominum saecula teneantur* [tFOC]]] (Cic. *rep.* 6,23)
 In that case the passage of the year can be explained. I hardly dare to say how many generations of men are contained within such a year.
- (20) *Id quod adversus hunc dicere audeo magis auderem adversus nutricem dicere* (Quint., *decl.*, 338,24)
 What I dare to say against him I should dare all the more to say against the foster mother.

In (18) the infinitive *dicere* is a Focus: its emphasis is due to the fact that it is in contrast to the other complex predicate “*nec praeterire (audeo)*”, with which it shares the same main verb; in (19) the infinitive *dicere* expresses new information, so it is a Focus and regularly precedes the finite verb; finally, in (31) Quintilianus creates a particular stylistic effect using *dicere audeo* [...] *auderem* [...] *dicere* in the same sentence in a chiasmic construction.

3 THE PROPERTIES OF THE ASPECTUAL VERBS

The two tests used to individuate a monoclausal structure with modal verbs - namely weak pronoun dislocation in the Wackernagel’s position and inversion of SOXV into SOVX order – also provide good results with aspectual verbs. This evidence confirms that aspectual verbs show the same syntactic structure of modal verbs and constitute another class of restructuring verbs. Let us consider the following three examples, which demonstrate the presence in Wackernagel’s position of weak pronouns as the object of the embedded verb:

- (21) *Quidem illum soleo hortari.* (Cic., *ad Att.*, 6,2,2)
 Of course I usually exhort him

(22) *Priusquam illa conor attingere quibus orationem ornari atque illuminari putem* (Cic., *de orat.*, 3,25)

Before attempting to deal with the qualities that seem to me to give ornament and brilliance to a discourse.

(23) *Ille mihi risum magis quam stomachum movere solet* (Cic., *Att.* 6,3,7)

He excites my laughter rather than my rage.

As in the above examples from (10) to (16), also in those from (21) to (23) it is possible to note a variability in the order of the elements in the sentence. In particular, there is evidence that pronoun collocation in Wackernagel's position is always found, while the verbal group can be placed in the final position in the sentence (21) or in the last position of the matrix clause, before the relative clause (22); finally the verbal group can also be attested in the unexpected order with the infinitive preceding the finite verb (23). Nevertheless, the following examples illustrate that the unmarked order with the aspectual verbs is that in which the infinitive follows the finite verb (as for the modal verbs):

(24) *Dona quid cessant mihi conferre?* (Plaut., *Maen.*, 129)

Why don't they all hurry up with the gift?

(25) *Haec illi soleo praecipere* (Cic., *Planc.*, 59)

I am accustomed to give him advice.

(26) *Non iis libentissime soleo respondere quos mihi videor facillime posse superare* (Cic., *Sull.*, 46)

I do not find the greatest pleasure in refuting those persons whom, I think, I can easily defeat.

(27) *Effigem conor efficere* (Plin., *epist.*, 3,10,6)

I am endeavouring to draw an immortal picture.

In addition to this, even though, generally speaking, BTL shows only few examples in which the aspectual verbs *cesso*, *soleo* and *conor* occur with an infinitive, the only possible word order is that in which the infinitive follows the finite verb:

(28) (a) *Cesso abire* 1 vs **abire cesso*

(b) *Soleo praecipere* 1 vs **praecipere soleo*

(c) *Soleo respondere* 1 vs *respondere soleo* 1

(d) *Conor efficere* 1 vs **efficere conor*

The single case in which the unexpected word order infinitive plus aspectual verb is found can be explained in terms of information structure:

(29) *Ego respondere soleo meis consiliis, periculis, laboribus patriam esse servatam, non tam sum existimandus de gestis rebus gloriari quam de obiectis confiteri* (Cic. *dom.*, 93)

I am in the habit of replying that it was by my forethought, at my risk, and through my exertions that my country was saved; it must be considered that I am not so much boasting of my own exploit, as stating facts in answer to charges.

It is quite clear, in fact, that in (29) *respondere soleo* is a contrastive Focus. Cicero creates, in fact, an opposition between two highly emphatic verbal groups by the preposing of the infinitive with respect to the finite aspectual verb and of *sum* with respect to the gerundive: “*ego respondere soleo [...] non tam sum existimandus*”. The correspondences between the features of the modal and aspectual verbs confirm the hypothesis that together they belong to the class of restructuring verbs form the class of the restructuring verbs.

4 SYNTACTIC ANALYSIS OF THE RSS WITH MODAL AND ASPECTUAL VERBS

In this section, I suggest a syntactic configuration for RSs based on Cinque’s (1999; 2004) framework. I will show that the syntactic behaviour of each of the functional verbs depends on the functional projection it occupies with respect to that of the Past Tense.

4.1 Syntactic analysis of restructuring structure with modal verbs

Let us consider the following two examples, in which there is a modal verb (*queo, possum*) (‘I can’) plus an infinitive:

(30) *Non queo durare* (Plaut., *Asin.*, 907)

I cannot tolerate it.

(31) *Nec tecum possum vivere, nec sine te* (Mart., *epigr.*, 12,46)

I cannot live with or without you.

From a semantic point of view, one notes that the two infinitives *durare* and *vivere* are two monoargumental verbs which should express a subject, but in these contexts the respective subjects are non-overt. At first glance one might think that the subject of the infinitive must be a PRO; such an analysis, however, creates at least two problems. First of all, PRO needs a ‘controller’ element which is absent in predicates with modal (and aspectual) verbs; these are, in fact, ‘light verbs’ which do not project any argument structure, in the same way as auxiliary verbs do. Furthermore, Cecchetto and Oniga (2002) demonstrate that PRO in Latin, as in English, is not compatible with a past infinitive; nevertheless, modal verbs in Latin can occur with a past infinitive, as in examples (32) to (35). Also in these sentences, the order of the element shows variability: in (32) and in (34) the finite verb precedes the infinitive, but it is not adjacent to it and in (33) and in (35) the order between the finite verb and the infinitive is not the expected one:

(32) *Nequeo nil commisisse nefandum* (Ov., *Met.*, 9,626)

I cannot now undo the wrong that I have done.

(33) *Tametsi statim vicisse debeo* (Cic., *S. Rosc.*, 73)

Although I must have immediately won.

(34) *Fortuna, sepulchrum/dicere Pompei, quo condi maluit illum/quam terra caruisse socer?*
 (Lucan., 8,793-795)

Is it the will of Fortune to call this the grave of Pompeius, this grave which Caesar preferred for his son-in-law to no burial at all?

(35) *Non discere debemus ista, sed didicisse* (Sen. *epist.* 88,2)

We ought not to be learning such things; we should have done with learning

The same compatibility of a modal verb with a past infinitive can also be found in Italian and in English:

(36) Io (non) posso/voglio/devo farlo

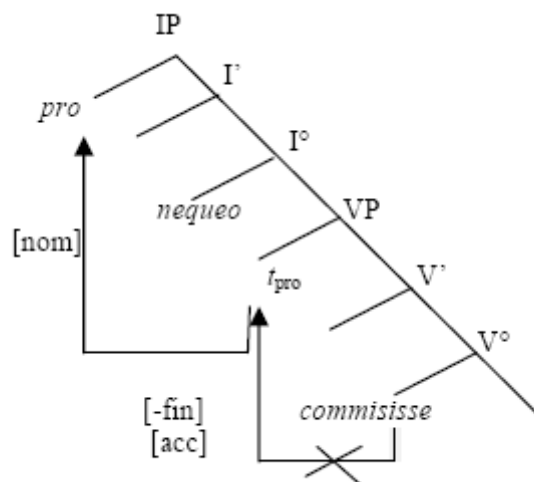
I can(not)/(do not) want/must (not) do it.

(37) Io (non) posso/voglio/devo averlo fatto

I can(not)/(do not) want/must (not) have done it.

According to Cecchetto and Oniga (2002), the presence in all these examples of a past infinitive confirms that PRO is excluded from predicates with modal verbs. This evidence suggests that, in order not to violate the Thematic Criterion, the argument realized as the subject of the modal verb must be the argument of the infinitive, namely of the one verb which projects its argument structure (given that the modals are 'light verbs'). The subject cannot receive case in its merge position (SpecVP of the infinitive), so it moves to the SpecIP of the matrix clause, the only position in which a finite verb can assign nominative case to it. This is shown in (38), which is the structure of (32):

(38) Figure 1



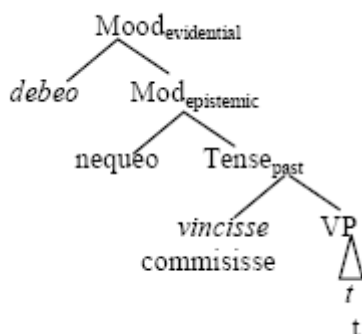
Although there is a large number of modals compatible with a past infinitive (*volo, nolo, malo, queo, nequeo, possum, debeo*), BTL offers two modals, namely *exopto* and *audeo* respectively,

attested in sentences (12) and (6), which cannot occur with a past infinitive, but only with an infinitive in the present tense. The corresponding Italian and English data show the same distinctive feature:

- (39) Io ti esorto a farlo *vs.* *Io ti esorto ad averlo fatto.
 I exhort you to do this *vs.* *I exhort you to have done this.
- (40) Io oso guardarti negli occhi *vs.* *Io oso averti guardato negli occhi.
 I dare to look in your eyes *vs.* *I dare to have looked in your eyes.

Semantic and syntactic motivation can explain this evidence: from a semantic point of view, ‘exhortation’ and ‘daring’ – different to ‘possibility’, ‘impossibility’, ‘duty’ and ‘preference’- are conceptually incompatible with the past. From a syntactic point of view, following Cinque (1999; 2004), I suggest that, in general, the problem of the compatibility or incompatibility of modals with a past infinitive is due to the existence of different classes of modal verbs, merged in different structural positions with respect to Tense merge position. As regards the above examples in which a modal occurs with a past infinitive, we note that they express different kinds of modal meanings. In particular, in the case of *nequeo commisisse* (32), the linguistic context favours an epistemic meaning, so the sentences in which it occurs could be explained by an expression such as: *I cannot have done such a thing, because generally it is not the kind of thing I would do*. On the other hand, as regards *vincisse debeo* (33), the linguistic context suggests that there is some evidence confirming the triumph to which Cicero refers. In his hierarchy, Cinque (1999; 2004) proposes that these two kinds of modalities correspond to specific functional projections merged in a structural position higher than Past Tense. This accounts for their compatibility with a past infinitive, as is shown in the following structure which illustrates the possibility for the infinitive to move in order to check Tense without any violation of the head movement constraint:

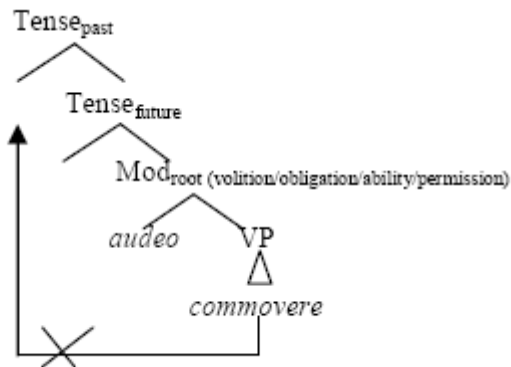
(41) Figure2



In contrast, the modals occurring with a present infinitive express neither evidential nor epistemic meaning: the examples found in the BTL which show modal verbs only occurring with a present infinitive express, in fact, volition - *malo (mavult dicere (14))*, *exopto (exoptabam dari (12))*

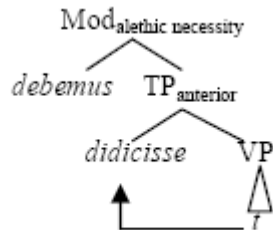
-, obligation - *debeo* (*debet seiungere* (11)) - ability *nequeo* (*nequeo mirari* (4) *nequeo noscere* (15)) and permission - *queo* (*queo locare* (3)), *audeo* (*audeo dicere* (6)). Cinque (1999) includes these semantic values under the label of ‘root modality’ and merges its functional projection below Tense. In this perspective it is possible to explain the incompatibility of ‘root modals’ with a past infinitive. In these cases, in fact, the infinitive cannot escape its position in order to check Tense without violating the locality constraints. In the following structure the infinitive *commovere* cannot check Tense without re-merging across the modal verb:

(42) Figure 3



As regards ‘root modals’, BTL offers an interesting example in which *debeo* occurs with the past infinitive *didicisse* (35); this, however, does not represent a counterexample. The linguistic context suggests, in fact, that such a sentence is highly emphatic. As a matter of fact, it is clear that what is involved is the contrast between Seneca’s declaration that we must not study now because we should have already *finished* studying. Such an observation suggests that the semantics of the modal verb is compatible with an idea of necessity, while that of the past infinitive is more aspectual than temporal. Thus, in this case, it is possible to merge the first in the functional projection $Mod_{alethic\ necessity}$ and to move the latter to $TP_{anterior}$. This is displayed in the following structure which accounts for the possibility for the infinitive to move in order to check $TP_{anterior}$:

(43) Figure 4



4.2 Syntactic analysis of restructuring structure with aspectual verbs

With respect to aspectual verbs, BTL shows that they are only compatible with a present infinitive. Let us consider the following examples:

- (44) *Non ego laudari curo* (Tib., 1,1,57)
I am not interested in being praised.
- (45) *Ergo ego, ne scribam, digitos incidere cunctor?* (Ov., Pont., 4,14,19)
Do I then hesitate to cut my fingers that I may not write?
- (46) *Hostis vivos rapere soleo* (Plaut., Pseud., 655)
I am used to capturing the enemies alive.
- (47) *De rerum natura pangere conor* (Lucr., 1,25)
I am about to fashion touching the Nature of Things.
- (48) *Iniuriam facere fortissime perseverat* (Cic., Quinct., 8)
[Dolabella] most manfully persevered in acting wrongfully.
- (49) *Maturat ab urbe proficisci* (Caes., Gall., 1,7,1)
[Caesar] hastens to leave Rome.
- (50) *Quem equidem cruci adfixum videre festino* (Curt., 6,3,14)
A man whom I, from my part, am in haste to see nailed to a cross.
- (51) *Orbis situm dicere aggredior* (Mela., 1,1)
I start to talk about the site of the city.
- (52) *De re publica dicere incipio* (Cic., Phil., 1,11)
I begin to speak on public affairs.

As for the intrinsic variability of word order in archaic and classical Latin data, the examples from (44) to (52) show that it is less predictable in the presence of an aspectual verb than in the presence of a modal verb. In the above cases, in fact, the infinitive precedes the finite verb and the verbal group is regularly placed in the final position in the sentence, except for (49) in both respects. Comparing the Latin with the Italian and English data, it can be noticed that, also in these languages, aspectual verbs always occur with a present infinitive:

- (53) (a) ?Non mi interesso di essere stato lodato.
 (b) *I am not interested in been praised.
- (54) (a) *Esito a essermi tagliato le dita per non scrivere.
 (b) *I hesitate to have cut my fingers that I may not write.
- (55) (a) *Sono abituato ad aver preso i nemici vivi.
 (b) *I am used to have captured the enemies alive.
- (56) (a) *Mi accingo ad aver scritto sulla natura delle cose.
 (b) *I essay to have fashioned touching about the nature of things.
- (57) (a) *Continua audacissimamente ad aver commesso un'ingiuria.
 (b) *He perseveres in have acted wrongfully.
- (58) (a) *Si sbriga ad aver lasciato Roma.
 (b) * He makes speed to have left Rome.
- (59) (a) *Certamente io ho fretta di averlo visto appeso alla croce.
 (b) *Of course I am in haste to have seen him nailed to a cross.
- (60) (a) *Esordisco con l'aver parlato del sito della città.
 (b) *I start to have talked about the site of the city.
- (61) (a) *Inizio ad aver parlato della Repubblica.
 (b) *I begin to have spoken on public affairs.

The absence of Latin, Italian and English sentences constituted by an aspectual verb occurring with a past infinitive can be explained by considering that an aspectual verb specifies an event taking place from an aspectual point of view, expressed by the infinitive. It implies that, in order to be aspectually specified, the infinitival event cannot be accomplished, so it has to be realized in the present tense (namely in the imperfective stem). Such semantic intuition appears to find confirmation from Cinque's (1999) perspective. In his hierarchy, the aspectual verbs are merged in a functional projection lower than Tense. Like the $Mod_{root(volition/obligation/ability)}$ verbs, the infinitive cannot escape its merge position in order to check tense.

5 CONCLUSIONS

In this paper I have put forward a syntactic analysis of infinitive structures with modal and aspectual verbs. First of all, following Rizzi (1976, 1978) and Cinque (2004), I provided further evidence in order to confirm that such infinitive structures are instances of RSs. In addition, I explained the problem of the (in)compatibility of modal and aspectual verbs with a past infinitive in the light of Cinque's (1999) theory on functional projections.

Furthermore, the different merge position of the classes of modal and aspectual verbs can account for this (apparent) idiosyncrasy. In this sense, the $Mod_{epistemic}$ and the $Mood_{evidential}$, being generated above Tense, can occur with a past infinitive, while the $Mod_{root(volition/obligation/ability)}$ and all aspectual verbs, being generated below Tense, cannot. In order to check Tense, these would have to escape their position and re-merge across the head of the matrix verb, violating the head movement constraint.

A syntactic analysis of the RSs also sheds light on the problem of Latin word order, at least in the period taken into account in the present research. The data demonstrate that there is great variability in word order in Latin, but this freedom is not unrestricted. BTL, in fact, clearly shows that in an unmarked order the predicate is placed at the end of the clause; furthermore, in the presence of a single, complex predicate, the unmarked order is that of ‘finite verb-infinitive’. These are both, in fact, statistically prevalent. These data allow us to conclude that also in Latin it is necessary to distinguish an unmarked word order which is predominant from a quantitative point of view.

BIBLIOGRAPHY

- Bortolussi, Bernard. forthcoming. “Generative Grammar and the Didactics of Latin. The use of Examples”. In R. Oniga, R. Iovino, G. Giusti (eds), *Formal Linguistics and the Teaching of Latin. Theoretical and Applied Perspectives in Comparative Grammar*, New Castle Upon Tyne, Cambridge Scholar Publishing.
- Burzio, Luigi. 1986. *Italian Syntax. A Government-Binding Approach*. Dordrecht: Reidel.
- Cardinaletti, Anna, and Michael Starke. 1999. “The Typology of Structural Deficiency: On the Three Grammatical Classes”. In H. van Riemsdijk (ed.), *Clitics in the Languages of Europe*, Berlin: Mouton de Gruyter, 145–233.
- Cecchetto, Carlo, and Renato Oniga. 2002. “Consequences of the Analysis of Latin Infinitival Clauses for the Theory of Case and Control”. *Lingue e Linguaggio* 1.151–189.
- Cinque, Guglielmo. 1999. *Adverbs and Functional Heads*. Oxford/New York: Oxford University Press.
- Cinque, Guglielmo. 2001. “‘Restructuring’ and the Order of Aspectual and Root Modal Heads”. In G. Cinque and G. Salvi (eds.), *Current studies in Italian syntax. Essays offered to Lorenzo Renzi*. Amsterdam: North Holland 137–155.
- Cinque, Guglielmo. 2002. “A Note on “Restructuring” and Quantifier Climbing in French”. *Linguistic Inquiry* 33.617–636.
- Cinque, Guglielmo. 2004. “‘Restructuring’ and Functional Structure”. In A. Belletti (ed.), *Structures and beyond. The cartography of Syntactic Structures. Vol 3*. Oxford/New York: Oxford University Press 132–191.
- Costantini, Francesco, and Luigi Zennaro. 2010. Some aspects of the syntax and semantics of the *verba voluntatis*, in M. Lenoble, D. Longrée et C. Bodelot (eds), *De linguae latinae usu*, Actes du 13e Colloque international de Linguistique.
- latine, Bruxelles et Liège, 4-9 avril 2005, (Bibliothèque d’Études Classiques), Leuven, Peeters.
- Devine, Andrew M., and Laurence D. Stephens. 2006. *Latin Word Order, Structured Meaning and Information*. Oxford: Oxford University Press.
- Rizzi, Luigi. 1976. “Ristrutturazione”. *Rivista di Grammatica Generativa*, 1.1–54.
- Rizzi, Luigi. 1978. “A Restructuring Rule in Italian Syntax”. In S. J. Keyser (ed.), *Recent Transformational Studies in European Languages*. Cambridge: The MIT Press.
- Salvi, Giampaolo. 2004. *La formazione della struttura della frase romanza. Ordine delle parole e clitici dal latino alle lingue romanze*. Tübingen: Niemeyer.

Salvi, Giampaolo. 2005. "Some firm point on Latin word order: the left periphery". In K. É. Kiss (ed.), *Universal Grammar in the reconstruction of ancient languages*. Berlin: Mouton de Gruyter. 429-456.

Zennaro, Luigi. 2004. *La sintassi dei verbi a ristrutturazione in latino*. PhD Thesis, University Ca' Foscari of Venice, ms. Available on line at <http://lear.unive.it/bitstream/10278/1337/1/ZENNARO%20LUIGI.pdf>.

ACKNOWLEDGMENTS

I would like to thank Anna Cardinaletti, Guglielmo Cinque, Giuliana Giusti, Renato Oniga, Anna Pompei and Imre Szilágyi for reading the text and providing helpful comments. I also thank my colleagues Michele Brunelli, Marco Coniglio, Francesco Costantini and Iulia Zegrean for discussing various data with me and for reviewing my English in a preliminary draft of this work. I am solely responsible for any remaining mistakes.

Margrét Jónsdóttir

Archaisms in comparative forms of Modern Icelandic adjectives

1 INTRODUCTION

The purpose of this article is to shed light on the use of archaic declensional forms of the comparative degree of adjectives in Modern Icelandic. The archaic comparatives appear mainly in four semantic groups of nouns:¹

- (1) a. Place names: Magnús átti heima á *Syðra-Hóli*.
Mod. language: ... *Syðri-Hóli*.
'Magnús lived on *Syðri-Hóll* ('Southern-Hill').'
- b. Kinship terms: *Eldra* soninn langar í ís.
Mod. language: ...*Eldri* soninn...
'The older son wants ice cream.'
- c. Body part nouns: Ég meiddist á *vinstra* fæti.²
Mod. language: ...*vinstri* fæti.
'I hurt my left foot.'
- d. Proper names: "Ég hef selt hann *yngra* Rauð..."
Mod. language: ...*yngrri* Rauð...
'I have sold the *younger* Red [the horse].'

According to, e.g., Saeed (2003:23-32), the difference between proper names and common nouns is found in definiteness: Proper names are inherently definite and they refer to individuals, to persons. Common nouns, on the other hand, refer to sets of individuals; they are indefinite, but can become definite depending on their syntactic context.

Van Langendonck (2007:202-203) claims that "place names are undoubtedly the most important category of names" after personal or proper names. His claim is clearly based on the

¹ The first three sentences in (1) are constructed of the article's material. Phrases like *Syðra-Hóli* (a), *eldra* soninn (b), and *vinstra* fæti (c) can easily be found on the web. The last sentence (d) is a line from a famous nineteenth century poem by Páll Ólafsson (1984:162).

² In Icelandic, *vinstri* 'left' and *hægri* 'right' behave as comparatives.

frequent identity of their syntactic behaviour. However, he claims that place names are more often used in prepositional phrases than proper names. That is the main difference, and a very important one.³

All the nouns in (1), it should be emphasized, can be said to be definite in some way or other, not just the place/proper names. This is related to the fact that in Icelandic, names for body parts and kinship terms largely exhibit the syntactic behaviour of definite nouns. With these nouns, the old declension of the comparative degree can still be used, although the sources reveal that the modern declension established itself in the eighteenth century at the very latest. Of course, the modern declension is also used, and may be more common, so there is choice between two forms. However, in the comparative degree form used with all ordinary nouns, the development has not been the same because the modern declension is used exclusively.⁴

The article is organized as follows: In section 2, the declension of adjectival comparative degree forms in older forms of the language and modern Icelandic are introduced and a survey of its history is attempted. Section 3 describes the theories this article is based on. The discussion introduces theories of analogy where a form, here a case, and a function, or meaning, play important roles. The main difference between common nouns and proper names in a wide sense is introduced. In section 4, the data are introduced and examined in the light of the theories introduced in section 3. Section 5 concludes the article with a summary of findings.

2 DECLENSION OF ADJECTIVAL COMPARATIVE DEGREE IN ICELANDIC

2.1 Comparative declension in Old and Modern Icelandic

The declension of the comparative degree in Old Icelandic is presented in (2); cf. Noreen (1923:297). In the masculine singular, there was a difference between nominative and the oblique cases. In feminine and neuter singular, there was only one ending for all cases. In the plural, the category of *gender* was neutralized and there were three endings. Let's look at the comparative degree *breiðari* of *breiður* 'broad':

(2)	Masculine	Feminine	Neuter
sg			
N	<i>breiðari</i>	<i>breiðari</i>	<i>breiðara</i>
A/D/G	<i>breiðara</i>	<i>breiðari</i>	<i>breiðara</i>
pl			
N/A		<i>breiðari</i>	
D		<i>breiðurum</i>	
G		<i>breiðari</i>	

³ Van Langendonck (2007:321) bases his treatment "on theory of proper names as they function in modern Western European languages, especially English and Dutch."

⁴ This is further investigated in Jónsdóttir (2010).

The declension of the comparative degree in Modern Icelandic is much simpler than the old one, cf. (3); here we have two endings in the singular but only one in the plural.

(3)	Masculine	Feminine	Neuter
sg	<i>breiðari</i>	<i>breiðari</i>	<i>breiðara</i>
pl		<i>breiðari</i>	

In the modern language, neuter singular has a special status: the ending is *-a*. All other forms have the ending *-i*. The change in declension from Old Icelandic to Modern Icelandic can be described as follows:

- (4) a. The opposition *masculine/feminine* is neutralized in the singular.
- b. The category *case* is neutralized.

It could be concluded from Noreen's (1923:298) words, in his Old Icelandic Grammar, that the declension of the comparative started to change early⁵, in other words, it began to be simplified. The change started in the dative plural.⁶ The whole process took a long time. According to Kvaran (2005:410), the new declension started to appear in texts in the fifteenth century. But from the Icelandic translations of the Bible it could be claimed that the sixteenth century is the time of change. In the *New Testament* (1540), the first Bible translation into Icelandic, the old declension is almost always used (Helgason 1929). On the other hand, in *The Bible* (1584) = *Guðbrandsbiblíja* (the Old and New Testaments), the new declension is used more extensively than before (Bandle 1956). According to Þórólfsson (1925:91), it is not until the eighteenth century that the new declension replaced the old one. However, Bandle (1956:321) claims that in eighteenth century literature there are still many examples of the old declension.

2.2 Comparative degree and other weak adjectival forms

While the new declension of the comparative degree was establishing itself, other weak adjectival forms also underwent change. Our sources show that the weak declension of the positive and superlative degree of the adjectives on the one hand and the present participle of verbs on the other also changed. There is some evidence that the comparative forms were given the same endings as the weak declension of positive and superlative degree. Some of these forms are still in use and, as a matter of fact, very common, cf. *meiru* 'more' and *minnu* 'smaller'.⁷ In the history of the language, there are also examples of mergers between the declension of the comparative degree and the present participle of verbs. Such examples, however, are not many.

2.3 Comparative degree in Modern Icelandic

On the basis of *Íslensk orðtíðnibók* (1991:1166-1171), it can be concluded that the comparative degree is the rarest degree; and of the comparative forms, the masculine gender is the rarest. In

⁵ Or in Noreen's words: " Im aisl. kommt selten (später häufiger) dat. pl. auf -e, -i vor; ..." This means earlier than 1350, cf. p. 8 in Noreen.

⁶ Despite this fact, the old dative plural comparative form *fleirum* (of *margur* 'many') is still in use.

⁷ See e.g.: Ferðamenn eyða *minnu/meiru* en áður 'tourists spend *less/more* (money) than before'.

Icelandic, according to Rögnvaldsson (1990:67), the positive degree is the most common; it is unmarked, while the comparative is the most marked one. This is because it is not used as much as the other two degrees. The declension of the comparative degree is only weak but the other degrees have both weak and strong declensions.

3 THEORETICAL BACKGROUND

This section focuses on the theories on which this article is based. The theories that will be discussed concern analogy where a form, e.g. a case form, and a function, or meaning, play important roles. The crucial difference lies somewhere in between common nouns and proper names in a wide sense.

3.1 Kuryłowicz's and Mańczak's theories of analogy

Kuryłowicz's (1947) fourth law of analogy describes linguistic change by analogy, specifically, the interplay between form and function and its effect on the development process. The fourth law is presented in this following formulation (5) in Collinge (1985:249):

- (5) Given a morphological derivation resulting in two differentiated forms, the derived form takes over the primary function and the old form is reserved for secondary function. (Collinge 1985:249)

The new form is used in a primary function, e.g., as a common noun. The old form is used in a secondary function, e.g., as proper name or place name. Some examples of Kuryłowicz's fourth law are shown in (6); cf. (1a).

(6)	Primary function	Secondary function
A	common noun in dative case: <i>hól-Ø</i> from <i>hóll</i> (N)	place name in dative case: <i>Hól-i</i> from <i>Hóll</i> (N)
B	comparative in dative case with common noun: <i>syðr-i</i>	comparative in dative case with common place name: <i>syðr-a</i>

Mańczak's (1958:388-401) so-called eighth tendency of analogical changes is more or less the same as Kuryłowicz's fourth law. In (7) A, Mańczak's eighth tendency is presented. In B, a translation by Hock (1991:232-233) is presented:

(7) A
 S'il y a une différence entre la flexion d'un nom géographique et celle d'un nom commun, par ailleurs semblables, les cas locaux présentent en général un caractère archaïque tandis que dans les cas non locaux il y a le plus souvent des innovations. (Mańczak 1958:388)

B
 If there is a difference between the inflection of a geographic noun and a common noun, which otherwise are similar, the local cases generally present an archaic character, while in the non-local cases innovations are more common. (Hock 1991:232-233)

The result of this can be seen in (8) where the archaic character is preserved in the dative case in two ways, cf. *syðra* and in the dative ending *-i*:

(8)	Common noun	Place name
N	syðri hóllinn	Syðri Hóll
A	syðri hólinn	<i>Syðra/Syðri Hól</i>
D	syðri hólnum	<i>Syðra/Syðri Hóli</i>
G	syðri hólins (‘the more southern hill’)	<i>Syðra/Syðri Hóls</i> (‘Southern-Hill’)

3.2 Croft's view

Croft's theory is based on semantics, where case ~ function stand in opposition. Croft (2003:165) discusses *markedness reversal*: If an unmarked nominative is semantically independent, the unmarked nominative is used in a primary function. However, if an unmarked dative is semantically dependent, the unmarked dative is in a secondary function. This means that the dative is unmarked or the default case in the category of local phrases.

3.3 Intermediate summary

Everything that has been said here is valid for Icelandic. As we have seen, common nouns and proper names for persons and places have a different status. This can be seen in the difference between the dative case *hól* when it is used as a common noun vs. *Hóli* as a place name, also in dative. The dative ending *-i*, cf. *Hóli*, is the older one and here it is preserved in a narrower function than the new ending, *-Ø*, which is used on the common noun.

4 OLD COMPARATIVE DEGREE FORMS: EVIDENCE FROM MODERN ICELANDIC

This section will be in six parts. The first five concern place names, proper names, kinship terms, words of inalienable possession and permanent relationship, and names for body parts. The old and new declension of the comparative form in question will be discussed and shown to be freely interchangeable. The last part concludes with a summary.

4.1 Place names

In (8), the declension of the common noun *syðri hóllinn* vs. *Syðri-Hóll* as a place name can be seen. In the oblique cases of *Syðri-Hóll*, either the new or the old form of *syðri* can be used.

According to Tiersma (1982:843), the dative case is the most common case for place names, and is therefore the default one. This is what Tiersma calls *local markedness*. That can easily be seen in *Íslensk orðtíðnibók* (1991:1156–1157): The dative case singular is the most common case for place names, or 56,2%.⁸ Therefore, place names and their markedness relations are in contrast with common nouns, for which nominative case is the default case.⁹ The question now arises if the nominative could be *Syðra-Hóll* due to the influence of the dative case, the most common one. Theoretically, this could be the case, but no examples of this particular form are found, although we have similar examples where the oblique form has been lexicalized.¹⁰

4.2 Proper names

In (9) are shown examples of comparatives with proper names. *Jón* is a man's name, cf. (9a).¹¹ (9b) contains the name of a horse, *Rauður*.¹² Theoretically, Páll Ólafsson, the poet, could have used the new declension, but the old one was acceptable and still is.

(9)		New
a.	Sannarlega snerti ránið Jón <i>eldra</i> (A). 'The robbery certainly concerned the elder Jón.'	<i>eldri</i>
b.	"Eg hef selt hann <i>yngra</i> Rauð (A) ..." 'I have sold the <i>younger</i> Red [the horse].'	<i>yngrri</i>

4.3 Words of kinship

So far, the behaviour of proper names and places names has been discussed. Kinship terms refer to persons, and are in that way similar to proper names because their reference is deictic (Anderson 2004:442). Therefore, they are definite in themselves and have the same status as proper names and place names. The two examples in (10) illustrate the two choices for the declension of the comparative:¹³

(10)		New
a.	Parna sérðu <i>eldra</i> (A) bróður minn.	<i>eldri</i>

⁸ The results for the other cases are: 10,4% (N), 16,4% (A), 17% (G).

⁹ The dative case is the most common one for place names because they are mostly used in prepositional phrases. In the Sagas and folktales, a name of a farm is in the dative with a preposition. A quotation from *Laxdæla saga* (1934:161): "Hvat heitir bær sjá?...sjá bær heitir í Tungu (D)." ('What is the name of the farm? ...The farm is called Tunga.'). In Modern Icelandic, the sentence would be: Þessi bær heitir *Tunga* (N).

¹⁰ A wellknown example of this is *Breiðafjörður* where the weak adjective form *breiða* (*breiður* 'broad') is an oblique case of *breiði*; the meaning is 'the broad fjord'. Furthermore, the same problem could be seen in *Borgarfjörður eystri* or *eystra*, see Jónsdóttir (2009).

¹¹ <http://www.itu.dk/people/astaolga/null/tyrkjaranid/heimildamynd/pdf-skjol/Stortidindafrasogn.pdf>

¹² The sentence (b) is a line from a famous nineteenth century poem, see note 1.

¹³ Phrases with *eldra* bróður, cf. (10), can be found on the web.

- ‘There you see my older brother.’
 b. Þarna sérðu Jón, *eldra* (A) bróður minn. *eldri*
 ‘There you see my older brother, Jón.’

4.4 Inalienable possession and permanent relationship

As Croft (2003:205) and others have suggested, it is well known and very common that words denoting *inalienable possession* have a special status. *Inalienable possession* means literally ‘a possession which is not alienable, cannot be taken or given away’. Or in Croft’s words: “Inalienable possession refers to a permanent relationship between two entities” (Croft 2003:205). Kinship terms and names for body parts are among the words denoting inalienable possession. They are connected to the “owner” (if that word can be used) through possessive markers or with an antecedent of some kind or another. We have seen how the old and the new declension of the comparatives are used with kinship terms. The same is also true for names of body parts.¹⁴

4.5 Names for body parts

As said in (4.4), words for body parts are most often connected to their owner. In Icelandic, they are never agents and used mostly in a locative meaning: We talk about our feet, teeth, arms, face, etc. and generally use the oblique cases to do so. It is a fact that most common cases for body parts are (singular and plural) the accusative (46,7%) and the dative (38,2%), or total 84,9%.¹⁵

In (11), the old and new declension of the comparatives with words for body parts can be seen. Names for body parts used metaphorically, cf. (11b) and (11c), are used in the same way.¹⁶

- | | | |
|------|---|----------------|
| (11) | | New |
| a. | Ég meiddi mig á <i>vinstra</i> (D) fæti.
‘I hurt my left foot.’ | <i>vinstri</i> |
| b. | Þeir eru í <i>vinstra</i> (D) armi flokksins.
‘They belong to the party’s right wing.’ | <i>vinstri</i> |
| c. | Það er eldur í <i>hægra</i> (D) væng þotunnar.
‘There is a fire in the right wing of the jet.’ | <i>hægri</i> |

Words for body parts are used as place names, or what Campbell (1999:240) refers to as *locative constructions*. Recall what was said about the dative case and place names in section 3.2: The dative case is the most common case because place names are mostly used in prepositional phrases in Icelandic. The scenario is very similar to that found in names of body parts.

¹⁴ In Icelandic syntax, words for body parts and those of kinship behave like definite nouns or proper names.

¹⁵ The results for the other cases are: 11, 3% (N) and 3,8 % (G) or total 15,1%. This can be concluded from *Íslensk orðtíðnibók* (1991). The frequency values cover the 15 most common words for body parts.

¹⁶ Modern examples of *vinstra* fæti, *vinstra* armi, and *hægra* væng, cf. (11), can be found on the web.

4.6 Intermediate summary

In this section, the focus was on the data and they were studied in the light of the theories presented in section 3. The use of comparatives with place names in 4.1 and proper names in (8) in 3.1 are clear examples of secondary vs. primary function, cf. Kuryłowicz's theories of analogy in 3.1. They can also be examined in the light of Mańczak's and Croft's interpretations. The oblique cases of the place name *Syðri-Hóll*, not only the comparative form but also the noun's case forms show this clearly. Let's take the sentence (1d). "Eg hef selt hann *yngra Rauð* (A) ... " as an example. *yngra Rauð* is the accusative; the nominative case is *yngrri Rauður*, where *Rauður* is a nominalized adjective, see *rauður* 'red'. The kinship terms and names of body parts can also be considered in the light of Kuryłowicz's, Mańczak's and Croft's views of primary vs. secondary function. However, it is also interesting to look at kinship terms and names of body parts from the viewpoint of Croft's inalienable possession and permanent relationship, cf. 3.3 and the examples in 4.3 and 4.5. In his opinion, they could be considered equivalent to proper names.

5 CONCLUSION

In this article, some examples of the old declension of comparatives were discussed, which are still used in the modern language. The discussion is based on theories of analogy where a form, here a morphological case, and a function, or in other words meaning, play important roles. Research in Icelandic by the present author and others has revealed that comparatives still used in Icelandic are more or less the same formally as in *The Bible* 1584, the first translation of the whole Bible into Icelandic.

Some of the comparatives show suppletion, that is when the positive and comparative degrees are from different roots. Comparatives like *minni* (of *lítill* 'small') or *eldri* (of *gamall* 'old') are examples of this. There are also examples of comparatives without either the positive or the superlative degree like, *vinstri* 'left' and *hægri* 'right', or without positive degree, e.g. *syðri* 'southern'.

The question presents itself if this twofold system makes the language more complicated than otherwise would have been the case. It does not have to be interpreted so. Preserving the old declension of comparison only in special semantic categories of nouns, denoting proper and place names, names for body parts and kinship terms, introduces clarity, and better defines their special status.

BIBLIOGRAPHY

- Anderson, John. M. 2004. On the grammatical status of names. *Language* 80.435-474.
Bandle, Oscar. 1956. *Die Sprache der Guðbrandsbiblíá*. Kopenhagen: Munksgaard.
Campbell, Lyle. 1999. *Historical linguistics. An introduction*. Cambridge, Massachusetts: MIT Press.
Collinge, N. E. 1985. *The laws of Indo-European*. Philadelphia: Amsterdam, John Benjamins.

- Croft, William. 2003. *Typology and universals*. Second edition. Cambridge: Cambridge University Press.
- Helgason, Jón. 1929. *Málið á Nýja testamenti Odds Gottskálkssonar*. Kaupmannahöfn: Hið íslenska fræðafélag.
- Hock, Hans Henrich. 1991. *Principles of historical linguistics*. Berlin, New York: Mouton de Gruyter.
- Íslensk orðtíðnibók. 1991. Reykjavík: Orðabók Háskólans.
- Jónsdóttir, Margrét. 2009. Á Borgarfirði eystri - á Borgarfirði eystra. Hvaða orðflokki tilheyrir eystra? Orð og tunga 11.143-158.
- Jónsdóttir, Margrét. 2010. Forn miðstigsbeyging í nútímamáli. To appear in *Íslenskt mál* 32.
- Kuryłowicz, Jerzy. 1947. La nature des procès dits analogiques. *Acta Linguistica* 5.15-37.
- Kvaran, Guðrún. 2005. *Orð. Handbók um beygingar- og orðmyndunarfræði*. Reykjavík: Almenna bókafélagið.
- Laxdæla saga. 1934. *Laxdæla saga*. Halldórs þættir Snorrasonar. Stúfs þátrr. Einar Ól. Sveinsson ritaði formála. Reykjavík: Hið íslenska fornritafélag.
- Mańczak, Witold. 1958. Tendances générales des changements analogiques. *Lingua* 7. 387-420.
- Noreen, Adolf. 1923. *Altisländische und altnorwegische Grammatik (Laut- und Flexionslehre)*. Vierte vollständig umgearbeitete Auflage. Halle (Saale): Max Niemeyer.
- Ólafsson, Páll. 1984. *Kvæði*. Hafnarfjörður: Skuggsjá.
- Rögnvaldsson, Eiríkur. 1990. *Íslensk orðhlutafræði*. Reykjavík: Málvísindastofnun Háskóla Íslands.
- Saeed, John I. 2003. *Semantics*. Second edition. Oxford: Blackwell.
- Tiersma, Peter Meijes. 1982. Local and general markedness. *Language* 58.832-849.
- van Langendonck, Willy. 2007. *Theory and typology of proper names*. Berlin, New York: Mouton de Gruyter.
- Þórólfsson, Björn K. 1925. Um íslenskar orðmyndir á 14. og 15. öld og breytingar þeirra úr fornálinu. Með viðauka um nýjungar í orðmyndum á 16. öld og síðar. Reykjavík: Fjelagsprentsmiðjan.

5. Language technology

Mojca Kompara

Automatic recognition of abbreviations and abbreviations' expansions in electronic texts – future development

1 INTRODUCTION

Abbreviations are difficult to deal with (Gabrovšek 1994) and represent a growing phenomenon present in all languages. The scope of this article is to present the insufficient and inefficient inclusion of abbreviations in dictionaries and the fact that neither paper nor online dictionaries are published or updated frequently enough to cover new abbreviations among their entries. On the other hand manual acquisition of abbreviation data and manual inclusion of abbreviations into online dictionaries is time consuming. The solutions are algorithms for automatic recognition of abbreviations and abbreviations' expansions in electronic texts. In the present paper such algorithm is presented.

2 ABBREVIATIONS IN DICTIONARIES

The scope of this chapter is to present the insufficient and inefficient inclusion of abbreviations in dictionaries by focusing on the inclusion of abbreviations in monolingual, bilingual, specialised and online dictionaries. Special emphasis is devoted to the microstructure of specialised abbreviations' dictionaries. Abbreviations are usually included among the entries or/and in supplements of monolingual dictionaries of English (*Collins COBUILD English Dictionary* (Sinclair 1999)), Italian (*Vocabolario della lingua italiana* (Zingarelli 2000)), German (*Deutsches Universal Wörterbuch* (Drosdowski 1989)), Spanish (*Clave: Diccionario de Uso del Español* (García Marquez 2002)) and French (*Le Nouveau Petit Robert* (Robert 1996)) but they are not included among the entries of the Slovene monolingual dictionary *Slovar slovenskega knjižnega*

jezika (Bajec *et al.* 1970–1991). As far as bilingual dictionaries are concerned, abbreviations are usually not included in older publications of dictionaries for encoding, such as *Slovensko-angleški slovar* (Grad and Leeming 1997), *Slovensko-italijanski slovar* (Kotnik 1992), *Slovensko-španski slovar* (Grad 2000), *Slovensko-francoski slovar* (Jesenik 2005) but the exceptions are some new publications, *Veliki slovensko-italijanski slovar* (Šlenc 2006) and *Veliki slovensko-nemški slovar* (Debenjak 2003). On the other hand, abbreviations are included in dictionaries for decoding, such as *Veliki angleško-slovenski slovar* (Grad 1998), *Veliki angleško-slovenski slovar Oxford-DZS* (Gabrovšek *et al.* 2005–2006), *Veliki italijansko-slovenski slovar* (Šlenc 1997), *Veliki nemško-slovenski slovar* (Debenjak 2001), *Špansko-slovenski slovar* (Grad 2001) and *Francosko-slovenski slovar* (Grad 2004). The main problem of all mentioned dictionaries is the fact that they are not published frequently enough to cover all new abbreviations among their entries and abbreviations are not covered extensively.

2.1 English dictionaries of abbreviations

In English I observed several dictionaries of abbreviations that are listed here: *Dictionary of Abbreviations* (Paxton 1983), *The New Penguin Dictionary of Abbreviations* (Fergusson 2000), *Abbreviations Dictionary* (De Sola 1986), *World Guide to Abbreviations of Organizations* (Buttress 1976), *Dictionary of Abbreviations and Acronyms* (Dale, Puttick 1999), *Abbreviations and Acronyms* (Geddes and Grosset 1999), *Acronyms, Initialisms and Abbreviations Dictionary* (Gale Research Company 2006), *The Barnhart Abbreviations Dictionary* (Barnhart 1995), *The Dictionary of Acronyms and Abbreviations in Applied Linguistics and Language Learning* (Jung 1991), *Elsevier's Dictionary of Acronyms, Initialisms, Abbreviations and Symbols* (Benedetto Mattia 1997). The characteristics of their microstructure are visible in table 1.

Table 1

Index: X – no; √X – yes (not consistently); √R – yes (rarely); √√ – yes (consistently); √E – yes (extensive); √ – yes; √TE – yes (too extensive)

Name	Entries	Suppl.	Foreign abbr.	Translation	Qualifier	Lang.	Encyclop. data
(Paxton 1983)	25,000	√	√	√√	√	√X	√
(Fergusson 2000)	40,000	√E	√	√√	√	√	√
(De Sola 1986)	23,000	√E	√	√√	X	√	√R
(Buttress 1976)	60,000	X	√	X	X	X	X
(Dale 1999)	20,000	√TE	√	√√	√	√	√
(Geddes and Grosset 1999)	10,000	X	√	√√	√	√	X
(Gale Research Company 2006)	1,000,000	X	√	√√	√	X	√
(Barnhart 1995)	60,000	X	√	√√	√	√	√
(Jung 1991)	50,000	X	√	X	√X	X	X
(Mattia 1997)	60,000	√	√	X	X	X	X

The analysed and presented dictionaries differ in size but also in microstructure, especially in terms of translations which are not always present in entries or not present at all. Translation of abbreviations is an essential data that should always be consistently included in specialised dictionaries. A dictionary without translations does not solve problems but causes new ones. As seen from the table above, some dictionaries do not have supplements while others have too extensive ones. There is also a lot of inconsistency in terms of qualifiers, language and encyclopaedic data. Reverse dictionaries or just supplements such as *Reverse Acronyms Initialisms & Abbreviations Dictionary* (Gale Research Company 2007) or *The Barnhart Abbreviations Dictionary* (Barnhart 1995) are extremely useful for users but still missing in most dictionaries. The main problem of all dictionaries is the fact that they are too old and not published frequently enough to include new abbreviations.

2.2 German, Italian, Spanish and French dictionaries of abbreviations

I observed some Italian, German, Spanish and French dictionaries of abbreviations that are listed below. *Das Wörterbuch der Abkürzungen* (Steinhauer 2005), *Großes Abkürzungsbuch* (Koblichke 1983), *Dizionario di Sigle Abbreviazioni e Simboli* (Righini 2001), *Dizionario delle Sigle e degli Acronimi* (Malossini 1999), *Dictionnaire des Abreviations et Acronymes* (Murith and Bocabeille 1992), *Diccionario General de Abreviaturas Españolas* (Galende 1997, 2001). The characteristics of their microstructure are visible in table 2.

Table 2

Index: X – no; √X - yes (not consistently); √R - yes (rarely); √√ - yes (consistently); √E - yes (extensive); √ - yes; √TE - yes (too extensive)

Name	Entries	Suppl.	Foreign abbr.	Translation	Qualifier	Lang.	Encyclop. data
(Steinhauer 2005)	50,000	√E	√	X	√	√X	X
(Koblichke 1983)	50,000	√E	√	√	√	√	√
(Righini 2001)	10,000	X	√	√	√	√	√
(Malossini 1999)	8,000	X	√	X	√X	X	X
(Murith and Bocabeille 1992)	20,000	X	√	X	X	X	X
(Galende 1997, 2001)	20,000	X	X	X	X	X	X

There are even more differences in terms of microstructure in dictionaries of abbreviations presented above compared to English dictionaries. This concerns especially translations that should be essential data in such dictionaries. All dictionaries are old and do not include new abbreviations. As far as Slovene is concerned there is still no updated abbreviation dictionary. The first one *Kratice, Mala izdaja* (Župančič) was written in 1948. Fortunately we have an online

dictionary of abbreviations *Slovarček krajšav*,¹ containing over 5.000 entries and providing translations of all foreign abbreviations. Abbreviations represent a growing phenomenon present in all languages. New ones arise suddenly and on one hand paper dictionaries do not dedicate enough space and attention to them among the entries. Online dictionaries such as *Slovarček krajšav*,² *Evroterm*,³ *Acronym Finder*,⁴ *The Free Dictionary*,⁵ and others represent an alternative to paper ones and are easily updated. But unfortunately abbreviations in online dictionaries are acquired mainly manually which is still time consuming and not precise enough. The alternative is automatic recognition of abbreviations and abbreviations' expansions from electronic texts.

3 ALGORITHM FOR AUTOMATIC RECOGNITION OF ABBREVIATIONS AND ABBREVIATIONS' EXPANSIONS

The pioneer in automatic recognition of abbreviations and abbreviations' expansion is Taghva (1998). Automatic recognition was dealt with also by Yeast (1999), Larkey *et al.* (2000), Pustejovsky *et al.* (2001), Schwartz and Hearst (2003), Park and Byrd (2001), Chang *et al.* (2002) and Zahariev (2004), his approach is considered special due to the fact that he is recognising abbreviations in several different languages. In the early stage of my research words of five capital letters written in brackets were used as abbreviations' candidates, e.g. (MDDSZ). The reference was the Slovene online newspaper *Delo* from 2007. It had 25,588 such words and some occurred more than once. In order to come across a desired amount of abbreviations for further research I excluded all abbreviations that occurred more than once and all abbreviations that are not acronyms and/or abbreviations written in capital letters. Words that are not abbreviations, such as proper names, names of places etc. were excluded, using the Slovene monolingual dictionary, after the exclusions the database had over 4,000 abbreviations. After the exclusion of all abbreviations not written in capital letters and considering also the context that should not be written in capital letters, the final list had 2,500 abbreviations. In order to obtain the abbreviations' expansions from the newspaper *Delo*, the left context was investigated, because usually expansions are placed before the abbreviations, e.g. *European central bank (ECB)*, but still not excluding the possibility of right context. To recognize expansions four types of abbreviations were used. The first type are the so called *covered abbreviations* where letters match the words in the left context, e.g. *FF* with the expansion *Filozofska fakulteta*. The second type are *abbreviations with expansions containing prepositions and conjunctions*, e.g. *FDV Fakulteta za družbene vede*. The algorithm takes into consideration also expansions with one or two additional words. The third type concerns *abbreviations composed of the first two letters*, e.g. *NAMA Narodni magazin*. The fourth type covers *abbreviations with prepositions*, e.g. *DZU Družba za upravljanje* where prepositions appear in the abbreviation and also in the expansion. Considering the above

¹ <http://bos.zrc-sazu.si/kratice.html>.

² <http://bos.zrc-sazu.si/kratice.html>.

³ <http://www.sigov.si/evroterm/>.

⁴ <http://www.acronymfinder.com/>.

⁵ <http://acronyms.thefreedictionary.com/>.

In figure 2 we can notice that the system does not recognise the pattern *Družbi za avtoceste v RS (Dars)*, also after entering the modified pattern *Družbi za avtoceste v RS (DARS)*, where *Dars* is capitalised. The problem might be the prepositions *za* and *v*, if I exclude the preposition *v* and add the expansion for *RS (Republika Slovenija)* the pattern is recognised. The algorithm has to focus on the possibility of abbreviations in expansions and more than just one preposition in the expansion.



Figure 2

I entered the following text:

Preiskava Urada za varstvo konkurence (UVK), predstavniki urada so v četrtek končali delo v CPM, ni imela vpliva na napredovanje del na gradbišču predora Markovec, pojasnjujejo v Cestnem podjetju Maribor (CPM). "Delo v družbi CPM zdaj poteka nemoteno," še pišejo. Z Darsa pa dodajajo, da bo "Dars kot naročnik zahteval, da se dela opravijo v pogodbenem roku, ter ne bo ničesar zaustavljal".

In figure 3 the system does not recognise the copy-pasted pattern *Urada za varstvo konkurence (UVK)* but recognises the pattern *Cestnem podjetju Maribor (CPM)*. If the pattern *Urada za varstvo konkurence (UVK)* is retyped the system recognises it.

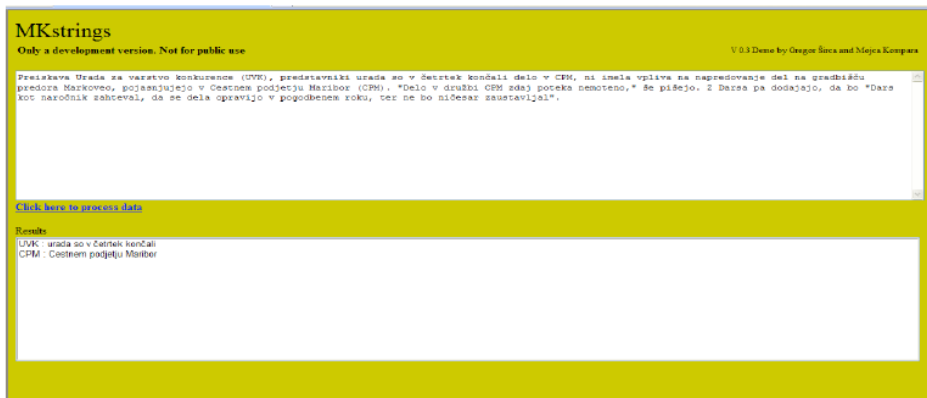


Figure 3

The algorithm does not recognise abbreviations with no expansions. Such examples are automatically excluded. An interesting issue are patterns composed of a foreign abbreviation and its Slovene expansion, e.g. *Združenje evropskih avtomobilskih proizvajalcev (ACEA)*. In the future much attention will be devoted to such patterns.

5 FOREIGN LANGUAGES

Among the analysed Slovene texts the algorithm recognised also some foreign abbreviations. To discover whether the algorithm is universal⁶ I used some English and Italian texts randomly selected and available online. In English texts the algorithm recognised patterns such as *Severe acute respiratory syndrome (SARS)*, *World Health Organization (WHO)*, *Center for Group Learning's (CGL)*, *Scottish Licensed Trade News (SLTN)* but did not recognise patterns having abbreviations in the expansion, e.g. *USB Implementers Forum (USB-IF)* and abbreviations written in a special way and/or expansions with punctuation, e.g. *Human-interface devices (HIDs)*. In Italian the situation is similar, the algorithm recognised the following patterns; *Confederazione Generale del Lavoro (CGdL)*, *Popolo della Libertà (PdL)*, *Centro Interdipartimentale di Studi Europei e Mediterranei (CISEM)* but it did not recognise patterns such as *Azienda Nazionale Autonoma delle Strade (ANAS)* because in the text *ANAS* is preceded by the definite article *L'ANAS*. The algorithm does not recognise patterns composed of articles combined with prepositions (*preposizione articolata*), e.g. *della, per la*, as seen in the following examples; *Dipartimento di Scienze del Linguaggio, dell'Interpretazione e della Traduzione (DSLIT)*, *Centro Interdipartimentale per la Ricerca Didattica (CIRD)*.

⁶ Such approach was taken also by Zahariev (2004).

6 FINAL RESULTS

In the final step, I applied modifications and improvements from the examples and enlarged the software in order to be able to filter larger amounts of data. I used a larger corpus composed of 60 million words (newspaper *Delo* from 2005 to 2009). The algorithm filtered the corpus in 30 minutes and gave 5820 abbreviation-expansion pairs. The obtained pairs were manually revised and verified using *Google*. Among the revised and verified pairs 4% of false pairs occurred, e.g. *PO predstavljenih podatkih o, NA na vse argumente, IN in novincev*. The precision of the algorithm is 96%. Among the revised and verified pairs were also genuine abbreviations not matching with the right expansions, e.g. *HIV virusom i hepatitisom in virusom* because the expansions were missing. Among the good expansions many occurred more than once and/or with tiny modifications, e.g. usage of different cases or spelling as seen in Example 1. In Example 1 just three expansions out of six are genuine.

MNZ	MNZ
1 ministrstva za notranje zadeve	1 ministrstva za notranje zadeve
2 medobčinskih nogometnih zvez	2 medobčinskih nogometnih zvez
3 ministrstvom za notranje zadeve	3 Muzej novejše zgodovine
4 Medobčinske nogometne zveze	
5 Muzeja novejše zgodovine	
6 Muzej novejše zgodovine	

Example 1

After the exclusion of false pairs, verification and revision of good pairs, 2665 genuine abbreviations-expansion pairs occurred. Among the good pairs there were also some foreign pairs although the recognition was focused only on Slovene texts. Among the foreign pairs some problems occurred in misrecognition of parts of expansions, e.g. *FEE for Environmental Education*, where *Foundation* is missing. Some Slovene expansions need accurate verification and help of experts, e.g. *DIIP 1 dokument identifikacije investicijskega projekta 2 dokument o identifikaciji investicijskega projekta*.

7 DICTIONARY PRODUCTION AND THE FUTURE

The 2665 genuine abbreviation-expansion pairs represent the data for the production of a dictionary of abbreviations. Simple entries usually consist of an abbreviation and an abbreviation expansion, such as *FF Filozofska fakulteta*, and can be produced entirely automatically using appropriate software that would convert such entries from the plain database to an attractive dictionary layout. A more accurate approach is needed for complex entries such as *EU angl.: European Union: Evropska unija*, where a translation and language qualifier are needed. But such entries can be even more complex, offering specific qualifiers and encyclopaedic data. In the future I will focus especially on such examples. I will also try to achieve a universal approach for identifying abbreviations by taking in consideration foreign

languages. Having had the opportunity to work in the translation unit in the European Institutions in Brussels I got the possibility to use a large corpus of EU texts in all 22 European languages that are translated on a daily base. The future aim of the algorithm is to be able to recognise abbreviations in those languages. Special emphasis will be dedicated also to some special abbreviation types that were not recognised by the algorithm at the present stage, such as abbreviations that are not capitalised. I will consider also different types of spelling of abbreviations, such as *KAD* or *Kad*, *SOD* or *Sod* or also a combination form, such as *BiH*, and the lexicalization of them. In further development I will focus also on abbreviation-expansion pairs where abbreviations occur in one language and expansions in another, e.g. *NATO*. The abbreviation *NATO* could be written in Slovene in two ways; *NATO* or *Nato*. In Slovene the translations *Organizacija severnoatlantskega sporazuma*, *Severnoatlantska pogodbeni zveza* or *Organizacija severnoatlantskega pakta* are used. The algorithm is able to recognise just the abbreviation-expansion pair *NATO*, *North Atlantic Treaty Organization*, but in the future I will focus on the possibility to recognise translations. The automatic recognition of abbreviations and abbreviations' expansions gives the possibility to produce a dictionary automatically or semi-automatically. The production depends on the type of the dictionary entry. In the future, I will focus mainly on the complex dictionary entries and try to automatically obtain some additional data, such as translations, qualifiers, encyclopaedic data etc.

8 CONCLUSION

Abbreviations are not something new or a fashionable way of communication; they were used even by Cicero (Kompara 2005). Abbreviations became part of our everyday life and are produced on a daily base. Their number is rapidly growing, but, as seen in the analysis, neither paper nor online dictionaries are published or updated frequently enough to cover new ones among their entries. Manual acquisition of abbreviation data and manual inclusion of abbreviations into online dictionaries is time consuming; the solutions are algorithms for automatic recognition of abbreviations and abbreviation's expansions in electronic texts. In the present paper such a system is described and the demo version of the algorithm is presented. In the paper the guidelines for future development of rules, algorithm and software development are presented. Also the universal approach of the algorithm that will be applicable for different foreign languages is taken in consideration. Such algorithm gives the possibility to create a semi-automatic dictionary of abbreviations and such a dictionary represents the future of electronic lexicography. Algorithms for automatic recognition of abbreviations present the link between the text and the semi-automatic production of a dictionary of abbreviations. That is why the production and further development of the algorithm is essential and useful.

REFERENCES

Acronym Finder. [URL: [http:// www.acronymfinder.com](http://www.acronymfinder.com)], accessed 10.10.2010.

- Acronyms, Initialisms and Abbreviations Dictionary. 2006, accessed 10.10.2010. [URL: <http://www.gale.cengage.com/pdf/samples/sp628573.pdf>].
- Bajec, Anton. *et al.* 1970-1991. Slovar slovenskega knjižnega jezika. Ljubljana: DZS.
- Barnhart, Robert K. 1995. The Barnhart Abbreviations Dictionary, accessed 10.10.2010. [URL: <http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0471571466.html>].
- Benedetto Mattia, Fioretta. 1997. Elsevier's Dictionary of Acronyms, Initialisms, Abbreviations and Symbols. Amsterdam: Elsevier science B.V.
- Buttress, Frederick Arthur. 1976. World Guide to Abbreviations of Organizations. Detroit: Grand River Books.
- Chang, Jeffrey T. *et al.* 2002. Creating an Online Dictionary of Abbreviations from MEDLINE. Journal of American Medical Informatics Association (JAMIA), IX(VI). 612-620.
- Dale, Rodney, and Steve Puttick. 1999. Dictionary of Abbreviations and Acronyms. Hertfordshire: Wordsworth.
- Debenjak, Božidar. 2001. Veliki nemško-slovenski slovar. Ljubljana: DZS.
- Debenjak, Božidar. 2003. Veliki slovensko-nemški slovar. Ljubljana: DZS.
- De Sola, Ralph. 1986. Abbreviations Dictionary, 7th edition. New York, Amsterdam, Oxford: Elsevier.
- Drosdowski, Günther. 1989. Deutsches Universal Wörterbuch. Mannheim: Duden verlag.
- Evroterm, Večjezična terminološka zbirka. [URL: <http://evroterm.gov.si/index.php?jezik=angl>], accessed 10.10.2010.
- Fergusson, Rosalind. 2000. The New Penguin Dictionary of Abbreviations; from A to ZZ. London: Penguin.
- Gabrovšek, Dušan. 1994. Kodifikacija angleškega jezika v specializiranih enojezičnih slovarjih: Too much of everything? Vestnik XXVII/I-II. 150-180.
- Gabrovšek, Dušan. *et al.* 2005-2006. Veliki angleško-slovenski slovar Oxford. Ljubljana: DZS.
- Geddes and Grosset. 1999. Abbreviations and Acronyms. New Lanark: Geddes and Grosset.
- Galende, Juan Carlos. 1997, 2001. Diccionario General de Abreviaturas Espanolas. Madrid: Editorial Verbum.
- Garcia Marquez, Gabriel. 2002. Clave: Diccionario de Uso del Español. Madrid: Ediciones Sm.
- Google. [URL: <http://www.google.si/>], accessed 10.10.2010.
- Grad, Anton, and Henry Leeming. 1997. Slovensko-angleški slovar. Ljubljana: DZS.
- Grad, Anton. 1998. Veliki angleško-slovenski slovar. Ljubljana: DZS.
- Grad, Anton. 2000. Slovensko-španski slovar. Ljubljana: DZS.
- Grad, Anton. 2001. Špansko-slovenski slovar. Ljubljana: DZS.
- Grad, Anton. 2004. Francosko-slovenski slovar. Ljubljana: DZS.
- Jesenik, Viktor. 2005. Slovensko-francoski slovar. Ljubljana: DZS.
- Jung, Heidrun. 1991. The Dictionary of Acronyms and Abbreviations in Applied Linguistics and Language Learning. New York: Peter Lang Publishing Group.
- Koblischke, Heinz. 1983. Großes Abkürzungsbuch. Leipzig: VEB Bibliographisches Institut Leipzig.
- Kompara, Mojca. 2005. Slovensko-italijanski glosar krajšav. Ljubljana: University of Ljubljana, Faculty of Arts, Department of Translation, ba.
- Kotnik, Janko. 1992. Slovensko-italijanski slovar. Ljubljana: DZS.

- Larkey, Leah S. *et al.* 2000. Acrophile: An Automated Acronym Extractor and Server. Proceedings of the fifth ACM conference on Digital libraries. 205-214.
- Malossini, Andrea. 1999. Dizionario delle sigle e degli acronimi. Milano: Avallardi.
- Murith, Jean, and Jean-Marc Bocabeille. 1992. Dictionnaire des Abreviations et Acronymes. Paris: TEC&DOC – Lavoisier.
- Park, Youngjy, and Roy J. Byrd. 2001. Hybrid TextMin-ing for Finding Abbreviations and Their Definitions. IMB Thomas J. Watson Research Center. 167-170.
- Paxton, John. 1983. Everyman's Dictionary of Abbreviations. London: J. M. Dent & Sons.
- Pustejovsky, James. *et al.* 2001. Au- tomation Extraction of Acronym-MeaningPairs from Medline Databases. In Proceedings of Medinfo. 371-375.
- Reverse Acronyms Initialisms & Abbreviations Dictionary. [URL: <http://www.gale.cengage.com/pdf/samples/sp628662.pdf>], accessed 10.10.2010.
- Righini, Enrico. 2001. Dizionario di Sigle Abbreviazionie e Simboli. Bologna: Zanichelli.
- Robert, Paul. 1996. Le Nouveau Petit Robert. Paris: Dictionnaires le Robert.
- Schwartz, Ariel S., and Marti A. Hearst. 2003. A simple algorithm for identifying abbreviation definitions in biomedical texts. Proceedings of the Pacific Symposium on Biocomputing. 451-462.
- Sinclair, John. (ed.) 1999. Collins COBUILD English dictionary. London: HarperCollins.
- Slovarček krajšav. [URL: <http://bos.zrc-sazu.si/kratice.html>], accessed 10.10.2010.
- Steinhauer, Anja. 2005. Das Wörterbuch der Abkürzungen. Mannheim: Duden verlag.
- Šlenc, Sergij. 1997. Veliki italijansko-slovenski slovar. Ljubljana: DZS.
- Šlenc, Sergij. 2006. Veliki slovensko-italijanski slovar. Ljubljana: DZS.
- Taghva, Kazem, and Jeff Gilbreth. 1998. Recognizing acronyms and their definitions. IJDAR I(IV). 191-198 .
- The Free Dictionary. [URL: <http://acronyms.thefreedictionary.com>], accessed 10.10.2010.
- Zahariev, Manuel. 2004. A (Acronyms). Ottawa: School of Computing Science, Simon Fraser University, PhD dissertation.
- Zingarelli, Nicola. 2000. Vocabolario della lingua italiana. Bologna: Zanichelli.
- Župančič, Jože. 1948. Kratice, Mala izdaja. Ljubljana: DZS.
- Yeates, Stuart. 1999. Automatic extraction of acronyms from text. Proceedings of the Third New Zealand Computer Science Research Students' Conference. 117-124.
- 24ur.com. [URL: <http://24ur.com/>], accessed 10.10.2010.

Sara Stymne

Using a grammar checker and its error typology for annotation of statistical machine translation errors

ABSTRACT

In this study I extend an error typology for machine translation based on an error typology for a grammar checker. I use this typology to classify errors in a Swedish text machine translated from English by a statistical system. I also investigate if the grammar checker can be used to automatically identify machine translation errors, and how its performance differs from that on several human text types. I found that the extended error typology was useful for annotating the machine translation output.

1 INTRODUCTION

Texts that are produced by machine translation (MT) systems are becoming increasingly more common in today's society, mainly because of the availability of free online machine translation systems such as Google Translate¹. Google Translate, and many current state-of-the-art research systems, are phrase-based statistical machine translation (SMT) systems, that are based on statistical models trained on human translations. One problem with such systems is that the output often is ungrammatical, since there generally is no linguistic knowledge in the systems.

Machine translation systems are often only evaluated using automatic metrics, such as Bleu (Papineni *et al.*, 2002), which compare the machine translation output to one or several human reference translations. Automatic metrics are a quick way to judge the performance of a system and they usually perform well for small changes in one system, or for comparisons of similar systems, while not always being fair for systems of different kinds. An alternative is human evaluation, which is often done to compare systems, either by ranking sentences, or by assigning

¹ <http://translate.google.com/>

numerical values to dimensions such as adequacy and fluency (see e.g. Callison-Burch *et al.*, 2007). Adequacy is used to describe the level of semantic equivalence between the source and target texts, and fluency is used to describe how fluent the source text is on its own, i.e. if it is a good source language text. Human evaluation is a good complement to automatic metrics, but it is very time consuming. A problem with both these evaluation types is that they do not give much information of what the particular problems in a system are. To address this issue, error analysis is needed.

Error analysis has been attempted before for MT. Vilar *et al.* (2006) described a general error typology that has also been used by several other authors. Their typology has five top categories of errors: *Missing words*, *Word order*, *Incorrect words*, *Unknown words*, and *Punctuation*, most of which are further subdivided. I find their suggestion problematic in several respects. First of all the typology mixes identification of errors with finding the cause for certain errors, e.g. by distinguishing unknown words by words that have a lemma that is known to the system, and words that are completely unknown. While such an analysis certainly can be useful, I think it is outside the scope of a general error typology. On the other hand the typology is rather shallow, and has several very large categories. The authors suggest that it can be extended for specific languages, and do so for Chinese, but I would prefer a more thorough base typology, even though I agree with the need for adding language specific elements, since target languages differ, for instance in the type of agreement they have. Moreover, there is an error type for *Wrong form* of words, but no further division as to the function of the word with wrong form, such as agreement errors or co-reference errors, both of which result in a word with the wrong form. Other authors have used their own error typologies in single studies, but to the best of our knowledge, there are no other generally used error typologies for machine translation.

Error analysis of machine translation can be useful in many respects. For machine translation system developers or researchers it is useful to know what the most common and/or serious errors are, in order to focus development and research on issues that are likely to really improve translation performance. For users of MT systems it can also be useful to know what the common error types are, in order to be able to easier understand the MT output. For buyers of MT systems it can be advantageous to know the weaknesses and strengths of different systems. The possibility to automatically identify, and possibly correct MT errors, is also useful both for system developers, who can use it to improve their systems, and for users, who could possibly use reading tools which highlight errors.

In this study I define an error typology for machine translation based on a previous error typology for a Swedish grammar checker, Granska (Knutsson, 2001). An error typology for a grammar checker, however, is purely monolingual, concerning the fluency of a monolingual text. For machine translation there are also adequacy errors, where the monolingual target text does not reflect the source text semantics, due to mistranslations. This means that I have to extend the typology to cope with adequacy errors. This study is an extension of Stymne and Ahrenberg (2010), where the grammar checker was used for automatic identification and post-editing of SMT output. In this study I focus on human error analysis using the extended Granska error typology. I also include an enriched discussion of the possibilities of automatic error annotation by Granska, based on a comparison with the human annotations. The Granska error typology is based on human texts and the Granska tool is developed for human texts. SMT

output is different from human texts in many respects, so this is also an investigation of how well typologies and tools that were developed for human texts can be applied to a completely different text type, SMT output, compared to different types of human texts.

2 TOOLS AND CORPUS

A statistical machine translation system is created by collecting statistics from a large corpus of texts translated by humans. It contains a number of statistical models that accounts for different aspects of the translation process. The two most important are the translation model, which accounts for adequacy, and contains probabilities for translations of short text segments, and the language model, which accounts for fluency, and contains probabilities for sequences of source words. I trained a standard phrase-based machine translation system for translation from English to Swedish, with the commonly used Moses decoder (Koehn *et al.*, 2007). The system is further described in Stymne & Ahrenberg (2010)². I automatically check the machine translation output with the Swedish grammar checker Granska (Knutsson, 2001), which is a hybrid, mainly rule-based grammar checker. Granska identifies errors and generates error descriptions and correction suggestions.

For both training and testing of the machine translation system I use texts from the Europarl corpus (Koehn, 2005), which is a collection of European parliament speeches, available in 11 languages. I use the English-Swedish portion of Europarl. The system was trained on 701,157 sentences, and the error analysis was performed on 500 sentences, with a total of 11,574 Swedish words.

3 ERROR TYPOLOGY

In this study I base an MT error typology on the error typology of Knutsson (2001), who investigated errors in Swedish human texts in order to develop and evaluate the Swedish grammar checker Granska (Knutsson, 2001). Knutsson performed a large scale investigation of errors in five types of Swedish texts: sport news, foreign news, popular science, government texts and student essays. Table 1 shows the error categories identified by Knutsson, their overall frequency, and their implementation status in Granska. Table 1 also shows the number of errors in each type in SMT output. As can be seen there are many more errors in the SMT output, despite the fact that the text is only 5.8% of the length of the human texts.

All error categories in Table 1 are errors that are found in monolingual texts produced by humans. Thus they only account for the fluency in a text. In machine translation, the target text should also correspond to the semantics of the source sentence. The MT output should not omit, change or add anything to the content of the source text. To be able to use the grammar checker based typology of Knutsson, I thus need to extend it with categories that account for adequacy. These categories are shown in Table 2. Overall fluency errors are much more common than

² The system was trained on 701,157 sentences, and used a part-of-speech sequence model.

adequacy errors, which is positive, since they generally are less problematic for understanding the target text, and they are also, in principle, possible to correct automatically.

Table 1

Type of error	Count, human texts (of 201,019 words)	Can be identified by Granska	Count, MT output (of 11,574 words)	Found by Granska in MT output
Noun phrase agreement	69	Yes	126	63
Predicative agreement	16	Yes	34	18
Verb form errors	89	Yes	148	41
Word order	8	To some extent	134	9
Split compounds	74	Yes	65	11
Wrong pronoun form	14	To some extent	-	-
Missing function word	56	To some extent	31	-
Coreference	1	No	26	-
Extra words	4	No	12	-
Wrong preposition	11	To some extent	10	-
Definiteness	4	No	10	-
Punctuation	2	No	4	-
Spelling error with gram-sem consequence	55	No	-	-
Contamination of expressions	4	To some extent	-	-
Upper/lower case	11	No	141	-
Foreign words	-	-	69	-
Total	418		810	142

Error types in human texts, their frequency in human texts, their implementation status in Granska (based on Knutsson, 2001), their total counts in SMT output, and the number of errors found by Granska in SMT output. These error types are all fluency errors.

Table 2: Adequacy error types and their counts in SMT output

Type of error	Count, MT output (of 11,574 words)
Mistranslated words	199
Missing words	88
Extra words	36
Wrong form	30
Wrong order	3
Total	356

4 ERROR ANALYSIS USING THE TYPOLOGY

In this section I will discuss the results of a manual annotation of 500 machine translated Swedish sentences, annotated for errors using the error typology described in Section 3. The annotation was performed by the author, who is a native Swedish speaker. The number of errors found of each type are shown in Table 1 and 2. In the following I will discuss and exemplify the error types that occurred in the machine translation output. The error types described in section 4.1-4.11 are fluency errors, i.e. they are problematic on the target side, while basically carrying the source semantics, whereas the last five categories, described in 4.12-4.16 are adequacy errors that are problematic with respect to the source side, while they can possibly be grammatical considering only the target. Three of the error types from Table 1 never occur in the MT output; these and the rare punctuation errors will not be discussed.

4.1 Noun phrase agreement errors

In Swedish there is noun phrase agreement between the article, noun and adjectives, on gender, number, and definiteness. Noun phrase agreement errors are very common in the SMT output. The majority of the erroneous NPs contain one adjectival modifier, errors on NPs without modifiers are very rare. There are also a number of more difficult cases, that are not identified by the grammar checker, including NPs with several adjectival modifiers, adjectives with adverbial modifiers, or NPs with conjunctions, exemplified in (1).

- (1) There needs to be a sensible and sensitive partnership in the external field
Det behövs *en förnuftig och känsligt partnerskap* på det utrikespolitiska området
It needs a-UTR sensible-UTR and sensitive-NEU partnership-NEU on the external field-
DEF
Det behövs ett förnuftigt och känsligt partnerskap på det utrikespolitiska området³

4.2 Predicative agreement errors

There is also agreement in Swedish predicatives between the subject and predicative on gender, number, and definiteness. Just like for NP agreement the simple cases are often found by the grammar checker, but there are many complex cases with conjunction or adverbial modifiers that are problematic, as in (2).

- (2) it must be clear and resolute
det måste vara *tydliga* och resolut
it-SG-NEU must be clear-PL and resolute-SG-NEU
det måste vara tydligt och resolut

³ Most examples are snippets of full sentences. All translation examples contain the English source text, the Swedish machine translation and its gloss, and a (close) correct Swedish translation. Italics are used to highlight relevant parts in the snippet. The following abbreviations are used for grammatical functions, when needed for clarity: SG: singular, PL: plural, DEF: definite, INDEF: indefinite, NEU: neuter gender, UTR: uter gender.

4.3 Verb form errors

This is a relatively diverse error type, with many subcategories, such as missing (finite) verb, form after aux/modal/cop, double supine/passive. It includes two main error types, either it detects that a verb, or infinitive marker, is missing, or that a verb has the wrong form. That a verb is missing does not always constitute a real error, since the test text contains many independent snippets and headlines, as (*applause from the UEN group*), which should not contain a verb. Many of the verb errors in the SMT output cannot be attributed to any of the verb error subcategories, showing that the types of verb errors produced by SMT, not surprisingly, are different from the verb errors in human texts. In many of these cases the target text erroneously keeps the structure of the English source text, as in (3), where the copula of the English progressive form is kept instead of using only the single main verb *ignorerar* in Swedish.

- (3) but that we in the committee are ignoring the fact
men att vi i utskottet *är att ignorera* det faktum
but that we in committee-DEF are to ignore that fact
men att vi i utskottet *ignorerar* det faktumet

4.4 Word order errors

Word order errors belong to the most common in the output, and they are hard to find for the grammar checker, which only has partial coverage of word order errors. The most common word order error is related to the fact that Swedish is a V2 language, and English is not, which leads to many cases where the verb and subject should be reversed in Swedish, but are not, as in (4). Another common word order error is misplacement of adverb, which is sometimes different in the two languages.

- (4) As a member of the Convention, I have to admit that
Som ledamot av konventet, *jag tillstår* att
As member of convention-DEF, I admit that
Som ledamot av konventet *tillstår jag* att

4.5 Split compounds

Splitting compounds into several words is a common writing error in Swedish and many other languages where compounds generally are closed, i.e. written as one word without spaces or other word boundaries between the parts. Writing them as two words changes the stress of the words, and often changes the meaning of the words, which is the basis of many jokes, e.g. *kycklinglever* (*chicken liver*) versus *kyckling lever* (*chicken is alive*). The grammar checker often finds these when the words have the correct form, but it is also common that words have the incorrect form, which makes them harder to identify, as in (5). This is also the category where the grammar checker has the highest number of false alarms.

- (5) the priority of EU climate change strategy
prioritering av eu:s *klimatförändringar strategi*

priority-DEF of eu's climate-changes strategy
prioritering av EU:s klimatförändringsstrategi

4.6 Spelling-related errors

Granska contains a spell checker that marks spelling errors. The SMT output does not have any spelling errors as such, since it can only produce words that are found in the parallel text it is trained on, or words that are present in the input to the system. There are, however, two types of errors that the spell checker can manage to find, problems with true casing and foreign words in the output. Examples of both these types of errors are shown in (6).

- (6) listened very carefully to Commissioner Verheugen' s words about 'selling' the EU to the candidate countries
lyssnat mycket noga till detta påpekande från kommissionär *verheugen* ord om 'selling' eu att kandidatländerna
listened very carefully to this point from commissioner verheugen about 'selling' eu that candidate-countries-DEF
lyssnade därför mycket noga till kommissionär Verheugens ord om 'att sälja' EU till kandidatländerna

4.7 Missing function words

Missing function words are considered fluency errors when they do not obscure the meaning of the sentence. These errors are quite rare, and generally concern either a missing article in noun phrases (7) or a missing preposition. These errors are only classified as fluency errors when the meaning of the sentence is clear even with the function word missing. If the loss of a function word makes the meaning unclear it has been classified as a missing word (see subsection 4.13).

- (7) for it to comply with the very specific decisions adopted in Helsinki
för att det skall överensstämma med mycket konkreta beslut som antogs i helsingfors
for that it shall comply with very concrete decisions which were adopted in helsinki
för att det skall överensstämma med *de* mycket konkreta beslut som antogs i Helsingfors

4.8 Coreference errors

Coreference is when a word, typically a pronoun, refers back to a previous expression. In the coreference errors the form of the coreferent is wrong. These errors are often third person pronouns with the wrong gender, as in (8), or the wrong choice of reflexive pronoun. The anaphora of the coreferent can generally be found in the same sentence, close to the referent, as in (8), but in a few cases it crosses sentence boundaries. These errors mainly occur when Swedish has distinctions that English does not have, such as the neuter and uter gender, in which case the SMT system tends to choose the most common translation of a word, unless the exact context has been seen before.

- (8) so that the Commission does not feel it is engaged on a major task
så att kommissionen inte anser att *det* är engagerad i en stor uppgift
so that commission-DEF-UTR not feel that it-NEU is engaged in a large task
så att kommissionen inte anser att den är engagerad i en stor uppgift

4.9 Extra words

This category includes extra function words that do not change the main meaning of the sentence, such as doubled prepositions (9) or other words, and spurious determiners or infinitive markers.

- (9) where marked differences of opinion were evident in the ministerial discussions
där stora skillnader var påtagligt *i på* ministernivå
there big differences-PL were evident-SG in on minister-level
där stora åsiktsskillnader var påtagliga på ministernivå

4.10 Preposition errors

Preposition errors are those errors where the wrong preposition is used, but where this does not obscure the meaning of the sentence, as in (10). Changing a preposition can also result in changed meaning, for instance if replacing *to* by *from*, in which case it has been classified as a mistranslated word (see subsection 4.12). Generally the erroneous preposition is a common translation of the English preposition, as in (10).

- (10) What has not been achieved in eleven months
Vad som inte har uppnåtts *i* elva månader
What which not has been achieved in eleven months
Vad som inte har uppnåtts under elva månader

4.11 Definiteness errors

In Swedish, definiteness is mostly expressed by a definite suffix on the noun, but are in some cases also marked by a definite article. This is different from English, where only the definite article is used for marking definiteness. This difference between the languages cause some definiteness errors, where the wrong way of expressing definiteness is chosen, as in (11), where an article is used instead of a suffix. In a few cases definiteness errors are also due to a different distribution of definiteness in English and Swedish.

- (11) because of the fragmentation of our decision-making processes
på grund av *den* uppdelning av våra beslutsprocesser
because of the fragmentation-INDEF of our decision-making-processes
på grund av uppdelningen av våra beslutsprocesser

4.12 Mistranslated words

This is the most common of the adequacy errors, and occurs when the wrong translation of a source word is chosen, given the context. Most of these errors are due to the ambiguity of words,

and the translation of a word is often correct in a difference context. But there are also some completely wrong translations. These errors occur for most parts-of-speech, and sometimes also change the part-of-speech, as in (12), where *aim* is translated with a noun instead of a verb.

- (12) If you aim for these targets, there is a curious range of percentages
Om ni *målsättning* för dessa mål, det är ett märkligt *rad* procentages
If you target for these goals, it is a curious row percentages
Om ni satsar på dessa mål, finns det en märklig skala av procentsatser

4.13 Missing words

These errors are also relatively common, and occur when a content word from the source is missing in the translation, as in (13). Like the previous category, this is really problematic for the understanding of the target text.

- (13) Will that *research* be incorporated in the sixth framework programme?
Kommer att införlivas i det sjätte ramprogrammet?
Will that be-incorporated in the sixth framework-programme-DEF?
Kommer *den forskningen* att införlivas i det sjätte ramprogrammet?

4.14 Extra words

In these errors a word that is not present in the source is inserted in the target, changing the meaning of the sentence. An example is shown in (14), which also has a missing word, *perspectives*.

- (14) In tandem with integrating equality perspectives into all policy areas
Tillsammans med integreringen av lika i alla politikområden *budgetplanen*
Together with integration-DEF of equal in all policy-areas budget-plan-DEF
Tillsammans med integreringen av jämställdhetsperspektiv i alla politikområden
budgetplanen

4.15 Wrong form of words, considering the source

These errors differ from the fluency form errors in that they might be grammatically correct, but that they do not reflect the intended meaning of the source, i.e., they can be in indefinite instead of definite, as in (15), which also has several other problems.

- (15) weaknesses in the systems of control
svagheter i *system* för kontroll
weaknesses in system-INDEF for control
svagheter i kontrollsystemen

4.16 Wrong order of words, considering the source

Like the previous category, these differ from the fluency word order errors in that they might be grammatically correct, but the word order is changed with respect to the source, rendering a

different meaning than in the source, as in (16). There are only three such errors in the analysed text.

(16) in which the money of poor and persecuted people ends up in the clutches of the mafia

där pengarna i fattiga människor slutar i och *förföljs* försämringen av maffian

where money-DEF of poor people ends up and persecuted degradation of mafia-DEF

där pengarna för fattiga och förföljda människor hamnar i maffians nypor

5 AUTOMATIC ERROR ANALYSIS

I compared the automatic identification performance of Granska on the five grammar checker categories that occurred in the statistical machine translation output, with two earlier evaluations on different types of text. One evaluation was on the 201,019 words of human adult texts in five text categories (sport news, foreign news, popular science, government texts and student essays) that the Granska error classification is based on (Knutsson, 2001). The other evaluation was performed on 29,812 words of text written by primary school children (Sofkova Hashemi, 2003, who does not report any results on split compounds).

The results are shown in Figure 1 and 2. Figure 1 shows precision, i.e. the proportion of the identified errors that really are errors, and Figure 2 shows recall, i.e. the proportion of the actual errors in the text that were identified. For all categories except verb form errors, the grammar checker has the highest precision. It has much lower recall than on adult texts, but approximately the same recall as on children's texts. The performance for the grammar checker varies with text type also for the adult texts, however, for instance the precision on NP agreement varies between 0.11-0.72 for the five genres. When the grammar checker does find an error in the SMT output, it tends to be correctly identified. The problem is that it only finds a low proportion of all errors, especially considering that it does not find any other error types, except some spelling related errors, than these five categories.

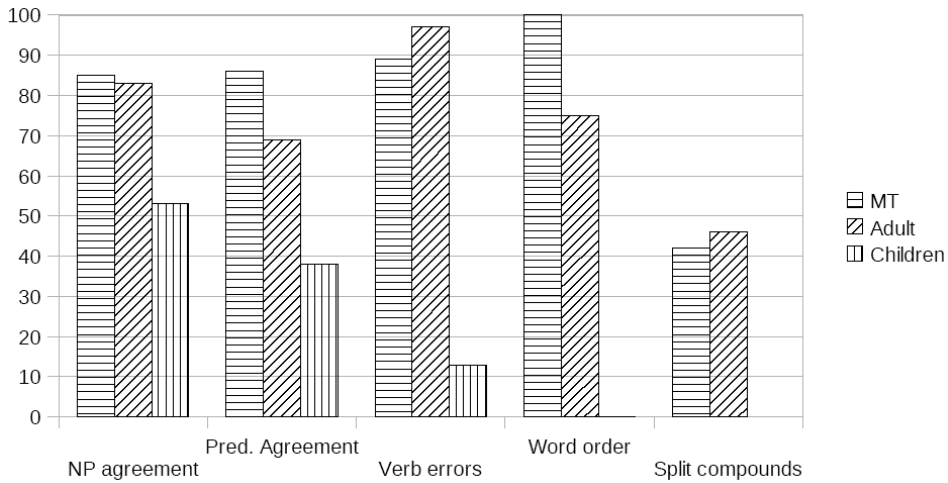


Figure 1: Precision for Granska error identification on three different text types

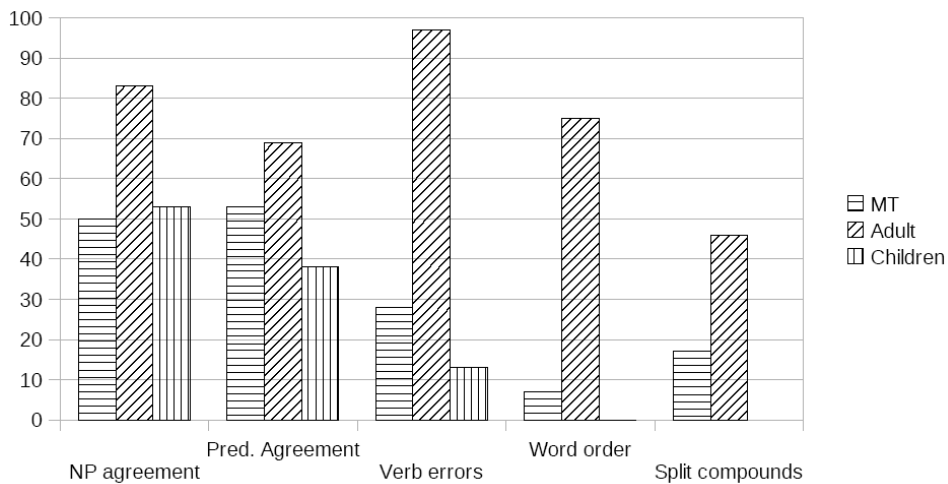


Figure 2: Recall for Granska error identification on three different text types

I also found an evaluation of Granska on texts written by learners of Swedish as a second language (Öhrman, 2000). This study uses different error categories and does not report precision on the individual categories, so it is not included in Figures 1 and 2. The overall precision of error identification in this study is 0.85, and overall recall is 0.35. The recall is very varied between categories, however, with a recall of 0.56 on NP agreement and of only 0.05 on word order. These results are actually quite similar to that of the SMT output. As for the SMT output, the spell checker was able to identify some foreign words and casing errors. Overall, the grammar checker only has a good recall on adult texts. The performance on SMT output is actually similar or sometimes better than for human texts with a high number of errors, as those written by children or learners.

6 CONCLUSION

The error typology of Granska, extended with MT adequacy categories, was useful for SMT error annotation. I could classify all errors into the categories, except that I needed additional subcategories for verb form errors. I also think that the results give quite a good picture of the strengths of our SMT system. As system developers it will certainly help us to focus our efforts on common problems. In some cases a further sub-classification would be desirable, for instance by describing which class of function words is missing for *Missing function word* errors. I plan to develop a general error typology for machine translation, both based on the lessons learned from using this typology, and from comparisons with previous work on error typology.

There is some promise in automatic identification of SMT errors, since the precision of the grammar checker is very high, it tends to only flag real errors. For most practical applications, however, it is problematic that only a small proportion of all errors is found, which is similar to the grammar checker behavior on non-adult texts. The human error analysis showed that the majority of the SMT errors were fluency errors, for which automatic error identification, and possibly also correction, at least in principle is possible. Using Granska in its current form is not enough, but I believe that it would be useful to investigate other postprocessing strategies.

BIBLIOGRAPHY

- Callison-Burch, Chris, Cameron Fordyce, Philipp Koehn, Christof Monz, and Josh Schroeder. 2007. (Meta-) evaluation of machine translation, ACL Workshop on Statistical Machine Translation, pages 136-158. Prague, Czech Republic.
- Knutsson, Ola. 2001. Språkgranskning av svensk text. Licentiate thesis, Royal Institute of Technology (KTH), Stockholm, Sweden.
- Koehn, Philipp. 2005. Europarl: A parallel corpus for statistical machine translation. Proceedings of MT Summit X, pages 79–86. Phuket, Thailand.
- Koehn, Philipp *et al.* 2007. Moses: Open source toolkit for statistical machine translation. of the 45th Annual Meeting of the ACL, demonstration session, pages 177–180. Prague, Czech Republic
- Öhrman, Lena. 2000. Datorstödd språkgranskning och andraspråksinlärare. Master thesis, Stockholms Universitet, Stockholm, Sweden.
- Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: A method for automatic evaluation of machine translation. Proceedings of the 40th Annual Meeting of the ACL, pages 311-318. Philadelphia, PA, USA.
- Sofkova Hashemi, Sofia. 2003. Automatic detection of grammar errors in primary school children's texts. A finite state approach. PhD thesis, Göteborg University, Gothenburg, Sweden.

- Stymne, Sara, and Lars Ahrenberg. 2010. Using a grammar checker for evaluation and postprocessing of statistical machine translation. Proceedings of the Seventh Conference on International Language Resources and Evaluation, pages 2175-2181. Valletta, Malta.
- Vilar, David, Jia Xu, Luis Fernando D'Haro, and Hermann Ney. 2006. Error analysis of statistical machine translation output. Proceedings of the Fifth International Conference on Language Resources and Evaluation, pages 697-702. Genoa, Italy.

6. *Discourse*

Jussi Niemi, Lidia Otsa, Aleksandra Evtyukova and John Niemi

Language and social engagement in Asperger discourse

1 INTRODUCTION

Social aspects of individuals with Asperger Syndrome have been characterized by delays in communication skills, difficulty with social interaction and behavioral symptoms involving repetitive behaviors and a limited range of interest in activities (e.g. Koegel, Schreibman, Good, Cerniglia, Murphy and Koegel, 1989; Crespi and Badcock, 2008; see also *ICD-10*). Focusing now more on the linguistic characteristics of language performance, Asperger output has been called “bookish” (Wing, 1981). As linguists, we would like to draw the reader’s attention to the fact that it is typical that autistic language impairments are impressionistically defined by clinicians, and clinicians in most cases are not language experts.

In assessing the present Asperger (AS) and typically developing (TD) data in the 100 consecutive turns per subject we used a checklist of circa 100 features of language, grammar and communication, ranging from lexical aspects to cognitive and affective features. In order to make the study methodologically as stringent as possible, we aimed at quantification of the data in as many features as possible. Thus, we tried to maximize our efforts in rendering the results to subsequent testing for validity and reliability. Below we list those features that turned out be relevant for the characterization of AS versus TD speech in the current context:

- (i) lexicon, e.g. distribution of lexical items across syntactic categories, content versus function word distribution, number of lexemes and word tokens,
- (ii) structural characteristics of turns, e.g. tri-partite sequential division into onset, core and coda, overlapping speech, narrative predictions,
- (iii) co-operation features, e.g. shared/non-shared elaboration of themes, attitude towards interlocutor elaboration, reaction to novel topic introduced by interlocutor, focus maintenance, politeness,
- (iv) prosody, e.g. non-conventional prosody, affective intonation,
- (v) cognitive aspects, e.g. involvement/commitment, world of discourse,

- (vi) affect features, e.g. expression of empathy, humor.¹

In a framework extending beyond linguistic considerations, we also discuss pragmatic failures in AS discourse and narratives as shortcomings in the social, interpersonal and intersubjective domains as defined by Peter Hobson and Jessica Hobson (see, e.g. Hobson and Hobson, 2008, for a synopsis, see also Barresi and Moore, 2008). The term social here refers to mere social engagement without any further commitments. For instance, other animals do show social behaviour with their conspecifics (and with humans), and we humans may treat our peers in this manner as well, for instance, when entering an elevator and greeting a fuzzy group of people unknown to us with a ‘Good morning’, or the like. When we, in addition to merely engaging in a common social activity, also share the mental space with our peers and treat them as individuals, we may speak of interpersonal sharing of the social. Finally, intersubjective engagement entails understanding (knowing) the minds of the others. We aimed at teasing out prototypical defects (“clear cases”) in the AS data that would show impairment in: (i) the mere engagement in the discourse with the addressee (“social”), (ii) power-relations and affective expressions with the addressee (“interpersonal”), and (iii) the epistemic mind-reading of the addressee by the AS subjects (“intersubjective”).

2 METHOD

2.1 Participants

The two male AS subjects were recruited from the autism pool of the Clinical Research and Rehabilitation Unit of the Honkalampi Foundation in Joensuu, Finland. Their typically developing (TD) controls were matched for their dialect background, socio-economic status, chronological age and sex with the AS subjects. In what follows we call the AS subjects AS1 (8 years of age) and AS2 (10), and their typically-developing (TD) controls TD1 and TD2.

2.2 Procedure

Methodologically speaking, and to the best of our knowledge, the present analysis of Asperger discourse and narrative is the first of its kind to adopt the following procedure and perspective: As for the procedure, in the selection of the spontaneous speech samples of the subjects, both Asperger and their controls, we randomly chose a sequence of 100 turns per subject and analyzed these samples as extensively and intensively as needed for the linguistic profiles of these speakers. By performing this exhaustive and in-depth analysis of randomly chosen long stretches of discourse we tried to avoid the typical procedure found in Conversation Analysis (CA), namely the selective sampling of allegedly interesting stretches of language output. It is granted that isolated examples may carry some value, if only in the heuristic sense of data-exhibition. Moreover, the “interesting exemplar” sampling procedure makes the use of any quantitative analysis conceptually impossible. A reflection in the field of autism studies of the

¹ A more comprehensive account of the present study (with e.g., cartoon description tasks, lexical analyses) is to be found in Niemi et al. 2010.

(unfortunate) canonicity of the “interesting exemplar” CA approach is seen, for instance, in the studies reported in the special issue of autistic discourse published in the journal *Discourse Studies* in 2004 (see, e.g. guest editors’ Introduction, Ochs and Solomon, 2004).

Finally, since the classification of speech output in typical speakers, not to speak of that in atypical speakers, is frequently a matter of great deliberation and open to analyst bias, we painstakingly went through each and every discourse turn and narrative event of the Asperger and control speakers as a collective. Only after having reached a consensus on a data point, whether it was lexical, syntactic, pragmatic, prosodic, or whatever, was the classification of that data point recorded in the files.

2.3 Data

The data reported are derived from video taped communicative environments where semi-structured dyadic interviews were conducted to elicit spontaneous discourse data on various topics, including school, free time and hobbies. For each subject, 100 consecutive therapist-child conversation turns were chosen for a deeper pragmatic/discourse analysis. (For other tasks, see Niemi et al., 2010.)

3 RESULTS

3.1 Lexically mediated social expressions

The two AS subjects show a considerably lower use of what are defined as social expressions than do the controls. Almost any linguistic expression is “social”, as language is, *per definitionem*, a social institution, but here terms like “lexically mediated social expression” refer to words that indicate activity socially shared with the people involved. Our analyses show that each AS subject produces 11 social expressions while controls produce about three times as many instances in the totality of 100 turns per speaker (TD1 34, TD2 32). Moreover, the internal distribution of these social expressions is skewed in a theoretically interesting manner, since the AS *never* use the pronoun forms for the socially inclusive ‘we’ and its variants, while their controls use them relatively frequently (TD1 12, TD2 8 instances). The AS subjects do use person pronouns, when, for instance, referring to themselves or to the co-teller (singular ‘you’), or to a third person. The non-use by AS of the socially inclusive first person plural forms does not, however, mean that the AS speakers do not refer to a social group in which they are members. For instance they can refer to their families by naming with nouns the individuals involved by saying ‘father’, and ‘mother’. It thus appears that the use of the contextually determined and socially contingent ‘we’ is avoided by the AS speakers.

3.2 Perceptual and mental use of locative proadverbs

In Finnish, like most probably in all languages, the so-called spatial proadverbs, like *tässä* ‘here-in’, can be used to refer to the immediate, perceptually shared space as well as to the mentally mediated space. We can construct a plausible hypothesis about the of the deictic proadverbs in the perceptual versus mental function in Asperger and TD speakers: on the basis of the Theory of Mind explanations of autism spectrum disorders (e.g. Baron-Cohen, Leslie and Frith, 1985,

Crespi and Badcock, 2008) one would predict that mentally mediated reference to the spatial coordinates would be of lesser use in the spontaneous language output of Asperger speakers than in that of typically developing individuals. Our results show that the spontaneous speech of all speakers carries spatial proadverbs both in the perceptual and mental function, but the relative use of these expressions differ across the AS – TD boundary in the manner predicted, as the TD subjects use the mental function proadverbs more than the AS speakers, both in absolute and relative terms (Figure 1).

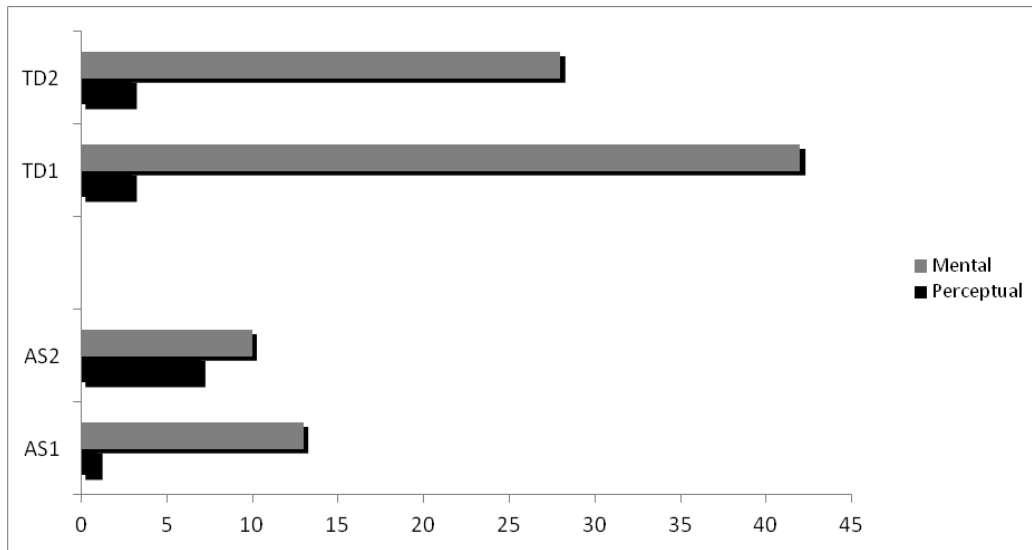


Figure 1. Perceptual and mental proadverbs (e.g., ‘here’, ‘there’) in AS and TD speech (N).

3.3 Turn Structure

A prototypical (default) turn in discourse is composed of three sequences: opening (onset), nucleus (core) and ending (coda), with their own structural (e.g. lexical, syntactic) and functional characteristics (e.g. Sacks, Schegloff and Jefferson, 1974). The analyses of the 400 turns of the four speakers show that the TD subjects are always able to verbalize the propositional content of their message, as their turns never lack the nucleus (Figure 2), while both AS subjects fail to produce the nucleus in six instances. Moreover, turn openings and endings are missing to a considerable extent in the AS turns, when compared to those produced by their controls (chi² tests: (a) openings: AS1 vs. TD1, p = 0.001, AS2 vs. TD2, p = 0.027, (b) endings: AS1 vs. TD1, p = 0.000, AS2 vs. TD2, p = 0.013).

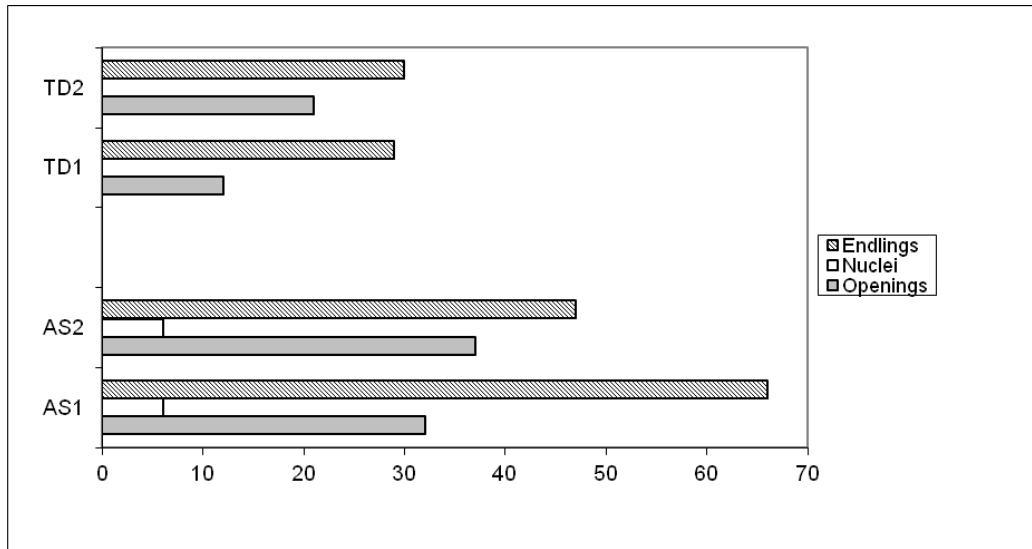


Figure 2. Turn construction failures, i.e., the number (N) of missing turn openings, nuclei and endings in AS and TD (N total per subject = 100).

Since prototypical dyadic discourse is in fact “co-telling”, the adjacency pairs are characterized by overlapping speech and predictions by both interlocutor(s). Not unexpectedly, our typically developing subjects show a marked amount of instances of both overlaps (TD1 45, TD2 40) and predictions (TD1 34, TD2 31) in their 100 turns per subject. These figures stand in a marked contrast to those of the Asperger subjects, as AS1 has 12 and AS2 19 instances of overlaps, and the former has eight instances of discourse prediction and the latter as few as three (chi² tests): (a) overlaps: AS1 vs. TD1, $p = 0.000$, AS2 vs. TD2, $p = 0.001$, (b) predictions: AS1 vs. TD1, $p = 0.000$, AS2 vs. TD2, $p = 0.000$). The significantly higher incidence of overlaps and predictions by the TD subjects, when compared to the AS subjects, reflects the general observations about these four speakers, viz., that the TD speakers are more actively constructing and sharing the dialogue with the interlocutor than the AS subjects.

What the quantitative data is, however, unable to reveal is that there is a qualitative difference between the controls and the AS subjects in interruptions: the immature interruptions and shifts of themes by the controls are not impolite, while those by AS tend to be that (e.g. (1), the Finnish original, e.g. with its word-order permutations, transformed into fluent English; the topic: cat of the AS family). The second turn of AS2 is totally out of place in relation to the topic of the discussion and the power-dynamics (politeness) of the situation.

- (1) AS2: It wasn't like any fine fur, but it was just a bit.
 Adult: Err, was it?
 AS2: At this moment many cats are thirty-seven years old. It is, it still lives, thirty-seven years, six months.

3.4 Content-related Features of Co-operation, World of Discourse and Affect

Both TD subjects were almost always able to elaborate on the themes of the discussion together with the interlocutor (TD2 always, TD1 failed on a few occasions), while the involvement in the expected sharing of themes by both AS speakers was usually unsuccessful. AS1's strategy was to list items, and AS2 elaborated on his favourite topics only, and even there his attempts were often infelicitous. A similar dividing-line is seen in the speakers' attitude towards the elaboration of the interlocutor as well as in focus maintenance. In the former sphere the TD subjects were always successful, while AS1 tended to criticise the interlocutor's elaborations and AS2 often simply rejected them. In focus maintenance, the TD performance was appropriate without exception, while both AS subjects tended to shift the focus inappropriately. In terms of politeness, AS1 was often, perhaps with some irony, correcting the adult's propositions, and AS2 often totally ignored his interlocutor's questions. In the child-adult context TD1 was sometimes hyper-polite and serious and TD2 relaxed but polite.

When we try to characterize the prototypical world of discourse of the subjects, our consensus is that AS1 is a speaker focused on exact knowledge and description of personal events, AS2 is interested in proper names, while TD1 centers on factual knowledge and his own experiences and TD2 is a narrator of his own experiences. Neither of the AS subjects shows humorous traits or signs of understanding humor. In contrast, TD1 uses humorous language, which is directed at third parties, and TD2 has himself, the interlocutor as well as third parties as targets of humorous expressions. In terms of the possible foci of affective language, TD1 uses few mildly affective expressions about school activities and TD2 directs his affective statements even to people. The two AS speakers have no focus of (positive) affective language. AS1 does use some affective expressions but they can be regarded as memorized strings of words, and AS2 shows some affect in his intonation, but only when he "knows better" or refuses to participate in the joint activities with the interlocutor.

4 DISCUSSION AND CONCLUSIONS

The AS subjects fail in several aspects of (oral) text output construction. For instance, the internal structure of the AS turns is definitely different from that of a typical speaker, since they very often lack turn openings and endings. In addition, even the nucleus of the turn may be missing in AS output (but never in the TD turns). The lowered dynamics of AS discourse is also reflected in the lowered use of overlaps and predictions, two features typical of a dyadic discourse that is socially shared and mutually constructed on-line by the participants. Associated with this we find atypical attitude to the interlocutor topics and lack of politeness in AS speech, features that are lacking in the TD data. The TD speakers also maintain the topic appropriately, while both AS show instances of inappropriate topic shifting. A somewhat more local indication of the lowered degree of social language is seen in the amount of what we call social expressions. The AS subjects show a considerably lower use of these lexical items compared to the TD subjects. Among the social expressions, it is most striking to find that the AS speakers never use the pronoun forms for the socially inclusive and contextually contingent 'we', while their controls use them in their narratives. The potential ambiguity of 'we' has been analyzed more in depth

by Wecshler (2010), who states that the two pronouns 'we' and 'you' /plural) have odd semantics among nominal's as they can refer to the following (combinations of) referents: speaker(s) [+addressee(s) + other(s)].

The absolutely and relatively low use of spatial proadverbs in the displaced, mental function in the AS speech output is taken to be an indication of non-mentalizing (or rather, lowly mentalizing) minds of individuals of autism spectrum disorder (e.g. Baron-Cohen, 1996; Crespi and Badcock, 2008). At a high level of abstraction, the current pattern of low production of mentally interpretable linguistic expressions in Autism Spectrum Disorders (ASD), including Asperger, may be seen as a reflection of the inability of the ASD individuals in second and third level theory of mind tasks, in which they are required to interpret other minds. The analysis of closed-class lexical items like proadverbs is methodologically superior to analyses of open-class physical versus mental terms, since open-class words may (or one should say, tend to) have other properties as well (e.g. differences in frequency of use, denotative semantics, morphological structure, abstractness, picturability), and these properties may become intervening variables in the physical vs. mental expression comparison in ASD or in any other language use. As a closed-class item (like the English 'here', 'there') a spatial proadverb has as such only and only its spatial semantics, irrespective of its use in the perceptual or mental function. In essence, through the functional differentiation inherent to these spatial proadverbs we are able to test the possible effects of the functional difference in the same lexical items. As for the perceptual function of the Finnish spatial proadverbs, we have previously shown that AS1 and AS2 (together with four other Asperger children of the same age range) are able to place items in the visual space in the manner identical to their TD peers (Niemi, Ravattinen and Niemi, 2008). In Niemi et al. (2008) each subject was to position, with a click of the computer mouse, a hedgehog on the screen according to the auditory instructions by the speaker to the addressee pictured on the screen standing face to face to each other at a distance of circa seven meters. The experiment comprised four types of sentences spoken by the female speaker, and the items began with 'Look, John ...' (the two figures were familiar to the AS subjects since they collected the data). Of the four stimulus types of the experiment, Type 3 is relevant to the present discussion. In this type the continuation to 'Look, John...' was 'from-me/from-you to the right/left is a hedgehog' and it thus tested the possible effects of the intrinsic (addressee) vs. relative (speaker) frame of reference. The last type was expected to reveal differences between the AS and TD groups, if any, as AS individuals, with their allegedly low capacity of reading other minds, were expected to be more unable than the TD subjects to shift from the speaker's space to that of the addressee. The results did not support this hypothesis, and the main result of this experiment is that in spite of their atypical use of the mental space proadverbs the two Asperger groups (and among them AS1 and AS2) show typical patterns in object placement using the same proadverbs in the perceptual space. The claim that AS individuals are able to perform like their TD controls in tasks of perceptual space finds corroborative evidence from these same Finnish subjects and their controls performing a maze task (van der Zee, Daley, Niemi, Lawson and Nikanne, 2007, Lawson, van der Zee and Daley, 2009, see also Caron, Mottron, Rainville and Chouinard, 2004).

Positioned within the meta-perspective of the social, interpersonal, and intersubjective domains presented in the Introduction, we would like to claim that the dyadic discourse and

narratives of the present AS speakers do show features of engagement in (and attention to) the social context. For instance, the AS subjects do not totally discontinue the discourse or in any other manner shun away from the social interaction. Moreover, the same AS subjects have been previously shown to follow orders in a typical manner in a maze task (van der Zee et al., 2007, Lawson, et al., 2009), and they have been shown to place objects in space in a normal fashion, using correctly both the speaker and the addressee as the landmarks (Niemi, et al., 2008). However, some of their linguistic features in the present discourse and narrative tasks can well be interpreted as manifestations of the AS speakers not sharing the interpersonal discourse space with their interlocutors. These features include the unexpected and abrupt topic-shifts, immature interruptions and the lack of (positive) affective language. Finally, shortcomings like the lower use of locative proadverbs in the mental function, the lower than typical frequencies of turn overlaps and predictions, the higher degree of missing turn onsets and nuclei as well as the non-use of 'we' are all here taken to indicate grave failures in AS in the most demanding aspect of social engagement, viz., in that of intersubjective understanding (knowing) the mental states of others.

To conclude, we hope to have shown that grammatically-based, quantitative case studies of long uninterrupted stretches of dyadic conversation and narratives are able to bring forth aspects of language processing and/or competence in Autism Spectrum Disorder that would (and should) also benefit those who construct ASD intervention and therapy programs, not to speak of those who are interested in the ways in which our cognitive and pragmatic domains interface with our grammars.

BIBLIOGRAPHY

- Baron-Cohen, S. 1996. *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: MIT Press.
- Baron-Cohen, S., A. Leslie and U. Frith. 1985. Does the autistic child have a theory of mind? *Cognition*, 21.37-46.
- Barresi, J. and C. Moore. 2008. The neuroscience of social understanding. In J. Zlatev, T. Racine, C. Sinha and E. Itkonen (eds.), *The shared mind: Perspectives on intersubjectivity* (pp. 39-66). Amsterdam: Benjamins.
- Caron, M., L. Mottron, C. Rainville, C. and S. Chouinard. 2004. Do high functioning persons with autism present superior spatial abilities? *Neuropsychologia*, 42.467-481.
- Crespi, B., and C. Badcock. 2008. Psychosis and autism as diametrical disorders of the social brain. *Behavioral and Brain Sciences* 31.241-261.
- Goodglass, H., and E. Kaplan. 1983. *The assessment of aphasia and related disorders*. Baltimore, Maryland: Waverly Press.
- Hobson, P., and J. Hobson. 2008. Engaging, sharing, knowing: Some lessons from research in autism. In J. Zlatev, T. Racine, C. Sinha and E. Itkonen (eds.), *The shared mind: Perspectives on intersubjectivity* (pp. 67-88). Amsterdam: Benjamins.

- ICD-10 (International Statistical Classification of Diseases and Related Health Problems, 10th Revision) 2007. World Health Organization. Available online: <http://apps.who.int/classifications/apps/icd/icd10online/> (accessed June 15, 2009).
- Koegel, R., L. Schreibman, A. Good, L. Cerniglia, C. Murphy and L. Koegel. 1989. How to teach pivotal behaviors to children with autism: A training manual. Santa Barbara: University of California, Santa Barbara.
- Lawson, S., E. van der Zee and L. Daley. 2009. Spatial language in computer mediated communication. In P. Turner and S. Turner (eds.), *Exploration of space, technology, and spatiality* (pp. 165-178). Hershey, PA: Information Science Reference.
- Niemi, J., L. Otsa, A. Evtjukova, L. Lehtoaro and J. Niemi. 2010. Linguistic reflections of social engagement in Asperger discourse and narratives: A quantitative analysis of two cases. *Clinical Linguistics and Phonetics* 24.928-940.
- Niemi, J., J. Ravattinen, J., and J. Niemi. 2008. Object placement in Finnish: an experiment with locative proadverbs in Asperger and typical populations. Paper presented at the Workshop on Spatial Categorization and Language Across Populations, NTNU/Stiklestad, Norway, 2008.
- Ochs, E. and O. Solomon. 2004. Introduction: Discourse and autism. *Discourse Studies* 6.139–146.
- Sacks, H., E. Schegloff and G. Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50.696-735.
- van der Zee, E., L. Daley, J. Niemi, S. Lawson and U. Nikanne. 2007. Finding our way around maps: perspective use in giving and receiving instructions. Paper presented at The First Conference of the Swedish Association of Language and Cognition, Lund, 2007.
- Wechsler, S. 2010. What ‘you’ and ‘I’ mean to each other: Person indexicals, self-ascription, and theory of mind. *Language* 86.332-365.

ACKNOWLEDGMENTS

The study has been financed through an Academy of Finland research project grant awarded to the first author (project title: *Grammatical, Pragmatic and Prosodic Performance in Autism Spectrum Disorder (ASD)*, Academy grant number 214151). The co-operation of the Clinical Research and Rehabilitation Unit of the Honkalampi Foundation (Joensuu) in recruiting the Asperger subjects was a sine qua non contribution to the execution of the whole project, including the present study on Asperger discourse. We would also like to thank Laura Lehtoaro for her assistance in analyzing the data.

Michael Szurawitzki

How to analyze the language of social networking sites – an analysis model¹

1 INTRODUCTION

(Mobile) electronic media undoubtedly shape our communicational habits. Computers, mobile phones or devices that combine these technologies are both working machines and an integral part of our everyday lives. Children growing up in the 21st Century understand this technology as self-evident. We can expect a greater shift in communication between human beings towards the use of electronic media since an increasing number of (young) people favor electronic forms of communication.

How can a linguist try to analyze this phenomenon of highly popular communication situated in electronic spheres? There are many possible ways of tackling this question. In German language research on this topic (we assume a Germanic philology perspective here), one might quote relatively recent studies such as Abfalterer (2007), Alvensleben (2007), Bartel (2008), Beißwenger (2000), Beißwenger (ed., 2001), Belik (2007), Bittner (2003), Bremer (2005), Bruns (2003), Dorta (2005), Eck (2007), LaRoche (2008), Runkehl/Schlobinski/Siever (1998), Stegbauer/Rausch (1999) Thiedecke (ed., 2003) or Verstraete (2004). Given the plentitude of such studies, it seems necessary to specify the focus of one's own efforts since rather commonplace statements trying to cover too broad a range of subject areas run the risk of being swiftly superseded.

The present paper consequently focuses on the analysis of the language of social networking sites on the Internet. It is a further development of the ideas developed in Szurawitzki (2010), in which an analysis model for the language used in *Facebook* and on other portals was developed and first published. Certain aspects will, however, be discussed in greater detail here. The paper is structured as follows: after a brief introduction (1.) the focus will lie on previous research on social networking sites (2.), after which the analysis model will be presented (3.), including a short analysis of a series of posts extracted from my own Facebook account. Finally, perspectives for further research are developed (4.). It must, however, be noted that the model developed is theoretical in nature and not based on large corpus studies. It can thus be understood as to

¹ Thanks to Roy Goldblatt for correcting my English.

initiate a discussion of how the language of social networking sites can be analyzed linguistically.

2 PREVIOUS RESEARCH ON SOCIAL NETWORKING SITES

In the development of the analysis model proposed in this paper, Naomi Baron's highly interesting monograph *Always On: Language in an Online and Mobile World* (2008) served as a point of departure. Baron writes a learned survey of the development of electronic (mobile) communication and also focuses on social networking sites. Disappointingly for linguists, however, she does not analyze the *specifics of the language* used in Facebook and other social networking sites such as *studivZ*, to name the most influential German language portal. Baron rather focuses on a sociological/historical perspective and thus she cannot be said to provide the cornerstone of the linguistic model developed here. However, Baron does provide us with interesting observations bearing on linguistically-orientated topics. She focuses on two aspects that at first glance seem relevant: the characteristics of so-called "away messages" in instant messaging programs (Baron 2008: 71-80) and the development of Facebook (Baron 2008: 80-98). Baron's approach is here, as in the whole book, more socio-cultural than linguistic in nature, but she does indicate that in 2008 social networking sites were already considered a topic worth researching, and we in international linguistics can concur with her in this respect. Interestingly enough, we currently lack monographs and/or larger corpus studies with a linguistic focus that analyze the language of social networking sites in any detail. This is also the case with Myers (2009), who sets out to analyze the language of blogs and wikis, a related subject, but stops short of an empirical study and thus is of relatively little use to our study here.

Due to the relatively slow nature of academic publishing, there might already be relevant but unpublished research results which have not been made available to the academic community. We can assume that for the time being there are many studies in progress, especially doctoral dissertations in the context of the German language area. Dorta (2005) was a comparable case, evolving from a Ph.D. dissertation on web chats at the University of Freiburg. Since we do not have empirical material in the discussed form to rely on, we proceed towards developing a model on theoretical grounds.

3 THE ANALYSIS MODEL

How could we analyze the language of social networking sites by linguistic means? There is a need for a methodological, systematic approach. The first questions we face in this respect are how to compile an adequate corpus for analysis and which research methodology to choose. Should the utterances (we call them "posts" in the context of this paper; cf. Baron 2008:83) of only one person be analyzed, or should we include more persons? Since social networking sites such as Facebook emphasize interaction, it would be more relevant to examine a series of posts in which people interact; thus, at least two persons/communicants are needed, even though we must also take into account that there could potentially be just "one-way traffic" in terms of

communication, i.e. one person posting utterances without receiving any reaction. If we assume that the more common case would be to expect more than one person engaging in communication, in what relation do the persons stand to each other? Are they connected as so-called “friends” via the analyzed social network and thus communicate directly? Possibly. They, however, do not have to be, depending on the privacy settings and whether they only allow ‘friends’ to comment on their own utterances. On the other hand, might these communicants have intended that their communication should not be seen by other social network users? This might sometimes be the case, even though the potentially worldwide visibility of posting on social networking sites should by now be common knowledge. Or do they use a communications channel which potentially enables all users who could establish contact with the communicants to join in the conversation? This last assumption can probably be assessed as being highly probable. We should rather speak of “discourse” than “conversation” in this respect. In terms of linguistic methodology, discourse analysis could prove a valuable tool for analyzing the structure of social network communication on the Internet (e.g. Brown/Yule (1983) or Schiffrin/Tannen/Hamilton (eds., 2001), to name but a few influential titles). We could first analyze the character of the discourse: do we look at a single post, or do we have a dialogue or, respectively, a polylogue? The analysis model developed here, focusing primarily on other aspects than discourse structure, can also be applied to single posts. With a series of posts of a dialogical or polylogical nature it might prove useful to apply methods of discourse analysis such as those found in the aforementioned relevant works, or, in the Germanic context, studies such as e. g. Brinker/Sager (2006) or Henne/Rehbock (2001). We could thus try to summarize the corpus compilation and possible discourse analysis application in the following table:

Table 1: Corpus compilation and possible discourse analysis

No. of communicants	No. of posts	Analysis/post	Discourse analysis
=1	=1 or >1	yes	no
>1	≥2	yes	yes

It does not seem useful at this point to try to make general assessments on how to compile a corpus in terms of its size. It is worth remembering that the size of the corpus correlates with the results and their validity: the larger the corpus, the greater the likelihood that the results will be generally accepted.

After a corpus has been compiled, an analysis could be conducted according to the model proposed in the following table:

Table 2: Analysis model (based on Szurawitzki (2010))

ANALYSIS MODEL FOR THE LANGUAGE OF ONLINE SOCIAL NETWORKING SITES

1. Determining the amount of text (number of posts, posts’ length, average length; words/characters)
2. Determining which language(s) is/are used; possibly focus on one language only (English?)
3. Grapheme analysis: are any capital letters used (especially according to the conventions of languages such as German)?

4. Analysis of semiotic elements such as a) emoticons, b) abbreviations/'internet slang' and c) embedded links; analyzing the functional pragmatics (emphasizing what?)
5. Syntax analysis (with special regard to the subject)
6. Lexical analysis with special emphasis on word formation processes and loan words (the latter especially in languages other than English; anglicisms)
7. Analyzing the orthography
8. Analyzing the punctuation (are full stops at the end of the sentence left out etc.?)
9. Analyzing the relation of oral vs. written elements
10. With discourses of a length of more than one post: discourse analysis

The steps proposed for the analysis of the language of Internet social networking sites can be understood as a top down model. However, there might be a need for modification due to the nature of specific corpora. Not all steps can necessarily be followed through; e.g. there is no need to analyze step 4.a) when there are no emoticons at all in the corpus. In the following, the steps within the analysis are discussed in greater detail.

1 Determining the amount of text

Here determining the amount of characters is as relevant as determining the number of words. This will serve to relate the results to other forms of electronic communication such as text messages on mobile phones (SMS) or so-called 'tweets', the utterances/posts transmitted via the online service *Twitter*.

2 Determining which language(s) is/are used

If we analyze a multi-language discourse, we could think a) of analyzing the interlingual ways of communicating within this special discourse or b) restrict the analysis only to posts in one particular language. The interlingual dimension can be potentially challenging, since one communicant could write his posts in different languages (one English, one German etc.) or even mix languages within the same post.

3 Grapheme analysis

This is an aspect derived from the background of Germanic linguistics, in which this model was first developed. German is a language with clear conventions e.g. capitalizing all nouns. This aspect is more interesting for German than for most other languages. However, for English we could, for example, analyze how Christian or place names are written and if the writing follows the conventions.

4 Analysis of semiotic elements

Are there emoticons, abbreviations and embedded links, and if so, in what pragmatic context (=emphasis?)? The meaning of the emoticons should always be explained clearly to avoid problems in understanding the argumentation. This can be even more challenging within the broad context of abbreviations, which might vary from language to language, with internationalized exceptions such as the abbreviation *LOL* [*laughing out loud*]. The links must be analyzed not only according to the actual link name, but also in regard to the content (pragmatic function).

5 Syntax analysis

What kind of syntax is used, and what is the subject? *Facebook*, for instance, practically places the name of the post's author in the subject position right at the beginning of the sentence and thus 'forces' the communicant to write about him-/herself using the Third Person Singular, as the following figure shows:



Figure 1: Facebook single post [June 14, 2010]

6 Lexical analysis

Can we determine recurring specific lexemes, possibly new products of word formation which have to be separately explained since they are not (yet) part of the lexicon of the language analyzed and are semantically only known to those participating in the discourse analyzed? English can be assumed to be the common denominator for all other languages since it is the *lingua franca* of electronic communication and of our age and generally sets the tone.

7 Analyzing the orthography

The analysis of the orthography serves to determine whether orthographic conventions are generally followed, or whether new forms of spelling or any other kinds of observable modifications of orthographic norms are deliberately used.

8 Analyzing the punctuation

Are full stops, question and exclamation marks, or semicolons used at all, and if so, are they used according to the conventions of the respective language analyzed?

9 Analyzing the relation of oral vs. written elements

Social networking sites are technically written communication, but – as a hypothesis – we can assume that we will find many elements of oral communication laid down within the language of the posts. This can be assessed by employing the results of the prior steps of the analysis. In a way, this part of the analysis model can be grasped as summarizing the results.

10 Discourse analysis

This step applies to discourses with a length of more than one post. Analyzing the discourse structure could be an initial step, with a possible focus on negotiating conflicts and/or relationships. One interesting aspect could be analyzing underlying discourse hierarchies. This

inevitably leads to theoretical challenges within traditional discourse analysis which would, on this level, most likely have to be adapted to the relatively new communication setting of 'social networking sites on the Internet'.

Example Analysis of a Set of Utterances Extracted From My Facebook Account

In the following, I will apply the analysis model developed above to a short set of utterances extracted from my own Facebook account [March 17, 2011, 7:54 PM GMT +1 – March 17, 2011, 10:36 PM GMT +1; extracted March 19, 2011]. I have anonymized the discourse between the two communications, one male and one female communicant. I call the male communicant DC and the female communicant TA. First, the discourse is printed in its entirety (English translations are provided in square brackets):

[1] DC: putzi putzi [cleany cleany] 7:54 PM

[2] TA: Wir überprüfen am Sonntag, ob alles auch ordentlich sauber geworden ist ;) [We are going to verify on Sunday whether everything became nice and clean ;)] 9:10 PM

[3] DC: ist klar, dass nur weibsvolk auf "gefällt mir" klickt... [it's obvious only females hit the "I like" button...] 9:58 PM

[4] TA: Ja und wen wolltest du denn mit deinem "putzi, putzi - Status" beeindrucken? [Yes and who was it that you wanted to impress with your "putzi putzi" status?] 10:25 PM

[5] DC: das weibsvolk [the ladies] 10:34 PM

[6] DC: :) 10:34 PM

[7] TA: Na also! [That's how I like it] 10:36 PM

The first step (1) determined in the analysis model developed above is the determination of the amount of text. For our analysis example we can state that all the included utterances consist of an amount of characters also suitable for use in SMS messages (≤ 160 characters) or 'Tweets' on *Twitter* (≤ 140 characters). In utterance [1] we have 11 characters (blanks included!), in [2] 74 characters, [3] consists of 56 characters, [4] of 76, [5] of 13, [6] of 2 and [7] of 8 characters. Since we only have German language posts, step (2) of the analysis here becomes unnecessary. In step (3), the grapheme analysis, we can state that DC uses only small letters and does not capitalize the initial words in his utterances or nouns [*weibsvolk*; 3, 5]. TA, in contrast, sticks to the orthographic conventions throughout her utterances. As for semiotic elements specific to internet language (step (4)), we have two smiley emoticons, ;) [2] and :) [6], representing smiling faces, with [6] further elaborating on [5]. When looking at the syntax (5), we can state that in all his utterances DC uses elliptical structures [1, 3, 5, 6], with both a finite verb and a subject missing in [1], the subject of the main clause missing in [3], the finite verb missing in [5] and verb as well as subject missing in [6]. AT, in contrast, uses complete sentences throughout her utterances; no deviations from standard German grammar can be detected here. In terms of lexicological features (6), the initial utterance *putzi putzi* [1] arouses our attention, since this is an example of an unusual word formation. We can clearly identify the free morpheme which *putz* stems from; this evolves from the communicational context, the verb *putzen* [*to clean*]. Its use with the bound morpheme *-i*, as is the case here, is not common in German. The *-i* can indicate a diminutive form such as in *Mausi*, *Schatzi* or the like, entailing a positive connotation. This can

thus also be assumed here, adding a positive connotation to the rather dull task of cleaning (keeping in mind that DC is male, even though most females also would describe cleaning as ‘dull’). In our analysis, there are no orthographic features (7) that need to be examined any further, since there are no deviations from standard German orthography. There are, however, deviations from standard punctuation (8). Only [4] and [7] use standard punctuation through the usage of a question mark in [4] and an exclamation mark in [7]; in all other utterances the conventionally obligatory full stops are missing. [6] does not need punctuation marks, since here we have an emoticon standing by itself. In providing an assessment of whether to attribute oral or written characteristics (9) to the utterances analyzed here, we can, on the basis of the observations above, state that DC’s utterances largely carry characteristics of orality, whereas TA’s utterances can clearly be characterized as highly resembling standard written German (if we for a moment ignore the omission of full stops as described above). If we look at the discourse structure (10), we can make the following observations: DC opens the conversation [1] with an unusual, stylistically not expectable utterance, *putzi putzi*, an expression of his being or having been engaged in cleaning (at home). This utterance is, however, as the context indicates, clearly understood by TA, who in [2] expresses her positive stance by humoristically, using formal language, commenting on [1] that the outcome of the cleaning will be checked during her next visit and reaffirming the positive statement with a smiley emoticon ;). DC refers in [3] to a feature outside the text, i.e. the “I Like”-button, and states that only *weibsvolk* (archaic, which can be perceived as offensive, for *females*) would indicate their empathy by pressing this button in relation to his utterance in [1]. TA in [4] takes up this statement and tries to evoke further clarification by DC on who should be impressed by the *putzi putzi* status update in [1]. DC’s clarification is a reutterance of the *weibsvolk* [5], which is stressed and possibly weakened in its potentially derogatory semantics through the smiley emoticon in [6]: :). TA concludes the conversation with a reaffirming *Na also!* [7]

When summarizing this rather short piece of analysis, we can conclude that based on the analysis and description model used here, we are able to research and describe the specifics of the language used in the passage extracted from my Facebook account without resorting to other means of analysis/description. This indicates that with the analysis model developed here we potentially have a tool for analyzing larger corpora.

4 PERSPECTIVES FOR FURTHER RESEARCH

It seems practical to test the model proposed and applied here by using it on suitable corpora. This model strives to include the most commonly known analysis aspects typical of the language of Internet social networking sites. This might still mean that there will be a need to modify this model in the future, depending on how social networking sites develop. This model serves to trigger a fruitful discussion within the relevant linguistic community, hoping to achieve two things: a) to possibly gain new insight into how to conduct analyses of the language(s) of social networking sites and b) to possibly provide researchers with a suitable model for use in corpus studies of the language of social networking sites.

BIBLIOGRAPHY

- Abfalterer, Erwin. 2007. Foren, Wikis, Weblogs und Chats im Unterricht. Boizenburg: Hülsbusch.
- Alvensleben, Roman von. 2007. w.w.Wahnsinn: eine Expedition in die virtuelle Welt der Chats. Frankfurt am Main: August von Goethe Literaturverlag.
- Baron, Naomi S. 2008. Always On. Language in an Online and Mobile World. Oxford/New York: Oxford University Press.
- Bartel, Rainer. 2008. Blogs für alle – das Weblog-Kompodium. Baar: SmartBooks.
- Beißwenger, Michael. 2000. Kommunikation in virtuellen Welten: Sprache, Text und Wirklichkeit. Eine Untersuchung zur Konzeptionalität von Kommunikationsvollzügen und zur textuellen Konstruktion von Welt in synchroner Internet-Kommunikation, exemplifiziert am Beispiel eines Webchats. Stuttgart: ibidem. [MA thesis, Universität Heidelberg].
- Beißwenger, Michael (ed.) 2001. Chat-Kommunikation. Sprache, Interaktion, Sozialität & Identität in synchroner computervermittelter Kommunikation. Perspektiven auf ein interdisziplinäres Forschungsfeld. Stuttgart: ibidem.
- Belik, Cornelia. 2007. Cyberstalking – Stalking im Internet, Foren, Newsgroups, Chats, per eMail. Norderstedt: Books on Demand.
- Bittner, Johannes. 2003. Digitalität, Sprache, Kommunikation. Eine Untersuchung zur Medialität von digitalen Kommunikationsformen und Textsorten und deren varietätenlinguistischer Modellierung. Berlin: Erich Schmidt Verlag.
- Bremer, Claudia. 2005. Chats im eLearning: Rollenspiele und andere didaktische Elemente in der netzgestützten Hochschullehre. In Michael Beißwenger, Angelika Storrer (eds.). Chat-Kommunikation in Beruf, Bildung und Medien: Konzepte - Werkzeuge - Anwendungsfelder. Stuttgart: ibidem: 89 - 100.
- Brinker, Klaus, and Sven F. Sager. 2006. Linguistische Gesprächsanalyse. Eine Einführung. 4. Auflage Berlin: Erich Schmidt Verlag.
- Brown, Gilian, and George Yule. 1983. Discourse Analysis. Cambridge: Cambridge University Press.
- Bruns, Axel. 2008. Blogs, Wikipedia, Second Life, and Beyond: from production to produsage. New York: Peter Lang.
- Döring, Nicola. 2003. Sozialpsychologie des Internet. Die Bedeutung des Internet für Kommunikationsprozesse, Identitäten, soziale Beziehungen und Gruppen. 2. Aufl. Göttingen: Verlag für Psychologie Hogrefe.
- Dorta, Gabriel. 2005. Soziale Welten in der Chat-Kommunikation. Untersuchungen zur Identitäts- und Beziehungsdimension in Web-Chats. Bremen: Hempen.
- Eck, Klaus. 2007. Corporate blogs: Unternehmen im Online-Dialog zum Kunden. München: Orell Füssli Verlag.
- Häusler, Sascha. 2007. Soziale Netzwerke im Internet. Entwicklung, Formen und Potenziale zu kommerzieller Nutzung. München: VDM Verlag Dr. Müller.

- Henne, Helmut, and Helmut Rehbock. 2001. Einführung in die Gesprächsanalyse. 4., durchgesehene und bibliographisch ergänzte Auflage. Berlin/New York: de Gruyter.
- LaRoche, Claire R. 2008. Employee blogs: protected speech or grounds for discharge. *Journal of Business & Economics Research* 6.2: 9–14.
- Myers, Greg. 2009. *The Discourse of Blogs and Wikis*. London/New York: continuum.
- Prommer, Elizabeth, Brücks, Arne, Mehnert, Julia, Neumann, Heino, Räder, Andy, Roßland, Franziska (eds.) 2009. »Real Life Extension« in web-basierten sozialen Netzwerken: Studie zur Selbstrepräsentation von Studierenden in studiVZ. Online: <http://opus.kobv.de/hff/volltexte/2009/58/> [4.9.2009].
- Runkehl, Jens, Schlobinski, Peter, Siever, Torsten. 1998. *Sprache und Kommunikation im Internet*. Opladen/Wiesbaden: Westdeutscher Verlag.
- Schiffrin, Deborah/Tannen, Deborah/Hamilton, Heidi E. (eds.) 2001. *Handbook of Discourse Analysis*. Oxford: Blackwell.
- Stegbauer, Christian, and Alexander Rausch. 1999. Ungleichheit in virtuellen Gemeinschaften. *Soziale Welt* 50: 93–110.
- Szurawitzki, Michael. 2010. Wie lässt sich die Sprache in sozialen Internet-Netzwerken untersuchen? Grundlegende Fragen und ein Vorschlag für ein Analysemodell. *Muttersprache* 120: 40-46.
- Thiedeke, Udo (ed.) 2003. *Virtuelle Gruppen. Charakteristika und Problemdimensionen*. 2. Aufl. Opladen: Westdeutscher Verlag.
- Verstraete, Claudia. 2004. *Virtuelle Marken-Communities: Newsgroups und Chats als Instrumente der Markenbindung*. Lohmar: Eul.

Oliver Winkler

Conflict patterns in dialogues between wife and husband in Arthur Schnitzler's Reigen

1 SCOPE OF THE STUDY

This paper discusses the literary construction of verbal conflicts between a husband and wife based on the German-language drama *Reigen*, written by the famous Austrian writer Arthur Schnitzler. This topic is an integral part of an ongoing dissertation project¹ in which the literary construction of verbal conflicts between married couples, based on a historical series of both Swedish and German-language plays beginning from 1900 up to the present time, is investigated. The article is structured as follows: Chapter two contains some theoretical thoughts on the characteristics of the literary dialogue as an object of historical dialogue analysis as well as a short discussion on the most common patterns as identification tools to identify verbal conflicts. Chapter three gives a short introduction to the situational as well as historical and socio-cultural background of the displayed literary dialogue. Thereafter follows a close analysis of the interaction between the husband and wife based on two dialogue excerpts. Finally, in chapter four, the conflict patterns obtained in the analysis are reflected against the background of the previously discussed conflict definitions.

2 THEORETICAL CONSIDERATIONS

The theoretical basis of the investigation is the historical dialogue analysis, which is a relatively young discipline within the field of linguistic pragmatics and which has received more attention, especially in the last twenty years (cf. e.g. Jucker et al. 1999, Kilian 2005, Sörlin 2008). Historical dialogue analysis focuses on the historical development of dialogic acting, its particular rules and conditions and it eventually aims at reconstructing different historical dialogue types. The most prominent sources for historical dialogue analysis are literary dialogues (cf. Kilian 2005: 43). From a historical perspective they provide an almost unlimited amount of dialogue-material. However, it is an incontrovertible fact that literary dialogues cannot be equated to

¹ Konfliktaushandlung zwischen Ehepartnern in deutsch- und schwedischsprachigen Dramen. Eine diachron-kontrastive linguistische Dialoganalyse.

authentic historical dialogues. Literary dialogues are always 'artificial' and moreover subject to aesthetic purposes. Nevertheless, the verbal interaction displayed in the literary dialogue is planned and "created by individuals who have grown up and participate in the society/culture they portray and to which they play." (Spitz 2005: 28) Hence, these individuals (i.e. the author and his contemporary recipients) share certain knowledge about the norms and conventions of verbal acting, which they have acquired through the process of socialization (cf. von Polenz 1998: 45). If the author wants his recipients to understand the dialogue he has to use this communicative knowledge:

"The principles, norms and conventions of use which underlie ordinary conversation are the resource that dramatists use to create dialogue in plays. Hence, the interaction in plays represents an internalised model or schema for the production of conversation – a competence model that speakers have access to." (Spitz 2005: 10)

However, authors are free to decide if they either want to adhere to the conventions of the time or if they prefer to deliberately break them. Nevertheless, also by deliberately breaking verbal rules, authors relate to the communicative competence of their contemporary recipients, namely their ability to identify a certain verbal behavior as rule-breaking. The bottom line, for the purpose of historical dialogue analysis, is that literary dialogues are an important source because they always reflect the communicative norms and conventions of the social-cultural context in which they are embedded. Since the following study focuses on verbal conflict patterns, a more specific description of this particular type of verbal activity is necessary.

Conflicts are commonly investigated from two different perspectives: Conflict researchers either focus on the eventual motives and reasons of the emergence of a conflict, or they look at the conflict as an ongoing event and focus on the process and its development (cf. Gruber 1996: 17). Although these two different perspectives do not exclude each other, conversation-oriented research mainly focuses on the processuality of verbal conflicts. Conflict is seen as a social and dynamic process, which is constituted mainly through the verbal interaction between two or more participants. Thus, following Spitz (2005), verbal conflicts ought to be investigated on the basis of a close analysis of verbal interaction:

"In keeping with this interactional view of language as the means for establishing and maintaining interpersonal relationships and for performing socially organised interactions between individuals, I view interpersonal conflict as an emergent process, which is jointly accomplished by the participants in and through talk-in-interaction. Therefore, rather than analyse why people conflict and with what outcomes, we should examine how people conflict, i.e. the interactional procedures participants employ to accomplish the activity [...]. Consequently, the analysis of the participants' verbal exchanges becomes the principal means of investigating conflict." (Spitz 2005: 44)

Research done on verbal conflicts (cf. the overview of Maynard 1985, Gruber 1996, Spitz 2005 and Sörlin 2008) focuses on the 'dissent organization of conversation' (Gruber 1996: 60) and emphasises the role of opposition within verbal interaction. As Maynard (1985) points out, the investigation of verbal conflicts demands a strictly interactional view, whereat "opposition is the second 'move' in any dispute episode" (Maynard 1985: 3). According to this view, a verbal conflict cannot be initiated by a single turn, but only by the oppositional reaction to an

antecedent turn. Furthermore, studies based on natural occurring conversation point out, that verbal conflicts cannot be identified with regard to single oppositional pairs, but must focus on longer sequential parts in order to obtain verbal conflict patterns. These studies have stressed the crucial role of the third position in a potential conflict sequence (Maynard 1985: 1). Commonly they agree on a three-sequence pattern, according to which a verbal conflict can be identified (cf. Spitz 2005: 89ff.). This pattern consists of at least three sequential turns, of which the second and third ones represent oppositional verbal or nonverbal reactions. Hence, if there are at least two subsequent oppositional turns, a verbal conflict is considered to continue. Figure 1 by Spitz (2005) shows this basic structure of verbal conflicts:

- | | |
|---|--|
| 1 A: antecedent event/
arguable action | i.e. verbal or nonverbal action |
| 2 B: initial opposition | i.e. disagreement with 1 |
| 3 A: counter-opposition | i.e. disagreement with 2 and/or supporting or insisting on 1 |

Figure 1: three-step opening-structure of arguments (Spitz 2005: 98)

Based on these theoretical considerations two dialogue samples from the marriage scene “Der Gatte und die junge Frau” in Arthur Schnitzler’s play *Reigen* are investigated in the following chapter. By focusing on the dyadic exchange of speech actions, verbal conflict patterns depicted in the dialogue are analysed and later on (chapter 4) contrasted with the conflict identification model discussed above.

3 CONFLICT PATTERNS IN SCHNITZLER’S REIGEN

For a better understanding of the following dialogue excerpts a few comments on the situational as well as socio-cultural embedment of the portrayed dialogue are necessary. The conversation between the married couple is situated within the bourgeois environment of fin-de-siècle Vienna. The marital relationship between the man (Karl) and the young woman (Emma) represents the ideal marriage in bourgeois society. On the one hand there is the elder, “experienced” and authoritarian man, while on the other is the premature wife who came into marriage without any possibility of self-determination. Thus, the distinct character of this marital relationship is its hierarchical structure and the inequality between the wife and husband. The wife is dominated by her husband and dependent on him in almost every conceivable aspect of life (Koch 1985: 445f.).

The couple’s interaction takes place in probably the most intimate and tabooed sphere of bourgeois society, the couple’s bedroom. The dialogue starts with the husband entering the bedroom to the astonishment of his wife who did not expect him to come so early. Through his verbal and nonverbal behavior Karl implicitly indicates that he wants to make love. Emma on the other hand tries to avoid these initial advances by engaging her husband in conversation. This conversation gradually turns into a discussion about young bourgeois women who end up as prostitutes (see line 38-41) and it becomes appreciably more conflictive when Emma starts to

question Karl's point of view (turn 39 and 41). The following dialogue analysis will focus on the continuation of the dialogue from line 42. Only the German dialogue is analysed. The English translation² is meant to help those who do not read German.

Dialogue Excerpt 1

- 38 Karl: [...] ihr kennt ja das Elend nicht, das die meisten von diesen armen Geschöpfen der Sünde in die Arme treibt.
[...] how can you know the misery that hounds these poor creatures into the arms of Sin?
- 39 Emma: So verkaufen sich denn alle?
But do they all sell themselves?
- 40 Karl: Das möchte ich nicht sagen. Ich mein ja auch nicht nur das materielle Elend. Aber es gibt auch – ich möchte sagen – ein sittliches Elend; eine mangelhafte Auffassung für das, was erlaubt, und insbesondere für das, was edel ist.
I wouldn't go so far as to say that. And I don't only mean material misery. There is also--I might say--a misery that is moral; an inability to grasp what is permissible, and more specifically, what is noble.
- 41 Emma: Aber warum sind die zu bedauern? – Denen gehts ja ganz gut?
But why are they to be pitied? They get along quite well!
- 42 Karl: Du hast sonderbare Ansichten, mein Kind. Du darfst nicht vergessen, daß solche Wesen von Natur aus bestimmt sind, immer tiefer und tiefer zu fallen. Da gibt es kein Aufhalten.
You have strange ideas, my dear. You mustn't forget that these creatures are destined by nature to sink lower and lower. There are no half-way stops for them.
- 43 Emma: *sich an ihn schmiegend* Offenbar fällt es sich ganz angenehm.
[*Cuddling up to him.*] The sinking seems to be rather pleasurable.
- 44 Karl: *peinlich berührt* Wie kannst du so reden, Emma. Ich denke doch, daß es gerade für euch anständige Frauen nichts Widerwärtigeres geben kann als alle diejenigen, die es nicht sind.
[*Pained.*] How can you say things like that, Emma? I've always thought that nothing could be more repulsive to respectable women like you than those who are not respectable.
- 45 Emma: Freilich, Karl, freilich. Ich habs ja auch nur so gesagt. Geh, erzähl weiter. Es ist so nett, wenn du so redst. Erzähl mir was.
Oh, of course, Karl, of course. I was only talking. Go on, tell me some more. It's so nice when you speak like that. Tell me things.
- 46 Karl: Was denn?
What about?

² The English translation can be found under: <http://www.theatrehistory.com/plays/reigen001.html> (04/2011)

- 47 Emma: Nun – von diesen Geschöpfen.
Well--about these creatures.
- 48 Karl: Was fällt dir denn ein?
Why on earth should I?
- 49 Emma: Schau, ich hab dich schon früher, weißt du, ganz im Anfang hab ich dich immer gebeten, du sollst mir aus deiner Jugend was erzählen.
Look, Karl, don't you remember, I begged you right from the beginning, many times, to tell me things about your youth.
Schnitzler (1972: 349-350)

To the preceding question of Emma in line 41, Karl, in line 42, responds with a critical and patronising tone. In the following turn, line 43, Emma indirectly opposes Karl's criticism with a verbally as well as nonverbally signalled sexual allusion. In doing so, Emma breaks the social convention, according to which only men are allowed to take sexual initiative. It becomes apparent that Emma's speech action is provoking her husband from the following reaction of Karl, in line 44, where he opposes his wife with a rebuke. The seriousness of Karl's turn is demonstrated by the changed form of address; changing from the expression *mein Kind*, line 42, to the actual name, *Emma*. Hereafter, in line 45, Emma does not react with a counter-argument but immediately withdraws the responsibility for her utterance and thus avoids the potential conflict situation. By portraying herself as an innocent woman who just talks she equates with the typical image of women of the time. Hereafter, Emma continues the turn with a pretended neutral request in line 45. Karl, in line 46, reacts by asking what he should talk about. This question enables Emma, in line 47, to reintroduce the highly delicate topic about the prostitutes. By trying to make her husband talk about his premarital sexual experiences, Emma touches social taboos and, thus, aggravates the conflict once again. This becomes obvious in line 48, where Karl opposes Emma again with a harsh rebuke. The structure of the rhetorical question and the modal particle *denn* clearly signals Karl's indignation about Emma's preceding behavior. Emma's following reaction, in line 49, where she explains – *Schau, ich hab dich schon früher, weißt du, ganz im Anfang hab ich dich immer gebeten, du sollst mir aus deiner Jugend was erzählen.* – can be seen as a step back onto safe ground. An important part in her role as a wife is to show interest in her husband's life. Hence, by taking on her role as the bourgeois wife, Emma once more mitigates the upcoming conflict, which was aggravated by herself just one turn ago.

To summarise the analyzed interaction sample, quite a tactical and dynamic verbal behavior by Emma, who is switching between conflict-aggravating and conflict-mitigating turns, can be observed. In doing so, Emma succeeds in controlling the dialogue although she is acting from a socially inferior position. So far it becomes obvious that although there are oppositional turns, the interaction does not develop into a verbal conflict in the form of a longer mutual opposition-sequence. The findings of this dialogue analysis shall now be contrasted by a second dialogue excerpt from the same dialogue.

Dialogue Excerpt 2

- 55 Emma: Aber eine Frage musst du mir beantworten ... sonst ... sonst ... ists

- nichts mit den Flitterwochen.
 But there's one question you've got to answer ... otherwise ... otherwise ... the
 honeymoon is out.
- 56 Karl: Du hast eine Art, zu reden ... denk doch, daß du Mutter bist ... daß
 unser Mäderl da drin liegt ...
 You certainly have a strange way of speaking . . . remember that you're a
 mother . . . that our little girl is sleeping right in there . . .
- 57 Emma: *an ihm sich schmiegend* Aber ich möchte auch einen Buben.
 [*Cuddling again.*] I'd like a little boy too.
- 58 Karl: Emma!
- 59 Emma: Geh, sei nicht so ... freilich bin ich deine Frau ... aber ich möchte auch
 ein bisschen ... deine Geliebte sein.
 Oh, don't act like that ... certainly I'm your wife ... but I'd like to be your
 sweetheart too ... just a weeny bit.
- 60 Karl: Möchtest du? ...
 Would you really?
- 61 Emma: Also – zuerst meine Frage.
 Well–answer my question first.
- 62 Karl: *gefügig* Nun?
 Well?

(Schnitzler 1972: 350)

In line 55 Emma suddenly puts pressure on Karl by threatening him that she will refuse the love act, metaphorically expressed as *Flitterwochen*, in case Karl does not answer her question. In doing so Emma not only reintroduces the highly delicate topic about the prostitutes but also breaks a social norm that a woman is not allowed to refuse sexual advances by her husband. The risk of this turn is textually indicated by the hesitation in Emma's utterance, which is observable from the repetition of the conjunction *sonst* as well as the graphematically indicated pauses. In the subsequent line 56 Karl opposes Emma again with a rebuke in which he refers to Emma's social role as a mother. Hereafter, in line 57, Emma aggravates the conflict, not only by another sexual advance towards Karl, which is indicated in the stage directions, but also verbally. For the understanding of this action it is important to consider the gender discourse of the time in which the only legitimate form of sexuality for the bourgeois wife was considered the sexuality for reproductive purposes (Koch 1985: 440f.).³ Thus, Emma insists on her wish for sexuality by taking on her role of a mother. Subsequently in 59, Karl opposes Emma with a harsh rebuke which is textually indicated in the form of an exclamation. At this critical point Emma does not react with a counter-argument but now mitigates the conflict situation by asking her husband not to be so strict. Subsequently she makes a quasi-compromise in which she portrays herself as

³ Behind this speech action stands the dichotomous perception on women typical of the time in which women are classified into two groups: On the one hand there is the compliant, undemanding wife and mother, on the other hand there is the sinful and depraved woman. This dichotomous view is represented maybe in its most radical form in the book *Geschlecht und Charakter* (1903), written by the Austrian philosopher Otto Weininger.

an 'innermarital' concubine. Karl, in line 60, falls for this explanation and, in his own interest, interprets Emma's action as a signal of willingness for the love act. Subsequently in line 61, instead of confirming Karl's question, which probably would lead to the initiation of the love act, Emma opposes Karl with the explicit request to answer her question. Karl accepts this and suddenly becomes compliant as the stage directions in line 62 describe.

At this point it becomes apparent that Emma is in control of the dialogue. She takes advantage of Karl's sexual desire in order to achieve her own goal, namely to make Karl talk about his sexual experiences. This second dialogue sample shows the same features as the first one. Emma tactically chooses between offensive and defensive actions. She succeeds in taking control over the dialogue and implicitly changes the power positions for a short while.

4 CONCLUSION

The findings of the two analysed dialogue samples raise the question whether the verbal interaction in question can be called a verbal conflict after all. According to a rather structure oriented definition of verbal conflicts, such as the above mentioned tripartite sequential model, the dialogue does not represent the classical features of a verbal conflict. It is not a conflict when it comes to ongoing agreement-disagreement sequences.

However, reconsidering socio-cultural aspects, namely the inequality between wife and husband in this specific context of the bourgeois society, it is not surprising that no open argument emerges. None of the classical features of an argument, such as open contradictions, refusals, denials or face threatening acts such as rebukes and personal insults are applicable for Emma who acts from a socially inferior position. Thus, Emma's verbal strategy of covertly controlling the dialogue through a dynamic change between aggravating and mitigating acting is probably the only means for her to enforce her verbal intentions and to oppose her dominant partner. Seen like this, the interaction in question clearly represents a verbal conflict even though there is no verbal battle on the surface of the dialogue.

The abovementioned tripartite structure, as a tool for identifying verbal conflicts, is based on synchronic investigations on naturally occurring conversations. One of the objectives of the historical dialogue analysis is to critically reflect on such definitions from the basis of historical material. Although the material is fictional it reminds us that verbal conflicts (as a specific form of verbal activity) are bound to historically and socio-culturally different norms and conventions of verbal acting. More investigations based on such historical dialogue material would give us a broader view on the variation and the development of verbal conflicts and their constitutive features and it would eventually lead to a more dynamic understanding and definition of verbal conflict.

BIBLIOGRAPHY

Drama

Schnitzler, Arthur. 1972. *Dramatische Werke*. Frankfurt a. M.: Fischer Verlag.

References

- Gruber, Helmut. 1996. Streitgespräche. Zur Pragmatik einer Diskursform. Opladen: Westdeutscher Verlag GmbH.
- Jucker, Andreas H. et al. (eds.) 1999. Historical Dialogue Analysis. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Kilian, Jörg. 2005. Historische Dialogforschung. Eine Einführung. Tübingen: Max Niemeyer.(=Germanistische Arbeitshefte, 41).
- Koch, Christiane. 1985. Wenn die Hochzeitsglocken läuten Glanz und Elend der Bürgerfrauen im 19. Jahrhundert. Frankfurt a.M.
- Maynard, Douglas W. 1985a. How children start arguments. *Language in Society* 14. 1-30.
- Sörlin, Marie. 2008. Att ställa till en scen. Verbala konflikter i svensk dramadialog 1725-2000. Uppsala.
- Spitz, Alice. 2005. Power plays. The representation of mother-daughter disputes in contemporary plays by women. A study in discourse analysis. Dissertation. <http://scidok.sulb.uni-saarland.de/volltexte/2006/595/pdf/dissFINALE.pdf> (04/2011).
- Weininger, Otto. 1903. Geschlecht und Charakter. Eine prinzipielle Untersuchung. Wien und Leipzig: Wilhelm Braumüller.

PROCEEDINGS OF THE SCANDINAVIAN CONFERENCE OF LINGUISTICS

1. Dahl, Östen, ed. 1974. *Papers from the First Scandinavian Conference of Linguistics, Kungälv, March 30-31, 1974*. Göteborg: Department of Linguistics, University of Göteborg, 1974.
2. Hovdhaugen, Even, ed. *Papers from the Second Scandinavian Conference of Linguistics, Lysebu, April 19-20, 1975*. Oslo: Department of Linguistics, University of Oslo, 1975
3. Karlsson, Fred, ed. *Papers from the Third Scandinavian Conference of Linguistics, Hanasaari, October 1-3, 1976*. Turku: Text Linguistics Research Group, Academy of Finland, 1976.
4. Gregersen, Kirsten, ed., Hans Basbøll & Jacob Mey, co-eds., *Papers from the Fourth Scandinavian Conference of Linguistics, Hindsgaol, January 6-8, 1978*. (Odense University Studies in Linguistics 3.) Odense: Odense University Press, 1978.
5. Petterson, Thore, ed. *Papers from the Fifth Scandinavian Conference of Linguistics, Frostavallen, April 27-29 1979. Part I: Section papers*. (Acta Universitatis Lundensis. Sectio 1, Theologica, juridica, humaniora, 30.) Stockholm: Almqvist & Wiksell International.

Petterson, Thore, ed. *Papers from the Fifth Scandinavian Conference of Linguistics, Frostavallen, April 27-29 1979. Part II: General sessions*. (Acta Universitatis Lundensis. Sectio 1, Theologica, juridica, humaniora, 31.) Stockholm: Almqvist & Wiksell International.
6. Hyltenstam, Kenneth & Moira Linnarud, eds. *Interlanguage. Workshop at the Fifth Scandinavian Conference of Linguistics, Frostavallen, April 27-29 1979*. Stockholm: Almqvist & Wiksell International.

Petterson, Thore, ed. *Aspectology. Workshop at the Fifth Scandinavian Conference of Linguistics, Frostavallen, April 27-29 1979*. Stockholm: Almqvist & Wiksell International.
7. Fretheim, Thorstein & Lars Hellan, eds. *Papers from the Sixth Scandinavian Conference of Linguistics, Røros, June 19-21, 1981*. Trondheim: Tapir, 1982.
8. Karlsson, Fred, ed. *Papers from the Seventh Scandinavian Conference of Linguistics, Hanasaari, Finland, December 17-19, 1982, 1-2*. (Publications [of the Department of General Linguistics, University of Helsinki], 9-10.) Helsinki: University of Helsinki, Department of General linguistics, 1983.
8. Tøgeby, Ole, ed. *Papers from the Eighth Scandinavian Conference of Linguistics, 1-2*. København: Institut for Nordisk Filologi, 1985.

9. Dahl, Östen, ed. *Papers from the Ninth Scandinavian Conference of Linguistics, Stockholm, January 9-11, 1986*. Stockholm: Institute of Linguistics, 1986.

Dahl, Östen & Anders Holmberg, eds. *Scandinavian syntax. Workshop at the Ninth Scandinavian Conference of Linguistics, Stockholm, January 1986*. Stockholm: Institute of Linguistics, 1986.

10. Rosén, Victoria, ed. *Papers from the Tenth Scandinavian Conference of Linguistics, Bergen, June 11-13, 1987*. Bergen: Department of Linguistics and Phonetics, University of Bergen.

11. Niemi, Jussi, ed. *Papers from the Eleventh Scandinavian Conference of Linguistics, 1-2*. (Kielitieteellisiä tutkimuksia / Joensuun yliopisto. Humanistinen tiedekunta, 14-15.) Joensuu: Yliopisto, 1989.

12. Ármann Sigurdsson, Halldór, ed., Þorsteinn G. Indriðason & Eiríkur Rögnvaldsson, assoc. eds., *Papers from the Twelfth Scandinavian Conference of Linguistics, Reykjavík, June 14-16, 1990*. Reykjavík: Linguistic Institute, University of Iceland, 1991.

13. Heltoft, Lars & Hartmut Haberland, eds. *Proceedings of the Thirteenth Scandinavian Conference of Linguistics [January, 1992]*. Roskilde: Department of Languages and Culture, Roskilde University, 1996.

14. Allwood, Jens, Bo Ralph, Paula Andersson, Dora Kós-Dienes & Åsa Wengelin, eds. *Proceedings of The XIVth Scandinavian Conference of Linguistics and The VIIIth Conference of Nordic and General Linguistics, August 16-21 1993. General session*. Vols. 1-2. (Gothenburg Papers in Theoretical Linguistics, 69.) Göteborg: Department of Linguistics, University of Gothenburg, 1994.

Hedlund, Cecilia & Anders Holmberg, eds. *Proceedings of The XIVth Scandinavian Conference of Linguistics and The VIIIth Conference of Nordic and General Linguistics, August 16-21 1993. Special session on Scandinavian syntax*. (Gothenburg Papers in Theoretical Linguistics, 70.) Göteborg: Department of Linguistics, University of Gothenburg, 1994.

Jokinen, Kristiina, ed. *Proceedings of The XIVth Scandinavian Conference of Linguistics and The VIIIth Conference of Nordic and General Linguistics, August 16-21 1993. Special session on pragmatics in dialogue management*. (Gothenburg Papers in Theoretical Linguistics, 71.) Göteborg: Department of Linguistics, University of Gothenburg, 1994.

Allwood, Jens, Frank Gregersen, Jussi Niemi, Dennis Day & Sture Ureland, eds. *Proceedings of The XIVth Scandinavian Conference of Linguistics and The VIIIth Conference of Nordic and General Linguistics, August 16-21 1993. Four special sessions: Spoken language, Neurolinguistics, Anthropological linguistics, Language contacts across the Baltic*.

(Gothenburg Papers in Theoretical Linguistics, 72.) Göteborg: Department of Linguistics, University of Gothenburg, 1994.

Junefelt, Karin, ed. *Proceedings of The XIVth Scandinavian Conference of Linguistics and The VIIIth Conference of Nordic and General Linguistics, August 16-21 1993. Special session on activity theory.* (Gothenburg Papers in Theoretical Linguistics, 73.) Göteborg: Department of Linguistics, University of Gothenburg, 1995.

Strömquist, Sven, ed. *Proceedings of The XIVth Scandinavian Conference of Linguistics and The VIIIth Conference of Nordic and General Linguistics, August 16-21 1993. Special session on language development.* (Gothenburg Papers in Theoretical Linguistics, 74.) Göteborg: Department of Linguistics, University of Gothenburg, 1994.

15. Moen, Inger, Hanne Gram Simonsen & Helge Lødrup, eds. *Papers from the XVth Scandinavian Conference of Linguistics, January 13-15 January 1995.* Oslo: Department of Linguistics, University of Oslo, 1995.
16. Haukioja, Timo, ed. *Papers from the Sixteenth Scandinavian Conference of Linguistics, Turku, November 14-16 1996.* (Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja, 60.) Turku: Turun yliopisto, 1998.
17. Lindberg, Carl-Erik & Steffen Nordahl Lund, eds. *Papers from the 17th Scandinavian Conference of Linguistics, Nyborg, Denmark, August 20-22, 1998, 1-2.* (Odense Working Papers in Language and Communication, 19.) Odense: Institute of language and Communication, University of Southern Denmark, 2000.
18. Holmer, Arthur, Jan-Olof Svantesson & Åke Viberg, eds. *Proceedings of the Eighteenth Scandinavian Conference of Linguistics, May 18-20, 2000, 1-2.* (Travaux de l'Institut de linguistique de Lund, 39:1-2.) Lund: Department of Linguistics and Phonetics, 2001.
19. Dahl, Anne, Kristine Bentzen & Peter Svenonius, eds. *Proceedings of the 19th Scandinavian Conference of Linguistics, Tromsø. January 10-12, 2002: 1-2.* (= Nordlyd. Tromsø University Working Papers on Language & Linguistics 31, nos. 1-2. 2003. Published online only: <http://www.ub.uit.no/munin/nordlyd/>.)

Dahl, Anne, Marit Richarsen Westergaard & Peter Svenonius, eds. *Proceedings of the 19th Scandinavian Conference of Linguistics, Tromsø. January 10-12, 2002: 3. Acquisition.* (= Nordlyd. Tromsø University Working Papers on Language & Linguistics 31, no. 3. 2003. <http://www.ub.uit.no/munin/nordlyd/>.)

Dahl, Anne & Peter Svenonius, eds. *Proceedings of the 19th Scandinavian Conference of Linguistics, Tromsø. January 10-12, 2002: 4. Finnic.* (= Nordlyd. Tromsø University Working Papers on Language & Linguistics 31, no. 4. 2003. <http://www.ub.uit.no/munin/nordlyd/>.)

- Jørgensen, Jens Normann, Anne Dahl & Peter Svenonius, eds. *Proceedings of the 19th Scandinavian Conference of Linguistics, Tromsø. January 10-12, 2002: 5. Bilingualism.* (= *Nordlyd. Tromsø University Working Papers on Language & Linguistics* 31, no. 5. 2003. [http://www.ub.uit.no/munin/nordlyd/.](http://www.ub.uit.no/munin/nordlyd/))
- Dahl, Anne & Peter Svenonius, eds. *Proceedings of the 19th Scandinavian Conference of Linguistics, Tromsø. January 10-12, 2002: 6. Supplement.* (= *Nordlyd. Tromsø University Working Papers on Language & Linguistics* 31, no. 6. 2004. [http://www.ub.uit.no/munin/nordlyd/.](http://www.ub.uit.no/munin/nordlyd/))
20. Karlsson, Fred, ed. *Proceedings of the 20th Scandinavian Conference of Linguistics in Helsinki, January 7-9, 2004.* (University of Helsinki, Department of General Linguistics, Publications, No. 36.) Helsinki: University of Helsinki, Department of General Linguistics, 2004. Published online under the address <http://www.ling.helsinki.fi/kielitiede/20scl/proceedings.shtml>. [The electronic proceedings contain 14 of the 48 section papers presented at the conference. The three plenary talks that were given at the meeting are not included.]
21. The papers from the 21st Scandinavian Conference of Linguistics in Trondheim, June 1-4, 2005, will remain unpublished.
22. Götzsche, Hans, ed. *Memory, Mind and Language: Proceedings of the 22nd Scandinavian Conference of Linguistics, Aalborg, 2006.* Newcastle, UK: Cambridge Scholars Press.
23. Saxena, Anju & Åke Viberg, ed. *Multilingualism: Proceedings of the 23rd Scandinavian Conference of Linguistics, Uppsala University, 1-3 October 2008.* Uppsala: Acta Universitatis Upsaliensis.
24. Tirkkonen, Jani-Matti & Esa Anttikoski, eds. *Proceedings of the 24th Scandinavian Conference of Linguistics.* (Publications of the University of Eastern Finland. Reports and Studies in Education, Humanities, and Theology, 5) Joensuu: University of Eastern Finland.

**PUBLICATIONS OF THE UNIVERSITY OF EASTERN FINLAND
REPORTS AND STUDIES IN EDUCATION, HUMANITIES, AND THEOLOGY**

1. Mervi Asikainen, Pekka E. Hirvonen ja Kari Sormunen (toim.). *Ajankohtaista matemaattisten aineiden opetuksen ja oppimisen tutkimuksessa. Matematiikan ja luonnontieteiden opetuksen tutkimuspäivät Joensuussa 22.–23.10.2009*. 2010.
2. Sinikka Niemi och Pirjo Söderholm (red.). *Svenskan i Finland. 15 och 16 oktober 2009 vid Joensuu universitet*. 2011.
3. Johanna Lintunen. *Taide ja hoiva rajapinnoille. Selvitys taide- ja kulttuurisisältöisten hyvinvointipalvelujen kysynnästä, tarjonnasta ja yhteistyöverkostoista Pohjois-Karjalassa*. 2011.
4. Hannu Kemppanen, Jukka Mäkisalo ja Alexandra Belikova (toim.). *Kotoista ja vierasta mediassa. Venäjästä suomeksi ja suomesta venäjäksi -workshop Joensuussa 7.–8.10.2010*. 2012.
5. Jani-Matti Tirkkonen and Esa Anttikoski (eds.). *Proceedings of the 24th Scandinavian Conference of Linguistics*. 2013.

**JANI-MATTI TIRKKONEN
& ESA ANTTIKOSKI (EDS.)**
*Proceedings of the 24th
Scandinavian Conference
of Linguistics*

The present collection of papers is based on the presentations given at the 24th Scandinavian Conference of Linguistics held at the University of Joensuu (now: University of Eastern Finland) August 25 – 27, 2010.

The papers cover many areas of Linguistics, ranging from grammar to discourse. All papers were peer-reviewed.



UNIVERSITY OF
EASTERN FINLAND

PUBLICATIONS OF THE UNIVERSITY OF EASTERN FINLAND
Reports and Studies in Education, Humanities, and Theology

ISBN: 978-952-61-0923-7 (NID.)

ISSNL: 1798-5641

ISSN: 1798-5641

ISBN: 978-952-61-0924-4 (PDF)

ISSN: 1798-565X (PDF)