

# **Re-thinking synonymy: semantic sameness and similarity in languages and their description**

---

Book of Abstracts

28.-30.10.2010, Helsinki

# Table of Contents

<b>LIST OF PARTICIPANTS:</b>	<b>1</b>
<b>INVITED SPEAKERS</b>	<b>3</b>
GEERAERTS DIRK	3
HASPELMATH MARTIN	4
LEVIN BETH	5
<b>SCIENTIFIC COMMITTEE</b>	<b>6</b>
<b>ORGANIZING COMMITTEE</b>	<b>6</b>
<b>WORKSHOP ON COMPUTATIONAL APPROACHES TO SYNONYMY</b>	<b>7</b>
GRAEME HIRST, KENTARO INUI AND STEDE MANFRED	7
<b>ABSTRACTS</b>	<b>8</b>
ANDERSSON MARTA	8
ANISHCHANKA ALENA, SPEELMAN DIRK AND GEERAERTS DIRK	9
ARPPE ANTTI AND DIVJAK DAGMAR	11
BACKUS AD AND MOS MARIA	13
BORIN LARS AND FORSBERG MARKUS	15
CALUDE ANDREEA AND PAGEL MARK	17
CAPPELLE BERT AND DESUTTER GERT	18
COPESTAKE ANN AND HERBELOT AURÉLIC	20
DALMAS MARTINE AND DOBROVOL'SKIJ DMITRIJ	22
DESHORS SANDRA C.	24
DEVOS MAUD	26
FAULHABER SUSEN	28
GAILLARD BENOÎT, NAVARRO EMMANUEL AND GAUME BRUNO	30
GLYNN DYLAN	31
GLYNN DYLAN AND LEVSHINA NATALIA	33
GOROKHOVA SVETLANA	35
HALDER GUIDO	37
HUUMO TUOMAS AND LEHISMETS KERSTEN	39
HUUMO TUOMAS AND OJUTKANGAS KRISTA	41
HÄUSLER SABINE	43
JÜRINE ANNI	44
KLAVAN JANE	46
KHOLODILOVA MARIA	48
KHOUTYZ IRINA	50
LURAGHI SILVIA	51
LIU DILIN	53
MARTÍ SOLANO RAMÓN AND RALUCA NITA	56
MATUSCHEK MICHAEL AND GUREVYCH IRYNA	58
MONTAZERI NILOOFAR AND HOBBS JERRY	60
MULLI JUHA	62
NACEY SUSAN AND EGAN THOMAS	64
NISSILÄ NIINA AND PILKE NINA	66
NEDJALCOV IGOR	68
O'CONNOR KATHLEEN AND CORTEEL CÉLINE	70

OVERSTEEGEN ELEONORE	72
PALOLAHTI MARIA	73
PAULSEN GEDA	75
PÄIVIÖ PIIA	76
RAKHILINA EKATERINA AND TRIBUSHININA ELENA	77
RAUKKO JARNO	79
RINGBOM HÅKAN	81
ROBERT STÉPHANE	82
ROCHE CHRISTOPHE AND CALBERG-CHALLOT MARIE	84
SAMEDOVA NEZRIN	86
SCHMEISER BENJAMIN	88
SIROLA-BELLIARD MAIJA	90
SOARES DA SILVA AUGUSTO	92
SUTROP URMAS	94
STEDE MANFRED	95
WANG TONG AND HIRST GRAEME	96
WIEMER BJÖRN AND SOCKA ANNA	98
VILKKI LIISA	100
VÄSTI KATJA	102
ZENNER ELINE, SPEELMAN DIRK AND GEERAERTS DIRK	104

## List of Participants:

<b>Last name</b>	<b>First name</b>	<b>Affiliation</b>
Andersson	Marta	Stockholm University
Anishchanka	Alena	Univeristy of Leuven (QLVL)
Arppe	Antti	University of Helsinki & Alberta
Backus	Ad	Tilburg University
Borin	Lars	University of Gothenburg
Calberg-Challot	Marie	Onomia - Savoie Technoloac
Calude	Andreea	University of Reading
Cappelle	Bert	Univesity College Ghent
Copestake	Ann	Computer Laboratory, University of Cambridge
Corteel	Céline	Université d'Artois
Dalmas	Martine	Université Paris-Sorbonne
Deshors	Sandra C.	University of Sussex
Desutter	Gert	University College Ghent
Devos	Maud	Royal Museum for Central Africa
Divjak	Dagmar	University of Sheffield
Dobrovol'skij	Dmitrij	Russian Academy of Sciences
Egan	Thomas	Hedmark Univsersity College
Faulhaber	Susen	Friedrich-Alexander-University Erlangen-Nuremberg
Forsberg	Markus	University of Gothenburg
Gaillard	Benoît	CLLE-ERSS, Univiversity Toulouse 2
Gaume	Bruno	CLLE-ERSS, Univiversity Toulouse 2
Geeraerts	Dirk	University of Leuven
Glynn	Dylan	University of Lund
Gorokhova	Svetlana	St Petersburg State University
Gurevych	Iryna	Technische Universität Darmstadt
Halder	Guido	The University of Texas at Austin
Herbelot	Arélic	Computer Laboratory, University of Cambridge
Hirst	Graeme	University of Toronto
Hobbs	Jerry	University of Southern California
Huumo	Tuomas	University of Tartu
Häusler	Sabine	Martin-Luther-Universität Halle-Wittenberg
Jürine	Anni	University of Tartu
Kholodilova	Maria	Saint-Petersburg State University
Khoutyz	Irina	Kuban State University
Klavan	Jane	University of Tartu
Lehismets	Kersten	University of Tartu
Levshina	Natalia	University of Leuven
Liu	Dilin	The University of Alabama
Luraghi	Silvia	Università di Pavia
<b>Last name</b>	<b>First name</b>	<b>Affiliation</b>

Marti Solano	Ramon	University of Limoges
Matuschek	Michael	Technische Universität Darmstadt
Miestamo	Matti	Helsinki Collegium for Advanced Studies
Montazeri	Niloofer	University of Southern California
Mos	Maria	Tilburg University
Mulli	Juha	University of Eastern Finland at Joensuu
Murphy	Lynne	University of Sussex
Nacey	Susan	Hedmark University College
Navarro	Emmanuel	IRIT, University Toulouse 3
Nedjalkov	Igor	St-Petersburg State University
Nissilä	Niina	University of Vaasa
O'Connor	Kathleen	Université de Lille 3
Ojutkangas	Krista	University of Turku
Oversteegen	Eleonore	University of Tilburg
Pagel	Markus	University of Reading
Palolahti	Maria	University of Helsinki
Paulsen	Geda	Åbo Akademi University
Pilke	Nina	University of Vaasa
Päiviö	Piia	University of Toronto
Rakhilina	Ekaterina	Russian Academy of Sciences
Raluca	Nita	University of Nantes
Raukko	Jarno	University of Helsinki
Ringbom	Håkan	Åbo Akademi University
Robert	Stéphane	CNRS-LLACAN, Paris
Roche	Christophe	University of Savoie
Samedova	Nezrin	Azerbaijan University of Languages
Schmeiser	Benjamin	Illinois State University
Sirola-Belliard	Maija	University of Tampere
Soares da Silva	Augusto	Catholic University of Portugal
Soares da Silva	Augusto	Catholic University of Portugal
Socka	Anna	University of Gdansk
Speelman	Dirk	University of Leuven
Stede	Manfred	University of Potsdam
Sutrop	Urmas	Institute of Estonian Language, Tallinn, University of Tartu
Tribushinina	Elena	University of Antwerp
Wakayama	Masayuki	Aichi Shukutoku University
Wang	Tong	University of Toronto
Wiemer	Björn	Johannes-Gutenberg-Universität
Vilkki	Liisa	University of Helsinki
Västi	Katja	University of Oulu & Helsinki
Zenner	Eline	University of Leuven

## Invited Speakers

Geeraerts Dirk  
University of Leuven

### Criteria for meaning discrimination and the identification of synonymy

In this talk, I will attempt to clarify the notion of synonymy against the background of the theoretical and methodological developments in contemporary (lexical) semantics. My starting point is the observation that the problem of synonymy and the problem of polysemy are essentially the same: if we have valid criteria for establishing whether we are dealing with one or more separate readings of a single item or construction, then the same criteria can be used for establishing whether two items or constructions illustrate one or more meanings.

It follows, in general, that the difficulties with the criteria for establishing difference of meaning that I analyzed in Geeraerts (1993) hold for synonymy as well as for polysemy. I will illustrate the point by looking more closely at how the three main criteria distinguished in the 1993 article might be applied to the identification of synonymy.

More specifically, zooming in on the so-called 'definitional criterion' will allow me to go beyond the 1993 framework by taking into account a number of important methodological developments of the last two decades. I will have a look at a referential approach to lexical meaning (as in Geeraerts, Grondelaers & Bakema 1994), a contextual feature approach (as in Speelman & Geeraerts 2009), and a word-space based distributionalist approach (as in Heylen, Peirsman, Geeraerts & Speelman 2008). For each of these approaches, I will try to indicate advantages and the drawbacks, an exercise that will lead to the conclusion that there is as yet no single and generally applicable method for establishing synonymy.

This conclusion will then be deepened by a critical look at the concept of synonymy itself: if we re-think synonymy, could it be that the absence of a methodological consensus reflects not just the weakness of linguistic methods, but a flaw in our conception of meaning ?

#### Bibliography

- Geeraerts, Dirk (1993). 'Vagueness's puzzles, polysemy's vagaries'. *Cognitive Linguistics* 4: 223-272.
- Geeraerts, Dirk, Stefan Grondelaers and Peter Bakema (1994). *The Structure of Lexical Variation. Meaning, Naming, and Context*. Berlin/New York: Mouton de Gruyter.
- Heylen, Kris, Yves Peirsman, Dirk Geeraerts and Dirk Speelman (2008). 'Modelling word similarity: An evaluation of automatic synonymy extraction algorithms'. *Proceedings of the Sixth International Language Resources and Evaluation Marrakech: European Language Resources Association*. (Link: [http://www.lrec conf.org/proceedings/lrec2008/pdf/818\\_paper.pdf](http://www.lrec conf.org/proceedings/lrec2008/pdf/818_paper.pdf).)
- Speelman, Dirk and Dirk Geeraerts (2009). 'Causes for causatives: the case of Dutch 'doen' and 'laten''. In Ted Sanders and Eve Sweetser (eds.), *Causal Categories in Discourse and Cognition* 173-204. Berlin: Mouton de Gruyter.

## **The role of homosemasy (semantic equivalence) for constructing comparative concepts in cross-linguistic studies**

Linguists who specialize in the comparison of languages routinely make use of homosemasy, i.e. semantic equivalence across languages, as a *tertium comparationis*. This is true not only for lexical comparison (whether historical-comparative or typological), but also for most cases of morphosyntactic comparison: purely form-based grammatical comparison is too difficult, because forms vary too much. In fact, typologists have often claimed that typological comparison of grammatical patterns is exclusively meaning-based, but this is too strong – in addition to semantic similarity, typological comparative concepts often involve some formal similarity as well.

An important observation is that just as there is no true synonymy, there is no true homosemasy: Meanings are often similar across languages, but rarely match perfectly. Cross-linguistic semantic variation is rampant. This means that cross-linguistic comparison must involve a set of special semantic concepts (which I call comparative concepts), which are not meanings of particular language forms, but are created specifically for the purpose of language comparison. I illustrate such comparative concepts from the World Loanword Database (Haspelmath & Tadmor 2009) and from their use in semantic maps (Haspelmath 2003). Unless one adopts an extreme decompositional approach, the comparative concepts cannot be used for the description of language-particular categories. Finally, I show that the same holds for morphosyntactic, formally defined entities such as cases, relative clauses or converbs: Comparative concepts are separate from descriptive categories, and language comparison and language description/analysis are more independent of each other than is often thought (cf. Haspelmath 2009, 2010).

### Bibliography

- Haspelmath, Mark, & Uri Tadmor (eds.). 2009. *World Loanword Database*. Munich: Max Planck Digital Library. <http://wold.livingsources.org/>.
- Haspelmath, Martin. 2003. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In Michael Tomasello (ed.), *The new psychology of language: Cognitive and functional approaches to language structure*, vol. 2, 211–242. Mahwah, NJ: Erlbaum.
- Haspelmath, Martin. 2009. La typologie des langues pourquoi est-elle possible? *Bulletin de la Société de Linguistique de Paris* 104. 17-38.
- Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in cross-linguistic studies. *Language* 86(3).

## **Synonymy and Arbitrariness in Linguistic Argumentation**

Among the questions posed in the call for papers for the conference is "What does synonymy (at any level/in any form) reveal about language?" A recurring answer is that synonymy --- or even near-synonymy --- can be used to argue for arbitrariness in the form-function (or semantics-syntax) mapping. Thus, researchers cite pairs such as "leaves" and "foliage", "mail" and "letters", or "poems" and "poetry" to argue for the arbitrariness of the mass/count distinction (e.g., Chierchia 1998, Ware 1975). Similarly, pairs like English "blush" and its Italian translation equivalent "arrossire" have been used to argue for the arbitrariness of the unaccusative/unergative verb distinction (e.g., Rosen 1984) and pairs like alternating "sour" and transitive-only "embitter" or alternating "shake" and intransitive-only "shiver" for the arbitrariness of the causative alternation (e.g., Farsi 1974).

In this talk, I will revisit some of these examples and suggest that they do not actually support arbitrariness in the form-function mapping for the associated phenomena. I will argue that these pairs at best represent near-synonyms so that their members do differ in meaning and that this difference in meaning is one that is critical to their difference in behavior. I will further suggest that such pairs arise because meanings are construals of the world, so that even if in one instance "leaves" and "foliage" might refer to the same entity -- the basis for the synonymy claim -- the two words capture different representations of this entity. In fact, it is this point that underlies Wierzbicka's well-known study of "oats" and "wheat" (1985). The conclusion, then, is that words with similar meaning constitute a rich domain for studying the semantics-syntax mapping and for developing finer-grained lexical semantic analyses.

## **Scientific committee**

Arppe Antti (University of Helsinki)

Austin Peter (SOAS, London)

Creissels Denis (University of Lyon)

Divjak Dagmar (University of Sheffield)

Goldberg Adele (Princeton University)

Gries Stefan (UCSB)

Huumo Tuomas (University of Tartu)

Janda Laura (University of Tromsø)

Jantunen Jarmo (University of Oulu)

Luraghi Silvia (University of Pavia)

Rice Sally (University of Alberta)

Siewerska Anna (University of Lancaster)

Wälchli Bernhard (University of Berne)

## **Organizing committee**

Arppe Antti (University of Helsinki)

Kittilä Seppo (University of Helsinki)

Kyröläinen Aki-Juhani (University of Turku)

Niemelä Maarit (University of Oulu)

Nikolaev Alexandre (University of Joensuu)

Rostila Jouni (University of Tampere)

Vartiainen Turo (University of Helsinki)

Visapää Laura (University of Helsinki)

## **Workshop on Computational Approaches to Synonymy**

Graeme Hirst, Kentaro Inui and Stede Manfred  
University of Toronto  
Tohoku University  
University of Potsdam

Many problems and applications in computational linguistics and natural language processing implicitly invoke, in various forms, the concept of synonymy or identity of meaning. In one way or another, they involve either determining identity (or non-identity) of meaning in different surface forms or creating different surface forms for a single meaning.

For example, paraphrase recognition is an important component of the more-general problem of recognizing textual entailment. Textual tailoring and personalization seeks to find the most effective linguistic realization of a message for a particular user; automatically simplifying texts and creating stylistic variations are special cases of this. Lexical choice in text generation tries to find the best word for a given meaning and to discriminate it from other words that are close in meaning but not synonymous in the context. Cross-lingual document retrieval and other cross-lingual applications such as, in particular, machine translation conflate the ideas of synonymy and translation equivalence.

But while there has been a large amount of research on computational methods for determining degree of similarity in lexical meaning and for recognizing paraphrase, little attention has been given to theoretical considerations of synonymy. Mostly, it is treated as a boolean property (two words are or aren't in the same synset; two sentences are or aren't mutual entailments) with little thought of any theoretical underpinning.

On the other hand, the real-world linguistic problems that natural language processing addresses provide useful test cases for linguistic theories of synonymy, and the computational methods developed are de facto theories of synonymy even if not intended as such.

This workshop will explore computational approaches to synonymy, with an emphasis on explicating their implicit theoretical notions and their implications for linguistic theory.

## **Abstracts**

Andersson Marta  
Stockholm University

### **Synonymy of result adverbials**

The purpose of this paper is to investigate semantic, syntactic, pragmatic, discourse and textual properties of five one-word adverbials (so, then, thus, therefore, hence) marking the specific function of the causal relation, i.e. result in written discourse of the BNC. These words are considered roughly synonymous, yet we should be skeptical about their full synonymy, even though the hypothesis is that they are all oriented towards a conclusion and cue a relationship of resultative nature. Yet the semantic space of result is not evenly divided between them and each of them can be understood to express another cognitive way of thinking about result, which is related to certain pragmatic, semantic, distributional and, possibly, discourse-textual differences.

The paper will demonstrate and discuss these differences and try to answer the question why English has all these different ways of licensing result relations. The words in focus will be compared as members of an onomasiological set, which yields not only similarities between them, but also certain differences, as some are more prototypical of result than others. Consequently, this may lead to the questioning of the aptness of the label “result” (which even intuitively seems to be very general) as describing the role of all the investigated words in managing information flow, text structure and social interaction. Thus an investigation of the similarities and differences in the behavior of so, therefore, hence etc. will not only be situated in the area of purely linguistic exercises, but will also touch upon the area of human cognition and social needs for making such distinctions.

## Co-extension and near-synonymy in color categorization

**Research question** – Continuing the line of research set out in Geeraerts (2006), Geeraerts et al. (1994), Geeraerts and Speelman (2010), the study addresses the interaction between concept-related semantic features and sociolectal parameters underlying lexical heterogeneity. By mapping the referential range of color terms in internet product catalogs on the basis of their RGB values, we can explore the conceptual and contextual factors that determine the extension of color categories of different specificity and the patterns of their referential overlap. Specifically, we focus on the phenomena of **co-extension** and **near-synonymy** as the major construal types in overlapping categories.

**Background** – The analysis applies the categorization model proposed by MacLaury (1992, 1997) that specifically addresses the multiple instances of referential overlap and near-synonymy in the color domain. Using vast empirical data, MacLaury identified four types of category construal - **inclusion, coextension, near-synonymy** and **complementation** - all having the same underlying cognitive mechanism of (dis)similarity emphasis. Focusing on co-extension as the most common type of referential overlap, he also demonstrated that the asymmetries in this type of construal license the use of near-synonymous names within the same referential range. With the growing recognition of (dis)similarity as a basic categorization mechanism (Jameson and Alvarado 2003, Langacker 2009, Taylor 2003), this model proves particularly insightful for the linguistic analysis of (near)-synonymy. Specifically, it allows accounting for salience effects, non-discreteness and multidimensionality of lexical meaning that had been revealed by cognitive linguistic studies and that provide major challenges for the notion of synonymy traditionally formulated in terms of ‘mosaic’ semantic fields, semantic features, or truth-conditional equivalence (Geeraerts et al. 1994, Geeraerts 2006, Divjak and Gries 2006, Glynn 2010).

**Method and design** – The study is based on an extensive corpus of color names and color samples (over 65 000 observations) used by US manufacturers for online marketing in four product categories (automobiles, clothing, make-up, and house paints). The available material provides linguistic, sociolectal and referential parameters of color naming in advertising and is particularly suited for an independent measurement of referential range of the color names using the automatically retrieved RGB values of the color samples.

Building on MacLaury’s criteria for inclusion, coextension, and near-synonymy, we explore semasiological and onomasiological variation in overlapping color categories, such as purple, mauve, lavender, lilac. Starting from a semasiological perspective, we first map the referential range of basic and non-basic color words in order to identify the semasiological structure of color categories and the patterns of co-extension and near-synonymy in their referential overlap. The second stage investigates whether the degree and the pattern of referential overlap significantly correlate with conceptual factors (such as vagueness and prototypicality of the color concept, entrenchment and complexity of the color name, specificity of the color name and color concept) and contextual factors (product category and type, target consumer). Taking the onomasiological perspective, we zoom in on the areas of referential overlap and explore the factors that influence the degree of lexical heterogeneity and the choices of basic and non-basic color names for a specific referential area. For instance, as a conceptual factor, is there more or less diversity in one area of the color spectrum than in the other, and, as an additional contextual factor, is this effect the same across the four product categories?

The multivariate (regression) techniques used in the analysis allow us to give an integrated multifactorial account for the interaction of conceptual and contextual factors in color category construal and thus take the notion of near-synonymy in color-naming beyond referential overlap.

## Bibliography

- Langacker, Ronald W. 2009: A dynamic view of usage and language acquisition. *Cognitive Linguistics* 20(3): 627-640.
- Langacker, Ronald W. 2009: *Investigations in Cognitive Grammar*. Berlin, New York: Mouton de Gruyter.
- Divjak, D. 2006. Delineating and Structuring Near-Synonyms. *Corpora in cognitive linguistics*, St. Th. Gries & A. Stefanowitsch (eds), Berlin: Mouton, 19-56.
- Arppe, A. 2008. *Univariate, bivariate and multivariate methods in corpus-based lexicography - a study of synonymy*. Dept. of General Linguistics, University of Helsinki, <http://urn.fi/URN:ISBN:978-952-10-5175-3>.

## Synonymy is *both Gradient and Context-Dependent*

Regardless of the semantic theory or model, in practically all treatments of synonymy the degree of semantic similarity among words is represented as a constant property – be it Boolean or gradient – holding among a selected set of words (often pairs), throughout an aggregate of all conceivable usage contexts. Consequently, little attention is paid to the degree of synonymy, semantic similarity, or interchangeability among purported synonyms in the underlying multitude of individual distinct contexts of usage. We argue that in addition to synonymy being a gradient phenomenon (cf. Miller & Charles 1991), synonymy can be seen to vary in natural usage from one context to another, such contexts being constituted by explicitly identifiable linguistic properties that can be analyzed according to some linguistic theory. This view is in line with a probabilistic theory of how language works.

Our view results from the computational analysis of four Finnish THINK verbs (3,404 sentences extracted from corpora; Arppe 2008) and six Russian TRY verbs (1,351 sentences; Divjak 2010). We explicitly studied the impact of a multitude of linguistic and extra-linguistic properties on the selection of one synonym over others in a large number of contexts. In our analysis, we applied polytomous logistic regression, which models the proportions of occurrence among a number of possible outcomes, given the occurrence of explanatory properties in various observable contexts (Arppe 2008). This statistical modeling approach implies that, given a linguistic analysis scheme, the distributions of expected probabilities of occurrence among possible alternative outcomes, i.e. synonyms, may in principle vary from one combination of explanatory properties, i.e. a context, to another. In practice, this variation ranges from a near-categorical preference for only one of the words in the synonym set, to near-equiprobable contexts where all synonyms are in principle equally likely to occur, with a variety of other recognizable scenarios of expected occurrence in between these extremes. Thus, in terms of the explanatory properties, in practice, some contexts do not allow for synonymy at all, whereas others do not provide much explicit distinction among the synonymous alternatives. In the latter case, more information about e.g. the general discourse situation and its participants, including “nuances” outside traditional linguistic analysis (Inkpen & Hirst 2006), would appear necessary.

Thus, synonymy clearly appears to be both a context-dependant and a contextually varying phenomenon (cf. Sinclair’s [1991] notion of variation of contextual preferences among word-forms of the same lemma). It does, nevertheless, vary systematically in terms of the context, and we will provide concrete example cases of this for both Finnish and Russian.

### Bibliography

- Arppe, Antti. 2008. *Univariate, bivariate and multivariate methods in corpus-based lexicography – a study of synonymy*. Publications of the Department of General Linguistics, University of Helsinki, No. 44.
- Divjak, Dagmar. 2010 (in press). *Structuring the Lexicon: a Clustered Model for Near-Synonymy*. Mouton de Gruyter: Berlin/New York.
- Inkpen, Diana & Graeme Hirst 2006. Building and Using a Lexical Knowledge-Base of Near-Synonym Differences. *Computational Linguistics* 32:2, 223-262.
- George A. Miller & Walter G. Charles. 1991. Contextual correlates of semantic similarity. *Language and Cognitive Processes*, 1-28.

Sinclair, John 1991. *Corpus, Concordance, Collocation*. Oxford : Oxford University Press.

## Near synonymy in partially specific constructions as evidence for the lexicon-syntax continuum

Rethinking the lexicon-syntax dichotomy as a continuum, as proposed in Construction Grammar and associated theories, raises the question to what extent there may be synonymy across what used to be seen as distinct modules. In this contribution, we discuss two Dutch constructions that express a very similar meaning, but are outwardly very different, one being more ‘morphological’ and the other more ‘syntactic’. The main questions we address is whether the same factors play a role in lexical and syntactic (near-) synonymy, particularly in the distribution of these two partially specific constructions, and to what extent differences in distribution reflect differences in constructional meaning.

Examples of the V-BAAR and IS TE V constructions are given in (1) and (2) respectively. In both cases, the construction expresses the (lack of) possibility that an entity X (expressed as the subject) undergoes event Y (expressed as the verbal stem in (1) and in the infinitival form in (2)).

1. Zijn handschrift is niet lees-baar  
His handwriting is not read-able
2. Zijn handschrift is niet te lezen  
His handwriting is not to read.INF

A corpus analysis of instantiations of the constructions shows that both are rather frequent (> 200 types for each construction in the 10-million word Corpus of Spoken Dutch). Out of these, 55 verbs occur with both constructions; instantiations with these verbs therefore potentially constitute pairs of synonyms. The corpus data and an experiment using novel instantiations showed that the main meaning difference lies in the speaker’s stance: the IS TE V construction tends to express the assessment of possibility, whereas the V-BAAR construction is often used for more factual statements.

Of the 55 verbs, a smaller subset shows signs of lexicalization, in one or both of the constructions. In such cases, the form has a specific meaning that cannot easily be deduced compositionally. For example, the verb *eten* (to eat) produces the forms *eetbaar* (‘edible, not poisonous’) and *is te eten* ‘is tasty’. Usually, the differences are less dramatic: the verb *bereiken* (can be reached) can refer to both physical and mental destinations, but *bereikbaar* (reachable) is nearly exclusively found with concrete locations.

The immediate syntactic context also exerts influence: the IS TE V construction is more commonly found with verbs that prefer clausal objects. The effect is that pragmatic associations differ for the two constructions, and to the extent that pragmatics is seen as part of meaning, this reduces the synonymy between the two. However, at least one factor that plays a role, priming, has nothing to do with meaning: the sheer occurrence of one of the constructions makes its reappearance later in the same conversation more likely.

We conclude two things. First, novel instantiations and some of the distributional data provide evidence that the two constructions show subtle semantic differences. Second, similar factors apply to both the morphological and the syntactic construction, which argues against maintaining a strict dichotomy between lexicon and syntax.



## Beyond the synset: Swesaurus -- a fuzzy Swedish wordnet

Swesaurus is a free Swedish wordnet currently under construction in our research unit. Swesaurus is made up in part by a combination of a number of pre-existing freely available lexical resources. Two central resources for the purposes of this presentation are SALDO and Synlex.

SALDO (Borin et al. 2008; Borin & Forsberg 2009; <<http://spraakbanken.gu.se/eng/saldo/>>) is a full-scale Swedish lexical-semantic resource with non-classical, associative relations among word and multiword senses. The senses in SALDO are identified by carefully designed persistent formal identifiers, and for this reason, SALDO has become the “pivot” resource of all our computational lexicon activities, including Swesaurus (Borin et al. 2010; <<http://spraakbanken.gu.se/eng/swefn/>>). We will say something about the relationship of SALDO’s associative relations to classical lexical-semantic relations and how we intend to incorporate the SALDO relations in Swesaurus.

Synlex (Kann & Rosell 2006) is a graded Swedish synonym list created by asking members of the public – users of an online Swedish-English dictionary – to judge the degree of synonymy of a random, automatically generated synonym pair candidate.

We will describe our experiments turning the graded synonymy relations of Synlex into fuzzy synsets in Swesaurus. The introduction of fuzziness into a wordnet raises many intricate methodological and theoretical questions, e.g., if  $w_1$  is a graded synonym of  $w_2$ , and  $w_1$  is a hyponym of  $w_3$ , what is the relation between  $w_2$  and  $w_3$ ? Similarly, if  $w_a$  is a synonym of degree 0.75 of  $w_b$  and  $w_b$  is a synonym of degree 0.9 of  $w_c$ , what – if any – is the degree of synonymy between  $w_a$  and  $w_c$ ?

If synonymy is seen as an all-or-none transitive and symmetric relation, the construction of synsets from synonym pairs is arguably straightforward: We can simply compute the transitive closure of the synonymy relation.

When graded synonymy enters the picture, which method to use for collecting synonym pairs into synsets becomes much less obvious, and especially how to assign a degree of synonymy to “derived” pairs, i.e., pairs not in the original list. We have experimented with two different approaches for turning Synlex synonym pairs into fuzzy synsets in Swesaurus, transitive closure and clique formation. We will present the outcomes of these experiments and discuss the merits and disadvantages of the two methods.

### Bibliography

- Borin, Lars and Markus Forsberg 2009. All in the family: A comparison of SALDO and WordNet. Proceedings of the Nodalida 2009 Workshop on WordNets and other Lexical Semantic Resources – between Lexical Semantics, Lexicography, Terminology and Formal Ontologies. Odense: NEALT. 7–12.
- Borin, Lars, Markus Forsberg and Lennart Lönngrén 2007. The hunting of the BLARK – SALDO, a freely available lexical database for Swedish language technology. Resourceful language technology. Festschrift in honor of Anna Sågvald Hein, Joakim Nivre, Mats Dahllöf and Beáta Megyesi (eds). Acta Universitatis Upsaliensis: Studia Linguistica Upsaliensia 7. 21–32.
- Borin, Lars, Dana Danélls, Markus Forsberg, Dimitrios Kokkinakis, Maria Toporowska Gronostaj 2010. The past meets the present in Swedish FrameNet++. Proc. of Euralex 2010.

Kann, Viggo and Magnus Rosell 2006. Free construction of a free Swedish dictionary of synonyms.  
Proc. of the 15th NODALIDA. Dept. of Linguistics, University of Joensuu. 105–110.

## **Selfish Words: what can words in competition tell us about language change?**

As speakers, we have a number of word choices at our disposal when it comes to expressing a given meaning, for example, in English, when thinking of the concept of “two”, one can refer to this by saying “pair”, “twin”, “two”, “couple” and so on, yet, most often we chose “two” over all other lexical items. What governs the choices we make and how can we characterise this linguistic behaviour? Using techniques devised for describing population genetics and evolutionary biology, our results show that like genes, words come under pressure (that is, biased in some way), rather than being used neutrally under drift.

In the paper, we explore lexical variation in two corpora of Spoken American English (namely from LAMSAS, Kretzschmar et al 1993, and LAGS, Pederson et al 1986) using statistical tools to capture word-use within populations of speakers. The data consist of spoken records characterising the use of words for everyday items, concepts or events such as “wife”, “chair”, “thunderstorm”, “driven”, or “three”. The participants were asked identical sets of questions like “what do you call the place where you store your clothes?” thereby aiming to elicit responses such as “closet”, “wardrobe” and other variants. Here we report on the frequencies of different words which participants select to refer to the same items, for example, “attic”, “loft”, “garret”, and so on. Our interest, located within a cognitive approach to language and cultural evolution is in whether the choice of word is proportional to how often each word is used among the population of speakers (termed here “neutral use”), or whether word use is biased in favour of the most frequently used words (“conformist use”).

Using Ewens Sampling Formula (1972), we find that the majority of these word-use frequency distributions deviate significantly from what would be expected of “neutral use”. In all of these cases one or a small number of variants (e.g., “attic”) dominates the distribution, being used far more often than expected by chance given the availability of other forms (such as “garret” or “loft”). In the remaining cases, the different word-variants (for instance, “fire-grates” or “andirons”) are used in frequencies expected under our neutral-use scenario and in no case were different words employed in equal or even approximately equal frequencies.

The current work explores the concept of “synonymy” from a usage-driven perspective, where words which occupy (at least) some overlapping semantic space (variants such as “attic” and “loft”) can be understood to compete against each other for the attention of speakers in a given population (much like viruses or other biological phenomena, Mufwene 2002). We show how this approach can sharpen our understanding of not just how language-use evolves within communities but why some variants come to dominate.

### Bibliography

- Ewens, W. (1972). The sampling theory of selectively neutral alleles. *Theoretical Population Biology* 3: 87–112.
- Kretzschmar, W., McDavid, V., Lerud, T.K. and Johnson, E. (ed.) (1993). *Handbook of the Linguistic Atlas of the Middle and South Atlantic States*. Chicago: The University of Chicago Press.
- Mufwene, S. (2008). *Language Evolution: Contact, Competition and Change*. Continuum.
- Pederson, Lee, Susan Leas, Guy Bailey, and Marvin Bassett. (1986). *The Linguistic Atlas of the Gulf States*. Vol. 1, Handbook. Athens: Univ. of Georgia Press.

## ***We ought to re-think synonymy, shouldn't we?*** **A corpus-based reassessment of two English modals' substitutability**

The English modals *should* and *ought to* “seem to be largely interchangeable” (Palmer 1990; cf. similar sentiments in Coates 1983: 69; Quirk et al. 1985: 227; Huddleston and Pullum 2002: 186). Some proposals for possible distinctions have nonetheless been made, based *inter alia* on (i) subjectivity vs. objectivity (Swan 1980: 550; Gailor 1983: 348; Aarts and Wekker 1987: 193; Myhill 1996; Collins 2009: 54) (ii) absence vs. presence of an implication of non-fulfillment (Close 1981: 121; Gailor 1983: 348-349; Westney 1996: 170) and (iii) relative frequency vs. infrequency of epistemic reading (Coates 1983; Palmer 1987: 134; Collins 2009).

To test whether these and a few other claims made in the literature could be verified, we conducted a corpus-based study involving over 1000 sentences with *should* or *ought to* in contemporary spoken and written British English and coded them for 28 parameters. Raw frequency and frequency per discourse mode could not be coded, but the data confirmed previous observations that *should* is used much more frequently than *ought to* and that the frequency of *ought to* in spoken language is higher than its frequency in written language. To find out which of the parameters investigated contribute significantly to the distribution of *should* and *ought to*, we fitted a logistic regression model, making sure any corruption of the model caused by multicollinearity was excluded.

Eight parameters were shown to exert a unique and significant impact on the choice of *should* vs. *ought to*, in order of decreasing strength: (i) inversion, (ii) a following contracted perfect infinitive, (iii) no adverb or a following rather than preceding adverb, (iv) verb-marked negation, (v) embedding by *suggest* or a similar item, (vi) reference to the non-past, (vii) no embedding by *think* or a similar cognition expression and (viii) third (vs. first or second) person subject. All other factors turned out not to be significant or were left out of the model because they correlated with one or more parameters in the model.

Our model thus provides a parsimonious and accurate description of the *should/ought to* variation. It also guides us towards a possible explanation, which we believe can be found in the different grammatical status of *should* and *ought to*, the former being a true modal auxiliary, the latter being a semi-auxiliary (a blend between a modal and a lexical verb). We argue that this difference can account for the four strongest distinctions in usage we revealed.

We further discuss the results of a subsequent distinctive collexeme analysis (Stefanowitsch and

Gries 2003) supporting the claim made above that *should* (compared to *ought to*) occurs relatively more frequently with situations which are likely to be carried out than with unlikely situations (cp., e.g., *It should be noted that...* vs. *He ought to be crucified*) but rejecting the above claim that *should* is more subjective (e.g. *Abstracts should / ??ought to be submitted by April 16, 2010*). Finally, we reflect on the implications of our study for the (non)existence of free variation in language.

## Bibliography

- Aarts, Flor and Herman Wekker. 1987. *A Contrastive Grammar of English and Dutch – Contrastieve Grammatica Engels/Nederlands*. Leiden: Martinus Nijhoff.
- Close, R. A. (1983) *English as a Foreign Language: Its Constant Grammatical Problems*, 3<sup>rd</sup> ed., London: George Allen & Unwin.
- Coates, Jennifer (1983) *The Semantics of the Modal Auxiliaries*. London and Canberra: Croom Helm.
- Collins, Peter (2009) *Modals and Quasi-Modals in English*. Amsterdam and New York: Rodopi.
- Gailor, Denis (1983) "Reflections on *Should*, *Ought to*, and *Must*," *English Language Teaching Journal* 37: 346-349.
- Huddleston, Rodney and Geoffrey K. Pullum (2002) *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Myhill, John (1996) "*Should* and *Ought*: The Rise of Individually Oriented Modality in American English." *English Language and Linguistics* 1: 3-23.
- Palmer, Frank Robert (1987) *The English Verb*, 2<sup>nd</sup> ed. London and New York: Longman.
- Palmer, Frank Robert (1990) *Modality and the English Modals*, 2<sup>nd</sup> ed. London and New York: Longman.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik (1985), *A Comprehensive Grammar of the English language*. London: Longman.
- Stefanowitsch, Anatol and Stefan Th. Gries. 2003. Collocations: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics* 8: 209-243.
- Swan, Michael (1980) *Practical English Usage*. Oxford: Oxford University Press.
- Westney Paul (1995) *Modals and Periphrastics in English: An Investigation into the Semantic Correspondence between Certain English Modal Verbs and Their Periphrastic Equivalents*. Tübingen: Niemeyer.

## **Synonymy in an approach to combined distributional and compositional semantics**

We outline a perspective on synonymy from a novel approach to combining distributional and compositional semantics, Lexicalized Compositionality (LC). In LC logical forms, the standard model-theoretic lexical predicates (e.g., *cat'*) are replaced with distributional concepts (notated as e.g., *cat*<sup>o</sup>). *cat*<sup>o</sup> corresponds to the set of all linguistic contexts (expressed as logical forms) in which the lexeme *cat* occurs. We consider hypothetical contexts, which are based on all the possible statements true in some world (allowing a comparison to model-theoretic approaches), and actual contexts, corresponding to an individual's linguistic experience.

The aim of LC is to investigate models of semantics which are psycholinguistically plausible in terms of linguistic information. Some utterances are directly grounded. In such cases, contexts can be paired with perceptual data, and linguistic concepts, such as *cat*<sup>o</sup>, related directly to real world entities. However utterances can be understood which are not immediately grounded at all or only partially grounded. We assume that speakers' knowledge of word meaning is primarily acquired implicitly, from actual contexts (grounded or ungrounded), but that there is also a more minor role for explicit 'expert' knowledge. Individual speakers will have somewhat different models of word meaning, and understanding may thus require negotiation.

Within this approach, the notion of synonymy is multifaceted. Classic synonymy corresponds to the situation where a hearer is (more or less) directly informed that a term means the same thing as one they are already familiar with. For example: The aubergine (eggplant) has to be one of my favourite vegetables. <http://www.earth.li/~kake/cookery/aubergine.html> (last accessed 10th August 2010)

The much commoner case of near-synonymy (Edmonds and Hirst, 2002) corresponds to similarities between context sets. We hypothesise that the process of understanding a new word involves an initial approximation to the nearest known linguistic concepts(s) (by context set comparison), followed by the acquisition of differentiating information. We define a notion of a characteristic context in terms of the differences between two context sets. For instance, a hearer to whom *rancid* is a relatively unfamiliar term may initially relate it to *off/bad/rotten* etc, but then additionally differentiate it using further contexts (perhaps by its preferential application to fatty foodstuffs compared to *bad*). Alternatively, or additionally, an 'expert' definition may be known. In normal communication, the coarse-grained notion of synonymy (i.e., that *rancid* relates to *off* etc) usually suffices for understanding.

We should emphasize the differences between our approach and the usual computational models of distributional similarity (Harper, 1965, onwards). Clearly our contexts are much more structured: the recent trend is towards higher structure (e.g., Erk and Padó (2008)), but our work is still distinct in aiming for a logical account that incorporates distributions, thus allowing for (relatively conventional) logical inference. Furthermore, in our account, the contexts come from a specific individual's experience. We expect this to give very different results from the standard approaches which utilise data from very heterogeneous sources.

Our work is at a very early stage and has primarily involved development of the theoretical account. Given the current lack of individualised corpora, empirical work involves approximations of linguistic contexts. However, we believe these ideas suggest new theoretical perspectives for computational experiments on synonymy.

## Bibliography

- Edmonds, Philip and Graeme Hirst (2002) 'Near-synonymy and lexical choice', *Computational Linguistics*, 28(2), 105–144.
- Erk, Katrin and Sebastian Padó (2008) 'A Structured Vector Space Model for Word Meaning in Context', *Proceedings of the EMNLP*, Honolulu, HI.
- Harper, K.E. (1965) *Measurement of similarity between nouns*, The RAND Corporation, Santa Monica.

## A corpus-based analysis of quasi-synonymous adjectives in German

The aim of this paper is to prove to what extent corpus-based statistical data is helpful for semantic and pragmatic discrimination of quasi-synonymous adjectives, even if these adjectives are represented in dictionaries as meaning absolutely the same. The analysis of combinatorial profiles of a group of quasi-synonymous German adjectives (namely *ausgezeichnet*, *hervorragend*, *exzellent*, *herrlich*, *vortrefflich* and *vorzüglich*) allows us to objectivate both the semantic structure of nomina modified by a given adjective and the discourse domain. Since these two parameters govern the use of every adjective in question such an analysis can provide an appropriate basis for the relevant semantic and pragmatic discrimination.

In principle, there are three ways to explain relevant specifics of the combinatorial profiles of these adjectives.

(a) There are combinatorial differences which go back to the semantic class of the modified nouns. In this case, the semantic structure of the adjective itself displays features that predict combinatorial restrictions and preferences, i.e. the specifics of co-occurrence are semantic in nature.

(b) There are differences which go back to given discourse domains. In this case, the specific properties of the combinatorial behaviour of every adjective go back to their preference for certain discourse types or registers as well as for certain thematic fields, rather than to the meaning of the adjective itself. So, there are words that are avoided in spoken colloquial language, but are frequently used in the feuilleton of big magazines. In the everyday language people talk about weather differently than in the weather report. On the other hand, the choice of an adjective from the given group of quasi-synonyms can be influenced by the subject area discussed in the text. For instance, people talk about sports in different terms than about antique arts or about food. It means that besides the discourse type and register also the thematic domain plays a role while choosing a fitting word. Often it is not possible to separate both phenomena from each other (a reportage on sport events is, e.g., a thematically bound discourse type). Therefore, we consider both phenomena within the same category, namely discourse domain. The meaning explanations of the adjectives in question do not display any differentiative semantic features. They can be discriminated from each other only by an additional comment, such as “preferably in discourse domain X”.

c) There are also combinatorial peculiarities which cannot be explained systematically, i.e. the corpus analysis clearly shows that a given adjective may have combinatorial preferences that go back neither to its semantic class nor to a certain discourse domain. Here we are dealing with individual combinatorial properties that have to be fixed as lists. What is extremely important is that the list-representation makes sense only if all other possibilities do not work.

Which of the three explanation ways (a), (b) or (c) fits the data in every concrete case can be decided only on the basis of corpus analysis. We used here the platform CCDB Mannheim (© C. Belica) and the tools of the DWDS-Wortprofil (Berlin-Brandenburg Academy of Sciences).

The first phase of our research (Dobrovol'skij 2004; Dalmas/Dobrovol'skij in print) has revealed significant combinatorial differences among the given lexemes, which can partly be explained via properties of their inner form fixed in their morpheme structures. Conceptual features of this kind have to be included in the semantic explanation.

## Bibliography

- Biber, Douglas/ Conrad, Susan/ Reppen, Randi (1998): *Corpus linguistics: investigating language structure and use*. Cambridge: Cambridge University Press.
- Dalmas, Martine/ Dobrovol'skij, Dmitrij (in print): Quasisynonymie bei Adjektiven: hervorragend und Co (eine corpusgestützte Untersuchung). In: Schmale, Günter (Hg.): *Das Adjektiv. Was wird wo, wie und wozu „dazugeworfen“?*. Tübingen: Stauffenburg.
- Dobrovol'skij, Dmitrij (2004): Corpusbasierte kontrastive Lexikologie: semantische und kombinatorische Aspekte. In: *Mehrsprachige Individuen – vielsprachige Gesellschaften*. Abstracts zur 35. Jahrestagung der Gesellschaft für Angewandte Linguistik in Wuppertal. Wuppertal: Bergische Universität Wuppertal. 33-34.
- Hanks, Patrick (2008): Lexical patterns: from Hornby to Huston and beyond. In: *Proceedings of the XIII EURALEX International Congress*. Barcelona: IULA, 2008. 89-129.
- Jones, Karen Sparck (1986): *Synonymy and semantic classification*. Edinburgh: Edinburgh University Press.
- НОСС – *Новый объяснительный словарь синонимов русского языка*. 2-е изд., испр. и доп. / Под рук. Ю.Д. Апресяна. Москва/Вена, 2004.

## **Semantic similarity in non-native English: the case of *may* and *can* in French-English interlanguage**

This work presents one possible empirical route to trace patterns of semantic behavior in a corpus of French learner English. I investigate to what extent the expression of the semantic domain of POSSIBILITY, as conveyed by *may* and *can*, differs in native English and French-English interlanguage. I will specifically discuss (i) how English native speakers use *may* and *can* in a more data-driven way than is usually adopted in studies on modality, and (ii) how the uses of *can* and *may* by French learners differs from that of native speakers and what these distributional differences suggest as to what motivates the different patterns.

At the intersection between corpus linguistics and cognitive semantics, Gries & Divjak (2009) and Gries & Otani (2010) argue in favor of a Behavior Profile (BP) approach to semantic analysis, which allows to explore how meanings and functions of lexical and syntactic elements are correlated with the distribution(s) of formal elements within their contexts. However, with only one exception, applications have focused on polysemy, synonymy, and antonymy to native-speaker data from one language only or, in the case of Divjak & Gries (2009), the study of phasal verbs in English and Russian.

The current work presents an application of the BP approach to the domain of interlanguage. I investigated 3700 instances of *may* and *can* in English and French-English interlanguage as well as French *pouvoir* in three different corpora: the French sub-section of the International Corpus of Learner English (ICLE), the Louvain Corpus of Native English Essays (LOCNESS) and the Corpus de Dissertations Françaises (CODIF). I annotated these instances for 22 semantic and morpho-syntactic variables, including a total of 98 features and resulting on tens of thousands of data points. The statistical treatment of the annotated data includes both a monofactorial analysis that involves assessing the behavior of *may* and *can* in relation to individual semantic and morpho-syntactic independent variables; and a multifactorial analysis which involves the statistical modelling of possible interactions between independent variables and their effects on *may* and *can*.

The results of the monofactorial assessment confirm the relevance of a contextually-grounded approach. Linguistic components such as the senses of *may* and *can*, the type of clauses they occur in, the type of lexical verbs they occur with or the type of (in)animacy of their subject referents have all been identified as highly influencing components within the corpus (each individual component showing a *p-value* smaller than 0.001). However, results also indicate that the frequencies of *may* and *can* differ considerably across the corpora (*chi-square*= 3716.93,  $p < 0.001$ , *Cramer's V*=0.71), thus suggesting that the identified semantic and morpho-syntactic features yield different interaction patterns in each individual sub-corpus. I am currently carrying out a multifactorial analysis to investigate behavioral patterns characteristic of *may* and *can* both in L2 and IL.

This work shows that, in addition to providing a data-driven methodology for linguistic patterns and their semantics in one language, the BP approach also provides a useful empirically-grounded way to study characteristics of interlanguage varieties.

### Bibliography

Gries, Stefan Th. and Dagmar S. Divjak. 2009. Behavioral profiles: a corpus-based approach to cognitive semantic analysis. In *New directions in cognitive linguistics*, ed. by Vyvyan Evans and Stephanie S. Pourcel, 57-75. Amsterdam, Philadelphia: John Benjamins.

Gries, Stefan Th. & Naoki Otani. 2010. Behavioral profiles: a corpus-based perspective on synonymy and antonymy. *ICAME Journal*.

## Two times ‘go’ in Shangaci: semantics or pragmatics?

Shangaci, a Bantu language spoken in Mozambique, has two seemingly synonymous verbs to express ‘go’: *-entta* and *-lawá*. In the question below the two ‘go’-verbs are perfectly interchangeable.

- (1) a. o-tt-étt’      ó-nlímpu  
 2SG-PRES-go 17-3.well  
 ‘are you going to the well?’
- b. o-ttí-láw      ó-nlímpu  
 2SG-PRES-go 17-3.well  
 ‘are you going to the well?’

However, the answers to the above questions suggest that the verbs differ in meaning and/or use.

- (2) a. waalá      ki-tt-étt’      ó-muúti  
 nor      1SG-PRES-go 17-3.town  
 ‘no, I am going home’
- b. kha-n-cúwí      nkháma      ki-ttí-laáwa  
 NEG-1PL-know      if      1SG-PRES-go  
 ‘I do not know whether I am going’

Moreover, *-entta* and *-lawá* follow very different grammaticalization paths. The verb *-entta* is used as an auxiliary to express predication focus (3) (Devos and van der Wal 2010), whereas the auxiliary *-lawá* is used for discourse deixis (4) (Bourdin 2008) indicating that an action takes place at some distance from the deictic centre. As persistence of the original lexical meaning is typical in grammaticalization (Hopper 1991), the different paths followed by *-entta* and *-lawá* suggest a difference in lexical meaning.

- (3) yoómbw’etó      kha-y-iital-éeni      y-étt-é      ó-khízeéy-a  
 7.pot 7.DEM      NEG-7-be.full-PERF      7-GO-PERF      15-be.half-full-INF  
 ‘that pot is not full, it is half full’
- (4) a-tthíir-i      m-muuti’      phúule      a-láa      a-cí-víith-i  
 2-run-PERF      18-3.town      18.DEM      2-GO      2-REF-hide-PERF  
 ‘They fled from that town, and hid themselves’

The paper aims at defining the nature of the difference between the two verbs (pragmatic or semantic) and its implication for their identification as synonymous or merely similar forms. A related theoretical question is whether lexemes can be considered synonymous when they are semantically identical but pragmatically different, or reformulated: should synonyms be defined in purely semantic terms?

Botne (2005) identifies a similar pair of ‘go’-verbs in the Bantu language Chindali. According to him the difference is of a purely pragmatic nature: both verbs express going somewhere but in one case the goal is salient, whereas in the other case the motion is salient. The Shangaci verbs could be said to involve the same pragmatic difference. In (2a) the goal is contrastively focused whereas in (2b) the going itself is questioned. Moreover, the grammaticalization of *-entta* as a marker of predication focus involves strengthening of the salience of goal. If the difference is solely pragmatic in nature, *-entta* and *-lawá* could be said to be synonymous forms sharing the same image-schematic structure.

However, the difference in use of *-entta* and *-lawá* could also be a consequence of a more basic semantic distinction. As argued by Wilkins and Hill (1995) not all ‘go’- verbs are deictic and Shangaci could be said to have a deictic (i.e., *-lawá*) as well as a non-deictic (i.e., *-entta*) ‘go’-verb. In the grammaticalization of *-lawá* as a marker of discourse deixis, it is the deictic component of its meaning (away from deictic centre) that is strengthened. Such a semantic difference would imply that *-entta* and *-lawá* are similar but not synonymous forms.

#### Abbreviations

DEM	demonstrative
INF	infinitive
NEG	negative
PERF	perfective
PL	plural
PRES	present
REF	reflexive
SG	singular

#### Bibliography

- Botne, Robert. 2005. “Cognitive schemas and motion verbs: COMING and GOING in Chindali (Eastern Bantu)”. *Cognitive Linguistics* 16(1).43–80.
- Bourdin, Philippe. 2008. “On the grammaticalization of ‘come’ and ‘go’ into markers of textual connectivity”. In *Rethinking Grammaticalization: New Perspectives*, López-Couso, Maria José and Elena Seoane (eds.). 37-59. Amsterdam/Philadelphia: John Benjamins.
- Devos, Maud and Jenneke van der Wal. “‘Go’ on a rare grammaticalization path to focus”. TiNdag 2010, Utrecht.
- Hopper, Paul J. 1991. “On some principles of Grammaticalization”. In *Approaches to Grammaticalization, vol. 1*. Traugott, Elizabeth Closs and Bernd Heine (eds.). 17-35. Amsterdam: Benjamins.
- Wilkins, David P. and Deborah Hill. 1995. “When “go” means “come”: Questioning the basicness of basic motion verbs”. *Cognitive Linguistics* 6(2/3). 209-259.

## To what extent is synonymy a good indicator for verb complementation? Challenging semantics-based accounts

This paper will discuss the relationship between **verb meaning** and **verb complementation** on the basis of a **corpus-based** comparison of **semantically similar verbs** (Faulhaber forthcoming). The question addressed is whether syntactic information can be deduced in a regular way from the meaning of verbs or has to be regarded as a matter of **storage**.

It is undoubtedly the case that meaning plays a role – and most likely even an important one – in how activities or situations are verbalized, i.e. with which formal complement types a verb can be combined or, alternatively, in which valency patterns it can occur. This is reflected in statements such as that “[k]nowing the meaning of a verb can be a key to knowing its behavior” (Levin 1993: 5), that there is “a principled interaction between the meaning of a word and its grammatical properties” (Dixon 1991: 6) and that “[s]emantically similar verbs show a strong tendency to appear in the same argument structure constructions” (Goldberg 2006: 58), which can be found in the context of various linguistic theoretical frameworks. Evidence provided for this assumption ranges from hand-picked examples to broad comparisons of verbs and their complementation patterns such as Levin’s *English Verb Classes and Alternations* (1993) or Hunston and Francis’s *Pattern Grammar* (2000), which aim to show that verbs which exhibit similar grammatical behavior – in the sense that they allow the same diathesis alternations (Levin 1993) or can occur in the same complementation patterns (Hunston and Francis 2000) – are also semantically similar. By implication, it could be (and has been) construed that synonymous or semantically similar verbs are also syntactically alike.

However, this assumption is clearly not without problems. For once, the conclusion that semantic similarity leads to syntactic similarity cannot be based on studies which take syntactic similarity as a starting point since differing uses are excluded *ex ante* (cf. Levin 1993 or Hunston and Francis 2000). Moreover, counter-evidence is not as “rare” as is often assumed, which becomes evident when large amounts of actual language data are taken into consideration. A comparison of more than twenty groups of semantically similar English verbs (all in all more than eighty verbs) as regards their complementation patterns – based on the corpus-based *Valency Dictionary of English* (Herbst, Heath, Roe, and Götz 2004) and further corpus research – clearly reveals that despite clear structural similarities a semantic determinist view of verb complementation is highly problematic: The amount of idiosyncratic restrictions, i.e. restrictions which can neither be accounted for on the basis of verb meaning (e.g. the verbs’ participants, lexical aspect, selection restrictions etc.) nor of the meaning of the complement types or patterns, is too big to be dismissed as peripheral (Faulhaber forthcoming). These results clearly emphasize the role of **conventionalization** in verb complementation, providing a strong case for rethinking a deterministic role of semantics in verb complementation and for strengthening the role of storage instead.

### Bibliography

- Dixon, Robert M.W. 1991. *A New Approach to English Grammar, on Semantic Principles*. Oxford: Clarendon Press.
- Faulhaber, Susen. forthcoming. *Verb valency patterns – Challenging semantics-based accounts*. Berlin/New York: Mouton de Gruyter.
- Goldberg, Adele. 2006. *Constructions at Work. The Nature of Generalization in Language*. New York: Oxford University Press.

- Herbst, Thomas, David Heath, Ian Roe and Dieter Götz. 2004. *A Valency Dictionary of English*. Berlin/New York: Mouton de Gruyter.
- Hunston, Susan and Gill Francis. 2000. *Pattern Grammar. A corpus-driven approach to the lexical grammar of English*. Amsterdam: Benjamins.
- Levin, Beth. 1993. *English Verb Classes and Alternations – A Preliminary Investigation*. Chicago: The University of Chicago Press.

## **From binary synonymy to near synonymy by optimal proxemy of lexical resources**

Dictionaries of synonyms are often encoded as binary links between words. This encoding raises many issues. For example, we show that there is very little agreement between different expert-built resources, whereas they represent a similar linguistic reality. This may be due to the fact that absolute synonyms are rare and that most of the synonymy described in the dictionaries is in fact a description of nuances of a general meaning shared by a set of near-synonyms (Edmonds and Hirst, 2002). The complexity and scale of these nuances leave room for interpretation when projecting them on the basic “synonym/not synonym” alternative, which in turn gives rise to discrepancies between electronic lexical resources at the most fine-grained level. However, we claim that, at a coarser level, the patterns drawn by sets of near-synonyms should be mostly independent from the resource used to describe the lexicon of a given language. We propose a model that leverages the binary encoding of resources to represent the synonymy structure at various levels of granularity. This is not a new merging method, but an attempt at a better modelling of the notion of synonymy. Based on this model, we find an optimum level of granularity for which resources are the most similar, therefore our approach seems appropriate in order to infer near-synonymy patterns from any particular binary encoded resource.

We studied the similarities of seven well-known french dictionaries, binarily encoded as graphs in which vertices are words and two vertices are linked by an edge if they are considered synonyms. We used the F-score between sets of edges to measure the similarities between the unweighted graphencoded dictionaries of synonyms and found that it does not exceed 0.5. To model these resources on coarser levels, we introduce a notion of proxemy between any two words, based on the probability of reaching one word from another after a random exploration of the graph. At each step of the exploration, a particle is modelled to move from one word to any of its neighbours with a probability inversely proportional to its degree. Each word is thus associated to a proxemy vector characterising its semantic similarity with the other words. This vector depends on the starting vertex and on the number of exploration steps. When the number of steps tends to infinity, all words have the same proxemy vector, which describes the importance of each word in the graph, at the highest level of generality (Gaume et al., 2010). Conversely, after only one step, proxemy vectors have the lowest level of generality, taking only the neighbours of each node into account. The distance between two resources is then defined as the mean euclidian distance between the proxemy vectors of the same word in the two graphs. We experimentally found an optimal number of proxemy steps for which the similarity of any two resources is maximal. This is an optimal level of generality, according to the proxemy model, that enables us to discover near-synonymy patterns that are similar across dictionaries whose binary synonymy patterns are very different.

### Bibliography

- P. Edmonds and G. Hirst (2002) Near-synonymy and lexical choice, *Computational Linguistics*, 28(2), pp. 105-144.
- B. Gaume, F. Mathieu and E. Navarro (2010) Building Real-World Complex Networks by Wandering on Random Graphs, *Information - Interaction – Intelligence*, (To appear).

## **Usage-Based Cognitive Models. Synonymy as an operationalisation of conceptual structure**

Synonymy could be the key to operationalising the study of culturally determined concepts. As such, this paper aims at developing a method for the corpus-driven analysis of cognitive models. It focuses on the concept of LIBERTY in 20C America. The description of linguistically encoded culturally rich concepts is a basic enterprise of Cognitive Linguistics (Wierzbicka 1985, Lakoff 1987, 1995). Despite the insights of such research, its method of analysis is notoriously weak - conceptual structures are proposed, based on sets of contextless expressions. Firstly, this method offers no means for the verification of results. Secondly, it offers no information about the how the expressions, and therefore the concepts, are used. The actual structure of the concept, relative to a speech community, is entirely unknown. These "Idealised" Cognitive Models are devoid of the complexity of social reality. A usage-based approach to concepts must resolve these two shortcomings. Drawing on Divjak (2006), Janda & Solovyev 2009, and Glynn's (2010) approach to lexical synonymy, the study operationalises the problem of conceptual description by examining the use of the lexemes associated with a given concept.

The analysis is based upon two lexemes, examples of which are extracted from Time magazine - 250 occurrences of liberty and 250 of freedom (25 randomly selected occurrences from the first three years of each decade). Time magazine is chosen because it represents mainstream readership in American culture. The diachronic dimension of the study will help capture the social variation that is inherent in a Usage-Based Cognitive Model. Adopting the type of feature analysis developed in Gries (2006) and Glynn (2009), the examples are annotated for a range of formal, semantic, and sociolinguistic phenomena. Important factors include the grammatical construction associated with the lexeme, its source domain in figurative uses, the kind of patient affected by LIBERTY, its agentivity, as well as the topic of discourse.

Multiple Correspondence Analysis helps identify how the conceptual structure varies. During the 20s, 30, and 40s, liberty is often used when discussing nations, this shifts during the 50s and 60s to discussions about society, and then this shifts again during the 70s - 2000s to discussions of individual liberty. The metaphors associated with the lexeme shift in a similar pattern beginning with 'liberty' being conceptualised as a simple 'object' then toward the end of the century being conceptualised as a 'struggle' or as a 'place'. Similar patterns are observed for the use of the lexeme freedom. An analysis of variance demonstrates that the most significant and important variation over the 20C is the kind of experiencer to which the lexemes refer. This runs parallel to the results of the Correspondence Analysis. A Regression Analysis focuses on the importance of the experiencer in the conceptualisation of LIBERTY and confirms that there is a statistically significant shift from abstract non-human experiencers of LIBERTY to concrete and individual experiencers. Could this represent a shift towards a more individualist conceptualisation or is it a merely a result of world politics and the political emphasis of the magazine? A comparative study, based on a different genre needs to be performed to answer this question. The study demonstrates the importance of the method for the study of cognitive models, both because it permits result verification and because it adds a usage-based dimension to those results.

### Bibliography

Divjak, D. 2006. Delineating and Structuring Near-Synonyms. *Corpora in cognitive linguistics*, St. Th. Gries & A. Stefanowitsch (eds), 19-56. Berlin: Mouton.

- Divjak, D. 2010. *Structuring the Lexicon: a Clustered Model for Near-Synonymy*. Berlin: Mouton.
- Glynn, D. & Fischer, K. (eds). 2010. *Corpus-Driven Cognitive Semantics. Usage-Based approaches to conceptual structure*. Berlin: Mouton.
- Glynn, D. & Robinson, J. (eds). In press. *Polysemy and Synonymy. Corpus methods and applications in Cognitive Linguistics*. Amsterdam: Benjamins.
- Glynn, D. 2010. Synonymy, Fields, and Frames. Developing usage-based methodology for Cognitive Semantics. *Cognitive Foundations of Linguistic Usage Patterns*, H.J. Schmid & S. Handl (eds), 89-118. Berlin: Mouton
- Gries, St. & Stefanowitsch, A. (eds). 2006. *Corpora in Cognitive Linguistics. Corpus-based Approaches to Syntax and Lexis*. Berlin: Mouton.
- Gries, St. Th. 2006. Corpus-based methods and cognitive semantics: the many meanings of to run. *Corpora in Cognitive Linguistics*. St. Gries & A. Stefanowitsch (eds), 57-99. Berlin: Mouton.
- Janda, L. & Solovyev, V. 2009. What Constructional Profiles Reveal About Synonymy: A Case Study of Russian Words for SADNESS and HAPPINESS. *Cognitive Linguistics* 20: 367-393.
- Lakoff, G. 1987. *Women, Fire, and Dangerous Things. What categories reveal about the mind*. London: UCP.
- Lakoff, G. 1995. *Moral Politics: What Conservatives Know That Liberal Don't*. London: UCP.
- Stefanowitsch, A. & Gries, St. (eds). 2006 *Corpus-based Approaches to Metaphor and Metonymy*. Berlin: Mouton.
- Wierzbicka, A. 1985. *Lexicography and Conceptual Analysis*. Ann Arbor: Karoma.

## **The Synonymy of Morphological Semantics. A usage-based study of long and short adjectives in Russian**

Synonymy is not unique to lexical semantics. From a Cognitive Linguistics perspective, where there is no theoretical distinction between lexis and syntax and all form is semantically motivated, syntactic and morphological alternations are, in fact, the study of (near) synonymy. The corpus-driven study of both synonymous syntax and lexis has a strong tradition in Cognitive Linguistics (Divjak 2006, Arppe 2008, Janda & Solovyev 2009, Glynn 2010, Heylen, 2005, Grondelaers et al. 2007, Divjak 2009, Speelman & al. 2010, Glynn & Robinson forthc.). However, Morphological semantics present certain hurdles that are not present in lexical and syntactic research. This study advances the state of the art by overcoming these hurdles and extending the corpus-driven Cognitive research to morphology proper in an analysis of long and short forms of predicative adjectives in Russian. The traditional grammars contrast the two forms stating that short adjectives profile inherent characteristics and are more typical of formal registers where the long form of adjectives profile more transient characteristics and tend to be more informal in use.

The study of morphological semantics presents two problems to the cognitive corpus linguist. Firstly, it is harder isolate morphological meaning from lexical meaning than it is for syntactic semantics. The large lexical variation typical of syntactic patterns makes constructional semantics more straightforward to isolate. This study resolves this problem by including a lexical semantic analysis and isolates the grammatical semantics by contrasting lexical - constructional pairings. In effect, the grammatical semantics are identified by determining what remains unchanged when we change the lexeme. Secondly, syntactic patterns often come in minimal pairs, or alternations, for which es-tablished statistical techniques (Logistic Regression Analysis) have been developed. These do not exist for looking at paradigms of several morphological variants. This study employs Multiple Correspondence Analysis, an exploratory multivariate technique that indentifies patterns of associations in the data even when there are more than two forms under investigation.

The data are taken from literature, news press, and online personal diaries. A total of 500 occurrences of the lexemes for happy and sad, in three forms - the short, the nominative long, and the instrumental long form, are manually coded. The analysis covers semantic, extralinguistic, and formal features. Grouping the nominative and instrumental forms of the adjectives together and contrasting them with the short forms results in a binary alternation, permits a Logistic Regression Analysis. The results of this analysis broadly confirm the hypothesis that there exists a difference in degree of inherentness of characteristics expressed by the two forms. However, there are clearly other variables affecting the distinction. Multiple Correspondence Analysis is used to examine the interaction of the three forms (nominative and instrumental sperately). The analyses reveal that for one of the two lexemes, the instrumental long form, behaves more like a short form of the adjective. It would seem, therefore that the traditional position is confirmed but with a qualification – instrumentals can, at least for certain lexemes, profile a more inherent characteristic. Clearly further research is needed, but the usefulness of Correspondence Analysis for the corpus study morphological semantics is demonstrated.

### Bibliography

Arppe, A. 2008. *Univariate, bivariate and multivariate methods in corpus-based lex-icography - a study of synonymy*. Dept. of General Linguistics, University of Helsinki

- Divjak, D. 2006. Delineating and Structuring Near-Synonyms. St. Gries & A. Stefa-nowitsch (eds), *Corpora in Cognitive Linguistics*, 19-56. Berlin: Mouton.
- Divjak, D. 2009. Mapping Between Domains. The Aspect-Modality Interaction in Russian. *Russian Linguistics*, 33: 249-269.
- Glynn, D. & Robinson, J. Forthc. *Polysemy and Synonymy. Corpus methods and applications in Cognitive Linguistics*. Amsterdam: Benjamins.
- Glynn, D. 2010. Synonymy, Lexical Fields, and Grammatical Constructions. A study in usage-based Cognitive Semantics. H.-J. Schmid & S. Handl (eds). *Cognitive Foundations of Linguistic Usage-Patterns*, 89-118. Berlin: Mouton.
- Grondelaers, S., Geeraerts, D. and Speelman, D. 2007. A case for a cognitive corpus Linguistics. M. Gonzalez-Marquez, et al. (eds), *Methods in Cognitive Linguistics*, 149-169. Amsterdam: Benjamins.
- Heylen, K. 2005. A Quantitative Corpus Study of German Word Order Variation. S. Kepser & M. Reis (eds), *Linguistic Evidence: Empirical, Theoretical and Computational Perspectives*, 241-264. Berlin: Mouton.
- Janda, L. & Solovyev, V. 2009. What Constructional Profiles Reveal About Synonymy: A Case Study of Russian Words for sadness and happiness. *Cognitive Linguistics* 20: 367-393.
- Speelman, D. & Geeraerts, D. 2010. Causes for causatives: the case of Dutch doen and laten. *Linguistics of Causality*. T. Sanders & E. Sweetser (eds). Cambridge: CUP.
- Szmrecsanyi, B. 2010. The English genitive alternation in a cognitive sociolinguistics perspective. D. Geeraerts & al. (eds), *Advances in Cognitive Sociolinguistics*, 141-166. Berlin: Mouton.

## Syntactic overplanning in language production: Evidence from speech errors

This paper explores the effect of parallel processing of synonymous syntactic structures based on an analysis of 104 naturally occurring Russian speech errors (slips of the tongue).

A well-known phenomenon resulting from parallel planning in natural language production is a speech error known as the lexical blend (e.g. *flown* + *driven* → *flivven*), when two near-synonyms (usually interpreted as contextual synonyms) are selected instead of one target lexeme. Word blends are normally thought to occur when two lemma nodes are activated to an equal level, and both are selected. Similarly, phrasal blends (e.g. *if you're not careful* + *if you don't watch out* → *if you're not watch out*) are regarded as different formulations of the same message intertwined in speech (Fay 1982; Coppock 2006), which can also be accounted for by the parallel selection of two competing lemma nodes.

While semantic blends are well-known and have been observed by many researchers of speech errors (Cutler, 1982; Fay, 1982; Garrett, 1980; Stemberger, 1990 etc.), little, if anything, has been said of purely *syntactic blends*, i.e. blends of two synonymous syntactic structures with the same set of lemmas.

An analysis of Russian speech error data reveals that such blends are not uncommon in spoken Russian (a highly inflected language, where grammatical features such as case, number etc. almost always surface as bound morphemes) and that, similarly to lemmas, syntactic structures do compete during sentence production, e.g.

- (1) Mozhno *ja* redisk-u *doe-m* + Mozhno *mne* redisk-u *doe-st'* →  
 may I:NOM radish-ACC finish-SG.FUT may I:DAT radish-ACC finish-INF  
 Mozhno *ja* redisk-u *doe-st'?*  
 may I:NOM radish-ACC finish-INF  
*May I finish the radish?*

(the error results from splicing the pronoun *ja*, the nominative case form of 'I', from the first phrase and *doest'*, the infinitive form of 'finish', from the second. Both phrases are a form of request).

- (2) On stal *xud-oj* + On stal *xud-ym* → On stal *xud-om*  
 he became thin-SG.NOM he became thin-SG.INS [non-existent word form]  
*He has grown thin.*

(the error is caused by blending two different case forms, the nominative and the instrumental, of the adjective *xudoj* 'thin'. Either adjective case form collocates with the verb 'become').

Errors of this kind indicate that parallel planning is not confined to the lemma level of sentence generation; instead, they suggest parallelism in syntactic planning at the positional level of production (a level where, after the lemmas have been retrieved, the order of elements within a phrase is specified and all inflectional processing takes place). The blends occur when the processing mechanism fails to choose between the two competing positional level structures. Thus, syntactic blends suggest that parallelism, or overplanning is a basic principle of language production, which, alongside lexical retrieval, is involved in syntactic processing.

An implication for cognitive resource consumption is that syntactic structures are not produced automatically; instead, syntactic blends indicate that generating a syntactic structure requires significant processing resources (cf. Ferreira & Engelhardt, 2006).

## Bibliography

- Coppock, E. 2006. Alignment in Syntactic Blending. *The State of the Art in Speech Error Research: Proceedings of the LSA Institute Workshop*. C. T. Schütze and V. S. Ferreira (eds.), Cambridge, MA: MIT Working Papers in Linguistics, 240-255.
- Cutler, A. 1982. The reliability of speech error data. *Slips of the Tongue and Language Production*. A. Cutler (ed.), Amsterdam: Walter de Gruyter/Mouton, 7-28.
- Fay, D. 1982. Substitutions and splices: A study of sentence blends. *Slips of the Tongue and Language Production*. A. Cutler (ed.), Amsterdam: Walter de Gruyter/Mouton, 717-749.
- Ferreira, F. & Engelhardt, P. 2006. Syntax and Production. *Handbook of Psycholinguistics*. M. Traxler and M. A. Gernsbacher (eds.), London: Academic Press, 61-91.
- Garrett, M. F. 1980. Levels of processing in sentence production. *Language production, Volume I: Speech and Talk*. B. L. Butterworth (ed.), London: Academic Press, 177-220.
- Stemberger, J. P. 1990. Word shape errors in language production. *Cognition* 35: 123-157.

## A Construction Grammar Approach to German Support Verb Constructions

Support verb constructions (henceforth: SVCs) are constructions consisting of a verb with a reduced meaning (when compared to the full verb) and a noun. For example, (1a) illustrates the full verb meaning of German *stellen* ‘to put’, while (1b) uses *stellen* in a SVC.

- (1) a. Peter stellt die Vase auf den Tisch.  
‘Peter puts the vase on the table.’
- b. Peter stellt die Frage zur Diskussion.  
‘Peter brings the question up for discussion.’

Previous analyses (e.g. von Polenz 1963, Winhart 2002) provide a detailed account of the function of the verb in SVCs. However, neither of the two approaches fully explains why certain verb-noun combinations are unacceptable. Compare, e.g., *geraten* ‘to get into’, which can combine with *Brand* ‘fire’ in (2a), but not with *Feuer* ‘fire’ in (2b), even though the two nouns are synonyms.

- (2) a. Das Haus gerät in Brand.  
‘The house catches fire.’
- b. \*Das Haus gerät in Feuer.  
‘The house catches fire.’

Von Polenz (1963) does not explain why such selectional restrictions occur at all and Winhart (2002) only provides an account for which arguments are obligatory, implicit, or optional.

This paper proposes a novel approach towards identifying selectional restrictions in German support verb constructions by applying insights from Frame Semantics (Fillmore 1985) and Construction Grammar. It differs from syntactic-centric and lexical-conceptual structure approaches in that frame-semantic information is shown to influence directly a verb’s and a noun’s ability to combine with each other. I argue that the nominalization in (2b) cannot combine with the support verb because the frame-semantic information evoked by *Feuer* is incompatible with the frame semantics of *geraten*. Thus, either the verb and/or the noun blocks the formation of a support verb constructions. My analysis shows that in order for the support verb and the noun to be able to combine, their frame-semantic information needs to be compatible (as in 1b). However, in some circumstances such as in (2a, b), SVCs need to be listed as idioms in the lexicon because there does not seem to be any compositional restrictions that allow *geraten* to combine with *Brand* ‘fire’ but not *Feuer* ‘fire’. Based on a corpus of more than 300 SVCs with *geraten* I show that there are different patterns of productivity and idomaticity that necessitate a network inheritance model. Some SVCs, such as *ins Rollen geraten* ‘to start rolling’ allow widespread replacement of the noun with near-synonyms. Other SVCs such as in (2a) do not allow such replacement. On this view, both the abstract meaning of a SVC (e.g., in *X geraten* ‘to get into X’) and item-specific knowledge needs to be captured to be able to account for the full range of SVCs headed by *geraten*. Therefore, I posit a new construction which captures all the meanings expressed by SVCs with *geraten*.

## Bibliography

- Fillmore, C. J. 1985: Frames and the Semantics of Understanding. *Quaderni di Semantica* 6: 222-254.
- Von Polenz, P. 1963: Funktionsverben im heutigen Deutsch. *Sprache der rationalisierten Welt. Wirkendes Wort*, Beiheft 5
- Winhart, H. 2002: Funktionsverbgefüge im Deutschen zur Verbindung von Verben und Nominalisierungen. Dissertation. Universität Tübingen.

## Same conceptual content, different construals: On the Postposition / Preposition variation of Finnish path adpositions

Finnish is a predominantly postpositional language which also has a few prepositions and some two-faced adpositions that can be used in both functions (for a discussion, see e.g. Grünthal 2003). Many of such two-faced adpositions indicate a path, e.g., *kautta* 'via', *läpi* 'through', *pitkin* 'along', *halki* 'across' and *yli* 'over [dynamic]'. In the literature it has been pointed out that there are differences between the variants that are related to their stylistical value: prepositions are more literal or archaic whereas postpositions are neutral. However, this characterization seems to be valid only in the spatial domain, since for instance temporal relations are usually more natural to be expressed by prepositions. It has also been pointed out (Leino 1993, Salmi 1994) that there are some semantic differences between the prepositional and postpositional variants of some Finnish path adpositions, most clearly *kautta* 'via' and *pitkin* 'along'. These differences are, however, related to the construal of the path than to the conceptual content of the expression (for the distinction, see e.g. Langacker 2008). One crucial contrast goes between the meanings of actual motion (most naturally expressed with postpositions but also by prepositions; example 1) and the directional representation of static existence or occurrence of entities along the path (most naturally expressed with a preposition, example 2).

- |    |  |                          |                       |                          |                 |                        |  |
|----|--|--------------------------|-----------------------|--------------------------|-----------------|------------------------|--|
| 1) | Juoks+i+mme<br>run+PST+IPL   | metsä+n<br>forest+GEN    | läpi.<br>through      | ~                        | läpi<br>through | metsä+n.<br>forest+GEN |  |
|    | 'We ran through the forest.'   |                          |                       |                          |                 |                        |  |
| 2) | Sien+i+ä<br>mushroom+PL+PAR  | kasv+o+i<br>grow+PST+3SG | mon+in<br>many+PL.INS | paiko+in<br>place+PL.INS | läpi<br>through | metsä+n.<br>forest+GEN |  |
|    | 'There were mushrooms growing in many places throughout the forest.' |                          |                       |                          |                 |                        |  |

This last-mentioned path type is also illustrated by Talmy's (2000:71) example 'There is a house every now and then through the valley'. According to Talmy, the example utilizes a sequential perspectival mode, which involves a moving proximal perspective point with a local scope of attention, i.e., a subjective scanning over a static configuration. In such an expression there is no mover traversing the path, and the construal of the path is based on a subjective operation by the conceptualizer. Our preliminary corpus analysis suggests that the usages of path postpositions concentrate in examples indicating actual motion, whereas prepositional path expressions are more flexible and heterogeneous in this respect – they can indicate actual motion (as in 1), but to indicate the meaning like that of example 2, a preposition is needed. From the point of view of synonymy, both variants thus indicate a path and convey a similar conceptual content, their difference lying in the way of construing the path – a preposition is more flexible than a postposition and favors a construal based on purely subjective operations.

### Bibliography

Grünthal, Riho 2003. Grünthal, Riho 2003. *Finnic adpositions and cases in change*. Suomalais-ugrilaisen seuran toimituksia 244. Helsinki: Suomalais-ugrilainen seura.

- Langacker, Ronald W. 2008 Langacker, Ronald W. 2008. *Cognitive grammar: A basic introduction*. New York: Oxford University Press.
- Leino, Pentti 1993 Leino, Pentti 1993. Polysemia – kielen moniselitteisyys. *Suomen kielen kognitiivista kielioppia 1. Kieli 7*. Helsinki: Helsingin yliopiston suomen kielen laitos.
- Salmi, Tiina. 1994 Spatiaalinen ja metaforinen yli. – Leino, Pentti ja Tiina Onikki (toim.): *Näkökulmia polysemiaan: suomen kielen kognitiivista kielioppia 2. Kieli 8*. Helsinki: Helsingin yliopiston suomen kielen laitos, s. 164–188.
- Talmy, Leonard 2000: Talmy, Leonard 2000. *Toward a cognitive semantics*. Volume 1: Concept structuring systems. Cambridge: MIT Press.

## Why do we need external and internal case forms for the gram *sisä-* in Finnish?

Finnish has a rich system of grammatical words, or *grams*, which have many different functions. These functions extend from conceptually dependent adpositions to autonomous adverbs. When used as adpositions, many grams take their complement in the genitive case, which reflects the historical background of the construction: the grams were once nouns and the complements were their genitive modifiers. This is why Finnish grams often have many local case forms in use, many of them utilizing the case inflection to a remarkable extent.

The Finnish local case system has two paradigms: the internal cases (expressing notions like ‘inside’, ‘into’, ‘out of’), and the external cases (expressing notions like ‘at’, ‘to the outside of’ and ‘from the outside of’). A pervasive feature in both systems is directionality: both case series have a locative (‘in’/‘at’) case, a lative (‘to’) case and a separative (‘from’) case. Some grams can be productively inflected in all local cases. For instance, the stem *sisä-* ‘in’, has a full paradigm: *sisä+llä* ‘in’ [ADESSIVE] ~ *sisä+ltä* ‘from in’ [ABLATIVE] ~ *sisä+lle* ‘into’ [ALLATIVE] vs. *sisä+ssä* ‘in’ [INESSIVE] ~ *sisä+stä* ‘from in’ [ELATIVE] ~ *sisä+än* ‘into’ [ILLATIVE]. However, unlike with proper nouns, it is not intuitively clear what the opposition between the internal and the external case forms of this gram are. Both express, after all, a relationship of containment. In contrast, the directionality opposition between different case forms of the gram is quite functional and transparent.

In cognitive linguistics, it has been assumed that basic spatial concepts include relations such as inclusion, separation, association and contact (Langacker 1987: 225–231). When applying such notions to case-inflected grams, we need to remember that they express two different relations simultaneously – one by the stem of the gram, another by the case ending. Another central assumption of cognitive linguistics is that full synonymy does not exist. Thus internal vs. external local case forms of a gram should not be synonymous either. However, the extensive literature on grammaticalization has not paid much attention to the internal morphology of grams. Even in typologically-based studies where the development of adpositions from nouns is described in detail (e.g., Heine & al. 1991; Svorou 1993), the internal morphology of grams is discussed only superficially if at all.

In our paper we study the spatial functions and actual usage of the gram *sisä-*, with special reference to a) the internal division of labor between the stem and the case ending, and b) the opposition between the internal vs. external case forms of the gram. We will show that in spite of their apparent synonymy in simple examples, there are grammatical uses and kinds of landmarks that are only compatible with one kind of a case form (internal or external) but not another. We will also study the usage of the grams in spoken and written corpora, showing how the apparently symmetric case paradigms are actually blended in usage: e.g., the lative meaning may be primarily expressed by the internal case form of the gram, whereas the locative and separative meanings are more often expressed by external case forms.

### Bibliography

- Heine, Bernd, Claudi, Ulrike & Hünemeyer, Friederike 1991: *Grammaticalization. A conceptual framework*. Chicago: The University of Chicago Press.
- Langacker, Ronald W. 1987: *Foundations of Cognitive Grammar*. Vol. 1: *Theoretical Prerequisites*. Berlin: Mouton de Gruyter.

Svorou, Soteria 1993: *The grammar of space*. Amsterdam: John Benjamins.

## **Synonymy, sameness and semantic differences in verbal possessive constructions**

The paper is dedicated to the problem of synonymy in syntax, i.e. whether competing possessive constructions in the verbal phrase are supposed to be synonym in a way that they are exchangeable. Working on possessive constructions a continuum between the advantage of linguistic economy and the need of expressing connotations provides a rich field for research, looking for the semantic overlap of these syntactic features.

The starting point of the investigation is a reconsideration of Heine's quotation 1997:132 f.: *“On the one hand, there are languages where there is only one construction (e.g. the ‘have’-construction in English), taking care of most or all of the spectrum of possessive notions, even if there are other constructions in addition to cover parts of the spectrum. In Manding and Ewe, however, the spectrum is divided up among different constructions. Whether this typological contrast is suggestive of basic differences in the way possession is conceptualized, remains to be investigated.”*

Our analysis however is based on a corpus consisting of 16 languages of Europe both dead (Gothic, Latin, Greek, Cl. Armenian, OCHS) and alive (Slavic and Baltic languages), using up to 5 different verbal possessive constructions each. All verbal possessive construction have been classified – following Heine's event-schemata – according to possessors, possessa, negation, special semantic meaning in a given context, so the question of synonymy and disambiguation can be answered both within a language, cross-linguistically and concerning historical developments starting with PIE. So the study goes far deeper than the works on competing possessive cases known within Classical Philology, by asking which possessive concepts are divided by non-synonym constructions and whether there are implicational hierarchies to be found within the disambiguation of these semantic groups.

A special consideration is dedicated to the question whether Indo-European languages use different possessive construction to express the difference between alienable and inalienable relations, what would contradict a concept of synonymy in possessive constructions. Although the opinion expressed so far is a very sceptical one (cf. Bauer 2000: 164: *“No Indo-European language distinguishes explicitly between alienable and inalienable possession. Yet reinterpretation of evidence from a number of daughter languages suggests that Indo-European at a very early stage may have had this distinction.”*), our answer is a positive one by giving evidence in a systematic difference expressed by alternating constructions following the principles of grammaticalisation.

### Bibliography

- Bauer, B.L.M. 2000: *Archaic Syntax in Indo-European*, Berlin/New York.  
Heine, B. 1997: *Possession, Cognitive Sources, forces and grammaticalisation*, Cambridge.  
Seiler, Hj. 1983: *Possession as an Operational Dimension of Language*, Tübingen.  
Seiler, Hj. 2008: POSSESSION: variation and invariance, in: Th. Stolz (ed.), Hansjakob Seiler. *Universality in Language beyond Grammar: Selected writings 1990-2007*, Bochum.

## The Misspelling of Adpositional Phrases in Estonian – Illiteracy or Manifestation of Semantic Difference?

In Estonian there is a noticeable tendency to misspell adpositional phrases. It is fairly common to come across adpositional phrases written as one word, although according to the Estonian orthography the adpositional phrases should always be written as two words. It is hypothesized that the spelling issues of Estonian adpositional phrases are not only due to speakers' poor knowledge of orthography, but are instead related to more basic tendencies operating in languages, such as the 'one form one meaning' principle. Therefore, we consider the orthographically correct analytic construction (Example 1a) and the misspelled synthetic construction (Example 1b) as two competing constructions with potentially distinct meanings.

- |        |                                       |               |          |       |           |
|--------|---------------------------------------|---------------|----------|-------|-----------|
| 1. (a) | Hoi-a                                 | dušš-I        | pea- Ø   | kohal |           |
|        | hold-imp.sg2                          | shower-sg.par | head-gen | over  |           |
|        | 'Hold the shower head over your head' |               |          |       |           |
| (b)    | Su-l                                  | o-n           | nüüd     | katus | *peakohal |
|        | you-sg.all                            | be-sg3 now    | roof.nom | over  |           |
|        | 'You have a roof over your head now'  |               |          |       |           |

This view is in accordance with many cognitive linguistic accounts, such as Goldbergian Construction Grammar, which refutes the possibility of full synonymy in a language, referring to the principles of No Synonymy and Maximized Expressive Power. Thus, if we consider the two constructions as syntactically distinct, there must be either semantic or pragmatic difference between them (Goldberg 1995; Goldberg 2006). It is hypothesized that orthographical incorrectness is related to the different meanings that these items may have for speakers – by 'incorrect' compounding speakers actually try to differentiate the abstract/metaphorical senses from the more concrete ones. The results of a forced choice task – during which the participants were also asked to explain their choice of construction in free form – conducted with 143 subjects confirmed this hypothesis.

The first reason for considering distinct meanings of these constructions is the fact that the orthographical issues arise when dealing with phrases that occur in metaphorical contexts (Example 1b) or express abstract functions, i.e. abstract relations in which motivation for using a particular adpositional phrase is not transparent anymore (Example 2). So it seems that different linguistic contexts trigger the use of different meanings: we may consider the adpositional phrases that occur in various contexts as polysemous, furthermore the misspelled phrases are believed to be lexicalizing (Lehmann 2002; Brinton, Traugott 2005).

- |    |                           |        |                     |       |                |
|----|---------------------------|--------|---------------------|-------|----------------|
| 2. | Ma                        | ole-n  | selle- <del>o</del> | peale | mõel-nud.      |
|    | I.nom                     | be-sg1 | this-gen            | on    | think-pst.ptcp |
|    | 'I have thought about it' |        |                     |       |                |

It seems that another major factor triggering the variation is analogy, which is considered a significant factor in language change (Aitchinson 2000: 216). As the studies I have conducted (Jürine 2009) have shown, language users are not able to formulate any rules they might use to make a selection between the two forms, but seem to rely on their feeling for language and use

analogies (they learn the orthography of some adpositional phrases and then extend the rule to everything that seems comparable). The category of Estonian compound adverbs may set the pattern to misspell. From this perspective, the current variation, i.e. spelling the Estonian adpositional phrases as either one or two words, may indicate the beginnings of new compound adverbs. In which context exactly the orthographically incorrect constructions are more likely to occur is to be clarified via corpus-based research.

## Bibliography

- Aitchinson, Jean 2000. *The Seeds of Speech: Language Origin and Evolution*. Cambridge University Press.
- Brinton, Laurel J., Elizabeth Closs Traugott 2005. *Lexicalization and Language Change*. Cambridge University Press.
- Goldberg, Adele E. 1995. *Constructions: a Construction Grammar Approach to Argument Structure*. Chicago: The University of Chicago Press.
- Goldberg, Adele. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Jürine, Anni 2009. Kas \*metsavahel tähendab muud kui metsa vahel? [Does \*metsavahel mean something else than metsa vahel?] – *Oma Keel*, 18.
- Lehmann, Christian 2002. New Reflections on Grammaticalization and Lexicalization. *New Reflections on Grammaticalization*. Typological Studies in Language 49. Eds. Ilse Wischer, Gabriele Diewald. Amsterdam; Philadelphia: John Benjamins Publishing Company, pp. 1–18.

## Synonymous Locative Constructions in Estonian

In Estonian, we have among other synonymous items, several ways to express spatial relations – we can use locative cases or adpositional constructions. The present paper builds on previous research on this topic (Klavan et al. to appear) and the aim is to determine what does the choice between the Estonian locative constructions with the adessive case (example 1a) and the adposition *peal* (example 1b) depend on. The language data analysed comes from two linguistic experiments – a production task and an acceptability task. The paper proceeds from the theoretical premises of both Construction Grammar (Goldberg 1995, 2006) and Cognitive Grammar (Langacker 1987, 2008), where one of the basic general assumptions is that of no-synonymy – when two constructions differ syntactically, then they also differ either semantically or pragmatically (Goldberg 1995: 67).

- (1) a. vaas            on            laual  
vase:NOM      be-PRS:SG3    table:ADE  
'the vase is on the table'
- b. vaas            on            laua            peal  
vase:NOM      be-PRS:SG3    table:GEN      on  
'the vase is on the table'

Previous research has shown (Klavan et al. to appear) that although the Estonian locative cases and adpositional constructions are said to “express more or less the same meaning” in traditional grammars of Estonian (Erelt et al. 1995: 34), there are differences in how language speakers actually use these constructions. Klavan et al. (to appear) report the results of a forced choice task and a production task carried out with 138 native speakers of Estonian, aged between 15 and 71. The results of these studies confirm that the type of relation between the Figure and Ground plays a role: when the relation is abstract, the preferred choice is the adessive case; when we have a spatial scene, where the relation is somewhat unusual or atypical (e.g. a book on top of an alarm-clock), the language users prefer the adpositional construction. However, with common, everyday spatial scenes (e.g. a vase on the table), the picture is not as clear-cut: both the adessive case and the adpositional construction were used in the two studies with more or less the same frequency.

In order to determine what factors play a role in stereotypical spatial relations, two further studies have been devised – a production task and an acceptability task. The production task tests whether the following factors have an effect: animacy and size of the Figure, animacy and type of the Ground, and the type of contact between the Figure and Ground. It is predicted, among other things, that when the Ground is a thing (e.g. a tray), the preferred choice is the adposition *peal* (‘on’) and when it is a place (e.g. a wall), the language users prefer the adessive case. Previous research (e.g. Bartens 1978, Klavan et al. to appear, Ojutkangas 2008) shows that there might be an effect of these factors, but in order to validate these claims, converging evidence needs to be gathered.

Another group of factors that may influence the choice of these constructions with stereotypical spatial relations are the syntactic ones. Although a small-scale corpus study conducted by Rannat (1991) shows that the preference of either the synthetic or analytic form does not depend on the syntactic composition of the clause (e.g. on the transitivity of the predicate verb, the type of verb used, the clause element the locative phrase expresses, word order), these claims need to be substantiated because her study is based only on her own intuition. In order to test whether there is

an effect of the syntactic composition in the choice between the Estonian adessive and the adposition *peal* or not, an acceptability rating task has been devised.

#### Bibliography

- Bartens, R. 1978. Synteettiset ja analyttiset rakenteet lapin paikanilmauksissa. Suomalais-ugrilaisen Seuran toimituksia 166. Helsinki: Suomalais-Ugrilainen Seura.
- Erelt, M., Kasik, R., Metslang, H., Rajandi, H., Ross, K., Saari, H., Tael, K. and Vare, S. 1995. Eesti keele grammatika I. Morfologia. Tallinn: Eesti Teaduste Akadeemia Eesti Keele Instituut.
- Goldberg, A. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. USA: University of Chicago Press.
- Goldberg, A. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Klavan, J., Kesküla, K. and Ojava L. To appear. Synonymy in Grammar: The Estonian Adessive Case and the Adposition *peal* 'on'. In S. Kittilä, K. Västi and J. Ylikoski (eds.), *Studies on Case, Animacy and Semantic Roles*. John Benjamins.
- Langacker, R. W. 1987. *Foundations of Cognitive Grammar. Volume I: Theoretical Prerequisites*. Stanford: Stanford University Press.
- Langacker, R. W. 2008. *Cognitive Grammar. A Basic Introduction*. Oxford: Oxford University Press.
- Ojutkangas, K. 2008. Mihin suomessa tarvitaan sisä-grammeja? *Virittäjä* 3, 382–400.
- Rannat, R. 1991. *Noomeni sünteetiliste ja analüütiliste vormide kasutus*. [The Use of the Synthetic and Analytic Forms of the Noun.] Unpublished BA thesis. University of Tartu, Institute of Estonian and General Linguistics, Tartu.

## Relativization of subjects in Russian: a case study of competition between syntactic synonyms

Russian uses two major strategies for relativizing on subjects, namely, active participle strategy and relative pronoun strategy, pronoun *kotoryj* ‘which’ being by far the most frequent in the latter case. In most of the cases in question the two structures are interchangeable without any obvious difference in meaning (cf. (1a–b)).

(1a)	Ja I spektakle stage.play lit.: ‘I know the actress playing in this play’.	znaju know	aktrisu actress	igrajuščuju playing	v in	etom this	
(1b)	Ja I spektakle stage.play ‘I know the actress who plays in this play’.	znaju know	aktrisu actress	kotoraja who	igraet plays	v in	etom this

The aim of the study was to specify the factors influencing the choice of relativization strategy. The following parameters have been shown to be correlated with the choice of relativization strategy: the mode of communication (oral vs. written); the tense, aspect and mood of the verb in the relative clause; the presence of the reflexive suffix on the verb; the characteristics of individual verbs as well as verb classes; the length of the relative clause; the presence of complements of various types; negation; the type of the relative clause (restrictive vs. non-restrictive); the syntactic position that the head of the relative clause occupies within the main clause; intonational separation of the main clause and the relative clause. All the data were obtained using the Russian National Corpus ([ruscorpora.ru](http://ruscorpora.ru)) and analysed with statistical methods.

It was found that there are a number of factors **unequivocally** determining the choice of relativization strategy. These are partly due to restrictions on participle formation as there are no participle forms inflected for conditional mood or future tense in standard Russian. Other restrictions include the lack or rareness of some participle forms for certain verbs, such as *byt’* ‘be’, *khotet’* ‘want’ and *moč’* ‘be able’. When there is a need to express corresponding grammatical or lexical meanings, *kotoryj*-clause is used almost invariably.

There are also some parameters that show strong **statistical** correlation with a particular relativization strategy. For instance, *kotoryj*-clauses are more frequent in longer relative clauses in which the verb has more dependents, especially if these are of clausal type (e.g. converbs). Participles tend to appear in restrictive relative clauses. They are also much more common in writing than in spoken language.

Both the absolute and statistical distinctions found in the study boil down to the following generalization: if compared to *kotoryj*-clauses, participial clauses are more nominalized (in terms of [Lehmann 1986]), that is, display more properties of a nominal (in the broader understanding of the term) rather than a clause. For example, constraints on tense and constraints on possible complements are both known to be characteristic of nominalized relative clauses [Lehmann 1986: 671–672]. While some of the former are absolute in Russian and the latter only show up in statistical correlations, there still seems to be a unifying macro factor behind both types of constraints that influence the choice between the two syntactic strategies.

## Bibliography

Lehmann, Ch. 1986. On the typology of relative clauses. *Linguistics*, 24 (4): 663–680.

## Synonyms in modern discourses: marking communicative choices

‘Synonymy, as a rule, is not complete equivalence’ (Jacobson 2002: 114): Roman Jacobson’s statement stresses the thought that synonyms create linguistic variety, not sameness. The theory of meaning views synonyms as the main manifestation of paradigmatic relationship (Кронгауз, 2002); whereas the discourse analysis concentrates on the fact that the use of synonyms is the condition for cohesion development which is achieved through words relating to each other within the text (Cutting 2002: 2).

This paper, however, views synonymy as the increase of speakers’ communicative choices which help them to express connotations and specify their attitudes to the topic of discussion. This quality can be very well inferred from the following example: ‘Your clothes are too cool and you have **purple** highlights.’ ‘**Molichino**, please!’ I cried. ‘**Purple** makes me sound like a ... a teenager. (Keyes 2008: 8). ‘Molichino’ is an Italian borrowing which means ‘purple’. Yet, as it is possible to see from the example, it is not just an exact foreign equivalent. When this word is used in English discourse, it acquires new connotations, shades of meaning. Thus, these adjectives can be described as synonyms.

The expansion of synonymic fields is mostly affected by nonlinguistic causes and currently illustrates the appearance of numerous borrowings which have been integrating into the language. These new words often have local equivalents. However, the new words possess a strong pragmatic core in their meaning (which is always at least slightly different from that of the local word) and mark the discourse as contemporary and topical. The users of modern borrowings signal their belonging to a world without boundaries, their willingness to be a part of common problems and achievements (Khoutyz 2010). Moreover, speakers’ preferences to a certain synonym can implicatively inform us about the genre of the discourse, its situational reference, speakers’ age, profession, social status, etc.

This paper is going to observe how synonymy marks the discourse on many levels of meaning transfer. To accomplish this task, first, I am going to discuss various approaches to studying synonymy and dwell whether borrowings can be defined as full-fledged synonyms to local words. Then, communicative properties such as modality, connotation, implicating, ideological preference, etc. of a synonym are discussed. These properties are revealed through the analysis of functions of a borrowed word in a contemporary discourse. In this case the use of a certain word serves as ‘the conventional signal for a recurrent coordination problem’ (Croft 2000: 176). In conclusion, the communicative choices expressed through the synonym use are classified.

### Bibliography

- Croft, W. (2000). *Explaining Language Change*. Oxford: Longman.
- Cutting, J. (2002). *Pragmatics and Discourse*. London, New York: Routledge.
- Jacobson, R. (2002). On Linguistic Aspects of Translation. In: *The Translation Studies*. London, New York: Routledge, pp. 113-118.
- Keyes, M. (2008). *This Charming Man*, Penguin Books, London.
- Khoutyz, I. (2010). The pragmatics of anglicisms in modern Russian discourse. In: *From international to local English – and back again*. Bern: Peter Lang, pp.197-208.
- Кронгауз, М.А. (2002). *Семантика*. Москва: Российский государственный гуманитарный университет.

## How do languages deal with synonymy?

As stated in Taylor (2002:271), synonymy “tends to be avoided” as “an extravagant luxury”: indeed, when words come to have the same meaning as a consequence of semantic change, “‘corrective’ mechanisms come into play—one of the words may fall into disuse, or the words become associated with different nuances, possibly of a stylistic or sociocultural nature.” (ib. 271-2). In my paper I would like to show how such avoidance of synonymy has consequences on the use of two prepositions, *aná* and *katá*, originally ‘upward’ and ‘downward’, in Ancient Greek. As early as in the first written sources, these prepositions display a similar meaning extension. They both mean ‘throughout’, ‘among’ in occurrences such as:

- (1) Murmidónas d' ár' epoikhómenos thó#e#xen Akhilleùs  
Myrmidon:ACC.PL PTC PTC walk:PART.PRS.M/P.NOM arm:AOR.3SG A.:NOM  
pántas anà klisías  
all:ACC.PL up hut:ACC.PL  
“but Achilles went to and fro throughout the huts and let harness all the Myrmidons”  
(Homer, Il. 16.155-156).
- (2) ou mèn gàr pot' áneu de#□o#n ên, allà kat' autoùs  
NEG PTC PTC ever without enemy:GEN.PL be:IMPF.3SG but down DEM.ACC.PL  
stro#phât'  
range:IMPF.M/P.3SG  
“for he was never away from the enemies, but ranged among them”  
(Homer, Il. 13.556-557).

According to the traditional view, the semantic difference between *aná* and *katá* in such occurrences lies in the fact that the former indicates that the trajector moves along a trajectory which touches all relevant points in the landmark, while the latter indicates random motion touching only part of the units included in the landmark (Eberling 1885, Luraghi 2003). More recently, George (2006) has argued that the two prepositions had become complete synonyms, and that the choice between the two only depended on their position in the verse. However, while it is true that some of the occurrences mentioned by George do not point toward a difference in meaning, it is also true that the distribution of the two preposition only partly depends on metric factors; in addition, there is evidence from later prose (e.g. from Herodotus, cf. Luraghi 2003: 242) that the meaning traditionally assumed for Homer was still preserved. The real difference between *aná* and *katá* lies, as already indicated by Spitzner (1831), in the more restricted scope of *aná* with respect to *katá* in Homeric Greek; later, *aná* became increasingly infrequent, as shown especially in Attic prose, where the preposition only survives in some idiomatic expressions (Luraghi 2003: 227; see further Bortone 2000). This development may be seen as an instance of ‘corrective mechanism’ in the sense of Taylor (2002): in Homer, the two prepositions could often occur in the same contexts, in which the semantic difference, even if existent, was irrelevant; this overlap turned into increasing synonymy, and the use of *aná* became restricted to idioms.

### Bibliography

- Bortone, P. 2000: *Aspects of the history of Greek prepositions*. Oxford: Hilary Term.  
George, C. 2006: The spatial use of *aná* and *katá* with the accusative in Homer. *Glotta* 82: 70-95  
Ebeling, H. 1885: *Lexicon homericum*. Leipzig.

- Luraghi, S. 2003: *On the meaning of prepositions and cases: the expression of semantic roles in Ancient Greek*. Amsterdam: Benjamins.
- Spitzner, F. 1831: *Dissertatio de vi et usu praepositionum aná et katá apud Homerum*. Wittenberg.
- Taylor, J. R. 2002: *Cognitive Grammar*. Oxford: OUP.

## **Construing and Conventional Usage in the Use of Synonymy: An Analysis of Two sets of (Near)-synonymous Nouns Using Both Corpus and Elicited Data**

Most corpus-based behavioral profile (BP) studies of (near)-synonyms so far have been on verbs or adjectives (Divjak 2006; Divjak & Gries 2006; Gries 2001; Hanks, 1996; Liu, 2010). No BP study appears to have examined synonymous nouns. Furthermore, corpus-based BP research, though uniquely valuable, has the limitation of not being able to directly inquire why the producers of the language data used a given linguistic item the way they did. The latter information may only be obtained via solicited language, a type of data that recent research has shown to be valuable in the study of synonymy (Arppe & Järvikivi, 2007; Geeraerts, 2010). Hence, this study examines two sets of (near)-synonymous nouns (authority vs. power vs. right; duty vs. obligation vs. responsibility), using both corpus and elicited data.

The study began with a BP analysis, using the Corpus of Contemporary American English (COCA). It investigated the collocational/colligational patterns of the (near)-synonyms in each set, including typical adjectives, post-nominal infinitives, and verbs used with the synonymous nouns. The results including those of a multi-factorial test called “hierarchical configural frequency analysis” (reported in tables below) indicate that pre-/post-nominal complements/modifiers were effective in catching most of the semantic differences among synonymous nouns and in delineating a “coarse” internal semantic structure of a synonymous-noun set. However, these examined distributional patterns seemed unable to reveal some fine-grained differences that appeared to exist, for the (near)-synonyms in each set, while displaying many different distributional patterns, also exhibited some identical patterns, e.g. they sometimes took the same pre-/post-nominal complements/modifiers (e.g. civic duty/obligation/responsibility; the authority/power/right to declare war); it was not clear whether the different nouns used in the same distributional context in each case had the same meaning.

To help answer this question, 32 of these difficult-to-distinguish uses in context were selected from COCA to form a questionnaire (with the synonyms removed; see Appendix). 42 native English speaking college students were asked to read each sentence and supply the missing word by selecting from the three synonyms in the relevant set. The subjects were also asked to explain the rationales for their choices. Statistical tests of the subjects’ choices (tabulated/reported in Appendix) show a clear lack of consensus in 18 (56%) of the 32 items. Also, in 18 items, the choice by the majority of the subjects differed from that used in the original COCA sentence. The analysis of the rationales the subjects gave for their selections reveals that there were two major competing motivating factors for their choices: 1) their construal of the situation (e.g. whether a subject construed a given question as an issue of authority, power, or right) and 2) conventional usage patterns (i.e. often a subject made a choice simply because it was “the idiomatic usage” he/she had often heard). Many subjects were often torn between the competing factors in their decision making. This new finding foregrounds the interface between construal and conventional usage as a key issue in understanding synonymy, an issue that should be of interest to us, especially cognitive/construction grammarians.

## Bibliography

- Arppe, A. (2008). *Univariate, bivariate, and multivariate methods in corpus based lexicography: A study of synonymy*. Helsinki: University of Helsinki.
- Arppe, A & Järvikivi, J. (2007). Every method counts: Combining corpus-based and experimental evidence in the study of synonymy. *Corpus Linguistics and Linguistic Theory*, 3.2, 131–159.
- Charles, W. G. (2000). Contextual correlates of meaning. *Applied Psycholinguistics*, 21, 505-524.
- Church, K. W., Gale, W., Hanks, P., & Hindle, R. (1991). Using statistics in lexical analysis. In U. Zernik (Ed.), *Lexical Acquisition: Exploring Online Resources to Build a Lexicon* (pp.115-164). Hillsdale, NJ: Lawrence Erlbaum.
- Church, K. W., Gale, W., Hanks, P., & Hindle, R., & Moon, R. (1994). Lexical substitutability. In B. T. S. Atkins and A. Zampolli (Eds.), *Computational Approaches to the Lexicon* (pp.153-177). Oxford: Oxford University Press.
- Divjak, D. (2006). Ways of intending: Delineating and structuring near synonyms. In S. ThGries and A.Stefanowitsch (Eds.), *Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis* (pp. 19-56.). Berlin and New York: Mouton de Gruyter.
- Divjak, D. & Gries, S. Th. (2006). Ways of trying in Russian: clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory*, 2, 23-60. The corpus of contemporary American English Online. (2010). compiled and provided by Mark Davies of Brigham Young University, available at <http://www.americancorpus.org/>.
- Cruse, D. A. (1986). *Lexical Semantics*. Cambridge: Cambridge University Press.
- Cruse, D. A. (2000). *Meaning in Language: An Introduction to Semantics and Pragmatics*. Oxford: Oxford University Press.
- Divjak, D. (2006). Ways of intending: Delineating and structuring near synonyms. In S. Th Gries and A.Stefanowitsch (Eds.), *Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis* (pp. 19-56.). Berlin and New York: Mouton de Gruyter.
- Divjak, D. & Gries, S. Th. (2006). Ways of trying in Russian: clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory*, 2, 23-60.
- Edmonds, P. & Hirst, G. (2002). Near synonyms and lexical choice. *Computational Linguistics*, 28, 105-144.
- Firth, J. R. 1957. *Papers in Linguistics, 1931–1951*. New York: Oxford University Press.
- Geeraerts, D. (1986). On necessary and sufficient conditions. *Journal of Semantics*, 5, 275-291.
- Geerarts, D. (2010). *Theories of lexical semantics*. Oxford: Oxford University Press.
- Gries, S. Th. (2001). A corpus linguistic analysis of English *-ic* vs *-ical* adjectives. *ICAME Journal*, 25, 65-108.
- Gries, S. Th. (2004). HCFA 3.2. A program for R
- Gries, S. Th. (in press). *Statistics for Linguists with R: A Practical Introduction*. Berlin: Mouton de Gruyter.
- Gries, S. Th. & Otani, N. (in press). Behavioral profiles: A corpus-based perspective on synonymy and antonymy. *ICAME Journal*.
- Hanks, P. (1996). Contextual dependency and lexical sets. *International Journal of Corpus Linguistics*, 1 (1), 75-98.
- Liu, D. (2010). Is it *chief, main, major, primary, or principal* concern? A corpus-based behavioral profile study of the near-synonyms. *International Journal of Corpus Linguistics*, 15, 56-87.
- Miller, G. A. & Charles, W. G. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes*, 6 (1), 1-28.

Murphy, M. (2003). *Semantic relations and the lexicon*. Cambridge: Cambridge University Press.

## Metaphorical phrasal quantifiers and synonymy in a cross-linguistic perspective

Phrasal quantifiers or quantifying of-constructions such as *a torrent of, mountains of, buckets of, a barrage of, legions of, a sea of*, etc. are numerous and tend to be regarded as synonymous. All of these complex determiners fit into the same collocational framework (*a*) *N1 + of + N2* and are used to designate not only big quantities but also small ones as with *a crumb of* or *a grain of*. If taken at face value, these quantifiers can be considered synonymous but once examined more closely two different levels of synonymy become noticeable. Firstly, there is a broader level of functional synonymy in which all of these constructions serve the same purpose, i.e. expressing either a very large or a very small indeterminate quantity. And secondly, a narrower level of conceptual synonymy in which only metaphorical expressions from the same source domain(s) can be interchangeable. Thus, both *armies of lawyers* and *legions of lawyers* are attested, recurrent and interchangeable but not *\*floods of lawyers*, or *a flood of information* and *a stream of information* but not *\*a legion of information*. This paper addresses the major issue of contextual synonymy in typological comparison. We claim that these quantifiers can be interchangeable in discourse only at the second level of synonymy. We explore the actual uses of these expressions in a total of six European languages, namely English, French, Italian, Portuguese, Romanian and Spanish. Our claims are borne out by substantial corpus evidence of a large number of occurrences of this open class of quantifiers. We argue that these conventional metaphorical patterns are shared by the aforementioned languages and that they tend to collocate with the same nominal bases. Special attention is drawn to the contrastive analysis of a selection of quantifiers which appear to be recurrent and systematic in all six languages. We show that metaphorical phrasal quantifiers are, for the most part, not language-specific as illustrated in the table below.

However, a detailed frequency analysis of the overall paradigm of quantifiers reveals interesting information as to marked combinatory preferences across languages. In English *a barrage of criticism* and *a wave of criticism* are far more frequent than *a shower of criticism* whereas in French and in Romanian *une pluie de critiques* and *o ploaie de critici* are, respectively, the preferred realisations. Spanish, Italian and Portuguese, in their turn, favour the metaphoricity of words such as *ola, oleada, onda, ondata, vaga*, all of them meaning *wave*.

### Bibliography

- Croft W. 1990: *Typology and Universals*. Cambridge: Cambridge University Press.
- Hanks P. 2006: Metaphoricity is gradable. *Corpus-based Approaches to Metaphor and Metonymy*, A. Stefanowitsch & S. Th. Gries (eds), Berlin/New York: Mouton de Gruyter, 17-35.
- Heine B. & Kuteva T. 2006: *The Changing Languages of Europe*. Oxford: Oxford University Press.
- Lakoff G. & Johnson M. 1980: *Metaphors We Live By*. Chicago/London: The University of Chicago Press.
- Leclère Ch. & Brisbois-Leenhardt J. 2004: Synonymie de mots et synonymie de phrases: une approche formelle. *Lexique, Syntaxe et Lexique-Grammaire*, Ch. Leclère, É. Laporte, M. Piot & M. Silberztein (eds), Amsterdam/Philadelphia: John Benjamins, 389-404.
- Renouf A. & Sinclair J. M. 1991: Collocational Frameworks in English. *English Corpus Linguistics*, K. Aijmer & B. Altenberg (eds), New York: Longman, 128-143.

<b>English</b>	<b>French</b>	<b>Italian</b>	<b>Portuguese</b>	<b>Romanian</b>	<b>Spanish</b>
A sea of people	Une mer de gens	Un mare di gente	Um mar de gente	O mare de oameni	Un mar de gente
An army of lawyers	Une armée d'avocats	Un esercito di avvocati	Um exército de advogados	O armata de avocati	Un ejército de abogados
A mountain of evidence	Une montagne de preuves	Una montagna di prove	Um monte de provas	Un munte de dovezi	Una montaña de pruebas

## Beyond the Synset: Synonyms in Collaboratively Constructed Semantic Resources

We present a comparative analysis of synonyms in collaboratively constructed and linguistic lexical semantic resources and its implications for NLP research. Our focus is on the Wiki-based resources constructed mostly by non-experts on the Web which rely on user collaboration for quality management, as opposed to conventional sources of synonyms such as WordNet or thesauri. The most prominent examples are Wikipedia<sup>1</sup> (a free Encyclopedia) and its dictionary spin-offs Wiktionary<sup>2</sup> and OmegaWiki<sup>3</sup>, where the latter has a strong focus on crosslinguality. We will examine three major ways how synonyms emerge in these resources, all of which imply a different operational definition of synonymy. We will then show how these synonyms can be mined and used building upon previous research in this field ((Zesch, Müller & Gurevych, 2008), (Wolf & Gurevych, 2010)), and we will also examine what theoretical conclusions about the notion of synonymy can possibly be drawn from our examinations.

The first part is the explicit encoding of synonymy, for example a link between word senses in Wiktionary, where it can be argued that the user community agrees that they are synonymous; this gives rise to a new notion of cognitive synonymy which is anchored in the “collective mind”. Following earlier work considering German resources (Meyer & Gurevych, 2010), we analyze in detail how synonyms are dealt with in different English resources, what problems arise for their exploitation (e.g. due to inconsistencies) and how this compares to conventional lexical resources.

The second part is the implicit encoding of synonyms, e.g. deducing synonymy through a transitive relation between two senses in Wiktionary. Another example is the redirect/link anchor structure in Wikipedia. Here, it can be claimed that synonyms link to the same article. Contrary to previous work (Nakayama et al., 2008), we show that this claim does not really hold, but some interesting observations can be made regarding the link structure and how it relates to the idea of synonymy. We also examine how links lead to insights about capital or subordinate traits of “distant relatives” (cf. (Cruse, 1986)) which might give us a better idea of why words are perceived as similar.

The third part does not rely on the structure of the resources but on the inference of synonymy from context. Two examples are mining synonyms from the Wikipedia revision history (cf. (Nelken & Yamangil, 2008)) as well as from an aligned corpus of Wikipedia and Simple Wikipedia. The hypothesis is that (apart from spelling corrections) terms could be synonyms if they have been replaced by each other in an article’s history (Kulesa, 2008) or if they are used interchangeably in the “normal” and “simple” versions of an article (Zhu, Bernhard, & Gurevych 2010). Both examples relate to the notion of propositional synonymy, but replacement of terms might also imply that they were deemed invalid somehow; this observation could be another path for future research.

To substantiate our work, we give illustrative examples of synonyms collected from the examined resources, and we provide statistical evidence about their structure and content.

---

<sup>1</sup> <http://www.wikipedia.org/>

<sup>2</sup> <http://www.wiktionary.org/>

<sup>3</sup> <http://www.omegawiki.org/>

## Bibliography

- Cruse, D.A. 1986: *Lexical Semantics*. Cambridge Textbooks in Linguistics, Cambridge University Press.
- Kulesa, S. 2008: *Mining Wikipedia's Revision History for Paraphrase Extraction (Master Thesis)*. Technische Universität Darmstadt.
- Meyer, C., & Gurevych, I. 2010: Worth its Weight in Gold or Yet Another Resource — A Comparative Study of Wiktionary, OpenThesaurus and GermaNet. *Proceedings of the 11th International Conference on Intelligent Text Processing and Computational Linguistics*, (pp. 38-49). Iași, Romania.
- Nakayama, K., Pei, M., Erdmann, M., Ito, M., Shirakawa, M., Hara, T., 2008: Wikipedia Mining: Wikipedia as a Corpus for Knowledge Extraction. *Proceedings of Annual Wikipedia Conference (Wikimania)*.
- Nelken, R., & Yamangil, E. 2008: Mining Wikipedia's Article Revision History for Training Computational Linguistics Algorithms. *Proceedings of the Wikipedia and AI Workshop at the AAAI Conference*. Chicago, USA.
- Wolf, E., & Gurevych, I. 2010: Expert-Built and Collaboratively Constructed Lexical Semantic Resources for Natural Language Processing. *Language and Linguistics Compass*. (to appear)
- Zesch, T., Müller, C. & Gurevych, I. 2008: Extracting Lexical Semantic Knowledge from Wikipedia and Wiktionary. *Proceedings of the 6th International Conference on Language Resources and Evaluation*. Marrakech, Morocco.
- Zhu, Z., Bernhard, D. & Gurevych, I. 2010: A Monolingual Tree-based Translation Model for Sentence Simplification. *Proceedings of The 23rd International Conference on Computational Linguistics*, Bei Jing, China. (to appear)

## Synonymy and Near-Synonymy in Deep Lexical Semantics

In an effort called “deep lexical semantics” (Hobbs, 2008), we characterize the meaning of word senses in terms of axioms anchored in core theories that explicate basic concepts. For example, a general senses of “give” can be defined by the axiom

```
(forall (x y z)
  (iff (give-vb1 x y z)
    (exist (e1 e2)
      (and (cause x e1)(changeTo' e1 e2)(have' e2 z y))))))
```

That is, **x** gives **y** to **z** (in WordNet’s verb sense 1) if and only if **x** causes a change to a situation in which **z** has **y**. The predicates **cause** and **changeTo** are explicated in a core theory of events and event structure (cf. Hobbs, 2005), and the predicate **have** is explicated in Hobbs (2008). (Our treatment of event verbs obviously owes much to the extensive body of work on lexical decomposition that began with Gruber (1965).)

In this framework, the set of word senses of the same word normally constitutes a radial structure, where the link between one word sense and the next is an incremental change in the corresponding axiom. For example, the most general sense of “hit” in WordNet is sense 4, meaning a change to being at a real or metaphorical location, as in “The thermometer hit 90 degrees.” The less general sense 2 adds to sense 4 that there was a sudden impact, as in “The car hit a tree.” Sense 1 adds to sense 2 that the impact causes motion, as in “He hit the ball out of the park.”

This framework suggests a very natural characterization of synonymy and near-synonymy. Synonymy occurs when two word senses of different words happen to be characterized by the same axioms. Near-synonymy occurs when the associated axioms differ only by a proposition or two, similarly to closely related word senses. For example, WordNet sense 1 for “provide” has the axiom

```
(forall (x y z)
  (iff (provide-vb1 x y z)
    (exist (e1 e2)
      (and (cause x e1)(changeTo' e1 e2)(have' e2 z y)(need z y))))))
```

which differs from “give” (vb1) only in that what is provided is needed, whereas what is given doesn't have to be. They are near-synonyms.

The axioms that explicate the meanings of words can capture subtle distinctions of meaning, but only if the appropriate core theories have been explicated. In Edmunds and Hirst's (2002) example, a blunder and a lapse are both errors but a blunder is caused by stupidity and a lapse by neglect. Thus, we need to have explicated a commonsense theory of intelligence, attention and other aspects of cognition (cf. Hobbs and Gordon, 2005). In the full paper we give more examples of closely related nouns.

But the characterization of near-synonyms as differing only by a proposition or two may be too broad. “raise” and “lower” differ only by a proposition; both mean to move something, but in raising the direction is up and in lowering it is down. Similarly, “raise” and “rise” differ only in one proposition, as raising is causing to rise. Are these pairs near-synonyms?

In the full paper we consider structural conditions on axioms that narrow the definition of near-synonyms. But it is more likely that near-synonymy is not a natural kind about which people have firm intuitions. Deep lexical semantics leads to a precise characterization of the relation between two similar words, and this is probably more important than judging whether or not they are near-synonyms.

## Bibliography

- Edmonds, Philip, and Graeme Hirst, 2002: Near-synonymy and lexical Choice. *Computational Linguistics* Vol. 22, No. 2, pp. 105--144.
- Gruber, Jeffery C., 1965. *Studies in Lexical Relations*, unpublished Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, Massachusetts.
- Hobbs, Jerry R., 2005: Toward a Useful Concept of Causality for Lexical Semantics. *Journal of Semantics*, Vol. 22, pp. 181-209.
- Hobbs, Jerry R., 2008: Deep Lexical Semantics. *Proceedings of the 9th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2008)*, Haifa, Israel, February 2008.
- Hobbs, Jerry R., and Andrew S. Gordon, 2005: Encoding Knowledge of Commonsense Psychology. *Proceedings, 7th International Symposium on Logical Formalizations of Commonsense Reasoning*, Corfu, Greece, pp. 107-114, May 2005.

## Observations on Synonymy and Lexical Variation of Body-part Idioms in German

In general, synonymy is understood as referring to similarities of sense between lexical items. However, the potential synonymy of idioms has been ignored in semantics, although it is implicated, for instance, in the practice of cross-referencing in special idiom dictionaries (Duden 2008, Schemann 1993). Despite extensive studies on the prototypical properties of idioms in different languages, we can conclude that only little attention has been paid to potential synonymy of idioms. Hence, the emphasis of the current literature suggests that synonymy is restricted to single words (cf. Crystal 2003). In brief, many German studies have taken notice of the ongoing large-scale electronic lexicography (e.g. DWDS-Korpus) in order to disclose contextual conditions of idioms (cf. Hümmer 2004). Despite these and further studies, it has often been emphasized that much more, for instance lexical information, is needed.

In general, lexemes denoting parts of the body (so-called body-part idioms) appear frequently in German idioms. According to Heringer (2004: 175), approximately 20 per cent of the idioms in German contain a body-part noun as their constituent, e.g. *einen kühlen Kopf bewahren* 'to keep a cool head'. Moreover, idioms that contain (at least) one body-part noun, account for the biggest group of German idioms in general (Schemann 1993). First, prototypical idioms are not only syntactically but also lexically restricted expressions containing more than one word. Second, idioms are per definitionem non-compositional, i.e. the meaning of an idiom cannot be derived from the meanings of its components. In addition, idioms are institutionalized, i.e. they are known to native speakers of the language and can be found in dictionaries (Nunberg et al. 1994). Nonetheless, there have also been attempts to make idioms relatively flexible (cf. Penttilä 2006) and studies related to these aspects have shown that idioms include more variation than assumed hitherto and much of the variation is systematic (Fellbaum 2007). The present study has three major aims. First, lexical variation of German body-part idioms will be examined. The ultimate goal is to find out whether the testing of native-speakers of German could provide novel aspects of the variation of body-part idioms. Second, the crucial question to be answered is: Is there only lexical variation, or, in general, should we consider this as synonymy? In other words, the present study sets out to see whether synonymy and lexical variation are measurable properties of body-part idioms in German. In brief, this paper tries to shed some light on the question where to draw the line between synonymy and lexical variation in German idioms.

### Bibliography

- Crystal, David 2003. *A Dictionary of Linguistics and Phonetics*. Malden, MA: Blackwell Pub.
- Duden. *Redewendungen. Wörterbuch der deutschen Idiomatik* 2008. Herausgegeben von der Dudenredaktion. 3., überarbeitete und aktualisierte Auflage. (Der Duden in 12 Bänden. 11). Mannheim: Dudenverlag.
- DWDS-Korpus: <http://www.dwds.de>
- Fellbaum, Christiane (ed.) 2007. *Idioms and Collocations: Corpus-based Linguistic and Lexicographic Studies*. London: Continuum.
- Heringer, Hans Jürgen 2004. *Interkulturelle Kommunikation. Grundlagen und Konzepte*. 2., durchgesehene Aufl. (UTB 2550: Sprachwissenschaften). Tübingen und Basel:

- Francke.Hümmer, Christiane 2004. A Corpus Based Approach to Near Synonymy of German Multi-Word Expressions. In: Petr Sojka, Karel Pala & Pavel Smrz (eds.) *Proceedings of the Second International WordNet Conference*. Brno: Masaryk University.
- Nunberg, Geoffrey, Ivan A. Sag & Thomas Wasow. 1994. Idioms. *Language*, 70, 3, 491-538.
- Penttilä, Esa 2006. *It Takes an Age to Do a Chomsky: Idiomaticity and Verb Phrase Constructions in English*. PhD thesis, University of Joensuu.
- Schemann, Hans 1993. *Deutsche Idiomatik. Die deutschen Redewendungen im Kontext*. Stuttgart: Klett.

### **The (near-)synonyms *begin* and *start*: evidence from translation corpora**

Non-finite complement constructions with the English matrix verbs *begin* and *start* exhibit a degree of similarity, if not outright synonymy, along two axes. On the one hand there is the contrast between the two matrix verbs. As Dixon puts it, “In many sentences *start* and *begin* may be substituted one for the other with little or no change in meaning” (Dixon 2005: 181). On the other hand, it is often difficult to discern substantive differences in meaning between pairs of constructions containing the same matrix verb but differing in choice of complement form. Thus Quirk *et al.* (1985: 1192), commenting on the choice of *to*- infinitive or *-ing* complements, note that in the case of many examples “there is no observable difference of meaning between the constructions”. Freed also underlines the similarity between the various constructions in certain contexts, writing “there are cases or contexts in which the difference between them does not matter” (Freed 1979: 75). This impression of mutual entailment is also emphasised by Egan (2008), who writes “There are many cases where a *begin* construction may be seen to entail a *start* construction, a *to infinitive* construction an *-ing* construction, and *vice versa* (i.e. there are many utterances of the type ‘she began to do X’ which, if true, guarantee the truth of ‘she started to do/doing X’)” (Egan 2008: 256).

Although all four of the authors mentioned above stress the mutual substitutability of the four constructions in certain contexts, they also all maintain that there exist subtle distinctions between them, in other words contexts in which the substitution of one for the other would lead to differences in interpretation. There is, however, considerable disagreement between these and other authors as to the exact nature of the relevant distinctions (see also, for example, Mair 2003).

In this paper we attempt to pin down more closely the extent and nature of the similarities and differences between the various constructions containing *begin* and *start* using translation corpora, a source of evidence which is as far as we know not previously mined. We look at how constructions containing *begin* and *start* are translated into Norwegian, a language which contains the cognate verbs *begynne* and *starte*. Our data for this part of the study come from the English-Norwegian Parallel Corpus (see Johansson 2007), which contains 426 tokens of *begin* and 277 of *start*. We then use the Oslo Multilingual Corpus to examine translation equivalents of constructions containing the Norwegian matrix verbs *begynne* and *starte* into French, German and English. We investigate the extent to which cross-linguistic similarities and differences in choice of translation options may mirror putative lexical, constructional, functional and formal similarities and differences in the original expressions. In so doing, we utilize translation corpora to shed light on the degree to which *begin* and *start* may be considered synonymous.

#### Bibliography

- Dixon, R. M. W. (2005). *A semantic approach to English grammar* (2nd. ed.). Oxford: Oxford University Press.
- Egan, T. (2008). *Non-finite complementation: a usage-based study of infinitive and -ing clauses in English*. Amsterdam: Rodopi.
- Freed, A. F. (1979). *The Semantics of English aspectual complementation*. Dordrecht: Reidel.
- Johansson, S. (2007). *Seeing through Multilingual Corpora : On the use of corpora in contrastive studies*. Amsterdam: John Benjamins.
- Mair, C. (2003). Gerundial complements after *begin* and *start*: Grammatical and sociolinguistic factors, and how they work against each other. In B. Mondorf & G. Rohdenburg

(Eds.), *Determinants of grammatical variation in English* (pp. 329-345). Berlin: Mouton de Gruyter.

Quirk, R., Sidney Greenbaum, Geoffrey Leech & Jan Svartvik (1985). *A Comprehensive grammar of the English language*. London: Longman.

## Synonymy in specialized communication – a terminological approach

Specialized communication sets different requirements on the language than everyday communication. Language for specialized purposes (LSP) is a tool for classifying, describing and mediating information within different subject fields. The language used in LSP communication should thus be characterized by unambiguousness, precision and logicity. The ideal situation for achieving precise and unambiguous communication is monosemy, i.e. one concept is represented by only one term. But it is obvious that polysemy, homonymy and synonymy have a significant role also in the communication between experts within the same subject field in a similar way as linguistic variants in general language. Existence of synonymous expressions is especially combined with fields where the development is fast and where overall coordination still is missing.

In this paper we first focus our attention on how synonymy is discussed in terminological literature. Terminology science is a field of study of concepts and terms based on systematic concept analysis. In general synonymy is seen as an unwanted phenomenon in terminology because it can prevent mutual understanding within the subject field. (Arntz & Picht 1989). There seem to exist somewhat different views among the researchers within the field of terminology. According to Lotte (1993) synonymy is actually dangerous: “Die Synonymie von Termini ist auch gefährlich, und sie sie muß eliminiert werden, d.h. jedem Begriff darf nur ein Terminus zugeordnet werden”. Irgl (1989) represents another view by pointing out that synonymy is caused by objective and subjective factors. This can be seen e.g. in the language of economics, where synonymy is surprisingly common. The developer of the General theory of terminology in the 1930s, Wüster (1991), states that an absolute requirement on precision is a theoretical ideal and gives a categorization of synonymies based on meaning (Sachbedeutung, Mitbedeutung, Sach- und Mitbedeutung) and uniformity (Spalt- und Systemsynonyme).

Secondly we analyze and categorize different kinds of synonymous expressions in vocabularies compiled using terminological methods. We intend to examine the term structures and the order of terms in terminological records including one or more qualified synonymic expression to the entry term both within and across languages. The synonymous terms in one language are studied also in relation to the aim or focus in the actual context due to so called dimensions (cf. Bowker1997) The terms entrance floor and ground floor, for example, represent the same concept but the choice between them depends on the point of view (entrance to the building vs. number of stocks). The analysis of the definitions in the vocabularies is based on a comparison between the terms (entry term, synonym/s) and the essential and delimiting characteristics in the description including the note section.

### Bibliography

- Arntz, R. & H. Picht (1989). *Einführung in die Terminologearbeit*. Hildesheim etc.: Georg Olms Verlag.
- Bowker, L. (1997). Your say “flatbed colour scanner” and I say “colour flatbed scanner” A descriptive study of the influence of multidimensionality on term formation and use with special reference to the subject field of optical scanning technology. In: *Terminology: international journal of theoretical and applied issues in specialized communication* vol 4 (2), 275–302. Amsterdam: Benjamins.

- Irgl, V. (1989). Synonymy in the language of business and economics. In: *Special language: from humans thinking to thinking machines*. 275–282. Eds. Laurén Ch. & M. Nordman. Clevedon, Philadelphia: Multilingual Matters.
- Lotte, D.S. (1993). Aufgaben und Methoden zur Regelung von technischer Terminologie. In: *Ausgewählte Texte zur Terminologie*. 190–220. Hg. Laurén, Ch. & M. Nordman. Wien: Termnet.
- Wüster, E. (1991). *Einführung in die Allgemeine Terminologielehre und Terminologische Lexikographie*. 3. Auflage. Bonn: Romanistischer Verlag.

## Translation variants as Another Type of Synonymy

Synonymy is a very important cognitive concept for describing a language lexicon and grammar both from the systemic-semantic and functional-pragmatic aspects since it enables linguists to detect minor semantic differences and combinability peculiarities of semantically similar words and constructions (cf. quick – fast – rapid – swift; The hunter killed the bear – The bear was killed by the hunter; He always comes to work in time – He is never late for his work; John bought the book from Mary - Mary sold the book to John) and, consequently, to find out common and differentiating contexts of semantically similar words and constructions. A broader perspective in approaching the concept of synonymy can be also achieved by including in the domain of synonymy analysis several translation variants of one and the same original text. The paper deals with the main types of English translation variants (both phrasal and clausal) of Russian fiction written by Fyodor Dostoyevsky, Leo Tolstoy, Anton Chekhov and Mikhail Bulgakov. These translation variants may be regarded synonymous since they convey one and the same meaning of the original text. Selected passages of the original texts contain non-trivial (either polysemous or occasional) words and phrases, as the verbs in the examples below:

- (1) On **blagopolučno izbegnul** vstreči s svojeju hoz'ajkoj na lestnice (Dostoyevsky. Crime and punishment) –
  - a. He **had successfully avoided** meeting his landlady on the staircase (C. Garnett);
  - b. He **was lucky to avoid** a meeting with his landlady on the stairs (D.Magarshack);
  - c. He **had been lucky** enough **to escape** meeting his landlady on the staircase (J.Katzer);
  - d. He **had succeeded in avoiding** an encounter with his landlady on the stairs (McDuff);
  - e. He **was fortunate** enough **not** to meet his landlady on the stairs (Anonymous translation; London 1997);
- (2) Vs'o **smeshalos'** v dome Oblonskih (Leo Tolstoy. Anna Karenina) –
  - a. Everything **was in confusion** in the Oblonskys' house (C. Garnett);
  - b. Everything **was in confusion** in the Oblonskys' household (D.Magarshack);
  - c. All **was confusion** in the Oblonskys' house (Pevear / Volokhonsky);
  - d. Everything **had gone wrong** in the Oblonsky household (R. Edmonds);
  - e. Everything **was upset** in the Oblonskys' house (L. and A. Maude);
  - f. The Oblonsky home **was in turmoil** (M. Wettlin);
  - g. Everything at the Oblonskys' **was topsy-turvy** (J. Carmichael).

The analysis of semantically identical or similar translation variants of the types presented above allows to take into consideration phrases and clauses expressing in different ways the same meaning of the original text and containing words which express similar concepts but belong to different parts of speech (cf. a Russian adverb **blagopolučno** and an English verb **succeed** in example (1) and a Russian verb **smeshalos'** and an English noun **confusion** in example (2)).

The second part of the paper analyses different English translations of Russian words expressing emotions with special attention to cases when original Russian words are rendered by the English corresponding lexemes belonging to another part of speech. The analysis included 320 Russian sentences and phrases of five Chekhov's plays (including stage directions), and since the number of English translations equaled five to six, the quantity of English translation variants was not less than 1650 sentences and phrases. In 110 English translations (approximately 7 % of all the translations) various changes in part-of-speech status of the original lexemes were registered. It should be noted that all twelve theoretically possible changes of part-of-speech status of the original Russian lexemes were present in 110 English translations: 1) Noun in the Russian text (RT) à Verbal form in the English translation (ET) (8 cases), 2) Noun in RT -> Adjective in ET (14 cases), 3) Noun in RT -> Adverb in ET (13 cases), 4) Verb in RT -> Noun in ET (6 cases), 5) Verb in RT -

> Adjective in ET (9 cases), 6) Verb in RT -> Adverb in ET (6 cases), 7) Adjective in RT -> Noun in ET (3 cases), 8) Adjective in RT -> Verb in ET (2 cases), 9) Adjective in RT -> Adverb in ET (7 cases), 10) Adverb in RT -> Noun in ET (13 cases), 11) Adverb in RT -> Verb in ET (10 cases), 12) Adverb in RT -> Adjective in ET (19 cases). It turns out that Russian adjectives are most stable (only 12 cases of part-of-speech status change), whereas Russian adverbs are most changeable (42 cases of part-of-speech status change, or about 38 % of all cases involving part-of-speech change). So, English words expressing the same emotive concepts but belonging to different parts of speech can be regarded as either synonymous or quasi-synonymous.

## **Autre – different vs. other – different: A contrastive approach**

In French, the relational adjectives *autre* 'other' and *différent* 'different' are binary predicates that establish a relationship of 'difference' between two entities (e.g. Noailly, 1999; Schnedecker, 2002). As might be expected, the two terms can sometimes be used interchangeably with very little, if any, contrast in meaning:

- (1) Marie est allée au Louvre, mais Pierre a préféré visiter un (autre) musée (différent).  
Mary went to the Louvre, but Pierre wanted to visit (another) (a different) museum.

Moreover, both adjectives establish a relationship between the phrase containing them and a second element elsewhere in the context (e.g. Van Peteghem, 1997), for example in their complement or to their left. In example (1), one of the arguments is part of the phrase containing the adjective (i.e. *musée*), whereas the second argument is to the left of this phrase (i.e. *Louvre*). In (2), the second argument is found in the complement to the phrase containing the adjective:

- (2) J'ai choisi un sac différent du tien/un autre sac que toi.  
I have chosen a different purse from yours/another purse than you.

These similarities can be seen to account for the fact that, in monolingual French dictionaries (e.g. *Le grand Robert* or the *Trésor de la langue française*), the entries for each term refer readers to the other.

However, sentences such as (1) where the two terms can be used interchangeably are in fact rare. There are a number of contexts in which only one of the two terms is possible without a change in interpretation:

- (3) Son mégot à peine écrasé, il alluma une autre cigarette/\*une cigarette différente.  
He had scarcely put out his cigarette when he lit another/a different one.

- (4) J'ai acheté trois robes différentes/#trois autres robes pour partir en vacances.  
I bought three different dresses/three other dresses for my vacation.

This brings us to the goal of the present paper. Using constructed and attested examples, we will describe the semantic and referential properties of *autre* and *différent* on the basis of the different possibilities for syntactically encoding their arguments. More specifically, it will be shown that *différent* functions as a symmetrical adjective with a reciprocal sense whose base meaning allows it to act as a determiner. *Autre*, on the other hand, appears to function more as a comparative than as a typical adjective. It has a negative meaning in that it excludes a given referent as a potential referent for the phrase in which it appears.

These conclusions will then be compared with the data concerning the English adjectives *other* and *different*. While superficially similar, both semantically and syntactically, the uses of the English adjectives do not line up perfectly with the French data, as shown in (5):

- (5) J'ai choisi un autre sac que toi.  
\*I have chosen another purse than you.

Thus, in English, it appears that the distribution of *other* is more restrained than in French, revealing that the terms are not synonymous cross-linguistically.

#### Bibliography

- Breban, T. 2010. *English adjectives of comparison: Lexical and grammaticalized uses* (Topics in English Linguistics). Berlin: Mouton de Gruyter.
- Dotlačil, J. 2010. Anaphora and Distributivity. A study of *same*, *different*, reciprocals and *others*. LOT dissertation series 23.
- Noailly, M. 1999. *L'adjectif en français*. Paris: Ophrys.
- Schnedecker, C. 2002. Présentation: Les adjectifs « inclassables »: des adjectifs du troisième groupe? *Langue française* 136: 3-19.
- Van Peteghem, M. 1997. Mécanismes anaphoriques sous-jacents aux 'indéfinis' *autre* et *même*. In W. DeMulder, L. Tasmowski-DeRyck and C. Vetters (eds), *Relations anaphoriques et (in)coherence* (pp. 187-200). Amsterdam: Rodopi.

## What 's in a synonym?

Synonymy lies at the heart of (lexical) semantics. The interest in this phenomenon, as visible in the history of linguistics, can be explained by the fact that insights in the equality or similarity of meaning may shed light on meaning itself. Nowadays, it is possible to obtain huge quantities of (NLP) applicable information about word usage by analyzing large data collections. This information, however, does not necessarily help us understand what is required, in terms of linguistic theory, for two words to be (nearly) synonymous. In our research, we adopt an enriched version of a decompositional approach to word meaning. Our main research question concerns the relative influence of six well defined word properties on (experimentally obtained) synonymy judgements.

In our research, we define a set of factors relevant to word meaning. These factors, envisaged as word properties, are not novel. In some form, they were introduced in studies of a syntactic, semantic or pragmatic nature. The composition of these factors in a instrument for measuring synonymy, however, is original. Three of them are coined paradigmatic, since they can be seen as a choice from a set of alternatives. The paradigmatic ones are: conceptual properties, register, and reading. The three others are labelled syntagmatic, since they refer to some structural pattern: selection restrictions, telic role, and agentive role (cf. Pustejovsky, 1995).

**The theoretical part:** A set of candidate word pairs is selected, existing of more or less synonymous words. For each word property, a value is attributed to the candidate word pairs, reflecting the degree of similarity of the pair. In order to warrant intersubjectivity in the analyses, a reliable dictionary of synonyms is used (...). For each word pair, the combined similarity scores for all word properties is defined as the score on semantic distance. This score ranges from zero (strong synonymy) to six (weak synonymy).

**The experimental part:** In three experiments, the correlation has been evaluated between criteria and distance score on the one hand, and judgements by native speakers on the other hand. In designing these experiments, special attention was given to three methodological issues: the selection of the set of word pairs, the linguistic context in which the words were to be presented, and the exact phrasings of the judgements asked for.

Judgements of synonymy appeared to be closely linked to the syntagmatic properties of words, with selection restrictions as the most determinative factor. This factor was followed by the syntagmatic properties telic role and agentive role. The paradigmatic properties played a surprisingly modest role.

Evidently, in deciding on synonymy, it's the deeper layers of linguistic knowledge about structural patterns that are consulted. The results seem to agree with the current view (e.g. Jackendoff, 2007), that the borderline between grammar and lexicon is not as rigid as we used to think. Moreover, the results engender a new set of questions about word meaning and the human mind. Why does the pragmatic property register play such a trivial role? What makes the telic role so important? Why is it more important than the agentive role?

## Re-thinking synonymy: a cognitive science perspective

The goal of my talk is to give a cognitive science perspective on some founding questions concerning the nature of synonymy, including the cognitive definition of synonymy, its potential usefulness as a theoretical concept in cognitive science, and translatability (similarity of meaning) across languages. My theoretical starting points are the philosophical notions of synonymy, the holistic view of meaning, and the indeterminacy of translation thesis ('the Gavagai argument') formulated by the philosopher W.V.O. Quine (1951, 1960). I will first briefly present these points, after which I will introduce some (partial) objections to these views. The objections also include two examples from current empirical language research where the notions of lexical as well as constructional synonymy would seem to play a theoretically important role.

The first example comes from the scientific study of bilingualism, where the theoretical concepts of synonymy and/or translational equivalence could be seen as potentially foundational. Consider, for example, the English words *pizza* / *boy* / *cognitive science* and their Finnish translation equivalents *pitsa* / *poika* / *kognitiotiede* and imagine then a fully bilingual speaker of English and Finnish. When words from two different languages are compared with each other this way, no usage-based techniques can be applied for identifying similarity of word meaning. However, Francis (2005) has argued, that for translation equivalents in general, the current empirical evidence from psycholinguistic studies on bilingualism strongly favours the scientific hypothesis of shared semantic systems and shared semantic/conceptual representations.

The second example comes from cognitive neuroscience. Consider the following Finnish sentences (and their English counterparts):

- (1) Poja-t            osta-isi-vat            pitsa-t            huomenna  
    boy-NOM.PL buy-COND-3PL pizza-ACC.PL tomorrow  
    '(The) boys would buy (the) pizzas tomorrow'
- (2) Pitsa-t            oste-tta-isi-in            huomenna poikien toimesta  
    pizza-ACC.PL buy-PASS-COND-AGR tomorrow by the boys  
    '(The) pizzas would be bought tomorrow by (the) boys'.

Sentence (2) is an example of the Finnish morphological ('impersonal') passive construction. The passivisation in Finnish is a morphosyntactic valency-changing operation, where the initial subject NP of the active sentence is either demoted or deleted, and the initial object NP (if there is any) is moved to sentence-initial position. At the same time the finite verb undergoes major morphological changes, including insertion of the passive marker *-(t)ta* and the suffix *-Vn*, which is a special grammatical person marker that appears in finite passive constructions. In a recent experimental study (Palolahti et al., in preparation) real-time semantic and grammatical processing of active and passive Finnish sentences were studied using event-related potentials (ERPs). The obtained results were then compared to the results of a previous ERP study by Kim & Osterhout (2005), where English periphrastic passives and active English sentences were used as experimental stimuli. Interestingly, the ERP results from these two studies were observed to be practically identical, regardless of the differences between languages and particular linguistic constructions in question.

## Bibliography

- Kim, A. & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, 52, 205-225.
- Francis, W.S. (2005). Bilingual semantic and conceptual representation. *Handbook of bilingualism: psycholinguistic approaches*. J. F. Kroll & A. M. B. de Groot (eds.). Oxford University Press.
- Quine, W.V.O. (1951). 'Two dogmas of empiricism', *The Philosophical Review*, 60, January 1951.
- Quine, W.V.O. (1960). *Word & Object*, Cambridge: M.I.T. Press.

## Resemblance and variation of causatives expressing social relations

In my paper, I discuss closely related languages Estonian and Finnish and their ways to encode and conceptualize social dominance. By social dominance I mean dominance and causation between human actors. There are both similar and different linguistic means in Estonian and Finnish for expressing social causation – synthetic and analytic. From the typological point of view, morphological causative derivation is not found in all languages. The morphological causatives derived with the suffix *ttA* are a widespread and idiomatic means of expression in Finnish; compared to that, the use of causative derivation with the corresponding morpheme *ta* is in Estonian estimated to be more restricted (Kasik 1989, 2001). However, some of the morphological causatives are used in rather similar contexts in these languages, compare the examples (1-2):

- (1) Hallitus juokсуттаа maanviljeljää (Fi.)  
Valitsus jookсутаб põllumeest (Est.)  
government-nom run-cause-3sg farmer-part  
'The government is running the farmers around'
- (2a) Liverpool kyykytti Chelsea 3:1 (Fi.)  
Liverpool-nom squat-caus-pst-3sg Chelsea-acc 3:1  
'Liverpool beat Chelsea 3:1'
- (2b) Liverpool seljatas Chelsea 3:1 (Est.)  
Liverpool-nom back-caus-pst-3sg Chelsea-acc 3:1  
'Liverpool beat Chelsea 3:1'

My presentation will focus on the morphological causatives expressing social dominance in these languages and leave the lexical and analytic outside the scope. How are social dominance relations encoded in these languages, what are the similarities and differences? How to capture the relations encoded between humans and the social hierarchies? What is the role of responsible conventional behaviour in causative constructions? I argue that causative verbs can have specialized uses where the expressions of social relations and attitudes have important roles. I approach the derivatives within the framework of conceptual semantics and argue for a prototype semantic interpretation for these derivatives. I suggest that the abstract type of prototype templates can be used for comparison of the verbs as well as for revealing the constructional extension and idiosyncrasy. I assume that there are partly overlapping prototype-based patterns and particular causative constructional patterns in these languages.

### Bibliography

- Kasik, Reet 1996. *Eesti keele sõnatuletus*. Tartu Ülikooli kirjastus.  
Kasik, Reet 2001. Analytic causatives in Estonian. *Estonian: typological studies V* (Mati Ereht, ed.).  
Publications of the department of Estonian of the University of Tartu.

## Re-Thinking Synonymy – Finnish Terminative Particles *asti* and *saakka* as Perfect Synonyms

The term **terminative** refers to a case marking typical in some Finno-Ugric languages, such as Estonian, Hungarian and the Permic languages. Terminative is generally used to designate a morphological case that indicates a border or a limit of motion or other kind of relation. Finnish does not have a terminative case; instead terminativity is generally expressed with the local cases and so-called terminative particles, such as *asti* and *saakka* ('until, all the way to, since, all the way from, as far as'). These two particles can be seen as very close synonyms (Kangasniemi [1997: 42] calls them perfect synonyms). If two words have same meaning, use and function, if they can be used in same connotative, affective environments, and if their quantitative appearance in different Finnish language corpora is identical – they are almost perfect synonyms.

In my presentation I will describe the use and distribution of *asti* and *saakka* from the oldest written (Agricola's) corpora to the contemporary writers of Finnish literature today (Päiviö 2007). I will discuss the reasons, why the meaning and use of *asti* and *saakka* has remained the same, and why for example one terminative particle has not developed new connotations and extensions to their meaning and use. Especially, I am interested in the question, why one terminative particle is not enough for the Finns, and what do the speakers of Finnish think about *asti* and *saakka* introspectively?

I will base my analysis on Geeraerts (1988) article Where does Prototypicality come from? Especially on the notion of avoiding monotony (Geeraerts 1988: 226). Additionally, I will describe the history and development of *asti* and *saakka* and aim to trace the grammaticalization path of the semantic meaning of these particles. The presentation offers a diachronic viewpoint to the assumed development of the terminative particles into their present functions.

### Bibliography

- Geeraerts Dirk 1988: Where Does Prototypicality Come From? – Brygida Rudzka-Ostyn (ed.) p.207-229.
- Kangasniemi Heikki 1997: *Sana, merkitys ja maailma. Katsaus leksikaalisen semantiikan perusteisiin.*  
Oy Finn Lecture Ab, Helsinki.
- Päiviö Pia 2007: *Suomen kielen asti ja saakka. Terminatiivisten partikkelien synonymia, merkitys, käyttö ja kehitys sekä asema kieliopissa.* TYSJYKLJ 75, Turku. (The *asti* and *saakka* of Finnish. The Synonymy, Meaning, Use, Development and categorization of the Finnish terminative Particles).
- Rudzka-Ostyn Brygida (ed.) 1988: *Topics in Cognitive Linguistics.* John Benjamins, Amsterdam.

## **On constructional synonymy: disentangling the Russian comparative constructions**

Cross-linguistically, it is common to express comparison by means of a similative construction (Haspelmath and Buchholz 1998). Interestingly, Russian possesses not only a similative construction (e.g. *siât' kak zoloto* 'shine like gold'), but also a seemingly synonymous construction where the standard of comparison is marked by the instrumental case (e.g. *siât' zolotom* 'shine gold-INS'). We hypothesized that the two constructions which are syntactically distinct must also be semantically distinct. To test this hypothesis, we investigated the distribution of the similative construction and the instrumental-of-comparison construction in the Russian National Corpus (ca. 140 million words).

The results indicate that the prototypical core of the instrumental-of-comparison construction are descriptions of shape as in *letet' klinom* 'fly in V-shape' (lit. fly wedge-INS). Since shape is a major visual cue to categorization (Friedrich 1970), the role of the standard of comparison in the instrumental construction may be compared to that of numeral classifiers. Such cases do not allow a similative in Russian. Likewise, case-free languages like English and Dutch use derivational morphology (compounds, denominal verbs) rather than similatives for this kind of shape descriptions as in *pruimmondje trekken* 'purse one's lips' (lit. pull a plum-mouth).

Realizations of the instrumental-of-comparison construction beyond the domain of shape (e.g. *letet' streloj* 'fly arrow-INS'; *vyt' volkom* 'howl wolf-INS') denote a comparison between two distinct entities. In such cases, the instrumental case is seemingly interchangeable with the similative construction (cf. *letet' kak strela* 'fly like an arrow'; *vyt' kak volk* 'howl like a wolf'). However, upon closer scrutiny it appears that the comparative instrumentals inherit several crucial properties from the more prototypical shape-related instrumental, which distinguish them from the corresponding similatives. One important property of the instrumental construction is monotonicity. Shape-related instances of the instrumental construction are intrinsically monotonous – they simply describe objects in terms of shape. Similarly, the semantics of the instrumental-of-comparison beyond the domain of shape is limited to only one parameter, such as sound, manner of looking or motion. In contrast, the similative construction often denotes comparison along several dimensions (e.g. *On vël sebâ kak podlec* 'He behaved like a scoundrel' vs. *\*On vël sebâ podlecom* 'He behaved scoundrel-INS').

Another crucial property of the instrumental-of-comparison construction is visualness: the comparee and the standard must be directly observable to the human eye. This explains why we cannot use the instrumental case to express comparisons in situations which are not visually observable as in *\*greet pečkoj* (gives.out.warmth stove-INS) and *\*plavaet ryboj* (swims fish-INS). The similative must be used in such cases.

Finally, this paper will focus on the relation between the instrumental-of-comparison and other constructions constituting an instrumental network (cf. Goldberg 1995, 2006; Kay and Fillmore 1999) and argue that the instrumental-of-comparison is semantically connected with the more prototypical instrumental of main predication (e.g. *byl učitelem* 'was teacher-INS') and the instrumental of additional characterization (e.g. *rabotal učitelem* 'worked teacher-INS'). We will suggest that less prototypical predicates for expressing identity require the attributed property to be monotonous and observable to accommodate the identity relation.

## Bibliography

- Friedrich, Paul. 1970. Shape in grammar. *Language* 46(2): 379–407.
- Goldberg, Adele. 1995. *Constructions: A Construction Grammar approach to argument structure*. Chicago: University of Chicago Press.
- Goldberg, Adele. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Haspelmath, Martin & Oda Buchholz. 1998. Equative and similative constructions in the languages of Europe. In Johan van der Auwera (ed.), *Adverbial constructions in the languages of Europe*, 277–334. Berlin & New York: Mouton de Gruyter.
- Kay, Paul & Charles J. Fillmore. 1999. Grammatical constructions and linguistic generalizations: The What's X doing Y? construction. *Language* 75. 1–33.

## Evaluating and analyzing semantic similarity and sameness in studies of polysemy and synonymy

When studying synonymy, the linguistic units whose semantic values are compared are non-identical forms. These items (in boldface) may be given in context or in isolation:

- (1a) **meaning**
- (1b) **semantics**
- (2a) What is the **meaning** of study?
- (3a) What is the **semantics** of study?
- (3a) I am interested in the **meaning** of synonymy.
- (3b) I am interested in the **semantics** of synonymy.

In the case of polysemy, the object of comparison is two or more instances of one (identical) form, often by default given in a context, or at least with contexts to be imagined:

- (4a) Oh, lawyers, that's just **semantics**!
- (4b) Oh, linguists, don't we all love **semantics**?
- (5) (Instruction:) Can you think of two different meanings for the word **semantics**?

The medium of comparison is a cline from semantic sameness to semantic difference. If it is hard to find two instances where the meaning of a form is not the same, the lexical item is not polysemous. In normal cases of polysemy, different instances and uses of the lexical item have sometimes different, sometimes similar meanings, and it should be the job of the semanticist to figure out "which distinctions matter the most"; what the different meanings of the polysemous word (e.g., the word *semantics*) are:

- (6a) *semantics*, 'the linguistic study of meaning'
- (6a') in many school linguistics, an area clearly separate from syntax
- (6a'') in Cognitive Linguistics, an area intertwined with syntax, pragmatics, etc.
- (6b) *semantics*, 'the meaning content of a linguistic unit'
- (6c) *semantics*, pejoratively used to refer to unnecessary or picky exploitation of the nuances of the meanings of words

If the semantic difference between two uses of an identical-looking form is very big – i.e., there is great semantic distance – one reason may be that the case is of homonymy, not polysemy. Another reason may be that a very polysemous word has evolved into so many directions that two extremes in its semantic extension do not have anything in common.

In other words, the study of polysemy needs to take into account the concept of semantic similarity in order to investigate semantic difference. In so doing, semanticists are on similar territory when studying synonymy and polysemy. There are some psycholinguistic methods that can be used for the purpose of covering the points of similarity and difference. One is similarity rating, as employed on polysemy in example (7):

- (7) On a scale from 0 to 5, how similar or different is the meaning of *semantics* in the following examples (0 = the same, 5 = very different)?

**Semantics** studies meaning.                    0 1 2 3 4 5    Every word has **semantics**.  
Every gesture has **semantics**.                    0 1 2 3 4 5    Every word has **semantics**.

The same method can be used for synonymy:

(8) On a scale from 0 to 5, how similar or different is the meaning of the words in boldface in the following examples (0 = the same, 5 = very different)?

Every gesture has **semantics**.                    0 1 2 3 4 5    Every gesture has **meaning**.

Another method is the sorting test. Informants are given ready-made instances of one word in different contexts, and they are asked to sort these instances into different categories according to semantic similarity and semantic difference. This method is more readily available for studies of polysemy, because it is easy to work on several instances of one word, but in the case of synonymy, the items compared should include several related words in the same context, possibly concentrating on different cases of synonymy in one study, which might distract the focus of attention in the informant.

When using these methods in studies of polysemy, one has to evaluate their reliability and compatibility. The concept of semantic similarity is operationalized in various ways, and in so doing, the concept has to be evaluated more thoroughly.

One question is whether informant behavior in such experimental settings varies according to the object of study. Do different forms suggest to us that there, by default, should be difference in meaning? Do similar forms encourage us to presuppose that there should be similarity or at least relatedness in meaning?

My talk will thus address the issue of semantic similarity (and sameness) from the two angles, synonymy and polysemy, and aim at finding out whether the concept has different nuances in the two cases by necessity.

## Cross-linguistic similarity and the foreign language learner

Linguists have generally been more interested in linguistic differences than in similarities. But in order to talk about difference in a meaningful way we must first consider similarity. For both concepts a comparison is always the basis. But if no similarity at any level can be found between two concepts, a comparison will be meaningless. It is futile to try to compare a stone with an ice hockey match: there simply is no common ground for making such a comparison. **In order to establish meaningful differences there must be an underlying similarity. Similarity is basic, whereas difference is secondary.**

Similarities can be either cross-linguistic or intra-linguistic. I shall confine myself to cross-linguistic similarities, but if we consider the language learner, the two go back to the same common strategy: the learner trying to make use of prior knowledge. Learners always try to relate new information to existing knowledge, and similarity facilitates the establishment of such relations.

For the FL learner we can distinguish three different cross-linguistic similarity relations:

(1) A **similarity relation** where the learner can establish a one-to-one relationship in the target language with another unit, usually in the L1. Across related languages there will be cognates facilitating both comprehension and learning, especially at the early stages, even though some of them may later turn out to be false, or partially false cognates. Grammatical structures across Germanic languages are largely similar, with the same basic categories.

(2) A **difference relation** where the learner can perceive both similarity and difference.

Native speakers of English learning a Romance language will encounter differences and similarities in varying proportions.

(3) A **zero relation** where learners can relate few or no aspects of the target language to prior knowledge. The zero relation does not mean that they cannot find anything at all that is relevant to the L1 as the learning progresses. There are, after all, some linguistic universals common to all languages. But the level of abstraction here is so high that an average language learner cannot easily notice features that a totally different target language has in common with L1. A learner who knows only Indo-European languages and starts learning Chinese will find it difficult to relate anything to his prior knowledge.

Cross-linguistic similarities can be **actual, perceived or assumed.**

Actual similarity is a concept which at least in theory could be objectively defined by linguistic terms. Perceived and assumed similarities, on the other hand, are psycholinguistic and fuzzy by nature. Learners first perceive similarities to their L1. This can be fairly easy and accurate if the L1 is related to the target language. When similarities cannot be perceived they are merely assumed, which often leads to errors.

The different ways in which the different types of similarities affect comprehension and production and the different language areas will be outlined. There is a lot of data from investigations comparing Finnish and Finland-Swedish learners of English over more than 30 years.

## Synonymy as a useful metalinguistic pipe dream

The aim of this paper is to show (1) that strict synonymy (i.e semantic identity between two different forms) does not exist within languages and only occasionally across languages, but (2) that, provided that one considers this limit, synonymy still can be a useful tool for linguistic analysis in two ways: first, it forces linguists to explain precisely in which way two terms can be synonyms while differing (or conversely), and therefore to refine the description of meaning; second, it reveals a fundamental property of language allowing speakers to construe the same situation in alternate ways (Langacker 1991a) and, therefore, to build up equivalences within and across languages.

In order to support these claims, I will present three components of meaning by which two lexemes sharing a common referential value can differ (Robert 2008) and therefore are not full synonyms:

- (1) different polysemic networks
- (2) different connotations
- (3) different referential paths

(1) illustrates the fact that synonymy is most of the time restricted to local synonymy; it can be exemplified by the interlinguistic contrast between colour adjectives used as nouns : in English the word *greens* can refer to village commons, leafy vegetables or members of a political party but the French equivalent of greens (*les verts*), can only refer to members of a political party or to a set game pawns.

(2) is of particular importance because connotations condition the context of use as well as the pragmatic effect produced by the usage of a term; e.g. *father* to designate a priest signifies that the speaker is a practicing catholic, contrary to using *priest* or *clergyman*; using *the heat* for “policeman” signals belonging to a certain age group and general ideology including a pejorative meaning in the word.

(3) is the source of important crosslinguistic variation : languages usually choose one of the referent’s properties to designate it, for example a physical or functional characteristic, and they usually differ in their strategies to gain access to reference; thus English designates a “used car” not by the fact that it is something one buys under favorable financial conditions, as in French (*une voiture d’occasion*), but rather by the fact that it was previously owned (or previously owned by only one other person as in *a second-hand car*).

Alternate ways to gain access to reference can be found also inside languages: for instance *pork* and *pig meat* are two synonyms in so far as they can designate the same reference but they differ in their referential paths. This holds true also for syntactic constructions as illustrated by the comparison between the ditransitive construction (*Bill sent Joyce a walrus*) and the *to* construction (*Bill sent a walrus to Joyce*) which accord relative salience to different facets of the construed scene (Langacker 1991b).

Finally I will show that the different referential paths of synonyms often trigger different constraints in their context of use. This can be illustrate by the differences between these two syntactic constructions as well as with the two locative prepositions “ in, inside ” of Tupuri (Ruelland 1998):

nen “ in, inside ” < “ eye ” →	compact domain	: * in a hole
bil “ in, inside ” < “ belly ” →	hollow interior	: * in the forest

## Bibliography

- Langacker, Ronald W. 1991a. *Foundations of cognitive grammar (vol.2)*. Stanford: Stanford University Press.
- Langacker, Ronald W. Cognitive Grammar, in *Linguistic Theory and Grammatical Description*, F. Droste et J. Joseph (eds), Amsterdam / Philadelphia, John Benjamins, 1991 : 275-306.
- Robert, Stéphane, 2008, Words and their meanings: principles of variation and stabilization, in Martine Vanhove (ed), *From polysemy to semantic change: towards a typology of lexical semantic associations*, Studies in Language Companion Series 106. Amsterdam : John Benjamins, 55-92.
- Ruelland, Suzanne, 1998. Je pense et je parle comme je suis: le corps, le monde et la parole entupuri. *Faits de langues* 11-12 : 335-58.

## **Synonymy in Terminology: the Contribution of Ontoterminology**

Defined as a “set of designations [a designation is a “representation of a concept by a sign which denotes it”] belonging to one special language” [ISO 1087-1], the main goal of terminology is to eliminate ambiguity from technical languages by means of standardization.

In order to achieve such an objective, the General Theory of Terminology (GTT) postulates several principles. One of them is bi-univocity. It means that a term denotes only one concept (monosemy) and a concept is denoted by only one term (mononymy). Other principles are the priority of concept over designation (term) and the universality of concept independent of the diversity of languages.

In theory, since terminology has to be normalized, there is no synonymy in terminology (from the GTT point of view). Nevertheless, in practice, as a term is used in texts as a word, a term owns a sense (signified) which must not be confused with a concept – it is important to bear in mind that the lexical structure extracted from corpora does not match the conceptual one directly built by experts in a formal language (it means that subsumption is not hypernymy). Terms in usage are bound into linguistic networks for different relationships, e.g. synonymy, metonymy, etc.

The aim of this article is to present the contribution of ontoterminology for understanding synonymy in terminology. An ontoterminology is a terminology whose conceptual system is a formal ontology – an ontology, from the knowledge engineering point of view, is a specification of a conceptualization; it means a shared description of concepts and their relationships about a domain expressed in a computer readable language. Ontoterminology separates and links the two different linguistic and conceptual systems of which terminology is composed. Concepts and terms in ontoterminology exist in their own right and definitions written in natural language (understood as linguistic explanations) are separated from definitions written in formal language (understood as formal specifications of concept). The conceptual level is language-independent and bi-univocity is not mandatory – polysemy is available and only normalized terms are monosemic. It is then possible to consider for synonymy the two dimensions, semantic and stylistic, along which the meaning of terms can vary.

In ontoterminology, two terms are said to be synonymous if they denote the same concept – difference in form generally corresponds to difference in usage (stylistic, sociolinguistic features). The total synonymy, i.e. same denotation (concept) and same connotation (usage), is not so rare in a technical domain. Ontoterminology allows to distinguish different kinds of plesionymy (structural, functional) according to the different types of relationships between the denoted concepts (is-a, part-of and function) and in relation to rhetorical figures (ellipsis, metonymy, meronymy, synecdoche). For example “wheel” is a structural plesionym of “turbine” since <wheel> is a part of <turbine>.

The article will emphasize the distinction between the linguistic and the conceptual dimensions of terminology (a term is not a lexicalized concept and a designation is a not a denomination). The paradigm of ontoterm (association of a concept and a term) will be introduced. The importance of formal ontology and the role of the conceptual relationships is-a, part-of and function for plesionymy will be detailed.

## Bibliography

- Budin, G. 2001: A critical evaluation of the state-of-the-art of Terminology Theory. *ITTF Journal*,12. Vienna: TermNet.
- Cabré, T. 2003: Theories in terminology. In *Terminology* 9(2): 163-199.
- Edmonds, P., Hirst, G. 2002: Near-synonymy and lexical choice. *Computational Linguistics archive*, 28(2): 105 - 144
- Felber, H. 1984: Terminology Manual, Unesco (United Nations Educational Scientific and Cultural Organization) – Infoterm (International Information Centre for Terminology).
- Hamon, T., Grabar, N. 2009: Exploring graph structure for detection of reliability zones within synonym resources: experiment with the gene ontology. *Human Language Technology, Proceedings of the Workshop on BioNLP*, Boulder, Colorado, June 2009: 89–96.
- Hirst, G. 1995: Near-synonymy and the structure of lexical knowledge. In *AAAI Symposium on Representation and Acquisition of Lexical Knowledge: Polysemy, Ambiguity, and Generativity*, Stanford: 51–56,
- ISO 1087-1:2000. Terminology work-Vocabulary-Part1: Theory and application. International Organization for Standardization.
- ISO 704:2000. Terminology work - Principles and methods. International Organization for Standardization.
- Pavel, S. & Nolet, D. 2001. Handbook of Terminology. Minister of Public Works and Government Services Canada 2001, Catalogue No. S53-28/2001.
- Roche, C., Calberg-Challot, M., Damas, L., Rouard, P. 2009. Ontoterminology: A new paradigm for terminology. *International Conference on Knowledge Engineering and Ontology Development*, 5-8 October, Madeira (Portugal).
- Wright, S.E., Budin, G. 1997: Handbook of Terminology Management, volume 1 and 2, John Benjamins Publishing Company.



**initiality (aspectual),**  
or perfectivity

•—————  
a syncretic process

Thus, the meaning of the perfective construction *stat'+INF* indeed contains two components.

**3.** Besides being a special interest for languages that have the category of aspect, the analysis enables to make the following fundamental conclusion. Linguistics does need the concept of synonymy.

As to the nature of the phenomenon, I believe the analyzed case has shed new light on it.

#### Bibliography

- Березин, Ф.М. 1998: Н.В. Крушевский – провозвестник лингвистики XX в. Н.В. Крушевский. Избранные работы по языкознанию. Moscow: Наследие. 4-24.  
Ломтев, Т.П. 1976: Общее и русское языкознание. Moscow: Наука.

## Synonymy in Contemporary United States Spanish

A topic of current interest in Spanish linguistics is that of lexical (i.e. ‘traditional’) synonymy in diachronic terms. Recent studies (Stale, 2009; Horcas Villareal, 2009) have concentrated on the occurrence of synonymy in diachronic terms. As Spanish evolved from Vulgar Latin, synonymy resulted due to word transmission. That is, **patrimonial words (voces populares)**, those words that underwent the predicted phonological processes, also maintained their **learned word (cultismo)** or **semi-learned word (semicultismo)** counterpart, as in (1).

The aforementioned studies have greatly increased our diachronic understanding of lexical synonymy in Spanish. That said, however, it is surprising to note the lack of research treating synchronic analysis of functional synonymy. For instance, Contemporary United States Spanish (henceforth, CUSP) is known for its abundance of synonymous lexical entries, as in (2). Unlike its diachronic counterpart, however, synonymy does not occur as a result of word transmission. That is, a peculiar trend emerges from the list above in (2) in that the words both on the left and the right already exist in Standard Spanish and have different meanings. For example, both *la carpeta* means ‘the folder’ and *la alfombra* means ‘the carpet’. However, in the case of CUSP, semantic extension as a result of contact with English has occurred, thus forming a **doublet (doblete)** in that *la carpeta* means both ‘the folder’ and ‘the carpet’. This is particularly relevant in that synonymy occurs in this variety from words generally that already exist in the language. For instance, *vacunar* exists in Standard Spanish as ‘to vaccinate’, with *pasar la aspiradora* meaning ‘to vacuum,’ yet in CUSP *vacunar* is used in its more functional meaning and *pasar la aspiradora* is understood, though generally not produced.

I conclude my treatment of the relationship between lexical and functional synonymy by suggesting that, although the transmissions were quite different, a similar pattern emerges; in diachronic terms, the **patrimonial word (voz popular)** is generally the more frequently used word, with the **(semi)learned word ((semi)cultismo)** less frequently used. In synchronic terms, the word that undergoes semantic extension as a result of contact with English is generally the more frequently used word, with the Standard Spanish word less frequently used.

In short, the current study is novel in that it is the first known study to consider the relationship between lexical and functional synonymy. In addition, it greatly adds to the field by exploring a linguistic process and a variety of Spanish that have both been inexplicably understudied. In the study, I illustrate how synonymy in Spanish has had two distinct patterns, namely i) during the formation of the language via word transmissions from Latin and ii) in a modern variety of Spanish that is in contact with English. Additionally, I offer a linguistic analysis that considers both the differences (e.g. word transmission vs. semantic extension) and the similarities (e.g. a trend for one of the lexical entries to be much more common) of both types of synonymy.

(1) patrimonial:	hostigar	(semi)learned:	fustigar	‘to pester’
	injertar		insertar	‘to graft’
	lindar		limitar	‘to restrict’
(2)	la luz		el semáforo	‘the traffic light’
	la carpeta		la alfombra	‘the carpet’
	vacunar		pasar la aspiradora	‘to vacuum’
	aplicar		solicitar	‘to apply’

## Bibliography

- Calvi, Maria Vittoria and Emma Martinell (1997). Los dobletes léxicos en la enseñanza del español a extranjeros. *ASELE Actas VIII*, 227-239.
- Horcas Villareal, José Mario (2009). El español: dobletes, cultismos, y neologismos. *Contribuciones a las Ciencias Sociales*, 1-5.
- Penny, Ralph (2004). *Variation and Change in Spanish*. Cambridge University Press: Cambridge.
- Stala, Ewa (2009). Dobletes etimológicos en español - su origen y evolución semántica: observaciones puntuales. *Studia linguistica*, 126, 113-127.

## The Finnish comitative case *-ine* and the postpositional *kanssa* construction – how synonymous are they?

The paper examines the relation between the Finnish comitative case *-ine* and the construction with the postposition *kanssa* 'with' that have been assumed to be synonymous. The core meaning of the comitative is Accompaniment, although cross-linguistically the same form can also be used for encoding Instrument, Possession or Inclusion, for example. Accompaniment can be expressed by adpositions, case affixes and serial constructions, among other means. (Stolz et al. 2009: 602f.) In Finnish, the principal means are the number-insensitive comitative case affix *-ine* which is accompanied by a possessive suffix when attached to a noun, as in example (1), and several postpositions governing the genitive case, the main variant being *kanssa*, as in example (2).

In recent literature it has been claimed that the functional domains of the comitative case and the postposition *kanssa* overlap (Stolz et al. 2006: 61, Karlsson 1982: 132; partly also Haarala et al. 1990–1994, Hakulinen et al. 2004: 942f., 1211f.) and that the comitative case is being replaced by the postposition *kanssa* (Stolz et al. 2005: 214, 2006: 61; cf. also Nau 1995: 133), which implies the assumption of synonymy between the two constructions. In my presentation, I will show that these claims are questionable.

It is clear that the two constructions have common functions, such as the prototypic function of Accompaniment. For the most part, however, newspaper corpus material (Pajunen 2003) shows that their functional domains are different: the frequencies of the common functions are notably different and, in addition, both constructions have some distinct functions that cannot be expressed by the other construction. Essentially, *kanssa* concentrates on expressing Accompaniment extensively while the case is used in more diverse functions (cf. Sirola 2008). In my paper, I will demonstrate this on the basis of corpus material.

The differences in the usage of the two constructions are directly based on their meanings. First, the construction with *kanssa* combines two symmetric, independent participants, whereas the comitative case represents an asymmetric relationship in which the companion (marked with the case affix) is subordinate to the accompanee. Consequently, it is impossible to encode for example a reciprocal action with the inflectional comitative while it is commonly expressed by *kanssa*. Second, since the comitative case marker is followed by a possessive suffix that refers (in most cases) to the accompanee, the participants are semantically bound to each other; the participants of the structure with *kanssa* do not have such a semantic limitation. On the one hand, this favors the postposition *kanssa*, as it can be used more widely for expressing Accompaniment, for example with proper nouns. On the other hand, the bond between the participants allows the comitative case to express diverse kinds of meronymic and hyponymic relations which cannot be expressed by the postposition *kanssa*.

Examples (stem from the HS2000 corpus, Pajunen 2003):

- (1) Anne Nordlund laps-**ine**-en            on palannut kuukausi sitten Suome-en.  
Anne Nordlund child-COM-POSS.3 has returned month ago Finland-ILL  
'Anne Nordlund has returned to Finland **with** her child(ren) a month ago.'  
(HS 1 578 563)
- (2) Äiti oli lapse-nsa            **kanssa** polkupyör-i-llä matka-lla uimaranna-lle.  
mother was child-GEN+POSS.3 with bicycle-PL-ADE way-ADE beach-ALL  
'Mother was on her way to the beach **with** her child by bicycles.'  
(HS 443 214)

## Bibliography

- Haarala, R.; Lehtinen, M.; Grönros, E.-R.; Kolehmainen, T. and Nissinen, I. (eds) 1990–1994. *Suomen kielen perussanakirja*. Kotimaisten kielten tutkimuskeskuksen julkaisuja 55. Helsinki: Kotimaisten kielten tutkimuskeskus.
- Hakulinen, A.; Vilkuna, M.; Korhonen, R.; Koivisto, V.; Heinonen, T.-R. and Alho, I. (eds) 2004. *Iso suomen kielioppi*. Suomalaisen Kirjallisuuden Seuran toimituksia 950. Helsinki: SKS.
- Karlsson, F. 1982. *Suomen peruskielioppi*. Suomalaisen Kirjallisuuden Seuran toimituksia 378. Helsinki: SKS.
- Nau, N. 1995. *Möglichkeiten und Mechanismen kontaktbewegten Sprachwandels unter besonderer Berücksichtigung des Finnischen*. Edition Linguistik 08. München – Newcastle: Lincom Europa.
- Pajunen, A. 2003. HS2000. 31 million word FDG-analysed Context Mill database of the newspaper Helsingin Sanomat, volumes 2000 and 2001.
- Sirola, M. 2008. Komitatiivi nykysuomessa: sijan typologista ja areaalista taustaa sekä sen ilmaiset merkitykset Helsingin Sanomien korpuksessa. Master's thesis. University of Tampere.
- Stolz, T.; Stroh, C. and Urdze, A. 2005. Comitatives and Instrumentals, *The World Atlas of Language Structures*, M. Haspelmath, M.S. Dryer, D. Gil, B. Comrie (eds), Oxford: Oxford University, 214–217.
- Stolz, T.; Stroh, C. and Urdze, A. 2006. *On comitatives and related categories: a typological study with special focus on the languages of Europe*. Berlin – New York: Mouton de Gruyter.
- Stolz, T.; Stroh, C. and Urdze, A. 2009. Varieties of comitative, *The Oxford Handbook of Case*, A. Malchukov and A. Spencer (eds), Oxford – New York: Oxford University Press, 601–608.

## **Competitions of synonyms through time: Conceptual and social salience factors and their interrelations**

This paper takes as a starting point three different hypotheses. First, the study of the diachronic development of synonymous forms reveals essential aspects about the nature and motivations of synonymy. Second, the emergence and competition of synonymous forms can either result from conceptual salience factors, namely prototypicality (semasiological salience) and entrenchment (onomasiological salience), or from social salience factors, i.e. sociolinguistic, stylistic or pragmatic prevalence, or even from interaction of both salience factors. Crucially, prototype-theoretical features of the concepts involved can determine the necessary differences between synonyms within and across lectal varieties or even across languages; and the lectal features of the items involved can not only determine the occurrence of synonyms across lectal varieties but also motivate the differences of prototypical structure between synonyms. The third hypothesis is that competitions of synonyms shed light about the processes of language variation and change, including convergence and divergence processes between lectal varieties and the processes of linguistic stratification and standardization.

These hypotheses will be tested with two different corpus-based case studies in lexical synonymics of Portuguese. The first case study is about the semantic development of the verb *deixar* 'to leave, to let' from Old to Modern Portuguese and its most competitive synonyms, namely *abandonar* 'to abandon' (a Gallicism) and *permitir* 'to allow, to permit' (a juridical Latinism). The late entry of *abandonar* and *permitir* into Portuguese (in the fifteenth century only) led to a situation of full conceptual and distributional synonymy with the two prototypical uses of *deixar* at the time. However, this situation rapidly caused a semantic dissimilation which took the shape of a prototype reorganization, mainly in the semasiological structure of *deixar*: the conceptual center of the subgroup of senses of *deixar* with verbal complement shifted from the active sense of 'to allow' to the passive sense of 'not to intervene' and the two prototypical centers of *deixar* climbed up to hierarchically more abstract levels. At the same time, nondenotational meaning differences, namely stylistic, emotional and interactional have facilitated these prototype reorganizations. Additionally, we will see that Romance cognate verbs, namely the Portuguese *deixar*, Spanish *dejar*, French *laisser*, Italian *lasciare* and Romanian *a lăsa* are not totally semantically equivalent, as they also have differences in prototype organization.

The second case study includes the development of four dozens of sets of denotational synonymous nouns selected from the lexical fields of football and clothing in European and Brazilian Portuguese in the last 60 years. The aim is to examine the impact of item-related features (i.e. endo-/exogenousness, foreign influence or loanwords, archaism, neologism) and concept-related features (i.e. prototypicality, vagueness, recent origin, semantic field) in the production of denotational synonyms within and across the two national varieties of Portuguese. Internal uniformity measures enable us to measure the lexical uniformity/diversity of a certain concept in a language variety: the internal uniformity value will decrease the more synonymous terms there are competing to denote the same concept, and the more dominant some of these terms become. These measures allow us to calculate increases/decreases of lexical homogeneity/heterogeneity in both national varieties of Portuguese and lexical standardization process as well, and they also permit to compare the degree of change. External uniformity measures enable us to know if the two national varieties of Portuguese have been going through a process of lexical convergence or divergence. The denotational synonyms studied show that the Brazilian variety has changed more than the European variety and that both varieties diverge from each other in the vocabulary of clothing.

The two diachronic studies stress the following characteristics of synonymy: the role of prototypicality in the differentiation of synonyms, the sufficient semantic similarity and the necessary semantic differences between synonyms, the superfluity and contingency of total synonymy, the influence and correlations of concept-related features and item-related features in the emergence and competition of synonyms within and across lectal varieties, and the interaction of conceptual and social factors in the occurrence of synonyms.

## **Synonymy of fuzzy and scalar concepts and terms**

Studying vocabulary one can see that different lexical domains behave differently. For example, colour terms correspond to colour concepts that are fuzzy “clouds” in the colour space, but temperature terms form scalar concepts. In both cases one can find synonyms, basic and non-basic terms. Even very simple categories such as white and black have synonyms – whitish, snow white, warm white, charcoal black, deep black, etc.

In an empirical study of Estonian sense perception terms 80 subjects were interviewed both for colour and temperature terms. In colour naming task each subject 65 standard Color-aid colour tiles were shown and asked to name them. As a result, 5,197 names were given to the tiles at all. Among these there were 638 different terms. As a mean 20.1 synonyms were given for each tile. The minimum number of synonyms (9) got tile Y (yellow), and the maximum number of synonyms (41) got ORO S3 (orange-red-orange shadow 3). It is clear that colour tiles with focal colour have fewer synonyms and non-focal colours have more synonyms. The distribution of synonyms for each tile corresponds to Zipf curve. Basic colour terms tend to be most frequent dominant synonyms and non-basic recessive synonyms occur with low frequency.

In the temperature term ranking task 80 subjects were asked to build up a temperature scale using temperature terms she or he had named in the earlier tasks. Every term in a scale got numerical value. Psychological zero (decided arbitrarily) got the value 0. As a result a progression... -2, -1, 0, 1, 2 ... follows. After that the subjective temperature scale was constructed over all individual scales. In the ranking task 659 terms were used in all. Among these there were only 85 different terms in comparison with the 142 different terms listed in the earlier tasks. Every individual temperature scale consisted of a mean 8.24 terms.

In the case of fuzzy colour concepts the synonymy was defined using standard colour tiles. All names given to a certain colour tile were synonyms. In the case of scalar temperature terms one can define the terms which are located “quite close” on the subjective integral temperature scale as synonyms. So ice cold and icy, suitable and tepid, and burning and hot are synonyms in Estonian. They do not denote the same temperature.

Stede Manfred  
University of Potsdam

**TBA**

## Associating Difficulty in Near-Synonymy Choice with Types of Nuance Using Core Vocabulary

There are arguably infinitely many dimensions along which members of a cluster of near-synonyms can differ (Cruse, 1986), and the nuances that differentiate near-synonyms along a dimension are often subtle and difficult even for native speakers. The diversity of near-synonym variation types has motivated the categorization of these dimensions of these variations. DiMarco, Hirst, & Stede (1993) proposed 38 dimensions for differentiating near-synonyms, which were further categorized into semantic and stylistic variations. Stede (1993) focused on the latter and further decomposed them into seven scalable sub-categories. Inkpen & Hirst (2006) organized near-synonym variations into a hierarchical structure, combining stylistic and attitudinal variation into one class in parallel to denotational differences.

Despite the variety of categorization methods, stylistic variation among near-synonyms is an important dimension that has been frequently addressed. In this study, we hypothesize that the stylistic nature of nuances correlates to the degree of difficulty in choosing between near-synonyms. Contrasting some recent studies that focus on contextual preferences of synonyms (e.g., Arppe & Järvikivi 2007), we elect to investigate the internal features of near-synonym nuances. Specifically, we adopt the notion of *core vocabulary* to associate stylistic variation in theory with the difficulty level of near-synonym choice in practice. Core vocabulary consists of “words that suffice to define all of the remaining vocabulary of a language” (Lehmann 1991). It was first related to stylistic variation among near-synonyms by Stede (1993). Carter (1987) listed ten features of core vocabulary, among which, *associationism* is the “bridging” dimension between stylistic variation and core vocabulary. It is characterized by scalable dimensions closely resembling those Stede used for characterizing stylistic variations. Carter claimed that core vocabulary words are relatively neutral on these scales, indicating fewer stylistic variations among them.

Notably, some of Carter’s features of CV are readily verifiable using computational linguistic techniques. The collocability of a word, for example, can be approximated by the number of co-occurring word types (normalized by the number of senses to eliminate the confounding factor of polysemy); neutrality in field of discourse can be verified by a word’s distribution across different genres in a balanced corpus. Multiple linguistic resources are combined in our study to achieve an empirical characterization of core vocabulary.

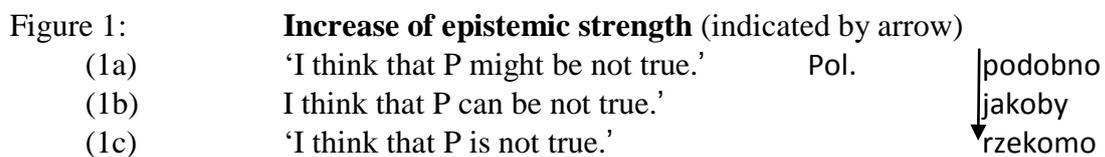
To test our hypothesis, a near-synonym lexical choice task (Edmonds 1997) is employed to measure difficulty levels. In this task, lexical gaps are created in sentences from a corpus by removing members of a near-synonym cluster. The sentences are then presented to subjects whose task is to determine from context which member of the cluster is the missing word. Experiments in existing studies have shown great variance in the performance (and hence in level of difficulty) on different near-synonym clusters (Edmonds 1997; Inkpen 2007). Our study shows that such variance is correlated with differing degrees of coreness of the near-synonyms, and in turn, different types of near-synonym variations. Counter to intuition, the seemingly subtle stylistic nuances are usually easier for subjects to distinguish than non-stylistic differences.

## Bibliography

- Arppe, Antti & Järvikivi, Juhani. 2007: Every method counts: combining corpus-based and experimental evidence in the study of synonymy. *Corpus Linguistics and Linguistic Theory*, 3(2), 131–159.
- Carter, Ronald. 1987: *Vocabulary: Applied Linguistic Perspectives*. Allen & Unwin.
- Cruse, D. A. 1986: *Lexical Semantics*. Cambridge University Press.
- DiMarco, Chrysanne; Hirst, Graeme; & Stede, Manfred. 1993: The semantic and stylistic differentiation of synonyms and near-synonyms. *AAAI Spring Symposium on Building Lexicons for Machine Translation*, 114–121.
- Edmonds, Philip. 1997: Choosing the word most typical in context using a lexical co-occurrence network. *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*, 507–509.
- Inkpen, Diana. 2007: A statistical model for near-synonym choice. *ACM Transactions on Speech and Language Processing*, 4, 1–17.
- Inkpen, Diana & Hirst, Graeme. 2006: Building and using a lexical knowledge base of near-synonym differences. *Computational Linguistics*. 32, 223–262.
- Lehmann, Hubert. 1991: Towards a core vocabulary for a natural language system. *Proceedings of the Fifth Conference of the European Chapter of the Association for Computational Linguistics*, 303–305.
- Stede, Manfred. 1993: Lexical choice criteria in language generation. *Proceedings of the Sixth Conference of the European Chapter of the Association for Computational Linguistics*, 454–459.

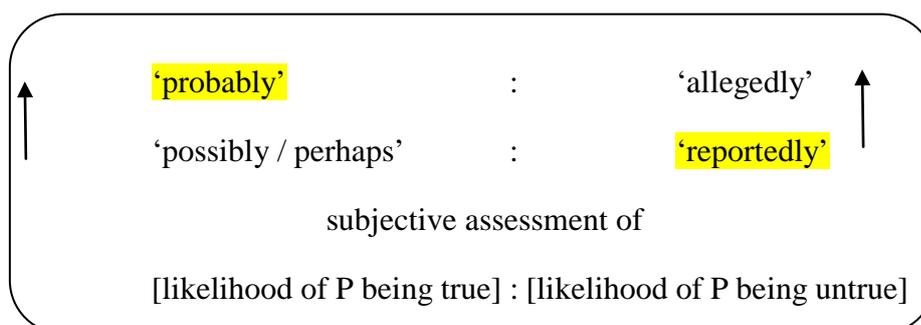
## How to do contrastive semantics with propositional modifiers: The case of hearsay adverbs

The paper will present preliminary results from two ongoing projects dealing with evidentiality markers. Specifically, our concern is the semantic description of a couple of propositional modifiers indicating hearsay in Polish and German (Pol. *rzekomo*, *jakoby*, *podobno*, Germ. *angeblich*, *vorgeblich*, *mutmaßlich*) which – though considered close synonyms – often cannot faithfully be translated into each other, nor by Eng. *allegedly*. These hearsay markers have been claimed to carry epistemic overtones by which the actual speaker transmits his/her doubts into the contents of the message referred to (P). Corpus-based research shows that these lexemes do so to a varying extent: with some of them, e.g. *podobno*, epistemic overtones arise only via conversational implicature, while for others, e.g. *rzekomo*, they seem to form part of their inherent semantics and are not cancellable in most contexts (Wiemer 2006). This analysis leads to an ordering along degrees of epistemic strength that resembles Horn-scales:



However, Socka (2009; 2010), in her corpus-based study on Polish-German translational equivalence, has shown that any of the aforementioned hearsay markers can become void of epistemic overtones in specific contexts; the „negative” epistemic default can be cancelled even for Pol. *rzekomo* and Germ. *angeblich* (see below). For a similar observation regarding Eng. *allegedly* cf. Ramat/Ricca (1998: 230), who speculated that the epistemic value of the respective sentence adverb is influenced mainly by the existence (or absence) of alternative hearsay adverbs in the language which „share“ into the stages of a scale corresponding to Fig. 1. They assume an analogy between merely epistemic adverbs and reportive adverbs (often carrying a negative epistemic commitment) to rest on Horn-scales, each with a marked and an unmarked member:

Figure 2: Horn-scale based analogy between epistemic and reportive adverbs



**marked member of opposition** (arrow indicates increase of negative stance as for P being true)

We can show that the epistemic load of the aforementioned hearsay markers cannot be captured by Horn-scales. Explanations based on the Q-principle (cf. Levinson 2000; Huang 2007) must fail unless they arise from clausal implicatures. Best suited are explanations based on the M-principle, applied in practice by Olbertz (2007) for Mexican Spanish *dizque*. Olbertz' pragmatic analysis, in turn, does not explain (and even contradicts) facts known from the distribution of such hearsay markers across discourse types: Germ. *angeblich* and Pol. *rzekomo* regularly lose their otherwise strong epistemic default in news reports, i.e. in discourse for which an indication of second-hand information is presupposed and, on this account, redundant.

Given these puzzles on the borderline between lexical semantics and pragmatics, we formulate a proposal of how epistemic and evidential meaning components should be accounted for in the representation of German and Polish hearsay adverbs (and particles). This case study also allows to highlight general problems raised by units with „subjective“ meanings that modify entities on the propositional and higher layers (in Dik et al.'s 1990 terms): since their description and (intra-lingual as well as cross-linguistic) comparison can be only intensional, translational equivalence is here even less sufficient to establish synonymy than in the case of figurative extensions of adjectives denoting physical properties or kinship terminology. We thus use our case study to illustrate how the problem of synonymy raised in lexical typology (cf. Koptjevskaja-Tamm 2008, Evans 2010) can be expanded to function words.

## Bibliography

- Dik, S.C., Hengeveld, K., Vester, E., Vet, C. (1990): The hierarchical structure of the clause and the typology of adverbial satellites. In: Nuyts, J., Bolkestein, M., Vet, C. (eds.): *Layers and levels of representation in language theory*. Amsterdam, Philadelphia: Benjamins, 25-70.
- Evans, N. (2010): Semantic typology. In: Jung Sung, J. (ed.): *The Oxford Handbook of Linguistic Typology* (to appear). Oxford etc.: Oxford U.P., 504-533.
- Huang, Y. (2007): *Pragmatics*. Oxford etc.: Oxford U.P.
- Koptjevskaja-Tamm, M. (2008): Approaching lexical typology. In: Vanhove, Martine (ed.): *From Polysemy to Semantic Change (Towards a typology of lexical semantic associations)*. Amsterdam, Philadelphia: Benjamins, 3-52.
- Levinson, S.C. (2000): *Presumptive meanings. The theory of generalized conversational implicature*. Cambridge, M.A.: MIT Press.
- Olbertz, H. (2007): *Dizque* in Mexican Spanish: the subjectification of reportative meaning. In: Squartini, M. (ed.): *Evidentiality between lexicon and grammar*. Italian Journal of Linguistics – *Rivista di linguistica* 19, 151-172.
- Ramat, Paolo, Ricca, Davide (1998): Sentence adverbs in the languages of Europe. In: van der Auwera, Johan, Baoill, Dónall P.Ó. (eds.): *Adverbial Constructions in the Languages of Europe*. Berlin, New York: Mouton de Gruyter, 187-275.
- Socka, A. (2009): Marker der reportativen Evidentialität im Deutschen und ihre polnischen Äquivalente. In: *Akten des 43. Linguistischen Kolloquiums „Pragmantax II. Zum aktuellen Stand der Linguistik und ihrer Teildisziplinen“*, Magdeburg, 10.-13.09.2008.
- Socka, A. (2010): Reportative Partikeln in kontrastiver Sicht (Polnisch – Deutsch). In: Kątny, A., Socka, A. (eds.): *Modalität / Temporalität in kontrastiver und typologischer Sicht*. Frankfurt/M. etc.: Lang, 239-264.
- Wiemer, B. (2006): Particles, parentheticals, conjunctions and prepositions as evidentiality markers in contemporary Polish (A first exploratory study). *Studies in Polish Linguistics* 3, 5-67.

## **The comparability of similar inferential meanings across languages**

Inferential expressions indicate the speaker's inference, based on some type of direct information source, such as observational or memorial source. Different kinds of inferential expressions show remarkable cross-linguistic semantic variation. However, genetically related languages or languages in contact often have inferential expressions that seem to indicate quite similar, even (near) synonymous meanings. The purpose of this paper is to discuss the comparability of inferential meanings in languages by means of comparative concepts (cf. Haspelmath 2010) and the degree of similarity between language-specific inferential meanings. The discussion is based on my typological investigation into the semantic domains of epistemic modality and inferentiality as they are grammatically represented in the world's languages. This study uses descriptive data from the sample of 130 languages. In this kind of study, it is not possible to consider in great detail contextual and stylistic differences between the compared meanings. Nevertheless, it is possible to deal with the notions of similarity and synonymy from various perspectives. In this presentation, I will firstly introduce the comparative concepts, distinguished for language-specific grammatical inferential meanings, and illustrate the relationship between the comparative concepts and meanings by several examples. Second, I will focus on the question of how differences between semantic networks arguably affect the degree of similarity between grammatical inferential meanings.

The central type of comparative concept, used in this study, is that of function. This notion is defined by Haspelmath (2003) as neutral between language-specific notions of sense and use (or meaning). According to Haspelmath, a function is distinguished if there is at least one language that has a formal expression for that function. Due to the differences in quality and details of information, provided by the descriptive material, the functions created in this study have to be somewhat more coarse-grained. However, they capture the main distinctions of the variation in meaning, and the proposed semantic description is more fine-grained than in the previous large-scale typological studies concerning inferentiality and epistemic modality. The pure inferential functions do not include epistemic properties, indicating degrees of the speaker's certainty. They have been elaborated on the basis of epistemological notions of direct sources of knowledge/information which can provide a foundation for an inference (e.g. Audi 2003). Inferential functions can also consist of inferential properties combined with epistemic properties. For instance, the English modal *must* represents several types of functions with combined properties. The basic purpose of functions is to identify similar meanings within languages and across languages. Actually, they cover meanings, ranging from (near) synonymous to roughly similar meanings. For example, the function 'inference from memory' represents the gamut of inferential meanings, such as 'inference based on previous experience' and 'inference based on general knowledge'. Inferential meanings of different languages are members of various types of semantic networks. I will argue that if the notion of (near) synonymy is needed at all instead of the notion of similarity in the comparison of inferential meanings across languages, it should be confined to those inferential meanings that belong to highly similar semantic networks.

### Bibliography

Audi, Robert 2003. *Epistemology. A Contemporary Introduction to the Theory of Knowledge*. New York, London: Routledge.

- Haspelmath, Martin 2010. Comparative Concepts and Descriptive Categories in Cross-Linguistic Studies. To appear in *Language* 86.
- Haspelmath, Martin 2003. The Geometry of Grammatical Meaning: Semantic Maps and Cross-Linguistic Comparison. *The New Psychology of Language*, Volume 2., Michael Tomasello (ed.), Mahwah, New Jersey: Lawrence Erlbaum Associates, 211-242.

## Searching for constructional synonymy: Paraphrasing as evidence against ellipsis

In this presentation, the concept of synonymy is used to signal essentially identical reference to particular event types by separate clause-level constructions. By “essentially”, it is meant that the events cover the same profiled participants (in the sense of cognitive grammar and frame semantics, see, e.g., Fillmore 1977, 1982; Langacker 1987). More precisely, I discuss the meaning of the Finnish ablative-initial (cf. 1) and allative-initial (cf. 2) verbless constructions that are functionally comparable to finite clauses and occur frequently in headlines. Since they lack a finite verb, it is challenging to indicate the event types they can refer to: traditionally, it is taken that event types are largely specified by verbs. However, paraphrasing is seen as one solution for describing the meaning of the constructions in question and finding synonymous finite constructions for them.

- (1) *Opiskelijo-i-lta*      *monikulttuurinen*      *vaatemallisto*  
student-PL-ABL      multicultural[NOM.SG]      apparel.collection[NOM.SG]  
lit. ‘From students a multicultural apparel collection.’
- (2) *Varka-i-lle*      *iso*      *saalis*      *ravirada-lta*  
thief-PL-ALL      big[NOM.SG]      haul[NOM.SG]      trotting-track-ABL  
lit. ‘To thieves a big haul from trotting-track.’

There are comparable finite constructions in Finnish, i.e., constructions with the same cases in the same positions and a finite verb (e.g., the passive construction, the necessary zero subject construction, and the possessive construction), but, based on an intuitive analysis, I have argued that the ablative-initial and allative-initial verbless constructions cannot be seen as elliptic clauses. That is, these verbless constructions denote different event types from the comparable finite constructions; it is not the case that a verb is simply omitted. To justify the argument also empirically, I have designed a paraphrase test for discovering how the ablative-initial and allative-initial verbless constructions are construed. In order for the verbless constructions to be elliptic, most of the paraphrases would have to be formally identical except for a “restored” verb: respondents would intuitively know the “missing” verb. 163 respondents participated in the paraphrase test concerning the allative-initial verbless constructions and 209 respondents in the one concerning the ablative-initial verbless constructions, and most of the yielded paraphrases were basic transitive or intransitive clauses (cf. 3) – that is, not the constructions that simply adding a verb would have yielded. This is crucial for establishing that they are not elliptic clauses but independent constructions (in the sense of, e.g., Goldberg 1995, 2006) that are verbless from the very beginning.

- (3) *varkaa-t*      *sa-i-vat*      *iso-n*      *saalii-n*      *ravirada-lta*  
thief-NOM.PL      get-PST-3PL      big-ACC      haul-ACC      trotting-track-ABL  
‘Thieves had a big haul from a/the trotting-track.’

Thus paraphrasing verbless constructions with finite clauses illustrates that synonymy may be defined as covering far broader domains than individual lexical elements; it applies to coding particular event types in different ways as well. In that case, synonymy must certainly be seen as a relatively coarse-grained equivalence between constructions. It does not cover, for example,

pragmatic aspects of meaning. However, from my point of view, this is not a problem but the desired state of affairs: in order for synonymy to be a useful concept in linguistics, it needs to be inclusive enough.

### Bibliography

- Fillmore, Charles J. 1977. Topics in Lexical Semantics. *Current Issues in Linguistic Theory*, Roger W. Cole (ed.), 76–138. Bloomington and London: Indiana University Press.
- Fillmore, Charles J. 1982. Frame Semantics. *Linguistics in the Morning Calm: Selected Papers from SICOL-1981*, The Linguistic Society of Korea (ed.). Seoul: Hanshin, 111–137.
- Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure* [Cognitive Theory of Language and Culture]. Chicago and London: The University of Chicago Press.
- Goldberg, Adele E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar [volume I]: Theoretical Prerequisites*. Stanford: Stanford University Press.

## Lexical Choice and Loanwords: the Use of Anglicisms in Dutch

**Synonymy and the Success of Loanwords:** In this paper we discuss the issue of synonymy from a contact linguistics angle. Our focus is on determining which features influence the relative preference for a loanword when a synonymous native alternative exists. This onomasiological perspective on borrowing is developed in two separate case studies, each dealing with English person reference nouns (e.g. designer, lover) in a Belgian Dutch and Netherlandic Dutch newspaper corpus (consisting of over 1 billion words ranging from 1999 to 2005).

**Case Study 1 – Conceptual Features:** The first study tries to capture the influence of conceptual features on the success of English person reference nouns. We define that success as the corpus frequency of the anglicism, relative to the the total frequency of the English loan word and its denotational synonyms:

Synonymous expressions for the concept HOOLIGAN	Token frequencies in the Dutch corpus
hooligan	9337
voetbalvandaal	611
success-rate for <i>hooligan</i> : $9337 / (9337 + 611) = 93.9 \%$	

Table 1

These success-rates are calculated semi-automatically for 100 English person reference nouns. In order to account for variation in these rates, we verify the effect of three conceptual parameters (lexical field, concept frequency, conceptual newness) and three additional parameters (age of the loanword, regional variation, donor language frequency). Mixed effect regression shows how the conceptual features, and most prominently conceptual newness, are most powerful in explaining the attested variation.

**Case Study 2 – Contextual Features:** This model does however not account for all variation. Although denotational synonymy is a prerequisite for the selection of expressions in case study 1, possible stylistic and pragmatic nuances were not attended to. A contextualized approach to the issue of lexical choice is thus a necessary addition when trying to fully understand how language users choose either the anglicism or a (semi-)synonym (Grondelaers & Geeraerts 2003, Geeraerts 2010). This is the purpose of the second case-study, in which we present an in-depth analysis of junk(ie) vs. verslaafde (“addict”). Relying on previous work (Geeraerts et al. 1994, Divjak & Gries 2006, Edmonds & Hirst 2002), we systematically scrutinize the contexts of use for 11,000 tokens by means of manual coding. Taking a broad definition of context, we focus on regional preferences (Belgian vs. Netherlandic Dutch) and register variation (popular vs. quality newspapers). Taking a more narrow definition of context, we include topic, verify the importance of collocational patterns (focusing on the object of the addiction for each token – e.g. drugs vs. fashion) and take text structuring into account (avoidance of repetition). Logistic regression analyses examine the interplay of all these features in determining lexical choice.

## Bibliography

- Edmonds, P. & G. Hirst. 2002. Near-synonymy and lexical choice. *Computational Linguistics* 28, 2, 105-144.
- Geeraerts, D., S. Grondelaers & P. Bakema. 1994. *The Structure of Lexical Variation. Meaning, Naming and Context*. Berlin/New York: Mouton de Gruyter.
- Geeraerts, D. 2010. Lexical Variation in Space. In Auer, P. & J. E. Schmidt (eds.), *Language and Space: Theories and Methods*, 821-837. Berlin/New York: Mouton de Gruyter.
- Gries, S. & D. Divjak. 2006. Ways of trying in Russian: clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory* 2, 1, 23-60.
- Grondelaers, S. & D. Geeraerts. 2003. Towards a pragmatic model of cognitive onomasiology. In Cuyckens, H., R. Dirven & J. Taeldeman (eds.), *Cognitive Approaches to Lexical Semantics*, 67-92. Berlin/New York: Mouton de Gruyter.
- Haspelmath, M. & U. Tadmor (eds.). 2009. *Loanwords in the World's Languages: A Comparative Handbook*. Berlin/New York: Mouton de Gruyter.