

## **SECTION PAPERS:**

### **1. How many “rains” do we have? Temporal and Spatial readings of the lexeme “rain” in Serbian and Kikuyu**

*Tijana Asic, University of Geneva – ISC CNRS Lyon, Switzerland*

There is, in Serbian (a South-Slavic language) and in Kikuyu (a Bantu language spoken in Kenya) quite a particular type of sentences in which we get the so-called *temporal reading of a sentence* (Asic, 2003a, 2003b, 2004). For example, the following sentences (first one in Serbian and the second one in Kikuyu) both mean that Dusan sleeps as it rains, but also that he is protected from the rain (so there is no physical contact between his body and the rain drops):

- 1) Dusan spava *po* kisi.  
Dusan sleeps over rain
- 2) Dusan akomete *na* mbura.  
Dusan sleeps with rain

It should be emphasized here that these sentences are marked, for they exemplify the violation of certain semantic rules stating what kind of prepositions (in Serbian) or connectors (in Kikuyu) should be used with a particular (spatially static) kind of verbs and particular kind of natural phenomena (Asic, 2003a, 2003b, 2004).

In non-marked sentences (with no rule-violation) we have the physical reading (poor Dusan is sleeping under the rain):

- 3) Dusan spava *na* kisi.  
Dusan sleeps on rain  
*Dusan is sleeping in the rain.*
- 4) Dusan akomete *mbura-ni*.  
Dusan sleeps rain- at  
*Dusan is sleeping in the rain.*

So, while in 3 and 4 “kisa” and “mbura” represent a kind of physical entity (rain drops falling on the dog), in 1 and 2 the same lexemes denote (at least at the first sight) an entity of a ontologically different nature – rain as a temporal event, or, more precisely as a temporal duration that (only temporally!) overlaps with the activity of sleeping.

One of the questions that arise here is: should we then consider that there is “kisa1” / “mbura1” (a physical phenomenon) and “kisa2” / “mbura2” (a temporal entity – an event like a storm)? My answer is negative: we do not have here to multiply the senses of the lexeme and talk about the polysemy here. Actually, it is possible to consider this phenomenon as a kind of coercion (Pustejovsky, 1995) but only if we accept that this generative mechanism is not exclusively syntactic by also semantic. It means that it affects not only the syntactic type but also the semantic interpretation of linguistic units (Asher and Pustejovsky, 2004). Thus, the interpretation of “rain” changes according to the preposition / connector that is used but, at the same time, the meaning of the preposition / connector depends on its arguments. Alternatively this phenomenon can be considered as a kind of co-composition. Obviously, in my approach, the meaning

composition is divorced from syntax. Hence, what we have here is a loss of the strictest forms of compositionality.

### References:

- Asher N. & Pustejovsky J. (2004), *Word meaning and metaphysics*, available from nasher@la.utexas.edu.
- Asic T. (2003a), « The *po-na-u* opposition in Serbian and its Equivalents in some Slavic languages and Kikuyu »; in: ed. by P. Kosta, J. Blaszczak, J. Frasek, Lj. Geist, M. Zygis . *Investigation into Formal Slavic Linguistics*, Frankfurt, Peter Lang, p. 783-797.
- Asic T. (2003b), « L'expression de la relation temporelle d'over-crossing dans certaines langues » *Cahiers de Linguistique Française* 25, 221-235.
- Asic T. (2004), *La représentation cognitive du temps et de l'espace: Etude pragmatique des données linguistiques en français et dans d'autres langues*, thèse de doctorat, Université de Genève et Université de Lyon2.
- Pustejovsky J. (1995), *The Generative Lexicon*, Cambridge (Mass.), The MIT Press.

## 2. Gender assignment and the structure of the lexicon

*Philipp Konzett, University of Tromsø, Norway*

In recent years, there has been an increasing theoretical interest in the issue of gender assignment. The gender of nouns is a lexical category. In his standard work, Corbett (1991:8) underlines that “[f]rom a theoretical point of view, assignment systems have important implications for attempts to determine the structure of the lexicon”. The description of how the lexicon is organised varies according to the (theoretical) approach chosen to gender assignment. The former assumption of arbitrary gender assignment postulated that for each noun, gender is stored individually and independently in the lexicon. This view is now largely given up. Corbett (op.cit.:70-104) gives convincing evidence for the psychological reality of gender assignment regularities. His conclusion is that gender is assigned by rules: “A noun has typically one value for the gender feature, which it brings with it from the lexicon (determined by the assignment rules)” (op.cit.:146). Most of the recent literature on gender assignment is written within a rule-based framework.

Turning back to lexical theory, we can, citing Bybee (1995:427), associate this approach with the “dual-processing model”: Only irregular gender assignment is lexically represented but regular assignment is accounted for by (a) symbolic rule(s) applying to the information about the noun stored in the lexicon. Assignment rules can refer to semantic, phonological and morphological information. The problem with rule-based theory on gender assignment is that it is only little embedded in the wider context of lexical storage and processing. Assignment rules presuppose the lexical information mentioned above without investigating further its status and structure. At a closer look, it is often unclear whether gender should be predicted from this information or whether gender itself is used to determine such information. Furthermore, using a “dual-processing model”, regular inflectional information e.g. is also provided by rules. This

means that assigning regular gender on the basis of morphological information can involve several rules at different areas of lexical information. Despite less memory load due to less stored information, such kind of processing will increase online computational load (cf. Butterworth 1983:259); thus regular assignment would take longer than access irregular gender stored in the lexicon. In an alternative model for lexical representation of morphological information proposed by Bybee (1985; 1988), there are no symbolic rules and hence no distinction is made between regular and irregular patterns. In this "network model", regularities between different information stored in the lexicon are accounted for by connections made among these related items (cf. Bybee 1995).

In my section paper, I would like to take a closer look at how different theoretical approaches to gender assignment embedded in the wider field of lexical representation and processing can shed light on the status of the lexicon in the theory of language. The presentation will be based on data from Indo-European, mainly Germanic, and from the initial results of my master thesis on gender assignment in Scandinavian.

### References

- Butterworth, B. 1983. Lexical representation. In B. Butterworth (ed.) *Language Production*. Vol. 2, 257-294. London: Academic Press
- Bybee, J. 1985. *Morphology: A study of the relation between meaning and form*. Philadelphia: John Benjamins.
- Bybee, J. 1988. Morphology as lexical organization. In M. Hammond & M. Noonan (eds.) *Theoretical morphology*, 119-141. San Diego: Academic Press.
- Bybee, J. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10(5): 425-455.
- Corbett, G. G. 1991. *Gender*. Cambridge: CUP.

### 3. Encoding of Verbs in English and Bulgarian

*Rositsa Dekova, NTNU, Trondheim, Norway*

One of the major problems concerning the Lexicon is to discover and represent the inner knowledge that native speakers have about the words in their language. This information is implicit, unconscious and demonstrated only in their ability to use the words. Our task is to make it explicit what information is encoded in the lexical representation of the words. The knowledge that speakers demonstrate suggests that there is something more than idiosyncratic word-specific properties, which needs to be considered for in the word representation. It is the information about the syntactic environment, in which the particular word can appear, that should also be part of the lexical encoding, and this is particularly relevant for verbs.

The format in which the semantic information is represented is also an important task. On one hand, there is mental representation - the way that information is possibly stored and accessed in brain. On the other hand, there is formal representation - the linguistic efforts to create a proper format resembling the mental representation that speakers have for words. Therefore we try to capture the way language is organized within the cognitive human capacities and represent it in a similar but formalized format.

The current paper focuses on the formal representation of the information that can be encoded in the verbs as lexical entries. I have used corpora data, as well as some test results, to determine what kind of participant information should be included in the lexical encoding of verbs. I have worked with English and Bulgarian building on the idea that there exists a common component in the meaning of verbs across languages. This existence originates in the similar cognitive perception and it is manifested in a number of semantic features shared between corresponding verbs from different languages.

With a case study on some verbs, a subgroup of *Verbs of Contact by Impact* (as defined in Levin, 1993), I show that verbs across languages do not differ much in the participant information they semantically encode, and therefore similar formal representations can be adopted and used in the Lexicon.

The formal representation is initially based on the Sign Model (Dimitrova-Vulchanova 1996/99), but it further employs a number semantic features, as well as encoding of special participant information, that can lead to predictions about the possible morpho-syntactic environment of a particular verb.

In conclusion, I try to show that breaking up the information to be encoded into relevant semantic features and using a suitable formal representation is crucial in finding a unified format of representing lexical entries across languages. Thus a more formal (and probably more accurate) comparison of verbs (their meaning and their syntactic behavior) can be achieved, as it would be possible to compare also verbs that encode more/less information than their correlates.

### **Bibliography:**

- Dimitrova-Vulchanova, M. 1996/99. *Verb Semantics, Diathesis and Aspect*. Doctoral dissertation, NTNU (University of Trondheim).
- Levin, B. 1993. *English Verb Classes and Alternations*. Chicago and London: University of Chicago Press.

## **4. On the organisation of the lexicon: evidence from the verb morphology in Greek**

*Alexandra Galani, University of York, Britain*

Data: An analysis of a rich inflectional language where information regarding the conjugational classes is also encoded in the system, requires the morphemes -consisting the forms- to represent syntactic, phonological as well as morphological features (class). In Modern Greek (MG) verbal forms, it could be argued that what follows the root is the theme vowel (TV) representing morphological features. These features additionally show that there is a certain degree of interrelation between TVs, further supported by the degree of frequency and regularity verbs -belonging to different conjugational classes exhibit. The importance of such features is further highlighted by what seems to be frozen suppletive stems in MG (deriving from the Ancient Greek counter-examples). An additional ultimate advantage is that a diachronic relation is also captured.

Issues: This particular analysis of the verbal morphosyntax in MG causes problems for a formulation of the account within Distributed Morphology (DM) (Halle and Marantz 1993). It seems that there is not a systematic way in line with which vocabulary items (VIs) are organised in the vocabulary (DM's lexicon). Halle and Marantz claim that VIs are arranged in terms of specificity; the most specified entry precedes items that are less specified. Nevertheless, this makes the wrong predictions about the degree of regularity, frequency and diachrony, one would want to account for MG. Clearly, there is not also space for any type of interrelation between the VIs. Overall, a simplified view of the DM Vocabulary does not allow us to account for all the features and processes under which word formation occurs.

An Alternative: I propose an extended version of DM's Vocabulary which is, nonetheless, more restricted than the function of the lexicon under other frameworks (Di Sciullo and Williams 1987, Chomsky 1992, Lieber 1992, Aronoff 1994). I abandon the idea that VIs are organised into blocks based on the degree of specification. Instead, I assume that VIs are hierarchically organised in tree structures rendering similarities with the hierarchical grouping of the structural levels (DS, SS, MS, PF) and also Head-driven Phrase Structure Grammar (HPSG, Koenig 1999). This hierarchy is based on a key feature of a language. As far as MG is concerned, the hierarchy is based on the degree of markedness of TVs determined by the degree of frequency and regularity TVs present in the forms in which they appear. The way the morphological and syntacticosemantic features are interpreted during the process of Vocabulary Insertion is based on the order in which each feature appears in the Vocabulary tree structure. If the features of the morphological node match the morphological features of the Vocabulary, the second set of features –the syntacticosemantic- compete for insertion. Once all features are checked and Vocabulary Insertion is complete, the structure enters PF.

Consequences: This extended view shows important advantages for the purposes of word formation and the treatment of allomorphy. These are explored in the final section of the paper in detail.

## **5. Integrating lexical and grammatical typology**

*Claudia Gerstner-Link, University of Munich*

The functional-typological approach to grammar is well-founded, yet the challenge remains to work out significant systematic typological properties of the lexicon. The author argues from a theoretical and practical viewpoint based on a chapter of "Lexical Typology" in her grammatical description of the Papuan language Kilmeri (Papua New Guinea).

One major issue concerns the range of classificatory strategies associated with nouns. Lexical classificatory devices in the nominal domain are noun classifiers and measure nouns. Noun classifiers are normally free lexical items standing next to the classified noun in a noun phrase; however, one also finds systems of petrified compounding structures classifying particular semantic domains of nouns. Measure nouns often build collocations with other nouns. They seem to exist in all languages, yet it is less than clear what semantic domains are generally classified by measure nouns and

what range of nouns in a language actually may serve as measuring means. It should be possible to approach this open class of nouns in terms of universal cognitive features.

Further systematic semantic issues arise in particular with special conceptual verb classes. Perceptual verbal expressions may be collocations and can be distinguished according to control and intensity. Motion verbs can be characterized according to their inherent components of PATH, MANNER, FIGURE, GROUND or combinations thereof. Physical, emotional and cognitive verbal expressions are very often constructions involving body part terms. Here one has to ask which body part terms are primarily referred to and what types of metaphorical structure and extension are found. All these lexical fields illustrate the opposition single lexical item (language A) vs. constructional pattern (language B) and point to the desirability or even necessity to include constructions into the lexical account of a language. This issue culminates in languages that have only a small number of generic verbs often leading to complex constructional patterns that give rise to specific schematic formulas. Here we are confronted with lexical-typological properties with direct consequences for syntactic analysis. The domain of epistemics extends to the question of what kind of information can be lexicalized - or vice versa: can be grammaticalized. Epistemic meaning and information is an important example for the interface of lexical and inflectional structure of a language. We find verbs, adverbs, or auxiliaries on the lexical side as against verbal suffixes on the grammatical-inflectional side. Are there languages without epistemic verbs and adverbs, since the epistemic domain is almost completely grammaticalized?

It is proposed here to think of the formal structure of (larger) semantic domains or (smaller) lexical fields in terms of a cycle (analogous to the well-known morphological cycle): from constructions to compounds or derivations to single words (to constructions). Compounding is the first step towards lexicalization; by obscuring the meaning of one of the compound elements we arrive at derivational patterns. By further obscuring the morphological structure and complexity the single word level is reached.

Illustrating examples are given from Papuan, Austronesian, and Australian languages including Kilmeri, the target of the author's fieldwork for several years.

## **6. Verbal polysemy vs. underspecification: the case of *open***

*Wilhelm Geuder, Universitaet Konstanz, Germany*

Meaning variation of verbs often correlates with the choice of different sorts of arguments, especially as direct objects. This phenomenon is investigated here for the case of the German verb *oeffnen* and its English counterpart *to open*, focusing on sense variation with concrete, spatial readings (without addressing metaphorical extensions).

*Open*-verbs have been described by some authors as polysemic networks of independently listed senses. For instance, Taylor (1992) posits a network of variants which clusters in three major subgroups: (a) "move apart": *open a book / one's eyes*; (b) "create aperture": *open the door / the lid*; (c) "make accessible": *open a room*. This looks like a case of polysemy in that there is apparently no uniform semantics which unites all the variants. However, Taylor considers verb+object combinations as the basic units of classification (and of learning), so it is not quite a standard case of polysemy, which would consist in a distinction of different entries of one lexeme, i.e. the verb *open*. The

question thus remains to which extent the pattern of variation involves one entry of the verb or several different entries.

In this paper, I argue that the verb *open* should be given a uniform representation in terms of an abstract schema that combines topological and functional features (drawing from work in cognitive semantics, e.g. Talmy 2000, and from formal models of spatial cognition, esp. Zwarts 2003). In a nutshell, the verb denotes the creation of a configuration in which the shape of a *bounded region* intrinsically projects a designated axis, connecting points inside and outside the region such that no point of the boundary is on this axis. Functionally, this axis yields a potential path for access to the bounded region.

A prototypical instance of this configuration is a container with one opening for things to pass, e.g. a bottle. Crucially, however, the topological parameters (container, boundary, access axis) are not to be equated with the arguments of the predicate; rather, the verb *open* can be applied to a concrete object as its theme argument if the properties of this object can be subsumed under the abstract spatial schema; this subsumption process depends on inferences from pragmatic principles and world knowledge. Hence, the modulation of senses (as in Taylor's classification) cannot easily be localized in lexical entries, but is due to compositional processes and inferences during composition. On a purely lexical-semantic level, no polysemy is needed.

After introducing the abstract schema that underlies the semantics of the verb "open", I show how it is applied to data discussed by Taylor for English and by Pause & Heitz (1998) for German "oeffnen". In particular, it is shown that the difference between Taylor's (b) and (c) types arises predictably in the process of subsuming the different objects under the same schema; we are not dealing with a "metonymic" alternation that was posited by Pause & Heitz. Likewise, variants involving "created objects" (like: "to open a gap") are shown not to be lexically independent readings.

## 7. Confixes in Modern Greek and Italian: Borderline cases between Lexicon and Grammar

*Giannoula Giannoulopoulou, Aristotle University of Thessaloniki, Greece*

This paper examines a word formation phenomenon, which lies between compounding and derivation, i.e. the 'confixation'. We define as 'confixes' the morphemes that:

- (a) do not coincide with stems, or when they do coincide with them, they differ from a semantic point of view e.g. Gr. *piis-i* 'poetry' vs. Gr.-*piis-i* '-ization'.
- (b) come from Classic Greek or Latin elements, and
- (c) are usually used in the International Scientific Vocabulary, but are also diffused in the general vocabulary (cf. Petrounias 1997).

In the present examination 43 morphemes are examined (e. g. the Greek *evro-*, *poly-*, *-odis* 'euro-, multi-, -oid' and the Italian correspondent ones *euro-*, *multi-*, *-oso*). The paper draws from the theoretical framework of grammaticalization and from the

theoretical framework of “Emergent Grammar” in particular (Hopper 1987, 1998b) and argues that confixes:

1. present lexical meaning (such as the constituents of the compounds) but they change the constituent order in the word so that the internal word order resembles the constituent order of derivation;
2. are created by secretion;
3. their specific meaning is derived from the “conventionalization of implicature”;
4. for their formation the diachronic factor plays a significant role since they are old morphemes that are re-introduced in the modern language through loan-words and neologisms;
5. are found in the International Scientific Vocabulary and in the vocabularies of mass media and politics and therefore their grammaticalization is observed in particularly restricted discourse contexts;
6. can be interpreted as products of a model-word because this function can describe the growing frequency of an item which is ex-free and ex-lexical and
7. constitute cases of ‘incipient grammaticalization’ (Hopper 1998a) because they are “emergent regularities that have the potential for being instances of grammaticalization” (Hopper 1991).

The main theoretical issues that the paper addresses are: 1. The formal approaches (lexicalist and syntactacist position within generative morphology) cannot describe satisfactorily the status of such morphemes because they deny to recognize that the linguistic categories are not separated with clear-cut distinctions. 2. The study of such morphemes challenges the *langue – parole* and synchrony – diachrony dichotomies. 3. Grammaticalization theory is an appropriate theoretical framework because has the possibility to describe the emergence of Grammar through historical and contextual pressures. 4. The study of the Lexicon –and especially of borderline cases such as confixed words- can be fruitful for the theory of language since these cases show how Lexicon and Grammar are not separate sectors but are in a continuous interplay.

## References

- Hopper P., 1987. “Emergent Grammar”, *Berkley Linguistics Society*, 13: 139-157.
- Hopper P. 1991. “On some principles of grammaticalization”. In C.E. Traugott & B. Heine (eds.) *Approaches to Grammaticalization*. Amsterdam: Benjamins, v. I., 17-35.
- Hopper P., 1998a. “The Paradigm at the End of the Universe”, in A. Giacalone-Ramat & P. Hopper (eds), *The limits of Grammaticalization*, Amsterdam: Benjamins, 147-158.
- Hopper, P. 1998b. “Emergent Grammar”. In M. Tomasello (ed.) *The new Psychology of Language. Cognitive and Functional Approaches to Language Structure*. London: Lawrence Erlbaum Associates, Publishers, 155 –175.
- Petrounias, Ev., 1997. “Loan Translations and the Etymologies of Modern Greek”, in *Greek Linguistics 95*. Salzburg: University of Salzburg, 791-801.

## 8. The Smell of Language

*Ilona Herlin & Laura Visapää, University of Helsinki, Finland*

*Linguistics of the senses*, as we call it, aims at describing the various ways in which the cultural and biological aspects of sensory experience are conceptualized in a natural language (Herlin & Visapää 2004). Language on the whole is strongly governed by embodiment, experientialism and anthropocentricity, which is illuminatingly exemplified by the cognitive domains of the senses (e.g. pitch, colour, temperature) and their linguistic encoding (see Koptjevskaja-Tamm & Rakhilina forthc.). By looking closely at the relationship between language and the senses, one comes more and more convinced of the fact that language – although clearly being connected to the physical parameters of sensory experience – does not reflect an “objective” reality but provides us with a set of conventional constructions that serve the means of human interaction and conceptualization.

Having compared temperature adjectives in Russian and Swedish, Koptjevskaja-Tamm and Rakhilina stress that the temperature domain is not conceptualized in relation to an objectively measurable scale and that the lexicon of temperature adjectives should be analyzed with respect to human interaction. More evidence of the experiential nature of “sensory” linguistic meaning is to be found in the domain of taste (Herlin & Visapää 2004). Despite the fact that the sensory experience of taste can be traced to the physical receptors in the nodules of the tongue, it is not the physical experience as such (‘sweet’, ‘salty’, ‘sour’, ‘bitter’, ‘umami’ etc.) that is worth mentioning but, normally, our personal preferences – and other practically motivated matters (e.g. comments on spoiled food).

In our presentation, we concentrate on the conceptualization of yet another sensory domain, that of smell. The linguistics of smell differs from the study of (e.g.) colour, taste and temperature in that smell is neither measurable nor analyzable into smaller units. Humans can distinguish around 10 000 chemicals, but for some reason languages seem not to be “interested” in encoding this variety. In our paper, we look at the small amount of Finnish adjectives of smell; the ways in which smell is lexicalized; and, finally, how smell sneaks into our everyday interactions. Scents seem not to provoke acts of description but those of affect and evaluation; smells *do* things to us (to be seen in the participial adjectives, e.g. *etova* ‘disgusting’ [from the verb *etoa*, ‘to disgust’]) and, furthermore, are often untraceable (e.g. *Täällä haisee* “it is smelling”).

Our paper not only addresses the linguistics of the senses but at the same time cross-cuts to the typologically interesting category of adjectives. Many authors have found adjectives a particularly weak category within and across languages (e.g. Givón 1984). Most of the typologists proceed from the adjectival meanings established by Dixon (1982: dimension, color, age, value, physical property, speed, human propensity). Our starting point is rather different: by concentrating on the fuzzy and marginal category of scent adjectives we try to trace down their interactional and cognitive motivation – what are the adjectives of smell *really* used for – and thereby approach the larger question of when and why human experience is conceptualized using adjectival forms.

## References

- Dixon, Robert 1982: Where have all the adjectives gone? – In Dixon, R. M. W. (ed.), *Where have all the adjectives gone? and other essays in semantics and syntax*. Mouton, Berlin; New York.
- Givón, Talmy 1984: *Syntax. A functional-typological introduction*. John Benjamins, Amsterdam.

Herlin, Ilona & Laura Visapää 2004: *Aistien lingvistiikkaa*. Kurssi Helsingin yliopiston suomen kielen laitoksella keväällä 2004. [Linguistics of the senses. A course given at the University of Helsinki in the spring of 2004]

Koptjevskaja-Tamm, Maria & Ekaterina V. Rakhilina forthc. "Some like it hot": on the semantics of temperature adjectives in Russian and Swedish. Manuscript.

## 9. Synonyms as contextual units of meaning

*Jarmo Jantunen, University of Joensuu, Finland*

It is widely accepted and recognised that absolute synonyms are extremely uncommon, if not impossible (e.g. Lyons 1968, Cruse 2000). This is often explained by cases which show that synonyms have differences in their emotional (Lyons 1968) or expressive meaning (Cruse 1986), or that they differ on the stylistic level (Saukkonen 1984, Cruse 2000). Corpus analyses have shown, in turn, that co-occurrence preferences and restrictions direct the use of synonyms and distinguish them from each other. Usually these analyses focus on collocational restrictions (Persson 1990, Stubbs 1995, Biber et al 1998; see also Cruse 1986), but also semantic (Stubbs 1995) and recently morphological restrictions (Arppe 2002) have been touched.

However, little attention has been focused on synonyms as contextual units of meaning, i.e. as a component of a larger syntagmatic pattern which includes both lexicogrammatical and semantic choices. Sinclair (1996, 1998) states that a lexical item consists of five components: the core itself, collocational and colligational preferences, and semantic preference and prosody. Furthermore Jantunen (2004) has noted in line with Arppe (2002) that morphological preference is also an essential contextual association pattern that should be taken into account when the aim is to describe contextual patterns in morphologically rich languages such as Finnish.

The purpose of this presentation is to find systematic differences at every level of co-occurrence tendencies between synonymous expressions and to show how these tendencies cut across all levels of syntagmatic patterning. By analysing the cotext I attempt to find reasons for why one of the synonymous expressions sounds more normal in a given cotext (1) than others (2, 3).

- 1) Hän ajaa *hyvin* hitaasti. ('She/he drives very slowly.')
- 2) Hän ajaa *kovin* hitaasti.
- 3) Hän ajaa *oikein* hitaasti.

The data for the present study consist of a corpus of two million words. The corpus is a subset of the Corpus of Translated Finnish compiled at the University of Joensuu and contains contemporary fictional and non-fictional texts written originally in Finnish. The contextual association patterns of synonymous expressions are studied analysing three semi-grammatical words namely the synonymous degree modifiers *hyvin*, *kovin* and *oikein* (all roughly meaning 'very').

## 10. Lexical structure of verbs: a comparison of models and formalisms

*Elisabetta Jezek, Università di Pavia, Dipartimento di Linguistica, Italy*

In light of the central role recently assigned to the lexicon in linguistic theory, I would like to discuss the most significant formalisms put forth within the formal and the functional framework to represent the various kinds of lexical information that verbs are assumed to have (i.e. argumental, aspectual, semantic, combinatorial information).

As is widely known, the description of verbs from a lexical perspective constitutes a difficult task, mainly because verbs are predicative elements and as such select arguments and impose constraints on this selection (unfortunately, these constraints are not easy to define, because of the frequent lack of correspondance between semantic and syntactic valency, because verbs are highly polisemic and because single verbs may allow multiple syntactic realizations of their arguments).

Starting from the assumption that verbs contain semantic, aspectual and syntactic information (among other kinds of information), I will firstly discuss each of these levels separately, and illustrate briefly the most significant formalisms that have been proposed in order to represent these properties. Secondly, I will discuss one of the main problems that these formalisms need to account for, i.e. the interrelation between the various levels (I will focus my attention on the interrelation between the semantic properties of a verb and its syntactic behaviour). Thirdly, I will raise the problem of establishing whether a specific piece of information (i.e. a semantic or an aspectual feature) is part of the lexical structure of the verb or is the result of compositional processes (this last part addresses the complex problem of the compositional nature of meaning).

Ultimately, I will discuss the application of the formalisms used to represent verb lexical properties to other word classes. In particular, I will discuss the application of aspectual and argument-structure parameters to the study of Nouns.

In the discussion I will make use of examples mostly taken from Italian, but the argumentation aims at capturing generalizations that are valid in a broader perspective.

The purpose of the presentation is to offer an account of the tools at disposal for the researcher to represent the lexical information for verbs, discuss the advantages and disadvantages of the various proposals, and outline the crucial problems which characterize this domain.

### References

- Dowty, D. (1979) *Word Meaning and Montague Grammar*, Dordrecht, Reidel  
Grimshaw, J. (1990) *Argument Structure*, Cambridge Mass., The MIT Press  
Jezek, E. (2003) *Classi di verbi tra semantica e sintassi*, ETS edizioni, Pisa  
Levin B. & Rappaport Hovav M. (1998) "Building verb meaning", in M. Butt & W. Geuder (eds.), *The Projection of Arguments: Lexical and Compositional Factors*, Stanford, CSLI Publications, 97-134  
Olbertz, H., Hengeveld K., Garcia J. S. (eds.) (1998) *The Structure of the Lexicon in Functional Grammar*, Amsterdam, Benjamins  
Pustejovsky, J. (1995) *The Generative Lexicon*, Cambridge Mass., The MIT Press

van Valin R.D. & LaPolla, R.J. (1997) *Syntax. Structure Meaning and Function*,  
Cambridge, Cambridge University Press

## 11. The limits of lexicon

*Päivi Juvonen, Stockholm University, Sweden*

Pidginization, i.e. the creation of a restricted language variety for specific language contact situations, minimizes not only the grammatical means speakers have at their disposal, but also the number of lexical items dramatically. Whereas an adult speaker of for example Swedish may have an active vocabulary of around 60 000-100 000 lexical items (The Academy of Sweden Word List has 120 000 entries), a speaker of a pidgin normally has only a couple of hundred words to his/her disposal. The questions addressed in this paper are 1) what is the character of this ripped-of or minimal lexicon in pidgin languages? and 2) can we find support for different theories of lexical organization and lexical universals in the organization of pidgin lexica?

More specifically, these questions will be discussed in light of a pilot study of the lexical organization of three pidgin languages: Chinook Jargon, Yokohama Pidgin Japanese and Plains Sign. The data, consisting of known word lists and vocabularies, but also some written texts, have been analyzed in terms of 1) part-of-speech organization, 2) the semantic primes of Wierzbicka (1996) and Goddard (2001) and, 3) in terms of typological hierarchies for lexicalization proposed by e.g. Talmy (1985) and Viberg (1993). The pilot study also raises a number of methodological questions that will be touched upon.

## References

- Goddard, Cliff 2001 Lexico-semantic universals: A critical overview. *Linguistic Typology* 5-1, pp. 1-66.
- Talmy, Leonard 1985 Lexicalization patterns: semantic structure in lexical forms. *Language Typology and Syntactic Description, vol. 3: Grammatical Categories and the Lexicon*, ed. by Timothy Shopen, 57-149. Cambridge: Cambridge University Press.
- Viberg, Åke 1993 Crosslinguistic perspectives on lexical organization and lexical progression. In: Hyltenstam, K & Å. Viberg (eds.): *Progression & regression in language*. 340-385 Cambridge: Cambridge University Press.
- Wierzbicka, Anna 1996 *Semantics : primes and universals*. Oxford: Oxford Univ. Press.

## 12. The anomaly of the verb "give": a study in lexical typology

*Seppo Kittilä, University of Turku, Finland*

"Give" is formally an anomalous verb in a number of languages. This has been discussed, for example, by Borg and Comrie (Borg and Comrie 1984), who clearly show that "give"

differs from all other trivalent verbs in Maltese. Borg and Comrie do not, however, make an attempt to explain this formal anomaly. An explanation is proposed in the present paper. The explanation is primarily formal, even though semantic aspects (lexical information) of "give" are also touched upon.

If we take a closer look at the anomaly of "give", we soon notice that the characteristic feature of "give" is its high formal transitivity: "give" is almost without exception among those verbs, which have two direct object-like arguments, if the language in question has double object constructions. This means that both the Theme and the Recipient arguments of "give" behave morpho-syntactically in the same way as the transitive Patient (direct object). The most evident manifestations of this are illustrated by cases in which both objects of "give" bear the same case marking as the transitive Patient (this applies both to marking on nominal arguments and to verbal cross-referencing of the arguments). This is either obligatory, as in Pitjantjatjara or Erromangan, or optional, as in English or Djaru (in the latter case "give" may have two direct objects via dative shift, while many other verbs cannot). Also other features of formal transitivity like verb morphology (basic vs. derived) and passivization render "give" a formally highly transitive verb. There are, however, also cases in which this analysis is not applicable, and in which "give" is either outranked in formal transitivity by other verbs or a given language lacks formally ditransitive constructions. However, "give" is without exception the only ditransitive verb in any given language, if a language has only one such verb. The number of such verbs varies cross-linguistically, as do the extent to which verbs have to deviate from "give" to be marked differently.

The present paper examines the anomaly of the verb "give" from a cross-linguistic perspective. The goal is to show that "give" is a formally highly transitive verb, which distinguishes from most or even all other trivalent verbs in many languages. Also semantic features lexicalized in "give" will be considered, and it will be shown that "give" semantically outranks many trivalent verbs in transitivity. "Give" does, for example, usually imply a higher affectedness of the theme and the distribution of the semantic roles of agent, theme and recipient necessitates three overt arguments.

### **13. Recategorization and constructions: The case of copular sentences with bare nouns in French**

*Peter Lauwers, Katholieke Universiteit Leuven, Belgium*

This paper deals with a series of copular constructions in French taking bare nominals as a predicative complement (*attribut du sujet*):

- Mon frère est très professeur. ('shows the qualities and defaults inherent to the status of a teacher; is very 'teacher-like')
- Ce film est très théâtre. ('shows the qualities and defaults inherent to theatre; is very 'theatre-like')
- Je ne suis pas fromage. (I don't like cheese, I don't eat cheese very often)
- Cet été sera très livre. ('during this summer, the focus will be on literature')

Within traditional grammar, and, more recently, in the work of Tesnière (1959) and of

those who endorse his concept of *translation*, this type of use of bare nouns has often been considered as (partly) adjectival.

After a brief sketch of the syntactic and semantic properties of these constructions, I will examine them in the light of some crucial theoretical questions. First, one has to call into question the status of these constructions in grammar. At first sight, a lexical treatment of the problem (subcategorization frame; selectional restrictions), in the form of a fine-grained valency-based analysis of *être* seems to be the best way to approach the issue. However, then other copular verbs can be inserted in this frame, such as *rester* ('stay'), *devenir* ('become'), *s'annoncer* (e.g. *la musique devient très 'Pierre et le loup'*; *l'été s'annonce très chanson*) and the restrictions (Givón 2001, Lamiroy-Melis, to appear) they in turn impose on the construction, have to be taken into account as well. This observation shows the need for a supra-lexical constructional treatment, which has to account for the incompatibilities (*\*je deviens fromage* vs *ce film devient très théâtre à la fin*) and interactions (e.g. the role of grammatical aspect: *je suis devenu (plus) fromage*) between the constructional frames and the whole range of (semi-)copular verbs.

Second, our analysis yields the need for a very nuanced view on 'adjectivization'. The constructions examined in this paper are very 'open' frames, in which almost every noun can be inserted (if, of course, it respects specific syntactic or semantic constraints, as well as constraints to do with general knowledge of the world). Some of the nouns involved are moving towards (lexical) adjectivization, as can be shown by a set of syntactic tests (e.g. *peuple*, *chatte*, etc.), whereas others are clearly not. The latter maintain part of their nominal properties and adopt some new adjectival characteristics. Only the former can be considered as truly 'adjectivized'. Both theoretical issues are concerned with the interface between lexical items (both nouns and copular verbs) and syntax (constructions). The second problem discussed in this paper raises the question of lexicalization and recategorization, whereas the first deals partly with grammaticalization.

## 14. *Affixoidhungrig? Skitbra!:* Comparing Affixoids in German and Swedish

*Torsten Leuschner - Kristin Ascoop, University of Gent, Belgium*

For several decades, linguists working with German have been debating the theoretical status of so-called "Affixoide" ('affixoids'). Affixoids are productive word-forming elements in either a prefixing or a suffixing function: *Riesen-* 'giant', *bitter-* 'bitter' and the (in)famous *Scheiß-* 'shit' are examples of the former (i.e. "Präfixoide"), *-freundlich* 'friendly' and *-papst* 'Pope' are examples of the latter (i.e. "Suffixoide"). Prefixoids often have an augmentative or qualifying function (as in *Riesenchaos* 'huge chaos', *bittersüß* 'bittersweet') and are based on nouns or adjectives. Suffixoids, on the other hand, are mostly based on adjectives (*benutzerfreundlich* 'user-friendly') and rarely on nouns (*Literaturpapst* 'leading authority in literary criticism').

What makes the affixoid such a thorny issue in linguistics is its apparent status in-between two well-established categories, viz. lexeme and affix. After detailed, mainly

synchronic discussion (Schmidt 1987), many authorities on German word-formation seem to have concluded that there is no valid category of "affixoid" after all (e.g. Fleischer / Barz 1995; but cf. Motsch 1996). Recently, however, Stevens (fc.) has pointed out that the transparent and systematic diachronic relationship between affixoids and their etyma, which he describes in terms of grammaticalization, does justify speaking of "affixoids". The purpose of our contribution is to defend this controversial view from a new perspective, viz. cross-linguistic comparison.

Comparing affixoids across languages is particularly instructive in the case of Swedish, not only because Swedish can have prefixoids on verbs (something impossible in German), but also because Swedish linguists recognize only "prefixlika förleder" (*jätte-* 'giant', *skit-* 'shit', etc.) and/or "suffixlika efterleder" (*-vänlig* 'friendly', *-hungrig* 'hungry') but have not so far considered any higher-level category corresponding to "Affixoid" (e.g. Söderbergh 1968, Lundbladh 2002). In our contribution, we would like to present the results of a recent corpus-based survey of affixoids in present-day German and Swedish (Ascoop 2004), including a detailed comparison of a few particularly interesting prefixoids and suffixoids in both languages, in order to defend the status of "affixoids" in the lexicon and suggest ways in which the German and Swedish traditions can learn from each other through a comparison of their terminology and data.

## References

- Ascoop, Kristin (2004): *Affixoidhungrig? Skitbra! "Affixoide", ihr Status und Gebrauch im Deutschen und Schwedischen, kontrastiv untersucht*. Master's thesis, German Dept., Ghent University, Belgium
- Fleischer, Wolfgang / BARZ, Irmhild (1995, in collab. with Marianne Schröder): *Wortbildung der deutschen Gegenwartssprache*. 2., durchgesehene und ergänzte Auflage. Tübingen: Niemeyer
- Lundbladh, Carl-Eric (2002): *Prefixlika förleder*. Available for download at Svenska Språkbanken, <http://spraakdata.gu.se/ordat/pdf/Ordat17.pdf>
- Motsch, Wolfgang (1996): "Affixoide. Sammelbezeichnung für Wortbildungsphänomene oder linguistische Kategorie?" In: *Deutsch als Fremdsprache* 33 (1996), 160-168
- Schmidt, Günter Dietrich (1987): "Das Affixoid. Zur Notwendigkeit und Brauchbarkeit eines beliebten Zwischenbegriffes der Wortbildung." In: *Deutsche Lehnwortbildung. Beiträge zur Erforschung der Wortbildung mit entlehnten WB-Einheiten im Deutschen*. Tübingen: Narr, 53-101
- Söderbergh, Ragnhild (1968): *Svensk ordbildning*. 2:a upplagan. Stockholm: Läromedelsförlagen
- Stevens, Christopher M. (fc.): "Revisiting the Affixoid Debate: On the Grammaticalization of the Word." To appear in: Torsten Leuschner / Tanja Mortelmans (eds., in collab. with Sarah De Groot), *Grammatikalisierung im Deutschen*. Berlin: de Gruyter

## 15. Constructions with the verb *ägn*a in Swedish

*Nina Martola, Svenska Litteratursällskapet i Finland rf*

Within two Swedish corpora of about 20 million words each the far most common verb in combination with the preposition *åt* 'to' is *ägna* 'devote'. There are two main constructions, the reflexive construction *ägna sig åt ngt* 'devote oneself to something' and the transitive construction *ägna ngt åt ngt/ngn* 'devote something to something/somebody'. However, it is apparent from the occurrences that there are restrictions as to what the object (*ngt*) in the transitive construction can be. The most typical cases seem to be expressions of time: *tid, mycket tid, mer tid, sommaren/helgen* etc., *X veckor/månader/år* etc. (time, plenty of time, more time, the summer/ the weekend, X weeks/months/years) where X stands for an expression of amount. Other recurrent objects are *sitt liv, uppmärksamhet, utrymme* (one's life, attention, space). Beside the transitive construction with an PP with *åt* the verb *ägna* is also constructed ditransitively *ägna ngn/ngt ngt*.

In my paper I will discuss the relation between the verb *ägna* and the constructions it is a constituent of and particularly the relations between the different constructions. Another main point will be the relation between syntax and lexicon. The verb *ägna* tends to appear in several fairly fixed expressions: *ägna tid åt ngt, ägna stort utrymme åt ngt, ägna en tanke åt ngt* etc. (spend time with somebody/ doing something, use a lot of space for something, give a thought to something). In other words, we have to do with more or less automated sequences (Pauly and Syder 1983, Bybee 2004), but are they constructions of their own or do they operate within constructions? In my analysis I will use Construction Grammar as theoretical basis.

### References:

- Bybee, Joan L. 2004: The Emergent Lexicon. [www.unm.edu/~jbybee/emergent.htm](http://www.unm.edu/~jbybee/emergent.htm). 2004-06-07.
- Pauly, Andrew and Frances Hodgetts Syder. 1983. Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. *Language and Communication*, ed. by Jack C. Richards and Richard W. Smith, 191-225. London and New York: Longmans.

## 16. Multiword expressions in the lexicon: the case of Italian phrasal verbs

*Francesca Masini, Università Roma Tre, Italy*

This paper focuses on multiword expressions (MWEs), or «constructions», i.e. lexical units larger than a word that can bear both idiomatic and compositional meanings. MWEs are one of the main issues in current theoretical linguistics: apart from being the central notion in the framework of Construction Grammar (Fillmore, Kay & O'Connor 1988, Goldberg 1995, Kay & Fillmore 1999), they are the main concern of recent works by Ackerman & Webelhuth (1998), Booij (2002) and Jackendoff (1997).

The question is: how are these constructions to be dealt with? And, even more interestingly, what is the relationship between these constructions and the lexicon? The general claim here is that some syntactic or phrasal «constructions» correspond to or compete with morphological complex words, in particular with derived words. We also

assume that these MWEs are lexically determined: even though they have a phrasal structure, they are not created in syntax proper, but in the lexicon.

In the present paper we apply this general view to verbal modification, and more precisely to Italian phrasal verbs. Phrasal and particle verbs are very common in Germanic languages and have inspired much work in recent years (Dehé, Jackendoff, McIntyre & Urban 2002 being a case in point). Quite to the contrary, this seems to be a scarcely productive pattern in Romance languages.

In this general framework, Italian phrasal verbs are an exception. They are constructions consisting of a verbal base followed by a locative adverb, such as the following:

- |     |                                       |                       |
|-----|---------------------------------------|-----------------------|
| (1) | [mettere]V [giù]ADV                   | 'to put down'         |
| (2) | [mandare] V [avanti] ADV (un'impresa) | 'to run (a business)' |
| (3) | [fare] V [fuori] ADV                  | 'to kill'             |

Their widespread use in present-day Italian led us to investigate them more closely, in an attempt to understand how they developed. The diachronic-typological analysis shed new light on their origin, providing evidence that these verbs appeared as early as in Ancient Italian as an alternative means to morphological derivation. Thereafter, they gradually took on and nowadays they are to be considered as a productive lexical template for the formation of new verbs in Italian.

## References

- Ackerman, Farrell & Gert Webelhuth (1998), *A Theory of Predicates*, Stanford, CSLI Publications.
- Booij, Geert (2002), *Constructional Idioms, Morphology and the Dutch Lexicon*, in «Journal of Germanic Linguistics» 14 (4), 301-329.
- Dehé, Nicole, Ray Jackendoff, Andrew McIntyre & Silke Urban (eds.) (2002), *Verb-Particle Explorations*, Berlin/New York, Mouton de Gruyter.
- Fillmore, Charles J., Paul Kay & Mary Catherine O'Connor (1988), *Regularity and Idiomaticity in Grammatical Constructions: the Case of Let Alone*, in «Language» 64 (3), 501-538.
- Goldberg, Adele (1995), *Constructions. A Construction Grammar Approach to Argument Structures*, Chicago/London, The University of Chicago Press.
- Jackendoff, Ray (1997), *The Architecture of the Language Faculty*, Cambridge (MASS.), The MIT Press.
- Kay, Paul & Charles J. Fillmore (1999), *Grammatical Constructions and Linguistic Generalizations: the What's X doing Y? construction*, in «Language» 75 (1), 1-33.

## 17. Lexical functions and semantic classes to organize semantic links between lexical units in a dictionary

Myriam Mortchev-Bouveret, Université de Rouen, France

This paper deals with phraseology and word combinations in general and specialized lexicography. Phraseology is the set of phraseological word-like units or sentence-like units of a language, that is to say units that lack a certain freedom of combination and can be formulae, idioms, fixed or semi-fixed collocations (cf Cowie 2001). Phraseology can also be seen more broadly. At one end, a phraseological unit can be a fixed, completely holistic expression, which requires a global interpretation. At the other end, the unit can be partly compositional and sometimes accepts synonyms for one of the units. Mel'čuk describes these global vs. compositional properties in a model of Explanatory and Combinatorial Lexicology (Mel'čuk et al. 1984, 1988, 1992, 1999, Mel'čuk et al. 1995, Polguère 2003) using Phrasemes and Lexical Functions. Full phrasemes and Quasi phrasemes are the more or less fixed set phrases or words. In a language the majority of collocations are semi-phrasemes. To describe these collocations, Mel'čuk proposes a set of about sixty Lexical Functions of two kinds: paradigmatic Lexical Functions like synonyms, antonyms and syntagmatic lexical functions like modifiers, semantic derivatives, syntactic derivatives, support verbs. We will illustrate phraseology and word combinations with six examples of verbs, adjectives and nouns taken in ten different 20<sup>th</sup> century French dictionaries and we will show the way they deal with it more or less systematically. Then, following research about Lexical Functions in specialized lexicography (Dancette et L'Homme 2002, L'Homme 2003), we will present the results on French Data of a multilingual project in progress (French-English-German dictionary of biotechnologies, 1300 entries, published in 2002). This dictionary encodes semantic links between terms by the means of Lexical Functions. It also includes verbs and organizes sets of derivatives around a key term. Thanks to Lexical Functions, semantic relationship between words can be formalized in a systematic way. Following the principles of systematic lexicography, we will try to broaden reflexion to « lexicographic types ». « A lexicographic type is a group of lexemes with a shared property or properties, not necessarily semantic, which are sensitive to the same linguistic rules and which should therefore be uniformly described in the dictionary », (Apresjan 2002, introduction, xii). We will thus explore a way of using the Explanatory and Combinatorial model to generalize systematic information to sets of terms sharing semantic properties. In this direction we will examine the notions of verb classes (Levin 1995) and semantic class.

## **18. How are paradigmatic relations represented? The contrary case of antonymy**

*Lynne Murphy, University of Sussex, Britain*

Murphy 2003 argues that paradigmatic relations such as synonymy, contrast and hyponymy are not represented in the lexicon. When such relations are strictly semantic in nature, they are derivable from the conceptual representations or meanings associated with the words. For example, one knows that *cat* is a hyponym of *animal* because of what one knows about cats, not what one knows about the word *cat*. Other sets of words may be related *qua* words (form-meaning pairs) rather than just as meanings. In this case, Murphy 2003 argues that the relation between the words is metalinguistic; that is,

knowing that *couch* and *sofa* are synonyms is part of what one knows (or derives) *about* those words, but is not part of their lexical representation. However, the monograph's conclusion grants that the lexical relation of antonymy (e.g., *cold/hot*, in contrast to the purely semantic relation of opposition—e.g., *frigid/hot*) may be a special case, in part because antonymy is syntagmatic as well as paradigmatic in nature.

This paper follows up on the suggestion made in Murphy 2003 that antonym relations might best be represented as discontinuous lexical items in a construction grammar. The argument for special treatment of antonyms is based on several types of evidence. First, corpus evidence shows that antonyms co-occur at higher rates than non-lexical opposition relations (Charles/Miller 1989, Willners 2001). Second, experiments demonstrate that antonyms are so strongly linked that their particular meanings become somewhat irrelevant to our recognising them as antonyms (Charles et al. 1994). Semantic development of antonyms further supports this. For example, the antonymic link between *black* and *white* is so well-learned that *white* is used as an opposite for *black* even when the intended referent is not literally (e.g., *black/white coffee*), or even metaphorically, white (e.g., *black/white box testing*). Lastly, there is constructional evidence. Antonyms co-occur in a particular range of syntactic frames (Jones 2002), which can be argued to have contrastive meaning themselves. These constructions (as in (1)-(3)) have two slots in which antonyms, or other contrastive word pairs, occur.

- (1) post-nominal *X or Y*:  
Firstly, none of the characters, **straight or gay**, are particularly respectable  
(<http://www.cryptoclast.org/Opinion/homo/gaytv.htm>)
- (2) *not X, but Y*:  
Tolkien's great act of genius [...] was to turn the convention of quest stories upside down and make the ring **not good but evil**. (<http://www.movies101.com/LORD.HTML>)
- (3) *more AdjX than AdjY*:  
...he still seems **more innocent than guilty**  
([www.orbitalreviews.com/pages/full/APlaceInTheSun.shtml](http://www.orbitalreviews.com/pages/full/APlaceInTheSun.shtml))

This paper proposes that the antonym pairs themselves are lexical constructions that can be inserted into the types of constructions in (1)-(3) and that can trigger other opposite pairings, as in (4).

- (4) Ancillary Antonymy (Jones 2002)  
The **rich** get to choose and the **poor** get the queues.

Word pairs like *good/evil* and *rich/poor* are recorded in the lexicon as antonym pairs. Such constructions are based on a more abstract antonym construction, which constrains the types of words that can be paired as antonyms and which specifies their pragmatic/discourse (rather than just semantic) role as antonyms.

## References

- Charles, W. G. & G. A. Miller. 1989. Contexts of antonymous adjectives. *Applied Psycholinguistics* 10:3.357-75. F

- Charles, W. G., M. A. Reed, & D. Derryberry. 1994. Conceptual and associative processing in antonymy and synonymy. *Applied Psycholinguistics* 15: 329-54. F  
Jones, Steven. 2002. *Antonymy: a corpus-based approach*. London: Routledge.  
Murphy, M. L. 2003. *Semantic relations and the lexicon*. Cambridge: CUP.  
Willners, C. 2001. *Antonyms in context*. Lund: Lund University.

## 19. Co-occurrence of Antonyms in a Context

*Yoshio Narisawa, Tohoku Gakuin University, Japan*

*Masaaki Machiya, Hachinohe Technical College, Japan*

Antonymy is a special lexical association between word pairs. That it is lexical and not simply semantic follows from the fact that different words for the same concept can have different antonyms: for examples, *big-little* and *large-small* are good antonym pair, but *large-little* is not.

Charles and Miller(1989) argue, contrary to psycholinguistic theory, that the primary source of these associations is a tendency they hypothesize for antonyms to co-occur within same sentences. This paper extends Charles and Miller (1989) co-occurrence hypothesis, which states that antonymous adjectives co-occur in the same sentence with frequencies far greater than predicted by chance and Fellbaum's *Co-occurrence and Antonymy* which extends Charles and Millers hypothesis to nouns, adjectives and verbs.

In searching the COBUILD CD-DOM for co-occurrence of semantically opposed concepts for any part of speech, we found that words expressing antonymous concepts regardless of any part of speech co-occur within a sentence as well as in a neighboring sentence with higher-than-chance frequencies.

## 20. What is behind Word Sense Disambiguation

*Zeynep Orhan, Zeynep Altan, Department of Computer Engineering, Istanbul University, Turkey*

Word sense disambiguation is an important intermediate stage for many natural language processing applications. This paper deals with the difficulties in word sense disambiguation task in general and discusses the typical problems in the WSD applications due to the structure of Turkish language.

Lexical ambiguity which is an important problem at the bottom level of NLP applications, does not have a perfect solution. Word sense disambiguation (WSD) is an important intermediate stage for many natural language processing applications. The ultimate goal of the WSD research is determining the most suitable sense of an ambiguous word among the possible set of senses for that specific word in a given text or discourse.

The meaning of the central word can be decided unambiguously if we are allowed to see N neighbouring words, however the decision function does not solely depends on

those N words in human language processing system. Additionally there is even no consensus on the definition of meaning of a sense, but there are many different approaches for the meaning of a word in the history.

In WSD the first step must be determination of the set of applicable senses of a word for a particular context. This set can be formed by using sources like dictionaries, thesauri, translation dictionaries etc. The next step will be the selection of features that can be helpful for resolving the ambiguity. Then the last step will be the determination of the most appropriate sense in the given context.

Sense classification and granularity is one of the main problems. Determining the syntactic features and semantic components that are affecting the sense distinctions are very critical.

Electronic resources, such as dictionaries, thesari, bilingual corpora, parsers, morphological analyzers, ontologies like Wordnet are the critical factors for WSD systems. Other than these, the properties of the specific languages must be taken into account, since the senses of a word and the factors that are affecting sense distinctions vary dramatically in different languages.

We believe in that the solution of all these discussions and the WSD process is understanding the human language processing system. Today we know only a little about this system. Psychologists says that when a human hears or reads an ambiguous word, more than one senses of the word can be activated initially. But in a very short period of time disambiguation takes place in the human brain and that person chooses the correct sense in the given context. This given context must have a mental representation reflecting the properties of it. This specific representation must not have only a linguistic component, but also situational context and general knowledge components. We must have a coherent and plausible representantion model for the entities in the world analogous to the humans. If this first and the most important step can be completed and the problems stated above for computational stage can be overcome then we can have a very powerful disambiguation system. Otherwise, we can achieve a partial success only.

Keywords: Word Sense Disambiguation, Lexical And Contextual Ambiguity, Feature selection

## **21. On the interpretation of scalar antonyms and their negated equivalents**

*Carita Paradis, Department of English, Lund University*

*Caroline Willners, Department of Linguistics, Lund University*

Antonymy and negation are phenomena that can be studied from many perspectives. In the literature, antonymy is recognized as the most robust of the lexico-semantic relations, important to both the mental organization of the vocabulary and the organization of coherent discourse (Cruse 1986, Fellbaum 1998, Murphy 2003, Willners 2001 and Jones 2002). "Depending on whom you ask, negation may be a logical operator or a type of speech act, a basic element of semantic representations or a pragmatically loaded form of

communicative interaction" (Israel 2004: 701). In other words, negation is commonly used as an expression of oppositeness as well as a hedging device in discourse, e.g. *I don't know, but I think it's a good idea to...* (Giora 2004, Tottie 1991, Tottie & Paradis 1982).

This experiment investigates the interpretation of eight Swedish scalar antonym pairs and their negated equivalents. Our hypothesis is that the negation suppresses the accessibility of the expressed adjective, i.e. *not narrow* = 'wide' and *not wide* = 'narrow'. The experiment is designed as follows. There are 32 test sentences and 32 distracters and E-prime is used as experiment software. The adjectives studied are *ljus-mörk* (light-dark), *kort-lång* (short-long), *smal-bred* (narrow-wide), *låg-hög* (low-high), *lätt-tung* (light-heavy), *lätt-svår* (easy-difficult), *varm-kall* (warm-cold). Each test item is presented with and without negation in a fixed context. The informants are asked to grade their interpretation of an expression on a given scale, e.g. *Vägen längs kusten är smal* (The road along the coast is narrow) on the scale 'stig-motorväg' (path-motorway).

Our pilot study shows that there is consensus across the informants concerning the interpretation of the expressions included in the test. The response times for the negated expressions are longer than for the non-negated items, which may show that they are more difficult to process. However, at the time of writing, our results are still inconclusive.

## References

- Cruse, A. 1986. *Lexical semantics*. Cambridge: Cambridge UP.
- Fellbaum, C. 1998. *WordNet: An Electronic Lexical Database*, Cambridge (Mass): MIT Press.
- Giora, Rachel (2004). Negation as positivity in disguise. In Colston, H.L. & A. Katz: *Figurative language comprehension: social and cultural influences*. Hillsdale, NJ: Erlbaum.
- Israel, M. (2004). The pragmatics of polarity. In Horn, L. & G. Ward: *The handbook of pragmatics*. Oxford: Blackwell.
- Jones, S. 2002. *Antonymy: a corpus-based perspective*. Amsterdam: Routledge.
- Murphy, L. 2003. *Semantic relations and the lexicon: antonyms, synonyms and other semantic paradigms*. Cambridge: Cambridge UP.
- Tottie, G. 1991. *Negation in English speech and writing*. San Diego, New York, London: Academic Press.
- Tottie, G. & C. Paradis, 1982. From function to structure. Some pragmatic determinants in impromptu speech. In Enkvist, N-E *Impromptu speech*, Publications of the Research Institute of Åbo Akademi Foundation nr 78, Åbo.
- Willners, C. 2001. *Antonyms in Context. A Corpus-based Semantic Analysis of Swedish Descriptive Adjectives*. Travaux de l'Institut de Linguistique de Lund 40. Lund, Sweden: Dept. of Linguistics, Lund University.

## **22. The historical development of nominalizations in German and English, with a crosslinguistic comparison**

*Monika Rathert, University of Frankfurt, Germany*

In this talk, I investigate nominalization patterns in German and English both with diachronic and typological methods.

Nominalizations may be more close to either nouns (and, because of this, close to the lexicon) or sentences both in morphosyntax and semantics. However, it is less clear where the borderline is between noun-like and sentence-like nominalizations: how clearly may a nominalization carry markers for finiteness or tense/ aspect/ modality to be still counted among the noun-like nominalizations? If a nominalization easily combines with adjectives or cannot be negated, might it nevertheless be sentence-like? German and English produced quite different nominalizations in their different historical stages. Typological research has also found out that a great variety of combinations of noun-like and sentence-like features in nominalizations is attested, cf. Koptjevskaja-Tamm (1993).

I investigate these features and also want to find out in which ways they are related; for instance, what is the relation between finiteness and tense-marking in nominalizations? Moreover, which feature bundles represent productive patterns of nominalizations crosslinguistically? And which feature bundles are always doomed to be marginal, and why? Finally, which nominalizations occur easily side by side within a single language, and how could that be explained?

## **23. Multilingual Corpora in the Lexicon Construction**

*Liudmila Rychkova, Grodno State University, Belarus*

The *lexicon* is usually viewed as a compendium of all the lexical items of language. It is not equal to wordlist but is rather a system of inter-related lexemes formally presented by one token or more in the case of compound lexemes.

It is argued that multilingual approach can facilitate the lexicon construction (as far as it is possible for fuzzy sets); varieties within a language are not taken into consideration.

Let us limit to two domains each of which represents quite contrary approach to study of lexical meaning. Lexical-centred approach is represented from the point of view of nominative derivation (V. Nikitevich) and text-centred one is illustrated from the point of view of lexical-semantic variability.

Any word as a nominative unit can condensate in itself even complex word-combination. There are no other limits of the initial word-combination complexity except the limits of nomeme's structure. On the other hand, word-combination of any degree of complexity, which is semantically equal to the word as a nominative unit, is a single nomeme with a multi-token form. The word-combination of this kind is not a syntactic unit, but a part of the lexicon. It can serve as a communicative equivalent to a "content" word constructed by means of derivational morphology and sometimes is the only one possibility, which exists in a certain language to express the particular derivative

meaning. The notion of “derivative form” introduced by V.Nikitevich, presupposes existence of different ways of nomemes’ structural expression except the traditional affixation. From this point of view certain languages can choose “analytical derivation” as a main model for formation of nominative units. If there is no opposition to a single-word nomeme in the system of a certain language it is almost impossible then to find such units with an initial degree of lexicalization but still strong realized syntactic and semantic valencies. Comparative or multilingual perspective can serve as a strong tool to facilitate search of such units for the lexicon construction, for instance by extracting translation equivalents to collocations.

It becomes really possible with development of corpus linguistics, corpus-driven lexicography, and multilingual electronic resources, among which the multilingual corpora are of the greatest importance.

Text-centred approach to the lexicon is connected “to the general rule that each word when used in a new context is a new word”, contextual meaning being considered as “the functional relation of the sentence to the processes of a context of situation in the context of culture” (J.Firth). Therefore lexical-semantic variability is inevitable even for monosemic words, though the various meanings of a word taken from a certain language “have no clear-cut demarcation lines but rather a kind of hazy fringe through which they imperceptibly merge into each other” (S.Ullmann). Multilingual concordances derived from multilingual corpora do allow the distributional analysis of meanings and from this point of view can serve for the lexicon construction as a source of syntactic, semantic, and pragmatic information for every language represented in the multilingual corpora under consideration.

## **24. Finnish as a Semitic Language**

*Pauli Salo, University of Helsinki, Finland*

The nature of the lexicon has been subjected to intensive investigation in cognitive science, using a wide array of research methods from linguistics, neuroimaging and behavioral measurement, among others. Currently, there is still a debate between two opposing views concerning the function of the lexicon within an array of other cognitive processes. The lexicalist position holds that the lexicon and word-formation (WF) constitutes, at least partially, a separate component or module in the brain, subject to its own principles, processes and perhaps also localized to a different region of the brain than the phrasal syntax. The nonlexicalist theory, on the other hand, begins from the view that the creation of complex words is part of the syntax proper and essentially no different from creation of complex phrases.

The lexicalist position is often defended by a powerful argument which aims to show that the syntactic component of grammar would get too complicated if the derivation of zero-level categories is explained by using the same principles as the derivation of phrasal categories.

I argue that this challenge for the nonlexicalist theory of word formation, even if true as it stands, can be met when it is assumed that the zero-level derivation is non-categorical and the lexicon/WF is freed from lexical categories. Thus, it is argued that derivational morphology, i.e. creation of complex words such as "juoksuttautua", is part

of the syntax, but deals with categorically indeterminate 'semitic' roots and affixes which are not inherently nominal, verbal or adjectival. That is, affixes such as juokse-, -tta-, -u- are not syntactically verbal, nominal or adjectival. Zero-level derivation thus becomes 'syntax with category indeterminate words.' This hypothesis is in agreement with recent speculations that the open-ended productivity of the human brain results from a single underlying process that is applied to various domains, e.g. morphology, syntax, semantics, art, music, social cognition, and so forth, rather than from several such processes.

The theoretical background of my work is in Distributed Morphology. More specifically, the idea that lexical category features (N/V/A) are part of syntax proper and thus more like inflectional affixes than lexical features comes from the recent work of A. Marantz and his colleagues and likewise from my own recent dissertation in cognitive science, both which are within the generative tradition that aims to formulate a psychologically realistic theory of language learning and language use. The theory is applied to Finnish morphology.

Keywords

Lexicalism; Lexical categories; Word-formation; Distributed morphology

## **25. The status of lexis in a semiotic-functional integrative model of language**

*Miriam Taverniers, Gent University, Belgium*

In this paper, which has a theoretical orientation, I will propose that the status of lexis in an overall model of language can be described and theorized in terms of four types of semiotic relationships:

(1) EXPONENCE: the relationship between an elemental level (where the maximum unit is a word with categorial meaning) and a constructional level (where the maximum unit is a clause with different layers of functional structure) within lexicogrammar.

(2) DELICACY or MICRO-INSTANTIATION: the relationship between schematic (less specified) grammatical signs and more delicate (more specified) lexical items within the elemental level; and between lexicogrammatical structures and collocational patterns within the constructional level of lexicogrammar.

(3) TACTIC INSTANTIATION: the relationship between elemental signs (lexemes and morphemes) and morphological rules (morphotactics).

In addition to these three semiotic relationships, which pertain to lexis in its interaction with other components of language (especially grammar), I will define two types of intra-lexical relationships:

(1) MICRO-STRATIFICATION or CODING: the relationship between lexical formal meaning (or sense) and lexical items (i.e. derivational morphemes and lexemes).

(2) SENSE RELATIONS, or the different types of semantic relations into which particular lexemes enter (synonymy, hyperonymy, and so on).

Each of these semiotic relationships will be defined within the context of an overall linguistic model which is designed according to a semiotic and functional approach to language in general, and which integrates insights from four different linguistic schools: Hjelmslev's semiotic theory of language, Halliday's Systemic Functional Grammar, Coseriu's Integrative Linguistic theory, and Langacker's Cognitive Grammar. I will focus on the theoretical and descriptive motivation behind each type of semiotic relationship, and on the way in which the different types of relationships interact with one another.

One important theoretical and descriptive advantage of the model which will be presented is that it offers a detailed theoretical framework for providing a clear and finegrained definition of a number of concepts which have traditionally been used to specify what is meant by  $\mathbb{E}$ lexis<sup>1</sup> in language:

- Grammatical  $\mathbb{E}$ meaning<sup>1</sup> vs. lexical  $\mathbb{E}$ meaning<sup>1</sup>.
- Inflectional morphemes, derivational morphemes, and lexemes. Within the category of derivational morphemes a further distinction will be made between categorial morphemes (such as *-ion* in *modification* or *-able* in *modifiable*) and classematic morphemes (such as *-let* in *leaflet* or *-ess* in *princess*); both types of derivational morphemes (and lexemes) will be defined as lexical items.
- The notion of a  $\mathbb{E}$ word<sup>1</sup>: a distinction will be made between morphemic words (e.g. articles and auxiliaries), categorial words (e.g. *I* or *mine*) and lexical words. It will be argued that, in terms of a conception of lexicogrammar as a continuum (as defined by means of the graded relationship of delicacy or micro-instantiation, cf. above), categorial words are intermediate between lexis and grammar.

## 26. Lexicon or Information Structure?

*Anita Vizsket, Hungarian Academy of Sciences - Research Institute for Linguistics*

*Peter Buranyi, University of Pecs, Hungary*

The behaviour of bare NPs (bare NP: NP without any determiner) is somewhat different in Hungarian than in some other languages. In West Greenlandic, bare NPs do not occupy argument positions, so they can be considered as incorporated elements of the verb. In Hungarian, just the opposite is true. In some languages, singular bare NPs rarely occur, in English plural forms can get generic interpretation, singulars are used only in lexicalized constructions (except mass nouns, which can be used generically in singular form as well):

- 1) Bears are intelligent
- 2) I was watching TV.
- 3) Sugar is sweet.

In Hungarian, whereas bare NPs function as arguments and do not necessarily get generic interpretation, their use is restricted with regard to the predicate. In the literature studying the Hungarian argument structure influenced by the so-called Definiteness Effect, the

widely accepted viewpoint is that the bare NP is the (most prominent) subclass of the non-specific NPs.

In our presentation, we wish to confute this traditional suggestion on the one hand, and on the other to sketch, what circumstances make it possible for bare NPs to take the argument positions (subject or object) of some predicates.

Information influencing the possibility of occurrence of bare NPs as subject or object arguments show a great variety in the range of the different levels of grammar. They are connected with the telicity of the predicate and the aspect of the sentence; the phonological properties of the verb; the word order of the sentence (the filling of the immediately preverbal position); pragmatic facts (whether the denotation of the NP-verb construction is a general, everyday activity or not). All of these not only permit the bare NP argument, but restrict its complements as well (in some constructions the bare NP cannot have any complement at all, it can occur as a "quasi-head" only, yet its semantic properties do not exclude natural complement types).

Besides the fact that much information must be considered simultaneously arriving from the very different components of the grammar, the behaviour of bare NPs in Hungarian is strongly influenced by another fact: to what degree has the construction been lexicalized. We do not have real tests to examine this fact, no real classes can be distinguished from each other. Hungarian has totally lexicalized and totally productive forms. For example 'fát vág' (tree-acc cuts, 'to log') is lexicalized to a high degree, but 'öltönyt hord' (suit-acc wears, 'to wear suit') to a lesser degree, and 'énekkart alapít' (choir-acc founds, 'to found a choir') is not a lexicalized form at all.

Applications worked out for analyzing Hungarian constructions usually do not separate the components of the grammar. The lexicalized and rule based constructions are stored in the same way. By examining bare NPs, we can conclude the same. In this regard, the question arises: Is information stacked up on the same level, which can be called the Lexicon, or should we suppose that the traditional generative grammatical components are collapsed into a single information structure.

## **27. Where Lexical Semantics meets Spatial Description: A framework for *klettern* and *steigen***

*Matthias Weisgerber, University of Konstanz, Germany*

We want to propose a formal framework for lexically encoding manner-path-verbs. Our framework's main goal is to derive information from the interplay between *Manner-* and *Path-*information and from the objects involved: It uses conceptual and world knowledge, but keeps that information separate, which leads to small lexical verb entries.

Consider, as an example, the German verbs *klettern* and *steigen*. On the first sight, both are equivalents of English *to climb*. However, they differ in meaning: while *klettern* is approximately equivalent to *clamber* and *climb*, *steigen* has the meaning of *climb*, *go up*, *ascend* and *increase*. This seems to be a systematic difference:

- (1) Das Flugzeug steigt. (The plane is climbing (go up).)

- (2) Der Affe klettert. (The monkey is climbing (clamber).)

While (1) only expresses an increase of height, (2) only expresses the manner of the monkey's movement. So far, the lexical entries would look like in (3), where *x* denotes the subject and *P* a path, and where the brackets denote that this information can be overwritten by other information:

- (3.a) *steigen(x,P)*  
go(x,P)  
[DIRECTION: upward]

- (3.b) *klettern(x,P)*  
go(x,P)  
MANNER: clambering

However, not all objects can be used as subject of *klettern* and *steigen*:

- (4) \* Der Affe steigt. (the monkey is climbing (go up).)  
(5) \* Das Flugzeug klettert. (the plane is climbing (clamber).)

In the theory we propose, example (4) can be explained like that: *x* is of type *animateObject*, therefore the *Manner*-information [repeat(standOn(Place\subsetOf(Path)))] is added via the lexical entry *go(x,P)*, where conceptual knowledge about moving objects is stored. Due to the definition of *Place* and *standOn*, there must be contact to an object. Since no such object exists in (4), the sentence is not valid. In (5) the Manner=clambering doesn't fit the non-animate object.

We will discuss in a similar way how information about direction influences or changes the *Path* entry, as well as due to the subject, the *manner* entry changes.

Technically, our framework uses formalizations of spatial representation, and in particular, the notions of *Path* (investigated in various approaches -- we only want to mention Zwarts (2003, 2004) whose formal path algebra will be of good use here, Zee (2003), Nam (1995) and Jackendoff (1983, 1991)) and *Surface*. This interplay of conceptual information locates our framework at the interface between what Jackendoff (2002) calls Conceptual Structure (CS) on the one hand and Spatial Structure (SpS) 'encoding the spaial understanding of the physical world' (ibid., p.346) on the other. Features and mappings like *Solid(Object)*, *Surface(Object)* and *Place(Path)* build up the relation between conceptual knowledge and spatial description. This seems to contradict Jackendoff's hypothesis that 'the grammatical aspects of language make reference only to CS, not o SpS' (2002, p.348, and 1996), which we will discuss and reformulate.

## References

- Jackendoff, R. S. (1983). Semantics and Cognition. Number 8 in Current studies in linguistics series. Cambridge, Mass./London: MIT press.  
Jackendoff, R. S. (1985). Multiple Subcategorization and the Theta-Criterion: The Case of CLIMB. Natural Language and Linguistic Theory 3, 271--295.  
Jackendoff, R. S. (1991). Parts and boundaries. Cognition 41, 9--45.

- Jackendoff, R. S. (1996). The architecture of the Linguistic-Spatial Interface. In: Bloom et al (eds.), *Language and Space*. Cambridge, Mass.: MIT Press, 1--30.
- Jackendoff, R. S. (2002). *Foundations of language: brain, meaning, grammar, evolution*. Oxford: Oxford Univ. Press.
- Nam, S.(1995, 2). *The semantics of Locative Prepositional Phrases in English*. Phd dissertation, UCLA, Los Angeles.
- van der Zee, E. and J. Slack (Eds.) (2003). *Representing direction in language and space*. Oxford: Oxford Univ. Press.
- Zwarts, J. (2003). *Vectors across Spatial Domains: From palce to size, orientation, shape, and parts*. In: van der Zee and Slack (2003), pp. 39--68.
- Zwarts, J. (2004). *Prepositional aspect and the algebra of paths*. Technical report, Utrecht Institute of Linguistics. unpublished paper, <http://www.let.uu.nl/users/Joost.Zwarts/personal/downloads.htm>.

## 28. Five Arguments for an Emergent Structure of Lexicon

*Bernhard Wälchli, Institut für Sprachwissenschaft, Switzerland*

According to DiSciullo & Williams' (1987: 3) listeme-model, the lexicon "is incredibly boring by its very nature...and those objects that it does contain are there because they fail to conform to interesting laws." This paper argues that the lexicon is everthing less than a list of unconnected boring facts and that an emergent model of the lexicon (Bybee 1998) is more appropriate to for its description than the boring-list model. Five arguments for an emergent model of the lexicon are presented:

**Continuum between permanent and temporary lexemes.** If the mental lexicon were a simple list, each lexeme would be equally well-entrenched in the speaker/hearer's memory. However, there is much evidence that only part of the lexicon is permanent, the most permanent component of the lexicon, as is well-known from comparative linguistics, being its core vocabulary, whereas every conversation, every text can have its individual temporary lexicon shared only by a small community of interlocutors with an exclusive common ground.

**Expressive relexification.** Not even the most permanent part of the lexicon is really permanent, in all languages there is constant on-going relexification, due to extravagancy, which affects form more than meaning, which is why the "inner form" of the lexicon can remain very stable diachronically.

**Lexical classes.** The lexicon of every individual language is full of clusters of lexemes which, while not being grammatical classes (grams), share specific formal and semantic properties and have certain specific domains of use. Examples are the middle (reflexive), causatives, diminutives, agent nominals, phrasal verbs, noun incorporation and various types of compounds (Wälchli 2003). Many lexical classes are traditionally treated in word formation (derivation and compounding), but lexical classes are not restricted to words, many lexical classes in many languages are intermediate between words and phrases. Characteristic for lexical classes is their Janus-faced character (idiosyncratic and regular at the same time). Well-entrenched members toward the

permanent pole of the lexicon are often non-compositional in meaning while lexical classes at the same time can be fully productive in the temporary lexica.

**Only partial compositionality.** When describing the lexical units of a certain functional domain, such as motion, it is often not sufficient to list items on one structural level (only stems, only words, or only collocations of words). The lexicon must thus have a multi-layered structure and the layers are connected.

**Floating categories.** Many words in many languages cannot unequivocally be assigned to a single wordclass, they float between several wordclasses (e.g., verb and adposition). Diachronically lexemes with floating categories often represent transitional stages in grammaticalization from one wordclass to another.

The data presented to exemplify the arguments come from a great variety of languages, including varieties of Finnish, and from various functional domains, including motion, body parts, and natural coordination.

## References

- Bybee, J. 1998. 'The emergent lexicon' — *Chicago Linguistic Society* 34.2: 421-435.  
DiSciullo, A.-M. & Williams, E. 1987. *On the definition of word*. Cambridge, Mass.: MIT.  
Wälchli, Bernhard. 2003. *Co-compounds and natural coordination*. Stockholm: Stockholm University, Department of Linguistics.

## **POSTERS:**

### **29. Recent Progress on EDR Electronic Dictionary**

*Hitoshi Isahara, Kyoko Kanzaki and Takano Ogino, National Institute of Information and Communications Technology, Japan*

As the National Institute of Information and Communications Technology (NICT) is the sole national laboratory for info-communications research in Japan, NICT has set as its goal to become a world-class core research institute by promoting comprehensive research and development related to info-communications. It conducts research in various fields, including research on natural language processing (NLP). As a part of NLP research, NICT is compiling linguistic resources for academic research. This task includes compiling two-million word Japanese learners' corpus of English and precisely annotated multilingual corpus among Japanese, English and Chinese.

And also, we recently obtained all copyright of the EDR Electronic Dictionary. The EDR Electronic Dictionary was developed for advanced processing of natural language by computers, and is composed of eight sub-dictionaries. NICT distributes the EDR Electronic Dictionary for both academic and business use.

The EDR Electronic Dictionary is a machine-tractable dictionary that catalogues the lexical knowledge of Japanese and English (the Word Dictionary [Japanese: 270,000 words, English: 190,000words], the Bilingual Dictionary [Japanese-to-English: 230,000 words, English-to-Japanese: 160,000 words], the Co-occurrence Dictionary [Japanese:

900,000 occurrence, English: 460,000 occurrence] and the Technical Terminology Dictionary), and has unified thesaurus-like concept classifications (the Concept Dictionary [410,000 Concepts]) with corpus databases (the EDR Corpus [Japanese: 200,000 sentences, English: 120,000 sentences]). The Concept Classification Dictionary, a sub-dictionary of the Concept Dictionary, describes the similarity relation among concepts listed in the Word Dictionary. The EDR Corpus is the source for the information described in each of the sub-dictionaries. The basic approach taken during the development of the dictionary was to avoid a particular linguistic theory and to allow for adoptability to various applications.

After NICT obtained the EDR Electronic Dictionary, we continue to improve the lexicon. We added around 17,000 words to the Japanese Word Dictionary, 13,000 words to the English Word Dictionary and 13,000 concepts to the Concept Dictionary. We developed the Lexicon Browser for EDR Electronic Dictionary, which enables users to extract necessary information in different dictionaries via common procedure. We are also using the EDR Electronic Dictionary as a knowledge source for our NLP researches, such as sentence generation and machine translation, and for our basic research on language. NICT established its own Thai Computational Linguistics Laboratory (TCL) in 2002 in Thailand and started a project of TCL Lexicon. TCL Lexicon is currently in Thai and English, however, we will extend it to Asian Languages. We are planning to use EDR as a seed lexicon of this project.

### **30. Imitative Representations of Bird Sounds in Finnish and Estonian**

*Maria-Magdalena Jürvetson, Tallinn Pedagogical University, Estonia*

This abstract investigates some aspects of imitative representations of bird sounds in Finnish and Estonian, such as typology of sound symbolism, symbolic value of some vowels and the role of imitatives in derivative system of language.

By Hinton et al (1994) there are four different categories of sound symbolism: corporeal, imitative, synesthetic and conventional. Imitative representations of bird sounds relate to imitative sound symbolism. Imitatives include many utterances that utilize sound patterns outside of conventional speech and are difficult to portray in writing.

One of the universals of imitative sound symbolism is symbolic value of front vowel *i*, which traditionally is connected to the height, clearness, weakness, gentleness etc. of voice. Thus, song representation of *Corn Bunting* (fi. 'harmaasirkku', est. 'halltsiitsitaja') in Finnish is monotonous chirr *tsi tsi-tsi-ri-riririssss* (est. 'tsük-tsik-tsik-tzrl'l'l'zr'r'r'r'), song of tiny *Goldcrest* (fi. 'hippiäinen', est. 'pöialpoiss') - a very high verse *tiitiityri-tiitiityri-tititiu* (est. 'tui-tii-tidi tui-tii-tidi tui-tii-tidi tui-tii-tidi tui-tiidjirruh'). Song of *Greenish Warbler* (fi. 'idänuunilintu', est. 'rohe-lehelind') is soft, short and fast verse *ti tslia-tslia-sli-titslia-tslia-slia-slisli-slililili* (est. 'ti titslia tslia ti-titslia tididi-tslia-slililili tslia').

Front vowel *i* has double symbolic value in the names of tiny birds, such as *Goldcrest* (fi. 'hippiäinen', est. 'pöialpoiss'), *Wood Warbler* (fi. 'sirittäjä', est. 'mets-

lehelind') or *Bohemian Waxwing* ('tilhi', est. 'siidisaba'), because at the same time it points to the high, gentle voice and to the tiny size of a bird.

Back vowels *a*, *o*, *u* point to the low and loud voice. Thus, call of male *Common Eider* (fi. 'haahka', est. 'hahk') is loud, melancholy *aouu* (est. 'auu-u') and the answer call of female bird - a rough, abrupt cackle *kok-ok-ok* (est. 'kok-ok-ok'). Call of *Eurasian Eagle-Owl* (fi. 'huuhkaja', est. 'kassikakk') is a powerful booming *huuu-uu* (est. 'uu-uu'). Loud, impressive screech of *Common Crane* (fi. 'kurki', est. 'sookurg') *kruu-kro* (est. 'krruu-ii(u)') resembles trumpet.

Back vowels *a*, *o*, *u* have also double symbolic value in the names of bigger birds, such as *Steller's Eider* (fi. 'allihaahka', est. 'kirjuhahk'), *Eurasian Eagle-Owl* (fi. 'huuhkaja', est. 'kassikakk') or *Whooper Swan* ('laulujoutsen', est. 'laululuik'), because at the same time it points to low, loud voice and the big size of bird.

Imitatives are also part of derivative system of language. There are lots of bird names (e.g. *sirittäjä* < *sip sip sip-sip-sip*, etc.) and verbs, derivated from imitatives. Thus, sound verbs with vowel *i* in first syllable represent different sounds of tiny birds. *lirittää*, *lirkuttaa*, *livertää*, *piipittää*, *piiskuttaa*, *sirittää*, *sirkuttaa*, *tiitittää*, *tilkuttaa*, *tirskuttaa*, *visertää* etc.

There are also bird names (e.g. *alli* < *a-alli*, *a-alli*, etc.) and verbs with back vowels *a*, *o*, *u* or front vowel *ä* in first syllable, derivated from imitatives and describing sounds of bigger birds: *huhuilla*, *kaakattaa*, *kakattaa*, *kuhertaa*, *kukkua*, *kurluttaa*, *käkättää*, *räkättää*, *raakkua*, *ronkkua*, *vaakkua* etc.

Key words: onomatopoeia, onomatopoeic verbs, sound symbolism, imitatives.

## 31. Analysis of Semantic Usage of Adjectives via Hierarchical: Self-Organizing Semantic Map

*Kyoko Kanzaki, Eiko Yamamoto, Qing Ma\* and Hitoshi Isahara, National Institute of Information and Communications Technology, \*Ryukoku University, Japan*

When compiling lexical databases, it is important to consider what rules or phenomena should be described as lexical meanings and how these lexical meanings should be formalized and stored electronically. The method of organization of word meanings is also a crucial issue with lexical databases. In current lexical databases and/or thesauri, abstract nouns indicating concepts are identified manually and words are classified in a top-down manner based on human intuition. This is a good way to make a lexical database for users with a specific purpose. However, word hierarchies based on human intuition tend to vary greatly depending on the lexicographer, and there is often disagreement as to the make-up of the hierarchy. If we could find an objective method to organize word meanings based on real data, we would avoid this variability.

Our purpose in this research is to extract word hierarchies from corpora automatically. Our initial task to this end is to determine adjective hyperonyms. In order to find adjective hyperonyms, we utilize abstract nouns. We constructed linguistic data by extracting semantic relations between abstract nouns and adjectives from corpus data and classifying abstract nouns based on adjective similarity using a self-organizing semantic

map (SOM), which is a neural network model. The relative proximity of words in the semantic map indicates their relative similarity. And also, to find an objective hierarchical word structure, we utilize the complementary similarity measure (CSM), which estimates a one-to-many relation, such as superordinate-subordinate relations. We propose an automated method for constructing adjective hierarchies by connecting strongly related abstract nouns in a top-down fashion within a semantic map, mainly using CSM and obtained a semantic map which indicates semantic usage of adjectives from real corpus data.

In our experience, we generated semantic map of 365 abstract nouns based on 35,173 adjectives automatically using 40 years-worth newspaper articles, and verify classifications on the map. We could observe the differences of the adjectives on the map depending on the lower level of hierarchies. For example, "gankona(stubborn)", "yasashii(kind)" et.al are related to the hierarchy, "koto(matter)", "men1(side)", "seikaku(character)", "seishitu(property)", "kishitsu(personality)", "kifuu(disposition)", and "kishou(temper)". On the other hand, "kagakutekina (scientific)", "igakutekina(medical)", "gijutsutekina(technical)" et.al are related to the hierarchy, "koto(matter)", "men1(side)", "sokumen(one side)", "imiai(meaning)", "kanten(viewpoint)", "kenchi(standpoint)", "bun'ya(domain)", and "ryouiki(area)". However, these adjectives are co-occurring with the same abstract noun, "men1 (side)", that is, these adjectives represent the same abstract concept at the upper level. From hierarchies on the map the second level of abstract concepts of adjectives are "men1(side)" and "joutai(state)". This means that each adjective has the differences of representation at the second level from the top; the one side of something and/or the state at some point in time.

In this paper we propose how to construct the semantic map with hierarchical distribution and describe the observation of the result on the Japanese adjectives and their concepts.

## 32. Etymology and the structure of the lexicon

*Katalin Sipőcz, University of Szeged, Hungary*

This poster discusses etymological practices that are related to the structure of the lexicon. The data are taken from the etymological literature of Uralic languages.

The usual practice in etymology is researching the origin of individual words and statements about the positions of the examined words within the lexicon are about conceptual groupings. The handbooks or the historical lexicological studies of Uralic languages typically provide such descriptions of the reconstructed vocabularies of the different proto-language levels (cf. Häkkinen, Kaisa: Suomen kielen sanaston historiallista taustaa, *Fennistica* 7, Turku 1985: 69–113; Hajdú, Péter: Bevezetés az uráli nyelvtudományba, Budapest 1966: 86–87).

It is commonly accepted that the existence of certain words implies the existence of some other words and such observations have been present in the etymological practice (cf. Häkkinen, op cit 76). The revolutionary step in the systematic research of the vocabulary was made by the structuralists, whose theory of semantic fields is based on the recognition that the meaning of a word depends on the meanings of other words of the

vocabulary. This approach, similarly to the more recent theories of lexical semantics, is primarily used in synchronic studies but they might have their roles in historical studies, too (cf. Trier's well-known diachronic studies). The so-called componential analysis has also evolved in synchronic research but several authors have pointed out its applicability to etymology (cf. pl. Fox, Antony 1995: *Linguistic Reconstruction. An Introduction to Theory and Method*. Oxford. 1995: 115). In Uralic studies, Bakró-Nagy Marianne has applied componential analysis to etymology (cf. e.g. Bakró-Nagy, Marianne: *Rokon nyelvi megfelelések szemantikai vizsgálata*, *Nyelvtudományi Értekezések* 89: 57–63, 1976).

Using typological approaches to the lexicon and semantics, I have proposed to revise the reconstructed basic meanings of several color names in *Uralisches Etymologisches Wörterbuch* (Sipőcz, Katalin: *A vogul nyelv színnevei*. *Studia Uralo-Altaica*, Supplementum 3. Szeged, 1994: 111–115) and I have applied notions like regularities and hierarchies within lexical structure in typological-etymological studies of body-part names (cf. Sipőcz, Katalin: *A jelentés rekonstruálásáról*, *Budapesti Uráli Műhely* II, 2001: 138–144; *The semantic study of body-part terminology in Uralic languages*, *FUF* 55, 1999: 59–92).

Studies by Bakró-Nagy and Sipos surpass the tradition with conceptual groupings (cf. pl. Bakró-Nagy: *Die Begriffsgruppen des Wortschatzes im PU/PFU*. *UJb.* 1992: 13–40, Sipos, Mária: *Az obi-ugor kor lexikális innovációi*. PhD-thesis. Budapest 2004).

This poster is an attempt to give an outline of the possibilities to connect etymology, historical lexicology and theories of lexical semantics or issues of typological studies concerning the lexicon. The data are from Uralic languages but they might be relevant generally for etymology, too.