

## **Wir bedanken uns sehr herzlich bei**

der Stadt Bielefeld, besonders dem Oberbürgermeister Eberhard David für  
den Empfang  
der Universität Bielefeld  
dem Rektor der Universität Bielefeld, Prof. Dr. Dieter Timmermann  
der Dekanin Prof. Dr. Lore Benz und dem Prodekan Prof. Dr. Rüdiger  
Weingarten  
der Fakultät für Linguistik und Literaturwissenschaft der Universität  
Bielefeld  
der Westfälisch-Lippischen Universitätsgesellschaft  
dem Studentenwerk Bielefeld

Dr. Angelika Wöllstein-Leisten (Universität zu Köln) für guten Rat  
Prof. Dr. Klaus-Ove Kahrmann und Jan Blum (Fakultät für Linguistik und  
Literaturwissenschaft / Kunst und Musik) für den Entwurf des  
Tagungsplakats  
der Software AG für die Bereitstellung der XML Datenbank Tamino  
Carolin Kram für die Pflege der Webseite  
Imke Veen für die Erstellung des Tagungsbandes und der Broschüre  
Dem HRZ Bielefeld für die technische Unterstützung  
Ben Hell für die Koordination

Allen Hilfskräften, besonders den freiwilligen studentischen Helfern und  
Helferinnen

### Organisationskomitee:

Dafydd Gibbon, Ben Hell, Gerhard Jäger, Barbara Job, Wiltrud Mihatsch,  
Thorsten Trippel

in Zusammenarbeit mit dem Programmkomitee der DGfS, besonders Dr.  
Anke Lüdeling und Prof. Dr. Richard Wiese sowie Prof. Dr. Rosemary Tracy



## **Inhaltsverzeichnis**

Programm der 28. Jahrestagung der DGfS .....	2
Abstracts zu den Vorträgen .....	3
Plenarvorträge.....	3
AG 1 Die Spannung zwischen Sprachbeschreibung und Sprachtypologie/.....	5
The tension between language description and language typology .....	5
AG 2 Syntax und Morphologie mehrdimensional .....	28
AG 3 How to recognize a sentence when you see one: methodological and linguistic issues in the creation of sign language corpora.....	41
AG 4 Current Issues in Areal Typology .....	53
AG 5 Parenthese/Parenthetical constructions .....	72
AG 6 Spracharchive – Standards, Erzeugung & Zugang / Language Archives, Standards, Creation and Access.....	88
AG 7 Beschreibung und Modellierung grammatischer Variation Describing and modeling variation in grammar .....	105
AG 8 Subordination' vs. ‚Koordination' in Satz und Text aus sprachvergleichender Perspektive / 'Subordination' vs. 'coordination' in sentence and text from a cross-linguistic perspective.....	126
AG 9 Sprache und Recht .....	147
AG 10 Fortgeschrittene Lernervarietäten .....	164
AG 11 Korpusbasierte Behandlung nichtkompositioneller Phänomen.....	180
AG 12 Adaptive Ontologien auf syntaktischen Strukturen .....	193
AG 13 Lexikalisch-semantische Ressourcen zur Sprachdokumentation und maschinellen Sprachverarbeitung .....	204
AG 14 Hybridialekte .....	220
Studentische AG .....	227
Postersession der Computerlinguistik.....	236
Verzeichnis der Vortragenden .....	251

# SPRACHDOKUMENTATION UND SPRACHBESCHREIBUNG

## Programm der 28. Jahrestagung der DGfS

### Dienstag 21.02.2006

13.30-17.00	Vorstandssitzung DGfS	D3-121
17.00-19.00	Lehramtsinitiative	D2-136
19.00	warming up	Brauhaus Joh. Albrecht Hagenbruchstrasse 8 33602 Bielefeld

### Mittwoch 22.02.2006

09.00-09.45	Begrüßung	Auditorium Maximum
10.00-11.00	Plenarvortrag Nikolaus Himmelmann <i>Daten und Datenhuberei</i>	Auditorium Maximum
11.00-12.00	Plenarvortrag Steven Bird <i>Linguistic Data Management with the Natural Language Toolkit</i>	Auditorium Maximum
12.00-13.45	Mittagspause <i>Währenddessen:</i> 12.30-13.30 --- Jahreshauptversammlung der Gesellschaft für bedrohte Sprachen in D2-136	
13.45-18.15	Arbeitsgruppen	Pause: 15.45-16.15
18.15-19.15	DGfS-Sektion Computerlinguistik	Halle, BLC
19.15-20.00	Verlagspräsentation zur Sprachdokumentation	Hörsaal 1
20.00	Geselliger Abend	Mensa-Markt

### Donnerstag 23.02.2005

09.00-12.30	Arbeitsgruppen	Pause: 11.00-11.30
12.30-14.00	Mittagspause <i>Währenddessen:</i> 12.30-14.00 --- Poster-/Demo-Session der Computerlinguistik --- Pressekonferenz in D3-121	
14.00-18.00	DGfS-Mitgliederversammlung	Auditorium Maximum
19.00	Empfang des Oberbürgermeisters	Ravensberger Spinnerei

### Freitag 24.02.2006

08.00-09.00	Plenarvortrag: Stephen Levinson <i>Enrolling other sciences in language documentation: describing an isolate language, Yéli Dnye, in Papua New Guinea</i>	Auditorium Maximum
09.00-10.00	Plenarvortrag: Arienne M. Dwyer <i>Linguist-Driven Tools: A Manifesto</i>	Auditorium Maximum
10.30-17.30	Arbeitsgruppen	Pause: 12.00-13.00

## Abstracts zu den Vorträgen

### Plenarvorträge

---

Nikolaus P. Himmelmann  
Ruhr-Universität Bochum  
**Daten und Datenhuberei**  
Mi 10.00-11.00

---

In den letzten 10-15 Jahren hat in der Sprachwissenschaft weltweit eine Rückbesinnung auf die empirischen Grundlagen eingesetzt, wie es sie in weiten Teilen des 20. Jahrhunderts nicht gegeben hat. Das ist eine begrüßenswerte Entwicklung, zu der die von Deutschland ausgehende Dokumentationslinguistik einen nicht unerheblichen Beitrag geleistet hat. Allerdings gibt es derzeit, wie so oft wenn Altes mit großem Enthusiasmus neu entdeckt wird, allerlei überzogene Erwartungen, falsche Oppositionenbildungen und begriffliche Verwirrungen, von denen einige in diesem Vortrag aufgegriffen werden. Dabei wird u.a. dafür argumentiert, dass die folgenden drei Datentypen zu unterscheiden sind: Rohdaten, Primärdaten und Strukturdaten. Für die linguistische Theoriebildung sind letztlich nur Strukturdaten von Interesse. Soweit aber linguistische Theorien empirischen Gehalt haben sollen, ist eine ernsthafte und sorgfältige Auseinandersetzung mit den beiden anderen Datentypen unerlässlich.

---

Steven Bird  
Department of Computer Science, University of Melbourne  
Linguistic Data Consortium, University of Pennsylvania  
**Linguistic Data Management with the Natural Language Toolkit**  
Mi 11.00-12.00

---

Recent technological advances are having a profound impact on the accessibility, quality and quantity of linguistic corpora. However, linguists wanting to reorganise, restructure, and reuse this data face unparalleled frustration, thanks to the proliferation of incompatible special-purpose tools and formats. This talk will present NLTK, the Natural Language Toolkit, a suite of Python program modules, corpus samples, and documentation. I will show how NLTK can be used by linguists to write simple programs for manipulating and analyzing annotated linguistic data, in support of descriptive research in phonology, lexicography, and syntax. Finally, we will see how NLTK can be used to confront some unique challenges in fieldwork data management.

(NLTK is freely downloadable from <http://nltk.sourceforge.net/>)

---

Stephen C. Levinson  
Max Planck Institute for Psycholinguistics  
**Enrolling other sciences in language documentation: describing an  
isolate language, Yéfi Dnye, in Papua New Guinea**  
Fr 08.00-09.00

---

This paper argues that there's no such thing as vanilla description or a commonsense level of language documentation. Good description needs all the theory it can get. Whose theory? Language description and documentation tend to be viewed by linguists as serving linguistic theory (or even just typology!), but in fact language documentation is far too important for that alone – it serves a much wider set of disciplines, including anthropology, archaeology, biology, psychology, even new sciences like geographic information theory (GIS). All these disciplines have an interest in the data. In turn they offer theory, method, questions and insight which can greatly sharpen and broaden linguistic description and documentation. If linguistics is to have a flourishing future, it has to look outside the confines of its narrow humanities origins and contribute to wider debates.

I will try and exemplify this by focussing on a single language, Yéfi Dnye, an isolate spoken offshore from Papua New Guinea. I will show how questions and methods from other disciplines deepen our knowledge of that language.

---

Arienne M. Dwyer  
University of Kansas  
**Linguist-Driven Tools: A Manifesto**  
Fr 09.00-10.00

---

Most of us linguists - myself included - have barely begun to use the potential power of computers and similar devices to assist us with data discovery and analysis. The reason is simply that there are far too few digital tools designed for linguistics, and those that exist are often not useful for the specific tasks at hand, not quickly learnable and adaptable, and not suited to a wide range of theories or field situations. Theoretical or descriptive linguists and software developers (or even computational linguists) rarely talk to one another, which has sometimes resulted in the well-intentioned creation of products that are regrettably minimally useful to most linguists.

Which types of tools would, then, not only be useful to but also be used by linguists of all stripes? I discuss tools assessment criteria, areas which are currently not well-served by extant tools (as exemplified by A/V field data), and suggest steps to stimulate more cooperative tool development efforts.

## Programm der Arbeitsgruppen (AGs)

### AG 1 Die Spannung zwischen Sprachbeschreibung und Sprachtypologie/ The tension between language description and language typology

Koordination: Walter Bisang, Martin Haspelmath

#### Programm

Mittwoch 22.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
13.45-14.15	Dik Bakker, Anna Siewierska	<i>What typologists always wanted from grammars, but never dared to ask</i>
14.15-14.45	Emily M. Bender, Scott Drellishak, Dan Flickinger, Jeff Good	<i>Typological coverage and descriptive precision in grammar engineering</i>
14.45-15.15	Walter Bisang	<i>Innovation by description – what description can see that typology and theory can't</i>
15.15-15.45	Raymond Boyd	<i>“Adjectives”: one descriptive linguist's struggle with a “typological generalization”</i>
15.45-16.15	Kaffeepause	
16.15-16.45	Sonia Cristofaro	<i>The organization of reference grammars: A typologist user's point of view</i>
16.45-17.15	Anaïd Donabedian	<i>Evidentiality: a challenge for both descriptive and typological linguistics</i>
17.15-17.45	Matthias Fritz	<i>Typology and reconstruction</i>
17.45-18.15	Martin Haspelmath	<i>Preestablished categories don't exist – consequences for language description and typology</i>

Donnerstag 23.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
09.00-09.30	Katharina Haude	<i>Labelling grammatical relations: the case of Movima</i>
09.30-10.00	Olesya Khanina	<i>Lexicon-grammar interaction in grammatical descriptions</i>
10.00-10.30	Bernhard Wälchli	<i>Descriptive typology vs. the extractingdescriptions-of-particular-languages method: Continuous variables in the typology of motion events</i>
10.30-11.00	Masha Koptjevskaja-Tamm	<i>Lexical typology and language descriptions</i>
11.00-11.30	Kaffeepause	
11.30-12.00	Barbara Stiebels	<i>On the typology of SOA-argument-taking predicates</i>
12.00-12.30	Mehrdad Naghzguye Kohan, Marjan Fanaei	<i>Phonological typology of Iranian dialects as evidence for the development of the New Persian vowel system</i>

Freitag 24.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
10.30-11.00	Björn Wiemer, Paolo Acquaviva	<i>The interplay of lexical semantics and grammatical categories in typology and language description</i>
11.00-11.30	Annie Montaut	<i>Describing a modal system regarding (the lack of?) evidential category: The example of Indo-Aryan languages</i>
11.30-12.00	Claire Moyse-Faurie	<i>A typological survey of Oceanic intensifiers and reflexive markers</i>



12.00-12.30	Frederick Newmeyer	<i>Typological rarity and grammatical complexity: Is there a correlation?</i>
12.30-13.30	Mittagspause	
13.30-14.00	Christiane Pilot-Raichoor	<i>Typological approach to the description of sentence types in a language</i>
14.00-14.30	Ron Schaefer, Francis Egbokhare	<i>Lexical typology and the practice of descriptive linguistics</i>
14.30-15.00	Frank Seifart	<i>Nominal classification in Miraña (North West Amazon): A challenge to the typological distinction between classifiers and noun classes</i>
15.00-15.30	Stavros Skopeteas	<i>Comparative production data in grammatical description</i>
15.30-16.00	Kaffeepause	
16.00-16.30	Josef Szakos	<i>Enabling field-workers: How the expertise of typologists can help us to become real professionals of language documentation</i>
16.30-17.00	Alice Vittrant	<i>How typology allows new analysis of verbal phrase in Burmese</i>

---

Anna Siewierska, Dik Bakker

Universität Lancaster, Universität Amsterdam

**What typologists always wanted from grammars, but never dared to ask**

Mi 13.45-14.15

---

Typologists are typically on the asking side of linguistic research. Their sources are grammars and grammatical descriptions provided by field linguists and other language experts rather than their own first hand observations on the languages of the world. This state of affairs creates a number of problems, running from the interpretation of example sentences in grammars to the actual absence in descriptions of facts crucial for the establishment of concrete typologies and theory building based on them. Arguably, progress could be made in this area if typologists without field experience themselves would become more aware of the problems of elicitation and writing a grammar, preferably by trying their hands on it themselves. However, linguistics is a dynamic branch of science. It seems to be illusory that a grammar might anticipate all the questions that might be asked about a language. In our paper, we will illustrate this point with several examples from our own typological research on agreement, alignment systems and person marking. We will suggest that there may be a partial way out of this dilemma via a threeway approach. A first step might be the expansion and adaptation of checklists like the one used by the Descriptive Grammar Series (DGS). Especially the categories semantics and pragmatics should get attention. In general, such a list should contain all categories which may be considered as parametric on the basis of the current state of the art in linguistic typology. A second, more refined checklist should be based on a dynamic, bottom up ontology of linguistic terms like the one coming out of the Typological Database System project (TDS). Its dynamic character keeps it in tune with ongoing linguistic research. Languages are in the heads and the mouths of their speakers, not in grammars. So, thirdly, for each grammar, extensive spoken corpora should be collected and glossed according to a generally accepted glossing system, such as the Leipzig version of the EuroTyp list.

---

Emily M. Bender, Scott Drellishak, Dan Flickinger, Jeff Good

University of Washington, University of Washington, CSLI Stanford,

MIP EVA

**Typological coverage and descriptive precision in grammar engineering**

Mi 14.15-14.45

---

For some languages, e.g. English or Japanese, modern machine-readable grammars are able to assign analyses to such a wide range of data that they could be considered actual grammatical descriptions—albeit of a very non-traditional sort. Researchers are now attempting to move beyond creating computational systems for individual

languages and develop systems that can be applied to data from typologically-diverse sets of languages.

One such system is The Grammar Matrix (Bender et al. 2002), based on HPSG. It is intended to serve as a typologically-informed “starter kit” for the development of a machine-readable grammar, containing the foundations of such a grammar as well as a series of configurable *modules* covering specific grammatical domains (e.g., word order, negation, coordination, etc.). In the modules, typology informs description by providing basic “parameters” which are expected to be useful for creating specific grammars. Ultimately, it is expected that grammatical phenomena in particular languages not well accommodated by a given module will serve as the basis for refinements, allowing specific descriptions to influence the Matrix's system of typology.

This talk will focus on the tension between typology and description as it has manifested itself in the development of Grammar Matrix modules, particularly the module for sentential negation. Negation has served as an interesting testing ground because its morphosyntactic realization interacts with a number of areas of a given language's grammar, making it challenging to encode computational analyses for each variant which are consistent with the full range of variation along other typological dimensions. The main area of difficulty in designing the module has been “unpacking” these typological categories into subcomponents to be encoded in different parts of the machine-readable grammar, and we will discuss lessons that we believe this process of unpacking raises for both typology and descriptive linguistics.

---

Walter Bisang

Universität Mainz

**Innovation by description — what description can see that typology and theory can't**

Mi 14.45-15.15

---

Complaints by theory-oriented linguists about language descriptions that do not provide data on what is relevant seem to be quite notorious. From such a perspective, it looks as if the description of individual languages always lags behind theory. In the present paper, I would like to sketch the opposite view by showing how a detailed description of individual languages may reveal new theoretical options. For that purpose, I will address the following two cases against basic tenets in linguistics:

1. *Fixed correlation between lexical items and syntactic categories like N and V:* In his description on parts of speech in Tongan, Broschart (1997) shows that lexical items are not preclassified for syntactic categories. In another language like Late Archaic Chinese (5<sup>th</sup> – 3<sup>rd</sup> century BC), the likelihood of object-denoting lexemes to occur in the V-position can be described in terms of stereotypical implicatures along a language-specific version of the animacy hierarchy.
2. *The relevance of semantic roles for the linking of syntax and semantics:* As pointed out by Foley & Van Valin (1984), reference seems to be more important for the assignment of lexical items to syntax than semantic roles in Tagalog. In Liangshan Nuosu (Tibeto-Burman, Loloish; Gerner 2004), it

seems to be impossible to describe the syntactic category of pivot in terms of restricted neutralization of semantic roles (cf. Role & Reference Grammar).

The above descriptions suggest that it is necessary to look for typological variation where neither typology nor theoretical linguistics would expect it.

Broschart, Jürgen. 1997. "Why Tongan does it differently: Categorical distinctions in a language without nouns and verbs." In: *Linguistic Typology* 1, 123-165.

Foley, William & Van Valin, Robert D. 1984. *Functional syntax and universal grammar*. Cambridge: UP.

Gerner, Matthias. 2004. "On a partial, strictly word-order based definition of grammatical relation in Liangshan Nuosu." In: *Linguistics* 42. 109-154.

---

Raymond Boyd  
CNRS-LLACAN, Paris

**"Adjectives": one descriptive linguist's struggle with a "typological generalization"**

Mi 15.15-15.45

---

1. This paper considers whether Zande (Niger-Congo, Ubangian subbranch) satisfies the specific typological generalization that all languages have a lexical "adjective" word class (Dixon 2004).

2. Prima facie, Zande has "noun-like" adjectives "with restricted functional possibilities for adjectives" (Dixon 2004: 28) and both a) Modified Nominal + Modifying Nominal genitive structures and b) "relational nouns" which *must* be followed by a modifying noun.

3. In reality, any two-term Zande noun phrase is "unoriented", i.e., has no intrinsic head. The head must be determined from context and customary usage.

4. Zande thus has, not a class, but adjectival and genitival *syntactic positions*, which only become recognizable once the head of any given noun phrase has been established. *A priori*, any nominal could appear in either of these positions.

5. This difficulty could be rejected on the basis of "statistical tendencies" and/or "fuzzy areas between classes" (Dixon 2004:45). The objection that Zande has been described by only a single scholar and that the data might be read otherwise is also tenable.

6. There is, however, a second difficulty: many adjectival notions are expressed in Zande by verbs only or by both verbs and etymologically unrelated nominals. But such verbs have two senses according to context: stative ('be + Attribute' ) and processive ('become, grow + Attribute'); "nominal adjectives" do not have this possibility. Dixon (2004) does not envisage this situation.

7. The difficulties arising from these data raise the question of whether large-scale generalizations can be reached without prior understanding of low-level situations by region and/or genetic unit.

Dixon, R. M. W., 2004, Adjective classes in typological perspective, in Dixon, R. M. W. and Alexandra Y. Aikenvald, *Adjective classes: a cross-linguistic typology*, Oxford: Oxford University Press, 1-49.

---

Sonia Cristofaro  
Università Pavia

**The organization of reference grammars: A typologist  
user's point of view**

Mi 16.15-16.45

---

This paper addresses the issue of the organization of reference grammars from a typologist's perspective, based on specific examples collected while carrying out cross-linguistic studies of grammatical domains such as clause linkage and the expression of tense, aspect, and mood.

Attention is focused on the following issues:

- (i) The distribution of information across different parts of a grammar.
- (ii) The parameters taken into account when providing information about the categories described in the grammar.
- (iii) The criteria used in identifying the categories described in the grammar.

The paper outlines two basic problems with the organization of reference grammars. First, the organization of the grammar may lead the user to miss pieces of information that actually are there, but are located in parts of the grammar where they may be not easily recoverable. Second, the grammar may actually fail to provide complete and typologically adequate information on the phenomena being described. This may be for two reasons. On the one hand, the grammar may fail to take into account parameters that interact with the categories being described and may determine variations in the structural properties of these categories. On the other hand, the description may be organized so as to fail to reveal the differences and similarities between the relevant categories and comparable categories in other languages.

The paper puts forward a number of proposals that are meant to enhance the typological adequacy of grammars, and concern organization of the content of a grammar in terms of functionally based categories, compliance with the results of typological research in terms of parameters taken into account in the analysis, use of elicited data as a means to supplement data from naturally occurring discourse, and standardization of the descriptive framework and terminology adopted in individual grammars.

---

Anaïd Donabédian-Demopoulos  
INALCO Paris

**Evidentiality : a challenge for both descriptive and typological linguistics**  
Mi 16.45-17.15

---

Since Chafe & Nichols 1986, evidentiality is a very popular category in both descriptive and typological studies, inspiring many cross-linguistic studies (Guentcheva 1996, Johanson and Utas 2000, Aikhenvald 2003, 2005) involving Balkanian, Caucasian, Amerindian, or tibeto-birman areas, as well as in French or English where evidentiality is not grammatically encoded.

This popularity may be explained by the very nature of evidentiality in languages where it was first discovered : ‘linguistic coding of epistemology’ (Chafe and Nichols), ‘abduction’ (Guentcheva), ‘sources of knowledge’ (Dendale & Tasmowski), etc.. This category is very fruitful for cognitive studies, and it opens doors for both semasiological (what is evidentiality in evidential-marking languages) and onomasiological (how is evidentiality encoded cross-linguistically) studies.

However, purely descriptive study of evidentiality in a language where it is grammatically encoded may reveal that the function of this category throughout the TAM system of the language is much more complex, and that the cognitive explanation does not catch the function of the marker. Our aim is to bring some data from languages like Western Armenian, Turkish, Albanian, showing that there is an alternative way to describe this category : ‘hearsay’ appears to be just a contextual meaning of the marker (mainly in *récit*) since the core meaning (closer to the mirative, appearing in *discours* - DeLancey 1997) may be described as a specific type of statement. Such an analysis may first seem less attractive for typology, but in fact, it opens the way for cross-linguistic analyses involving different levels of analysis (Bonnot-Donabedian 1997-1999). Typology rarely takes into consideration such studies, maybe because it lacks models to combine so many parameters (different languages, different levels of encoding, different categories). Some recent proposals (cf. Plungian - Van der Auwera 1998) probably opens interesting perspectives.

---

Matthias Fritz  
FU Berlin

**Typology and reconstruction**  
Mi 17.15-17.45

---

Reconstruction is a part of the work being done by Linguists who deal with prehistoric stages of language. The linguistic discipline that did language reconstruction first was Indo-European Linguistics two centuries ago and it is Indo-European Linguistics, too, any kind of modern linguistics developed from. The basic

method of research to get an insight into the prehistoric evolution of two different but similar items is comparison.

In the 19<sup>th</sup> century there was almost no consciousness of typological aspects in historical linguistics. Indeed there was general linguistics just beginning to gather features of different languages, but those collections scarcely influenced historical linguists. Indo-European linguistics developed from Classics. Therefore its linguistic method is language description. If you compare two described languages, then the result of your reconstructing calculation won't be a described language, but a merely reconstructed one. Yet even when Indo-European linguistics didn't yet recognize typology, there were decisions to be made over which typologically different features to take for reconstructing the proto-language.

The question now is how typological parameters can increase the verisimilitude of the results of language reconstruction. As there are, however, languages that show at least one feature not matching the regular scheme of typology, you can always find real evidence against any feature that seems typologically universal.

In this lecture I will try to show how language reconstruction might take advantage from language typology regarding for instance the number of dual that can be reconstructed as a grammatical category of Proto-Indo-European. The difficult part of this is to answer the question of which parts of speech really had the dual in Proto-Indo-European, because some Indo-European languages have the dual with every inflecting part of speech, others have it only connected with the category of person and some else have it only with the noun. So which one is the original system, how can typology help this question and what's the contribution that reconstruction can make to typology?

---

Martin Haspelmath

MPI-EVA Leipzig

**Pre-established categories don't exist —  
consequences for language description and typology**

Mi 17:45-18:15

---

Structural categories of grammar (such as clitic, affix, compound, adjective, pronoun, dative, subject, passive, diphthong, coronal) have to be posited by linguists and by children during acquisition. This would be easier if they simply had to choose from a list of pre-established categories. However, existing proposals for what such a list might be are still heavily based on the Latin and English grammatical tradition.

Thus, descriptive linguists still have no choice but to adopt the Boasian approach of positing special language-particular categories for each language. Theorists often resist it, but the cross-linguistic evidence is not converging on a smallish set of possibly innate categories. On the contrary, almost every newly described language presents us with some "crazy" new category that hardly fits existing taxonomies. Although there is thus no good evidence for pre-established categories, linguists still often ask:

- (i) Is the Romanian definite article a clitic or a suffix? (Ortmann & Popescu 2000)
- (ii) Are Mandarin Chinese property words adjectives or verbs (McCawley 1992)
- (iii) Is the Tagalog *ang*-phrase a subject or a topic? (Schachter 1976)
- (iv) Is German *er* a pronoun or a determiner? (Vater 2000)

A consequence of the non-existence of pre-established categories for language description is that such questions are pointless. Instead of fitting observed phenomena into the mould of currently popular categories, the linguist's job is to describe the phenomena in as much detail as possible.

A consequence of the non-existence of pre-established categories for typology is that comparison cannot be category-based, but must be substance-based, because substance (unlike categories) is universal. This has been recognized in the Greenbergian approach, though it is often hidden by widely practiced terminology ("noun-genitive" order, "verb-object" order, etc.).

Katharina Haude  
RU Nijmegen

**Labeling grammatical relations: the case of Movima**  
Do 09.00-09.30

This paper deals with the problem of how an apparently unique alignment system can be described by using the canonical terminology. Movima (unclassified, Bolivia) displays a hierarchical alignment system. Unlike better-known systems of this type (e.g. Algonquian), it does not operate with prefix- and suffix positions, but with word order and phonological cliticization. The argument higher on the animacy hierarchy is obligatorily realized and encliticized to the predicate; the second argument is optional and morphologically independent. Markers on the predicate indicate which argument is A and which one is O. Thus, in the transitive direct construction in 0, the first argument NP represents A and the second NP represents O. In the inverse construction in 0, the first NP is O and the second is A.

*sal-ná=kus itila:kwa (kos pa:ko)*  
 search-DR=the man (the dog)  
 ‘The man looks for it (the dog).’  
*sal-kayá=kus itila:kwa (kos pa:ko)*  
 search-INV=the man (the dog)  
 ‘It (the dog) looks for the man.’

When compared with other constructions in Movima, the second argument of a transitive clause can be identified as the subject. The argument of intransitive clauses is also optionally realized and not cliticized; and only the second argument can be relativized.

However, labeling the second argument of a transitive clause “subject” can be misleading. It lacks some properties commonly associated with the subject status,



such as the encoding of the actor role, higher animacy or topicality of the referent, and obligatory realization.

The proposed solution is to describe the grammatical relations distinguished in Movima by using language-specific terminology, and to complement this by indicating how terms such as “subject” can be applied. In this way, the confusion with other conceptions of the term “subject” is avoided, while it is made clear at which points the Movima system is comparable with other alignment patterns.

---

Olesya Khanina

Moscow State University

**Lexicon-grammar interaction in grammatical descriptions**

Do 09.30-10.00

---

Theoretical questioning about language structures is hardly possible without systematic exploitation of descriptions of particular languages. However, almost every new insight is followed by a plenty of further questions that were previously never asked. At the same time, a lot of languages are endangered, and a researcher writing his/her grammar might be the last linguist to talk directly to the speakers. Questions (s)he has not asked might remain unanswered forever, so the problem of collection of data for questions that will arise only later is quite important. I am presenting a set of hints how the task can be done for one topic that is coming to the scene of linguistics nowadays: structural rules of lexicon and their interaction with those of grammar. This kind of information is worth being represented in descriptive grammars, and thus the manner of doing it without turning them into dictionaries must be worked out.

As a typologist, I have studied a linguistic problem situated on the borderline between grammar and lexicon: morphosyntax and semantics of desideratives. In order to collect the data I checked hundreds descriptive grammars and a few dictionaries: the procedure was far from being easy as only purely grammatical topics tend to be mentioned in the descriptions, and wanting does not belong to pure grammar in many languages. Being at the same time a descriptive linguist, I understand it is unrealistic to ask the grammar authors to give full information on wanting regardless its pertinence to the structure of the language under description. The situation with wanting is not unique though; one can imagine many other upcoming studies struggling with the lack of typological data, just because the topic is not purely grammatical. Using the example of desideratives cross-linguistically, I am suggesting some heuristics about how a grammatical description could regularly refer to the language objects that are neither purely grammatical, nor purely lexical.

---

Bernhard Wälchli  
MPIEVA Leipzig

**Descriptive typology vs. the extracting-descriptions-of-particular-  
languages method**

**Continuous variables in the typology of motion events**

Do 10.00-10.30

---

During the last decades, typology has focused especially on further developing its analytic tools while mostly outsourcing its descriptive component to grammar writing. Here we subsume under the notion of descriptive typology all typological data collection processes based on primary sources, i.e. written or spoken texts (questionnaires, parallel texts, original texts, retelling of picture stories or films). While the feature values obtained from the two methods—grammar (and dictionary) extracting and extracting information directly from texts—are not necessarily different in practice, the data obtained by the two methods are often of a fundamentally different kind. Unlike the extracting-descriptions-of-particular-languages method, description in descriptive typology is based on cross-linguistic comparison from its very beginning and makes statements about particular texts and not about languages directly. Its strong point is the investigation of continuous variables (frequency) and generally the study of language use. Because of their different advantages and disadvantages, the extracting grammar method and descriptive typology can be used to mutually evaluate their results (data evaluation generally being a strongly underdeveloped component of typology). This requires that research questions in both frameworks are designed in such a manner that their results are as comparable as possible. The paper discusses how this can be done and how analytic typology might profit from this mediation between two methods.

The paper investigates several continuous variables in the domain of motion events (source-goal indifference, displacement/path verbs, multi-verb constructions) with values obtained in a parallel text based descriptive typological investigation (in a world-wide sample of 150 texts/150 languages) and to what extent these features can be translated into traditional structural features (presence or absence of a value) such as they can be extracted from grammars, dictionaries or specialists' personal communication.

---

Maria Koptjevskaja-Tamm  
Universität Stockholm  
**Lexical typology and language descriptions**  
Do 10.30-11.00

---

Compared to the impressive progress within typological research in grammar and phonetics / phonology, typological research on lexicon has been relatively modest. Significantly, it is largely underrepresented in two of the most extensive typological enterprises - EUROTYPE and the WALS, the latter with only 6 of the 142 chapters on lexical semantics (COLOR, HAND/ARM, FINGER/TOE), which cover a large portion of our knowledge about large-scale genetic and areal distribution of lexico-semantic phenomena. Likewise the standard textbooks in typology do not say much on the lexicon. Conversely, Talmy (1985) is the only chapter on lexical semantics in Shopen's three volumes.

I will here concentrate on three possible foci in lexical typology: 1. Stratification of semantic domains; 2. Polysemy, semantic associations and semantic shifts: intra-field ('see' > 'hear' within PERCEPTION) vs. cross-field ('hear' > 'remember', from PERCEPTION to COGNITION), including grammaticalisation; 3. Interaction between vocabulary and grammar (e.g., special behaviour of BODY-PART terms). A recurrent problem in research on these issues is that the relevant information is either totally missing from available descriptions or is difficult to extract from them. Thus, words belonging to the same semantic domain are normally scattered across an ordinary dictionary, and the translations do not often help to figure out their mutual relations. A thesaurus is a help, but it misses largely on semantic associations and semantic shifts of the words, and on their morphosyntactic properties. Grammars often omit details on idiosyncratic behaviour of particular groups of words or, conversely, are not specific about which groups of words show a particular type of behaviour.

The problems with availability of data that can be used in lexical typology from language descriptions do not only inhibit this kind of research, but also partly stem from its relative limited insights. This vicious circle should be broken.

---

Barbara Stiebels  
ZAS Berlin

**On the typology of SOA-argument-taking predicates**

Do 11.30-12.00

---

Many grammars and linguistic descriptions treat the role of predicates that select ‘state of affairs’ arguments (SOAA) more or less superficially and unsystematically. Consequently, the empirical base for the research on the role of SOAA-taking predicates for clausal embedding is quite limited – in theoretical as well as typological linguistics. Control and raising predicates, for instance, have been mainly studied from an Indo-European perspective; therefore, the role of infinitival complementation with these predicates has been highlighted, whereas other patterns of subordination have received much less attention – with consequences for the theoretical modeling of complement control. The typological research (e.g. Givón 1980, Cristofaro 2003, van Valin & LaPolla 1997) has mainly tried to establish correlations of predicate classes and their degree of fusion/bonding with the dependent clause; the specific structural and semantic properties of the various items of the predicate classes, which are relevant for the language-specific analysis, are often left aside.

In my talk I will focus on predicates that select structures of complement control. I will discuss a) the syntactic and semantic parameters of clausal embedding that have to be dealt with in order to account for the specific role of the SOAA-taking predicates and b) the interplay of lexicon and grammar from a typological perspective (e.g. as inherent vs. structural control). I will show that detailed language-specific analyses of the competition of subordination structures with a given predicate are more relevant for the typology of SOAA-taking predicates than aspects of fusion/bonding.

Cristofaro, S. 2003. *Subordination*. Oxford University Press.

Givón, T. 1980. The binding hierarchy and the typology of complements. *Studies in Language* 4.3, 333-377.

Van Valin, R. & R. J. LaPolla. 1997. *Syntax. Structure, meaning and function*. Cambridge University Press.

---

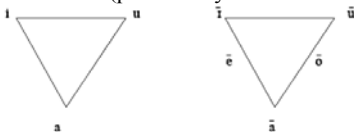
Mehrdad Naghzguy Kohan, Marjan Fanaei  
Bu Ali Sina U, Hamadan; Islamic Azad U, Shahreza<sup>1</sup>  
**Phonological typology of Iranian dialects as evidence for the  
development of the New Persian vowel system**

Do 12.00-12.30

---

New Persian (NP) begins with the invasion of Arabs and introduction of Islam to Persia. Although from this period onward the language is traditionally called New Persian, we should not assume that Persian has remained the same during this long span of time. In fact, one can observe many changes in morphology, phonology, syntax and lexicon of the language. Therefore, in some studies a distinction is made between Classical Persian (CP) of the post-Islamic period (up to 13th century), and Modern Persian (MP) of today. We adopt the same view in this study.

Mostly based on the written record, and in spite of uncertainties, researchers distinguish two different vowel systems for CP and MP. As for CP, it is argued that during the period the *quantity* of vowels was a distinctive feature; therefore it is distinguished from MP in which *quality* of vowels is distinctive. The vowel system reconstructed for CP looks like this (phonetic symbols used here are traditional ones):



This vowel system is neither symmetric, nor a stable one. There is no short counterpart vowel for /ei/ (*jaay-e majhul*) or /oi/ (*vaav-e majhul*). In contrast, the vowel system of MP contains six vowels: /i e a u o â/. Here one can quickly see the introduction of /e/ and /o/, and disappearance of /ei/ and /oi/ in the system.

In this article we have tried to see whether the related existing dialects could give more support to the proposed vowel system for CP. To fulfil this end, we have examined the vowel systems of 34 Iranian dialects within the geographical area of the present Iran. Here are the results:

1. Only Kurdish and Baluchi, are exactly like the proposed system for CP. This is statistically low (it means %5.88 of the total), but it is of great significance; since these are among the most conservative Iranian dialects.
2. Only %8.82 of the dialects, like MP, do not distinguish between long and short vowels, but the remaining portion of dialects do.

Like CP, %96.77 of the systems that distinguish between long and short vowels are asymmetric.

---

<sup>1</sup> This study was done by the financial support of Hamadan & Shahreza Islamic Azad Universities.

---

Paolo Acquaviva, Björn Wiemer  
U. Dublin, U. Konstanz

**The interplay of lexical semantics and grammatical categories in  
typology and language description**

Fr 10.30-11.00

---

Our contribution centres on the relation between grammar and the lexicon. Typology can provide generalizations about how grammatical categories are affected by lexical semantics, based on invariant semantic principles and on the encoding of grammatical properties in single languages. In language-specific descriptions, these generalizations act as hypotheses to falsify and guide the categorization of data. We take as illustration two related grammatical categories: number on nouns and iterativity / frequentativity on verbs. These often affect the interpretation of a lexical base, and can display a high degree of fusion between lexical and grammatical meaning. Typological research must account for this in a framework allowing for systematic cross-linguistic comparison, distinguishing instantiations of general typological patterns from truly idiosyncratic phenomena.

Many lexicalised instances of number are rooted in the conceptualisation of part structure: interpretive contrasts like mass - sort / portion, abstract - concrete, count - mass, all go back to basic mereological concepts (part and whole, kind and instances, boundedness, identity). Similarly, the particular meanings of iterative affixes can be reduced to part-whole relationships of (sub)intervals to the state of affairs denoted by the verb stem. Often the products of *Aktionsart*-changing derivation are lexicalised, as in the German *lachen* 'to laugh'  $\Rightarrow$  *lächeln* 'to smile'.

Language descriptions should clarify what lexicalised readings appear crosslinguistically with these number and aspect affixes, what types of stems are involved, and to what extent this follows typological tendencies. Typological analysis can clarify to what extent the patterns of lexicalisation attested in different languages are related, revealing for example implicational tendencies. This requires, however, analysing distinct instances of lexicalisation of grammatical morphemes from a unified perspective, which is what we argue for.

---

Annie Montaut  
INALCO Paris

**Describing a modal system regarding (the lack of?) evidential category:  
The example of Indo-Aryan languages**

Fr 11.00-11.30

---

The category of evidentiality has in the last twenty years emerged as a significant typological and areal feature. Since many languages display various meanings under the same marker (hear-say, inference, doubt, surprise, polemic contrast), attempts have been made to unite them into a consistent semantic area, mapped as an epistemic modality. At the same time, since markers often derive from a perfect, it has been suggested that inference, and indirect access to the source (hence hear-say), is the core value, mirativity being some kind of marginal, derived side-effect. However, the immediacy of new first-hand information is crucial in the balkanic markers, hence the claim for a specific “mirative” category distinct from evidentiality.

A description looking for these categories will generally ignore it if there is no specific verbal marker conveying the notion ‘distance from the information source’ or ‘surprise’, or both. Such is the case of Indo-Aryan languages, with the exception of Nepali. The reason of this exception is probably that the language has a special form of the perfect called “inferential”, which acquired a number of evidential/mirative meanings, whereas Hindi/Urdu, although attaching most of the “mirative” meanings to the aorist, never get it recognized in descriptions. Unbiased descriptions are required in order to test the hypothesis of a common operation (rather than a common semantic core) underlying these similarities. ‘Unbiased’ : critical regarding the traditional or modern labels used, which often convey implicit categories not always fit for the language. For example, Nepali verbal system of is described as having an aorist (and two perfects). Hindi has the same form (morphologically and etymologically: a resulting past mapped into a possessive syntactic pattern) labelled ‘perfective’, or ‘simple past’, a possible reason why its non-past non-perfective meanings are almost never mentioned. I will present Hindi/Urdu mirative and polemic meanings, alongside with other evidential markers (putative), compared with Nepali, and explore their relation to the aoristic scheme in terms of abstract operations rather than concrete meanings.

---

Claire Moyse-Faurie  
CNRS-Lacito, Paris

**A typological survey of Oceanic intensifiers and reflexive markers**  
Fr 11.30-12.00

---

It is generally asserted that Oceanic languages “have morphological markers used to encode reciprocal and certain other situations, but not reflexive situations” (Lichtenberk, 1999:31). Other authors do mention reflexive markers but no attempt has been made yet to present a typological survey on this topic.

In most of the Oceanic languages, there are indeed overt means to disambiguate between a normally other-directed action, and the same action done to oneself. Even if the markers involved are not obligatory in non-ambiguous contexts, their quasi systematic occurrence with third person arguments is an indication of their on-going grammaticalization.

I have begun a systematic survey of the different types, origins and functions of intensifiers and reflexive markers found in Oceanic languages and will present the result of my own fieldwork (in New Caledonia and Polynesia) along with what may be found in the available grammars in the light of the typological studies. Besides the confirmation of the importance of body parts in the development of intensifiers and reflexive markers, I found several other sources: they mainly come from verbs (‘true’; ‘return’; ‘be like’), adverbs (‘alone’, ‘again’, ‘backwards’), noun (‘responsibility’), or modal particles. Each of these markers exhibits different degrees of grammaticalization. Links with reciprocals will also be taken into account, as intensifiers or reflexive markers may extend their use to the reciprocal domain. Finally, since some of these markers have spread across languages due to contact, their distribution and development will also be considered from an areal perspective.

LICHTENBERK, Frank, 2000. Reciprocals without reflexives, in Z. Frajzyngier and T.S. Curl (eds), *Reciprocals. Forms and Functions*. Amsterdam/Philadelphia: J.Benjamins, Typological Studies in Language 41, p. 31-62.

---

Frederick J. Newmeyer  
U Washington

**Typological rarity and grammatical complexity: is there a correlation?**  
Fr 12.00-12.30

---

The purpose of this paper is to investigate the common claim that crosslinguistically rare features require more complex descriptive devices or supplementary statements not needed for common ones. I argue that such is sometimes true, but it is not generally the case, and I go on to discuss some consequences of this result.

There are in fact a number of syntactic phenomena that support a correlation between rarity and complexity, for example the rare V2 order, which is usually assumed to



require a special statement in the grammar. However, more often a complexity-rarity correlation is *not* supported. I focus on clausal negation and demonstrate that symmetric negation is descriptively the simplest. But since only a minority of languages employ that strategy, the idea of a complexity-rarity correlation is not fulfilled.

An immediate consequence of the failure of the correlation to hold means that theoretical-descriptive linguists and typologists have less to benefit from each other's results than might be expected (or hoped). But there is no reason to despair that the findings of typologists are in general irrelevant to other subfields of linguistics. I point to results that suggest that a significant percentage of typological generalizations have explanations based in language processing (Hawkins 2004; Newmeyer 2005). If such is correct, then typologists should look more to psycholinguists — and vice-versa — than has generally been the case.

---

Christiane Pilot-Raichoor  
CNRS-Lacito, Paris

**Typological approach to the description of sentence types in a language**

Fr 13.30-14.00

---

This paper aims to show that the methodology and concepts developed for cross-linguistic typology can be fruitfully adapted to the description of intra-language variation. We illustrate it in the domain of sentence structures in Badaga, a Dravidian language. Traditionally it is said, as Steever (1998) put it for Dravidian in general, that a 'simple sentence consists of a subject and a predicate' and that 'by combining two kinds of subjects [in the nominative case or in the dative case] and two kinds of predicate [a finite verb or a nominal predicate], 'four basic sentence types are generated'. However, in some cases, the subject properties are distributed onto two distinct NPs, questioning the relevance of notion of subject. Similar –and other– difficulties prompted some typologists such as Bhat (1991) or Kibrik (1997) to claim that subject-predicate is not the only way sentences can be structured. Several other modes of structuration, along different 'dimensions' (Kibrik): semantic roles, discourse flow, deixis..., are also common in languages. In the line of the 'differentiating typology' (Bhat), we claim that not four, but three 'maximally different' types of sentence -with coherent sets of features in terms of coding, semantic domain, etc.- can be identified in Badaga. Interaction and 'competition' (Stassen 1997) between these basic types provide the necessary tools to explain more complex or non-canonical structures. In return, this analysis of the Badaga types question Kibrik's generalisation assuming a dichotomy between 'subject-oriented' and 'subjectless' languages as both types of structuring are present in this language.

Bhat, D.N.S., 1991. *Grammatical Relations. The evidence against their necessity and universality*, London-New York, Routledge.

Kibrik, A.E., 1997. "Beyond subject and object : toward a comprehensive relational typology". *Linguistic Typology* 1-3, p. 279-346.

Stassen, L., 1997, *Intransitive Predication*, Oxford, Clarendon Press.

---

Ron Schaefer, Francis Egbokhare  
Southern Illinois U, University of Ibadan  
**Lexical typology and the practice of descriptive linguistics**  
Fr 14.00-14.30

---

This paper examines the relation between lexical typology and linguistic description as found in dictionaries. Benue-Congo languages are frequently cited as syntactically deficient (Welmers 1973, Foley and Olson 1985, Lord 1993, Crowley 2002). Their lack of prepositions and inflectional morphology, while typologically accurate, provides little insight into what grammatical resources are active or how they pattern. In the lexical typology of Talmy (1985) championed by Nichols (1992), precedence is a relation holding between verb complement elements moving object (*hay*) and goal (*wagon*) in the basic precedence construction *John loaded hay onto the wagon*. English allows basic precedence as well as reversed precedence (*John loaded the wagon with hay*). The nature and scope of precedence can become evident when lexical entries pay rigorous and systematic attention to verb complement alternations. Emai (Edoid, Benue-Congo) favors basic precedence for all verbs and disallows reverse precedence: *òjè óón àmè ó vbi ògú* [Oje pour water CL LOC bottle] ‘Oje poured water into the bottle’ but not \**òjè óón ògú àmè* [Oje fill bottle water]. Emai has no motion verbs akin to ‘fill.’ With noun entries, precedence becomes evident in the cause domain. A minimal entry for an abstract noun like *ùìn* ‘fever’ might simply list a definition plus phrasal illustration *ùìn éliyú* ‘that kind of fever.’ A more robust lexical entry incorporating clausal expression reveals strict precedence: causing condition must precede causee in *ùìn gbé òjè* [fever kill Oje] ‘A fever killed Oje,’ with no allowance for the order of ‘Oje died of fever.’ It is thus the absence of verb complement type among verb entries or the nature of clause level illustration in abstract noun entries that registers a significant typological generalization not only within a domain but across domains.

---

Frank Seifart  
Ruhr-Universität Bochum  
**Nominal classification in Miraña (North West Amazon):  
A challenge to the typological distinction between classifiers and noun  
classes**  
Fr 14.30-15.00

---

The system of nominal classification in Miraña (Witotoan, North-West Amazon) consists of an internally heterogeneous set of classifying morphemes (about 70) that occur in a variety of morphosyntactic contexts where these forms fulfill a variety of functions, e.g. derivation and unitization in nouns, formation of demonstratives and numerals, and cross-reference on verbs. This system shares characteristics of almost

every type of nominal classification system recognized in current typological approaches (e.g. Aikhenvald 2000; Grinevald 2000). It shares with “noun classes” (as one of these types) the overt marking in nouns and the agreement marking in expressions such as demonstratives, verbs, etc. With “classifiers” (as another type) it shares the large number of classifying morphemes and their specific semantic content. With individual subtypes of classifiers, it shares above all the morphosyntactic contexts, among them nouns (“noun classifiers”), numerals (“numeral classifiers”), and verbs (“verbal classifiers”). With numeral classifiers as one subtype it also shares the semantic profile (shape distinctions) and the unitizing function. In addition, there is a subset of class markers in Miraña—the general class markers—that resemble typical gender systems with respect to the semantic distinctions (natural gender) and the marking of number in addition to noun class. The correlations of characteristics for individual types predicted by current typological approaches (e.g. agreement and a small number of classes for the type “noun classes”) are thus not met by the Miraña data. These challenging characteristics are also displayed by a number of neighboring languages.

Aikhenvald, Alexandra Y. 2000. *Classifiers. A typology of noun categorization devices*. Oxford: Oxford University Press.

Grinevald, Colette. 2000. A morphosyntactic typology of classifiers. In *Systems of nominal classification*, ed. by G. Senft. Cambridge: Cambridge University Press. 50-92.

---

Stavros Skopeteas  
U Potsdam

### **Comparative production data in grammatical description**

Fr 15.00-15.30

---

This presentation outlines the importance of comparative production data in grammatical descriptions. I illustrate the methodological approach by means of an experiment on “given patients & new agents”, a discourse condition that induces marked constructions in different languages. In the presented experiment, patient givenness is combined with two further discourse variables, namely identifiability of the agent and animacy of the patient. A grammatical description can learn from this experiment the following:

- *Which construction is induced by a given discourse condition in the object language?* E.g., the condition “given patient and new agent” induces passive in German and Yucatec Maya, patient-agent-verb order in Georgian and Prinmi, and (predominantly) a marked prosodic structure of the basic word order in Greek.
- *Which discourse conditions are appropriate for the use of a given construction of the object language?* E.g., passive sentences with a preverbal patient subject are strongly preferred in Yucatec Maya in the experimental condition “animate patient”; when the patient is inanimate, then it occurs in the postverbal argument position.

- *Which preferences determine language production under particular discourse conditions?* Experimental data allow for quantitative generalizations which give an idea about the preferences in the language-specific discourse. Passive voice in Mayan occurs in 72% of the collected data in the condition “given patient and new agent” and in 100% in the condition “given patient and non-identifiable agent”; in German we gained 40% in the first condition and 63% in the second.

This data illustrates how the use of standardized experimental methods in grammatical description will not only support the individual descriptions with exact observations on controlled conditions, but will also allow for the typological characterization of the object language on the basis of a directly comparable type of linguistic evidence.

---

Josef Szakos

National Dong Hua University, Taiwan

**Enabling field-workers: How the expertise of typologists can help us to become real professionals of language documentation**

Fr 16.00-16.30

---

Based on decades of data collection in Formosan languages, I outline some points where clear typological statements and support is needed to save disappearing materials.

1. Typological interviews paired with recording discourse: Here I wish to offer ideas, how earlier questionnaires (Lingua, Eurotyp, other ethnological ones) could be updated to enhance to efficiency of field collection. Multimedia forms of questionnaires should be tested and the power of internet communication exploited.
2. Typologists should consider up-dating successful handbooks (like Shopen), but at a level, where beginning researchers and linguists with non-European background (Chinese e.g.) would be able to use them.
3. The meta-language of ontologies (like in GOLD) should be explained in a culture specific way, so that generalizations, as well as specific items, distinctions could be marked.
4. Typologists could be more involved in training people who spend time overwhelmingly in gathering data and they could raise consciousness for the applicability of typological observations (feedback to the field).
5. There should be a discussion on how to raise the prestige of the devoted field-workers in the general field of Linguistics, so that ongoing funding is ensured. This is where typologists and field-workers would learn to mutually respect and accept one another.
6. A DVD library of the World's languages should be prepared where we would have about 6 hours (or more) of speech, annotated, translated, tagged, in unified XML format and each typologist would be cooperating with a field-worker, where the further practical problems would be a good lesson for all participants.

This presentation is aimed at viewing the documentation and typological research from the alternative angle of participants in the field, so that we can achieve more in the short time left for saving languages.

Abraham, W. 1995. *Deutsche Syntax im Sprachenvergleich*. Tübingen: Narr

Abraham, W. 2001. „Modals: toward explaining the epistemic ‚non-finiteness-gap‘.“ In: R. Müller & M. Reis (hg.) *Modalität und Modalverben im Deutschen*. Hamburg: Buske, 7-36.

Erb, Chr. 2001. *Auxiliaries in German*. Dissertation Tilburg.

---

Alice Vittrant  
CNRS Lacito, Paris

**How typology allows new analysis of verbal phrase in Burmese**  
Fr 16.30-17.00

---

Burmese has been studied by westerner scholars since 18<sup>e</sup> century. Recent works containing good description of the language (or part of the language) are Allott (1965), Okell (1969), Bernot (1980), Wheatley (1982), Bernot et al (2001), Allott & Okell (2001). However comparison of the description made by these authors leads to different analysis of the verbal phrase. For instance the list of optional verbal morphemes (particles and auxiliaries) varies depending of the criteria used by the author for his analysis.

Moreover, as in many Asian languages, groups of verbs not separated by connectors exist in Burmese. These verbal phrases composed by several verb roots have identical form in surface. However, they cannot be analyzed in the same way.

They can correspond to lexical expressions, or compounds and be listed in dictionaries. The verbal phrase may be analyzed as a prototypical serial verbs construction (SVC), here is to say, part of one clause only which refers to a unique event. Third possibility : some of the verb roots can be grammaticalized and be considered as auxiliaries.

Given the identical surface forms and the different under-laying structures, we look for a model allowing us to treat together phenomenon traditionally analyzed separately. Therefore, the notion of serial verbs construction (CVS) recently developed by Déchaine (1993), Durie (1997) or, Aikhenvald & Dixon (in press) among others help us to build the adequate frame for a more global approach of the VP in Burmese.

## AG 2 Syntax und Morphologie mehrdimensional

Koordination: Monika Budde, Andreas Nolda, Oliver Teuber

### Programm

Mittwoch 22.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
13.45-14.15	Monika Budde, Andreas Nolda, Oliver Teuber	<i>Multiple dimensions in syntax and morphology</i>
14.15-14.45	Patrizia Noel	<i>Wackernagel's law: on the interaction of linguistic subsystems</i>
14.45-15.45	Aslı Göksel	<i>A syntactic template for Turkish</i>
15.45-16.15	Kaffeepause	
16.15-17.15	Bernd Wiese	<i>Zwischen Paradigma und Morphem: Zur polnischen Substantivflexion</i>
17.15-17.45	Nanna Fuhrhop	<i>Das Zusammenspiel von Morphologie und Syntax in der Schreibung des Deutschen</i>
17.45-18.15	Katrin Mutz	<i>Französisch "se": Zur Rolle der Argumentstruktur</i>

Donnerstag 23.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
09.00-10.00	Sebastian Drude	<i>Nominale Prädikation oder Modus? "-(t)u"-Formen im Awetí</i>
10.00-10.30	Marie-Hélène Viguiier	<i>Kategorielle Polysemie und semantische Kohärenz: Morphologische, syntaktische und lexikalische Faktoren bei der Bestimmung einer Gesamtbedeutung für das französische Imparfait</i>

10.30-11.00	Harald Stamm	<i>Kompositionalität in einem multidimensionalen Ansatz: Der Aufbau von Nominalgruppenbedeutungen in der Integrativen Satzsemantik</i>
11.00-11.30	Kaffeepause	
11.30-12.30	Hans-Heinrich Lieb	<i>Wortbildung auf morphologischer und Wortbildung auf syntaktischer Grundlage</i>

Freitag 24.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
10.30-11.30	Louisa Sadler	<i>Single conjunct agreement and multidimensionality</i>
11.30-12.30	Berthold Crysmann	<i>A coanalysis approach to Polish past tense agreement</i>
12.30-13.00	Andreas Nolda	<i>Nounless nominals in German: ellipsis, use as noun, nominalization</i>

---

Monika Budde, Andreas Nolda, Oliver Teuber  
Technische Universität Berlin, Humboldt-Universität zu Berlin, Universität  
Osnabrück

**Multiple Dimensions in Syntax and Morphology**

Mi 13.45-14.15

---

The properties of morphosyntactic objects can be accounted for in two different ways: *mono-dimensionally* or *multi-dimensionally*. Mono-dimensional approaches represent all of those properties jointly in a single structure without proper parts corresponding to the different properties. A paradigm case for this kind of approach is the Minimalist Program: the constituent structure encodes not only the hierarchical structure of the sentence, but also syntactic relations, word order, inflectional properties, and even—according to Rizzi (1997) and others—aspects of information structure. A multi-dimensional approach, on the other hand, represents some or all of those properties in different, though interrelated, substructures. Lexical Functional Grammar, for instance, describes sentences jointly by means of a c[onstituent]-structure, a f[unctional]-structure, an a[r]gument]-structure, a p[rosodic]-structure and a semantic or  $\sigma$ -structure (Bresnan 2001). And in Integrational Linguistics, words and sentences are assigned a constituent structure, a marking structure, a lexical interpretation, an intonational structure, and a relational structure (Lieb 1983).

In this introductory talk, we shall justify this classification of approaches into mono-dimensional ones and multi-dimensional ones. For the latter, a set of questions raised by their multi-dimensionality will be specified.

Bresnan, J. 2000. *Lexical Functional Syntax*. Oxford: Blackwell.

Lieb, H.-H. 1983. *Integrational Linguistics*. Bd. 1: *General Outline*. Amsterdam: Benjamins.

Rizzi, L. 1997. „The fine structure of the left periphery.“ In: L. Haegeman (Hg.) *Elements of Grammar*. Dordrecht: Kluwer, 281–337.

---

Patrizia Noel Aziz Hanna

Ludwig-Maximilians-Universität München

**Wackernagel’s Law: On the Interaction of Linguistic Subsystems**

Mi 14.15-14.45

---

According to Wackernagel (Wackernagel’s Law 1892), “light” elements occur in second position (second constituent or second word). Wackernagel’s Law presumably refers to a very old PIE syntactic structure. The Wackernagel complex, which consists of conjunctions, certain adverbs, pronouns and place words, is not a stable syntactic configuration. It shows both restructuring and movement from second position (cf. Ivanov 1999, Schrijver 1997). The syntactic configuration might even be called a “syndrome”, since the interaction of linguistic subsystems in this structure (syntax, phonology, and information structure) leads to characteristic results in individual languages.





The central role of focus and stress becomes apparent in sentences with non-case marked arguments. The form of such sentences (cf. (2a) – Erguvanlı 1984) is taken as one of the main pieces of evidence for SOV as the underlying structure of Turkish:

(2)a. Sağlık mutluluk getirir. b. SađLIK getirir mutluluk. c. Sađlık getiRİR mutluluk. d. GetiRİR sađlık mutluluk.

health	happiness	brings	O	V	S	S/?O	V	O/?S	V
S/O	O/S								

‘Health brings happiness.’ ‘Happiness brings health.’ ‘Health brings happiness.’ ‘Health brings happiness.’  
 ?‘Happiness brings health.’ ‘Happiness brings health.’

However, (2a) is nothing more than the natural outcome of the structured nature of the preverbal position as is claimed here. (2b) and (2c) show that identifying an NP as having a particular grammatical function has to do not only with its position but also with the position of focal stress. As for (2d), when these two arguments cooccur in the PVP where they cannot be stressed, the result is ambiguous. Under our analysis, these are expected results of incorporating focus and stress into a description of syntax in free word order languages where syntax itself is hierarchically and linearly structured.

Erguvanlı, E. E. 1984. *The Function of Word Order in Turkish Grammar*. Berkeley: UC Publications.  
 Göksel, A. & A. S. Özsoy 2000. „Is there a focus position in Turkish.“ In: Göksel, A. & C. Kerslake (eds.) *Studies on Turkish and Turkic Languages*. Wiesbaden: Harrassowitz, 219-228.  
 Göksel, A. 2005. Flatness and structure in the PVP. 38th Meeting of the SLE, Valencia.  
 Kornfilt, J. (undated) Directionality of identical verb deletion in Turkish. Ms.  
 Kural, M. 1997. „Postverbal constituents in Turkish and the linear correspondence axiom.“ *LI* 28, 498-521.  
 Rooth, M. 1992. „A theory of focus interpretation.“ *Natural Language Semantics*, 75-117.  
 Takano, Y. 2005. LSA Lectures on Japanese Syntax, MIT.

Bernd Wiese  
 Institut für Deutsche Sprache, Mannheim  
**Zwischen Paradigma und Morphem: Zur polnischen Substantivflexion**  
 Mi 16.15-17.15

Der Beitrag greift die Frage auf, wie sich funktionsbezogene (syntaktisch basierte) Unterscheidungen von Kasusformen („distributional cases“ im Sinne von Comrie) zu formbezogenen (morphologisch basierten) Unterscheidungen verhalten (Comries „formal cases“). Am Beispiel des Polnischen – einer Sprache, deren Kasusendungen teils Charakteristika des agglutinierenden Typs, teils des fusionierenden Typs zeigen, und die damit eine besondere Herausforderung darstellt – wird gezeigt, wie die beiden Aspekte von Kassystemen so aufeinander bezogen werden können, daß die in weiten Bereichen prima facie verwirrend uneindeutigen Form-Funktions-Zusammenhänge transparent gemacht werden können. Hilfreich ist die Annahme

eines hierarchisch gestuften Kasussystems, dessen Endpunkte die bekannten, syntaktisch bestimmten Kasus (wie Lokativ, Dativ usw.) bilden, das aber im Unterschied zu traditionellen ‚flachen‘ Kasusklassifikationen weitere Kategorien höherer Stufe (non-terminale Kategorien, z.B. Obliquus) liefert, die für eine angemessene Erfassung der Formseite der Kasus wesentlich sind. Werden non-terminale Kategorien in paradigmatischen Formenspezifikationen zugelassen, so können Homonymien und Synonymien bei den Endungen weitgehend aufgelöst. Im Ergebnis kann ein (im Vergleich zur Zahl der zu besetzenden Paradigmenpositionen) wenig umfangreiches, deklinationenübergreifendes Inventar von Endungen aufgestellt werden, aus deren Zusammenspiel sich die Vielfalt der Deklinationsmuster ableitet.

[Zur Unterscheidung von Form- und Funktionsseite bei Kasus kategorien:] Comrie, Bernard. 1991. „Form and function in identifying cases.“ In: F. Plank (Hg.) *Paradigms. The Economy of Inflection*. Berlin/New York: Mouton de Gruyter, 41-55.

[Zur Konzeption der zugrunde gelegten Kategoriensysteme:] Wiese, Bernd. 2004. „Categories and Paradigms. On Underspecification in Russian Declension.“ In: G. Müller, L. Gunkel & G. Zifonun (Hg.) *Explorations in Nominal Inflection*. Berlin/New York: Mouton de Gruyter, 321-372.

---

Nanna Fuhrhop  
Universität Potsdam

## **Das Zusammenspiel von Morphologie und Syntax in der Schreibung des Deutschen**

Mi 17.15-17.45

---

Dieser Vortrag liefert einen Beitrag zur empirischen Fundierung des AG-Themas, zunächst aus dem Bereich der Graphematik oder auch der Orthographie. Mit zunehmender Erforschung sowohl der Groß- und Kleinschreibung des Deutschen als auch der Getrennt- und Zusammenschreibung im Deutschen stellt sich deutlich heraus, dass sowohl morphologische Kriterien als auch syntaktische Kriterien ihren Beitrag zur jeweiligen Schreibung leisten. Typische Zweifelsfälle können anhand der Kriterien beschrieben werden. So sind zum Beispiel Verbindungen aus Adjektiven und Verben (von *grün streichen*, über *leeressen*, *sich totlachen* bis *krankschreiben*) voneinander abzugrenzen, vom Syntagma, über Reihenbildung bis zur Rückbildung. Bei einer solchen Beschreibung wird jeweils ein Aspekt hervorgehoben (und so sind auch die hier gewählten – aber nicht einzig möglichen – Schreibungen gewählt). Anders als sonst sind in der Schreibung Entscheidungen nötig (es gibt keine HalbgröÙschreibung und auch keine Halbzusammenschreibung) und die Entscheidungen sind auch sichtbar.

Zunächst wird die Interaktion an konkreten Fällen aus der Graphematik vorgestellt. In dem Vortrag soll dann auch getestet werden, inwieweit die gewonnenen Einsichten auf andere Bereiche in der Grammatik zu übertragen sind, ein besonders strittiger Fall sind hier sicherlich die Wortarten, aber auch Fälle wie Funktionsverbgefüge können durch die konkrete Betrachtung sowohl der morphologischen als auch der syntaktischen Eigenschaften und ihres Zusammenspiels adäquater beschrieben werden.

---

Katrin Mutz  
Universität Bremen  
**Französisch *se*: zur Rolle der Argumentstruktur**  
Mi 17.45-18.15

---

Wie in vielen Sprachen der Welt tritt auch der französische Reflexivmarker *se* in einer Reihe von Konstruktionen mit je unterschiedlicher Funktion auf, vgl. u.a. *il se lave* „er wäscht sich“, *ils se regardent* „sie schauen sich an“, *la fenêtre s'ouvre* „das Fenster öffnet sich“, *la tour se voit de loin* „der Turm ist von weitem zu sehen“. Des Weiteren weist das Französische eine große Anzahl von lexikalisierten *se*-Verben auf, z.B. *s'évanouir* „ohnmächtig werden“, *se (mê)ffier* „(miss)trauen“, *s'amuir* „verblassen“.

Nach der Darstellung und Analyse der verschiedenen Verwendungen von nfrz. *se* soll auf der Grundlage einer diachronen Untersuchung der *se*-Konstruktionen und *se*-Verben der Frage nachgegangen werden, wieviele und was für Ebenen der Repräsentation anzunehmen sind, um die *se*-Konstruktionen des Französischen sowohl synchron als auch diachron adäquat beschreiben zu können: In dem Vortrag soll expliziert werden, was für eine Rolle der Argumentstruktur (AS) bei der Analyse und Repräsentation der verschiedenen *se*-Konstruktionen und *se*-Verben zukommt, wie die Argumentstruktur beschaffen sein muss – es wird von einer mehrdimensionalen AS ausgegangen, die mindestens zwei Ebenen, eine semantische und eine syntaktische, umfasst – wie *se*, je nach Funktion, die AS in unterschiedlicher Weise affiziert und wie sich schließlich die Interaktion zwischen Argumentstruktur und morphologischer bzw. syntaktischer Realisierungsebene, zwischen AS und lexikalisch-semantisch-konzeptueller Struktur und AS und der pragmatischen Komponente (Informationsstruktur) manifestiert.

Desweiteren soll anhand von komplexen Wörtern des Typs *s'autodétruire* „sich selbst zerstören“ und *autodestruction* „Selbsterstörung“ bzw. *s'entraider* „sich gegenseitig helfen“ und *entraide* „gegenseitige Hilfe“ und ihrer Situierung in der Syntax das Zusammenspiel bzw. die Arbeitsteilung zwischen Morphologie und Syntax beschrieben werden und auf dieser Basis Rückschlüsse gezogen werden auf die „interne Strukturierung“ des Morphologie- bzw. Syntaxmoduls und die herausragende Rolle der AS.

- Ackema, P. & A. Neeleman. 2004. *Beyond morphology. Interface conditions on word formation*. Oxford: OUP.
- Alsina, A. 1996. *The role of argument structure in grammar*. Stanford: CSLI Publications.
- Booij, G. 1992. „Morphology, semantics and argument structure“. In: I. M. Roca (Hg.) *Thematic Structure: Its Role in Grammar*. Dordrecht: Foris, 47-67.
- Bresnan, J. 2001. *Lexical-Functional Syntax*. Oxford: Blackwell Publishers.
- Grimshaw, J. 1990. *Argument structure*. Cambridge: MIT Press.
- Mutz, K. 2004. „Reflexivierung und Wortbildung“. In: M. Hummel & R. Kailuweit (Hg.) *Semantische Rollen in der Romania*. Tübingen: Narr, 355-374.
- Mutz, K. 2005. „Reflexiva und Verwandtes im Französischen und französisch-basierten Kreolsprachen. Ein Vergleich“. In: C. Sinner & G. Veldre (Hg.) *Diathesen im Französischen / Les diathèses en français*. Frankfurt: Peter Lang, 115 - 136.

---

Sebastian Drude  
Freie Universität Berlin & Museu Goeldi, Belém (Brasilien)  
**Nominale Prädikation oder Modus? *-(t)u*-Formen im Awetí**  
Do 9.00-10.00

---

Im Awetí, einer Tupi-Sprache in Zentralbrasilien, erscheinen neben normalen finiten Verbformen auch Formen mit einem Suffix *-tu* oder *-u*, die sowohl nominale als auch verbale Eigenschaften aufweisen:

- Die Personenpräfixe, die bei *-(t)u*-Formen auftreten, sind weitgehend dieselben, die bei Substantiven die Person des ‚Besitzers‘ kennzeichnen.
- Anstelle eines Personenpräfix‘ kann auch ein Nominalausdruck vor der *-(t)u*-Form stehen, der bei der entsprechenden normalen Verbform eines intransitiven Verbs das Subjekt und bei der eines transitiven Verbs das Objekt ist. Die so gebildete Konstruktion gleicht der Possessiv-Juxtaposition (Besitzer-Besessenes) bei Substantiven.
- Andererseits werden die *-(t)u*-Form-Ausdrücke wohl nie oder fast nie referentiell gebraucht sondern prädikativ, ohne erkennbare Unterschiede in der Proposition.
- Auch sonst verhalten sie sich syntaktisch wie ihre verbalen Entsprechungen, z.B. hinsichtlich der Valenz oder dem Auftreten von Tempus- und anderen Partikeln.

Interessanterweise verhalten sich die beiden Hauptvarietäten des Awetí (der von Männern und der von Frauen gebrauchte Generolekt) leicht unterschiedlich hinsichtlich des wichtigsten morphologischen Kriteriums, der Personenpräfixe. Danach könnte eine scheinbar unbedeutende Eigenschaft wie die Nasalisierung eines Vokals sehr weitreichende Auswirkungen haben.

Zwischen beiden möglichen Analysen – als spezielle Substantive in meist prädikativer Funktion oder als ein Modus, also als zum Verbparadigma gehörig (siehe zu dieser Analyse ausführlicher Drude i.E.) – kann aber nur entschieden werden, wenn nicht nur morphologische Eigenschaften betrachtet werden. Es muß vielmehr geprüft werden, für welche der Möglichkeiten eine detaillierte Analyse der verschiedenen syntaktischen Ebenen (insb. Konstituentenstruktur und grammatische Relationen zwischen den Konstituenten) vorgenommen werden kann, die mit Satzbedeutungen vereinbar ist, die weitgehend äquivalent mit der von Sätzen mit normalen Verbformen ist und sich ausgehend von den anzusetzenden Wortbedeutungen (insb. der der *-(t)u*-Formen) konstruieren läßt.

Für ein solches Vorhaben benötigt man eine umfassende Sprachtheorie, die die morphologische und syntaktische Beschreibung eines Satzes mit seiner Semantik in Zusammenhang bringt. Erste Schritte einer solchen Analyse im Rahmen der Integrativen Sprachwissenschaft sollen in diesem Beitrag vorgestellt werden.

Sollten schließlich für beide Möglichkeiten mögliche Analysen plausibel sein, muß erwogen werden, daß beide in unterschiedlichen gegenwärtigen Idiolektsystemen (oder gar in denselben Idiolektsystemen) verankert sind. Dies würde einen möglichen Sprachwandel anzeigen oder sogar eine über längere Zeit stabile Situation von alternativen, aber äquivalenten Strukturen.

Drude, Sebastian. i.E. „Tense, Aspect and Mood in Aweti Verb-Paradigms: Analytic and Synthetic Forms.“ In: D. Harrison, A. Dwyer & D. Rood (Hg.) *A world of many voices: Lessons from documented endangered languages*. Amsterdam etc.: Benjamins (geplant für 2006).

---

Marie-Hélène Viguière  
Berlin

**Kategorielle Polysemie und semantische Kohärenz:  
morphologische, syntaktische und lexikalische Faktoren bei der  
Bestimmung einer Gesamtbedeutung für das französische Imparfait**  
Do 10.00-10.30

---

Das Imparfait steht wegen seiner widersprüchlichen Verwendungen im Zentrum der meisten Diskussionen um das französische Tempussystem. Die für jede kohärente Kommunikation vorausgesetzte Devise „eine Form – eine Bedeutung“ scheint hier außer Kraft gesetzt zu sein. Dieser Beitrag zielt darauf ab, am Beispiel irrealer Imparfaitvorkommen zu zeigen, daß die semantische Kohärenz einer Kategorie nur unter Berücksichtigung morphosyntaktischer, intonationaler und lexikalischer Parameter gesichert werden kann, die teilweise im gesamten Satz des Vorkommens verteilt sind. Den Rahmen der Untersuchung bildet die Integrative Sprachwissenschaft (vgl. Lieb 1983, 1993), wo für jeden Satz eine Wortfolge, eine mehrgliedrige syntaktische Struktur und eine lexikalische Interpretation angenommen werden. Im Vordergrund der Präsentation stehen die Interaktionen zwischen diesen Dimensionen und die formale Darstellung von kategoriellen Gesamtbedeutungen.

Lieb, H.-H. 1983. *Integrational Linguistics. Vol I, General outline*. Amsterdam: Benjamins.

Lieb, H.-H. 1993. „Integrational Linguistics.“ In: J. Jacobs [et al.] (Hg.) *Syntax: Ein internationales Handbuch zeitgenössischer Forschung / An International Handbook of Contemporary Research*. Vol.1. Berlin etc.: de Gruyter, 430-468.

---

Harald Stamm  
Heinrich-Heine-Universität Düsseldorf

**Kompositionalität in einem multidimensionalen Ansatz: Der Aufbau von  
Nominalgruppenbedeutungen in der Integrativen Satzsemantik**  
Do 10.30-11.00

---

In dem Vortrag wird der Bedeutungsaufbau von Nominalgruppen innerhalb der Integrativen Semantik, einem 'multidimensional basierten' Ansatz, exemplarisch am Deutschen beschrieben (Lieb 1983). Im Kompositionsprozess wird systematisch auf Informationen zurückgegriffen, die durch die unterschiedlichen Beschreibungsebenen zur Verfügung gestellt werden. Zu diesen Informationen gehören morphosyntaktische Informationen zu Numerus, Kasus, Definitheit, lexikalische Informationen zu Wortbedeutung, Wortartangaben (z.B. DEMONSTRATIVPRONOMEN) und syntaktische

Informationen zur Konstituentenstruktur, den bestehenden syntaktischen Funktionen und zur Intonation.

Am Beispiel definitiver Possessivkonstruktionen des Deutschen (z.B. *der Garten des Mannes, der Garten von dem Mann, sein Garten, das Gewicht des Mannes, die Väter der Kinder*) wird die Verarbeitung heterogener Information im Kompositionsprozess nachgewiesen. Possessivkonstruktionen stellen eine notorische Schwierigkeiten für eine kompositionale Semantik dar, was durch die aktuelle Literatur unterstrichen wird (Kim et al. 2004). Der Beitrag wird Bezug nehmen auf aktuelle Debatten um Fragen der Kompositionalität (Werning et al. 2005). Als Ausblick kann die vorgeschlagene Analyse mit Analysen in konkurrierenden, nicht-mehrdimensionalen Ansätzen verglichen werden.

- Kim, J.-Y., Y. A. Lander & B. H. Partee (Hg.) 2004. *Possessives and Beyond: Semantics and Syntax*. University of Massachusetts Occasional Papers in Linguistics 29. Amherst, Mass. USA: GLSA Publications.
- Lieb, H.-H. 1983. *Integrational Linguistics*. Vol. I: *General Outline*. (Current Issues in Linguistic Theory, 17.) Amsterdam usw.: Benjamins.
- Machery, E., M. Werning & G. Schurz (Hg.) 2005. *The compositionality of meaning and content*. Vol. II: Applications to Linguistics, Psychology and Neuroscience. Frankfurt: Ontos Verlag.
- Werning, M., E. Machery & G. Schurz (Hg.) 2005. *The compositionality of meaning and content*. Vol. I: *Foundational Issues*. Frankfurt: Ontos Verlag.
- 

Hans-Heinrich Lieb  
Freie Universität Berlin

**Wortbildung auf morphologischer und Wortbildung auf syntaktischer  
Grundlage**  
Do 11.30-12.30

---

Es wird eine Auffassung von Wortbildung skizziert, die den folgenden Grundsätzen genügt:

1. Wortbildung ist die Bildung von lexikalischen Wörtern (bestehend aus einem Wortparadigma und einem Begriff), deren Formen syntaktische Einheiten sind; Wortbildung gehört deshalb in die Syntax und nicht in die Morphologie.
2. Wortbildung kann auf morphologischer Grundlage erfolgen, durch Stammformen-Bildung (diese gehört in die Morphologie) oder auf syntaktischer Grundlage (z.B. als syntaktische Ableitung bei Partikelverben).
3. Für Wortbildung ist (anders als in der Integrativen Linguistik bisher angenommen) eine IP-Auffassung angemessen, rekonstruiert nicht durch Algorithmen, sondern durch Funktionen bestimmter Art.

Die Skizze verwendet Beispiele hauptsächlich aus dem Deutschen, akzeptiert den theoretischen Rahmen der IL und verbindet im Zusammenhang mit Wortbildung WP-, IA- und IP-Konzeptionen.

---

Louisa Sadler  
University of Essex  
**Single Conjunct Agreement and Multidimensionality**  
Fr 10.30-11.30

---

Despite the considerable attention that agreement phenomena have received in recent years in the theoretical linguistic literature, rather less attention has been paid to the analysis of single conjunct agreement (SCA) patterns, where the agreement controller is not the coordinate structure as a whole but one of the conjuncts (but see especially Moosally 1998, Ackema and Neeleman 2004).

This paper considers the consequences of a range of distinct single conjunct agreement patterns for our view of the role and location of agreement processes in the grammar, arguing that agreement is a multi-dimensional phenomenon, making reference both to linear position and syntactic relations (such as subject, object, adjunct and the like). We first briefly show how predicate argument agreement in Welsh (and in Celtic languages in general) systematically targets the ‘closest’ conjunct, but that other agreement processes target the features of the coordinate structure as a whole. This indicates that a notion of resolution (a grammatical mechanism which ‘calculates’ the set of resolved features to be associated with the coordinate structure as a whole), is not sufficient to account for SCA agreement patterns: in such languages both the resolved features and those of the distinguished conjunct must be syntactically accessible.

We then turn to a discussion of SCA agreement patterns within NP, that is, concord between determiners, adjectives and nouns, or head-modifier agreement (in contrast to predicate-argument agreement), focusing on coordinated NPs in Portuguese. On the basis of a relatively large corpus study, we present the complex picture of agreement patterns, within and beyond NPs, which emerges from the corpus data.

The data highlights in particular the importance of a linearity condition in formulating agreement generalizations: the data shows that in addition to agreement by resolution in coordinate NPs, both leftmost and rightmost single conjunct agreement occur relatively frequently within NPs, and indeed both can co-occur within the same coordinate structure. We consider the implications of the Portuguese data for an analysis in LFG, where agreement is modeled at f-structure. Reference to the agreement features of both initial and final conjuncts can of course be achieved at f-structure by positing additional LAGR (left agreement) and RAGR (right agreement) features in the f-structure of the coordinate structure, as copies of the features of the leftmost and rightmost conjunct, alongside the resolved features of the coordinate structure as a whole. However the patterns of target-controller associations in both the Portuguese and Welsh data are strongly suggestive of an explicit linear order constraint on choice of controller, and thus of a multidimensional approach to agreement, which should be directly coded in the grammar.

Ackema, P. & A. Neeleman. 2004. *Beyond Morphology: Interface Conditions on Word Formation*. Oxford: OUP.



Berthold Crysmann  
Deutsches Forschungszentrum Künstliche Intelligenz, Universität des  
Saarlandes  
**A Coanalysis Approach to Polish Past Tense Agreement**  
Fr 11.30-12.30

---

The status of Polish past tense agreement markers has been a matter of considerable debate (Spencer 1991; Borsley & Rivero 1994; Borsley 1999; Bąnski 2000; Kupść 2000; Kupść & Tseng 2005). Past tense agreement is expressed by a set of bound forms that either attach to the past participle, or else “float off” to a host further to the left. Despite this relative freedom of attachment, it is often noted in the literature (e.g., Borsley 1999; Kupść & Tseng 2005) that the combination of verbal host and agreement marker forms a wordlike unit, suggesting that the agreement marker functions essentially like an inflectional affix: they observe a strict adjacency requirement when appearing to the right of the verb, trigger morphophonological idiosyncrasies on their verbal host, and fail to take wide scope over a coordination of hosts. Within HPSG, current approaches to the morphosyntax of Polish past tense agreement markers appear to emphasise either the affix-like properties (e.g. Kupść & Tseng 2005) or syntactic placement (Borsley 1999). I shall propose to augment Borsley’s linearisation-based approach to clitic placement with a linearisation-based theory of morphologically discontinuous elements (Crysmann 2003). This approach assumes that lexical signs, just like phrasal ones may correspond to more than a single object in linear structure (Reape 1994; Kathol 1995). Morphological rules of exponence may then distribute the phonological yield across different and possibly discontinuous order domain objects contributed by a lexical sign. I will show that the division of labour between morphological introduction and syntactic placement is sufficient to give a unified treatment to both syntactic and morphological properties of the Polish past tense agreement marker. Under the perspective of lexically introduced discontinuous affixal phonology, no additional book-keeping is required since the markers are part of the lexical representation of the verb, having full access to the verb’s valence information. Likewise, since floating markers originate in the lexical representation of a (complex) verb, unique exponence follows directly. Finally, a linearisation-based approach to morphosyntactic coanalysis can capture quite naturally the intuition that Polish agreement markers are indeed verbal inflection simply “floating off”.

Bąnski, P. 2000. Morphological and prosodic analysis of auxiliary clitics in Polish and English. Ph.D. thesis, Uniwersytet Warszawski, Warsaw.

Booij, G. & J. Rubach. 1987. „Postcyclic vs postlexical rules in lexical phonology.“ In: *LI* 18, 1-44.

- Borsley, R. 1999. „Weak auxiliaries, complex verbs and inflected complementizers in Polish.“ In: Borsley & Przepiórkowski (Hg.) *Slavic in Head-Driven Phrase Structure Grammar*. Stanford: CSLI Publications, 29-59
- Borsley, R. & M. L. Rivero. 1994. „Clitic auxiliaries and incorporation in Polish.“ In: *NLLT* 12, 373–422.
- Crysmann, B. 2003. *Constraint-based Coanalysis. Portuguese Cliticisation and Morphology–Syntax Interaction in HPSG*. Saarbrücken: Saarland University and DFKI LT Lab.
- Kathol, A. 1995. Linearization-based German syntax. Ph.D. thesis, Ohio State University.
- Kupść, A. 2000. An HPSG grammar of Polish clitics. Ph.D. thesis, Polish Academy of Sciences and Université Paris 7.
- Kupść, A. & J. Tseng. 2005. „A new HPSG approach to Polish auxiliary constructions.“ In: S. Müller (Hg.) *Proceedings of the 12<sup>th</sup> Intl' Conference on HPSG*. Stanford: CSLI publications, 253–273.
- Miller, P. 1992. *Clitics and Constituents in Phrase Structure Grammar*. New York: Garland.
- Reape, M. 1994. „Domain union and word order variation in German.“ In: Nerbonne et al. (Hg.) *German in Head-Driven Phrase Structure Grammar*. Stanford: CSLI Publications, 151–197.
- Spencer, A. 1991. *Morphological Theory*. Oxford: Blackwell.

Andreas Nolda  
Humboldt-Universität zu Berlin  
**Nounless Nominals in German:**  
**Ellipsis, Use as Noun, Nominalization**  
Fr. 12.30-13.00

In German, there is reason to distinguish between three kinds of *nounless nominals* (‘elliptical noun phrases’):

1. nounless nominals like *rosa* in (1),
  2. nounless nominals like *rosanen* in (2), and
  3. nounless nominals like *Rosanes* in (3).
- (1) *Peter glaubt, dass Claudia rosa und Florian blauen Stoff mag.*  
Peter believes that Claudia pink and Florian blue ACC SING MASC cloth likes  
‘Peter believes that Claudia likes pink cloth and that Florian likes blue cloth.’
- (2) *Florian mag blauen Stoff, und Claudia mag rosanen.*  
Florian likes blue ACC SING MASC cloth and Claudia likes pink ACC SING MASC  
‘Florian likes blue cloth and Claudia likes something pink denotable by a German masculine mass noun.’
- (3) *Claudia mag Rosanes.*  
Claudia likes pink ACC SING NEUT  
‘Claudia likes pink things.’

Presupposing the multi-dimensional framework of Integrational Linguistics, I shall argue in my talk that those nominals are nounless in different respects:

1. in (1), the head noun is *elided by phonological reduction*;
2. in (2), the head is an adjective *syntactically used as noun*;
3. in (3), the head is a *lexically nominalized adjective*.

### **AG 3 How to recognize a sentence when you see one: methodological and linguistic issues in the creation of sign language corpora**

Koordination: Onno Crasborn und Els van der Kooij (Kurz-AG)

#### **Programm**

Donnerstag 23.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
09.00-09.30	Onno Crasborn, Els van der Kooij	<i>Introduction: what's the problem?</i>
09.30-10.00	Invited speaker: Elisabeth Engberg-Pedersen	<i>Intertwined ellipsis – a multicausal construction in Danish Sign Language</i>
10.00-10.30	Irene Greftegreff	<i>Where Shall We Cut? Issues in Analysing Norwegian Sign Language Texts</i>
10.30-11.00	Josep Quer	<i>Transcription as heuristic tool</i>
11.00-11.30	Kaffeepause	
11.30-12.00	Martje Hansen, Jens Heßmann	<i>Reanalysing sentences in German Sign Language (DGS) texts</i>
12.00-12.30	Marion Blondel	<i>Even more complex than that! The case of bimodal utterances before 2 years old</i>

Freitag 24.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
10.30-11.00	Kristin Mulroony	<i>Where does the period go? Defining a line for narrative analysis</i>
11.00-11.30	Ulrike Wrobel	<i>How to utter a complex sentence by using a single sign</i>
11.30-12.00	J. Fenlon, T. Denmark, B. Woll	<i>Seeing sentence boundaries</i>
12.00-13.00	Mittagspause	
13.00-13.30	Felix Sze	<i>Identifying sentence boundaries: a case study of Hong Kong Sign Language discourse data.</i>
13.30-14.00	Brenda Nicodemus	<i>Prosody and utterance boundaries in ASL interpretation</i>
14.00-14.30	Trevor Johnston, Adam Schembri	<i>Identifying clauses in three signed languages: applying a functional approach</i>
14.30-15.00	Final discussion	<i>What about the sentence in signed languages...</i>

---

Elisabeth Engberg-Pedersen  
University of Copenhagen  
**Intertwined ellipsis - a multiclausal construction in Danish Sign  
Language?**  
Do 9.30-10.00

---

In contrast to Neidle et al.'s claim about ASL (2000) I find that space plays a minor role in the syntactic organization of clauses in Danish Sign Language. Basic clauses in Danish Sign Language are generally of two kinds, NVN- or NNV-clauses with the more active participant - A - before the less active participant - P and NV-clauses which are pragmatically organized as topic - comment structures, i.e. SECOND LOOSE means 'the second one lost' or 'the second one was lost' as the case may be. Zero anaphora is widespread in Danish Sign Language and seems to be partly pragmatically influenced: in a topic chain a referent need not, but may, be mentioned with later verbs. In the following example, however, there seems to be prosodic and semantic reasons for talking about a grammatical construction with zero anaphora:

LATER BUS General-object-Pm+move-in-zigzag-from-centre-forward-right / TOWN  
NAME

Salbach / PRON+fr /

'Later we went by bus to a town called Salbach.'

[TOWN NAME Salbach] can be analysed as an internally headed relative clause with TOWN being the A argument of CALL and the zero anaphor being the Goal argument of General-object-Pm+move-in-zigzag-from-centre-forward-right. Further, a zero topic can alternate its semantic relation (A or P) to a series of verbs, and two referents may even both be left unmentioned when they alternate their relations (A and P) to a series of verbs. It is this latter type of sequence I tentatively call intertwined ellipsis, and in my talk I shall discuss possible criteria for describing it as a grammatical construction.

Neidle, C., J. Kegl, D. MacLaughlin, B. Bahan & R.G. Lee. 2000. *The syntax of American Sign Language: functional categories and hierarchical structure*. Cambridge, Mass.: The MIT Press.

---

Irene Greftegreff  
Møller Resource Centre

**Where Shall We Cut? Issues in Analysing Norwegian Sign Language  
Texts**

Do 10.00-10.30

---

For over a decade our centre has been producing materials for sign language instruction that have been aimed at varying groups of users. In recent years the rapid spread of relatively cheap and powerful computers have enabled us to produce teaching materials where video is linked with instructive text. This new opportunity to segment the video stream immediately presented us with the burning issue of how to perform the segmentation, the problem of Where Shall We Cut?

In the process following we have found formal syntactic theories to be of relatively little use. A likely explanation is that these theories have developed through analysis of written language sentences, and have largely avoided the additional complications that occur in spoken language texts, such as rhythm, tempo tone and illustrative gestures. The concepts that we have found workable have come instead from discussions of sign language genres, narrative theory, discourse analysis, and from functional grammar. Also, the demand that we translate our teaching material into Norwegian has turned out to be helpful in unsuspected ways, as it has forced us to constantly ask ourselves exactly what meaning is coded and in which form, and to look for analogous phenomena in the target language. The presentation will contain several illustrative examples.

---

Josep Quer  
Universitat de Barcelona  
**Transcription as heuristic tool**  
Do 10.30-11.00

---

Defining the boundaries of (SL) sentences obviously requires parallel evaluation of a multiplicity of cues provided by various grammatical levels, crucially including lexical, morphophonological, syntactic, semantic and prosodic information. Ideally, the task should be carried out by a linguistically trained native signer who has a solid background on all aspects of the grammar of the target SL. However, important parts of SL grammars are still to be characterized properly. It thus becomes necessary to use the information obtained from the transcription task itself in order to complete and refine our knowledge about the structure of the language. Two concrete examples of this sort of two-way feedback will be discussed: (i) the ambiguous status of the possessive sign OF/POSSESS in LSC and (ii) the predicative status of a localizing index (Zwitserslood 2003, van Gijn 2004, Crasborn p.c.). The central issue addressed is that the occurrence of a predicative element is a necessary condition for defining

sentence boundaries, but still an insufficient one: not only do we still have to resolve the issue of complex sentences containing embedded clauses, but we also must take into account that predicative relations occur within the nominal domain parallel to the clausal one (Abney 1987, Kayne 1994).

Abney, S.P. 1987. *The English Noun Phrase in Its Sentential Aspect*. Doctoral dissertation, MIT.

Kayne, R.S. 1994. *The Antisymmetry of Syntax*. Cambridge, Mass.: MIT Press.

van Gijn, I. 2004. The quest for syntactic dependency. Sentential complementation in Sign Language of the Netherlands. Utrecht: LOT.

Zwitzerlood, I. 2003. *Classifying hand configurations in Nederlandse Gebarentaal (Sign Language of the Netherlands)*. Utrecht: LOT.

---

Martje Hansen, Jens Hessmann  
University of Applied Sciences Magdeburg-Stendal (FH)  
**Reanalysing sentences in German Sign Language (DGS) texts**  
Do 11.30-12.00

---

Hessmann (2001) assembles a range of information to allow users to access a corpus of natural DGS data. The corpus consists of ten spontaneously signed texts, totalling 56 minutes (all of the signed texts and some of the transcripts and commentaries are available at <http://www.let.ru.nl/sign-lang/echo>). Some notion of ‘sentence’ is central to this work, as all the information given, such as transcriptions, commentaries, annotations, and translations, refer to “meaningful units that consist of sequences of mostly small numbers of manual signs” (p. 119). Signals often mentioned in the literature such as rest position or prolonged holding of the hands, eye blinks, etc., are found to be relevant but not conclusive in establishing sentence boundaries. Rather, Hessmann (2001) opts for a pragmatic notion of sentence: Breaking down the flow of manual and nonmanual activity into meaningful chunks is part and parcel of the process of understanding signed texts.

Recently, we have started to reanalyse the division into sentences found in Hessmann (2001) on the basis of a distinction between topics and predicates in signed texts. Topics are linguistic expressions that correspond to what an utterance is about; often, though not in every instance, they refer to ‘old information’. Interestingly, the joint articulation time of topics and predicates in signed utterances conforms to a temporal pattern which corresponds to the timing of spoken utterances in natural discourse as described in the literature: Overall lengths of signed and spoken ‘sentences’ are similar, which is probably due to limitations of the working memory. Topics are generally considerably shorter than predicates, and this cannot be explained by articulatory constraints concerning single signs, but may be caused by the constraints of cognitive speech processing: Old information occupies less processing time than new information. On the whole, the temporal patterning of topics and predicates in discourse supports the semantically based segmentation of signed texts into sentences.

Our presentation will demonstrate in what way this kind of reasoning refines, though not overturns the sentence division originally presented in Hessmann (2001). Though arguably the term sentence as used here corresponds more closely to the traditional notion of ‘clause’, some indication will be given of units that may be considered complex sentences.

Jens Hessmann (2001), *GEHÖRLOS SO!: Materialien zur Gebärdensprache*. 2 Volumes. Hamburg: Signum.

---

Marion Blondel  
CNRS, Université de Rouen

**Even more complex than that! The case of bimodal utterances before 2 years old**  
Do 12.00-12.30

---

My presentation will question the sentence domain through a very young bilingual child production.

As for any signed production, we need to segment the gestural flow in order to offer a suitable transcription; moreover, to understand how morphosyntax is emerging in a young signer we have to delimitate the domain in which units interact at a morphosyntactic level.

My presentation is based on the data we analyzed for the following specific purpose: to determine both status and role of the pointing gestures by a bilingual and bimodal (LSF-French) child between 9 and 22 months with the hypothesis that pointing gestures start from a prelinguistic status to a grammatical one (Blondel & Tuller, submitted paper). Hence we face the well-known questions in child production: to know for instance whether two consecutive signs belong to the same sentence or constitute two independent sentences (Pt1 LEAVE ‘I leave’ or ‘It’s me, I leave’ for instance). The challenge will be to deal with two modalities, prelexical units like miam-miam ‘ium-ium’ and the combinations of prelinguistic gesture with linguistic units. I will propose to confront our data with criteria already exposed in Baker et al. (1999) such as the use of pauses, changes in facial expression or eye-gaze direction “in combination with grammatical unity and semantic cohesion” (p16). This is the opportunity to submit unsolved cases.

Baker A., van der Bogaerde B., Coerts J., Woll B. (1999) « Methods and procedures in sign language acquisition studies ». Paper presented at the 4th Intersign Workshop London. Manuscript.

Blondel M. & Tuller L. (submitted paper) «Pointing in bimodal bilingual acquisition: a longitudinal study of LSF-French bilingual child», TISLR 8 in Barcelona, september 2004.



---

Kristin Mulrooney  
Gallaudet University

**Where does the period go? Defining a line for narrative analysis**  
Fr 10.30-11.00

---

Linguistic research on narratives in American Sign Language has focused primarily on elicited or performance narratives Gee and Kegl (1983), Bahan and Supalla (1995), Rose (1996). This study analyzes twelve personal narratives that occur in conversation in an attempt to describe the structure of narratives in ASL. Traditional line-based transcripts were made, typical of the first step in a narrative analysis. The transcripts were enhanced by frame shots of all signs and gestures produced in the narratives. This was imperative for an effective analysis of a language that depends on space and facial expressions in addition to lexical signs to convey meaning. The production of the transcript required a decision to be made regarding what should be considered a 'line' in the transcript. The concept of intonation units as defined by Chafe (1994) was used to define what this smallest unit of analysis would be. The criteria for identifying these units were expanded to include direction of eye gaze and body position, eye blinks, and extended pauses between signs. This presentation explains how the concept of intonation unit can be applied to sign language data and why it may provide different insight to data than using a sentence as the line boundary.

---

Ulrike Wrobel

Ludwig-Maximilians-Universität München

**How to utter a complex sentence by using a single sign**  
Fr 11.00-11.30

---

A well-known and often used sentence concept is the structural one: a minimal sentence is defined to be a core structure consisting of subject and predicate. Although this concept has proved its worth for the analysis of spoken languages, its application on sign language linguistics poses several theoretical problems (cf. BOUCHARD/DUBUISSON (1995), cf. ERLINKAMP (2000) and DOTTER (2001) for DGS). One of them is the observation that there are single signs that seem to function as whole sentences. Evidence for this is given by the following examples of DGS: the so-called Agreement Verbs (like Ich-frage-sie), the "Classifier Verbs" (like Zweiachsiges-Fahrzeug-fährt-in-einer-Rechtskurve-den-Berg-hoch), and the so-called Idioms (MALLY 1993) or "Special Signs" (HILLENMEYER 2000) (like Zum-einen-Ohr-rein-zum-anderen-wieder-raus, Worte-ignorieren or Verwechseln/sich irren/ verwirrt sein). These sequences of movements are linguistic devices to express one "sentence" by using a single sign – without being elliptical constructions. Since the latter also possess some properties that contradict the standard assumptions

made by sign language morphophonology, the discussion of these borderline cases is one of the means for gaining new theoretical insights in sign languages. A functional-pragmatic approach will be introduced to revise the sentence concept by emphasizing the actional character of language.

- BOUCHARD, DENIS/ DUBUISSON, COLETTE, 1995. Grammar, order and position of wh-signs in Quebec Sign Language. *Sign Language Studies* 24; 99-139.
- DOTTER, FRANZ, 2001. Gebärdensprachforschung. In: MARTIN HASPELMATH, EKKEHARD KÖNIG, WULF ÖSTERREICHER (eds.). *Language Typology and Language Universals 1*. Berlin, New York: Walter de Gruyter; 141-154.
- ERLENKAMP, SONJA, 2000. Syntaktische Kategorien und lexikalische Klassen. *Typologische Aspekte der Deutschen Gebärdensprache*. München: Lincom Europa.
- HILLENMEYER, MARGIT, 2000. DGS. *Deutsche Gebärdensprache [Lehrbücher für Bayern]*. München: Margit Hillenmeyer.
- MALLY, GERTRUD, 1993. *Redewendungen in der Deutschen Gebärdensprache*. Hamburg: Signum.
- 

J. Fenlon, T. Denmark and B. Woll  
University College London  
**Seeing sentence boundaries**  
Fr 11.30-12.00

---

Prosody has many functions in languages. Intonation in spoken languages identifies clause boundaries and can distinguish between an utterance intended as a statement or question. Little is known about prosody in British Sign Language. Linguists have identified both pragmatic and syntactic functions when examining the distribution of non-manuals (see Sutton-Spence & Woll 1999). However, there is little knowledge regarding how native signers may interpret non-manuals with respect to sentence boundaries.

The study described here, which is part of a larger study of visual prosody, is a pilot aiming to clarify whether there are any prosodic cues that indicate sentence boundaries in BSL; whether such cues can be used when looking at an unfamiliar sign language; and whether visual prosodic cues are unique to sign language

Participants watch a video retelling of the story of 'The Tortoise and the Hare' and are asked to press a button when they see a sentence boundary. Responses are timed and analysed to examine their correspondence with manual cues, non-manual cues, or both.

Data are being collected from two groups of participants on the following tasks:

1. Deaf fluent signers of BSL watching the story in BSL
2. Deaf fluent signers of BSL watching the story in Swedish Sign Language
3. Deaf fluent signers of BSL watching the story in English
4. Hearing non-signing speakers of English watching the story in BSL
5. Hearing non-signing speakers of English watching the story in English without a soundtrack

This study will provide a greater understanding of the perception of sentence boundaries in BSL and visual English, as well as determining the extent to which non-manuals provide an understanding of sentence structure in BSL.

---

Felix Sze  
University of Bristol  
**Identifying sentence boundaries: a case study of Hong Kong Sign  
Language discourse data.**  
Fr 13.00-13.30

---

Despite the fact that numerous linguistic theories have been built upon the notion of 'sentence', there has been no consensus among linguists as to what constitutes a sentence (Beaugrande 1999). Finding sentence boundaries in sign language presents an even bigger challenge because sign language data collected through naturalistic settings may be full of hesitations, pauses, and speech errors. It has been suggested that constituent structures or sentence boundaries in sign language may be followed by relatively longer pauses, blinks, gaze change, or an across-the-board change of non-manual features (Grosjean and Lane 1977, Baker and Padden 1978, McIntire 1980, Wilbur 1994, Sandler 1999). In this presentation, I will discuss the extent to which these suggestions are borne out empirically by spontaneous production data of Hong Kong Sign Language. My preliminary observation suggests that none of the above features may serve as a reliable indicator of sentence boundary on its own. While a typical sentence boundary may be marked by most or all of these features, their absence does not necessarily entail an absence of sentence boundary. Hence sign linguists need to make use of the syntactic, semantic as well as prosodic information in the data in order to identify sentence boundaries.

- Baker, C. & Padden, C. (1978) 'Focusing on the nonmanual components of American Sign Language.' In P.Siple (ed.) *Understanding Language through Sign Language Research*. New York: Academic Press.
- Beaugrande, R (1999). Sentence first, verdict afterwards: On the remarkable career of the "sentence". *Word, Journal of the International Linguistic Association* 50, 1-31.
- Grosjean, F. and Lane, H. (1977). Pauses and syntax in American Sign Language. *Cognition*, 5, 101-117
- McIntire, M.L. (1980) *Locatives in American Sign Language*. Unpublished Ph.D dissertation, University of California.
- Sandler, W. (1999) The medium and the message: Prosodic interpretation of linguistic content in Israeli Sign Language. *Sign Language and linguistics* 2:2, 187-215
- Wilbur, R. (1994) 'Eyeblinks & ASL Phrase Structure'. *SLS* 84. pp.221-240

---

Brenda Nicodemus  
University of New Mexico  
**Prosody and utterance boundaries in ASL interpretation**  
Fr 13.30-14.00

---

Research in a number of spoken languages has shown that speakers use a variety of means to indicate the organization of a message into phrasal units, and that listeners are sensitive to this organization (e.g., Wightman et al. 1992, van Donzel 1999). The organization of the message may be especially critical when language is being transmitted through an interpreter since the message is conveyed via a third party, rather than directly from speaker to listener. In this study, we investigated the communication of phrasing in English-to-ASL interpretation by asking Deaf ASL users to indicate perceived boundaries as they viewed a video of an interpreted lecture. The locations of these perceived boundaries were then examined to see what prosodic cues might be responsible for the perception of a boundary. The study is unusual in focusing on the use of phrasing in interpretation, and in also contributing to the extremely limited literature on the use of prosody in signed languages (e.g., Wilbur 1994, Boyes-Braem 1999, Sandler 1999, Winston and Monikowski 2003).

The source material for this study was a videotape of a 15-minute lecture, presented in English by a university lecturer. Five certified, experienced signed language interpreters were then videotaped as they interpreted the lecture. Each interpreter was recorded on video separately with the same Deaf person serving as audience during all five interpretations. The five interpretations were transferred to a Macintosh PowerBook and loaded into iMovie software. The material was then presented to Deaf ASL users, who were asked to press a key each time they perceived a “sentence” boundary. Before viewing the interpreted lecture, the Deaf judges were shown a video in ASL giving instructions explaining their task, and had a three-minute practice session. The instructions explicitly stated that the “sentences” to be identified did not have to correspond to a complete sentence as in written English. The term “sentence” was used because it was the most appropriate term that was likely to be familiar.

Ten different judges viewed each of the five interpretations. The number of boundaries indicated averaged 66 to 85 per interpreter. Clusters of agreement were identified, where a majority of the judges who viewed a given interpretation had indicated a boundary within an interval of one second. Examination of the recorded interpretations during the times where boundaries were perceived suggest that the interpreters used a variety of prosodic markers to indicate boundaries. These included the use of multiple articulators: hands clasping in front of the body, prolonged duration of final handshape, eye blinks, head nods, and shift of the body back to neutral signing space. Of these, hand clasps appear to be the most reliable boundary marker. In addition, a number of facial gestures were associated with perceived boundaries, including pursed lips, furrowed brows, and eye gaze shifts.

The analysis so far demonstrates that Deaf participants are able to identify sentence boundaries in ASL interpretation, although with varying levels of agreement. The Deaf participants' ability to perform this metalinguistic task may be facilitated by the interpreters' use of multiple articulators to indicate prosodic boundaries, which provides further evidence of 'layering' (Wilbur 2000) in ASL. Even more than in spoken language prosody, the signed language modality gives the language user many choices in the production of prosody.

- Boyes-Braem, P. (1999). Rhythmic temporal patterns in the signing of deaf early and late learners of Swiss German sign language. *Language and Speech* 42, 177-208.
- Sandler, W. (1999). Prosody in two natural language modalities. *Language and Speech* 42, 127-142.
- van Donzel, M. (1999). Prosodic aspects of information structure in discourse. The Hague: Holland Academic Graphics.
- Wightman, C., Shattuck-Hufnagel, S., Ostendorf, M. & Price, P. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 92, 1707-1717.
- Wilbur, R. (1994). Eye blinks and ASL phrase structure. *Sign Language Studies* 84, 221-240.
- Wilbur, R. (2000). Phonological and prosodic layering of nonmanuals in American Sign Language. In K. Emmorey & H. Lane (eds.), *The Signs of Language Revisited*. Mahwah, NJ: Erlbaum. Pages 215-244.
- Winston, E., & Monikowski, C. (2003). Marking topic boundaries in signed interpretation and transliteration. In Melanie Metzger, Steven Collins, Valerie Dively, and Risa Shaw (eds.) *From topic boundaries to omission: New research on interpretation* (pp. 187-227). Washington, D.C.: Gallaudet University Press.

---

Trevor Johnston, Adam Schembri

**Identifying clauses in three signed languages: applying a functional approach**  
Fr 14.00-14.30

---

In this paper, we revisit data collected from 12 native signers of three signed languages: Australian Sign Language (Auslan), Flemish Sign Language (Vlaamse Gebarentaal or VGT) and Irish Sign Language (Vermeerbergen et al., 2004). We re-examine the data in light of Van Valin and La Polla's model of the "layered structure of the clause", an analysis that draws on universal distinctions in clause structure (Van Valin & LaPolla, 1997). In this approach, clauses may be divided into core and peripheral elements, with the core further subdivided into a nucleus and one or more core arguments (note that 'sentences' may be a single clause or a clause complex). We explore constructions that are discussed in the literature and appear in our data, such as split sentences, verb sandwiches, verb doubling, and topicalised elements. Prosodic features and their possible role in determining clause boundaries will also be mentioned. We suggest that the distinction between core and peripheral arguments appears not to be highly grammaticalised in Auslan (i.e., is not consistently marked by the use of prepositions or other grammatical elements), and discuss some of the implications of this finding for recognising clause boundaries in discourse and the annotation of a digital Auslan corpus.

- Van Valin, R.D. & LaPolla, R. J. (1997). *Syntax: Structure, meaning and function*. Cambridge: Cambridge University Press.
- Vermeerbergen, M., Leeson, L., Schembri, A. & Johnston, T. (2004). Considering cross-linguistic analysis of constituent ordering in sign languages: A presentation in three parts. *Workshop on Sign Languages: A Cross-Linguistic Perspective* at the *Annual Meeting of the German Linguistic Society (DGfS)*, Mainz, Germany, February 25-27.

## AG 4 Current Issues in Areal Typology

Koordination: Michael Cysouw & Balthasar Bickel

### Programm

Mittwoch 22.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
	<b>General Session</b>	
13.45-14.15	Jeanette Sakel, Yaron Matras	<i>Prototypical linguistic areas and borrowing hierarchies</i>
14.15-14.45	Nick Enfield	<i>Lessons of areal typology: population thinking for problems of language, cognition, and culture</i>
14.45-15.15	Søren Wichmann	<i>Pre-neolithic linguistics</i>
15.15-15.45	Michael Cysouw	<i>Identifying linguistic areas from typological surveys</i>
15.45-16.15	Kaffeepause	
	<b>Session on Europe</b>	
16.15-16.45	Thomas Stolz	<i>The European north-west – an area with non-European features</i>
16.45-17.15	Johannes Helmbrecht	<i>On the development of Standard Average European (SAE) - the case of polite pronouns</i>
17.15-17.45	Anna Giacalone, Andrea Sanso	<i>The demise of indefinite man constructions in European languages: An areal perspective</i>
	<b>New tools for areal linguistics</b>	
17.45-18.15	Hans-Jörg Bibiko	<i>Visualisation of geographical patterns</i>

Donnerstag 23.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
	<b>Session on Asia</b>	
09.30-10.00	Don Stilo	<i>Can linguistics and genetics work together to unravel the pre-history of the peopling</i>

10.00-10.30	Brigitte Pakendorf	<i>of the Iranian Caspian seacoast?</i>
10.30-11.00	Balthasar Bickel	<i>Genes and languages in Siberia</i>
11.00-11.30	Kaffeepause	<i>Statistical issues in areal typology</i>
	<b>Session on Asia (continued)</b>	
11.30-12.00	Kristine Hildebrandt	<i>The areal and genealogical dimensions of hiatus</i>
12.00-12.30	Peter Hook, Kaoru Horie, Shinji Ido, Joungmin Kim, Qing-Mei Li, Hsin-hsin Liang, Colin P. Masica, Yutaro Murakami, Prashant Pardeshi, Kingkarn Thepkanjana	<i>Areal distribution and semantic evolution of EAT in the languages of Asia</i>

Freitag 24.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
	<b>Session on Africa &amp; South America</b>	
10.30-11.00	Tom Güldemann	<i>Towards the synchronic macro-areal profile of Africa</i>
11.00-11.30	Mily Crevels, Pieter Muysken, Hein van der Voort	<i>The Guaporé-Mamoré region: a linguistic area?</i>
11.30-12.00	Dmitry Idiatov	<i>The identity of WHO? and WHAT? in South America: an areal feature?</i>
12.00-13.00	Closing Discussion	



---

Janet Sakel, Yaron Matras  
University of Manchester  
**Prototypical linguistic areas and borrowing hierarchies**  
Mi 13.45-14.15

---

Linguistic areas often go back to long-standing historical-cultural contacts between speech communities in the region in question. They are characterised by clusters of isoglosses. The method for identifying linguistic areas has so far been to 1. identify some similarities between languages, 2. identify the region as a cultural-historical contact zone and 3. identify in more detail distinguishing features of the area. This was Weinreich's (1953, 1958) agenda of the study of 'convergent areas', and is well exemplified by the numerous studies devoted to well-known linguistic areas such as the Balkans and South Asia.

In recent years, a somewhat more generous or flexible understanding of linguistic areas has emerged: Areas are firstly identified based on a pre-conception of the region as a historical-cultural contact zone. They are then examined in order to establish whether or not structural features are shared among the languages in that region – thus, in effect, skipping phase 1 (i.e. basing the investigation on tenable evidence for contact). This approach has been referred to as 'areal typology' (cf. Koptjevskaja-Tamm & Dahl (2001) on the Circum-Baltic area, and Ramat & Stolz (2002) on the Mediterranean region).

A still more recent trend attempts a 'de-construction' of the notion of linguistic areas, arguing that they are not fundamentally different from a mere density of shared isoglosses that go back to the diffusion of features between pairs of languages (cf. Matras, McMahon & Vincent eds., in press; especially the contributions by Campbell & Bakker). If this is indeed the case, and convergent areas are phenomena that are not fundamentally different from any process of borrowing, then the relation between the kind of borrowing that leads to the emergence of areas, and other borrowing, is of interest. Specifically, this notion can be put to the test by comparing the structural categories that are likely to display shared features in a linguistic area, with categories that are considered likely candidates for structural borrowing in cases of one-to-one language contact.

We take a new approach on the issue of areas. We set out to examine whether there is such a thing as a 'prototypical linguistic (convergent) area', in terms of the structural profile, specifically the domains of structure (grammatical categories) in which shared features are likely to be found (e.g. word order, tense-aspect, modality, phonology, clause combining, or morphological typology). In order to investigate this, we examine a sample of recognised areas. For each area, we examine the structural domains in which shared features appear. Structural domains (categories or functions) that re-occur in the sample are considered as having a high 'convergent profile'; they may be said to be a more typical manifestation of a linguistic area than others.

Our sample consists of languages representing around a dozen linguistic areas from different parts of the world, complemented by a sample of languages in contact that are not usually considered to be part of a convergent area. Drawing on a

database of contact- induced features, which we developed as part of the research project on Language Convergence and Linguistic Areas in Manchester, we present our preliminary findings concerning typical profiles of linguistic areas, and categories that are susceptible to borrowing in other contact situations, and address the questions of convergence-prone functions of language, as well as that of the ‘prototypical’ cluster of shared features in a linguistic area.

---

Nick J. Enfield

Max Planck Institute for Psycholinguistics, Nijmegen

**Lessons of areal typology: population thinking for problems of language, cognition, and culture**

Mi 14.15-14.45

---

This paper argues that areal typology yields critical insights not only for linguistics generally, but also for research in both cognitive science and anthropology. The puzzle of areal typology is how and why geographically proximal languages come to share typological structure, independent of ‘genealogy’. The key approach to solving this puzzle is to apply population thinking to our understanding of linguistic structures and processes. Such thinking gives rise to a framework which recognizes four basic units of analysis: 1. linguistic items (words, constructions, etc.); 2. individual speakers and their mental representations (i.e. of individual linguistic items and of structured systems of such items); 3. actual communicative interactions among speakers; 4. populations of speakers in social association. (‘Languages’ are secondary at best as units with any causal role or analytic value.) The properties of these elements in combination impact upon the historical and geographic course of linguistic diffusion, bringing about the kinds of patterns which areal typologists observe. By this conception of language, linguistic processes are necessarily played out in a ‘micro-macro’ nexus. That is, questions of individual linguistic cognition cannot be understood without considering interactions among individuals, and, in turn, emergent population level effects. And questions of higher level convention (properties of ‘languages’) cannot be understood without considering properties of individual cognition, as well as the actual occasions of linguistic interaction among individuals. Population thinking makes this explicit.

I discuss consequences of this kind of understanding for questions in two disciplines closely related to linguistics. First, in cognitive science, a major controversy centers on the nature of linguistic categorization and semantic representation as located in the individual mind, notably on whether such representations are ‘innate’ or ‘constructed’. Areal typology supports a view by which conceptual representations of linguistic categories are constructed based on evidence available almost exclusively in occasions of face-to-face interaction. This has predictions for the kinds of representations which will be constructed given the distinct types of interactional contexts associated with ‘language contact’ situations versus in-group social situations. Second, in much of sociology and social anthropology, there is a tension between the analytic reliance on macro notions such as ‘a culture’ or ‘a society’, on the one hand, and the rejection of social/cultural essentialism, on the other. The paradox can be worked out by adopting the kind of

population thinking which areal typology demands. The trick is to see how individual cognitive representations and specific occasions of communicative interaction feed into the higher level aggregation of convention (as observed in ‘languages’). The required conceptual framework is readily available in research in sociology on the diffusion of behavioral innovation.

The broad message of this paper is that areal typology demands a particular conception of linguistic process and structure (i.e. population thinking), which not only brings general linguistics to a more realistic understanding of language, but can illuminate our understanding of linguistic questions in related disciplines.

---

Søren Wichmann  
Max Planck Institute for Evolutionary Anthropology, Leipzig  
**Pre-neolithic linguistics**  
Mi 14.45-15.15

---

The kind of areal-typological research pioneered by Johanna Nichols has revealed the existence of continent-wide or even cross-continent linguistic areas. ‘Super-areas’ might be a fitting term for this phenomenon. Data gathered under the auspices of the World Atlas of Linguistic Structures project (Haspelmath et al. 2005) has further contributed to the possibility of identifying super-areas. Such areas include Eurasia, the Pacific Rim, the Americas, Sahul, and a zone transversing central Africa. In order to explain the phenomenon, one might consider effects of both linguistic inheritance and language contact, but I would argue that the considerable time depth that such areas presuppose do not allow for a resolution fine enough to distinguish between the two. Moreover, as long as we can identify zones of human interaction and begin to make archaeological and genetic correlation this is already a big step ahead.

The aim of the paper is to identify the major super-areas in the world and to explore the possibility of an association with pre-neolithic contact zones. The logic is that neolithic revolutions around the world would have slowed down the extent of movements of human groups such that it may be inferred that very wide-spread features should date to pre-neolithic times. An obvious question is why the areas are distributed the way they are. It seems to be the case that pre-neolithic population spreads are not inherently directional; the expansion rather seems to continue until it is simply blocked by natural barriers. This pattern is in rather stark contrast to neolithic expansions, which tend to follow latitudinal climatic zones.

In addition to the distribution of typological features, I will draw upon a geographical mapping of linguistic diversity as a clue to prehistory. In the case of the Americas, for instance, such a mapping gives some internal structure to the general picture of a linguistic super-area, suggesting a rapid pre-neolithic north-to-south expansion along the west coast succeeded by slower expansions from west to east, presumably to a large extent confined to neolithic times. The question of what types of features are involved will also be addressed. Is there a difference in nature between typological features that are very widespread and ones that are more restricted? If the most widespread typological features are due to relatively rapid movements of populations they should presumably be easily borrowed, whereas neolithic expansions

inducing stable bilingual situations would allow for the spread of ‘deeper’ grammatical patterns.

Haspelmath, M., Matthew S. Dryer, David Gil, and Bernard Comrie. 2005. *The World Atlas of Language Structures*. Oxford: Oxford University Press.

---

Michael Cysouw  
Max Planck Institute for Evolutionary Anthropology, Leipzig  
**Identifying linguistic areas from typological surveys**  
Mi 15.15-15.45

---

Already for some time, typologists are finding large areal consistency in the world-wide distribution of linguistic characteristics. With the appearance of the World Atlas of Language Structures (Haspelmath et al. 2005) such areal patterns can now clearly be shown. However, the human eye is not a very trustworthy source for the identification of areal patterns. I will discuss some possibilities to analyze geographical patterns more consistently. When geographical patterns are found, the next question is how to analyze them. I will argue that it is possible to differentiate between patterns of long-term static contact and patterns of recent spread (either because of genealogical expansion or political dominance). When there is a strong correlation between the geographical distance and the linguistic diversity, then this is a sign of long-term static contact. More dynamic histories will be characterized by a mismatch between the geographical and the linguistic diversity.

Haspelmath, Martin, Matthew S. Dryer, David Gil & Bernard Comrie [eds.], 2005. *The world atlas of language structures*. Oxford: Oxford University Press. (<http://www.wals.info>)

---

Thomas Stolz  
Universität Bremen  
**The European North-West - an Area with Non-European Features**  
Mi 16.15-16.45

---

Recent research on the areal linguistics of Europe has shown that there is a distribution of features which reflects a centre-periphery pattern. Languages situated in the western central part of the continent participate in numerous isoglosses whereas languages outside this region share a decreasing number of features with the languages in the core of the linguistic area. Languages which typically fail to side with the so-called SAE-languages are the ones spoken in the remote north western part of the continent stretching from Scandinavia to Ireland. Icelandic, Faroese, Norwegian and other mainland Scandinavian languages join the members of the Goidelic branch of the Celtic phylum not only in the sense that they do not display certain features which are characteristic of other European languages. What is more,

the North Germanic languages and the Celtic languages have a number of properties in common which set them apart from the bulk of the languages of Europe. In a way, the languages in the North- West behave more like languages spoken outside of Europe. Some similarities between North Germanic and Celtic were already discussed in the 1940-1960ies by scholars like Harald Wagner and others. What makes these similarities especially interesting for areal linguistics is the fact that they cannot be attributed to the common Indo-European ancestry of the languages, meaning the shared features have not been inherited from a common proto-stage but developed much later. However, not all of the ideas expressed in the earlier literature are fit to stand the test according to the standards of modern areal linguistics.

In my talk I will present a new piece of evidence for a existence of a North-West European area - a feature which so far has not gained sufficient attention by areal linguists. North Germanic languages (especially the two insular varieties Icelandic and Faroese) and Celtic languages converge in the realm of possession as they display similar possession-splits based on an alienability-like correlation. These distinctions are systematic and next to obligatory (with obligatoriness diminishing on the Scandinavian mainland). Grammaticalised possession-splits of this kind are exceptional in Europe. According to Haspelmath (2001) the absence of alienability-distinctions is a (negative) trait of the SAE-languages. The talk will sketch the grammar of possession in the languages of the European North-West. It will also focus on the question of how and when these possession-splits have come about. Moreover, I will discuss the crucial problem of whence the phenomenon has come. For this purpose, the oldest available texts of the languages under scrutiny are checked. In addition, cultural and factual history of the region of Europe, the migrations and the demographic development will be given due attention.

Haspelmath, Martin (2001): The European linguistic are: Standard Average European. In: Haspelmath, Martin et al., *Language Typology and Language Universals. An International Handbook*. Vol. 2. Berlin: Walter de Gruyter, 1492-1510.

Wagner, Harald (1964): "Nordeuropäische Lautgeographie", in: Beneš, Birgit et al. (2002), *Heinrich Wagner. Beiträge zur typologischen Sprachgeographie*. Bern: Peter Lang.

---

Johannes Helmbrecht

Universität Erfurt

**On the development of Standard Average European (SAE) - the case of polite pronouns**

Mi 16.45-17.15

---

Recent research on Europe as a large linguistic area (cf. for instance the papers in Bechert et al. 1990; Haspelmath 1998, 2001; van der Auwera 1998; and some others) identified several grammatical features that are rather rare on a global perspective but sufficiently frequent in European languages in order to qualify as criteria for a linguistic area. These criteria include the existence of definite and indefinite articles, relative clauses with relative pronouns, the so-called have-perfect, nominative marking of experiencer subjects, passive constructions with passive participles, the

dominance of anti-causatives and a few others. Surprisingly, none of the publications in this research mentioned polite pronouns as another areal feature of Europe.

It is the goal of the proposed paper to show that politeness distinctions in personal pronouns like German *du/ Sie* and French *tu/ vous* are another linguistic criterion for Standard Average European. In order to prove this a combination of typological, areal-typological and historical methods of research are employed. This includes that the various types of politeness distinctions in personal pronouns are distinguished and mapped with regard to their geographical distribution in Europe and the world. In addition, the historical development of the politeness types from the origin in the late phase of Latin until the 20th century will be reconstructed.

The results demonstrate that there are two dimensions of diffusion involved. One dimension describes the geographical spread of polite pronouns from the territory of today's France and Germany to the borders of Europe. The other dimension describes the social spread of polite pronoun usage from the royal courts and the aristocracy to the bourgeoisie and the lower classes. Three waves of a historical spread can be discerned and only the first one with the greatest historical depth reached the borders of the area. The conclusions are: politeness distinctions are an areal feature of SAE. Their emergence and historical development can be determined with more precision than for any of the other features of SAE mentioned above. The time of the first appearance and the geographical center from where the innovations started to spread coincides with what Van der Auwera (1998) called the "Charlemagne Sprachbund".

- Bechert, Johannes & Bernini, Giuliano & Buridant, Claude (eds.) 1990. *Towards a Typology of European Languages* (Empirical Approaches to Language Typology 8). Berlin: Mouton de Gruyter.
- Haspelmath, Martin 1998. "How young is Standard Average European?". In: *Language Sciences* 20:271-87.
- Haspelmath, Martin 2001. "The European linguistic area: Standard Average European." In: Haspelmath, Martin & König, Ekkehard & Oesterreicher, Wulf & Raible, Wolfgang (eds.) *Language typology and language universals. (Handbücher zur Sprach- und Kommunikationswissenschaft)* Berlin: de Gruyter, 1492-1510.
- Helmbrecht, Johannes 2005a. "Typologie und Diffusion von Höflichkeitspronomina in Europa." In: Lehmann, Christian (ed.) *Arbeitspapiere des Seminars für Sprachwissenschaft der Universität Erfurt* (ASSidUE) Nr. 18. Erfurt: University of Erfurt. Available as a PDF file: <http://www.db-thueringen.de/servlets/DerivateServlet/Derivate-6136/ASSidUE18.pdf>
- Helmbrecht, Johannes. 2005b. "Politeness Distinctions in Personal Pronouns." In: Martin Haspelmath, Matthew S. Dryer, David Gil, Bernard Comrie (eds.) *The World Atlas of Language Structures*. Oxford: Oxford University Press, 186-90.
- van der Auwera, Johann 1998. "Conclusion". In: van der Auwera, Johann (ed.) *Adverbial Constructions in the Languages of Europe*. (Empirical Approaches to Language Typology - EUROTYP 20-3). Berlin: Mouton de Gruyter, 813-36.

---

Anna Giacalone, Andrea Sansò  
Università di Pavia

**The demise of indefinite *man* constructions in European languages: An areal perspective**

Mi 17.15-17.45

---

Indefinite *man* constructions (also known as impersonal *man* constructions, ultra-indefinites [Koenig 1999], *man-Sätze*, ‘generic’ subject constructions, etc.) are a well-attested feature of a number of Standard Average European languages, while being quite rare across languages (the construction is attested in 23 languages out of 100 in Haspelmath’s 1997 sample, cf. Haspelmath 1997: 324ff.). In some languages (e.g. French) this construction type appears to be fully grammaticalized (with “man” patterning like a full pronoun), whereas in other languages “man” is still syntactically a noun phrase. More importantly, in the history of single languages the range of uses of such constructions tends to contract, and they eventually disappear, as is the case in English (Los 2005: Ch. 5), Italian (Kontzi 1958: 100-103; Salvi n.d.; Brambilla Ageno 1964: 168; Barrett Brown 1936) and Italian varieties, Catalan (Kaufmann 2002: 2173), and Spanish (Barrett Brown 1931).

Different hypotheses have been formulated in order to explain the retraction of such constructions. Some of them are of a structural, “systemic” nature: Los (2005), for instance, supposes that the demise of *man* was significantly aided by the loss of  $V_2$  in Middle English. Other scenarios explicitly invoke the role of language contact in both the emergence and the demise of these constructions: Harris (1978: 191), for instance, suggests a Frankish influence on the rise of this construction in French; Kontzi (1958: 101) explains the rise of this construction in Ancient Italian by postulating “fremde Einflüsse ... aus der französischen Übersetzungsliteratur und aus der provenzalischen Lyrik”. Other pieces of evidence point towards a multiple genesis of this construction in different varieties: it is in fact attested in some Central and Southern Italian varieties (Romanesco; Abruzzese: Rohlf’s 1949: 272ff.; D’Alessandro and Alexiadou to appear) for which the hypothesis of a French influence is implausible, if tenable at all.

In this paper we explore the diffusion and retraction of *man* constructions in European languages in a systematic way, with special focus on Romance, Germanic, and Slavonic languages. The results of our investigation confirm the idea that the impersonal *man* construction has spread throughout Europe through language contact, but, unlike other Standard Average European features, it is a “recessive” construction, whose vital cycle has come to an end quite early in the Middle Ages.

Barrett Brown, Ch. 1931. „The disappearance of the indefinite *hombre* from Spanish“. *Language* 7: 265-277.

Barrett Brown, Ch. 1936. „*Uomo* as an indeterminate pronoun“. *Language* 12: 35-44.

Brambilla Ageno, F. 1964. *Il verbo nell’italiano antico*. Milano-Napoli: Ricciardi.

D’Alessandro R., and Alexiadou, A. To appear. „The syntax of the indefinite pronoun *nome*“. *Probus*.

Harris, M. 1978. *The evolution of French syntax*. London: Longman.

Haspelmath, M. 1997. *Indefinite Pronouns*. Oxford: Clarendon Press.

- Kaufmann, A.B. 2002. „La passiva i les construccions que s’hi relacionen“. In: J. Solà, M.-R. Lloret, J. Mascaró, M. Pérez Saldanya (hg.), *Gramàtica del català contemporani*, Vol. 2: *Sintaxi*. Barcelona: Empúries, 2111-2179.
- Koenig, J.P. 1999. „On a tué le président! Ultra-indefinites and the nature of passives“. In: B. Fox, D. Jurafsky, and L. Michaelis (hg.), *Cognition and Function in Language*. Stanford: CSLI Publications, 235-251.
- Kontzi, R. 1958. *Der Ausdruck der Passividee im älteren italienischen*. Tübingen: Max Niemeyer Verlag.
- Los, B. 2005. *The Rise of the To-Infinitive*. Oxford: Oxford University Press.
- Rohlf, G. 1949. *Historische Grammatik der Italienischen Sprache und ihrer Mundarten*, Bd. II: *Formenlehre und Syntax*. Bern: Francke Verlag.
- Salvi, G. n.d. La realizzazione sintattica della struttura argomentale. Manuscript. Progetto *ItalAnt* – *Grammatica dell’italiano antico*: <http://ludens.elte.hu/~gps/konyv/indice.html>.
- 

Hans-Jörg Bibiko  
Max Planck Institute for Evolutionary Anthropology  
**Visualizations of geographical patterns**  
Mi 17.45-18.15

---

The WALs (The World Atlas of Language Structures; Haspelmath et al. 2005) is a large database of structural (phonological, grammatical, lexical) properties of languages gathered from descriptive materials (such as reference grammars) by a team of more than 40 authors (many of them the leading authorities on the subject). Additionally with this book a computer program ‘The Interactive Reference Tool’ is released which allows to view the maps in a variety of different forms, as well as to combine features, i.e. to generate compound features and to display these as well. By using this program to visualize the world-wide distribution of linguistic phenomena it turned out that such tools are very helpful to understand the issues of areal patterns.

In my presentation I want to show different solutions to prepare such maps by means of different approaches such as by using the Interactive Reference Tool of WALs, by using Filemaker’s native possibilities, by using the SVG (scalable vector graphics) markup language, and finally by using a work in progress solution called WALs3D (a 3D modeling software) which uses real satellite images of the world in conjunction with height field data, areal data about e.g. botany, population density, soil conditions etc. WALs3D could help to find maybe possible correlations between language structures and their distributions in respect of geography.

Haspelmath, Martin, Matthew S. Dryer, David Gil & Bernard Comrie [eds.], 2005. *The world atlas of language structures*. Oxford: Oxford University Press. (<http://www.wals.info>)  
SVG - scalable vector graphics. <http://www.w3.org/Graphics/SVG/>



---

Don Stilo

Max Planck Institute for Evolutionary Anthropology, Leipzig; Swedish  
Collegium for Advanced Study in the Social Sciences

**Can Linguistics and Genetics Work Together to Unravel the Pre-history  
of the Peopling of the Iranian Caspian Seacoast?**

Do 9.30-10.00

---

According to very recent data from the Genetics Department of the Max Planck Institute for Evolutionary Anthropology, analyses of DNA samples of the Iranian populations (Gilaki, Mazanderani) of the southern Caspian littoral (Nasidze et al., to appear) indicate a rather unexpected pattern in the history of this area. The mtDNA of these populations indicate that the female line is wholly consonant with local Iranian populations south of the Caspian. The Y-Chromosomes of these two populations, however, have yielded a surprise result in the male line: their molecular biological analyses show that they derive from the south Caucasus. Historical data allude to no such population movements that would yield this mixed population type. Where did these men who introduced this male line derive from and at what historical period: Azerbaijan (Turkic; pre-Turkic; pre- or post-Christian times), Armenia (pre-Indo-European; Indo-Europeanized Armenians; pre-Christian Zoroastrian or post-Christian times), Georgia? Are there any linguistic features that may help unravel this mystery?

The linguistic patterns of modern Gilaki and Mazanderani (Caspian subgroup of Northwest Iranian languages) show typological features that are inconsistent with the features of their Iranian relatives to the south, particularly Persian (Southwest Iranian). On the other hand, Stilo [2005] shows that the typically “OV-features” of these Caspian languages (SOV, N-postposition, gen-N, adj-N, etc., all of which are the opposite of Persian) are completely consistent with the typology of Caucasian and Turkic languages. While it is tempting to use implicational universals (OV, dem-N, adv-VB, sub-MAIN CL, etc.) to help identify a Caucasian strain in the history of the Caspian, they in fact cannot serve these purposes since they are also typical of all the languages of *an unbroken chain* extending from Turkey through the Caspian to Korea and from Finland to Alaska (excepting intrusive Russian). That is, these features are shared not only by the languages of the Iranian Caspian and the south Caucasus (including Armenian, Kartvelian, Daghestanian), but also by (almost) all Turkic languages, all Uralic languages (with some exceptions in the extreme western Uralic zone), by Ket, Mongolic, Tungusic, Chinese (for the most part), and by all the “Paleo-Asiatic” languages. This set of isoglosses continues on into North America. My presentation will include an isogloss mapping of these common features across northern Eurasia.

Since these major patterns are of no direct help, we must seek evidence in minor features of phonology and grammar to help us determine any Caucasian influence and, if possible, specifically from what part of Caucasus. From this list, we must identify (and set aside) any features that could simply be universal patterns or tendencies. I hope to discuss some linguistic features shared by the Iranian Caspian and the south Caucasus as potential candidates in certain parallelisms in verbal aspect formations, typology of reflexives, etc.

- Nasidze, Ivan, et al. to appear. "Mother Tongue: Concomitant Replacement of Language and MtDNA of South Caspian Populations of Iran"
- Stilo, Donald, 2004. "Iranian as Buffer Zone Between the Universal Typologies of Turkic and Semitic," [in] Éva Ágnes Csató, Bo Isaksson, Carina Jahani (eds.), *Linguistic Convergence and Areal Diffusion: Case studies from Iranian, Semitic and Turkic*, London: RoutledgeCurzon, pp. 35-63.
- 

Brigitte Pakendorf  
Max Planck Institute for Evolutionary Anthropology, Leipzig  
**Languages and Genes in Siberia**  
Do 10.00-10.30

---

Siberian languages share some typological features such as (S)OV word order, predominantly suffixing and often agglutinative morphology, a large number of nominal cases (including distinctive Comitative vs. Instrumental, and Allative vs. Dative cases), head-marking possessive suffixes, and a lack of gender marking (Comrie 1981, Anderson 2004). Although these features are not individually rare world-wide, the combination indicates the existence of a linguistic area. A notable exception within this area is Ket, a linguistic isolate spoken along the Yenisey river, and to a lesser degree Nivkh, an isolate spoken at the Amur river mouth and on Sakhalin island.

From a molecular-genetic perspective, Siberian populations are characterized by some shared polymorphisms as well, such as Tat C or M86 on the Y-chromosome, or the mitochondrial haplogroups C and D (Karefet et al 2002, Starikovskaja et al. 2005, Derenko et al. 2003).

This paper examines the degree to which the linguistic and the genetic areas overlap, and whether there is evidence for shared prehistory/admixture in the region. Furthermore, the question of whether the linguistic isolates Ket and Nivkh are also genetically isolated from their neighbours will be investigated.

- Anderson, G. D. S. 2004. "The languages of Central Siberia: Introduction and overview." In: E. Vajda (Ed) *Languages and Prehistory of Central Siberia*. Amsterdam: John Benjamins Publishing Company, 1-119.
- Comrie, B. 1981. *The languages of the Soviet Union*. Cambridge et al.: Cambridge University Press.
- Derenko, M.V.; Grzybowski, T.; Malyarchuk, B.A.; Dambueva, I.K.; Denisova, G.A.; Czarny, J.; Dorzhu, C.M.; Kakkakov, V.T.; Miscicka-Sliwka, D.; Wozniak, M.; Zakharov, I.A. 2003. Diversity of mitochondrial DNA lineages in South Siberia. *Ann Hum Genet* 67/5, 391-411.
- Karafet, T.M.; Osipova, L.P.; Gubina, M.A.; Posukh, O.L.; Zegura, S.L.; Hammer, M.F. 2002. High levels of Y-chromosome differentiation among native Siberian populations and the genetic signature of a boreal hunter-gatherer way of life. *Hum Biol* 74/6, 761-789.
- Starikovskaya, E.B.; Sukernik, R.I.; Derbeneva, O.A.; Volodko, N.V.; Ruiz-Pesini, E.; Torroni, A.; Brown, M.D.; Lott, M.T.; Hosseini, S.H.; Huoponen, K.; Wallace, D.C. 2005. Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. *Ann Hum Genet* 69/1, 67-89.

---

Balthasar Bickel  
University of Leipzig  
**Statistical Issues in Areal Typology**  
Do 10.30-11.00

---

Linguistic areality is not a categorical but a gradual phenomenon, and like all gradual phenomena it requires quantitative analysis and statistical assessment. In this paper I address a number of methodological issues that arise during statistical testing of linguistic area hypotheses, illustrated by recent research on the Eurasian and Circum-Pacific macro-areas (Bickel & Nichols 2003, 2005a, 2005b, 2005c, 2006).

Areality presupposes that observed similarities between languages are not induced by universal preferences or genealogical relationship; therefore, universals and genealogy are the key variables that need to be controlled for when testing hypotheses about areas. Universal preferences can usually be factored out by world-wide matching samples. Genealogical inheritance effects cannot be distinguished from areal effects if they are old and result from deep, non-demonstrable genealogy. (I propose the term ‘historical effect’ for all effects that are induced by contact or non-demonstrable inheritance in such a non-distinguishable way.)

But more shallow inheritance effects, resulting from genealogy that is demonstrable by the Comparative Method, can and should be controlled for. The standard strategy for such control in statistics is sample stratification so that each genealogical unit (stock, genus) is evenly but randomly represented. Because many genealogical units have only one representative (isolates) and for many only one is described, random sampling within strata is impossible in practice. The alternative to this, proposed by Dryer (1989, 2000), is genus-based (or stock-based) sampling, where each datapoint represents the whole genus rather than a language within that genus (or stock). Genus-based samples ideally contain all genera within the areas of interest. Janssen, Bickel & Zúñiga (2005) have shown that this sampling technique leads to exhaustive samples, which are logically incompatible with classical statistics. In response, they propose and adapt randomization-based (Monte-Carlo) and exact statistics for typology. I will showcase the application of these methods for scalar and count variables.

- Bickel, Balthasar & Johanna Nichols (2003). Typological enclaves. Paper presented at the 5th Biannual Conference of the Association for Linguistic Typology, Cagliari, September 18.
- Bickel, Balthasar & Johanna Nichols (2005a). Areal patterns in the World Atlas of Language Structures. Paper presented at the 6th Biannual Conference of the Association for Linguistic Typology, Padang, July 24.
- Bickel, Balthasar & Johanna Nichols (2005b). Inclusive/exclusive as person vs. number categories worldwide. In Elena Filimonova (ed.) *Clusivity*, 47 – 70. Amsterdam: Benjamins.
- Bickel, Balthasar & Johanna Nichols (2005c). Inflectional synthesis of the verb. In Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie (eds.) *World atlas of language structures*. Oxford: Oxford University Press.
- Bickel, Balthasar & Johanna Nichols (2006). Oceania and the Pacific Rim linguistic area. *Proceedings of the 32nd Annual Meeting of the Berkeley Linguistics Society*.

- Dryer, Matthew S. (1989). Large linguistic areas and language sampling. *Studies in Language* 13, 257 – 292.
- Dryer, Matthew S. (2000). Counting genera vs. counting languages. *Linguistic Typology* 4, 334 – 350.
- Janssen, Dirk, Balthasar Bickel & Fernando Zúñiga (2005). Randomization tests in language typology. Ms. University of Leipzig ([www.uni-leipzig.de/~bickel/research/papers](http://www.uni-leipzig.de/~bickel/research/papers)).

---

Kristine A. Hildebrandt  
University of Manchester  
**The Areal and Genealogical Dimensions of Hiatus**  
Do 11.30-12.00

---

The goal of Optimality Theory is to account for language variation, thereby revealing universal (violable) constraints and their cross-linguistic rankings. The FAITHFULNESS constraints (MAX and DEP) have recently been discussed within the context of Positionally Sensitive Faithfulness (PSF), i.e. certain (morphological, semantic, phonological) prominent positions maintain phonological contrasts more readily (McCarthy & Prince 1995). One specific account of PSF is a cross-linguistic analysis of Hiatus phenomena (Casali 1995, 1997).

In Hiatus, two or more vowels become adjacent via morphological or syntactic concatenation, as in Persian (Indo-European, Mahootian 1997) *mi-xa=æm* > [*mi.xam*] (DUR-want=1SG) ‘I want’. Here, the V-V adjacency across the root+enclitic is resolved via elision of the enclitic vowel, with no subtractive process affecting the semantically prominent root. Across languages, hiatus environments may stay unresolved (heterosyllabified). It is claimed though that many languages do not tolerate Hiatus (Casali 1996; 1997), with resolutions including vowel elision, consonant glide insertion, glide formation, and coalescence.

In a sample of largely Niger-Congo languages, Casali (1997) maintains that several scenarios of vowel elision are frequently attested and are guided by PSF principles. Specifically, the choice of which vowel is elided is frequently determined by the relatively active OT constraint MAXLEX, where there is a greater tendency to maintain the phonological material in elements that encode greater semantic content (i.e. in root/lexical morphemes). For example, in the Persian example above the elision targets the enclitic only, with no subtractive process to the root.

This study exhaustively examines Hiatus phenomena in a sample of 41 languages. While many languages are Sino-Tibetan and Indo-European (Eurasia), other stocks/areas are also represented. A survey of: (a) (potential) Hiatus types and resolutions and (b) the material targeted by the resolution reveals some interesting patterns.

1. In the larger sample, Hiatus is just as likely to be tolerated as resolved, with 54% of languages resolving Hiatus in some way. But, in European languages, the degree and variety of Hiatus resolutions is significantly greater than for S/SEA languages (Fisher’s Exact  $p = .005$ ). Also, where Hiatus is resolved via vowel elision, there is no significant trend for respecting MAXLEX or MAXWD (i.e. no significant preference to preserve material in roots or word-initially).

2. Looking at family (I-E vs. S-T) and area (Europe vs. S/SEA), Hiatus is resolved via vowel ELISION significantly more in I-E & European languages than in S-T & S/SEA ones ( $p = .005$ ). In S-T & S/SEA, elision is significantly DISPREFERRED to other resolution strategies (insertion or alternation).

3. In I-E/European languages, where elision is frequently attested, there is also no significant pattern of vowel elision indicating a respect for MAXLEX or MAXWD (i.e. elision does not significantly favor retention of material in root or word-initial position).

4. For other resolutions besides elision, morpheme alternations and segment insertion are almost exclusively properties of I-E & European languages; morpheme alternations are infrequent in S-T & S/SEA languages. Rather, consonant insertion is the significantly preferred strategy.

Early findings suggest that the degree to which hiatus is tolerated and the possible resolutions depend, among other things, on the degree of coherence that the targeted morphemes share within a context of the overall prosodic domain profile of the language. A consideration of this in tandem with the overall prosodic domain profiles of language families and areas, suggests that while there may be a bias towards preserving word-initial contrasts (MAXWD) in resolution scenarios, the word boundaries, and their locations in morpho-syntactic strings, will vary genealogically and geographically.

Casali, R. 1996. Resolving Hiatus. ROA 215-0997, Rutgers Optimality Archive, <http://roa.rutgers.edu/>

Casali, R. 1997. Vowel elision in hiatus contexts: Which vowel goes? *Language* 73.3 493-533.

Maahotian, S. 1997. *Persian*. London: Routledge

McCarthy, J., and A. Prince. 1995. Faithfulness and reduplicative identity. *Papers in optimality theory*, ed. by J. N. Beckman, L. Walsh, and S. Urbanczyk, 249-384. (University of Massachusetts occasional papers in linguistics 18). Amherst, MA: Graduate Linguistics Students Association.

Prashant Pardeshi, Peter Hook, Shinji Ido, Hajar Babai, Dileep Chandralal, Kaoru Horie, Joungmin Kim, Omkar N. Koul, Qing-Mei Li, Hsin-hsin Liang, Colin P. Masica, Yutaro Murakami, Kingkarn Thepkanjana  
Kobe University, University of Virginia

**The areal distribution and semantic evolution of EAT-expressions in the languages of asia**  
Do 12.00-12.30

Many of the languages of Asia have families of expressions in which a verb meaning 'eat' takes as direct object a noun expressing an action or process. Formally active, these expressions sometimes have meanings roughly equivalent to English 'suffer' or 'undergo':

- |     |  |            |             |                |            |                   |            |
|-----|--|------------|-------------|----------------|------------|-------------------|------------|
| (1) | <i>tuu</i>                             | <i>aaj</i> | <i>maar</i> | <i>khaa-ne</i> | <i>par</i> | <i>lagii.huii</i> | <i>hai</i> |
|     | you                                    | today      | beating     | EAT-INF        | on         | keen              | are        |
|     | 'You are looking for a beating today!' |            |             |                |            |                   |            |
| (2) | <i>đán</i>                             | <i>ăn</i>  | <i>đòn</i>  | <i>lám</i>     |            |                   |            |

worthy EAT thrashing well  
'(He) needs a good beating.'

In other cases they may have meanings of exhibition of an attitude or presence of a state:

(3) *ti* *khup-ats* *bhaav* *khaa.te*  
she much-EMPH price EAT  
'She thinks very highly of herself.'

(4) *xué* *diàn.nǎo* *hěn* *chī* *xiāng*  
study computer very EAT fragrance  
'Computer science is very hot (now).'

Although EAT-expressions are encountered across Asia in nearly every language family, their types, numbers, and semantic range vary. Regarding their distribution as an instance of areal diffusion our paper lists and categorizes EAT-expressions in selected languages of West Asia (Arabic, Farsi, Turkish), South Asia (Hindi-Urdu, Marathi, Kashmiri, Telugu), Southeast Asia (Thai, Vietnamese), East Asia (Chinese, Japanese, Korean) and Central Asia (Uzbek, Tajik). From uniform datasets, as in (5), comparisons are made with a view to determining geographical origins and the paths of semantic evolution.

(5)	<i>(men)</i>	<i>ota.m</i>	<i>dan</i>	<i>gap</i>	<i>ye.dim</i>
	<i>māi.ne</i>	<i>pitaajii</i>	<i>se</i>	<i>DāāTh</i>	<i>khaa.i</i>
	<i>(nae.ka)</i>	<i>ebeci</i>	<i>hanthey</i>	<i>yok</i>	<i>mek.ess.ta</i>
	<i>(boku.wa)</i>	<i>tittoya</i>	<i>no</i>	<i>kogoto-wo</i>	<i>kut.ta</i>
	<i>*(mæn)</i>	<i>æz</i>	<i>baba.m</i>	<i>tobikh</i>	<i>khor.dæ.m</i>
	<i>*wǒ</i>	<i>bèi</i>	<i>bàba</i>	<i>chī.le</i>	<i>chì.mà</i>
	<i>*ja</i>	<i>ot</i>	<i>otca</i>	<i>s'e.l</i>	<i>bran'</i>
	I	father	ATE	scolding-ACC	ATE

'(I) got a scolding from Dad.'

Preliminary results from listing and mapping EAT-expressions in Farsi, Marathi, Hindi-Urdu, Uzbek, Russian, Korean, Mandarin, Vietnamese, and Thai, indicate that they have a Central Asian origin and that they have spread into Southwest, South, Southeast and East Asia in a pattern more uniformly annular than that found by Masica for the distribution of other typological features defining the Indo-Turanian linguistic area in his pioneering 1976 study. This finding is not unexpected given the greater salience in the speaker's consciousness of idioms and turns of phrase (when compared to structural phenomena like word order, dative "subjects", and the morphological causative). It also suggests that the more lexical features of linguistic areas (such as compound verbs) should travel across both genetic lines and bundles of typological isoglosses more easily than other, more abstract grammatical features do.

Masica, Colin P. 1976. *Defining a linguistic area: South Asia*. Chicago: University of Chicago Press.

---

Tom Güldemann  
Max Planck Institute for Evolutionary Anthropology, Leipzig  
**Towards the synchronic macro-areal profile of Africa**  
Fr 10.30-11.00

---

The last studies concerned with macro-areal linguistics on the African continent like Greenberg (1959, 1983) and Heine (1976) are now more than 20 years old. Since then, no new attempts have been published in which the major entities of areal typology in Africa are identified and empirically substantiated. To a considerable extent, this is related to the so far unproven hypothesis that African languages can be classified into just four language families. The present paper tries to give a new survey of linguistic macro-areas in Africa under the assumption that some of these alleged families can be viewed alternatively as potential areal units.

Since a number of African languages and language groups are still deficiently documented, such a study cannot yet be based on a fully representative survey of sufficiently many linguistic features. I will argue in this paper that a promising way out of this dilemma is to concentrate at the beginning on features that are typologically highly marked and at the same time susceptible to transfer by contact. A number of such features - ranging from speech sounds like clicks and labial velars to syntactic patterns like logophoricity or certain word order types - turn out to be good diagnostics for linguistic macro-areas on the continent. Provided the present results will be substantiated by more extensive data, the strategy to start areal surveys by focusing on "quirky" linguistic properties promises to yield fruitful results in other continental areas of the world.

- Greenberg, Joseph H. 1959. "Africa as a linguistic area". In: William R. Bascom and Melville J. Herskovitz (eds.), *Continuity and change in African cultures*. Chicago: Chicago University Press, 15-27.
- Greenberg, Joseph H. 1983. "Some areal characteristics of African languages". In: Dihoff, Ivan R. (ed.), *Current approaches to African linguistics 1* (= Publications in African Languages and Linguistics 1). Dordrecht/ Cinnaminson: Foris, 3-22.
- Heine, Bernd. 1976. *A typology of African languages based on the order of meaningful elements* (= Kölner Beiträge zur Afrikanistik 4). Berlin: Dietrich Reimer.

---

Mily Crevels, Pieter Muysken und Hein van der Voort  
Radboud Universiteit Nijmegen  
**The Guaporé-Mamoré region: a linguistic area?**  
Fr 11.00-11.30

---

The Guaporé and Mamoré are two great rivers of the Southwestern Amazon region. The Guaporé river forms the border between the Brazilian federal state Rondonia and the Bolivian departments Santa Cruz and Beni, where the river is called Iténez. The Guaporé and Mamoré rivers together drain a part of the tropical lowlands where traditionally over 50 different indigenous languages are spoken. Since these languages represent numerous language families and unclassified languages that may be isolates, it is one of the linguistically most diverse regions of the world. This linguistic diversity, which includes, besides languages from the Arawakan, Chapacuran, Jabutian, Nambikwaran, Panoan, Tacanan, and Tupian families no less than 11 unclassified languages, is highly threatened with extinction. The speakers' numbers have dwindled under the ecological, physical, social and Western cultural pressures of the national societies. Nowadays more than half of the languages have less than 50 speakers and one third less than 10 speakers.

Recent archaeological, genetic and ethno-botanical research suggests that this is one of the core diffusion areas on the South American continent. The existing classifications of the Guaporé-Mamoré languages are often based on thin evidence, since the languages used to lack thorough professional documentation and description. Fortunately, the past decennium has seen a considerable increase in the number of linguists conducting fieldwork in the region, producing grammars and dictionaries of the different languages. Several initiatives (e.g. those of Muysken and of Wetzels in The Netherlands, Moore in Brazil, Dixon and Aikhenvald in Australia) have led to long-term descriptive projects on Guaporé-Mamoré languages, and although many preliminary results have been attained, our ideas about the relationships between the languages of the region still remain in the exploratory stage. It appears that, apart from new insights about certain genetic relationships, the accumulated data also point to relationships of linguistic diffusion. In our presentation, we will show that the Guaporé-Mamoré region is a complex of areal clusters rather than a linguistic area in the traditional sense. After reviewing comparative lexical data and formal correspondences of grammatical elements, we turn to an overview of 25 grammatical features for the languages involved which lead to the cluster analysis. We will try to link this analysis to ethno-historical facts.



---

Dmitry Idiatov

University of Antwerp

**The identity of WHO? and WHAT? in South America: an areal feature?**

Fr 11.30-12.00

---

It is generally assumed that the distinction between the interrogative pro-nouns WHO? (an interrogative pro-noun for the ontological category PERSON) and WHAT? (an interrogative pro-noun for the ontological category THING) is basic to natural languages (see, e.g., Wierzbicka 1996: 38-39). However, in a world-wide convenience sample of some 500 languages, around 70 languages were found that do not oblige their speakers to distinguish WHO? from WHAT?. (This number does not include the well-known cases of partial formal overlap of the paradigms of WHO? and WHAT? of the type found in Latin or Tsez). Almost half of these languages are spoken in South America and a third in Australia. This distribution indicates a certain areal bias on a world-wide scale.

In South America, it appears possible to further reduce this areal pattern to a region that can be broadly defined as the Amazonia-Andes border. Interestingly, an overwhelming majority of the South American languages with identity of WHO? and WHAT? in my sample belong to only two families, the Arawakan (Maipuran) family and the Mataco-Guaicuru family. The areal distribution of the rest of the South American languages with one form for both WHO? and WHAT? and the fact that they typically represent isolated instances within their respective families seem to point to Arawakan languages as the main factor responsible for the attested areal patterning within South America. This hypothesis will be considered against the background of what is known at present about the history of dispersion of Arawak-speaking societies, which are believed to have been involved in extensive trade and intermarriage relations with other groups.

Finally, I will argue that the explanation for the development of identity of WHO? and WHAT? in Arawakan and Mataco-Guaicuru languages should be sought in the lexicalization patterns of the interrogative pro-nouns in these languages, in the peculiarities of their gender systems and gender assignment (first of all, in the ways they treat inanimate referents) and in the predominance of split ergative alignment (most often, of an active-stative type) in these languages.

Wierzbicka, A. 1996. *Semantics: Primes and Universals*. Oxford: Oxford University Press.

## AG 5 Parenthese/Parenthetical constructions

Koordination: Nicole Dehé, Yordanka Kavalova

### Programm

Mittwoch 22.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
13.45-14.15	Nicole Dehé, Yordanka Kavalova	<i>Introduction to parentheticals</i>
14.15-14.45	Franz Josef d'Avis	<i>How close is enough - parentheticals as adjuncts</i>
14.45-15.15	Gunther Kaltenböck	<i>Parenthetical clauses in spoken English</i>
15.15-15.45	Stefan Schneider	<i>Reduced parenthetical clauses as mitigators: an overview</i>
15.45-16.30	Kaffeepause	
16.15-17.15	Markus Steinbach	<i>Integrated parentheticals and assertional complements</i>
17.15-18.15	Tanja Kiziak	<i>Extraction or parenthesis? Testing competing accounts.</i>

Donnerstag 23.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
09.00-10.00	Francesca Del Gobbo	<i>On the syntax and semantics of appositive relative clauses</i>
10.00-10.30	Anita Fetzer	<i>The conjunctive function of disjuncts</i>
10.30-11.00	Yordanka Kavalova	<i>Parenthetical clauses</i>
11.00-11.30	Kaffeepause	
11.30-12.00	Mark de Vries	<i>Parentheses as B-merged adverbial phrases without scope</i>
12.00-12.30	Christian Fortmann	<i>On (clausal) parentheticals</i>

Freitag 24.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
10.30-11.30	Jörg Peters	<i>Syntactic and prosodic parenthesis</i>
11.30-12.00	Rick Nouwen	<i>Conventional implicatures and dynamic binding</i>
12.00-12.30	Rudy Loock	<i>Parentheticals vs. appositive relative clauses in English discourse</i>
12.30-13.00	Sandra Döring	<i>Parentheticals in modern German. Prosodic aspects.</i>
13.00-13.30	Nicole Dehé	<i>Prosodic aspects of parentheticals in English</i>

---

Franz Josef d'Avis  
Universität Göteborg  
**How Close is enough - Parentheticals as Adjuncts**  
Mi 14.15-14.45

---

Parentheticals are -in common understanding- expressions that are inserted into another clause, and in the most cases, they seem to be structurally independent from that clause, s.(1).

(1) Dieser Wagen -ich habe schon viele davon verkauft- ist wirklich billig.

*'this car -I have already many oft-it sold- is really cheap'*

In spoken language, parentheticals are marked by a special intonation: (i) clear breaks before and after, (ii) spoken faster or slower than the main clause, (iii) lower pitch of the voice. The intonation contour of the main clause is continued after the end of the parenthetical. Intuitively it is obvious what parentheticals are: Expressions, ranging from single words to complex clauses, are built in into a surrounding clause and this insertion is different from the conventional construction of a clause. In particular, it seems to be the case that the parenthetical element is not necessary for the grammaticality of the whole construction. Parentheticals have certain properties also in other areas that distinguish them from other kinds of clause combining: (i) their behaviour w.r.t. the positions they can occur in, (ii) scope of sentence mood operators, (iii) focus-background-structure, (iv) variable binding. The peculiar behaviour of parentheticals w.r.t. the above mentioned points leads to the question of whether, and if so, how they are related to other, more standard ways of clause combining. One could think that parenthetical constructions could indeed be classed in a continuum of dependencies or embedding between two sentential units, (see Lehmann 1988 for such a continuum). I will argue that this is not the case, but that parenthetical clause combining of different strength is ordered parallel to the standard ways of building two clauses together.

Nonetheless, the question remains how the structural relation between parenthetical and main clause can be described. I am going to discuss the proposal of Espinal (1991), who offers an analysis based on a three-dimensional conception of syntax. I will compare this approach to a more conservative view on syntax, treating parentheticals like in (1) as adjunction structures, the defining properties of parenthetical constructions leading to an interpretation on a semantic-pragmatic level that is independent of and parallel to the interpretation of the host clause.

Espinal, T.M. 1991. The representation of disjunct constituents. *Language* 67, 3. S. 726 – 762.

Lehmann, C. (1988). "Towards a Typology of Clause Linkage." In: J. Haiman & S. Thompson (eds.) *Clause combining in discourse and grammar*. Amsterdam: Benjamins, 181-225.

---

Gunther Kaltenböck  
Universität Wien  
**Parenthetical clauses in spoken English**  
Mi 14.45-15.15

---

In this workshop contribution I will discuss the class of spoken parentheticals (PCs), typical examples of which are given in (1), using data from the British component of the *International Corpus of English* (ICE-GB)

- (1) a. You could *I suppose* commission some prints of you yours yourself (s1a-015-37)  
b. What I've done here *I hope you don't entirely disapprove* is try and limit the time taken on by this item by putting it in writing (s1b-075-180)

The presentation consists essentially of 2 parts. The first part is devoted to the delimitation of the class in question and a discussion of its internal stratification, resulting in a taxonomy of different types of PCs. Exact delimitation is important as the term 'parenthetical' has come to be used in a variety of different ways in the literature and often lacks a clear definition. For a non-circular definition it is necessary to clearly separate formal from functional criteria and adopt a definition based on purely syntactic criteria, viz. syntactic form, lack of syntactic attachment, and positional flexibility. With the help of these three criteria it is possible to successively narrow down the class of PCs and delimit it from related categories such as question tags, discourse markers, and anacolutha.

The second part of the presentation homes in on one particular type of PC, viz. Reduced Parenthetical clauses (RPCs) or comment clauses, as in (1a). On the basis of a detailed syntactic and semantic analysis of the corpus data I will focus especially on the link between the RPC and its host construction. This is of particular interest because of the lack of a syntagmatic link between them and raises questions regarding the (semantic-pragmatic) scope of the RPC as well as the diachronic development of RPCs from corresponding matrix clauses.

---

Stefan Schneider  
Karl-Franzens-Universität Graz  
**Reduced parenthetical clauses as mitigators: an overview**  
Mi 15.15-15.45

---

The presentation is based on the results of a corpus-based study of reduced parenthetical clauses (RPCs) in contemporary spoken French, Italian, and Spanish (cf. Schneider, forthcoming). In spoken discourse, these expressions have several functions: They may be phatic devices, reporting expressions, illocutionary force indicators or self-correction markers. We focus on mitigating RPCs, that is, on parenthetical clauses that reduce the speaker's communicative responsibility.

Based on a distinction regarding the structure of utterances made by Hare (1970) and based on Caffi's (2001) mitigation frame-work, we distinguish four classes of RPCs that reduce communicative responsibility. The first one is composed of clauses operating on the propositional content of the host. Its most prominent representatives are Fr. *disons*, It. *diciamo*, and Sp. *digamos*. The second class consists of clauses indicating the illocution and operating on the propositional content or on the speaker's commitment. These are plain or hedged performatives, e. g., Fr. *je veux dire*, It. *voglio dire*, and Sp. *quiero decir*. The clauses of the third class directly mitigate the speaker's commitment. They contain belief verbs (e. g., Fr. *je crois* and *je pense*) or negated forms of the verbs Fr. *savoir*, It. *sapere*, and Sp. *saber* that express absence of knowledge on behalf of the speaker. Finally, there is a class of clauses indirectly mitigating the speaker's commitment. It comprises evidential clauses and forms of Fr. *savoir*, It. *sapere*, and Sp. *saber* that attribute knowledge to the addressee or to everybody. The evidential parentheticals derive from verbs expressing sensory perception (e. g., Fr. *voir*) or the obtaining of knowledge (e. g., Fr. *je me rappelle*) and from verbs expressing the semblance or appearance of a state of affairs (e. g., Fr. *sembler*).

RPCs may operate on more than one utterance component. They either alternate between two components (e. g., propositional content or speaker commitment) or operate on two at the same time (e. g., illocution indication and propositional content, illocution indication and speaker commitment).

Caffi, C. 2001. *La mitigazione. Un approccio pragmatico alla comunicazione nei contesti terapeutici*. Münster: LIT.

Hare, R. M. 1970. Meaning and speech acts. *Philosophical Review* 79: 3-24.

Schneider, S. forthcoming. *Reduced parenthetical clauses as mitigators. A corpus study of spoken French, Italian, and Spanish*. Amsterdam - Philadelphia: Benjamins.

---

Markus Steinbach  
Johannes Gutenberg-Universität Mainz  
**Integrated Parentheticals and Assertional Complements**  
Mi 16.15-17.15

---

Verb-first parentheticals like *glaubt er* in (1) differ from other kinds of parentheticals at least in four respects: (i) they are prosodically integrated into their host clause and both form one focus-background-structure; (ii) they have verb-first order, which is very uncommon for declarative clauses in German; (iii) they share crucial properties with matrix clauses licensing V2-complements and with *was*-parentheticals; (iv) the proposition expressed by the host is linked to an implicit propositional argument of the parenthetical predicate.

- (1) Martin [glaubt er] möchte [glaubt er] das Theorem [glaubt er] beweisen [glaubt er]

Martin (believes he) wants-to (believes he) the theorem (believes he) prove  
(believes he)

This talk will focus on properties (ii), (iii), and (iv). Concerning argument linking, two alternative analyses come to mind. The propositional argument of the parenthetical predicate may either be linked to a pronominal expression dropped in sentence-initial topic position (canonical licensing) or it may not be linked in syntax at all (non-canonical licensing). In the latter case, the argument is a syntactically inactive implicit object coreferent with the proposition denoted by the host clause. According to the first analysis, verb-first parentheticals are common V2-structures. By contrast, the second analysis predicts that verb-first parentheticals are genuine V1-structures. It will be argued that both analyses are necessary to account for two different kinds of verb-first parentheticals.

Canonical licensing will turn out to be relevant for the analysis of parenthetical constructions that are hosted by interrogative clauses, which are derived in the same way as corresponding *was*-parentheticals. This analysis also explains why parenthetical predicates like *glauben* can be inserted into interrogative clauses although they do not select interrogative complement clauses.

Non-canonical licensing, on the other hand, is essential to the interpretation of the implicit propositional argument in declarative parenthetical constructions. Declarative parenthetical constructions obey the same restriction as V2-complement constructions. In both cases, the propositional complement of the parenthetical predicate must be assertional. By contrast, topics must not be assertional because they are related to prominent discourse referents. Consequently, an analysis that assumes topic drop is not available and the propositional argument of the parenthetical predicate must be licensed non-canonically.

---

Tanja Kiziak  
Universität Tübingen  
**Extraction or parenthesis? Testing competing accounts.**  
Mi 17.15-18.15

---

For the German construction in (1) two analyses have been discussed in the literature:

- (1) *Welchen Bewerber glaubst du stellen wir im Juni ein?*  
‘Which applicant do you think will we employ in June?’

On the one hand, such constructions have been analysed as long *wh*-extractions from embedded verb-second-clauses (e.g. Grewendorf 1988). On the other hand, they have been argued to be matrix interrogatives with verb-first parenthetical inserts (e.g. Reis 1996).

- (2a) Extraction analysis: *Welchen Bewerber<sub>i</sub> glaubst du [t<sub>i</sub> stellen wir t<sub>i</sub> im Juni ein]?*  
(2b) Parenthetical analysis: *Welchen Bewerber [glaubst du] stellen wir im Juni ein?*

The disagreement has remained unresolved because it is difficult to find any clear evidence which distinguishes between the two accounts.

We used the magnitude estimation methodology (Bard et al 1996) to elicit judgements of sets of strictly controlled example materials. We compared the controversial construction in (1) with clear long extractions from clauses with complementizer *dass* (‘that’) on the one hand and clear parenthetical constructions on the other. Reis (1996) claims that some predicates can appear as matrix predicates in long extractions as in (3a), but that they do not occur in prosodically integrated parentheticals as e.g. in postsubject verb-first-parentheticals in (3b). For the extraction/parenthesis debate it is decisive how the controversial construction in (1) behaves in this respect. We tested as range of predicates with these three structures.

- (3a) *Welchen Bewerber glaubst du, dass wir im Juni einstellen?*  
‘Which applicant do you think that we will employ in June?’  
(3b) *Welchen Bewerber stellen wir, glaubst du, im Juni ein?*  
‘Which applicant will we employ in June, do you think?’

The judgements of the extraction construction in (3a) and the controversial construction in (1) display statistically significant differences. Prima facie this result speaks against an extraction and in favour of a parenthetical analysis of the criterial construction. However, we will also address other possible explanations.

Bard, E./Sorace, A. (1996) „Magnitude Estimation of Linguistic Acceptability”. *Language* 72(1). 32-68.

Grewendorf, G. (1988) *Aspekte der deutschen Syntax*. Tübingen: Narr.

Reis, M. (1996) „Extractions from Verb-Second Clauses in German?” In: Lutz, U. & Pafel, J. (hg.) *On Extraction and Extraposition in German*. Amsterdam: Benjamins.



---

Francesca Del Gobbo  
Universita' di Venezia, Ca' Foscari  
**On the Syntax and Semantics of Appositive Relative Clauses**  
Do 09.00-10.00

---

Appositive relative clauses have been analyzed as independent sentences, or “assertions-to-follow” (Sells 1985, Demirdache 1991), given that a certain parallelism between appositives and main clauses seems to hold. I show that such parallelism breaks down when a wider range of data is taken into consideration.

I propose that appositive relative clauses denote propositions (type *t*) (see Partee 1976) and that the appositive relative pronoun is an E-type pronoun (Sells 1985, Demirdache 1991). I adopt an approach to E-type anaphora that is both grammatical (Heim 1990) and pragmatic (Cooper 1979, Heim and Kratzer 1998), and is built on the theory of NP-deletion elaborated by Elbourne (2001).

Following Haegeman’s (2003) distinction between event-conditionals and premise-conditionals, I propose that appositives are attached outside of the matrix domain. Namely, while restrictives are adjoined to NPs, appositives are adjoined to the matrix CPs. More precisely, appositives are merged to DPs in narrow syntax, but after Spell-Out they are demerged (see Aoun, Choeiri and Hornstein 2001, Fukui and Takano 1998) and merged to CPs.

The crucial claim I make is that at the discourse level, namely after the appositive is remerged to the matrix and a discourse structure is created, both PF and LF representations are relevant for the resolution of the appositive pronoun.

My proposal accounts for the new data and makes a strong empirical prediction: no language with prenominal relatives should allow appositive relative clauses.

- Aoun, J., Choueri, L. and Hornstein, N. 2001. „Resumption, movement and derivational economy.“ *Linguistic Inquiry* 32, 371- 403.
- Cooper, R. 1979. „The interpretation of pronouns.“ In: Heny, F. & H. S. Schnelle (hg.) *Selections from the Third Groningen Round Table (Syntax and Semantics 10)*. New York: Academic Press, 61-92.
- Demirdache, H. 1991. *Resumptive Chains in Restrictive Relatives, Appositives and Dislocation structures*. Dissertation, MIT, Cambridge.
- Elbourne, P. 2001. „E-type anaphora as NP-deletion.“ *Natural Language Semantics* 9, 241-288.
- Fukui, N. and Y. Takano. 1998. „Symmetry in syntax: Merge and Demerge.“ *JEAL* 7, 27-86.
- Haegeman, L. 2003. „Conditional Clauses: External and Internal Syntax.“ *Mind and Language*, 18, 317-339.
- Heim, I. 1990. „E-type pronouns and Donkey Anaphora.“ *Linguistics and Philosophy* 13, 137-177.
- Heim, I. and A. Kratzer. 1998. *Semantics in Generative Grammar*. Malden, Mass.: Blackwell.
- Partee, B. 1976. *Montague Grammar*. Academic Press, New York.
- Sells, P. 1985. „Restrictive and Non-Restrictive Modification.“ CSLI Report No. CSLI-85-28. Stanford, California.

---

Anita Fetzer  
Universität Lüneburg  
**The conjunctive function of disjuncts**  
Do 10.00-10.30

---

The goal of this paper is to examine the functional-grammar concepts of multiple themes (Halliday 1994), disjuncts (Greenbaum & Quirk 1990) and thematic coherence (Givón 1993) with respect to their formal status as parentheticals and with respect to their functional status as a speaker comment about how the utterance is meant and how the addressee is intended to interpret it. Furthermore, the context-preserving and context-modifying functions of disjuncts and their connectedness with context and discourse common ground are investigated.

Multiple themes, extended multiple theme zones (Gómez-González 2000) and disjuncts have a decisive function in the construction of discourse coherence regarding its textual, interpersonal and topical planes. Halliday defines theme as a speaker-oriented notion and specifies its discursive function as *'the element the speaker selects for grounding'* (Halliday 1994:34). But does this also hold for multiple themes? While topical themes are assigned a propositional status and attributed to the sentence-internal domain, textual themes are assigned a conjunctive function whose referential domain may be both sentence-internal and sentence-external. Unlike topical and textual themes, an interpersonal theme's referential domain is sentence-external. For this reason, it is assigned the status of a disjunct which can be looked upon as a parenthetical par excellence. Like a parenthetical, a disjunct plays a superior role to sentence elements, being somewhat detached from and superordinate to the rest of the sentence. Regarding their distribution, disjuncts are more frequent in spoken discourse where they form a separate tone group. Regarding their functions, their prime function lies in the intensification of the pragmatic force. They trigger a process of inferencing to retrieve the speaker-intended meaning and, for this reason, they can be seen as functionally equivalent to illocutionary force indicating devices whose illocutionary effects are defeasible by definition (Sbisa 2002).

Givón, T. (1993): English grammar: a function-based introduction. Amsterdam: Benjamins.

Gómez-González, M. (2000): The theme-topic interface. Evidence from English. Amsterdam: Benjamins.

Greenbaum, S. & R. Quirk (1990): A student's grammar of the English language. London: Longman.

Halliday, M.A.K. (1994): Introduction to functional grammar. London: Arnold.

Sbisa, M. (2002): Speech acts in context. In *Contexts of Social Action. A special issue of Language and Communication* 22(4), A. Fetzer and V. Akman (eds.), 421-436.

---

Yordanka Kavalova  
University College London  
**Parenthetical clauses**  
Do 10.30-11.00

---

This paper focuses on clausal constructions that are inserted into the syntactic string of a host structure and are introduced by the connective *and*. I refer to these as *and*-parenthetical clauses, e.g.,

- (1) Unfortunately for him, *and it is a serious thing at his age*, he'd fractured a leg. (DL-J04, #53)
- (2) Madam, will the Minister confirm that come the single uh Common Market that three hundred million EEC nationals could, *and I emphasise could*, seek employment in this country without the need to obtain a work permit. (s1b-059, #40)

Following a well-established line of analysis of parenthetical constructions, I will argue that parenthetical *and*-clauses are not part of the hierarchical structure of the host. On the other hand, it cannot be denied that these constructions are contained on the linear axis of the host string. To account for this, I will employ Ackema and Neeleman's (2004) approach to parentheticals as insertions of a syntactic representation in a nonterminal node of another syntactic representation.

The second part of this paper will address the speaker's motivation to implement such a cumbersome and lengthy interpolation in his utterance. Crucially, I will propose that the insertion of parentheticals is not to be seen as a disfluency phenomenon, but as part of the speaker's pursuit of achieving optimal relevance—more effects, less cost (Sperber and Wilson 1995). The intended relevance of the utterance lies in the cognitive effects recovered from the host plus the parenthetical. Only together do they offer a faithful representation of the thoughts of the speaker and contribute collectively to the processes of utterance interpretation; hence the *and*.

Finally, the syntactic and pragmatic analyses of these parentheticals will pave a way to argue effectively that the place of interpolation is not ad hoc, but is instead linked to information packaging. I hypothesise that the different timing of the merger of the parenthetical *and*-clause with the host determines its semantic/pragmatic contributions. In addition, based on my data set, I will distinguish two key groups of *and*-parenthetical clauses depending on the way they relate to the host and the effects they contribute to, *viz.* 'speaker-oriented' and 'sentence-oriented'.

---

Mark de Vries  
University of Groningen  
**Parentheses as B-Merged Adverbial Phrases without Scope**  
Do 11.30-12.00

---

There is conflicting evidence as to whether parentheses are syntactically integrated with the host sentence. I argue that they are, but not in the usual way — there are clear differences between parentheses and other ‘non-essential’ sentence material (adverbial phrases), namely in scope, intonation, etc. In particular, I claim that parentheses are b-merged adverbial phrases without scope. This will be explicated below.

On the whole, there are three main types of paratactic (nonsubordinative) constructions: general coordination (including clausal/phrasal coordination and juxtaposition), general apposition (including appositions, nonrestrictive relative clauses, and repair), and parentheses (including parentheticals, hedges, etc.). Some important properties that distinguish these types from each other and also from adverbial phrases are independency, variability in position, intonation, and scope.

I will advance an *invisibility hypothesis*, which says that parentheses do not interact with the host in terms of c-command-based relations. It is tested on the basis of phenomena such as movement, quantifier binding, idioms and polarity items. For example, Q-binding into a restrictive relative clause (which involves subordination) is possible, but Q-binding into an appositive relative (parataxis) is not.

Quite clearly, a generative grammar with Merge as its only basic operation is not able to create syntactic structures with the properties required for parentheses in general. I contend that the intuitive distinction between hypotactic and paratactic hierarchies can be incorporated into the theory of grammar if we allow for a second type of inclusion relation (next to dominance), which has the side-effect of blocking c-command. It follows that there will be a second type of merge, called b-merge.

Parentheses, then, are b-merged adverbial phrases. Therefore, they are syntactically integrated in the host sentence (i.e. they can be detected by constituency tests, etc.), but at the same time ‘invisible’ in terms of scope phenomena.



---

Jörg Peters  
Radboud University Nijmegen  
**Syntactic and prosodic parenthesis**  
Fr 10.30-11.30

---

Traditionally, parentheticals were characterized as syntactic units that are structurally independent of the surrounding sentence. More recent studies suggest that we cannot define parentheticals in purely syntactic terms. Some parentheticals are syntactically linked to the surrounding sentence or even form a syntactic constituent of it. Therefore, it is now popular to delimit parentheticals from other grammatical phenomena by a common communicative function such as adding ‘secondary information’. The aim of the present paper is twofold. First, we present German spontaneous speech data that raise doubts about the existence of a communicative function which is unique to parentheticals. Second, we show that parentheticals can be defined in purely grammatical terms if we distinguish between two kinds of parenthesis, syntactic parenthesis and prosodic parenthesis. According to this view, prosodic features do not signal syntactic structure. Rather, syntactic parenthesis and prosodic parenthesis are operations on different levels of grammatical organization, which independently contribute to information structure. Most of the prototypical parentheticals cited in the literature result from a coincidence of syntactic and prosodic parenthesis. Other cases, in which the status as parenthetical is more controversial, involve either syntactic parenthesis without prosodic parenthesis or prosodic parenthesis without syntactic parenthesis. Finally, we show how an analogous analysis can solve problems in other areas of grammatical description including appositions and relative clauses.

---

Rick Nouwen  
Utrecht University  
**Conventional Implicatures and Dynamic Binding**  
Fr 11.30-12.00

---

In this paper, I take parenthetical constructions to be a source of conventional implicatures. While what is conveyed by a parenthetical phrase is entailed by the utterance it occurs in, this information is separate in the sense that it has widest scope and is therefore, in principle, completely insensitive to embedding. This view is popular, but not entirely unproblematic. Wang et. al. (2005) observe that in (1) the appositive content does not seem to scope out, in contrast to (2). This contrast raises the question as to how the conventional implicature interacts with referential aspects of the host utterance. Del Gobbo (2003) shows that appositive relatives can sometimes, but not always, occur with a quantificational anchor (cf. 3&4). This observation generalises to other parenthetical constructions.

- (1) If a professor, a famous one, publishes a book, he will make a lot of money.
- (2) If John, a famous professor, publishes a book, he will make a lot of money.
- (3) #Every/#No Dutch boxer, a fierce fighting machine, is afraid of Jake.
- (4) Every Dutch boxer, all fierce fighting machines, are afraid of Jake.

I will show that the data are reminiscent of the phenomena involving pronouns that refer to a non-referential antecedent. I will propose a dynamic semantic analysis where the parenthetical constrains the same discourse referent as its host statement. It is well-known that constraints on reference within the scope of a quantifier differ from those outside the scope. Consequently, a conventional implicature that originates from within the scope of a quantifier, but which yields a widest scope effect, will conflict with the conditions contributed by the quantificational statement itself, unless reference is made to something other than the quantified variable. It follows from the dynamic analysis that conventional implicatures might potentially end up incorporated in at-issue semantics. I will argue that in cases of case quantification (as in (1)) this is essentially what happens.

del Gobbo, F. 2003. „Appositives and Quantification.“ In: *Proceedings of the 26th Annual Penn Linguistics Colloquium*.

Wang, L., Reese, B. & McCready, E. 2005. „The projection problem of nominal appositives.“ In: *Snippets* 10.

Rudy Loock  
 Université de Lille III  
**Parentheticals vs. Appositive Relative Clauses in English Discourse**  
 Fr: 12.00-12.30

What this paper suggests is a comparison between parenthetical constructions and appositive relative clauses (henceforth ARCs) in English from an information-packaging point of view (Vallduví 1992, 1993; Lambrecht 1994, Birner and Ward 1998 among others). Appositive Relative Clauses have often been compared to, or even assimilated to, parenthetical structures not belonging to the syntactic structure of the sentences in which they appear (e.g. Safir 1986, who treats ARCs and parentheticals as orphan constituents or Fabb 1990 and his radical orphanage approach for ARCs). We indeed would like to show that there are reasons for bringing them together not only from a syntactic point of view but also from the point of view of their discourse functions. Using contextualised examples, we shall first show that the two structures fulfil close, sometimes similar, roles in discourse. We will show that parentheticals and ARCs are then to be contrasted with independent clauses, whether juxtaposed or coordinated. However, parentheticals and ARCs can be opposed as, in context, it is not always possible to replace one structure with the other to convey the same informational content. We will show that, beyond register variation and syntactic constraints (e.g. the presence/absence of an anaphoric marker), the differences lay in the way the information is presented to the addressees, in particular word order.

- BIRNER, B.J. and G. WARD. 1998. *Information Status and Noncanonical Word Order*. Philadelphia: John Benjamins Publishing .
- FABB, N. 1990. « The Difference between English Restrictive and Nonrestrictive Relative Clauses », in *Linguistics* 26 : 57-78.
- LAMBRECHT, K. 1994. *Information Structure and Sentence Form*. Cambridge University Press.
- SAFIR, K. 1986. « Relative Clauses in a Theory of Binding and Levels », in *Linguistic Inquiry* 17, 663–698
- VALLDUVÍ, E. 1992. *The Informational Component*. Ph.D. thesis, University of Pennsylvania.
- VALLDUVÍ, E. 1993. « Information packaging : a survey », rapport préparé pour WOPIS (*Word Order, Prosody and Information Structure*) : <http://www.hrcr.ed.ac.uk/publications/rp-44.ps.gz>

Sandra Döring  
 University of Kent at Canterbury, Universität Leipzig  
**Parentheticals in Modern German: Prosodic Aspects**  
 Fr. 12.30-13.00

Prosodic features are often mentioned in the relevant German literature on parentheticals (Altmann 1981; Brandt 1994; Öhlschläger 1996; Pittner 1995; Schreier 1988) but they do not play a significant role in the analysis of parentheticals.

Taking into consideration the methodical reflections on prosody in Pierrehumbert (1980/87), Nespor/Vogel (1986) and Grice/Baumann (GToBI, 2002), I will explore the hypothesis that the crucial feature of parentheticals is to be found in the prosodic domain. The investigation will be based on a representative number of examples taken from the oral corpus of German debates in the Bundestag (Lower House of Parliament) in 2004/2005.

The following aspects will be analysed by a choice of typical examples:

- a) The parenthetical has its own intonation pattern, which is significantly different from its anchor clause.
- b) The parenthetical is marked by a different speech rate, register, and intensity.
- c) The end of the first part of the anchor clause is prosodically clearly marked for the following interruption, but not necessarily by a pause.
- d)

- Altmann, H. 1981. *Formen der 'Herausstellung' im Deutschen: Rechtsversetzung, Linksversetzung, freies Thema und verwandte Konstruktionen*. Tübingen.
- Brandt, M. 1994. "Subordination und Parenthese als Mittel der Informationsstrukturierung in Texten". In: *Sprache und Pragmatik* 32: 1-37.
- Döring, S. In press. "Parentheticals in Modern German. Prosodic aspects". In: *PLUM. Proceedings of the 14<sup>th</sup> Postgraduate Conference in Linguistics, 4<sup>th</sup> March 2005*.
- Grice, M. & S. Baumann. 2002. "Deutsche Intonation und GtoBI". In: *Linguistische Berichte* 191: 267-298.
- Nespor, M. & I. Vogel. 1986. *Prosodic Phonology*. Dordrecht: Foris.
- Öhlschläger, G. 1996. "Überlegungen zur Funktion von Parenthesen". In: V. Hertel et al. (Hg.) *Sprache und Kommunikation im Kulturkontext: Beiträge zum Ehrenkolloquium aus Anlaß des 60. Geburtstages von Gotthard Lerchner*. Ffm. 315-324.
- Pierrehumbert, J. 1987. *The Phonology and Phonetics of English Intonation*. Bloomington.



- Pittner, K. 1995. "Zur Syntax von Parenthesen". In: *Linguistische Berichte* 156: 85-108.
- Schreier, G. 1988. Die parenthetische Struktur in der deutschen Gegenwartssprache. Habilitationsschrift, Jena.
- 

Nicole Dehé  
FU Berlin, University College London  
**Prosodic aspects of parentheticals in English**  
Fr 13.00-13.30

---

In the previous literature on parentheticals in English, a number of prosodic characteristics have been identified. For example, they are set off from the host by pauses, they are lower in pitch than the matrix sentence, and they often have a rising terminal (Bolinger 1989 among others). In general, it has been argued that parentheticals interrupt the prosodic flow of the host utterance.

In this paper, I will try to systematically compare the prosodic properties of different types of English parentheticals, such as sentential parentheticals, non-restrictive relative clauses, comment clauses (*I think, you know*) and one-word parentheticals (*what, say, like*). I will specifically address pitch movement/ pitch level and prosodic phrasing. The aim is to make a substantial contribution towards a prosodic definition of parentheticals.

Given the diversity among parenthetical constructions, it is little surprising that there is a lot of variation in the prosodic features of different types of parentheticals, depending, e.g., on their length/prosodic weight, but also within certain types of parentheticals. Here are a couple of examples of the variation I have in mind:

- Sentential parentheticals very often have a rising terminal, while one-word parentheticals never seem to have a rising terminal.
- Within one-word parentheticals, there can be an obligatory pause / prosodic phrase boundary after the parenthetical, as in the case of *what* (Dehé & Kavalova 2005), or the pause can be optional both before and after the parenthetical, as I will show for *like*.

Given the properties of some one-word parentheticals, even the seemingly uncontroversial assumption that parentheticals in general interrupt the prosodic flow of the host utterance must be questioned. In addition to the purely prosodic aspects, I will try to relate the prosodic features of certain parentheticals to their linear position and their pragmatic function.

Bolinger, Dwight. 1989. *Intonation and its Uses: Melody in Grammar and Discourse*. London: Arnold.

Dehé, Nicole & Yordanka Kavalova. 2005. "The syntax, pragmatics and prosody of parenthetical *what*." Ms., University College London.

## AG 6 Spracharchive – Standards, Erzeugung & Zugang / Language Archives, Standards, Creation and Access

Koordination: Helen Aristar Dry, Jost Gippert, Sophie Salffner, Romuald Skiba, Thorsten Trippel, Peter Wittenburg

### Programm

Mittwoch 22.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
13.45-14.15	Kathrin Kirsch	<i>Working Cross-Platform: a case Study in Coding, Sharing and Analyzing Corpora</i>
14.15-14.45	David Nathan	<i>Protocol and the language data life-cycle at ELAR</i>
14.45-15.15	Ulrike Zeshan	<i>Archiving sign languages</i>
15.15-15.45	Stefanie Dipper, Michael Götze, Stavros Skopeteas	<i>A Typological Language Archive for Reseraching Information Structure</i>
15.45-16.15	Kaffeepause	
16.15-16.45	Peter Berck, Hans-Jörg Bibiko, Marc Kemps-Snijders, Albert Russel, Peter Wittenburg	<i>Language Archive Utilization</i>
16.45-17.15	Mihaela Vela, Silvia Hansen-Schirra	<i>The CroCo Translation Archive</i>
17.15-17.45	Jost Gippert	<i>The TITUS text archive: 10 years of Unicode application revisited</i>
17.45-18.15		<i>Summary and Discussion</i>

Donnerstag 23.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
09.00-10.30	Gary Simons, Helen Aristar-Dry	<i>E-MELD: openness, ontologies and interoperability</i>
10.30-11.00	Daan Broeder, Andreas Claus, Freddy Offenga, Romuald Skiba, Paul Trilsbeek, Peter Wittenburg	<i>Language Archives at MPI</i>

11.00-11.30	Kaffeepause	
11.30-12.00	Adam Saulwick, Rob Godemans, Alexis Dimitradis, Menzo Windhouwer	<i>Architecture and procedures for the integration of linguistic databases in the Typological Database System</i>
12.00-12.30	Mirjam Kessler Thierry Declerck	<i>Data Categories as the link between Langauge archives and NLP tools</i>

Freitag 24.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
10.30-11.00	Monika S. Schmid	<i>The Language Attrition Research Project</i>
11.00-11.30	Jochen Rehbein, Kai Wörner	<i>Mehrsprachige Korpora geschriebener und gesprochener Sprache(n) am Sonderforschungsbereich Mehrsprachigkeit – Modelle, Werkzeuge und Daten</i>
11.30-13.00	Panel discussion, Moderator: Thorsten Trippel	<i>The Standard issue</i>
13.00-13.30		<i>Final discussion an future plans</i>

---

Kathrin Kirsch  
Max Planck Institute for Psycholinguistics  
**Working Cross-Platform: a Case Study in Coding,  
Sharing and Analyzing Corpora**  
Mi 13.45-14.15

---

Linguistic research in the area of first and second language acquisition relies heavily on the analysis of spoken language corpora. Creating such corpora is a time and work intensive task, and is often specific to the type of input data available, to the type of analysis to be performed, and to time span available to the researcher.

That all leads to the very common situation, that the various steps necessary to create a corpus from raw data is shared between different individuals and institutions, is repeated and refined over time, with different tools and varying target objectives. Depending on the results of the analysis and on the development of research interests, that process is iterative, and open.

The present paper describes a very typical corpus development. The Stoll corpus is a collections of 364 audio and video recordings of Russian children, and was originally recorded by Sabine Stoll from 1994 to 1997 in St. Petersburg, Russia. The corpus is now part of the MPI IMDI database. The paper describes the real life cycle of the corpus data, inclusive the description of the used tools, transcription techniques, corpora branches

etc; it further describes the advantages and disadvantages for the chosen techniques for these data.

The paper also motivates a potentially more efficient and flexible work flow. The estimated effort for that hypothetical course is compared to the effort really spent for the Stoll corpus. This theoretical course further motivates the design of interoperable tools; transcription, encoding and version control standards; and the need for auditing and reproducibility of the corpus life cycle processes.

---

David Nathan  
SOAS, University of London  
**Protocol and the language data life-cycle at ELAR**  
Mi 14.15-14.45

---

There is wide acknowledgement that linguistic diversity is one of humanity's cultural assets and that language endangerment is a global social crisis. However, there is much less agreement to be found amongst those attending to language endangerment about how to address it, and such agreement that exists is confined to certain areas such as an incipient discipline of language documentation, and some technical aspects of data encoding and archiving.

The Endangered Languages Archive (ELAR) is one programme of the Hans Rausing Endangered Languages Project at the SOAS, University of London. ELAR aims to

address some of the gaps we see between the issues that currently occupy centre-stage in language endangerment, and what we perceive to be the services that a future archive can offer. Firstly, along what we could call the “resource” axis, current emphasis is on resource discovery, i.e. making institutionally-held language data identifiable (and in many cases deliverable) to a wider public. At the other end of this axis is what seems to us to be a demonstrably greater need - promoting the development of documentation materials, especially those that can be used to combat language endangerment. The set of roles that archives can legitimately and effectively play across this spectrum is currently an open question, given convergences in the digital domain and the important role that archives such as DoBeS are increasingly played in bridging data creation, preservation, and distribution. Along a second axis more closely associated with the internal processes of archives, current emphasis is on data preservation, e.g. through methods for maximising its interoperability, surely an ever more crucial activity in a world of volatile digital data; yet the opportunities of multimedia and the realities of resource distribution suggest that archive institutions can contribute increasingly to supporting effective and innovative methods of delivering resources.

One of the issues that we are pursuing at ELAR is “protocol” - sensitivities, relationships and restrictions that apply to language data and how these are described and implemented. Protocol could be regarded as *non*-interoperability at the social level. The field of endangered languages is imbued with sensitivities (often due to the same reasons that cause languages to be under threat), and we are exploring how to bring them into the centre of a methodology for documentation and archiving. Several relevant archives recognise the issue, but none has yet made a comprehensive implementation. Protocol is not merely an outpost of the “metadata” that supports discovery and management of materials. It is a core part of the data production, preservation and distribution chain, reflecting the full and dynamic lifecycle of data and the human lifeblood that ties data to its potential for language revitalisation. It also resonates with wider issues of our time, such as changing approaches to intellectual property, and digital rights management.

The presentation will look at protocol issues, from elicitation to encoding, to how we plan to implement them at ELAR through policies, technologies, and relationships.

---

Ulrike Zeshan  
Max Planck Institute for Psycholinguistics  
**Archiving sign languages**  
Mi 14.45-15.15

---

The creation and archiving of corpora of individual sign languages is not a new undertaking. For instance, in the 1990s the *100,000 Sign Corpus of New Zealand Sign Language* was set up at Victoria University of Wellington, the Institute for German Sign Language and Communication of the Deaf at Hamburg University has developed tools for managing lexicographical databases of German Sign Language signs (most recently *GlossLexer*), and the American Sign Language Research Group at the University of Boston keeps upgrading its *SignStream* annotation program. However,

corpora and archives comprising a number of different sign languages were previously nonexistent.

The Sign Language Typology Research Group at the Max Planck Institute for Psycholinguistics is pioneering the compilation of a sign language corpus representing a large number of geographically and typologically diverse sign languages. Sub-nodes in the corpus so far include sign languages from South Korea, mainland China, India, Turkey, Germany, Russia, the Netherlands, Hong Kong, and a deaf village in Bali, Indonesia. The sign language corpus uses the metadata and annotation standards of the MPI's Browsable Corpus (e.g. *IMDI Editor*, *ELAN*). However, the archiving of sign language data poses specific problems in terms of accessibility and ethical standards that do not present themselves in the same way for spoken language corpora. Some sign languages, especially those used in small local village communities with a high percentage of hereditary deafness, are highly endangered, and the expansion of the archive will have to take special note of these cases.

The next steps of working with the sign language corpus will be to extract information from the corpus in the form of a searchable sign language typology database, where information can be stored, managed and retrieved about grammatical characteristics of the various sign languages in the corpus. Initial comparative data sets will include color terms, interrogatives, kinship terms, negatives, numbers and numeral incorporation, possession and existence, and signs related to sensory perception. This raises questions about the relationship between such a searchable database and the existing and expanding sign language corpus, both at technical and at conceptual levels.

This presentation discusses the current set-up and specific challenges in relation to a large sign language corpus, as well as discussing the said future directions of development.

---

Stefanie Dipper, Michael Götze, Stavros Skopeteas  
Universität Potsdam

**A Typological Language Archive for Researching Information Structure**  
Mi 15.15-15.45

---

Information structure (IS) concerns the organization of linguistic information within utterances, i.e., the way a speaker presents information to some hearer by taking into account factors like background knowledge and context. In 2003, a long-term research center (SFB 632) has been established at Potsdam University and Humboldt University of Berlin in order to investigate the various facets of IS from very different perspectives and to contribute to a broader and more general understanding of the phenomena by collecting and integrating data from (a) different linguistic disciplines such as Psycholinguistics or Theoretical Linguistics, (b) different layers such as phonology or syntax, and (c) different languages such as English or Primi.

The data is elicited by means of a universally applicable, typological questionnaire, which mainly consists of a collection of 30 production experiments that are designed to induce different realizations of IS by manipulating the discourse and context conditions. By means of this questionnaire, data is currently elicited for 15 typologically diverse languages.

The collected data is annotated by detailed information, at the layers of phonology, morphology, syntax, semantics, and IS. Annotations make use of standardized tag sets, for which we provide detailed guidelines that aim at being largely language-independent and theory-neutral.

The data with its annotations is mapped to a standardized, XML-based standoff representation. This XML data is fed into a web-based tool, ANNIS, which visualizes the annotated data and allows the user to search information at multiple levels. This way, putative interactions and correlations between individual factors can be investigated in a systematic and straightforward way.

Our talk will present the typological archive and its structure with emphasis on the development of standardized annotations for a cross-linguistic corpus and exemplify their functionality for various research questions. The database ANNIS and annotated data of different languages will be made available to the research community by the time of the DGfS conference.

---

Peter Berck, Hans-Jörg Bibiko, Marc Kemps-Snijders, Albert Russel, Peter Wittenburg  
Max Planck Institute for Psycholinguistics  
**Language Archive Utilization**  
Mi 16.15-16.45

---

The emerging language archives covering resources such as annotated media files and lexica and the virtual integration of such archives already being implemented in some projects have to support a variety of access methods to the user that fully use the modern capabilities of the web. We can distinguish four levels: (1) Usage of metadata to discover language resources. (2) Direct play or visualization of individual language resources using browser plug-ins. (3) Web-based download of language resources. (4) Utilization of a selection of language resources from one or several archives via web-based methods.

The first three levels are all available when using the MPI archive. The metadata descriptions are open and created according to the IMDI<sup>1</sup> XML-schema. The IMDI framework offers a range of navigation options such as browsing in the distributed domain of XML files, browsing in a HTML domain created by on-the-fly-transformations, searching by using structured information to increase the precision, searching by using full-text capabilities on all metadata fields, providing reduced metadata descriptions to satisfy Dublin Core<sup>2</sup> and OLAC<sup>3</sup> service providers and finally by using typical search engines as Google. When access is granted every resource found can be played or visualized when appropriate tools or plug-ins can be invoked from the browser including MPEG4 video streaming. A very important

---

<sup>1</sup> IMDI is a detailed metadata set for the discovery and management of language resources.

<sup>2</sup> Dublin Core is a flat metadata set of 15 elements.

<sup>3</sup> OLAC = Open Language Archives Community, a metadata set based on the DC set.

option is to download not only individual resources, but whole sub-archives. This option is relevant for many usage scenarios, in particular when setting up for example regional data centers.

However, an increasing number of users want to be able to create a virtual selection of resources (basket), define it as their temporary working domain and carry out some operations on it with web-based methods. Web-based frameworks that allow to handle complex language resources such as annotated media or lexica, to easily compare various resources, to make use of an interaction between lexica and annotations, to carry out combined metadata and content searches on such resources, to create statistics and to present results such as concordances are of great relevance. The MPI already offers first versions of such web-based utilization frameworks for working on annotated recordings and multimedia lexica. In general it has to be assumed that such baskets will contain annotations and lexica in different formats. To cope with that heterogeneity a number of well-known and standard formats are supported.

It may be assumed that the selected resources will also exhibit a large variety in linguistic terminology describing linguistic phenomena, i.e. interoperability mechanisms at the semantic level have to be offered. In a collaboration between the Max-Planck-Institutes for Psycholinguistics and evolutionary Anthropology an infrastructure was suggested and is now being implemented that takes care both of the bottom-up driven needs of the researchers and of the emerging centralized ontology resources. The bottom-up driven mechanism allows the researcher to easily create new and manipulate existing concept and relation registries for the selected set of resources and to integrate references to concepts registered centrally such as in the emerging ISO TC37/SC4 Data Category Registry<sup>1</sup> or in GOLD<sup>2</sup>.

The talk will present the different access methods in detail, describe the way ontologies are used and discuss further needs.

---

Mihaela Vela, Silvia Hansen-Schirra  
Universität des Saarlandes  
**The CroCo Translation Archive**  
Mi 16.45-17.15

---

The main goal of the CroCo project is the creation of a multi-lingual language resource with multi-layer annotation. Archiving translations is an important issue in translation studies and practice. A translation database can be used for translation quality assessment on the one hand as well as for teaching purposes and translation practice on the other. Similarly to a huge online dictionary, it offers translators and translation students a large amount of examples in context and can thus help to find translation solutions. Recently, translation corpora have also become subject of investigation in translation research. On the basis of aligned parallel corpora as well as monolingually comparable corpora of translated and original texts, researchers

---

<sup>1</sup> ISO TC37/SC4 is a subgroup within ISO dealing with language resource management.

<sup>2</sup> The GOLD ontology is being worked out within the E-Meld project.



investigate the nature of translated text and its influence on the target language. Since there is no reliable resource for German translators and researchers, it is our aim to create a translation resource and to make it accessible for national and international translation, teaching and research purposes.

In the CroCo project, German-English as well as English-German translations are collected. Archiving translations for research as well as translation training and practice involves three main issues:

- Corpus representation: The corpus file headers in the CroCo project are encoded in XML on the basis of the Text Encoding Initiative (TEI) standard. This guarantees the exchangeability and searchability of the corpus and its mark-up. Each text is encoded in terms of a header that provides meta-information on title, author, publication, register information, etc. For a translation corpus, it is important to encode information on the translator, the translation, the translation process, the author of the original and the source language text. This information is important to filter the corpus and divide it into subcorpora. The text body is annotated for headings, sentences, paragraphs etc.
- Alignment: For the exploitation of a parallel corpus, the units of translation (i.e. source language text units and their translational equivalents) need to be aligned. The most commonly implemented technique is sentence-by-sentence alignment. For many purposes, however, also smaller translation units are to be compared. Thus, we develop an algorithm which enables the alignment of clauses, phrases as well as words. The aligned words, phrases, clauses and sentences are stored in separate files, linking the sentences through XML IDs and IDREFs
- Linguistic annotation: In order to offer a grammatical background for the translated texts, our translations are annotated for morphological information, parts-of-speech and chunks. Each annotation is realized in a different layer because different units are annotated per layer (phrases, words, clauses, etc.). The annotation layers are kept separate using XML stand-off mark-up. They are linked through unique word IDs.
- Since not only translation units are to be compared, but also the annotation assigned to the translation units (e.g. English noun phrases are contrasted to their German equivalents), a new representation standard is developed which allows the alignment of annotated translation units. The annotation units are linked through IDs and IDREFs.

The representation standard developed in the CroCo project for the annotation of translation corpora allows not only the alignment of raw text sequences but also of annotation units. This facilitates the exploitation of the corpora since the extraction of corresponding translation units as well as their annotation units is a straightforward step in the querying process. On the basis of simple XSLT queries, text segments, translation units and/or annotation units of several annotation layers can be extracted and compared.

---

Jost Gippert  
Universität Frankfurt  
**The TITUS text archive: 10 years of Unicode application revisited**  
Mi 17.15-17.45

---

The TITUS server (<http://titus.uni-frankfurt.de>) was established in 1995 with the aim to provide, among other materials, the primary texts collected in the data base of ancient Indo-European and adjacent languages (“Thesaurus indogermanischer Text- und Sprachmaterialien”) which had been founded in 1987. To cover the languages involved (e.g. Sanskrit, Avestan, Middle Iranian languages, Ancient Greek, Old Germanic languages) in a decent way, we decided right from the beginning to use Unicode encoding as the default encoding for the textual materials, thus meeting both the requirements of scholarly work and the facilities of modern computational technology. The problems to be solved during the successive phases of the development of the Unicode standard consisted, among others, in the representation of different systems of diacritics used with Latin and other scripts, the application of non-Latin scripts, and the interchange of language-specific data through browsers. The paper will discuss some of the problems and the solutions achieved within the TITUS server project since 1995.

- Bunz, C.-M. / Gippert, J. 1997: *Unicode, Ancient Languages and the WWW*. <http://titus.uni-frankfurt.de/personal/jg/unicode/unicode.htm>
- Bunz, C.-M. 1998: *Der Thesaurus indogermanischer Text- und Sprachmaterialien (TITUS) – ein Pionierprojekt der EDV in der Historisch-Vergleichenden Sprachwissenschaft*. <http://titus.uni-frankfurt.de/texte/sdv198.pdf>
- Gippert, J. 1999. „Language-specific encoding in multilingual corpora: Requirements and solutions.“ In: J. Gippert (hg.) *Multilingual Corpora: Codierung, Strukturierung, Analyse. 11. Jahrestagung der GLDV*. Prague: Enigma, 371-384.
- Gippert, J. 2001. „Der TITUS-Server: Grundlagen eines multilingualen Online-Retrieval-Systems.“ In: G. Willée / B. Schröder / H.-Chr. Schmitz (hg.) *Computerlinguistik: Was geht, was kommt? / Computational Linguistics: Achievements and Perspectives. Festschrift für Winfried Lenders*. Sankt Augustin: Gardez!, 81-86. See also: <http://www.zdv.uni-tuebingen.de/tustep/prot/prot831-titus.html>
- <http://titus.uni-frankfurt.de/texte/texte.htm>
- <http://titus.uni-frankfurt.de/unicode/unitest.htm>

---

Helen Aristar Dry, Gary F. Simons  
Eastern Michigan University, SIL  
**The Experience of the E-MELD Project in Developing and Evaluating  
Digital Archiving Recommendations**  
Do 9.00-10.30

---

As early as 1996, a special US Taskforce on Digital Archiving drew attention to the fact that changes in coding, formats, software programs, and operating systems have made much valuable data inaccessible to modern computers, and therefore useless. They warned that we may inadvertently be creating digitized material that “will be even less accessible to future generations than the printed material it was intended to replace.” To combat the ‘relentless obsolescence of digital formats and platforms,’ archivists and information technology specialists have been developing ‘recommendations of best practice’ in the creation and preservation of digital material. And initiatives like E-MELD (Electronic Metastructure for Endangered Languages Documentation), a 5-year NSF-sponsored project, are adapting the recommendations to language documentation and promulgating them within the linguistics community.

Now in its fifth and last year, E-MELD has sponsored 5 workshops bringing together documentary linguists and language technology experts in order to develop both recommendations of best practice and the tools and facilities necessary to support them. Nevertheless, given the relative inaccessibility of digital expertise and appropriate software within the discipline, flawless implementation of ‘recommendations of best practice’ may not always be practical for the documentary linguist. Hence, one outgrowth of E-MELD workshops has been the recognition of a hierarchy of goals which digital documentation practices can strive for. Listed in increasing order of utility (which corresponds with increasing effort to achieve), the results of digital language documentation should be:

- preservable (Good practice)
- interpretable (Better practice)
- interoperable (Best practice)

A practice is “good” if it creates a resource with the potential for long term preservation, “better” if it also promotes easy interpretation by future generations, and “best” if it, in addition, facilitates automated interoperation with resources developed by other linguists. Thus, creating an archival copy of an annotated text in ASCII format is acceptable practice since it is a format that should remain readable long into the future, but adding well-defined meta-explanation (such as documenting the phonetic value of the characters and the meaning of grammatical abbreviations) is even better practice, since it facilitates interpretation of the resource and reduces the probability of “creating another Rosetta Stone” that will take generations to decipher. Using machine-interpretable standards to encode the structure and analysis of the text approaches “best” practice because it also facilitates the interoperation of the resource with similar resources that have been independently archived.

In this two-part presentation, the recommendations that have come out of the E-MELD workshops over the years (e.g., use of Unicode, OLAC metadata descriptions, Ethnologue language codes, XML descriptive markup, the GOLD ontology) will be discussed within this framework. First, Helen Aristar Dry will present the hierarchy and demonstrate how it may be used as a set of criteria by which particular archived resources as well as archival practices in general can be evaluated. Then, Gary Simons will focus on the criteria of interoperability and illustrate how the E-MELD recommendations for best practice are facilitating interoperation across resources and across the linguistics community.

---

Daan Broeder, Andreas Claus, Freddy Offenga, Romuald Skiba, Paul  
Trilsbeek, Peter Wittenburg  
Max Planck Institute for Psycholinguistics  
**Language Archives at MPI**  
Do 10.30-11.00

---

The introduction of digital technology revolutionized our view on archiving which traditionally was about storing physical objects for an unlimited time. In contrast to most traditional physical objects digital media have a very short life-time. Therefore, we understand that we don't have to take care anymore of the original recording be it on a cassette tape, on CDROM or notes on paper. We can exclusively focus on the content, since creating digital copies is in general free of losing quality. At the Max-Planck Institute for Psycholinguistics (MPI) this fundamental change was realized very early with the consequence that since several years all material that has a long-term relevance was systematically digitized and stored on computer systems. The MPI researchers soon realized the great advantages of operating in the digital domain. The effect, however, was that the increasing amount of digital material every individual researcher had to maintain resulted in a chaotic situation: Digital copies were stored everywhere, people did not know where to find data from colleagues who left, the state of the different copies was not clear etc.

To compensate for these deficits the technology team at the MPI started to develop a complete metadata framework. It should allow researchers to describe the various types of language resources, to group them into manageable structures, to discover them via search and browse operations and to integrate them into a central repository that takes care of technology developments. With the help of European projects (ISLE<sup>1</sup>, INTERA<sup>2</sup>, ECHO<sup>3</sup>) and the DOBES programme<sup>4</sup> the IMDI<sup>1</sup> metadata

---

1 ISLE = International Standards for Language Engineering, a project funded by the C and NSF

2 INTERA = Integrated European language Resource Area, a project funded by the EC

3 ECHO = European Cultural Heritage Onlien, a project funded by the EC

4 DOBES = Documentation of Endangered Languages funded by the VolkswagenFoundation.

framework and the archive concept were further developed and refined. We can now speak about mature and stable frameworks that allow to manage and access the big MPI archive with its more than 150.000 digital objects requiring a capacity of about 11 TeraByte. All resources are discoverable via web technologies using the IMDI metadata infrastructure tools. A downloadable browser operating in the linked XML metadata files and allows to easily combine IMDI domains coming from different archives. Web applications written in Java allow to browse in the metadata domain via normal web browsers. Structured and unstructured search, DublinCore/OLAC<sup>2</sup> gateways and searching via Google are supported.

With increased amount of data and contributors from different areas of linguistic studies (endangered languages, child and second language acquisition, sign language) another bottleneck emerged: the archive managers could not carry out anymore the task to check the correctness of all deposited data and integrate it into the well-organized archive. Therefore, the MPI team built the LAMUS<sup>3</sup> tool which is a full-blown content management system specifically tailored for language resources. Users that want to work independently of archive managers can now integrate new and update existing resources including metadata descriptions and structure the archive with the help of web-based interfaces.

The availability of LAMUS allows us to open up the archive services to other linguists. On request work spaces can be opened and individual linguists or projects could make use of the MPI facilities. Metadata has to be created and once integrated it is open to the public. The access to the resources can be defined by the depositor.

---

1           IMDI = a detailed metadata set for language resources worked out by many linguists and language engineers

2           Dublin Core is a flat metadata set with 15 elements for the discovery of all types of web resources; OLAC added two special elements to better describe language resources

3           LAMUS = Language Archive Management and Upload System

Adam Saulwick, Menzo Windhouwer, Rob Goedemans, Alexis Dimitriadis  
 University of Amsterdam, Leiden University, Utrecht Institute of Linguistics  
**Architecture and procedures for the integration of linguistic databases  
 in the Typological Database System**

Do 11.30-12.00

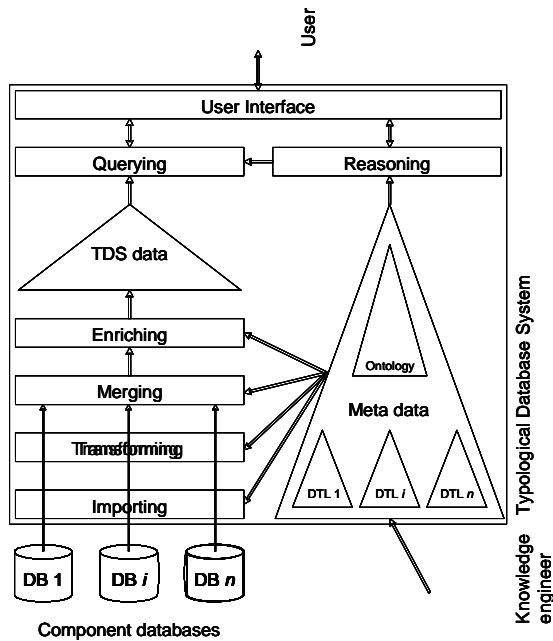
The Typological Database System (TDS) provides an online interface to multiple independently developed typological databases. It allows unified querying by means of an integrated ontology. Constructing such a system requires careful consideration of both the form and the content of the information contained in linguistic archives. We show that it is sometimes necessary to perform special types of data manipulation in order to represent the information from diverse sources in a coherent and faithful manner. The data manipulations may be trivial in nature, such as standardizing values, or may involve more substantial transformation, such as the regrouping of fields into relevant superordinate categories. In all cases we aim to preserve the integrity of contentful variation between the databases, such as different theoretical commitments, while striving to eliminate differences in form.

We present our methodology for the integration of databases into the system. The diagram below gives an overview of the integration process. We focus on a step-wise description of theoretical

and technical issues involved. These include: (i) data re-structuring, (ii) data addition, (iii) data representation, (iv) mapping to ontological concepts which have to be defined for each specialised domain, (v) database documentation and (vi) representing concepts in a network of theoretically motivated hierarchies.

The broader relevance of the TDS is that the system architecture, tools and techniques are robust, and hence adaptable to the needs of other data-warehousing applications.

Access to the TDS is open to everyone, and



requires only a web browser of relatively recent vintage. A working prototype of the system is available at <http://language.link.let.uu.nl/tds/>.

Contact: [a.g.saulwick@uva.nl](mailto:a.g.saulwick@uva.nl)

---

Mirjam Kessler, Thierry Declerck  
Universität Mainz, DFKI

**Data Categories as the link between Language archives and NLP tools**  
Do 12.00-12.30

---

Our contribution is about an European Project, in the eContent framework, the LIRICS project: Linguistic Infrastructure for Interoperable Resources and Systems. This project addresses the needs of today's information and communication society where globalisation and localization necessitate multilingual communication creating an increasing need for new standardization as well as urgent recognition of existing de facto standards and their transformation into de jure International Standards.

In this, LIRICS provides ISO ratified standards for language technology to enable the exchange and reuse of multilingual language resources. LIRICS also provides for guidelines supporting the implementation of these standards for end-users by providing an open-source implementation platform, related web services and test suites building on legacy formats, tools and data. As for the resources, LIRICS covers the ISO standardisation of computational lexicons, morpho-syntactic, syntactic and semantic annotation for the language industry.

The model LIRICS is adopting for specifying and representing the various annotation schemes concerned -- morpho-syntax, syntax and semantic -- is based on the general principles of the Linguistic Annotation Framework, which is an on-going project within the ISO committee TC 37/SC 4. Those principles consider a class of semi-structured documents that can be specified through the combination of, on the one hand, a meta-model that informs the general practices in organizing information in a given application domain, and, on the other hand, a selection of data categories (DCS), that characterizes the elementary information units that can be attached to the various sub-structures (or components) of the meta-model. Indeed, the components in the meta-model should be considered as elementary linguistic abstractions that reflect the granularity of the description intended by the meta-model.

The data categories in the domain of language resources will be defined in LIRICS independently of any specific annotation scheme or language processing application while describing as wide a variety of linguistic contexts as possible. This dynamic resource will act as a reference for anyone describing an information structure in the domain of language resources.

Using this Data Category framework we see a way to ensure the interoperability between the metadata describing the huge variety of annotated language data available in the kind of archives described by the call for contributions of the WG6 "Language Archives: Standards, Creation and Access" on the one hand and language processing tools on the other hand. This interoperability can be based on mapping or filtering mechanisms between the descriptions of language units in use in the archives and the input/output specifications of language processing tools.

Another issue would be to add directly and possibly in an automated fashion to the data contained in the language archives a description layer consisting of the kind of data categories being defined for example in the context of LIRICS (but to be extended also to other ISO activities dealing with language).

To validate the feasibility and practicality of an on-line server environment accessing the relevant data categories, a task has been defined in LIRICS for integrating a reference set of data categories for metadata descriptions of language resources as they are used for example in IMDI and OLAC.

---

Monika S. Schmid  
Faculteit der Letteren, Vrije Universiteit Amsterdam  
**The Language Attrition Research Project**  
Fr 10.30-11.00

---

This talk will present the ongoing efforts to standardize research designs and build a multilingual resource and data archive for first language attrition research. The project aims to assess and re-evaluate the methodological and theoretical frameworks for first language attrition research, and to develop a widely applicable research design in order to provide a more unified and solid basis that subsequent studies can draw upon. There are a number of theoretical and methodological problems which such a research design has to overcome. Typically, attrition studies ‘borrow’ their experimental designs from other areas of linguistic investigation (e.g. SLA, foreign language teaching, aphasia). This selection is usually based on the formal requirements of how the theoretically formulated hypothesis could be tested, but does not take into account the problem that the attrition situation might require a different approach. It has not been established whether the experiments which demonstrably test certain things in the areas of investigation from which they were borrowed actually test the same things in language attrition.

An international group of senior scholars and graduate students has attempted for the past 3 years to develop a common research design, apply this in their own studies, and collect a body of language attrition data that can be shared across the research group. The experiments thus developed are currently being tested and assessed for their validity across a range of languages and settings.

This presentation will discuss some of the problems concerned with the standardization of these data, with making them available to other researchers, and with how to design such experiments for maximal usefulness across a wide range of research questions.



---

Jochen Rehbein, Kai Wörner  
Vrije Universiteit, Universität Hamburg  
**Mehrsprachige Korpora geschriebener und gesprochener Sprache(n) am  
Sonderforschungsbereich Mehrsprachigkeit – Modelle, Werkzeuge und  
Daten**  
Fr 11.00-11.30

---

Am SFB 538 wird seit Juli 2000 im Projekt Zb „Mehrsprachige Datenbank“ (seit Juli 2005: Projekt Z2 „Computergestützte Erfassungs- und Analysemethoden multilingualer

Daten“) an Konzepten und Werkzeugen für die Verarbeitung von Korpora gesprochener

und geschriebener Sprache gearbeitet. Während in der Entwicklung sprachtechnologischer Anwendungen der Rückgriff auf große Korpora von Speech mittlerweile die Regel ist, befinden sich die nicht-technologisch orientierten Bereiche, die mit der Erforschung gesprochener Sprache zu tun haben, diesbezüglich noch in einer Experimentierphase.

Für die Arbeit am SFB 538 wurde das XML-basierte System EXMARaLDA (Extensible Markup Language for Discourse Annotation) entworfen und in Form von mehreren aufeinander aufbauenden Datenformaten und miteinander interagierenden Software-Werkzeugen zur Verarbeitung von Transkriptionen gesprochener Sprache (hauptsächlich: ein Transkriptionseditor, ein Korpus-Manager und ein Suchwerkzeug) implementiert.

Die entwickelten Werkzeuge wurden und werden genutzt, um Korpora gesprochener Sprache zu erstellen und zu pflegen. SFB-Projekte, die an Daten geschriebener Sprache arbeiten, wurden ebenfalls bei der Methoden- und Korpusarbeit unterstützt, hier wird vorwiegend auf mit TEI-kodierten Texten gearbeitet.

Dabei entstanden und entstehen ein- und mehrsprachige Korpora in einer großen Anzahl von Sprachen, sowohl aus schriftlichen als auch aus mündlichen Verwendungszusammenhängen. Im Projekt Z2 wird eine möglichst breite Vergleichbarkeit dieser Korpora angestrebt.

Der Vortrag soll, nach einem kurzen Überblick über die EXMARaLDA-Umgebung einige der am SFB entstandenen Korpora exemplarisch vorstellen. Abschließend sollen einige Fragen zum Zugang und zur Langzeitarchivierung der Daten kurz angerissen werden.

Mit den Sonderforschungsbereichen 441 „Linguistische Datenstrukturen“ und 632 „Informationsstruktur“ werden zudem ab dem 1.12.2005 in einem gemeinsamen DFG-geförderten Projekt Richtlinien und Verfahren zur Langzeitarchivierung und Verfügbarmachung von linguistischen Daten erstellt. Diese sollen auf die am SFB gesammelten Daten angewendet werden, um diese im Rahmen der rechtlichen Möglichkeiten für andere Forscher sichtbar und verfügbar zu machen.

Abraham, W. 1995. *Deutsche Syntax im Sprachenvergleich*. Tübingen: Narr U2-232

---

Panel Discussion  
**The Standard Issue**  
Fr: 11.30-13.00

---

Many different initiatives develop and discuss standards, among them are E-MELD, ISO TC 37 SC 4 and individual projects. This Panel Discussion is intended to be an open forum at the end of the workshop targeting and discussing questions including but not restricted to the following:

- What use are standards in the creation and access of language archives?
- Which standards are to be worked with?
- In which areas do we need more of a standard?
- Where did standards go wrong, in the sense that they are not used, cannot be used, or should not be used?
- If standards are important, how can they be promoted to a wider use? Do we need to hide standards from a user?
- How can developers for tools be encouraged to use standards?
- What kind of infrastructure do we need to foster creation, access and archiving of language resources?

More or other questions may emerge during the workshop and should be added for a general discussion.

The ideal outcome of this part of the workshop is an action plan for the various groups, and initiatives, each given to someone in attendance to take the initiative or lobby at one or the other group.

## **AG 7 Beschreibung und Modellierung grammatischer Variation** **Describing and modeling variation in grammar**

Koordination: Andreas Dufter, Jürg Fleischer, Guido Seiler

### **Programm**

Mittwoch 22.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
13.45-14.45	Bernd Kortmann, Verena Haser	<i>Modeling variation in English dialect syntax: New prospects for syntactic theory</i>
14.45-15.15	Bert Cappelle	<i>Clean up your room or clean your room up: is free choice an illusion?</i>
15.15-15.45	Said Sahel	<i>Was steuert die Variation der Adjektivflexion nach Pronominaladjektiven?</i>
15.45-16.15	Kaffeepause	
16.15-16.45	Judith Berman	<i>Das Prädikativ als Quelle grammatischer Variation</i>
16.45-17.15	Ljudmila Geist	<i>Variation im Bereich der prädikativen Adjektive im Russischen</i>
17.15-17.45	Kris Heylen	<i>The order of nominal subject and pronominal object in the German middle field. A case of free syntactic variation?</i>
17.45-18.15	Gert De Sutter, Dirk Speelman, Dirk Geeraerts	<i>A multivariate approach to the description and modelling of word order variation</i>

Donnerstag 23.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
09.00-09.30	Georg A. Kaiser	<i>Impersonal pronouns in Romance null-subject-languages</i>
09.30-10.00	Martin Elsig, Shana Poplack	<i>Sprachtheorie und grammatische Variation: Die Bildung von Fragesätzen im</i>

10.00-10.30	Uwe Vosberg	<i>Französischen The influence of extra-semantic factors on the choice of non-finite complement variants in Modern English</i>
10.30-11.00	Eva Berlage	<i>Cognitive complexity and horror aequi as determinants of grammatical variation: Marked vs. unmarked infinitives following transitive uses of help</i>
11.00-11.30	Kaffeepause	
11.30-12.00	Jóhannes Gísli Jónsson, Höskuldur Thráinsson, Thórhallur Eythórsson	<i>Syntactic Variation without Regional Dialects</i>
12.00-12.30	Gerson Klumpp	<i>Variation in Komi dialectal object marking</i>

Freitag 24.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
10.30-11.00	Faye Chalcraft	<i>Variation as lexical choice: have, got and the expression of possession</i>
11.00-11.30	Philip Wallage	<i>Jespersen's Cycle in Middle English: evidence from diachronic data for a Minimalist syntax of negation</i>
11.30-12.00	Katarina Klein	<i>Valenzvariation des Kontroll- und Halbmodalverbs versprechen</i>
12.00-12.30	René Schiering	<i>Von der Dokumentation zur grammatischen Beschreibung: Ruhrdeutsche Präpositionalphrasen</i>
12.30-13.00	Martin Neef	<i>Zehn Arten, Gelispele auszusprechen, und eine einheitliche (deklarative) Analyse hiervon</i>
13.00-13.30	Richard Wiese	<i>The grammar and typology of plural noun inflection in varieties of German</i>

---

Bernd Kortmann, Verena Haser  
Universität Freiburg i. Br.

**Modelling variation in English dialect syntax: New prospects for syntactic theory**  
Mi 13.45-14.45

---

Having suffered a long period of relative neglect, the study of morphosyntactic variation has come into its own in recent years, due in part to a growing awareness among both functionalists and formalists that studies of dialect morphosyntax are highly relevant to syntactic theorizing. Our talk will survey major developments in this area. Focusing on morphosyntactic variation in English, especially English dialect syntax, we will glance back at previous work in the field and provide a look ahead, sketching some fruitful avenues for future research. There are essentially two ways in which the study of morphosyntactic variation can pose a challenge to syntactic theories. First, specific constructions found in dialects but absent in "standard" languages are frequently left unaccounted for in mainstream frameworks. Second, the fact that a single speaker may possess a rich array of alternative constructions expressing the same meaning presents problems to some well-known syntactic models. We will have a closer look at those approaches that seem to be particularly well (or ill-) suited to model syntactic variation (both in the sense of specific phenomena found in dialects and in the sense of intraspeaker variation). Particular emphasis will be placed on the ways in which corpus studies and syntactic theorizing can complement each other. Approaches to be taken into consideration include generative syntax, construction grammar, and optimality theory. Most examples will be drawn from previous research and future projects conducted by the Freiburg research group on English dialect syntax.

- Kortmann, B. 2002. „New prospects for the study of dialect syntax: Impetus from syntactic theory and language typology.“ In: H. Bennis & S. Barbiers (eds.). *Syntactic microvariation*. Amsterdam, 185-213.
- Kortmann, B., T. Herrmann, L. Pietsch & S. Wagner. 2005. *A comparative grammar of English dialects: Agreement, gender, relative clauses*. Berlin/New York: Mouton de Gruyter.
- Kortmann, B. (ed.). 2004. *Dialectology meets typology: Dialect grammar from a cross-linguistic perspective*. Berlin/New York: Mouton de Gruyter.

---

Bert Cappelle  
K. U. Leuven

***Clean up your room or clean your room up: is free choice an illusion?***

Mi 14.45-15.15

---

*Background.* The position of verbal particles in English has been a puzzle for at least over a century. It is by now generally agreed, though, that particle placement does not exhibit completely free variation. Specifically, information structure plays an important part (cf., e.g., Dehé 2002, Gries 2003). In accordance with general word order principles, light and accessible objects typically precede the particle (even obligatorily so in the case of unstressed pronouns; cp. *pick {it up / \*up it}*), while heavy objects typically follow the particle (cp. *pick {up the book I'm reading / ?the book I'm reading up}*). In an attempt to capture both the similarities and the dissimilarities between the two patterns, Cappelle (to appear) proposes a model in which the two orderings are seen as 'allostructions', that is, as different structural realizations of an information-structurally underspecified transitive verb-particle construction. Neither ordering is considered as derived from the other, nor are the two orderings completely isolated entities, since they are linked via a common hyperonymic construction.

*Goal/Approach.* In the present paper, I further explore how Cappelle's model can accommodate lexical preferences (e.g. *put {up a fight / \*a fight up}*; *put {\*together two and two / two and two together}*) as well as higher syntactic and pragmatic generalizations. I then outline a sorting task and a priming experiment designed to find out whether or not speakers perceive the two orderings as related. These experiments hinge on the use of a superficially similar 'decoy' pattern with a preposition (e.g. *turn on the radio / turn the ladies on / turn [on one leg]*).

*Conclusion.* The proposed model sheds light on the appearance of free choice between the two discussed orderings. I claim that this appearance is caused by the joint fact that the two orderings are mentally related in the grammar and that they are not in complementary distribution. Thus, in principle (i.e. barring lexical restrictions, provided the NP is nonpronominal and relatively light, etc.), both orderings can occur when there is focus on the entire VP (e.g. *What did you do today? – Well, I cleaned {up my room / my room up}*) or on the NP only (e.g. *What did you clean up, your room or the kitchen? – I cleaned {up my room / my room up}*). Free choice may not be an illusion here.

Cappelle, Bert. To appear. Particle placement and the case for 'allostructions'. *Constructions* [Special issue with selected papers of the Third International Conference on Construction Grammar, Marseille, 7-10 July 2004].

Dehé, Nicole. 2002. *Particle Verbs in English: Syntax, Information Structure, and Intonation*. (Linguistik Aktuell/Linguistics Today, 59.) Amsterdam and Philadelphia: John Benjamins.

Gries, Stefan Thomas. 2003. *Multifactorial Analysis in Corpus Linguistics: A Study of Particle Placement*. (Open Linguistics Series.) London: Continuum.

---

Said Sahel  
Universität Bielefeld  
**Was steuert die Variation der Adjektivflexion nach  
Pronominaladjektiven?**  
Mi 15.15 – 15.45

---

Gegenstand des Beitrags ist ein Fall innersprachlicher Variation im Deutschen. Es geht um die Variation der Adjektivflexion nach bestimmten Lexemen wie *all-*, *einig-*, *manch-*, *sämtlich-*, *solch-*, *viel-*..., die im Folgenden Pronominaladjektive genannt werden. Die Auswertung eines aus Zeitungstexten bestehenden Korpus mit mehr als 100 Millionen Wortstellen hat ergeben, dass nach ein und demselben (flektierten) Pronominaladjektiv sowohl starke als auch schwache Adjektivformen auftreten. Allerdings handelt es sich hierbei nicht um eine völlig freie Variation. Vielmehr hat die Datenanalyse gezeigt, dass diese Variation von sprachsystematischen Faktoren wie Kasus und Numerus (des jeweiligen Pronominaladjektivs) gesteuert wird.

In dem Beitrag werden die Ergebnisse dieser Studie präsentiert und ihre Relevanz für die Debatte um die grammatiktheoretische Erfassung innersprachlicher Variation diskutiert.

- Berruto, G. 2004. The problem of variation. *The Linguistic Review* 21, 293-322.
- Klein, W. 1998. Ein Blick zurück auf die Varietätengrammatik. *Sociolinguistica* 12, 22-38.
- Paolillo, J. C. 2002. *Analyzing Linguistic Variation. Statistical Models and Methods*. Stanford: CSLI Publications.
- Sahel, S. (eingereicht). Variationssteuernde Faktoren der Adjektivflexion nach Pronominaladjektiven.
- Wiese, B. 2004. Zur Systematisierung der Schwankungen zwischen starker und schwacher Adjektivflexion nach Pronominaladjektiven. <http://www.ids-mannheim.de/gra/texte/wi5.pdf>, 2.2.2005.

---

Judith Berman  
Ruhr-Universität Bochum  
**Das Prädikativ als Quelle grammatischer Variation**  
Mi 16.15 – 16.45

---

Das Prädikativ weist im Deutschen verschiedene Besonderheiten auf. So können manche prädikativ gebrauchte Substantive mit oder ohne Artikel auftreten wie in *Er ist Arzt/ein Arzt*. Artikellos verwendet werden die prädikativen Substantive häufig als adjektivierte Substantive oder auch als Adjektive analysiert. Dies ist insofern gerechtfertigt, als sie typische substantivische Eigenschaften verlieren und andererseits Eigenschaften mit Adjektiven teilen. So können sie nur beschränkt attribuiert werden, die Negation kann mit *nicht* erfolgen, sie können adjektivtypisch modifiziert werden und wie prädikative Adjektive im Deutschen 'unflektiert', d.h. ohne Numerusmarkierung auftreten. Aus dieser Eigenart, weder der Wortart Substantiv noch der Wortart Adjektiv eindeutig zugeordnet werden zu können, ergeben sich verschiedene grammatische Variationen, wie z.B. bei der Negation in *Er kann kein/nicht Arzt werden* oder der Numerusmarkierung in *Wir sind Zeuge/Zeugen eines dramatischen Niedergangs*.

Noch adjektivischer erweisen sich prädikative 'Substantive' wie *schuld, angst, pleite, feind* usw. Sie sind nicht artikelfähig und haben teilweise eine andere Valenz als das entsprechende Substantiv. Im Mittelhochdeutschen konnten einige 'Substantive' in prädikativer Verwendung sogar kompariert werden und zeichneten sich dadurch noch deutlicher als Adjektive aus.

In der prädikativen Verwendung lässt sich somit eine fortschreitende Konversion vom Substantiv zum Adjektiv beobachten. Die Möglichkeit der Umkategorisierung hängt mit der Eigenschaft der Kopula zusammen, dem Prädikativ keine kategorialen Restriktionen aufzuerlegen. Das Prädikativ kann als NP, AdjP, AdvP, PP und als Satz realisiert werden. Einige Lexeme wie *quitt, schade* oder *los* kommen ausschließlich als Prädikativ vor und verwehren sich jeglicher Kategorisierung.

Die These, die in diesem Beitrag vertreten wird, ist somit, dass die kategoriale Unterdeterminiertheit der prädikativen Position es ermöglicht, dass Lexeme ihre wortarttypischen Eigenschaften verlieren können, was dann zur Entstehung von grammatischer Variation führt.



---

Ljudmila Geist  
Universität Stuttgart  
**Variation im Bereich der prädikativen Adjektive im Russischen**  
Mi 16.45-17.15

---

**Das Phänomen.** In der Grammatik des Russischen gibt es zwei konkurrierende Varianten des prädikativen Adjektivs. Das prädikative Adjektiv kann im Hinblick auf Kongruenz mit dem Subjekt in einer sog. Langform (LF) auftreten, die mit dem Subjekt in Genus, Numerus und Kasus kongruiert, oder in einer Kurzform (KF), die Genus und Numerus-Kongruenz, jedoch keine Kasus-Kongruenz aufweist, vgl. (1):

- |                     |                    |                  |   |
|---------------------|--------------------|------------------|---|
|                     | <i>byla</i>        | <i>prost-aja</i> | <i>/ prost-a.</i>                           |
| (1) <b>Zagadk-a</b> | Rätsel.FEM.-SG.NOM | war              | einfach-(LF)SG.FEM.NOM / einfach-(KF)SG.FEM |

**Analyse.** In der Literatur wurden einige Faktoren festgestellt, die die Wahl der KF begünstigen, es konnte bisher aber kein Zusammenhang zwischen diesen Faktoren identifiziert werden. Festgestellt wurde z.B., dass die KF im Gegensatz zur LF Komplemente im Akkusativ haben kann, nur mit der Kopula und nie mit Vollverben kombiniert wird, mit der Höflichkeitsform *Иы* ‚Sie‘ grammatische Kongruenz im Plural aufweist, und nicht die semantische Kongruenz im Singular wie die LF. Diese und weitere Unterschiede zwischen LF und KF lassen nach unserer Meinung auf kategoriale Unterschiede zwischen den Formen schließen und sind ein Zeugnis dafür, dass die KF eine verbale Kategorie ist, während die LF typische adjektivische und zum Teil auch nominale Eigenschaften aufweist (vgl. auch Corbett 2004). Die kategorialen Unterschiede zwischen KF und LF können in dem Modell der Distribuierten Morphologie (Halle & Marantz 1993) mithilfe von kategorialen Merkmalen nach Steinitz (1997) und Wunderlich (1996) dargestellt werden.

**Schlussfolgerung.** Die Unterschiede zwischen KF und LF zeugen unserer Meinung nach von einer Aufspaltung der hybriden [+V,+N]-Kategorie „Adjektiv“ in zwei separate Kategorien, wobei im kategorialen Kontinuum zwischen V und N sich die KFn in Richtung zum verbalen Pol, die LFn zum nominalen Pol entwickelt haben. Der Unterschied zwischen den beiden Formen ist somit als Reflex eines kategorialen Wandels innerhalb der Kategorie „Adjektiv“ anzusehen.

---

Kris Heylen  
K. U. Leuven

**The order of nominal subject and pronominal object in the German middle field. A case of free syntactic variation?**

Mi 17.15-17.45

---

Word order variation in the so-called ‘middle field’ of the German clause has been a long standing issue in syntax research. More particularly, syntacticians with different theoretical backgrounds have grappled with the complex interplay between grammatic and pragmatic factors that influence the relative order of the case-marked verb arguments. One case of variation in particular has been defying adequate modelling, viz. the relative order of a nominal subject and a pronominal object. The pronominal accusative or dative object can either precede or follow the nominal subject:

1. Gestern hat **ihn der Lehrer** gestraft / Gestern hat **der Lehrer ihn** gestraft.
2. Heute hat **ihm der Lehrer** geholfen / Heute hat **der Lehrer ihm** geholfen.

The variation has been discussed both within generative (e.g. Lenerz 1994; Müller 2002) and functional frameworks (e.g. Hoberg 1997), but up to now no factors have been identified that lead to a clear preference of one order over the other. The question then arises: Is this a genuine case of free syntactic variation?

In this study, we try to find out whether there are any factors at all, intra- or extra-linguistic, that do influence object-subject ordering. As typical with presumed cases of free variation, the two variants do not noticeably differ in meaning or grammaticality, so that traditional grammaticality judgements cannot differentiate between them. We therefore apply an innovative methodology relying on quantitative corpus analysis, whereby multivariate statistics are used to validate factor effects in actual usage data. Results indicate, among other things, that pronominal objects follow nominal subjects relatively more often in subordinate compared to main clauses, in spoken compared to written language and in German compared to Austrian and Swiss usage. We will argue that such interplay of intra- and extra-linguistic factors is quite natural in usage-based grammar models (Barlow & Kemmer 2000).

- Barlow, M. & S. Kemmer (ed.). 2000. *Usage-based models of language*. Stanford: CSLI.
- Hoberg, U. 1997. „Die Linearstruktur des Satzes“ In: G. Zifonun, L. Hoffmann & B. Strecker: *Grammatik der deutschen Sprache*. Berlin: de Gruyter, 1495-1577.
- Lenerz, J. 1994. „Pronomenprobleme“ In: B. Haftka (ed.): *Was determiniert Wortstellungsvariation?* Opladen: Westdeutscher Verlag, 161-173.
- Müller, G. 2002. „Harmonic Alignment and the Hierarchy of Pronouns in German“ In: H. J. Simon & H. Wiese (eds.) *Pronouns – Grammar and Representation*. Amsterdam: Benjamins, 205-232.

---

Gert De Sutter, Dirk Speelman, Dirk Geeraerts  
K. U. Leuven

**A multivariate approach to the description and modelling of word order variation**  
Mi 17.45-18.15

---

In the present paper, we argue that grammatical phenomena in general, and word order phenomena in particular, are essentially multivariate in nature, and that a systematic-empirical, quantitative, multivariate treatment is needed in order to describe and model grammatical variation adequately. This does not just involve getting the right data, or starting from articulate and appropriate linguistic models, but it crucially involves linking data to models (i.e. operationalising linguistic hypotheses in ways that make them empirically testable), and statistically analysing the data by means of sophisticated multivariate techniques.

These theoretical and methodological claims are demonstrated on the basis of a case study, which tackles one of the most intriguing cases of grammatical variation in Dutch, viz. word order variability in bipartite clause-final verb clusters, consisting of a past participle and an auxiliary verb:

[...] dat ik een boek gekocht<sub>PART</sub> heb<sub>AUX</sub> / heb<sub>AUX</sub> gekocht<sub>PART</sub>  
'[...] that I have bought a book'

Both word order variants *gekocht heb* and *heb gekocht* are equally grammatical, and do not seem to have any obvious semantic-pragmatic correlates.

Building on a large-scale representative corpus of contemporary Dutch, this study will try to find out (i) how different types of factors relate to each other in their effect on the choice of word order, and (ii) what the global effect and predictive power of these factors is. We will do this by statistically analysing the effect of 9 language-internal factors. To achieve that goal, a set of multivariate statistical techniques is used. The results of the analyses show among other things that the kind of word order variation at hand is *indeed* multivariate in nature; that, however, not all of the proposed factors have a unique contribution to the explanation and prediction of the variation; that the semantic factor in the model is the most powerful one; and that the model is able to explain and predict the lion's share of the variation.

---

Georg A. Kaiser  
Universität Konstanz  
**Impersonal pronouns in Romance null-subject languages**  
Do 9.00-9.30

---

In typological and generative approaches it is assumed that there is a strong – parametric – correlation between the null-subject property and the lack of impersonal (expletive) pronouns (Haider 2001). This correlation holds for *all* Romance null-subject languages whose standard varieties ignore any kind of expletives. However, studies of colloquial and dialectal varieties of some of these languages have provided evidence for the occurrence of (apparent) expletive subjects, as for example in Dominican Spanish (Henríquez Ureña 1939):

- (1) (a) *Ello* estaba lloviznando un poco  
it was drizzling a little  
(b) *Ello* hay muchos mangos este año  
it there-is many mangoes this year  
(c) *Ello* llegan guaguas hasta allá  
it arrive buses until there

The acceptability of those utterances, also found in dialects of Portuguese, Galician, Catalan, Occitan and Italian, is highly controversial. While they are categorically rejected by speakers of standard and other varieties, even speakers of dialects exhibiting expletives often accept them only with some reluctance.

In my talk I discuss how to get reliable data from speakers of these dialects. I will show that this is possible by using questionnaires and by analyzing corpora from colloquial speech. It turns out that the use of expletives in these varieties is highly restricted by specific grammatical and pragmatic factors and differs thus from the use of expletives in non-null-subject languages. Therefore there is no correlation between the emergence of such pronouns and a possible loss of the null-subject property of these languages. A crucial question is whether this analysis also holds for French, which changed from a null-subject into a non-null-subject language. The study of Old French corpora shows that Old French already exhibits (some) subject pronouns in impersonal constructions and it seems that the increasing use of these pronouns correlates to the (gradual) loss of the null-subject property.

- Haider, H. 2001. „Parametrisierung in der Generativen Grammatik.“ In: M. Haspelmath et al. (eds.), *Language Typology and Language Universals. An International Handbook. Volume 1*. Berlin: de Gruyter, 283-293.  
Henríquez Ureña, P. 1939. „Ello.“ *Revista de filología hispánica* 1, 209-229.

---

Martin Elsig, Shana Poplack  
Universität Hamburg, University of Ottawa  
**Sprachtheorie und grammatische Variation: Die Bildung von  
Fragesätzen im Französischen**  
Do 09.30-10.00

---

Thema des Vortrags ist die Variabilität in Entscheidungsfragen des Québec-Französischen. Vier syntaktische Varianten konkurrieren seit Jahrhunderten: Subjekt-Verb-Inversion, Frageintonation, die postverbale Partikel *tu* und satzinitiales *est-ce que*. Untersuchungen zum europäischen Französischen berichten von einer Einschränkung der Variabilität zu Gunsten der Intonation. Inversion kennzeichnet Förmlichkeit und Schriftsprache. Stigmatisierung verhindert die Ausbreitung von *tu*. Der Gebrauch von *est-ce que* ist eher selten.

Wir stellen die diachrone Entwicklung der Entscheidungsfragen im *Québécois* anhand von umgangssprachlichen Aufnahmen aus dem 19. Jh. (*Récits du français québécois d'autrefois*, Poplack & St-Amand 2002) und aus dem 20. Jh. (*Ottawa-Hull French Corpus*, Poplack 1989) vor. Die Ergebnisse widersprechen Beschreibungen zum europäischen Französischen: Im 19. Jahrhundert wurde Inversion in fast der Hälfte aller affirmativen Fragesätze verwendet (beschränkt auf Subjekte der 2. Person). Sie kennzeichnet einen umgangssprachlichen Stil und kann somit nicht der Grammatik eines *français standard* zugewiesen werden. Fragesätze mit anderen Subjekten favorisieren *tu*. Intonation korreliert mit Förmlichkeit.

Diachrone Veränderungen betreffen lediglich Häufigkeiten im Gebrauch, nicht aber die Konditionierung. Inversion schwindet auch in Québec. Allerdings entwickelt sich nicht die Intonation, sondern *tu* zur häufigsten Variante. Ein Ergebnis dieser Entwicklungen, das weitverbreiteten Beschreibungen von *transplanted dialects* widerspricht, ist ein strukturell komplexeres Fragesatzsystem als das der *Input*-Varietät. Die Ergebnisse deuten an, dass das nach Kanada übertragene Französische zunächst alle vier Varianten aufwies und dass Veränderungen erst nach der Trennung zwischen den Dialekten eintraten. Wir behaupten, dass sich das Fragesatzsystem des europäischen Französischen innovativ verhält und dass die Variabilität im *Québécois* eine Beibehaltung des ursprünglichen Systems ist.

Poplack, S. (1989). The care and handling of a mega-corpus, in: R. Fasold & D. Schiffrin (Hg.), *Language Change and Variation*, Amsterdam: Benjamins, 411-451.

Poplack, S.; St-Amand, A. (2002). A real-time window on 19<sup>th</sup> century vernacular French: *the Récits du français québécois d'autrefois*. Paper presented at CLA, University of Toronto.

---

Uwe Vosberg  
Universität Paderborn

**The influence of extra-semantic factors on the choice of non-finite  
complement variants in Modern English**

Do 10.00-10.30

---

The establishment of the English gerund as a sentential complement variant is unique among all European languages and has resulted in a re-arrangement of the entire system of complementation: Since Early Modern English we observe a long-term and general development leading to the replacement of *to*-infinitives by *-ing* forms. For some verbs (such as *decline*), however, the ‘Great Complement Shift’ has even been reversed, presumably for semantic reasons.

Large-scale analyses of historical as well as present-day electronic corpora now reveal that semantic tendencies are very often accompanied (either reinforced or weakened) by major (and potentially universal) extra-semantic factors. This paper examines three (morpho-) syntactic constraints likely to accelerate or delay the rise or fall of *-ing* complements after certain verbs, adjectives and verb-noun collocations: (1) The infinitival complement option will tend to be preferred to the gerundial one in environments where a complement of the subordinate clause is extracted (by relativization etc.) from its original position and crosses clause boundaries (cf. Hawkins 1999). (2) The complexity principle postulates a higher degree of explicitness for grammatical structures in contexts involving a heavier processing load. Discontinuous constructions (caused by material inserted between a matrix predicate and its complement) can be seen as one manifestation of an increased cognitive complexity and are very often found to trigger the (presumably more explicit) *to*-infinitive rather than the *-ing* form. (3) The *horror aequi* principle describes an aversion against the use of formally (near-) identical and (near-) adjacent (non-coordinate) grammatical elements or structures. Thus, *-ing* complements are particularly slow to develop in environments where they (immediately) follow a verb which is realized as an *-ing* form itself. Similar observations can be made for sequences of (mainly directly linked) *to*-infinitives.

---

Eva Berlage  
Universität Paderborn

**Cognitive complexity and *horror aequi* as determinants of grammatical variation: Marked vs. unmarked infinitives following transitive uses of *help***

Do 10.30-11.00

---

This paper focuses on the rivalry between marked and unmarked infinitives complementing transitive uses of *help*. It shows that the distribution can largely be explained in terms of two functional concepts, the Complexity Principle (cf. Rohdenburg 1996; 1999; 2000) and *horror aequi*, with the latter describing the tendency to avoid formally (near-)identical and (near-)adjacent grammatical elements or structures (cf. Rohdenburg 2003; Vosberg 2003). My paper argues that these two factors tend to overshadow existing stylistic and semantic tendencies.

The Complexity Principle, which correlates cognitive complexity with grammatical explicitness, is accounted for by the parameters length (cf. Hawkins 1990; 1994) and structure of the indirect object NP (cf. Rickford et al. 1995; Wasow 2002). Large-scale analyses of British and American English electronic corpora reveal that both length and structure can be interpreted as distinct determinants of infinitival variation. Their effects are strengthened where they co-occur as in the case of the 6-word NP in (1), which contains a relative clause.

- (1) An attempt was made **to help** those who lost their jobs first **to find** work elsewhere. (*The Guardian* 2000)

As regards *horror aequi*, my analyses illustrate that a succession of two marked infinitives is avoided where the intervening NP is relatively short as in (2).

- (2) A task force was set up **to help** those quarantined **buy** food. (*The Independent* 2003)

If, however, the NP is as complex as in (1), complexity effects are likely to outweigh *horror aequi*. Even though American English generally shows a lower degree of infinitival marking than British English, it is equally sensitive to the effects of cognitive complexity and *horror aequi*.

---

Jóhannes Gísli Jónsson, Höskuldur Thráinsson, Þórhallur Eythórsson  
University of Iceland, Reykjavík  
**Syntactic Variation without Regional Dialects**  
Do 11:30 – 12:00

---

In this paper we discuss the results of a recent pilot study of syntactic variation in Icelandic where 187 speakers (in three age groups) were tested on 29 syntactic constructions. This study suggests that there are very few regional varieties in Icelandic syntax or differences according to social group. This is not surprising in view of the relative homogeneity of the Icelandic linguistic community and the near absence of regional dialects. However, robust variation between age groups was found in six constructions, some of which are currently undergoing a rapid change (e.g. the so-called “extended progressive”: *Ég er ekki að skilja þetta* ‘I am not understanding this’). On the methodological side, the study indicates that written questionnaires are just as reliable in most cases as oral interviews in data elicitation. Thus, the advantages of written questionnaires can be used in future studies to get a good overview of syntactic variation in Icelandic.

The results of the Icelandic pilot study raise the following question which we will address: Can intra-language variation be analysed in the same way as inter-language variation, i.e. as a choice between different values of binary parameters? Studies comparing two closely related varieties of the same language offer something close to a laboratory situation where the smallest units of syntactic variation can be isolated. Clearly, this is more easily attained in languages with little variation and therefore quite limited clustering of variants. Thus, Icelandic provides an ideal testing ground for the view that all intra-language variation is parametric, a view that has been challenged by various authors (see e.g. Jónsson & Eythórsson 2005 and Barbiers 2005).

Barbiers, Sjef. 2005. Word order variation in three-verb clusters and the division of labour between generative linguistics and sociolinguistics. Leonie Cornips and Karen P. Corrigan (eds): *Syntax and Variation: Reconciling the biological and the social*, pp. 233-264. John Benjamins, Amsterdam.

Jónsson, Jóhannes Gísli & Þórhallur Eythórsson. 2005. Variation in subject case marking in Insular Scandinavian. *Nordic Journal of Linguistics* 28.



---

Gerson Klumpp  
Universität München  
**Variation in Komi dialectal object marking**  
Do 12.00-12.30

---

Komi is a Uralic language in Northern Russia. Differential object marking in Komi patterns roughly as follows: A discourse-new inanimate direct object is unmarked (nominative:  $-\emptyset$ ). A given inanimate object is marked for accusative 3SG ( $-s\ddot{o}$ ). A discourse-new animate object is unmarked or marked for accusative ( $-\ddot{o}s$ ). A given animate object is marked for accusative or accusative 3SG. All given objects can also be marked for accusative 2SG ( $-t\ddot{o}$ ), which expresses a hearer's acquaintance with the object referent.

In addition there is another direct object marking strategy in Komi dialects, namely dative case ( $-ly$ ) in Vym', Izhma, Luza-Letka and Kosa-Kama. With the exception of Vym' and Izhma, these dialects areas are not adjacent. From a historical perspective the question arises whether the dative object is a remnant from Proto-Komi or rather an innovation in several dialects. Interestingly, the frequency of dative marked objects is quite different across these dialects and also for different speakers of a single dialect. The use of dative-marked objects ranges from a marginal phenomenon to a widely applied object marking strategy in functions otherwise reported for the different accusatives.

Up to now the function of dative marked objects has remained unclear. Text-external criteria (e.g. animateness) do not seem to play a role. It is rather discourse-pragmatic reasons which are relevant. The different frequencies of dative objects in the different dialects/idioms point to different functional spectra. The research corpus consists of dialect texts, most of them from narrative folklore.

- Baker, R. 1985. *The Development of the Komi Case System: A dialectological investigation*. Mémoires de la Société Finno-Ougrienne 189. Helsinki.
- Baker, R. 1986. „The Role of Animacy in Komi Direct Object Marker Selection.“ *Ural-Altäische Jahrbücher N.F.* 6. 47–60.
- Batalova, R. M. 1975. *Komi-permyackaya dialektologiya*. Moskva.
- Saharova, M.A. & Sel'kov, N.N. 1976. *Izhemskij dialekt komi jazyka*. Syktyvkar.
- Zhilina, T.I. 1985. *Luzsko-letschij dialekt komi jazyka*. Moskva.
- Zhilina, T.I. 1998. *Vymskij dialekt komi jazyka*. Syktyvkar.

---

Faye Chalcraft  
University of Cambridge  
**Variation as lexical choice: *have*, *got* and the expression of possession**  
Fr 10.30-11.00

---

Dialects of English are replete with variation in the forms used to encode possessive meaning, *have*, *have got* and *got* among those available:

- (1)     a. I had a lovely maroon pushchair.  
       b. She'd got a coach-built Silver Cross pram.  
       c. Ours is better than the one they got.

While previous research has addressed such variation from a sociolinguistic viewpoint (Tagliamonte 2003), or has considered the status of particular variants in isolation (e.g. LeSourd 1976), this paper uses data from a northern dialect of British English to consider how they arise as part of competence.

In the Chomskyan paradigm, interlingual variation is said to result from contrasting parameter settings. Intrapersonal variation is often handled in like fashion, yielding accounts involving optional features which are effectively versions of the competing grammars hypothesis (Kroch 1994). Stressing the value of a holistic approach which combines sociolinguistic and generative insights, this paper argues for a view on which variation reduces to lexical choice: speakers have just one grammar with an inventory of competing lexical items. In (1), there is a choice between two main verb synonyms: *have*, which behaves in the expected fashion (cf. *Do you have any wool?*) and *got*, which is anomalous both in being supported by *have* (cf. *Have you got any wool?*) and *do* (cf. *Do you got any wool?*), and in having two allomorphs, one phonologically contentful, the other phonologically null (cf. *Have you any wool?*).

LeSourd, P. 1976. "Got insertion." *Linguistic Inquiry* 7: 509-516.

Kroch, A. 1994. "Morphosyntactic variation." *Papers from the 30th regional meeting of the Chicago Linguistics Society*, 180-201.

Tagliamonte, S. 2003. "Every place has a different toll: Determinants of grammatical variation in cross-variety perspective." In: G. Rohdenburg and B. Mondorf (eds.) *Determinants of Grammatical Variation in English*. Berlin: Mouton de Gruyter, 531-554.

---

Philip Wallage  
 University of Manchester  
**Jespersen's Cycle in Middle English: evidence from diachronic data for  
 a Minimalist syntax of negation**  
 Fr 11.00-11.30

---

The change in the expression of sentential negation known as Jespersen's Cycle is seen in early English.

Stage One: Negation is expressed by a single negative marker (*ne*):

- |     |                     |     |           |      |                     |
|-----|---------------------|-----|-----------|------|---------------------|
| (1) | we                  | ne  | mugen þat | don  |                     |
|     | we                  | NEG | can       | that | do                  |
|     | 'we cannot do that' |     |           |      | (CMTRINIT,103.1370) |

Stage Two: Negation is expressed using two negative markers (*ne...not*):

- |     |                     |     |     |     |       |                      |
|-----|---------------------|-----|-----|-----|-------|----------------------|
| (2) | I                   | ne  | may | nat | denye | it                   |
|     | I                   | NEG | may | not | deny  | it                   |
|     | 'I may not deny it' |     |     |     |       | (CMBOETH,435.C1.262) |

Stage Three: *ne* is lost. Negation is marked by a single negator (*not*).

- |     |                           |      |     |     |       |                     |
|-----|---------------------------|------|-----|-----|-------|---------------------|
| (3) | I                         | know | nat | the | cause |                     |
|     | I                         | know | not | the | cause |                     |
|     | 'I do not know the cause' |      |     |     |       | (CMMALORY,627.3549) |

This paper proposes a syntactic account of Jespersen's Cycle using Minimalist syntactic theory and a grammar competition model of change (Kroch 1989). The account uses quantitative data from the PPCME2 corpus (Kroch & Taylor 2000). Patterns within these diachronic data demonstrate how competing grammars relate to each other, hence inform an account of the parametric differences between the competing grammars. Different syntactic analyses make different parametric options available, which entail empirical differences in the structure of competition over time. Frisch (1997) analyses the loss of *ne* and the introduction of *not* as independent changes, which intersect at stage two. However, this model does not account for observed variation in the diachronic data. I propose a different account. The change from *ne* > *ne...not* is a precondition for the loss of *ne*. *ne* is eliminated only when it is redundant to the expression of negation. The diachronic data support this model.

This model requires two distinct types of *ne*: one a negative marker, the other redundant, which are subject to different changes. The Minimalist notions of valued and unvalued morphosyntactic features make the required distinction. This allows the dependency between *ne* and *not* at stage two to be parameterised, reducing the Neg-criterion to a feature checking dependency, with implications for the phrase-structure status of the functional projection NegP.

Frisch, S. (1997) "The Change in Negation in Middle English: A NegP licensing account". *Lingua* 101, 21-64  
 Jespersen, O. (1917) *Negation in English and other languages*. København: Enjar Munksgaard.  
 Kroch, A.S (1989) "Reflexes of Grammar in Patterns of Language Change". *Language Variation and Change* 1, 199-244.  
 Kroch, A.S & A. Taylor (2000) *The Penn-Parsed Corpus of Middle English* (second edition). Philadelphia: Department of Linguistics, University of Pennsylvania.

---

Katarina Klein  
Ruhr-Universität Bochum  
**Valenzvariation des Kontroll- und Halbmodalverbs *versprechen***  
Fr 11.30-12.00

---

Das Verb *versprechen* (ähnlich: *drohen*) kann nicht nur als Subjektkontrollverb (1), sondern auch halbmodal verwendet werden (2).

(1) Er versprach (ihm), weitere Mitarbeiter einzustellen.

(2) Die Einstellung von weiteren Mitarbeitern versprach Abhilfe zu schaffen.

Anders als Anhebungsverben restringieren Kontrollverben ihr Subjekt sowohl syntaktisch als auch semantisch. Die Halbmodalverben *drohen* und *versprechen* nehmen eine Position zwischen Anhebung und Kontrolle ein (Kiss 2005). Sie gehen – wie Anhebungsverben – keine thematische Beziehung zu ihrem Subjekt ein (3), beschränken aber – wie Kontrollverben – ihr Subjekt syntaktisch: Die Einbettung von subjektlosen Verben ist zumindest markiert (vgl. Gunkel 2000); die Einbettung eines unpersönlichen Passivs nicht zugelassen (4).

(3) Es scheint/verspricht/\*hofft hier Ratten zu geben.

(4) Hier scheint/\*verspricht/\*hofft gearbeitet zu werden.

Das halbmodale *versprechen* lässt wahlweise ein infinites oder ein nominales Komplement zu (5). Das Vollverb erlaubt weitere, sententiale Varianten. Die Valenzvariation von *drohen* ist komplexer (vgl. Askedal 1997).

(5) a. Die Regierung in Algier hat allen Untergrundkämpfern eine Amnestie versprochen.

b. Beide Kandidaten legen Dossiers vor, die ein Turnier in neuen Dimensionen versprechen.

Korpusuntersuchungen weisen darauf hin, dass das halbmodale *versprechen* fast dreimal so häufig mit nominalem Komplement verwendet wird wie mit infinitem Komplement, Kontrollverben dagegen nur 1,5-mal so oft (vgl. Kiss 2005). Entscheidend für die Interpretation ist, unabhängig von der Valenz, nicht allein die Intentionalität des eingebetteten Ereignisses (Gunkel 2000) oder die Belebtheit des Subjekts, sondern das Zusammenspiel verschiedener Faktoren. In einer möglichst redundanzfreien Grammatik sollten die genannten Varianten daher voneinander abgeleitet werden.

Askedal, J. O. 1997. „*drohen* und *versprechen* als sogenannte ‚Modalitätsverben‘ in der deutschen Gegenwartssprache.“ *Deutsch als Fremdsprache* 34:1, 12-19.

Gunkel, L. 2000. „Selektion verbaler Komplemente. Zur Syntax der Halbmodal- und Phasenverben.“ In: R. Thieroff et al. (Hg.) *Deutsche Grammatik in Theorie und Praxis*. Tübingen: Niemeyer, 111-121.

Kiss, T. 2005. „Subjektselektion bei Infinitiven.“ In: J. F. Marillier & C. Rozier (Hg.) *Der Infinitiv im Deutschen*. Tübingen: Stauffenburg, 115-132.

---

René Schiering  
Universität Konstanz  
**Von der Dokumentation zur grammatischen Beschreibung:  
Ruhrdeutsche Präpositionalphrasen**  
Fr 12.00-12.30

---

Das Ruhrdeutsche fällt bei der Kasuszuweisung in Präpositionalphrasen durch ein erhöhtes Maß an morphosyntaktischer Variation auf (Mihm 1982). In den Beispielen (1) bis (3) variieren die Verschmelzungsformen *ins*, *in'n* und *im* bei der direktionalen Präposition *in* mit Nomina im Neutrum. In (4) und (5) steht die Form *in'n* im Kontext derselben Präposition mit maskulinen Nomina neben der Form *im*.

- (1) Und dann sind wir auch pünktlich zum Mittachessn **ins** Dorf eingefahrn.
- (2) Braun' wa nie **in'n** Cafe gehn.
- (3) Ich spring do au **im** Wasser un kann schwimmen.
- (4) Komm, wir gehn **in'n** Taubmschlach.
- (5) Auf eima krich' ich 'n Strahl **im** Arsch rein.

Unstrittig können diese und ähnliche Phänomene als ein Symptom diastratischer und diatopischer Variation verstanden werden. Darüber hinaus stellt sich dennoch die Frage, ob sich auf der Grundlage dieser Daten auch ein grammatisches System modellieren lässt. In diesem Beitrag soll eine Methodologie in der Erforschung solcher Fragen diskutiert werden. Die Untersuchung geht dabei von einer weit gefassten Dokumentation der regionalen Umgangssprache im Ruhrgebiet aus, die mehrere Altersgruppen aus zwei Städten umfasst. In einem zweiten Schritt werden die variationssteuernden Faktoren Alter und Sprechsituation bestimmt und die Variation in dieser Hinsicht eingegrenzt. In einem dritten Schritt werden die konkurrierenden Varianten in ihrer Häufigkeit erfasst. Durch eine Konzentration auf die hochfrequenten Formen und eine situationsabhängige Einordnung der minderfrequenten Formen lässt sich für die älteren Sprecher in nicht-öffentlichen Sprechsituationen das folgende System präpositionaler Flexion etablieren (Schiering 2005). Die Präposition kongruiert in Bezug auf die Merkmale Genus und Numerus mit dem regierten Nomen. Bei Maskulina und Neutra im Singular wird dabei die Form *-m* an den präpositionalen Stamm suffigiert, bei Feminina im Singular und bei Pluralen aller Genera die Form *-e*. Dieses System bildet ein Extrem auf dem Dialekt-Standard-Kontinuum, dessen Relevanz im Kontext des Ruhrdeutschen abschließend diskutiert werden soll.

Mihm, Arend 1982. „Zur Entstehung neuer Sprachvarietäten: Ruhrdeutscher Kasusgebrauch und seine Erklärung“. *Zeitschrift für Germanistische Linguistik* 10: 263–294.

Schiering, René 2005. „Flektierte Präpositionen im Deutschen? Neue Evidenz aus dem Ruhrgebiet.“ *Zeitschrift für Dialektologie und Linguistik* 72: 52–79.

---

Martin Neef  
Technische Universität Braunschweig  
**Zehn Arten, *Gelispel* auszusprechen, und eine einheitliche (deklarative)  
Analyse hiervon**  
Fr: 12.30-13.00

---

Eine Variationsgrammatik sollte idealerweise den Umstand widerspiegeln, dass wir intuitiv zwischen höheren und niedrigeren Stilregistern unterscheiden können. In einer deklarativen Konzeption wird die Grammatikalität eines sprachlichen Ausdrucks durch unverletzliche Beschränkungen modelliert, wobei eine grammatische Kategorie dadurch gekennzeichnet ist, welchen Beschränkungen sie unterliegt. In diesem Vortrag möchte ich zeigen, wie eine grammatische Kategorie, die über variierende Realisierungsformen verfügt, in deklarativer Weise so analysiert werden kann, dass die stilistisch höhere Version mehr Beschränkungen unterliegt als die stilistisch niedrigere Version, dass zugleich aber alle Beschränkungen, die für niedrigere Versionen gelten, auch für höhere Versionen bindend sind. Sich eines einfacheren Registers zu bedienen bedeutet damit, weniger Bedingungen zu erfüllen.

Als Gegenstand dient die durch Wörter wie *Gehopse*, *Gejammere* und *Gelispel* charakterisierte Lexembildungskategorie. Nach einer morphembasierten Analyse besteht das Wort *Gelispel* aus dem Wurzelmorphem *lispel* und dem derivierenden Zirkumfix *Ge...e*. In einer Realisierungsvariante desselben Worts kann ein Teil des Affixes wegfallen, und zwar erstaunlicherweise jenes, das unter wortsyntaktischer Perspektive als Kopf fungiert, nämlich das finale *e* in der Form *Gelispel*. In einer anderen Realisierungsvariante kann ein Teil des Wurzelmorphems entfallen, wie die Form *Gelispel* belegt. Wenn man Morphologie nicht als Buchstabenwissenschaft betreibt, sondern auf gesprochene Formen bezieht, kann man erkennen, dass sich hinter diesen drei schriftsprachlichen Varianten zehn unterschiedliche sprechsprachliche Varianten verstecken. Für entsprechende Bildungen zu anderen Verbstämmen sind weniger Varianten möglich bis hin zu Fällen wie *Gezeichne*, die nur in einer einzigen Art realisiert werden können. Die Analyse der teilweise phonologisch, teilweise morphologisch bedingten Variation ergibt eine Hierarchie von unterschiedlich komplexen Realisierungsformen.

---

Richard Wiese  
Universität Marburg

**The grammar and typology of plural noun inflection in varieties of  
German**

Fr 13.00-13.30

---

This presentation discusses a central sub-class of noun pluralisation in German. By analysing the much-discussed plural morphology of Standard German within a wider range of varieties of German, and by making use of the adequate division of labour between morphology and phonology, the analysis establishes the status of final plural schwa as in *Fisch-e* ‘fish, PL’. On the basis of several empirical results, I first argue that final schwa in German plural nouns is not, as generally assumed, an inflectional suffix. Rather, an optimality-theoretic constraint-based analysis of final schwa in plurals leads to the conclusion that this segment in noun plurals of Standard German arises as a vowel of epenthesis, as the result of a specific constraint interaction.

In a second part of this talk, noun pluralisation is studied in a few diverse dialects of German. Morphological and prosodic constraints, through the well-known mechanism of differences in constraint-ranking in Optimality Theory, derive the appearance or non-appearance of word-final plural schwas in these dialects. The constraints will include those which refer to properties of whole paradigms of word forms, not just to phonological properties of individual words. As an over-all descriptive result, a micro-typology of plural formation in varieties of German emerges.

## **AG 8 Subordination' vs. ‚Koordination' in Satz und Text aus sprachvergleichender Perspektive / 'Subordination' vs. 'coordination' in sentence and text from a cross-linguistic perspective**

Koordination: Cathrine Fabricius-Hansen, Wiebke Ramm

### **Programm**

Mittwoch 22.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
13.45-14.15	Cathrine Fabricius-Hansen	<i>Begrüßung und Einführung / Welcome and introduction</i>
14.15-14.45	Laurence Delort	<i>Clause Subordination and Discourse Relations</i>
14.45-15.15	Christelle Cosme	<i>A corpus-based perspective on clause linking patterns in English, French and Dutch</i>
15.15-15.45	Mary Carroll, Antje Roßdeutscher, Christiane v. Stutterheim	<i>Subordination in Erzählungen. Ein Beitrag zum Sprachvergleich</i>
15.45-16.15	Kaffeepause	
16.15-16.45	Nicole Baumgarten	<i>Text-forming in a cross- linguistic perspective: Hierarchical and non- hierarchical discourse structuring in English and German texts</i>
16.45-17.15	Kåre Solfeld	<i>Satzteilung in der Übersetzung – Beibehaltung der Diskursstruktur</i>
17.15-17.45	Wiebke Ramm	<i>Satzteilung in der Übersetzung – Veränderung der Diskursstruktur?</i>
17.45-18.15	Manfred Stede	<i>RST Revisited: Disentangling Nuclearity</i>



Donnerstag 23.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
09.00-09.30	Joyce E. Stavick	<i>Subordinating Narrative Design for Syntactic Simplicity: An Examination of Punctuation, Subordination, and Coordination in Two Editions of Graham Greene's The Power and the Glory</i>
09.30-10.00	Michael Franke	<i>Pseudo-Imperatives: A Case-Study in the Ascription of Discourse Relations</i>
10.00-10.30	Ingo Reich	<i>From Discourse to "Odd Coordinations" – On Asymmetric Coordination and Subject Gaps in German</i>
10.30-11.00	Rosemarie Lühr	<i>Sätze zwischen Subordination und Koordination im Altindischen</i>
11.00-11.30	Kaffeepause	
11.30-12.00	Maria Averintseva	<i>To the right of the clause: right dislocation vs. afterthought</i>
12.00-12.30	Anke Holler	<i>Detached in Syntax – Attached in Discourse</i>

Freitag 24.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
10.30-11.00	Svetlana Petrova, Michael Solf	<i>Rhetorical Relations an Verb Placement in Early Germanic Languages. A Cross-Linguistic Study</i>
11.00-11.30	Olav Hackstein	<i>Another look at main-clause phenomena in German subordinate clauses</i>
11.30-12.00	Hardarik Blühdorn	<i>Adpositionen, Adverbien, Subjunktionen und Konjunktionen. Die Interaktion syntaktischer, semantischer und textueller Verknüpfungseigenschaften in vier unterschiedlich grammatikalisierten Konnektorklassen</i>
12.00-12.30	Sara Wischer	<i>Die Konzession als Diskurs-Relation</i>
12.30-13.00	Angelika Wöllstein	<i>Kontrafaktizität als semantisch-konzeptuelle Basis der Satzkonnexion</i>
13.00-13.30		<i>Abschlussdiskussion</i>

---

Cathrine Fabricius-Hansen  
Universität Oslo  
**Einführung/Introduction**  
Mi 13.45-14.15

---

Die Unterscheidung zwischen ‚Unterordnung‘ und ‚Nebenordnung‘, die in der Syntax, d.h. auf die Satzstruktur bezogen, eine lange Geschichte hinter sich hat, ist in neueren diskurstheoretischen Ansätzen auf die Beschreibung von Diskurseinheiten und Diskursrelationen übertragen worden. Die Beziehungen zwischen den diskurs- und den satzbezogenen Begriffspaaren sind jedoch relativ wenig erforscht. Diese Beziehungen und die sprachliche Kennzeichnung ‚subordinierender‘ und ‚koordinierender‘ Diskursrelationen bilden das Thema der Sprachen vergleichend orientierten AG 8. In der Einführung sollen zentrale Fragestellungen kurz vorgestellt und die Struktur der AG kommentiert werden.

In my introduction, I shall briefly present the main issues motivating this workshop and outline the structure of the workshop.

The presentation will be in German (accompanied by slides/handout in English) or in English, depending on the demands of the audience.

---

Laurence Delort  
Univ. Paris 7  
**Clause 'Subordination' and Discourse Relations**  
Mi 14.15-14.45

---

It's tempting to correlate clause combining and discourse structure because both display subordination and coordination. Particularly, it seems assumed in (Matthiessen & Thompson, 1988) that subordination at the clause level (a subordinating conjunction) lexicalizes subordination at the discourse level (a subordinating relation in Segmented Discourse Representation Theory, (Asher & Lascarides, 2003), or a nucleus-satellite relation in Rhetorical Structure Theory, (Mann & Thompson, 1988)).

The goal of this study is to draw a distinction between clause 'subordination' and discourse relations, by examining two French subordinating conjunctions: *avant que* (English *before*) and *parce que* (English *because*). They usually introduce subordinate clauses which denote circumstances of eventualities described in main clauses. That is, they convey temporal or causal relations which occur at the conceptual level, not at the discourse level. On the other hand, they may not introduce subordinate clauses, but main clauses, (Green, 1976). In this case, they convey (structural) discourse relations which occur at the discourse level: *avant que* triggers the coordinating

relation *Narration*, (Le Draoulec, 2005; Delort, 2006), and *parce que* triggers the subordinating relation *Explanation*, (Debaisieux, 2005; Delort & Danlos, 2005). It seems that there is no parallel between clause 'subordination' and discourse subordination, because a subordinating conjunction can express either a conceptual relation or a discourse relation, and if a discourse relation is conveyed, it can be either coordinating or subordinating.

- Asher, N. & Lascarides, A. (2003). *Logics of Conversation*. Cambridge University Press.
- Debaisieux, J.-M. (2005). Les conjonctions de subordination: mots grammaticaux ou mots de discours? Le cas de *parce que*. *Revue de Sémantique et de Pragmatique*. Forthcoming.
- Delort, L. (2006). *Avant que* en contexte discursif. *Cahiers Chronos*. Forthcoming.
- Delort, L. & Danlos, L. (2005). Coordination of Causal Relations in Discourse. In *Proceedings of the Symposium on the Exploration and Modelling of Meaning (SEM-05)*, p. 75-84, Biarritz, France.
- Green, G. M. (1976). Main Clause Phenomena in Subordinate Clauses. *Language*, 52(2), 382–397.
- Le Draoulec, A. (2005). *Avant que/de*: Possibles passages à la connexion temporelle. *Journal of French Language Studies*, 15(2), 131–151.
- Mann, W. C. & Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. *Text*, 8(3), 243–281.
- Matthiessen, C. & Thompson, S. A. (1988). The Structure of Discourse and 'Subordination'. In J. Haiman & S. A. Thompson (eds), *Clause Combining in Grammar and Discourse*, volume 18 of *Typological Studies in Language*, p. 275–329. Amsterdam/Philadelphia: John Benjamins.
- 

Christelle Cosme

Univ. catholique de Louvain, Belgique

## **A corpus-based perspective on clause linking patterns in English, French and Dutch**

Mi 14.45-15.15

---

The two traditional clause linking devices, viz. coordination and subordination, are very interesting to study from a cross-linguistic perspective. Principles of information packaging, of which clause linking forms an integral part, have indeed been claimed to be – at least partly – language-specific (Hasselgård *et al.* 2002). Fabricius-Hansen (1998), for instance, observes that some languages (e.g. German) favour a hierarchical (i.e. vertical) discourse information structure making intensive use of subordinating devices, while other languages (e.g. English and, to a larger extent, Norwegian) prefer an incremental (i.e. horizontal) discourse information structure, thus favouring coordinating devices.

This paper takes a cross-linguistic view on sentence-level coordination and subordination in English, French and Dutch. The aim is to test a number of intuition-based hypotheses made in the contrastive literature. Chuquet & Paillard (1987:151), for instance, point out that the explicit linking of clauses is often realized through coordination in English and through subordination in French (see Hannay & Mackenzie (1996) for similar claims for English and Dutch). The claims will be

tested empirically on the basis of a multilingual parallel corpus. Shifts from coordination to subordination and, conversely, from subordination to coordination will be examined in three translation directions: (a) from English to French; (b) from French to English and (c) from English to Dutch (see Cosme (forthcoming) for the corpus findings). Another implicit aim of the paper is to gain new insights into the relationship between coordination and subordination. It will be argued, in line with Cristofaro (2003) among others, that clause linking is best treated as a gradient phenomenon.

- Chuquet, H. & Paillard, M. 1987. *Approche linguistique des problèmes de traduction anglais >< français*. Paris: Gap / Ophrys.
- Cosme, C. forthcoming. Clause combining across languages. A corpus-based study of English-French translation shifts. To appear in: *Languages in Contrast*.
- Cristofaro, S. 2003. *Subordination*. Oxford: Oxford University Press.
- Fabricius-Hansen, C. 1998. Informational Density and Translation, with Special Reference to German-Norwegian-English. In: S. Johansson & S. Oksefjell (eds) *Corpora and Cross-linguistic Research: Theory, Method, and Case Studies*. Amsterdam & Atlanta: Rodopi, 197-234.
- Hannay, M. & Mackenzie, J. 1996. *Effective writing in English: a resource guide*. Groningen: Martinus Nijhoff.
- Hasselgård, H., Johansson, S., Behrens, B. & Fabricius-Hansen, C. 2002. *Information Structure in a Cross-Linguistic Perspective*. Amsterdam & New York: Rodopi.

---

Mary Carroll, Antje Rossdeutscher, Christiane von Stutterheim  
Univ. Heidelberg  
**Subordination in Erzählungen. Ein Beitrag zum Sprachvergleich**  
Mi 15.15-15.45

---

Ziel des Vortrags ist es, die unterschiedlichen Befunde über Motivation und Umfang von Subordination als Folge von sprachspezifischen globalen Strategien zur Bewältigung von Erzählaufgaben zu erklären. Die empirische Basis ist ein Erzählkorpus mit den Sprachen Deutsch, Englisch, Französisch und Italienisch. Wir verstehen diese Strategien als durch die syntaktischen Strukturen der Einzelsprachen motiviert, die je andere Hierarchien von zu treffenden Entscheidungen der Selektion, Linearisierung und Perspektivierung mit sich bringen. Erfordert die grammatische Struktur einer Sprache für jede Satzprädikation eine Entscheidung, so treffen die Sprecher diese Entscheidung global. Relevante Faktoren, die die Strategien bestimmen, sind (i) freie Wortstellung, Verbzweit vs. feste Wortstellung (ii) Grammatikalisierung von Aspekt, +,-. Deutsch, Französisch und Englisch instantiiieren diese Unterschiede.

In dem Vortrag sollen zunächst kurz die sprachspezifischen globalen Strategien erläutert werden. Dann soll gezeigt werden, wie die Subordination funktional als grammatisches Mittel eingesetzt wird, um den Erhalt global gesetzter Topikvorgaben auf der Ebene der assertierten Äußerungen (Hauptstruktur) systematisch zu gewährleisten.

Zur Erklärung der sprachspezifischen Strategien soll ein Entscheidungsmodell präsentiert werden, das aus einer Menge von Ereignissen unter der Annahme unterschiedlicher Hierarchien von zu treffenden Entscheidungen die Befunde zu rekonstruieren erlaubt.

---

Nicole Baumgarten  
Univ. Hamburg

**Text-forming in a cross-linguistic perspective: Hierarchical and non-hierarchical discourse structuring in English and German texts**

Mi 16.15-16.45

---

The paper presents work in progress from the project Covert Translation currently carried out inside the Collaborative Research Center on Multilingualism (Sonderforschungsbereich "Mehrsprachigkeit") at the University of Hamburg. The project investigates whether ongoing processes of globalization and internationalization and the concurrent status of the English language as a global *lingua franca* in key areas of public life such as science and economy impact on language specific norms and communicative conventions of translations from English into German and monolingual text production in German.

According to e.g. Fabricius-Hansen 1999, English and German differ in their preferences for discourse structuring, with German favouring hierarchical and English non-hierarchical patterns. Evidence from diachronic-contrastive qualitative and quantitative analyses of translation and comparable text corpora of popular scientific texts and business communication suggests that English-German language contact influences patterns of text-forming in German translations and comparable texts in German.

By the example of the use of the sentence-initial *and* and *und* and the use of *ing*-participle clauses and their translations in German texts, I will, first, describe the ways in which these preferences have developed over the course of the past 25 years. While *and* and *und* are roughly equivalent linguistic means in English and German, the *ing*-participle does not have a corresponding element in German. Thus, secondly, I will discuss whether a shift in German conventions of discourse structuring in the direction of the ones preferred in English texts is dependent on the presence of equivalent linguistic means in the two languages.

Fabricius-Hansen, C. 1999. Information packaging and translation: Aspects of translational sentence splitting (German - English/Norwegian). In: Doherty, M. (ed.): *Sprachspezifische Aspekte der Informationsverteilung*. Berlin. 1999. 175-214.

---

Kåre Solfjeld  
Hochschule in Østfold, Halden, Norwegen  
**Satzteilung in der Übersetzung – Beibehaltung der Diskursstruktur**  
Mi 16.45-17.15

---

Die typische Informationsdichte deutscher Sachprosa löst beim Übersetzen ins Norwegische oft eine Satzteilung aus (Fabricius-Hansen 1999, Solfjeld 2000), wobei Information in Adjunkten auf VP- oder NP-Ebene im Original in der Zielversion koordiniert – als Sequenz selbständiger Sätze oder als Konjunkte einer koordinativen Struktur – auftritt. Ein wichtiges Signal zur Kennzeichnung der Information als diskursmäßig subordiniert (keine Antwort auf die ‘Quaestio’ des Textes, von Stutterheim 1997) geht somit verloren. Ausgehend von authentischen Übersetzungen Deutsch-Norwegisch wird gezeigt, dass die Wahl von Zielstrukturen als das Ergebnis von Strategien erklärt werden kann, die auf die Beibehaltung der Diskursstruktur des Originals abzielen. Explizite Signale zur Vermittlung der Diskursstruktur werden dann in der Zielversion eingebaut, wenn die Propositionen der geteilten Sätze Interpretationen erlauben, die vom Original abweichen. Die aus der Informationsspaltung nach links und rechts (Fabricius-Hansen 1999) resultierenden Satzteilungen bedingen verschiedene Strategien zum Erhalt der Diskursstruktur. Bei Informationsspaltung nach links werden z.B. die Inferenzmechanismen in Bezug auf temporale Folge- und/oder (zumindest schwache) Kausalrelationen bei koordinativen Strukturen (Blakemore 1987) ausgenutzt. Koordination erscheint somit als ein systematisches Mittel zur Kompensierung der geringeren Auswahl (oder Ausnutzung) subordinierender Strukturen im Norwegischen. Primäre theoretische Basis bildet von Stutterheim (1997). Zugleich werden Arbeiten innerhalb so genannter Relevanztheorie (Blakemore 1987, Carston 2002) und SDRT-Theorie (Asher 1993, Asher und Vieu 2005) herangezogen. Der Beitrag ist im Rahmen des SPRIK-Projekts, Univ. Oslo, entstanden.

Asher, N 1993. *Reference to Abstract Objects in Discourse*. Dordrecht: Kluwer

Asher, N & Vieu, L 2005. Subordinating und coordinating discourse relations. *Lingua* 115,4, 591-610.

Blakemore, D 1987. *Semantic Constraints on Relevance*. Oxford: Basil Blackwell.

Carston, R 2002. *Thoughts and Utterances. The Pragmatics of Explicit Communication*. Oxford: Blackwell

Fabricius-Hansen, C. 1999. „Information packaging and translation: Aspects of translational sentence splitting (German–English/Norwegian)“. In: M. Doherty (hg.) *Studia grammatica* 47 Berlin: Akademie-Verlag, 175-213.

Solfjeld, K 2000. *Sententialität, Nominalität und Übersetzung* (OBG 26) Frankfurt, Peter Lang.

von Stutterheim, C. 1997. *Einige Prinzipien des Textaufbaus. Empirische Untersuchungen zur Produktion mündlicher Texte* (RGL 184). Tübingen: Niemeyer.

---

Wiebke Ramm  
Univ. Oslo

**Satzteilung in der Übersetzung – Veränderung der Diskursstruktur?**  
Mi 17.15-17.45

---

In diesem Beitrag soll es um die Grenzen der in Solfelds Beitrag skizzierten Übersetzungs-strategien zur Kompensation fehlender/unterschiedlich ausgeprägter grammatischer Res-sourcen in der Zielsprache gehen. Typologische Kontraste machen oft eine Reorganisation der deutschen Satzstruktur in der norwegischen Übersetzung erforderlich, was zu Veränderungen der Segmentierung des Textes in Sätze bzw. Satzgefüge führen kann. Die Diskursrelationen zwischen den (Teil-)Sätzen, ändern sich dadurch nicht unbedingt, doch ändert sich oft die Art der referentiellen Bezüge bzw. ihre Erkennung, und damit ein konstitutives Kohärenzmerkmal eines Textes. Nicht-restriktive Relativsatzkonstruktionen (nomenbezogene wie auch Satzrela-tivsätze) im Deutschen sind ein typischer, typologisch bedingter, Auslöser für die Satzteilung im Norwegischen, d.h. die „Aufwertung“ eines syntaktisch subordinierten Satzes zu einem selbständigen Hauptsatz, ggf. als Element einer Satzkoordination oder einer durch Komma /Semikolon getrennten parataktischen Satzfolge. Charakteristisches Merkmal sog. „weiterfüh-render Relativsätze“ ist ihre Ähnlichkeit mit einer Hauptsatzfolge. Als Unterschied wird meist angeführt (z.B. Holler 2005:211), dass die Relativsatzkonstruktion kohärenter ist, da weniger Inferenzaufwand zur Bestimmung der anaphorischen Bezüge notwendig ist. Dies betrifft auch den nachfolgenden Satz, bei dem bei einer Hauptsatzfolge Diskursreferenten ggf. explizit wiedereingeführt werden müssen, da das Signal zur syntaktischen Hierarchisierung und Einheitsbildung im Vorgängersatz fehlt. Doch dafür enthält die Satzfolge auch weniger (ggf. komplex/hierarchisch strukturierte) Informationseinheiten pro Verarbeitungseinheit (also durch Punkt begrenztem Satz), der Einzelsatz ist also ggf. für den Hörer/Leser einfacher zu verstehen. Vor dem Hintergrund der Frage, inwieweit solche Hierarchisierungs-/Segmen-tierungsunterschiede sprachspezifische Strategien der Diskursorganisation reflektieren (Fabricius-Hansen 1999, Doherty 2002) soll diskutiert werden, wie Konzepte wie „Text-kohärenz“ oder „Textverständlichkeit“ in dieses Bild passen, bzw. ob auch diese ggf. relativ zu (teilweise typologisch motivierten) Diskursorganisationsstrategien zu evaluieren sind.

Doherty, M. 2002. Strategy of incremental parsimony. [Ms.], auch als SPRIKreport Nr. 25: <http://www.hf.uio.no/forskningsprosjekter/sprik/docs/pdf/md/MDohertyReport25.pdf>

Fabricius-Hansen, C. 1999. Information packaging and translation: aspects of translational sentence splitting (German – English/Norwegian). In M. Doherty (Hg.) *Sprachspezifische Aspekte der Informationsverteilung*. Berlin: Akademie-Verlag, 175-214.

Holler, A. 2005. *Weiterführende Relativsätze. Empirische und theoretische Aspekte*. Berlin: Akademie Verlag.



Among the theories of discourse structure, Mann & Thompson's Rhetorical Structure Theory is the most *empirical* and has attracted much attention in corpus-oriented text linguistics and in computational linguistics. It took the bold step of stipulating a set of coherence relations applicable to basically any text; in order to achieve this, the notions are defined only semi-formally at best, and by necessity stated on a high level of abstraction. This holds for the relations as well as for the notion of *nuclearity*. Mann & Thompson characterized the nucleus of a relation as being „more central to the writer's purposes“ and furthermore suggested as a diagnostic test the procedure of removing from a text all satellites; the resulting sequence of nuclei would still be a comprehensible text. This notion does have an intuitive appeal, but it hides the fact that „being more central to the writer's purposes“ can be a very different thing from case to case. In our annotation work with the *Potsdam Commentary Corpus*, we asked annotators on what grounds they had made their decision of nucleus assignment between adjacent text spans. These are the reasons they responded with:

- i) Recurrence: One of the two spans, or part of it, is taken up later in the text.
- ii) The connective explicitly signals the nuc/sat difference, for example German *zwar-aber*.
- iii) Other (non-connective) lexical marking of „importance“, such as *I emphasize that...*
- iv) One element turns out to be more supportive of the text's overall purpose.
- v) Syntactic structure: Main clause has more weight than subordinate clause.
- vi) Information structure: One clause contains only Given information, the other offers New.
- vii) The relation is stated by RST to regularly assign nuclearity, as with *Purpose* or *Condition*.

When investigating these cases, it seems that annotators decide on nuclearity by weighing a number of different, possibly conflicting factors, in some unpredictable way. I conclude that RST's nuclearity is not very helpful for explaining coherence. As an alternative, I sketch a framework of *multi-level representation* that treats coherence not as a correlate of a single tree structure but as resulting from the combination of layers such as thematic development, intentional structure, and text-type specific distributions of „Haupt-/Nebenstruktur“. In such a framework, *nuclearity* would not vanish from the scene but be restricted to the layer of intentional structure, with one consequence being that relations do not „automatically“ assign nuclearity, and thus annotators are not forced to choose a nucleus in every situation.

---

Joyce E. Stavick  
North Georgia Coll.; State Univ., Dahlonega, GA, USA  
**Subordinating Narrative Design for Syntactic Simplicity: An  
Examination of Punctuation, Subordination, and Coordination in Two  
Editions of Graham Greene's *The Power and the Glory***  
Do 9.00-9.30

---

This paper will examine the linguistics of point of view in Graham Greene's *The Power and the Glory*. by analyzing editorial revision of the colon in the 1940 edition to simpler punctuation in the 1971 edition. The comparison shows that punctuation of internal thought presentation plays a significant role in development of point of view, particularly when such revision alters coordinate and subordinate structures.

Although several scholars have studied thought presentation in third-person point-of-view, notably G. Leech and M. Short (1981) and M. Toolan (1988), few have examined the cloudier distinctions in narratives in which thought presentation is developed through the interplay of punctuation with subordinate and coordinate structures. The revision of Greene's novel, however, provides an opportunity for comparative study of this nature.

Using several passages from both editions of the novel, the paper will analyze thought presentation, suggesting that the colon helps to establish control of coordinating and subordinating structures in narrative point of view, but when it is revised to another mark, the subsequent changes in coordination and subordination affect the narrative process. For instance, the 1940 Viking edition provides this passage: "How often the priest had heard the same confession – Man was so limited; he hadn't even the ingenuity to invent a new vice: the animals knew as much" (131). The 1971 Penguin edition deletes the first colon to transform the second clause from coordinate to subordinate: "How often the priest had heard the same confession – Man was so limited he hadn't even the ingenuity to invent a new vice: the animals knew as much" (97). In the original edition, in which the priest laments man's squalid existence while he listens to Confessions, the explanatory and compensatory colons serve to show the priest's thoughts in an improvised state, while the deletion of the explanatory colon in the revised edition gives the sense of a studied, predetermined thought by subordinating "he hadn't even the ingenuity . . ." to the preceding clause, although the thought appears to be too spontaneous for such a structured passage of thought presentation.

This very brief example and others suggest that the colon works with subordinate and coordinate structures to influence thought presentation in narrative.

---

Michael Franke  
Univ. Amsterdam

**Pseudo-Imperatives: A Case-Study in the Ascription of Discourse  
Relations**

Do 9.30-10.00

---

Pseudo-imperatives (PIs) are compound sentences where an imperative clause is conjoined or disjoined with a declarative clause (comp. Clark 1993). Depending on the desirability of the variable *Y*, a PI with conjunction as in (1) can be interpreted to ask for performance or omission of *X*.

(1) Do *X*, and *Y* will happen/be the case/be done.

PIs with disjunction (2) require an instantiation of *Y* which is undesirable for the addressee and cannot be used to request omission of action *X*.

(2) Do *X*, or *Y* will happen/be the case/be done.

In contrast to English, the Japanese adverbial conjunction *soushite* behaves like English disjunction in that (3) is never understood as a request for omission of *X*.

(3) *X* shiro. *soushite Y* da/ga okoru/wo okonawareru.

The question to be asked is: Why can English (1) mean "Don't do *X*!" in the light of the fact that Japanese (3) and English (2) cannot?

Recently, Lascarides & Asher (2004) have proposed an analysis for PIs with conjunction (1) in terms of discourse relations. They argue that the non-veridical coordinating discourse relation *Def-Consequence* holds between the conjuncts in (1) on the informational level, while the veridical subordinating meta-talk relation *Explanation\** holds on the intentional level just in case the conjoined declarative sentence states a desirable consequence for the addressee. Against Lascarides & Asher this paper holds that recourse to meta-talk relations is not explanatory, but also not necessary. It is argued that on the basis of the correct informational discourse relations associated with English *and* in (1) and *or* in (2) appropriate meta-talk relations are to be derived, thereby (i) consolidating the coordination/subordination divergence and (ii) contrasting *and* and *soushite*. In particular, English conjunction *and* is said to realize an informationally coordinating, speech-act subordinating discourse relation *Circumstance*, while Japanese conjunction *soushite* is described as a speech-act coordinator. English *or* is treated as an informational coordinator which presents epistemic alternatives.

Clark, Billy (1993) 'Relevance and "Pseudo-Imperatives"', In: *Linguistics and Philosophy*, 16, 79-121.

Lascarides, A. & N. Asher (2004) 'Imperatives in Dialogue', To appear in: P. Kuehnlein, H. Rieser & H. Zeevat (eds) *The Semantics and Pragmatics of Dialogue for the New Millennium*, Benjamins.

---

Ingo Reich  
Univ. Tübingen  
**From Discourse to »Odd Coordinations« —**  
**On Asymmetric Coordination and Subject Gaps in German**  
Do 10.00-10.30

---

Asymmetric coordinations (ACs) like »*Wenn du nach Hause kommst und der Gerichtsvollzieher steht vor der Tür, dann ...*« and subject gap coordinations (SGCs) like »*Hoffentlich sieht er uns und nimmt uns mit*« are known (since Höhle 1983) to share two crucial properties, neither of which has been satisfactorily accounted for so far: (i) the finite verb in non-first conjuncts always needs to be fronted (verb-first or verb-second), (ii) ACs and SGCs are always [!] interpreted in some »unifying« way (in a sense to be specified). In this talk, I investigate the hypothesis that these two properties are in fact interrelated: fronting the finite verb in German, I argue, is to be taken as a syntactic reflex of semantically and pragmatically linking a »non-dependent« utterance to a previous sentence. (This is spelled out in terms of a functional projection RhetP sitting on top of TP in case of non-initial conjuncts.) It is shown that this assumption straightforwardly accounts for the above mentioned behavior of ACs and SGCs.

Building on this proposal, I then present an approach to SGCs that combines this analysis with insights from Höhle (1990) (asymmetric syntax) and Kratzer (1996) (severing the external argument from its verb), the crucial additional assumption being that the presence of syntactic subjects in sentence grammar essentially depends on a semantic condition:

- (1) **Subject Condition:** Every linguistically relevant event requires a syntactic subject.

It is shown that (1) works well in standard cases, and that it furthermore predicts the possibility of SGCs in German: given a (syntactically asymmetric) coordination, where (i) the subject of the first conjunct c-commands the second conjunct, and (ii) the whole coordination gets a »one complex event«-interpretation, the second conjunct may remain subjectless.

Höhle, T. N.: 1983, *Subjektlücken in Koordinationen*. Ms., Universität Tübingen.

Höhle, T. N.: 1990, 'Assumptions about Asymmetric Coordination in German'. In: J. Mascaro and M. Nespors (eds.): *Grammar in Progress. Glow Essays for Henk van Riemsdijk*. Dordrecht: Foris, pp. 221-235.

Kratzer, A.: 1996, 'Severing the External Argument from its Verb'. In: J. Rooryck and L. Zaring (eds.): *Phrase Structure and the Lexicon*. Dordrecht: Kluwer, pp. 109-137.

Rosemarie Lühr  
 Universität Jena  
**Sätze zwischen Subordination und Koordination im Altindischen**  
 Do 10.30-11.00

Im Altindischen ist im Hauptsatz das Verb in der Regel unbetont; hat es ein Präfix, trägt dieses den Ton. Demgegenüber ist das Verb im konjunktorial oder mit Relativpronomen eingeleiteten Nebensatz betont und ein eventuelles Präfix unbetont.

(1) RV V,31,6

(a) **prá** te púrvāNi káraNāni **vocam**  
 PFX von dir-GEN.SG früh-ACC.PL.N Taten-ACC.PL.N ich will  
 verkünden-1.SG.INJ.AOR.ACT

prá nūtanā maghavan  
 PFX neuest-SUPERL.ACC.PL.N Gabenreicher-VOC.SG

(b) yá **cakārtha** /  
 welche-REL-ACC.PL.N du hast getan-2.SG.IND.PFV.ACT

(c) sáktīvo yád **vibhārā** ródasī ubhé  
 Mächtiger-VOC.SG dass-SUB du teiltest-2.SG.SBJV.PRES.ACT Welten-ACC.PL.N  
 beide-ACC.PL.N

„Deine früheren Taten will ich verkünden, deine neuesten, die du Gabenreicher getan hast, dass du Mächtiger beide Welthälften trenntest“

Nun gibt es im Altindischen auch Sätze, die wie der subordinierte Satz betont sind, aber keine Konjunktion oder Relativpronomen als Einleitewort haben:

(2) RV V,45,3

a) ví párvato **jīhita** sádhata  
 auseinander-Pfx Berg-ACC.PL.M er öffnet sich-3.SG.INJ.PRES.MED  
 er gelangt zum Ziel-

dyáur  
 3.SG.INJ.PRES.MED Himmel-ACC.SG.M

b) avivasanto dasayanta bhúma  
 zu gewinnen suchend-PRT.PRES.ACT sie zermürben-3.PL.INJ.PRES.MED  
 Erde-ACC.SG.M

„Der Berg tut sich auf, er gelangt zum Himmel. Um die Erde zu gewinnen, zermürben sie (die Aṣgiras) sich (mit Sprüchen).“

In der Indogermanistik werden Sätze wie (2)(a) mit einem finalen betonten Verb (**jīhita**) vor einem satzinitialen (regulär betonten) Hauptsatzverb (**sádhata**) kontrovers beurteilt: Teils betrachtet man den Akzent auf dem ersten Verb als Zeichen von Emphase, teils als Zeichen der Subordination. Beide Thesen sind jedoch für das vorgelegte Material nicht haltbar. So gibt es unter den altindischen wie ein Nebensatz akzentuierten Sätzen keine Sätze mit einem Kontrast-Fokus auf dem Verb. Kontrast ist aber auf jeden Fall ein bedingender Faktor für die Verbbetonung im ersten Satz. Dies machen bestimmte Strukturen deutlich, die aus einem vorangestellten Satz mit Nebensatzakzent auf dem Verb und einem Hauptsatz bestehen. So liegt regelmäßig

Verbbetonung im ersten Satz vor in Fällen mit *vā – vā* ‚entweder – oder‘, *ca – ca* ‚sowohl – als auch‘ oder mit *anyá- - anyá-* ‚der eine – der andere‘ (im selben Vers). Im Vortrag wird erstens gezeigt, dass solche Sätze Kontrast-Topiks enthalten, deren Referenten durch Kontext oder Weltwissen bekannt sind. Zweitens wird der Akzent „7“ auf dem konnektfinalen Verb des ersten Konnektivs als Tonerhöhung gedeutet, die die Erwartung eines zweiten Konnektivs signalisiert.

---

Maria Averintseva

HU Berlin, Univ. Potsdam

### To the right of the clause: right dislocation vs. afterthought

Do 11.30-12.00

---

Cases like (1) and (2) in German are traditionally analysed as right dislocation ('Rechtsversetzung', Altmann (1981)):

(1) (Anna und Petra haben sich gestritten.) Dann ist *sie<sub>i</sub>* weggelaufen, (ich meine *die Anna<sub>i</sub>*,  
(Anna and Petra quarrelled.) Then is *she<sub>i</sub>* ran-away, (I mean) the *Anna<sub>i</sub>*.)

(2) (Den Tag, den vergess' ich nicht,) *der<sub>i</sub>* war viel zu schön, *der Tag<sub>i</sub>*.  
(That day, I'll never forget it,) *it<sub>i</sub>* was much too wonderful, the *day<sub>i</sub>*.

[Altmann (1981:129)]

A thorough analysis reveals that (1) and (2) are two different constructions, *right dislocation proper* (2) and *afterthought* (1). These constructions are shown to differ in their discourse function:

- Right dislocation marks the referent of the right-peripheral phrase as being especially salient for the following discourse segment (i.e. its 'discourse topic'), and is thus a global discourse structuring strategy.
- Afterthought resolves an unclear pronominal reference in its host sentence, and is thus a local repairing strategy.

The functional distinction correlates with formal differences between right dislocation and afterthought. In short, these are the following: Formally seen, right dislocation proper is prosodically and syntactically real part of its host sentence. Afterthought, on the contrary, is an 'orphan' (cf. Haegeman (1991)), a prosodically and syntactically independent unit that gets integrated into its host sentence only at the level of the discourse via a discourse relation. The discourse relation *Afterthought* is formulated for this case, and its subordinating character is specified.

It is argued that the right edge of the sentence can cross-linguistically be used either for marking the 'discourse topic' for the following segment (global discourse structuring strategy) or for supplying additional information (local repair strategy). It is expected that languages differ with respect to their using both possibilities, some using both and some preferring only one of the options. This prediction is tested with the material from Russian and French.

Altmann, Hans (1981): *Formen der "Herausstellung" im Deutschen: Rechtsversetzung, Linksversetzung, Freies Thema und verwandte Konstruktionen*. Tübingen: Niemeyer (= Linguistische Arbeiten 106).

Haegeman, Liliane (1991): "Parenthetical adverbials: The radical orphan approach." In: Chiba, S. et al. (eds): *Aspects of modern English linguistics: Papers presented to Masatomo Ukaji on his 60th birthday*. Kaitakusha. 232–254.

---

Anke Holler  
Univ. Heidelberg  
**Detached in Syntax – Attached in Discourse**  
Do 12.00-12.30

---

In generative grammar, it is commonly assumed that clauses standing alone differ grammatically from ones that are dependent on a matrix clause, which forms the basis for the distinction between main and subordinate clauses. This difference is often expressed by a boolean feature called ROOT (or alike), and by analyzing + ROOT - clauses as syntactically highest clauses. The stipulation of a ROOT feature has been motivated by an observation going back to Emonds (1970) whereby clauses vary in admitting of so-called root phenomena. Whereas + ROOT clauses support these phenomena, - ROOT clauses disallow them.

The talk discusses certain non-canonical clause linkage phenomena that challenge any approach implementing a twofold differentiation between main and subordinate clause types. While considering German as main example, it argues that clauses can be distinguished by grammatical means in terms of the degree they are integrated into a potential host clause. What is further shown is that typical non-integrated clauses are detached in syntax, but provide the context for the interpretation of their host. Adopting an approach to peripheral adverbials by Haegeman (1991), non-integrated clauses are analyzed as ‘orphan’ constituents. They serve to form the discourse frame against which the proposition expressed in the host clause is evaluated. Hence, there is no modification relation established in syntax, but rather at the level of utterance interpretation, which means that orphaned clauses are attached to their host in discourse only.

Recent Head-driven Phrase Structure Grammar seems well equipped to handle these facts. Thus, a constraint-based analysis is sketched that makes use of the idea that feature structures describing clause types can be organized according to the way the respective clause is linked to its syntactic surrounding. In addition, the sign-based architecture of Head-driven Phrase Structure Grammar qualifies to implement discourse attachment without having to refer to any syntactic information: In order to combine a non-integrated clause with its host, the semantic information of an orphaned constituent is unified with the background or context information of its host.

Emonds, Joseph. 1970. “Root and Structure-Preserving Transformations.” Ph.D. thesis, MIT, Cambridge.

Haegeman, Liliane. 1991. “Parenthetical Adverbials: The Radical Orphanage Approach.” In Shuji Chiba, Akira Ogawa, Yasuaki Fuiwara, Norio Yamada, Osamu Koma, and Takao Yagi, eds., *Aspects of Modern English Linguistics*, Kaitakusha, Tokyo, 232–254.

---

Svetlana Petrova, Michael Solf  
HU Berlin  
**Rhetorical Relations and Verb Placement in the Early Germanic  
Languages: A Cross-Linguistic Study**  
Fr 10.30-11.00

---

In a previous investigation, Hinterhölzl et al. (to appear) pursue an approach to the syntax of Old High German that especially concentrates on the relationship between the informational status of discourse referents and the placement of the finite verb in the OHG Tatian (9<sup>th</sup> century). The investigation reveals two striking tendencies. V2 order is found in categorical sentences containing referents suitable for topics whereas V1 placement appears in typicalthetic or all-focus sentences with no explicit topic constituent or rather no topic-comment structure (cf. Drubig 1992, 146). Nevertheless, in both V1 and V2 structures the finite verb occurs at the beginning of the domain of new-information focus.

However, a problem arises with evidence of V1 sentences containing discourse-given referents suitable for topics. Such sentences occur regularly at the beginning of a new narrative sequence as well as in clauses with verbs of motion, verbs of cognitive or sensual perception as well as verbs of saying. Hinterhölzl & Petrova (2005) attempt at providing a unified account for these occurrences of V1 in sentences with both discourse-new and discourse-given referents from the perspective of text hierarchy and rhetorical relations. It is argued that V1 vs. V2 order correlates with types of discourse relations attributed to the broad distinction between coordination vs. subordination as proposed in Asher & Lascarides (2003). This in turn gives rise to the conclusion that the position of finite verb in this text functions as a device in marking the type of rhetorical relations between units in discourse.

The approach presented so far is a first attempt at describing word order variation in a dynamic model of discourse relations that has to be extended over the other early Germanic languages as well. The findings concerning the situation in OHG are now being compared with data from the Old English, Old Saxon and Old Icelandic period.

Asher, N. and Lascarides A. 2003. *Logics of Conversation*. Cambridge: University Press.

Drubig B. 1992. „Zur Frage der grammatischen Repräsentation thetischer und kategorischer Sätze.“ In: J. Jacobs (hg.): *Informationsstruktur und Grammatik*. Linguistische Berichte. Sonderheft 4, 142-195.

Hinterhölzl, R. and Petrova, S. 2005. “Rhetorical Relations and Verb Placement in Early Germanic Languages: Evidence from the Old High German Tatian Translation (9<sup>th</sup> century).” In: M. Stede et al. (hg.) *Salience in Discourse. Multidisciplinary Approaches to Discourse*. Münster: Nodus Publikationen, 71-79.

Hinterhölzl, R., Petrova, S. and Solf, M. To appear. “Diskurspragmatische Faktoren für Topikalität und Verbstellung in der ahd. Tatianübersetzung (9. Jh.)” In: S. Ishihara et al. (hg.): *Interdisciplinary studies on information structure*. Potsdam: Universitätsverlag.



---

Olav Hackstein  
Univ. Halle

**Another look at main-clause phenomena in German subordinate clauses**  
Fr 11.00-11.30

---

Ongoing language change in contemporary German has given rise to a second variety of German causal and concessive clauses in which the verb takes the second position, the so-called German *weil*-verb-second/*obwohl*-verb-second (*weil*-V-2, *obwohl*-V-2). This phenomenon has been attracting the attention of many Germanicists for at least the past two decades. Most articles devoted to the subject have concentrated on a pragmatic and functional description of the new constructions. Few have ventured a historical explanation.

Perhaps the most influential theory concerning *weil*-V-2 is the thesis that *weil*-V-2 is the result of a replacement process involving syntactic analogy: *weil* replaces *denn*-V-2, which is on the wane in oral language, while retaining the V-2-syntax of *denn*. This explanation, however, is far from assured. Most importantly, it has been shown that in the northern varieties of German, the loss of paratactic *denn* is not the conditioning factor for the spread of *weil*-V-2. The indications are that, conversely, the spread of *weil*-V-2 causes the paratactic *denn*-construction to disappear, cf. Pasch, *Deutsche Sprache* 25,3 (1997), 256f.

My presentation will seek a comparative approach to main-clause phenomena in German. By examining parallel cases of the grammaticalisation of paratactic causal clauses in both ancient and modern IE languages, I will point out historical mechanisms that favor the spread of main clause phenomena in subordinate clauses. A demonstration that such mechanisms may also play a role in the emergence of the German *weil*-V-2 clauses will end my presentation.

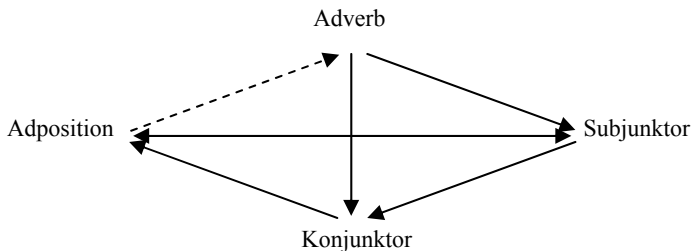
---

Hardarik Blühdorn  
IDS Mannheim

**Adpositionen, Adverbien, Subjunktionen und Konjunktionen.  
Die Interaktion syntaktischer, semantischer und textueller  
Verknüpfungseigenschaften in vier unterschiedlich grammatikalisierten  
Konnektorklassen**  
Fr 11.30-12.00

---

Der Vortrag gibt einen Überblick über die Verknüpfungseigenschaften der vier übereinzelsprachlich wichtigsten Konnektorklassen:



Konnektoren aller vier Klassen stellen semantische Relationen zwischen einem zu relationierenden Objekt E und einem Bezugsobjekt R her. Ihre semantischen Relate können Raumobjekte, Zeitobjekte (Sachverhalte), epistemische Objekte (Propositionen) oder deontische Objekte (Akte) sein. Adverbien und Konjunktionen können Relate aller vier Kategorien verknüpfen. Subjunktionen können keine Raumobjekte verknüpfen. Adpositionen eignen sich nur bedingt für die Verknüpfung von Propositionen und Akten.

Adverbien verknüpfen auf Textebene. Sie genießen relative syntaktische Stellungsfreiheit. Adpositionen und Subjunktionen verknüpfen auf Satzebene. Sie sind in ihrer syntaktischen Stellung beschränkt. Konjunktionen können von der Wortgruppe an aufwärts auf allen Strukturebenen verknüpfen. Ihre syntaktische Stellung ist fest.

Adpositionen verlangen Nominalkomplemente, die sie kasusregieren. Adverbien und Konjunktionen haben keine Rektion. Subjunktionen verlangen Satzkomplemente, die sie modusregieren. Adpositionen und Subjunktionen verknüpfen hierarchisch, Adverbien und Konjunktionen verknüpfen nicht-hierarchisch.

Zwischen den Klassen sind Übertritte durch Konversion in bestimmten Richtungen häufig. Diese gehen mit spezifischen syntaktisch-semantischen Prozessen einher: Intransitivierung, Retransitivierung, Rektionswechsel, Rektionsverlust, Stellungsverfestigung. In den Übergangsbereichen zum Konjunktor gibt es Mischformen: stellungsbeschränkte Adverbien und stellungsfeste Subjunktionen (Postponierer).

---

Sara Wischer  
Univ. Paderborn  
**Die Konzession als Diskurs-Relation**  
Fr 12.00-12.30

---

Konzessionen in argumentativen Zusammenhängen werden als sprachliche Strategien mit unterschiedlichen Intentionen eingesetzt:

- Überzeugen des Opponenten durch Vorwegnahme eines Gegenarguments
- moderates Widersprechen durch Betonung der Übereinstimmung mit der Meinung des Opponenten
- Bewahrung von Handlungsspielraum durch das Abwägen unterschiedlicher Standpunkte

Insbesondere bezogen auf die ersten beiden Intentionen würde eine sprachliche Realisierung der Konzessionen durch Subordination der Auffassung der RST von Mann/Thompson (1988) entsprechen, die die *concession* als eine *nucleus-satellite relation* definieren, bei der dem *nucleus* stärkeres argumentatives Gewicht zukommt.<sup>1</sup> Eine Untersuchung eines Textkorpus von über 250 politischen Reden im Rahmen meiner Dissertation hat aber gezeigt, dass Konzessionen hauptsächlich durch koordinierende Strukturen realisiert werden. Diese syntaktische Realisierung ist insbesondere auf die Komplexität der auszudrückenden Sachverhalte zurückzuführen.<sup>2</sup> Trotz des Vorliegens parataktischer Konstruktionen bei Konzessionen kann dem *satellite* einer Konzession jedoch diskursbezogen ein geringeres argumentatives Gewicht gegeben werden. D. h. zwischen der Informationspräsentation auf Satzebene und Textebene herrscht keine Entsprechung.

Welche sprachlichen Strategien für eine solche Hierarchisierung der Informationsstruktur auf Textebene – trotz parataktischer Realisierung – verwendet werden, soll im Zentrum des Vortrages stehen. Dazu gehören u. a. die Kombination typischer konzessiver Konnektive mit bestimmten Adverbien (z.B. *natürlich*, *gewiss*) oder Partikeln sowie sprachliche Konstruktionen, die eine Distanzierung des Sprechers zu dem konzedierten Sachverhalt signalisieren.

Mann/Thompson (1988): Rhetorical Structure Theory: Toward a functional theory of text organization. In: *Text* 8, S. 243-281.

---

<sup>1</sup> Mann/Thompson 1988, S. 270: „Nuclearity [...] has its obscure and borderline cases, but grammaticization of nuclearity in hypotaxis confirms a strong pattern.”

<sup>2</sup> Im Falle des Abwägens unterschiedlicher gleichwertiger Standpunkte entspricht diese syntaktische Realisierung ja auch der damit ausgedrückten Funktion.

---

Angelika Wöllstein  
Univ. Köln

**Kontrafaktizität als semantisch-konzeptuelle Basis der Satzkonnexion**  
Fr 12.30-13.00

---

Ausgehend von den Verbstellungstypen und Annahmen über ihnen spezifisch zugeordnete phrasenstrukturelle Abbildungen gelten insbesondere Verb-End-Konnekte generell als syntaktisch subordiniert. Damit kommt es zu einer vermeintlich grundlegenden Unterscheidung von ‚Matrix-+Komplementsatz‘- und ‚Matrix-+Adverbialsatz‘-Konstruktion-en einerseits und der Satzkoordination andererseits. Nur syntaktisch independente, nicht auf Subkategorisierung aufbauende Konnexionen unterschiedlichen Typs – konnektorlose ebenso wie solche, die von s.g. koordinierenden oder s.g. subordinierenden Konnektoren eingeleitet werden – weisen die gleiche kausale bzw. konditional Interpretation auf:

- (1) *Singt Hans, lacht Maria.* ‚WENN Hans singt, DANN lacht Maria.‘
- (2) *Hol den Arzt, oder der Hase stirbt.* ‚WENN du den Arzt nicht holst, DANN stirbt der Hase.‘
- (3) *Wir müssen uns beeilen, dass wir den Zug noch kriegen.* ‚WENN wir uns beeilen, DANN kriegen wir den Zug noch.‘
- (4) *Der Schuh ist zu klein, um zu passen.* ‚WENN der Schuh zu klein ist, DANN passt er nicht.‘

Wie kommt es bei so unterschiedlichen Konnexionstypen und -verfahren wie in (1)-(4) zu einer so weitgehenden Übereinstimmung ihrer Bedeutung, und welche Konzepte stehen für die Interpretation einer Satzkonnexion ohne overt Konnektor zur Verfügung? Die traditionell angenommenen Konnexionsverfahren stellen keine Interpretationsverfahren zur Verfügung, um die Bedeutung einer Satzkonnexion zu erschließen. Demgegenüber werde ich vorschlagen, dass alle syntaktisch independenten Konnexionen semantisch mittels eines Default-Interpretationsverfahrens verknüpft werden, welches eine semantische Dependenzrelation zwischen den beiden Konnekten etabliert. Diese Dependenzrelation basiert auf einem nicht-sprachlichen Konzept von Kausalität genauer auf einem intrinsisch asymmetrischen, kontrafaktischen kausalen Schlussverfahren, das auf Hume zurückgeht und u.a. von Lewis aufgegriffen wurde. Diese semantisch-konzeptuelle Basis motiviert und lizenziert die weitere Annahme, dass syntaktisch independente Konnexionen in der Syntax auf eine asymmetrische, binär verzweigende und von einem funktionalen Kopf projizierte Konnexionsphrase abgebildet werden. Eine weitere Implikation der semantischen Dependenzthese betrifft die Funktion der Konnektoren. Sie besteht darin, die Default-Interpretation entweder zu modifizieren, zu spezifizieren oder zu überschreiben.

## AG 9 Sprache und Recht

Koordination: Günther Grewendorf, Monika Rathert

### Programm

Mittwoch 22.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
13.45-14.45	Rainer Dietrich, Julia Hillebrand, Yvonne Stakič	<i>Vom allmählichen Verstehen juristischer Texte beim Lesen</i>
14.45-15.15	Sandra Hansen, Stella Neumann	<i>Syntaktische Komplexität in Rechtstexten. Tragen Reformulierungen zur Verständlichkeit von Gerichtsentscheidungen bei?</i>
15.15-15.45	Kent Lerch	<i>Weltverbesserungsmaßnahmen. Von der Verständlichkeit des Rechts und den Grenzen der Rechtslinguistik</i>
15.45-16.15	Kaffeepause	
16.15-16.45	Ralph Christensen	<i>Holistische Argumentationsstrukturen in der Grundrechtstheorie</i>
16.45-17.15	Ludger Hoffmann	<i>Diskursive Verständlichkeit</i>
17.15-17.45	Thomas M. Seibert	<i>Der Ertrag aus Mikroanalysen zur Gerichtsverhandlung</i>
17.45-18.15	Ulrike Schultz	<i>Zur Sprache des Geschlechts im Recht</i>

Donnerstag 23.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
09.00-10.00	Angelika Braun	<i>Forensische Phonetik und Linguistik</i>
10.00-11.00	Sabine Ehrhardt, Sabine Schall	<i>Das Kriminaltechnische Informationssystem Texte (KISTE)</i>
11.00-11.30	Kaffeepause	
11.30-12.00	Michael Jessen	<i>Forensische Phonetik und die Auswirkungen sprechstilistischer Unterschiede auf Globalmaße der Grundfrequenz</i>
12.00-12.30	Stephan Walter	<i>Computerlinguistische Methoden für die Rechtsterminologie</i>

Freitag 24.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
10.30-11.00	Karin Luttermann	<i>Mehrsprachigkeit in der Europäischen Union</i>
11.00-11.30	Isolde Burr	<i>EU-Rechtssprache(n): Risiko und Chance</i>
11.30-12.00	Thomas Gergen	<i>Sprachliche vs. juristische Übersetzungsprobleme im französisch-deutschen Rechts- und Sprachvergleich</i>
12.00-12.30	Tunde Opeibi	<i>Aspects of language use in a selected courtroom discourse in Lagos, Nigeria</i>
12.30-13.00	Andreas Lötscher	<i>Gesetzesredaktion innerhalb der Schweizer Mehrsprachigkeit</i>
13.00-13.30	Silke von Lewinski	<i>Mehrsprachigkeitsprobleme auf europäischer und internationaler Ebene beim Urheberrecht</i>

---

Rainer Dietrich, Julia Hillebrand, Yvonne Stakic  
Humboldt-Universität Berlin  
**Vom allmählichen Verstehen juristischer Texte beim Lesen**  
Mi 13.45-14.45

---

Der Aufbau von Wissen beim Lesen geschieht kontinuierlich. Eine Sequenz von im Text ausgedrückten Propositionen wird im Arbeitsspeicher zu einer größeren Einheit zusammengefügt und für eine gewisse Zeit gespeichert. In einem weiteren Schritt wird sie mit einer neu hinzukommenden Sequenz verbunden und, bei Erreichen der Kapazität des Kurzzeitspeichers in den Langzeitspeicher überführt. Im günstigen Fall ist die überführte Struktur eine kohärente und in sich widerspruchsfreie hierarchische komplexe Proposition. Treten bei diesem Prozess Lücken oder Unvereinbarkeiten auf, wird versucht, sie durch Regress auf zuvor verarbeitete Propositionen oder schließlich unter Einbezug von älterem Wissen zu überbrücken oder zu beseitigen; Ballstaedt u.a. (1981: 41 ff).

Becker u.a.(2004) haben nun eine auffällige Resistenz gegen Revisionen beim zyklischen Leseverstehen juristischer Texte (Allgemeine Versicherungsbedingungen) beobachtet. Ein einmal installiertes Verständnis von Begriffen wird angesichts später im Text auftretender unvereinbarer Informationen durchweg **nicht** revidiert und aktualisiert.

Wir haben versucht, diesen Befund in einem Lesexperiment zu replizieren. Versuchspersonen haben Belegtexte aus dem Korpus von Becker et al. und Kontrolltexte gelesen. Es wurden Augenbewegungs- und Fragebeantwortungsdaten erhoben und verglichen.

Ballstaedt, S.-P., Mandl, H., Schnotz, W. & S.O. Tergan. 1981. *Texte verstehen, Texte gestalten*. München: Urban & Schwarzenberg.

Becker, A., Paradi, Z. & F. Witzmann. 2004. *Wie ist das nun wieder gemeint? Empirische Untersuchung zum Verstehen eines Verbrauchervertrages durch Laien und Experten*. Forschungsbericht der Interdisziplinären Arbeitsgruppe „Sprache des Rechts“. Ms.. Berlin: Berlin-Brandenburgische Akademie der Wissenschaften.

---

Stella Neumann, Sandra Hansen  
Berlin, Universität des Saarlands  
**Syntaktische Komplexität in Rechtstexten**  
**Tragen Reformulierungen zur Verständlichkeit von**  
**Gerichtsentscheidungen bei?**  
Mi 14.45-15.15

---

In diesem Beitrag berichten wir über Modifikationen der syntaktischen Komplexität in Gerichtsentscheidungen. Seit längerem liegen Arbeiten zu syntaktischen Merkmalen von Rechtstexten vor. Genannt werden u.a. der Nominalstil und überlange, komplexe Sätze. In Arbeiten zur Verständlichkeit für Laien kommen neben einfachen Lesbarkeitsformeln auch aufwändigere psycholinguistische Methoden zum Einsatz. Schlussfolgerungen für eine verbesserte Verständlichkeit lassen sich jedoch nicht unbedingt aus Originaltexten ziehen.

Der vorliegende Beitrag stellt eine Vorgehensweise vor, die über die Untersuchung von Originaltexten hinausgeht. Die in einer korpuslinguistischen Analyse als Verständlichkeitshindernisse identifizierten Strukturen werden variiert, um dann die Auswirkung auf die Verständlichkeit mit psycholinguistischen Methoden zu überprüfen. Die Arbeit erfolgt in drei methodischen Schritten. Im ersten Schritt werden Gerichtsentscheidungen, Pressemitteilungen und Zeitungsberichte zu den Entscheidungen mit einem topologischen Parser, mit einer manuellen XML-Annotation und mit der Abfrage von Konkordanzen korpuslinguistisch ausgewertet. Es ist davon auszugehen, dass die Realisierung syntaktischer Strukturen in den Zeitungsberichten stark am Verstehen des Laien orientiert ist. Daher dienen die Charakteristika der Zeitungsberichte als Richtwert für Reformulierungen der Rechtstexte. Im zweiten Schritt werden ausgehend von Originalsätzen aus den Urteilen bei gleich bleibendem juristischem Gehalt zwei reformulierte Versionen gebildet. Die erste Reformulierung orientiert sich an den Realisierungen der Zeitungsberichte, während die zweite Reformulierung auf eine maximale Vereinfachung der syntaktischen Struktur abzielt. Die Versionen werden dann im dritten Schritt einem *self paced reading*-Experiment unterzogen. In die Auswertung gehen nicht nur die Lesezeiten ein, sondern auch die korrekte Beantwortung von Verständnisfragen und die Dauer bis zur Beantwortung der Fragen.

Die zuvor nur exemplarisch beschriebenen Besonderheiten von Rechtstexten können am Beispiel von Urteilen auch empirisch bestätigt werden. Die Analyse zeigt, worin sich die Komplexität der Strukturen genau manifestiert. Im Experiment schneidet die erste Reformulierung am besten ab. Daraus können Schlüsse für eine systematische Veränderung von Gerichtsentscheidungen mit dem Ziel einer verbesserten Verständlichkeit für den Laien gezogen werden.



---

Kent D. Lerch  
Mannheim  
**Weltverbesserungsmaßnahmen**  
**Von der Verständlichkeit des Rechts und den Grenzen der**  
**Rechtslinguistik**  
Mi 15.15-15.45

---

Das Recht ist nicht zu verstehen – darüber herrscht Konsens. Nicht nur die Bürger, auch die Juristen beklagen allenthalben die mangelnde Verständlichkeit des Rechts. Kritik wird an der Sprache des Rechts geübt, seit die Aufklärung die Verständlichkeit der Gesetze zu ihrem Anliegen gemacht hat. Mit den großen Kodifikationen des Rechts im ausgehenden 19. Jahrhundert hat die Kritik am angeblich schlechten, unverständlichen Juristendeutsch allerdings eine besondere demokratietheoretische Legitimation bekommen: es wird argumentiert, dass das Recht, da ihm alle unterworfen seien, auch für alle verständlich sein müsse. Diese Sprachkritik sucht seit den siebziger Jahren vermehrt bei der Linguistik um Rat, wie denn eine bessere Allgemeinverständlichkeit des Rechts verwirklicht werden könne. Getragen von dem Glauben an die Machbarkeit einer „sprachlichen Optimierung“ der Gesetzestexte kam es seither zu einer Reihe weithin beachteter Projekte für eine „bürgerfreundlichere“ neue Gesetzessprache. An verschiedenen staatlichen Stellen der Gesetzesproduktion wurden Sprachdienste geschaffen, in denen Linguisten und Juristen an der Redaktion von Gesetzen mitarbeiten. Vorsichtige Stimmen sowohl aus der Rechts- wie aus der Sprachwissenschaft verweisen allerdings darauf, dass sich die Sprachwirklichkeit der Rechtswelt solchen Bemühungen gegenüber als recht resistent erwiesen hat. Die Grenzen der Möglichkeit, Recht zu verstehen, werden daher neu ausgelotet.

- Busse, D. 1992, *Recht als Text. Linguistische Untersuchungen zur Arbeit mit Sprache in einer gesellschaftlichen Institution*. Tübingen: Niemeyer.
- Hoffmann, L. 1992, Wie verständlich können Gesetze sein, in: G. Grewendorf (Hg.), *Rechtskultur als Sprachkultur. Zur forensischen Funktion der Sprachanalyse*. Frankfurt am Main: Suhrkamp, 122-154.
- Lerch, K. D. 2004, Gesetze als Gemeingut aller. Der Traum vom verständlichen Gesetz, in: K. D. Lerch (Hg.), *Recht verstehen. Verständlichkeit, Missverständlichkeit und Unverständlichkeit von Recht*. Berlin: de Gruyter, 225-237.
- Lerch, K. D. 2005, Justitia im Bett des Prokrustes. Sinn und Unsinn der linguistischen Analyse von Rechtstexten, in: K. D. Lerch (Hrsg.), *Recht vermitteln. Strukturen, Formen und Medien der Kommunikation im Recht*. Berlin: de Gruyter, 169-182.

---

Ralph Christensen  
Mannheim

**Holistische Argumentationsstrukturen in der Grundrechtstheorie**  
Mi 16.15-16.45

---

Drittwirkung oder horizontale Wirkung nennt man die Geltung der Grundrechte im Konflikt zwischen Privatrechtssubjekten. Die Begründung der Drittwirkung erfolgt bei allen an der Diskussion Beteiligten dadurch, dass auf das Ganze der Rechtsordnung verwiesen wird. Wenn Ansprüche miteinander kollidieren, kann man ihre Reichweite innerhalb der Gesamtrechtsordnung bestimmen.. Die Rechtsordnung wurde bei der Begründung der Drittwirkungslehre als Wertpyramide oder Wertrangordnung angesehen, an deren Spitze die Grundrechte und Art. 1 GG stehen. Von dieser Spitze her lassen sich dann Konflikte zwischen einzelnen Elementen entscheiden. Das ist eine spezifische Form des holistischen Arguments. Man könnte auch vom Einzelnen her auf das Ganze hin argumentieren. Stattdessen wird hier vom Ganzen her argumentiert. Diese gegensätzlichen Strategien für holistisches Argumentieren im Recht sollen hier analysiert werden. Idealtypisch sind zwei Konzepte zu unterscheiden: Das divisionale Konzept geht davon aus, dass man das einzelne Recht vertikal ins Ganze der Rechtsordnung einfügt. Das Ganze ist dabei bestimmbar und beherrschbar und man kann von der Sinnmitte einer geordneten Totalität her dem einzelnen Anspruch seinen Platz zuteilen. Das ökologische Konzept begreift das Ganze der Rechtsordnung nicht als geordnete Totalität, sondern als offenen Horizont. Die Lösung des Konflikts kann nicht vom Ganzen her zugeteilt werden. Aber im Verfahren werden die Ansprüche der Parteien auf den offenen Horizont der Gesamtrechtsordnung bezogen und damit durch wechselseitige Einwände neu kontextualisiert. Dies zwingt zu einer neuen Rahmung und damit zur selbstreflexiven Praxis. Während das erste Modell auf das Gesetz und die Rechtserkenntnis setzt, muss sich das zweite Modell auf die Leistung des Verfahrens verlassen und ist damit stärker rechtsprechungsbezogen.

---

Ludger Hoffmann  
Dortmund  
**Diskursive Verständlichkeit**  
Mi 16.45-17.15

---

Ein Text gilt als verständlich, wenn das, was der Autor meint, so formuliert ist, dass es in seiner wörtlichen Bedeutung und mit relevanten Implikationen von den Adressaten verstanden werden kann. Problematisch ist der Zugang zum Verstehen größerer, inhomogener Gruppen von Adressaten. Im Diskurs hingegen wird ein bestimmtes Verständnis häufig explizit gemacht, wir finden Prozesse des Problematisierens von Verstehen, des Verständlich-Machens und des Streits um ein angemessenes Verständnis, daneben aber auch Indizien für tiefgreifende Missverständnisse und Verstehenskonflikte.

In der Gerichtsverhandlung gilt das „Mündlichkeitsprinzip“. Was dem Urteil zugrunde gelegt werden soll, muss in einem bestimmten Handlungsraum unter Präsenz der Entscheider und Betroffenen in deutscher Sprache (evtl. übersetzt) gesagt oder verlesen worden sein. Können sich die Akteure verständlich machen, so dass ihre Beiträge entscheidungsrelevant werden? Welche Verstehensprozesse und –konflikte zeigen sich im Prozess? Werden Verstehensschwierigkeiten bearbeitet? Wie kommen divergierendes Wissens- und Kulturhintergründe ins Spiel? Ist das Mündlichkeitsprinzip bloße Fiktion? Fragen dieser Art wird im Blick auf empirische Beispiele nachgegangen.

---

Thomas-Michael Seibert  
Landgericht Frankfurt  
**Der Ertrag aus Mikroanalysen zur Gerichtsverhandlung**  
Mi 17.15-17.45

---

Mikroanalysen befassen sich mit kleinen, vorzugsweise mikroskopisch kleinen Teilen der Gerichtsverhandlung, die wegen ihrer Detailbegrenzung der juristischen Aufmerksamkeit meist entgehen oder als juristische Ereignisse gar nicht thematisiert werden. Das Material dafür bilden Transkriptionen empirischer Abläufe, die seit den 60er Jahren des 20. Jahrhunderts im Gerichtssaal mit Hilfe von Tonaufzeichnungen gewonnen werden. Untersuchungsmethoden dafür sind in der linguistischen Pragmatik, in der Konversationsanalyse und in der soziologischen Handlungsforschung entwickelt und erprobt worden. - Erträge aus solchen Analysen gewinnt man nur, wenn der Untersucher wenigstens versucht, einen Mehrwert zu erzeugen. Das ist nicht selbstverständlich. Nicht selten scheint der Reiz für Analysen im Gerichtssaal schon in der Besonderheit des Ortes zu liegen. Mögliche Erträge sehe ich jenseits dieser nur thematischen Bindung in Einsichten über die Fremdheit und Gewalttätigkeit des Rechts ebenso wie in Veränderungsperspektiven für das Verhandeln des Rechts. Der eine Ertrag ist grundlagentheoretisch und rechtskritisch

(oder auch „sozialwissenschaftlich“), der andere wirkt in die Rechtspraxis hinein und gestaltet sie um. - Analysen über den Gewaltcharakter des Rechts liegen aus unterschiedlichen Verhandlungsfeldern vor. Es bot sich an, Vergewaltigungsprozesse als Modell für die Gewaltförmigkeit des Rechts zu untersuchen, und es erschien von Anfang an wahrscheinlich, dass männliches Dominanzgebaren in der Gesellschaft gerade in den Dominanzansprüchen der Rechtsverhandlung zum Ausdruck kommen würde. Wesentliche Erträge dazu bieten amerikanische Untersuchungen (Matoesian), deutsche haben empirischen Nachholbedarf.

Die Rechtspraxis hat sich im Laufe der seit ca. 1965 begonnenen Verhandlungsanalysen verändert, ohne dass man schlüssig behaupten könnte, sie habe dies gerade wegen der analytischen Ergebnisse getan. In dem Maße, in dem der sog. „ordentliche Rechtsstreit“ seine Entscheidungsfunktion nicht mehr erfüllen kann, gilt die Aufmerksamkeit hier der Vergleichspraxis und der Mediation, die neuerdings ein eigenes, institutionell ausdifferenziertes Verfahrensfeld bietet (Haft, v. Schlieffen).

Egli, U. 1996. *Vergleichsdruck im Zivilprozess*. Berlin: Duncker & Humblot.

Friedman, G.J. 1993. *A Guide to Divorce Mediation*. New York: Workman.

Haft, F., Gräfin v. Schlieffen, K. *Handbuch Mediation*, München: Beck

Matoesian, G.M. 1993. *Reproducing Rape: Domination through Talk in the Courtroom*. Chicago: UCP.

---

Ulrike Schultz  
FernUniversität Hagen  
**Zur Sprache des Geschlechts im Recht**  
Mi 17.45-18.15

---

Bestehen in der fachspezifischen Kommunikation Unterschiede zwischen Juristen und Juristinnen? Die empirischen Arbeiten der Linguistin Deborah Tannen zum Sprachstil von Männern und Frauen legen es nahe. Sie charakterisiert die männliche Sprache als Berichtssprache, die von Frauen als Beziehungssprache. Es gibt auch Feststellungen dazu, dass Frauen im Studium und im Beruf weniger zu Wort kommen, weniger Anerkennung für ihre Beiträge finden, häufiger unterbrochen werden. Weibliche Körpersprache unterscheidet sich in vielen Elementen von der männlichen. Wie wirken sich diese Unterschiede im juristischen Bereich aus? Haben Frauen, die bis vor 30 Jahren nur in geringem Umfang an juristischen Tätigkeiten beteiligt waren, inzwischen aber zu fast einem Drittel in den juristischen Berufen vertreten sind, die juristische Sprache, den juristischen Diskurs beeinflusst? Welche Auswirkungen haben dabei die Forderungen nach geschlechtsneutraler Sprache gehabt? Lange waren Frauen mit dem Argument, dass sie nicht zur Objektivität fähig seien, aus den juristischen Tätigkeiten herausgehalten worden. Hat also der gestiegene Frauenanteil Auswirkungen auf Rechtsprechung und –praxis gehabt? Verhandeln Frauen anders, sprechen sie anders Recht oder gar unrecht? Wie werden sie wahrgenommen, wie wird ihre Sprache, ihr Auftreten bewertet? Es soll eine erste vorsichtige Bestandsaufnahme und Annäherung an das Thema versucht werden. Dabei werden Erkenntnisse feministischer, linguistischer und kommunikationswissenschaftlicher

Arbeiten verwendet und die eigenen praktischen Erfahrungen als Kommunikationstrainerin für JuristInnen einbezogen.

- Braun, F. (2003). Rechtssprache – gerechte Sprache. In: Ministerium für Gesundheit, Soziales, Frauen und Familie NRW (Hg.) Frauen und Recht. Reader für die Aktionswochen der kommunalen Gleichstellungsbeauftragten. Düsseldorf 2003, S. 128-141
- Grabrucker, M. 1993. Vater Staat hat keine Muttersprache. Frankfurt: Fischer
- Schultz, U. 2002. Frauen im Recht. Aspekte der Kommunikation. In: Kruck, P., Rothe, C. & Schäfer, G. (Hg.) *Kommunikation im Gespräch. Festschrift für Franz R. Stuke zum 60. Geburtstag*. Münster: Daedalus, S. 137-143
- Schultz, U. 2004. Richten Richterinnen richtiger. In: Ministerium für Gesundheit, Soziales, Frauen und Familie NRW (Hg.) Frauenbilder. Reader für die Aktionswochen der kommunalen Gleichstellungsbeauftragten. Düsseldorf 2003, S. 117-126
- Tannen, D. 1991. *Du kannst mich einfach nicht verstehen. Warum Männer und Frauen aneinander vorbeireden*. Hamburg: Kabel

---

Angelika Braun  
Universität Marburg  
**Forensische Phonetik und Linguistik**  
Do 9.00-10.00

---

Dieser Beitrag eröffnet den Blick auf zwei hoch spezialisierte und gleichermaßen gesellschaftlich relevante Anwendungsgebiete der Phonetik und Linguistik. Dabei stehen Forensische Phonetik und Linguistik in einer symbiotischen Beziehung: Beide beinhalten die Anwendung phonetischen und linguistischen Wissens auf Analyse und Vergleich von Ton- und Schriftdokumenten, die für Ermittlungs- und Gerichtsverfahren relevant oder gar entscheidend sind.

Die Analyse von Täteranrufen schließt neben genuin phonetischen auch linguistische Befunde ein; die linguistische Analyse von Täterschreibern rekuriert im Wege der Fehleranalyse u.a. auf phonetisches Wissen. Wenn – wie häufig in Erpressungs- und Entführungsfällen – sowohl mündliche (Anrufe) als auch schriftliche (Briefe) Täteräußerungen vorliegen, gewinnt die Zusammenarbeit von Phonetikern und Linguisten an Bedeutung.

Der Beitrag versteht sich als eine Einführung in den Themenkreis zur Vorbereitung der spezialisierten Referate. Er umreißt zunächst das Spektrum der Aufgabenstellungen in der forensisch-phonetischen Sprechererkennung und beschreibt die zu ihrer Bewältigung angewandten Methoden. Dabei stehen Stimmenanalyse und Stimmenvergleich im Fokus; die eher "technischen" Bereiche *speech enhancement* und Tonträgerauthentisierung werden nur kurz gestreift.

In einem zweiten Schritt werden Grundlagen der forensisch-linguistischen Textanalyse erläutert. Hier geht es primär um das Spannungsfeld zwischen Idiolekt und Textsorte. Dies involviert im konkreten Fall eine – textsortenunabhängige – Fehleranalyse und eine – textsortenabhängige – Stilanalyse. Abschließend wird die Problematik der Ergebnisformulierung (Wahrscheinlichkeitsaussagen) in phonetischen und linguistischen Gutachten diskutiert.

- Braun, A. (1995). Procedures and Perspectives in Forensic Phonetics. In: *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Stockholm, vol. 3, 146-153.
- Champod, Chr. / Meuwly, D. (1998). The inference of identity in forensic speaker recognition. In: *Proceedings of RLA2C*. Avignon, 125-134.
- Rose, Ph. (2002). *Forensic Speaker Identification*. London: Taylor and Francis.
- 

Sabine Ehrhardt, Sabine Schall  
BKA Wiesbaden

## **Das Kriminaltechnische Informationssystem Texte (KISTE)**

Do 10.00-11.00

---

**Zur Forensischen Linguistik (FL) im Bundeskriminalamt:** Es gehört zu den Aufgaben der FL im BKA 1) durch die Erstellung von Gutachten der Polizei sowie der Staatsanwaltschaft und den Gerichten zuzuarbeiten, 2) die nationale Sammlung inkriminierter Schreiben zu führen, 3) Forschungsarbeit auf dem Gebiet der FL zu leisten. Bei allen drei Aufgabenstellungen werden die Linguisten des BKA von dem Computerprogramm KISTE durch dessen umfangreiche Sammlungs- und Recherchemöglichkeiten unterstützt.

**Zur Funktionsweise von KISTE:** KISTE ermöglicht die Speicherung eines Textes sowohl mit seinen administrativen Daten als auch mit sämtlichen linguistischen Informationen. Nach Eingabe der Verwaltungsdaten beginnt die eigentliche Texterfassung mit der manuelle Eingabe des Tatschreibens mit exakt allen seinen sprachlichen Eigenschaften. Anschließend erfolgen Korrektur, Deflexion, Wortartenbestimmung sowie grammatische und syntaktische Analyse. Den Abschluss der linguistischen Bearbeitung eines Textes bildet die Markierung aller Fehler und Auffälligkeiten sowie eine Interpretation dieser Befunde. Über die eingegebenen Informationen lassen sich Statistiken erstellen und sie können zu einem Befundprotokoll zusammengestellt werden.

Die sehr detaillierte Erfassung der inkriminierten Schreiben in KISTE ist die Voraussetzung für umfangreiche Suchroutinen. Die administrativen Daten können ebenso Gegenstand einer Recherche werden wie konkrete Formen des Wortbestandes, einzelne Fehler und Stilmerkmale. Zudem kann mithilfe von Verknüpfungsoperatoren und Abfragefunktionen im gesamten Korpus oder nur in einer ausgewählten Menge an Texten gesucht werden und mehrere Recherchen miteinander verknüpft werden.

**Ausblick: Das Forschungspotential von KISTE:** Insbesondere die umfangreichen Recherchemöglichkeiten innerhalb der Sammlung inkriminierter Schreiben, die KISTE bietet, besitzen ein hohes Forschungspotenzial für die FL im Besonderen und die Angewandte Linguistik im Allgemeinen. Für die Autorenerkennung spielen vor allem Hintergrundstatistiken zu Fehlervorkommen eine große Rolle, aber auch die Untersuchung von Merkmalen, die potenziell alters-, bildungs- oder regionsabhängig sind. Ebenfalls aussichtsreich ist die Forschung zu bisher kaum untersuchten Textsorten (z. B. Erpresserschreiben, Drohbrief, Beleidigungsschreiben).

---

Michael Jessen  
BKA Wiesbaden  
**Forensische Phonetik und die Auswirkungen sprechstilistischer  
Unterschiede auf Globalmaße der Grundfrequenz**  
Do 11.30-12.00

---

Bei der Erstellung von Stimmenvergleichsgutachten im Rahmen der forensischen Sprecher-Erkennung werden phonetische und linguistische Parameter untersucht, die eine hohe interindividuelle Variation aufweisen und deshalb für die probabilistische Individualisierung von Sprechern geeignet sind. Im Bereich der Grundfrequenzparameter haben sich zwei F0-Globalmaße als in dieser Hinsicht nützlich erwiesen: der F0-Mittelwert und die F0-Standardabweichung, wobei die F0-Standardabweichung als Variationskoeffizient ausgedrückt werden sollte (Künzel, 1987; Kraayeveld, 1997; Jessen et al., 2005). Erschwerend für die Stimmenvergleichsanalyse ist der Umstand, dass zahlreiche Faktoren auch die intraindividuelle Variation der Grundfrequenz erhöhen können, so u.a. die Sprechlautstärke, Emotionen, Stimmverstellungen oder der linguistische Kontext (Braun, 1995). Durch Wissen über die Art und das Ausmaß der Einflüsse solcher Faktoren auf die Grundfrequenz können solche Quellen intraindividuelle Variation gutachterlich gewürdigt werden. Im vorliegenden Vortrag geht es in erster Linie um den Einfluss, den der stilistische Unterschied zwischen gelesener und spontan gesprochener Sprache auf die genannten Globalmaße der Grundfrequenz hat. Dieser Einfluss ist forensisch u.a. dann relevant, wenn in der Tonaufzeichnung spontan gesprochen wurde, das Vergleichsmaterial aber auf gelesener Sprache beruht. Anhand eines im BKA erstellten Korpus von 100 männlichen erwachsenen Sprechern des Deutschen wurde der Unterschied zwischen Leseaussprache und Spontansprache untersucht. Das Ergebnis zeigt insgesamt geringe Unterschiede, aber einen statistisch nachweisbaren Trend zur Erhöhung des F0-Mittelwertes und des F0-Variationskoeffizienten in Spontansprache relativ zur Lesesprache.

- Braun, A. 1995. „Fundamental frequency – How speaker-specific is it?“ In: A. Braun & J.-P. Köster (hg.) *Studies in Forensic Phonetics*. Trier: Wissenschaftlicher Verlag, 9-23.
- Jessen, M., Köster, O. & Gfroerer, S. 2005. Influence of vocal effort on average and variability of fundamental frequency. Erscheint in: *The International Journal of Speech, Language and the Law*, Ausgabe 2, 2005.
- Kraayeveld, J. 1997. Idiosyncrasy in prosody: Speaker and speaker group identification in Dutch using melodic and temporal information. Dissertation Katholische Universität Nijmegen.
- Künzel, H.J. 1987. *Sprechererkennung: Grundzüge forensischer Sprachverarbeitung*. Heidelberg: Kriminalistik-Verlag.

---

Stephan Walter  
Universität des Saarlandes  
**Computerlinguistische Methoden für die Rechtsterminologie**  
Do 12.00-12.30

---

Gesetzgebung und Rechtsprechung finden in natürlicher Sprache statt. Die Anwendung von Rechtsvorschriften auf die Realität der Fälle hängt entscheidend von der Bedeutung der Rechtsbegriffe ab. Diese werden jedoch durch Definitionen im Gesetzestext meist nicht erschöpfend bestimmt, sondern nur partiell festgelegt. Über die Anwendbarkeit auf den konkreten Rechtsfall kann nur nach weiterer Präzisierung entschieden werden. Deshalb sind Definitionen auch in Gerichtsurteilen wesentliche Bestandteile. Obwohl formal nicht bindend, haben solche begrifflichen Festlegungen in Urteilsbegründungen eine hohe faktische Bindungswirkung für die anschließende Rechtssprechung.

Bei der Erstellung rechtsterminologischer Ressourcen stellen Urteilsbegründungen daher - neben Gesetzestexten - eine zweite wichtige Wissensquelle dar. Bei deren Nutzung ist die Unterstützung durch automatische Verfahren allein aufgrund der enormen Menge des zu berücksichtigenden Textmaterials kaum verzichtbar. Das in diesem Referat vorgestellte DFG-Projekt CORTE ("Computerlinguistische Methoden für die Rechtsterminologie", PI 154/10-1) beschäftigt sich mit der Verwendung computerlinguistischer Technologie zur automatischen Extraktion und Verarbeitung von Definitionen in Urteilstexten.

Die kaskadierte Kombination verschiedener flacher Verfahren erlaubt es, auch so komplexen Text wie Gerichtsentscheidungen robust verhältnismäßig tief zu analysieren. Dies trägt zunächst zur Verbesserung der Suche nach Definitionen bei. Da neben der Oberflächenform der Texte auch die linguistische Struktur (z.B. Phrasen, Modus und Tempus) berücksichtigt werden kann, lassen sich akkuratere Suchmuster formulieren, und es steht eine erweiterte Grundlage für die Klassifikation mittels automatischer Lerner zur Verfügung.

Außerdem ist durch die linguistische Analyse eine weitere Verarbeitung von Fundstellen möglich. So können Fundstellen z.B. in Definiens und Definiendum segmentiert oder auf explizite Geltungsbeschränkungen untersucht werden. In manchen Fällen können zwischen Fundstellen überdies Bedeutungsrelationen (wie Hyponymie und Antonymie) und rechtstheoretische systematische Zusammenhänge (Fallgruppen, widerstreitende Meinungen, Revision) ermittelt werden. Durch die Nutzung solcher Information können extrahierte Fundstellen strukturiert und gewichtet präsentiert werden, um beispielsweise die Erweiterung einer terminologischen Datenbank hinsichtlich aktueller Rechtsprechung zu unterstützen.



---

Karin Luttermann  
Ingolstadt  
**Mehrsprachigkeit in der Europäischen Union**  
Fr 10.30-11.00

---

„La parole a été donnée à l’homme pour déguiser sa pensée.“ Das sagte angeblich der Franzose *Charles de Talleyrand* zum Gesandten *Izquiero*, der ihn an Versprechen gegenüber *Karl IV.* von Spanien erinnerte. Für ein Europa im 21. Jahrhundert taugen diese Worte aber nicht mehr. Vielmehr braucht das Recht der Europäischen Union und ihrer fünfundzwanzig Mitgliedstaaten klare Aussagen. Doch in welcher Sprache? In welchen Sprachen? – Die Menschen in Europa sprechen viele Sprachen. Gegenwärtig sind 20 Amtssprachen gleichberechtigt: Dänisch, Deutsch, Englisch, Estnisch, Finnisch, Französisch, Griechisch, Italienisch, Lettisch, Litauisch, Maltesisch, Niederländisch, Polnisch, Portugiesisch, Schwedisch, Slowakisch, Slowenisch, Spanisch, Tschechisch und Ungarisch.

Daraus ergeben sich weit mehr als 300 Sprachkombinationen. Mit dem Beitritt von Bulgarien und Rumänien ab 2007 müßte in 22 Sprachen gearbeitet werden, 462 Sprachenpaare wären vorzusehen. Das erinnert an babylonische Sprachverwirrung und ist jedenfalls kaum noch beherrschbar. Praktisch ist auf die Grenzen der Übersetzungsleistung und deren Kosten zu blicken. Die Rechtspraxis der Europäischen Union begrenzt zuweilen schon den durchgängigen Gebrauch mitgliedstaatlicher Muttersprachen, um effizient handeln zu können. Not tut ein tragfähiges Sprachenrecht für das Gemeinschaftsrecht.

Kern des Vortrags ist darum ein neues Sprachenmodell, das auf den Grundlagen von Muttersprache und Amtssprachen aufbaut (sog. Referenzsprachenmodell). Mit dem Europarecht, besonders der rechtssprachenvergleichenden Praxis des Europäischen Gerichtshofes, wird es entwickelt und begründet. Das Referenzsprachenmodell wahrt die Vielsprachigkeit, mindert den Übersetzungsaufwand und ermöglicht den Unionsbürgern zugleich, ihre eigene Sprache zu gebrauchen. Wegweisend steht die Rechtsprechung des Europäischen Gerichtshofes, der die einheitliche Auslegung der Sprachfassungen hervorhebt und diesen Ansatz weiter differenziert. Das zeigt den Sprachenvergleich als die zentrale Aufgabe für Rechtsangleichung in den Mitgliedstaaten.

---

Isolde Burr  
Universität Köln  
**EU-Rechtssprache(n): Risiko und Chance**  
Fr 11.00-11.30

---

Die Problematik und Risikohaftigkeit der Sprachenvielfalt im EU-Recht wird nicht erst jetzt, seit es zwanzig als authentisch erklärte Sprachen gibt, immer wieder thematisiert. Die verschiedenen Rechts- und Sprachkulturen innerhalb der Europäischen Union erhöhen das Risiko sprachlicher Divergenzen um ein Vielfaches, die von linguistischer Seite aus oftmals präferierte „Urfassung“ eines EU-Rechtstextes und das damit verbundene Konzept der Übersetzung für alle anderen Sprachfassungen, ist nur scheinbar eine Lösung aus dem Dilemma. Dass auf nationaler Ebene EU-Rechtstexte meist nur unter Berücksichtigung der jeweiligen Amtssprachen umgesetzt werden, birgt insofern ein enormes Konfliktpotenzial, als die daraus resultierende Auslegung bisweilen von der mehrsprachigen Auslegung des EuGH abweicht. Je nach Auslegung wird beispielsweise das wirtschaftliche Handeln innerhalb der EU ganz unterschiedlich geprägt und stößt bei Divergenzen zwangsläufig auf weit reichende Konsequenzen. Die Gleichberechtigung der offiziellen EU-Rechtssprachen ist nicht nur ein Identitätsmerkmal der EU. Sie ist auch eine politische Vorgabe, die die Entwicklung des EU-Rechts überhaupt erst auf den Weg bringen kann. Dem steht auch die Quantität der zu berücksichtigenden Sprachen nicht im Wege. Dies lässt sich an den jüngsten Entscheidungen des EuGH belegen. Ein besonderes Augenmerk verdient die Frage, inwieweit juristische Sprachtheorien mit ihren daraus abgeleiteten Forderungen an den Rechtstext im Auslegungsprozess und als verständliches gesellschaftliches Verhaltenssteuerungsinstrument mit linguistischen Sprachtheorien übereinstimmt. Eine Genese von EU-Texten, die nicht auf einem Ausgangstext mit anschließender Übersetzung in alle anderen Sprachen basiert, zeigt, dass die große Zahl der Sprachfassungen nicht sosehr als Problem der Auslegung zu sehen ist, als vielmehr als Chance zum Einblick in den ursprünglichen Regelungswillen. Die Sprachenvielfalt unterstützt die Normierung der EU-Rechtssprache, sofern die neue Versprachlichung eines neuen Konzepts in einer engen Verknüpfung von Auslegung und Textgenese besteht.

- Braselmann, P. 1992. „Übernationales Recht und Mehrsprachigkeit. Linguistische Überlegungen zu Sprachproblemen in EuGH-Urteilen.“ In: EuR H.1 1992, 55ff.
- Müller, F. / Burr, I. Hrsg. 2004. *Rechtssprache Europas. Reflexion der Praxis von Sprache und Mehrsprachigkeit im supranationalen Recht.*. Berlin: Duncker & Humblot.
- Schübel-Pfister, I. 2004. *Sprache und Gemeinschaftsrecht. Die Auslegung der mehrsprachig verbindlichen Rechtstexte durch den Europäischen Gerichtshof.* Berlin: Duncker & Humblot.

---

Thomas Gergen  
Universität des Saarlandes  
**Sprachliche vs. juristische Übersetzungsprobleme im französisch-  
deutschen Rechts- und Sprachvergleich**  
Fr 11.30-12.00

---

Jeder Rechtsvergleich muss sich den Hürden des Sprachvergleichs stellen, und die Probleme, die die juristische Übersetzung mit sich bringt, mit linguistischer als auch juristischer Präzision angehen. Der einfache Zugriff auf Terminologie-Datenbanken ist natürlich eine große Hilfe. Doch geht es bei der Übersetzung juristischer Termini nicht ohne Vergleich auch der Rechtssysteme.

In unserem Beitrag sollen Schwierigkeiten und Lösungen aus eigenen französisch-deutschen Übersetzungserfahrungen dargelegt werden. So gehört zu einer ersten Erfahrung die Übersetzung zentraler, für den französischen Juristen einschlägiger Begriffe wie „Conseil d’Etat“ oder die „voie de fait“. Gewiss kann man „Conseil d’Etat“ nicht mit „Staatsrat“ übersetzen, war doch dies gerade ein politischer Begriff in der DDR. Schon bei „Conseil d’Etat“ und „voie de fait“ genügt eine einfache Übersetzung nicht, sodass der Übersetzer verpflichtet ist, eine adäquate Erklärung zu liefern, um den Rechtsbegriff zu paraphrasieren. Zudem können für unsere Untersuchung Beispiele aus der Übersetzung wichtiger Begriffe des deutschen BGB ins Französische geliefert werden. Hierbei werden Begriffe herangezogen, die sich um die notarielle Beurkundung (*authentification par-devant notaire*) und um die Insolvenzverwaltung (*faillite, syndics*) ranken.

Ein großes Problem trat schließlich noch auf bei der Übersetzung früherer und neuer Tatbestände des französischen Scheidungsrechts, die der deutsche Jurist bei der Anwendung französischen Rechts selbstverständlich beachten muss. Auch hier würde eine blinde Übernahme juristischer Termini ins Leere führen, sodass ein Vergleich der Rechtssysteme stets angebracht ist: „Scheidung auf Französisch“ ist eben doch nicht so einfach!

- Gergen, T. 2005. „Le statut juridique du syndic (Insolvenzverwalter) en droits allemand et français“. *Petites Affiches. Le Quotidien Juridique – La Loi* 159, 4-8.
- Gergen, T. 2004. „Juristisches Übersetzen: Französisch-Deutsch“. *Revista de Llengua i Dret* 41, 309-313.
- Gergen, T. 2003. „L’authentification des actes juridiques en droit civil allemand par le notaire et d’autres officiers publics“. *Petites Affiches. Le Quotidien Juridique – La Loi* 200, 6-10.

---

Tunde Opeibi  
Lagos  
**Aspects of language use in a selected courtroom discourse in Lagos,  
Nigeria**  
Fr 12.00-12.30

---

The law court is a sociolinguistic province where language is used to accomplish specific purposes. The courtroom is particularly differentiated from other settings because of the peculiar linguistic exchanges that take place among the participants in this universe of discourse. As an institutional setting, laid-down procedures usually dictate the conducts of the interactants with respect to what they can say and how they say what they can say.

To the average citizens who are laymen, language use in court proceedings is mystified, coded and 'strange' (Tiersma, 1999). The situation becomes more worrisome when one considers the fact that the law, which is expected to be a communal property, is couched in a language that alienates about 70% of Nigerian population.

This paper examines the significant features of language use in legal proceedings/courtroom discourse in Lagos, Nigeria. The approaches of discourse analysis and selected concepts from the Speech Act Theory (Austin, 1962 & Searle, 1969) provide the framework for the description and analysis of the linguistic exchanges in the courts. It utilises extracts from texts recorded from criminal and/or civil proceedings from Yaba Magistrate Court in Lagos metropolis.

The study provides insights into the communicative strategies in an L2 courtroom setting. Different discourse acts are identified and coded. The analysis and interpretation of other lexico-semantic and non-linguistic features found in the texts produced during the proceedings demonstrate the technicalities that characterise legal language. The survey conducted within and outside the court premises confirms the assumption that much of the Nigerian legal system is far removed, language wise, from the generality of the population whom these laws govern (Adetugbo, 1990).

The study thus raises the question on the viability and appropriateness of the choice English as the language of legal proceedings in non-native environments. It argues that a more acceptable linguistic code should be adopted to perform this very important regulatory function in the society.

---

Silke von Lewinski  
MPI München  
**Mehrsprachigkeitsprobleme auf europäischer und internationaler Ebene  
beim Urheberrecht**  
Fr 13.00-13.30

---

Auf dem Gebiet des Urheberrechts, einer Materie, die schon immer sehr international ausgerichtet war, zeigt sich besonders gut, wie eng die Sprache mit dem Recht verbunden ist und wie wichtig es ist, in welcher Sprache man sich ausdrückt. Die Existenz zweier grundsätzlich unterschiedlicher Rechtssysteme in der Welt – im Allgemeinen des anglo-amerikanischen und des kontinental-europäischen Rechtssystems – ist im Urheberrecht in Gestalt des Copyright Systems und des Urheberrechts- oder Droit d’auteur-Systems allgegenwärtig. Dieser Beitrag wird konkrete Beispiele dafür geben, wie gleiche Rechtsinstitute in verschiedenen Sprachen unterschiedlich ausgedrückt werden und wie dies unmittelbar mit dem jeweiligen Rechtssystem zusammenhängt. Daran anknüpfend sollen die Probleme bei der Rechtsharmonisierung in der Europäischen Gemeinschaft anhand von Beispielen gezeigt werden.

Darüber hinaus wird ein Augenmerk auf die Dominanz der englischen Sprache bei internationalen Verhandlungen, aber auch bei internationalen wissenschaftlichen und anderen Konferenzen sowie auf dem Publikationsmarkt geworfen und über mögliche Folgen für das Rechtsdenken und die Weiterentwicklung des Rechts in diesem Bereich weltweit nachgedacht.

## AG 10 Fortgeschrittene Lernervarietäten

Koordination: Patrick Grommes, Maik Walter

### Programm

Mittwoch 22.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
13.45-14.15	Patrick Grommes, Maik Walter	<i>Fortgeschrittene Lernervarietäten: Einführung in die AG</i>
14.15-15.15	Joybrato Mukherjee	<i>Referenzlernerkorpora und lokale Lernerkorpora als komplementäre Ressourcen in der</i>
15.15-15.45	Christiane Bongartz	<i>Lernersprachenanalyse Cross-linguistic comparability of advanced learner varieties – some methodological considerations</i>
15.45-16.15	Kaffeepause	
16.15-16.45	Anke Lüdeling	<i>Mehrdeutigkeiten und Kategorisierung: Probleme bei der Annotation von Lernerkorpora</i>
16.45-17.15	Eva Lippert	<i>“Der” oder “ein”? Probleme von Nichtmuttersprachlern mit der Artikelsetzung im Deutschen</i>
17.15-17.45	Ursula Maden-Weinberger	<i>Modality in Learner German</i>
17.45-18.15	Ulrike Gut	<i>Phonology in advanced learners of German – an approach based on the LeaP corpus</i>

Donnerstag 23.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
09.00-09.30	Heide Wegener	<i>Was wird zuletzt gelernt und warum?</i>
09.30-10.30	Christine Dimroth, Stefanie Haberzettel	<i>Verbflexion in der L2 Deutsch: Warum sind kindliche L2 Lerner schneller "fortgeschritten"?</i>
10.30-11.00	Josje Verhagen	<i>The Acquisition of the Non-modal Auxiliary 'Hebben': The Key to Syntactic Finiteness in Dutch as a Second Language?</i>
11.00-11.30	Kaffeepause	
11.30-12.00	Nicole Schumacher	<i>The Role of Grammatical Aspect in the Acquisition of Tense: Evidence from Advanced Learner Varieties</i>
12.00-12.30	Natascha Sakhonenko, Barbara Schmiedtowa	<i>L1 Italian – L2 German The role of grammatical aspect in event encoding: A comparison of Czech and Russian advanced learners of German</i>

Freitag 24.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
10.30-11.30	Marcus Callies, Konrad Szczesniak	<i>Argument Realization and Syntactic Weight: The Use of the Dative Alternation by Advanced German and Polish Learners of English - A Learner Corpus-Based Study</i>
11.30-12.30	Heidi Byrnes, Hiram Maxim	<i>Developing Advanced German Writing Ability: A Curriculum-based Perspective</i>

12.30-13.30

Norbert Dittmar, Joel Walters

*The Acquisition of  
Pragmatic Markers in  
the Corpus of Native  
Russian Advanced  
Learners of German  
and Modern Hebrew*



---

Joybrato Mukherjee  
Justus-Liebig-Universität Gießen  
**Referenzlernerkorpora und lokale Lernerkorpora als komplementäre  
Ressourcen in der Lernaltersprachenanalyse**  
Mi 14.15-15.15

---

Dieser Vortrag stellt zum einen die deutsche Komponente des gesprochenen Lernaltersprachenkorpus LINDSEI-Ger (*Louvain International Database of Spoken English Interlanguage – the German Component*) vor, das aus 50 Interviews mit fortgeschrittenen Lernern des Englischen, deren Muttersprache Deutsch ist, besteht und ca. 100,000 Wörter umfaßt. Dieses Korpus wird damit das größte gesprochene Referenzlernerkorpus darstellen, auf dessen Grundlage sich die typische Sprachverwendung eines abstrakten und idealisierten "archetypal learner" (Granger 1998: 7) beschreiben läßt. Beispielsweise läßt sich mit Hilfe von LINDSEI-Ger empirisch zeigen, inwieweit sich die *spoken grammar* fortgeschrittener deutscher Lerner (z.B. im Bereich der Verwendung von *discourse markers*) bereits an die muttersprachliche Zielnorm angenähert hat und inwiefern noch Verbesserungsbedarf in der Lernaltersprache besteht. Zum anderen wird gezeigt, daß es ebenfalls sinnvoll ist, wenn in konkreten Lerngruppen – beispielsweise in der Sekundarstufe II – "local learner corpora" (Seidlhofer 2002: 213) erstellt werden, die den Sprachstand der eigenen Lerner zu unterschiedlichen Zeitpunkten im Lernprozeß erfassen. Solche lokalen Lernerkorpora können beispielsweise durch die Computerisierung von Klausuren oder Facharbeiten vergleichsweise einfach kompiliert werden und nicht nur in bezug auf die gesamte Lerngruppe, sondern auch im Hinblick auf individuelle Lerner und/oder spezifische Fehlertypen untersucht werden. Am Beispiel des in Kooperation mit dem Hainberg-Gymnasium Göttingen kompilierten *Giessen-Göttingen Local Learner Corpus of English* (GGLLC) wird gezeigt, daß solche kleinen lokalen Lernerkorpora durchaus wertvolle Ressourcen für Unterrichtende und Lernende "vor Ort" darstellen (z.B. für die Analyse des eigenen lernaltersprachlichen Gebrauchs von *discourse markers*), während große Referenzlernerkorpora vor allem für die Erstellung und Verbesserung von Unterrichtsmaterialien, Lernerwörterbüchern usw. relevant sind.

- Granger, S. 1998. "The computer learner corpus: a versatile new source of data for SLA research." In: S. Granger (hg.) *Learner English on Computer*. London: Longman, 3-18.
- Seidlhofer, B. 2002. "Pedagogy and local learner corpora: working with learning-driven data." In: S. Granger, J. Hung & S. Petch-Tyson (hg.) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins, 213-234.

---

Christiane Bongartz  
Universität zu Köln

**Cross-linguistic comparability of advanced learner varieties – some  
methodological considerations**

Mi 15.15-15.45

---

The acquisition of linguistic competence in a second language always occurs in interdependence with the learners' first language. Constructs such as Klein and Perdue's (1997) Basic Variety capture in a very general way what learners have in common at the early stages of acquisition, with no strong differentiating role for L1 influence. To conceptualize learner knowledge in Advanced Learner Varieties, however, it is methodologically mandatory to factor in how different L1s impact acquisition. Especially relevant in this respect are ceiling effects, where L2 learners acquire near-native competence in most areas of morpho-syntax, but systematically show L1-related developmental gaps in other areas.

A case in point is the acquisition of the English definite article *the*, whose overt distribution depends on the intersection of syntactic, semantic, and pragmatic properties that are left underdetermined in the input. The role of the learner's L1 in the early development of the English interlanguage article system is well established (Master, 1988; Parrish, 1987). However, to date, there exists no evidence that clearly shows L1-related ceiling effects in article acquisition and that also shows that such effects constitute isolated gaps in otherwise near-native ALV. The reason for this lies in two methodological challenges imposed by criteria that apply to data which can inform a meaningful cross-linguistic comparison. One obstacle consists in establishing baseline proficiency scores, and the other in the different routes learners from different L1 backgrounds have taken towards acquisition. Differences in exposure cannot easily be separated from L1 effects.

My workshop contribution discusses these challenges in detail. Drawing on data from a cross-linguistic studies of ALV English with four L1 backgrounds (French, German, Hungarian, and Russian) I discuss the methodology involved in putting together a data corpus of mixed character. Much can be done to increase comparability, I argue, by analyzing learner production and grammaticality judgments from a variety of sources (interview data, a questionnaire, and story-retelling in this case). In a very general way, one has to think about more relaxed demands on comparability in order to capture ceiling effects and their place in ALVs.

---

Anke Lüdeling  
Humboldt-Universität zu Berlin  
**Mehrdeutigkeiten und Kategorisierung:  
Probleme bei der Annotation von Lernerkorpora**  
Mi 16.15-16.45

---

Lernerkorpora spielen in der Untersuchung von Fremdspracherwerb zunehmend eine größere Rolle (Granger, Hung & Petch Tyson 2002; dies gilt zumindest für das Englische als Fremdsprache – für das Deutsche als Fremdsprache gibt es bisher kaum größere, frei verfügbare Lernerkorpora). Während unannotierte Lernerdaten für statistische Vergleichsuntersuchungen (Contrastive Interlanguage Analysis) genutzt werden können, können Lernerdaten, die mit Fehlerannotationen versehen sind, für detailliertere qualitative und quantitative Studien herangezogen werden. Bei der Fehlerannotation gibt es allerdings mehrere grundsätzliche Schwierigkeiten

(a) Fehlerdefinition: was als Fehler gewertet werden soll, ist unklar (siehe die Diskussion in Corder 1974 und viele andere)

(b) Zielhypothesen: für eine fehlerhafte Lerneräußerung kann es unterschiedliche Zielhypothesen geben

Dazu kommen praktische Probleme in der Annotation und bei der Erstellung der Tagsets

(c) Fehlerexponent: in vielen Tagsets ist der Fehlerexponent unklar.

(d) Mehrdeutigkeit: je nach theoretische

(e) Annotationsqualität: um eine gewisse Qualität (Inter-Annotator Agreement) bei der Annotation sicherzustellen, muss ein Tagset relativ grobkörnig sein

In meinem Vortrag wird Falko (für fehlerannotiertes Lernerkorpus), ein frei zugängliches Lernerkorpus mit Texten von fortgeschrittenen Lernern des Deutschen als Fremdsprache vorgestellt (siehe <http://www.linguistik.hu-berlin.de/korpuslinguistik/projekte/falko/>, siehe auch Lüdeling et al. 2005). Um den angesprochenen Problemen bei der Fehlerannotation zu begegnen, wird in Falko eine mehrschichtige so genannte standoff-Architektur verwendet, in der unterschiedliche Zielhypothesen angegeben werden können und jede Fehlerebene unabhängig von anderen annotiert wird.

Corder, S. P. 1974. Error analysis, in J. P. Allen and S. P. Corder (eds.) *The Edinburgh Course in Applied Linguistics*, Vol. 3. London: Oxford University Press

Granger, S.; Hung, J.; Petch-Tyson, S. (eds) 2002. *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins

Lüdeling, A.; Walter, M.; Kroymann, E. & Adolphs, P. 2005. Multi-level error annotation in learner corpora. In: *Proceedings of the Corpus Linguistics 2005*, Birmingham.

---

Eva Lippert  
Humboldt-Universität zu Berlin  
**Der oder ein? Das ist die/eine Frage. Probleme von Lernern mit  
Artikelwörtern**  
Mi 16.45-17.15

---

An Korpusstudien hat sich gezeigt, dass Lerner des Deutschen als Fremdsprache Fehler machen, die Muttersprachler in Vergleichskorpora nicht machen. Eine Untergruppe dieser Fehler, „Definitivitätsfehler“, sollen hier vorgestellt werden. Als „Definitivitätsfehler“ in der Korpusuntersuchung wurden Nominalphrasen mit fehlerhafter Artikelwortsetzung verstanden. Also z.B. ?? *Im Text wird aber auch Fragezeichen vor die oben genannte Erklärung gestellt.*

Definitivität wird hier als Phänomen sowohl der Syntax als auch der Semantik angenommen. Hauptaugenmerk liegt auf der Systematisierung der häufigsten Fehlerursachen.

Auf syntaktischem Gebiet kann man Aussagen über falsch verwendetes Inventar treffen. Das Problem insbesondere der Nicht-Artikelsetzung auf der einen Seite und der Übergeneralisierung des bestimmten Artikels auf der anderen Seite hat sich als Haupt-Fehlerursache auf syntaktischem Gebiet herausgestellt.

In der Semantik dagegen hat der größte Teil aller vorkommenden Fehler mit einer nicht oder falsch verstandenen generischen Lesart zu tun. Fehler treten hier bei definiten ebenso wie bei indefiniten Artikeln oder artikellosen Nominalphrasen auf. Über Ursachen und Verteilung der Fehler kann eine korpuslinguistische Untersuchung Aufschluss geben. Mit ihr kann können verschiedene Fehlerursachen untersucht werden, z.B.:

- Haben Sprecher von Artikelsprachen es leichter als Sprecher artikelloser Sprachen?
- Welche Muster werden aus der Muttersprache übernommen?
- Treten Definitivitätsfehler häufiger an Subjekten oder Objekten auf, und inwiefern lässt sich dies auf die Muttersprache des Schreibers zurückführen?

Voraussetzung für diese Untersuchungen ist/war eine multi-level-standoff-Architektur, d.h. eine Architektur, in der die Hypothese darüber, was die korrekte Form gewesen wäre, explizit festgelegt wurde. Die Architektur hat neben dem Vorteil, Syntax und Semantik als Teilaspekte des Phänomens „Definitivität“ abgleichen zu können, das große Plus, dass jederzeit neue Aspekte, wie z.B. Kongruenz, mit aufgenommen werden können.

Löbner, S. 1985. *Definites*. In: Journal of Semantics 4: 279-326.

Lüdeling, A. et al. 2005. *Multi-level error annotation in learner corpora*. In: Proceedings of the Corpus Linguistics 2005, Birmingham.

Lyons, C. 1999. *Definiteness*. Cambridge: Cambridge University Press.

---

Ursula Maden-Weinberger  
Lancaster University  
**Modality in Learner German**  
Mi 17.15-17.45

---

This paper reports on the compilation and annotation of a 200,000-word corpus of advanced learner German written by British undergraduate students at Lancaster University. The main focus is a study on modality in the students' argumentative writing.

The compilation of learner corpora and analysis of learner language (also in comparison to the language of native speakers) has proven useful in research on learner errors, second language acquisition and the improvement of language learning and teaching (e.g. Granger/Hung/Petch-Tyson, 2002; Kettemann/Marko, 2002).

Modality plays a pivotal role in all texts and discourses that deal with the discussion or examination of practical or theoretical problems (Zifonun et al. 1997) and thus forms an integral part of students' argumentative writing. Furthermore, the various linguistic manifestations of modality (modal verbs, modal adverbials, subjunctive mood and other lexical markers of modality) are high frequency phenomena in the German language. Learners are therefore confronted with them early on in the acquisition process and encounter them frequently as soon as they begin to deal with authentic texts. However, they appear to have difficulties in this area, even in the advanced stages. Working with a learner corpus allows for an extensive investigation of all the different linguistic manifestations of modality in order to give a comprehensive picture of the learners' use of modality features, not just in terms of their respective frequencies, but also in their respective pragmatic contexts. On the basis of a thorough corpus-based contrastive analysis of L1, L2 and various stages of interlanguage (based on Granger's (1996) Integrated Contrastive Model), including some manual annotation of the data, patterns of over-, under- or misuse of modal expressions can be established. This can help finding explanations for the differing language use, which in turn can be give useful insights with regards to the development of better teaching materials.

Granger, S. 1996. "From CA to CIA and Back: An Integrated Approach to Computerized Bilingual and Learner Corpora". In: Aijmer, K.; Altenberg, B. & Johansson, M. (hg). *Languages in Contrast*. Papers from the Symposium on Text-based Cross-linguistic Studies. Lund 4-5 March 1994. Lund: Lund University Press.

Granger, S.; Hung, J. & Petch-Tyson, S. (hg.) 2002. *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins.

Kettemann, B. & Marko, G. (hg.) 2002. *Teaching and Learning by Doing Corpus Analysis*. Amsterdam: Rodopi

Zifonun, G.; Hoffman, L. & Strecker, B. 1997. *Grammatik der deutschen Sprache*. Schriften des Instituts für deutsche Sprache Bd. 7. Berlin-New York: De Gruyter.

---

Ulrike Gut  
Albert-Ludwigs-Universität Freiburg  
**Phonology in advanced learners of German – an approach based on the  
LeaP corpus**  
Mi: 17:45 – 18.15

---

An increasing number of studies have reported on exceptionally advanced learners who have successfully acquired the phonology of a foreign language even as adults and whose speech is rated as unaccented by native speakers (Schneiderman & Desmarais 1988, Moyer 1999, Bongaerts et al. 1997). Current research in second language phonology favours experimental data, tends to be based on a limited number of participants and one particular speech style. Recently, it has been suggested that a corpus-linguistic approach should be introduced into research in language acquisition. It is widely argued that a corpus-based methodology can complement the current research methods in second language learning and possibly compensate some of their weaknesses (e.g. Granger, Hung and Petch-Tyson 2002).

This talk presents the recently completed LeaP corpus, a fully text-to-tone aligned corpus of learner English and learner German. The corpus consists of about 12 hours of phonetically annotated recordings adding up to 73.941 words. It comprises four different types of speech: free speech in an interview situation, reading passage style, retellings of a story and readings of nonsense word lists. The German subcorpus contains recordings with 59 non-native speakers with 24 different native languages, comprising advanced learners, learners before and after a training course in prosody, and learners before and after a stay abroad.

The corpus analysis shows the range of variation in L2 German phonology and systematic differences between native and non-native phonology. Furthermore, it provides evidence for distinct structural differences between advanced learners' and less advanced learners L2 phonologies. This will be exemplified in the areas of consonant cluster reduction, syllabification, vowel reduction, speech rhythm, intonation and fluency in general. Finally, this talk describes factors of advanced phonological attainment by identifying learner variables and contextual correlates of exceptionally successful second language phonology learning.

- Bongaerts, T., van Summeren, C., Planken, B. & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition* 19, 447-465.
- Granger, S., Hung, J. & Petch-Tyson, S. (eds.), 2002. Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching. Amsterdam: Benjamins.
- Moyer, A. (1999). Ultimate attainment in L2 phonology. *Studies in Second Language Acquisition* 21, 81-108.
- Schneiderman, E. & Desmarais, C. (1988). The talented language learner. Some preliminary findings. *Second Language Research* 4, 91-109.

---

Heide Wegener  
Universität Potsdam  
**Was wird zuletzt gelernt und warum?**  
Do 9.00-9.30

---

Schaut man sich die Daten fortgeschrittener DaZ- oder auch DaF-Lerner an, so fällt auf, dass es unabhängig vom Alter der Lernenden und von ihrer L1 sprachliche Phänomene gibt, die regelmäßig besonders spät, unzureichend oder auch gar nicht erworben werden.

An von mir in einem Zeitraum von 3-4 Jahren erhobenen Daten aus dem natürlichen DaZ-Erwerb von Grundschulkindern aus Polen, Russland und der Türkei einerseits, an Daten erwachsener DaF-Lerner aus Burkina Faso und aus Spanien andererseits soll gezeigt werden, um welche Phänomene es sich dabei handelt. Nicht zufällig finden sich in meinen "späten" Daten genau die Phänomene, die nicht zur Basic Variety gehören, im wesentlichen also Bereiche der (Flexions)Morphologie, wie z.B. attributive Adjektive, und der komplexen Syntax, wie z.B. Passivsätze.

Zur Erklärung dieser fortgeschrittenen Lernervarietäten und des späten Erwerbs dieser unterschiedlichen Phänomene wird gezeigt, dass es einerseits die Komplexität der Regeln innerhalb der deutschen Nominalflexion ist, die den Spracherwerb so schwierig macht. Andererseits kann der späte Erwerb bzw. die hohe Fehleranfälligkeit aber auch durch bestimmte Lernstrategien erklärt werden, etwa dürfte das Beibehalten der Akkusativmorphologie für das Passivsubjekt auf dem Lernen mithilfe von Chunks beruhen, die vom Aktivsatz aus gebildet wurden.

---

Christine Dimroth, Stefanie Habertzettl  
MPI für Psycholinguistik Nijmegen, Universität Bremen  
**Verbflexion in der L2 Deutsch:**  
**Warum sind kindliche L2 Lerner schneller "fortgeschritten"?**  
Do 9.30-10.30

---

Phänomene wie die Basisvarietät (Klein & Perdue 1997) oder hartnäckige Fossilierungen auf einer Stufe, auf der nur die wichtigsten kommunikativen Bedürfnisse mit dem vorhandenen Repertoire erfüllt werden können, sind in erster Linie für erwachsene L2 Lerner belegt. In der Tat erreichen kindliche L2 Lerner oft fortgeschrittenere Varietäten bis hin zur *near nativeness* (vgl. z.B. die Lernerin *Nora* aus Wong Fillmore 1976). Kinder scheinen bereits in den elementaren Phrasen des ungesteuerten Erwerbs auf morphologische Details aus dem Input zu achten und eher bereit zu sein, diese zunächst unanalysiert zu reproduzieren. Diese Eigenschaft teilen sie mit Kindern im Erstspracherwerb, jedoch scheinen kindliche L2 Lerner weniger

Zeit zu benötigen, um aus dem Fundus an Formen, den sie angesammelt haben, zielsprachliche Kookkurrenzen zu abstrahieren.

Dies zeigen wir anhand von umfangreichen Langzeitstudien des Verbflexionserwerbs in der Zweitsprache Deutsch durch drei Kinder (7 – 9 Jahre) mit L1 Russisch. Der Vergleich mit entsprechenden Ergebnissen aus der Erstspracherwerbsforschung (vgl. Bittner et al. 2003) macht das hohe Tempo deutlich, mit dem das Formenrepertoire der Zweitsprache erworben und adäquat verwendet wird. Dies wirft die Frage auf, ob kindliche L2 Lerner auf dem Weg zu fortgeschrittenen Lernervarietäten zwangsläufig eine Art Basisvarietät durchlaufen müssen.

Bittner, D. ; Dressler, W. U. & Kilani-Schoch, Marianne 2003 (eds.). *Development of verb inflection in first language acquisition. A cross-linguistic perspective*. Berlin/New York: de Gruyter.

Klein, W. & C. Perdue 1997. *The Basic Variety (or: Couldn't natural languages be much simpler?)*. *Second Language Research* 13 (4): 301-347.

Wong Fillmore, L. 1976. *The Second Time Around: Cognitive and Social Strategies in Second Language Acquisition*. Ph D. Thesis, Stanford University. Ann Arbor, MI: Xerox University Microfilms.

---

Josje Verhagen

MIP Nijmegen, Free University Amsterdam

**The Acquisition of the Auxiliary 'Hebben': The Key to Syntactic Finiteness in Dutch as a Second Language?**

Do 10.30-11.00

---

In this presentation, evidence is presented for the important role of the auxiliary 'hebben' in the acquisition of Dutch as a second language. It is argued that the acquisition of this auxiliary is a crucial step in learners' development towards a Post-Basic Variety stage, where the finite verb is placed in second clausal position.

Previous research showed that L2-learners of English and German initially use pre-verbal negation with lexical verbs, whereas they use post-verbal negation with auxiliaries from the beginning (Parodi 200, Becker, 2005). Moreover, it was found that L2-learners of German extend their use of post-verbal negation to lexical verbs after they have started to produce the auxiliary 'haben' (Becker 2005). A similar pattern was found for inversion: L2-learners of Dutch use inverted structures only after they have started to produce 'hebben' (Jordens 2004).

The present study seeks to further investigate the relation between the acquisition of 'hebben', on the one hand, and inversion/post-verbal negation - or syntactic finiteness - (Lasser, 1997), on the other. For this purpose, an experiment was carried out that measured production and comprehension. The experiment contained free and controlled production tasks (film retelling, picture description, sentence imitation) and an implicit, reaction-timed judgment task (sentence matching). Subjects were Moroccan and Turkish learners of Dutch.

The results support the idea that 'hebben' enables learners to reach a Post-Basic Variety stage, where utterances are syntactically finite. In the production tasks,



most learners producing 'hebben' also used inversion and post-verbal negation, while none of the learners not producing 'hebben' did so. The important status of 'hebben' did not become apparent from the production data alone, however: the data from the imitation and sentence matching tasks showed that learners not producing 'hebben' may show passive knowledge of inversion and topicalisation in sentences with 'hebben', but not in sentences without 'hebben'.

- Becker, A. 2005. The semantic knowledge base for the acquisition of negation and the acquisition of finiteness. In H. Hendriks (Ed.), *The structure of learner varieties*. Berlin/New York: Mouton de Gruyter.
- Jordens, P. 2004. *Systematiek en dynamiek bij de verwerving van finietheid*. *Toegepaste Taalwetenschap in Artikelen*, 71, 2, 9-22.
- Parodi, T. 2000. Finiteness and verb placement in second language acquisition. *Second Language Research*, 16, 4, 355-381.

---

Nicole Schumacher  
Humboldt-Universität zu Berlin  
**The Role of Grammatical Aspect in the Acquisition of Tense:  
Evidence from Advanced Learner Varieties L1 Italian – L2 German**  
Do 11.30-12.00

---

One of the main characteristics of advanced learner varieties is the non-native language use rather than the violation of norms of the target language. In their production, advanced learners often select L2 items which are grammatical, but wouldn't be selected by a native speaker. In their comprehension, they often prefer interpretations of items which are possible, but wouldn't be preferred by a native speaker. Non-native language use is a challenge for teachers in L2 classrooms, since it is difficult to explain what it is that makes a certain language sound "natural".

This paper seeks to investigate the use of the German past tenses by advanced Italian learners. Contrastive analyses (cf. Schumacher 2005) have shown that the German and Italian tense systems correspond in form, meaning and use in many respects, but differ substantially in the realization of certain aspectual meanings. Following studies which provide evidence for the fact that grammatical categories in the L1 lead speakers to rely on certain conceptual categories more than others in language production (cf. Stutterheim 1999), a prediction is made for L2 language comprehension: Advanced Italian learners of German prefer non-native habitual readings of the German preterite (Präteritum) with telic verbs and/or terminative duration adverbials.

The aim of my paper is to test this hypothesis. Test instrument is a questionnaire with acceptability judgement tasks and sentence completion tasks. Participants are advanced Italian learners of German. Test items include sentences of the following kind: *Wir standen früh auf* (We got up early) and *Sie tanzte bis Mitternacht* (She danced until midnight). Results of this study are presented and discussed in two ways: within the framework of the discussion on the interaction between tense, aspect and actionality in SLA (cf. Bardovi-Harlig 2000) and in the light of L2 classroom problems.

- Bardovi-Harlig, K. 2000. *Tense and Aspect in Second Language Acquisition. Form, Meaning and Use*. Oxford: Blackwell.
- Schumacher, N. 2005. *Tempus als Lerngegenstand. Ein Modell für Deutsch als Fremdsprache und seine Anwendung für italienische Lernende*. Tübingen: Narr.
- von Stutterheim, Chr. 1999. "How Language Specific are Processes in the Conceptualizer?" In: R. Klabunde & Chr. von Stutterheim (eds.): *Representations and Processes in Language Production*. Wiesbaden: Deutscher Universitätsverlag, 153-179.
- 

Natascha Sakhonenko, Barbara Schmiedtová  
Ruprecht-Karls-Universität Heidelberg

**Die Rolle des grammatischen Aspekts in der Ereignis-Enkodierung: Ein Vergleich zwischen tschechischen und russischen Lernern des Deutschen**  
Do 12.00-12.30

---

Frühere Studien über Zeitstrukturen der Ereignisse, die semitische, germanische und romanische Sprachen untersuchten, zeigten sprachspezifische Effekte in der Art und Weise, wie Ereignisse aufgebaut werden (event construal). Es wurde gezeigt, dass die zugrundeliegenden Prinzipien durch die Perspektive bestimmt werden, aber auch starke Verbindungen zu grammatischen Mustern aufweisen (vgl. Carroll & v. Stutterheim, 2003, Meyer-Klabunde und v. Stutterheim, 1999). Zusätzlich haben neuere L2-Studien bewiesen, dass sogar sehr fortgeschrittene Lerner bei der Versprachlichung temporaler Ereignisse in der Zielsprache auf die Konzeptualisierungs-Strategien ihrer Muttersprache zurückfallen (vgl. Carroll und Lambert, 2003). In unserem Vortrag konzentrieren wir uns auf die Rolle des Aspekts in der Ereignis-Enkodierung durch tschechische und russische fortgeschrittene Lerner des Deutschen. Sowohl die beiden Aspektsysteme weisen einen grammatikalisierten Unterschied zwischen dem perfektiven und dem imperfektiven Aspekt auf. Das Deutsche dagegen kennt keine grammatikalisierten Aspektformen, drückt aber den lexikalischen Aspekt aus. Zusätzlich haben Tschechisch und Russisch feine, aber wichtige Unterschiede beim Gebrauch des perfektiven Aspekts und Präsensformen. Für die Datenerhebung dieser Studie wurde eine Reihe von kurzen Videoclips mit alltäglichen Situationen verwendet. Verschiedene aspektuelle Bedeutungen, z.B. abgeschlossen vs. im Verlauf begriffen, wurden in den Situationen dargestellt. Wir untersuchen und stellen mehrere Forschungsfragen zur Debatte: (1) Inwieweit unterscheiden sich das tschechische und russische Aspektsystem? (2) Was sind die Folgen dieser Unterschiede für die Konzeptualisierung zeitlicher Ereignisse im Deutschen als L2? Das bedeutet (a) verwenden tschechische und russische fortgeschrittene Lerner des Deutschen Konzeptualisierungsstrategien ihrer jeweiligen L1 oder (b) machen sie von Strategien Gebrauch, wie sie deutsche Muttersprachler typischerweise anwenden?

- Carroll, M. & Lambert, M. 2003. "Information structure in narratives and the role of grammaticised knowledge: A study of adult French and German learners of English." In: Ch. Dimroth, M. Starren (Hrsg.). *Information Structure and the Dynamics of Language Acquisition*. Amsterdam/Phil.: Benjamins, S. 267-287.

- Carroll, M. & v. Stutterheim, Chr. 2003. "Typology and information organisation: perspective taking and language-specific effects in the construal of events." In: A. Ramat (Hrsg.), *Typology and Second Language Acquisition*. Berlin: de Gruyter, S. 365-402.
- Meyer-Klabunde, R. & v. Stutterheim, Chr. 1999. *Representations and Processes in Language Production*. Wiesbaden: DUV.
- 

Konrad Szczesniak, Marcus Callies  
Silesian University Sosnowiec, Philipps-Universität Marburg  
**Argument Realization and Syntactic Weight: The Use of the Dative  
Alternation by Advanced German and Polish Learners of English - A  
Learner Corpus-Based Study**  
Fr 10.30-11.30

---

This paper investigates the use of the English dative alternation (DA) by advanced German and Polish learners of English in written production taken from the International Corpus of Learner English and comparable native speaker writing. The analysis focuses on 15 highly frequent DA-verbs identified in the pertinent literature and checked through comparison with a large reference corpus of English. The study examines the frequency of use of the two alternative postverbal constituent orderings that are possible: the double object construction as in (1), and the prepositional construction, exemplified in (2) below:

- (1) Although he was unable to change the constitution into a fully presidential one he did manage to include two clauses which gave the president a lot of power.
- (2) The ambiguities such as in Article 34 and 37 over the way legislation is past also gave wide reaching powers to the President.

Moreover, we aim to identify reasons for the occurrence of either variant. Descriptive, corpus-based and experimental studies on verb alternations and word order variants such as the DA have shown that their appearance is influenced by factors such as information structure (IS) and syntactic weight (e.g. Wasow and Arnold 2003). Thus, the DA in particular is not only interesting in terms of argument realization, but also from the point of view of IS since it has been shown to be a weight-sensitive construction (e.g. Wasow 1997).

While the acquisition of argument structure alternations has received considerable attention within UG-based approaches to Second Language Acquisition (SLA) (Juffs 2000), functional syntactic aspects of these phenomena have largely been neglected. There is evidence that advanced learners, who have usually mastered the formal system of the L2, still have problems with certain form-function mappings in the target language, for example with respect to areas where syntax, semantics and pragmatics interact. Information-structural concepts such as the principle of end weight are not well-examined topics in the sparse SLA research on advanced learners. By studying learners from two different L1s, the present study aims to provide evidence for the question in how far such principles are operative in the advanced learner variety.

Juffs, A. 2000. „An Overview of the Second Language Acquisition of Links between Verb Semantics and Morpho-Syntax“. In: J. Archibald (hg.) *Second Language Acquisition and Linguistic Theory*. Malden: Blackwell, 187-227.

Wasow, T. 1997. „Remarks on Grammatical Weight“, *Language Variation and Change* 9:1, 81-105.

Heidi Byrnes and Hiram H. Maxim

Georgetown University

**Developing Advanced German Writing Ability: A Curriculum-based Perspective**

Fr 11.30-12.30

---

For instructed L2 learners, attainment of advanced abilities is increasingly seen as related to literacy phenomena, particularly the development of reading and writing abilities. With regard to L2 writing development, research has offered substantive predictions in terms of patterns of syntactic complexity. However, analysis has typically relied on learner performances obtained from isolated and instructionally decontextualized composing tasks whose generalizability for a developmental trajectory, particularly toward advanced abilities, remains to be substantiated.

In order to understand adequately the relationship between instruction and learner development and to produce findings that are relevant for improving the same, this presentation reports on writing as investigated in a multi-year longitudinal and cross-sectional study of the development of German L2 syntactic complexity. It is based on writing data collected across the four years of an integrated undergraduate program at a U.S. university. The data comprise a corpus of nearly 250,000 words involving over 300 participants.

It will address issues related to the development of advanced writing abilities by

1. Contextualizing the research according to the learners, curriculum, and instructional approaches within an explicitly literacy-oriented program, wherein writing development toward advanced levels of ability is driven by the use of genre-based writing tasks throughout its sequence;

2. Outlining and justifying the research methodology employed for investigating syntactic complexity within L2 writing development, from cross-sectional, longitudinal, and task-based perspectives; and

3. Analyzing and interpreting data in terms of curricular predictions, instructional practices, and program-external research findings, with a focus on advanced levels.

*Developing multiple literacies: A curriculum renewal project of the German Department at Georgetown University, 1997-2000.* 2000

(<http://data.georgetown.edu/departments/german/programs/curriculum/>).

Byrnes, H. 2002. The role of task and task-based assessment in a content-oriented collegiate foreign language curriculum. *Language Testing*, 19(4), 419-437.

Cooper, T. C. 1976. Measuring written syntactic patterns of second language users of German. *Journal of Educational Research* 69, 176-183.

Ortega, L. 2000. *Understanding syntactic complexity: The measurement of change in the syntax of instructed L2 Spanish learners*. Unpublished doctoral dissertation, Honolulu, HI: University of Hawaii.

Ortega, L. 2003. Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics* 24, no. 4: 492-518.

Norbert Dittmar, Joel Walters  
Freie Universität Berlin, Bar-Ilan University Ramat Gan  
**The acquisition of pragmatic markers in the corpus of native Russian  
advanced learners of German and Modern Hebrew**  
Fr 12.30-13.30

---

In a joint project of the German Israeli Foundation (GIF) for three years (1997-2000) we conducted longitudinal research on the sociolinguistic integration of Russian Jews into German (Berlin) and Hebrew (Tel Aviv) target society. A detailed overview of research aims, sample, and linguistic stages in the acquisitional process can be found in (Dittmar, Spolsky & Walters, 2002). Our presentation will focus specifically on the use of connectives and particles in a sub-sample of speakers from this project.

Focussing on the German data, ND will give a detailed account of the syntactic positions the connective *also* can take in learner varieties and explain these positions in terms of acquisitional constraints and pragmatic functions. Are there particular domains of use which are characteristic of advanced learners and for which they show preferences? Beyond the investigation of *also*, ND will formulate hypotheses about formal and functional prerequisites of the pragmatic role of connectives and particles in advanced learner varieties presenting and discussing relevant data from the corpus.

Analyzing *az* (roughly equivalent to *also*) in Hebrew data, JW will show that there are two kinds of connectives and particles: (a) those limited to semantic and discourse organizational functions (*az* is a prime candidate) and (b) those with broader pragmatic and affective functions. JW will show that advanced stages of learning are marked by individual variation in the use of a small set of connectives and particles, which are used for sociopragmatic, affective and psycholinguistic functions, allowing learners to express subtle socio-communicative purposes.

It is argued that in order to define advanced learner varieties, we should pay more attention to discourse markers as indicators of acquisition, rather than reducing connectives and particles to their syntactic and semantic form. We will discuss our data in light of recent work by Halliday, Fraser, König, Blakemore, Pasch et al. (among others).

## AG 11 Korpusbasierte Behandlung nichtkompositioneller Phänomen

Koordination: Andrea Kowalski, Manfred Pinkal

### Programm

Mittwoch 22.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
13.45-14.15	Manfred Pinkal, Andrea Kowalski	<i>Begrüßung und Einführung</i>
14.15-14.45	Tibor Kiss, Katja Keßelmeier	<i>The syntax and semantics of seemingly ungrammatical phrasal combinations</i>
14.45-15.45	Ulrich Heid	<i>Syntagmatic and paradigmatic aspects of noun+verb-collocations: proposals for acquisition and representation</i>
15.45-16.15	Kaffeepause	
16.15-17.15	Arne Zeschel	<i>Corpus perspectives on constructional phraseology</i>
17.15-18.15	Jürgen Reischer	<i>Word and phrase blends from a (Corpus-)linguistic point of view</i>

Donnerstag 23.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
09.00-10.00	Angelika Storrer	<i>Corpus-based investigations on German support verb constructions</i>
10.00-10.30	Natalie Schluter	<i>On the applicability conditions of a support verb paraphrase operation</i>
10.30-11.00	Kow Kuroda, Keiko Nakamoto, Hajime Nozawa, Hitoshi Isahara	<i>Situations are „attractors“ of semantic interpretations</i>
11.00-11.30	Kaffeepause	
11.30-12.30	Ekaterina Stathi	<i>Idioms as lexical units: a corpus-based approach</i>

Freitag 24.02.2006

<b>Zeit</b>		<b>Titel</b>
10.30-11.30	Christiane Fellbaum	<i>German VP Idioms: Tough Nuts to Crack</i>
11.30-12.00	Gergely Pethő, Péter Csatár	<i>Recognition and processing of figurative language</i>
12.00-12.30	Aljoscha Burchardt, Katrin Erk, Anette Frank, Andrea Kowalski, Sebastian Pado, Manfred Pinkal	<i>Challenges in lexical semantics: Non-compositionality in SALSA corpus annotation</i>
12.30-13.30	Abschlussdiskussion	

---

Tibor Kiss, Katja Keßelmeier  
Ruhr-Universität Bochum  
**Potentially Ungrammatical Multiword Tokens and  
Part-of-Speech Tagging**  
Mi 14.15-14.45

---

Computational linguistics assumes a logical separation between the tokenization and categorization of words – typically as part of a part-of-speech tagger – and the processing of phrases based on the tokenization and separation.

Multiword tokens consisting of more than a single *orthographic* word pose an obvious challenge for the aforementioned separation. Apart from consisting of more than a single word, idioms and collocations also show a conventionalized semantics which resists a simple compositional analysis. Yet, tagging systems make mostly use of purely orthographic conventions and lists to tokenize and categorize atomic units for further syntactic and semantic processing.

As an immediate consequence of these considerations, part-of-speech tagging may result in category sequences which do not correspond to the grammar of a language. This problem has been addressed in early stages of generative grammar, cf. Chafe (1968). Surprisingly, Chomsky (1980) as well as di Sciullo/Williams (1987) assume that idiomatic multiword tokens may show deviant semantics, but definitely no deviant syntax. Nunberg et al. (1994) and Fleischer (1997) assume that the relevant tokens are ‘*words with spaces*’ and can be listed as such. Such listing proposals seem to be corroborated by the unproductivity of the pertinent expressions.

Unfortunately, this does not seem to be correct for the majority of the problematic cases. As an illustration, consider prepositions (and other particles) which productively combine in an otherwise ungrammatical fashion with a *noun*, or a combination of parts of an NP in (1).

- (1) *an Stelle, auf Anhieb, aus Erfahrung, gegen Ende, in Anlehnung, mit Ausnahme, nach Ablauf, ohne Ausnahme, seit Beginn, unter Berufung, von Bedeutung, wider Willen, zu Beginn*

The combination of P+N in (1) is productive. Still grammarians refuse to treat the sequences on a par with P+NP-combinations. Our talk will be concerned with a corpus-based analysis of the expressions illustrated in (2), focussing on syntactic productivity and the consequences for compositionality. We will particularly consider which constraints are imposed by syntactic productivity for a deviant semantic treatment, as e.g. suggested in terms of *decomposability* by Nunberg et al. (1994).



- Chafe, Wallace. 1968. *Idiomaticity as an anomaly in the Chomskyan paradigm*. Foundations of Language 4, S. 109-127.
- Chomsky, Noam. 1980. *Rules and Representations*. London: Blackwell.
- Di Sciullo, Anna Maria, Edwin Williams. 1987. *On the Definition of Word*. Cambridge/London: MIT Press.
- Fleischer, Wolfgang. 1997. *Phraseologie der deutschen Gegenwartssprache*. Tübingen: Max Niemeyer Verlag.
- Helbig, Gerhard, Joachim Buscha. 1998. *Deutsche Grammatik. Ein Handbuch für den Ausländerunterricht*. Leipzig: Langenscheidt.
- Nunberg, Geoffrey, Ivan A. Sag, Thomas Wasow. 1994. *Idioms*. Language 70/3, S: 455-489.
- Sag, Ivan A., Timothy Baldwin, Francis Bond, Ann Copestake, Dan Flickinger: *Multiword expressions: A pain in the neck for NLP*. Proceedings of the 3<sup>rd</sup> International Conference on Intelligent Text Processing and Computational Linguistics. S. 1-15.
- 

Ulrich Heid  
IMS Stuttgart

**Syntagmatic and paradigmatic aspects of noun+verb-collocations:  
proposal for acquisition and representation**

Mi 14.45-15.45

---

The proposed contribution deals with noun+verb-collocations, such as support verb constructions, or verbs with collocationally typical subjects and complements. Our working definition of the term "collocation" is based on Hausmann's (2004) lexicographic notion: for us, a collocation is a lexically determined binary word combination; the elements (Base and Collocate) have different status: the collocate is "synsemantic" (i.e. receives its semantic interpretation only in the collocation), whereas the base is "autosemantic". The semantics of the combination is thus not compositional, even though the base is used in (one of) its regular senses. German examples of the kinds of word combinations we analyse are "Vortrag+halten" (give+talk), or "aus dem Amt scheiden" (quit one's office). In an NLP-oriented dictionary, collocations should receive an equally detailed description as "single word lexemes". This concerns both their syntagmatic and their paradigmatic aspects. The former include morphosyntactic preferences (definiteness, modifiability, number preferences, etc.) and the function of collocations as predicates (subcategorization). The paradigmatic aspects (mainly relevant for text generation and information extraction) include morphological and lexical semantic relations between collocations and other lexical items, be they "single" or "multi-word". We present a requirements analysis towards a detailed model for a computational collocation lexicon and a few remarks on an implementation in the formalism of OWL-DL. The quantity of material to be covered (Sag et al. 2002 assume it to be equal in size to the "single word lexicon") calls for (semi-)automatic data acquisition. We report on experiments towards the automatic extraction of the syntagmatic properties and preferences of collocations from chunked German text corpora. Along with each collocation instance, we extract morpho-syntactic context parameters, e.g. tense, number and definiteness values. Summing and grouping over all instances from a given corpus allows us to derive corpus-specific preferences. First experiments with prenominal

participle constructions (e.g. "der gehaltene Vortrag", the talk [which was] given) seem to indicate that the method can be used for automatic pre-classification of collocation data.

- I. Sag, T. Baldwin, F. Bond, A. Copestake & D. Flickinger. "Multiword expressions: A pain in the neck for NLP". In: A. Gelbukh (hg.): CICLing-2002. Heidelberg/Berlin: Springer, 2002.
- F. Hausmann. 2004. Franz Josef Hausmann: "Was sind eigentlich Kollokationen?". In: K. Steyer (hg.) *Wortverbindungen – mehr oder weniger fest* [= IDS: Jahrbuch 2003]. 309 - 334.
- 

Arne Zeschel  
Universität Bremen  
**Collostructions: Investigating low-level schemas in usage-based  
grammar**  
Mi 16.15 -17.15

---

The present paper demonstrates a recent corpus-linguistic approach to grammatical constructions (or 'constructional idioms', a class of phenomena that challenge widespread ideas about lexical licensing and syntactic compositionality). This approach, known as *collostructional analysis*, consists of a family of methods geared towards investigating the interaction between lexical items and constructions as conceived in sign-based theories such as Cognitive Grammar and Construction Grammar (e.g. Stefanowitsch and Gries 2003, 2005; Gries and Stefanowitsch 2004). Using the construction in (1) below as an example, I give a short introduction to the method and argue for its rich potential for construction-based approaches to grammar in general:

- (1) a. X...sprüht vor Ideen, trieft vor Pathos, platzt vor Stolz, knistert vor Spannung...  
b. X... seethes with hate, explodes with delight, burns with ambition, sparkles with wit...

Specifically, I show how the method can be used for the precise identification of (i) fine-grained semantic restrictions associated with particular constructional templates and their conventional instantiations, (ii) instances of language-internal schema conflict and usage-pattern mixing between subtypes of a given construction, and (iii) crosslinguistic differences in the way that particular conceptual resources are tapped by equivalent constructions in two languages. Within a larger perspective, therefore, my study is intended to illustrate how work in cognitive semantics can benefit from quantitative corpus linguistics in interesting ways.

- Dowty, D. 2000. "The garden swarms with bees" and the fallacy of "argument alternation". In: Y. Ravin and C. Leacock (eds.), *Polysemy. Theoretical and computational approaches*. Oxford: OUP. 111-28
- Gries, S.T. and A. Stefanowitsch. 2004. Extending collostructional analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics*. 9.1: 97-129.
- Stefanowitsch, A. and S.T. Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8.2: 209-43.

Jürgen Reischer  
Universität Regensburg  
**Word and Phrase Blends from a (Corpus)Linguistic Point of View**  
Mi 17.15 – 18.15

---

Word blends like Germ. *Teuro* or *Mechatroniker* (Engl. *motel* or *basketball*) and phrase blends like *Tour de Lance* or *Brave Neuro World* are constructed neither formationally nor informationally in a regular fashion. The structures of these expressions seem to follow no conventional morphological or syntactical patterns as it is the rule for compounds and derivations, phrases and sentences. The lack of established construction rules for blending renders their automatic analysis difficult, especially as the combinatorial potential is tremendous.

However, due to the pragmatic restriction that blends have to be reconstructable by the recipient in the context of a (mainly written) communication, we must presume at least some formation regularities. Otherwise it would be impossible for the recipient to understand the intended meaning of the blend thereby losing its communicative effect. In order to abstract these formation patterns we need a data basis comprehensive enough to contain (almost) the entire palette of possible structural blending models for word and phrase blends. Such a corpus including a formational typology is now provided electronically by WordNexus, an application comprising several data bases and an interface for browsing the corpora.

At present, WordNexus maintains and manages a German and an English corpus of blend words and phrases containing more than 1000 items with each entry analyzed for its linguistic and non-linguistic properties (origin of an item; syntactic, semantic, and pragmatic features; structural analysis; meaning paraphrases); additionally, due to their formational identity with intentional blends, also slip of the tongue blends (contaminations) are provided.

From a theoretical point of view the question raises how blends are related to ‘normal’ word formation mechanisms like compounding, derivation, or abbreviation. A clear definition must provide a catalogue of checkable criteria able to define and (dis)qualify an expression as blend.

Further questions of interest concern the commonalities of word and phrase blends, or the semantic peculiarities of blends outperforming the informational potential of normal compounds.

Reischer, J. (2004): „The Compositionality of Blending – More than Compounding?“ Talk held at the interdisciplinary conference in Cognitive Science ‚Compositionality, Concepts & Cognition‘, 28.2.-3.3.2004, Düsseldorf. <http://www.das-phaenomen-sprache.de/Blending.PDF>

Schmid, H. U. (2003): „Zölibazis Lustballon“. *Muttersprache*, 3, 265-278.

Schulz, M. (2004): „Jein, Fortschritt und Ehrgeizhals. Wortkreuzungen in der deutschen Gegenwartssprache.“ *Zeitschrift für germanistische Linguistik*, 32.2, 286-306.

WordNexus (2005): <http://www-nw.uni-regensburg.de/~rej03046.infwiss.sprachlit.uni-regensburg.de/>

---

Angelika Storrer  
Universität Dortmund  
**Corpus-based Investigations on German Support Verb Constructions**  
Do 9.00-10.00

---

Support verb constructions (SVC), such as *to take a risk* or *to set into motion*, are composed of a support verb component (SV, e.g. *take, set*) and a predicative noun component (PredN, e.g. *risk, motion*). With respect to semantic compositionality, German SVCs cover a wide range: on the one hand, there are non-compositional idiomatic SVCs, such as *in Verzug geraten* (engl. *to fall behind*) or *Folge leisten* (engl. *to obey an order*). On the other hand, we find SVCs, such as *in Kontakt treten* (engl. *to get in touch*), in which the meaning of the whole SVC may be reconstructed by the meaning of their components (the SV and the PredN), even though in these cases the meaning of the SVs are semantically “reduced” or “light”. Additionally, the PredNs are subject to morphosyntactic constraints when they form part of an SVC, while the the same PredNs are not constrained when used together with other verbs. Thus, it is not surprising that one of the crucial issues in the literature on German SVC has been the search for features and tests to delimit SVC from prototypical idioms (of the type *kick the bucket*) on the one hand, and from “compositional” constructions (e.g. *take the soup*) on the other hand. German grammars discuss a set of classification and delimitation features, most of which are based on the hunch that predicative nouns in SVCs do not refer to discourse referents, but that they rather bear the semantic load in complex predicates formed together with the “semantically light” SVs. These features and tests are demonstrated with prototypical examples, but have not yet been evaluated using large text corpora. In my paper, I first want to report on several corpus-based case studies, in which some of these features (determiner and number variation in Npreds, anaphoric reference to Npreds in the subsequent context, modification by adjectives and relative clauses etc.) were evaluated on the basis of the German DWDS corpus (cf. [www.dwds.de](http://www.dwds.de)). The results of these studies will show that it is worthwhile to investigate the properties of SVC on a broad empirical basis and to consider the textual context of these constructions. In the second part of my paper, I will propose a typology of SVC that reflects the findings of these case-studies and discuss some methodological problems and issues that occur in connection with the corpus-based investigation of SVC.

---

Natalie Schluter  
Dublin City University, Université de Montréal  
**On the Applicability Conditions of a Support Verb Paraphrase  
Operation**  
Do 10.00-10.30

---

The modeling of meaning by the establishment of paraphrase rules such that one is able to predict acceptable results following the application of a paraphrase rule on any arbitrary sentence is essential for the optimisation of the utility of the paraphrase in Semantics and Computational Linguistics. In order to do this, one would have to first determine the undesired effects of paraphrase operations on arbitrary sentences. We propose to contribute towards this modelling from a new angle. We carried out the blind, manual application of a well-recognised paraphrase rule to a defined corpus of 832 sentences in English. In particular, we carried out a study of the problematic surrounding the paraphrastic operation that transforms a verb of a verb phrase into a support verb followed by a nominalisation of the original verb (and other original elements of the phrase).

The paraphrase operation we study involves the articulation or ‘breaking up’ of a main verb into a support verb and nominalisation of that main verb. The nominalisation becomes the second syntactic actant of the support verb, and the grammatical subject of the original sentence remains the grammatical subject of the support verb in the paraphrase

In our investigation, we do not examine the precise form that the paraphrase takes; we only verify if a result of the paraphrase operation exists, according to certain conditions previously laid out. Only where the paraphrase is unacceptable will we be trying to detect reasons for this unacceptability based on form and meaning. The sentences of our corpus were paraphrasable by Rule 18, with a rate of 38.8%. Of these attempted paraphrases, an unacceptable paraphrase was encountered with a rate of 29.4%.

The analysis of the resulting sentences of this paraphrase operation, the attempted paraphrases, was carried out with the hypothesis that linguistic aspect and nominal types play a significant role in the explanation of the unacceptability of some unacceptable paraphrases. In addition, we have found that the unacceptability of other sentences may be explained by the presence of other sentential syntactic or semantic properties. A total of eleven categories of these properties have been detected and distinguished throughout this analysis of unacceptable paraphrases.

---

Kow Kuroda, Keiko Nakamoto, Hajime Nozawa, and Hitoshi Isahara  
NICT Japan, Kyoto Universität, NICT Japan, NICT Japan  
**Situations are “Attractors” of Semantic Interpretation**  
Do 10.30-11.00

---

A “frame-based” semantic theory like Fillmore’s Frame Semantics predicts that successful interpretation of  $s = w_1 w_2 \dots w_n$  ( $W(s) = \{w_1, w_2, \dots, w_n\}$ ) does not require that the lexical semantics of every word in  $W(s)$  be fully specified, as far as enough information is supplied by its co-occurents. This predication can be (dis)confirmed if we examine how  $s$  is interpreted when  $s$  contains a nonce word  $w^*$ :  $s$  is successfully interpreted if the rest words of  $w^*$  in  $s$ , denoted by  $W(s) - w^*$ , evokes a reasonable “frame”  $F$  (or more specifically, an idealized “situation” denoted by  $F$ ), and if  $w^*$  accepts the “semantic role” that  $F$  assigns to it. For example in English, sentences like “**The gang of two** attacked  $y$ ” evoke a situation of <Bank Robbery> even when  $y$  is a nonce word. But sentences with more generic subjects, e.g., “**They** attacked  $y$ ”, don’t show this attraction-to-frame effect strongly enough. In cases where strong attraction-to-frame effects are observed, we can say that frame  $F$  “attracts” a reading of  $s$  and suppresses other possible readings. We call this the **attraction-to-frame/situation hypothesis**. In our study, we tested and positively confirmed this hypothesis by a psychological experiment, selecting a verb *osou* (attack, hit).

Procedure: (Step 1) two linguists of us did a corpus-based analysis modelled on BFN practices (Fontenelle, ed. 2003).<sup>1</sup> For this, we collected 400+ examples of sentences in which *osou* was used from a bilingual corpus of 150,000 pairs of Japanese and English sentences (Utiyama and Isahra 2003). Roughly 15 situations were identified that cover the full range of interpretations for the *osou*-sentences in the corpus.<sup>2</sup> (Step 2) Two psychological experiments were done later to evaluate the analysis with positive results. The first showed that linguist’s identification had a good fit to a layperson’s semantic intuition, as reported in Nakamoto et al. (2005). The second confirmed the attraction-to-situation effect to be reported in this work.

Fontenell, T. ed. 2003. *FrameNet and Frame Semantics: International Journal of Lexicography*, 16 (5).

Nakamoto, K., K. Kuroda, and H. Nozawa 2005. „Proposing the feature rating task as a(nother) powerful method to explore sentence meanings.“ *Japanese Journal of Cognitive Psychology*, 65-81.

Utiyama, M. and H. Isahara 2003. „Reliable measures for aligning Japanese-English articles and sentences.“ In: *Proceedings of the ACL*, 72-79.

---

<sup>1</sup> We did a much finer-grained analysis than BFN on purpose, being aware of optimal granularities.

<sup>2</sup> It is unclear whether it is appropriate to say that 15 frames cover the range of the senses of the target *verb*. Note that we classified the meanings of the sentences in which *osou* is used. This doesn’t necessary give the senses of a target *verb*.

---

Ekaterini Stathi  
BBAW, Freie Universität Berlin  
**Idioms as Lexical Units: a Corpus-based Approach**  
Do 11.30-12.30

---

Idioms are usually defined as multi-word units with non-compositional meaning, characterized by the stability of their constituents. Based on a large German corpus (about 980 million words) we examine VP idioms, i.e. idioms that consist of a verb and at least one noun (NP, PP), such as *ins Fettnäpfchen treten* ‘drop a brick’, *etw. unter den Teppich kehren* ‘sweep sth. under the rug’, *jmdn. auf den Arm nehmen* ‘pull so.’s leg’, *ins Gras beißen* ‘bite the dust’, *jmdm. die Leviten lesen* ‘read the riot act to so.’). We examine the following questions:

- (1) What can be inferred from the textual, semantic and syntactic, behaviour of idioms regarding their nature and their status as lexical items?
- (2) How should an entry in the lexicon look like in order to account for this behaviour?

As to (1), the answer presupposes that the lexical properties of the idioms can be determined from a sufficiently large amount of data. Traditional analyses have based their investigations primarily on the sentence level. We believe that such an approach cannot fully explain the grammar of idioms. Corpora, on the other hand, provide larger contexts which show the behaviour of idioms in discourse and provide new insights about their lexical and grammatical properties.

Linguistic, psycholinguistic, and computational approaches have largely ignored idiom use in context. Therefore descriptions of idioms are problematic in many ways. Starting from the data and the use of idioms, we propose a theory of the idiom lexicon. Moreover, existing theories (idioms as “dead metaphors”, idioms as “long words” etc.) are reviewed and correlated with the corpus findings.

Finally, we address question (2) about the lexicographic treatment of idioms. A traditional dictionary cannot deal with our corpus-based results. Can discourse and textual phenomena be integrated into a lexicon and into a theory of the idiom lexicon? And if this is the case, how should an entry look like?

- Moon, Rosamund. 1998. *Fixed Expressions and Idioms in English. A Corpus-Based Approach*. Oxford: Clarendon.
- Nunberg, Geoffrey, Ivan A. Sag, & Thomas Wasow. 1994. “Idioms.” *Language* 70, 491-538.

---

Christiane Fellbaum  
Princeton University, BBAW  
**German VP Idioms: Tough Nuts to Crack**  
Fr 10.30-11.30

---

Idioms are commonly assigned two complementary properties that distinguish them from ad-hoc, freely composed phrases: semantic non-compositionality and syntactic frozenness. Nunberg et al. (1994) and Dobrovolskij (1999) are among those who argue for a correlation between these two properties. If this correlation were robust, syntactically fixed idioms could be treated as lexical units--long words--while syntactically free idioms should be neatly decomposable into metaphors--single lexemes--and components from the free language. Metaphors (like *nut* in *tough nut to crack*) would require standard lexicographic treatment, no different from entries for other words. Frozen idioms and metaphors would be stored and retrieved as units from speakers' mental lexicons and be available for morphological and syntactic operations via the standard mechanisms.

Corpus data contradict this rather oversimplified picture. First, semantically non-compositional idioms are syntactically flexible: *...der kleine Mann findet um sich nur Stränge, und er haut ueber sie*. Second, opaque idiom constituents can undergo modification and lexical substitution: *...den sie mehrmals an dem Abend im Kerzenschein auf den musikalisch-erotischen Arm nahm. hier tanzt/stept/tobt der Bär*.

Their complex behavior indicates that the place of idioms in the grammar extends beyond the lexicon. Idioms defy the traditional notion of "lexical unit" and pose great challenges for the creation of a lexicographic resource, starting with the quest for a citation form.

This talk describes the design and contents of the corpus-based and corpus-linked idiom database constructed at the Berlin-Brandenburg Academy of Sciences. We discuss how a wide range of phonological, morphological, syntactic, and lexical phenomena must be accommodated to fully account for the grammar of idioms. While the database is painstakingly being constructed one idiom at a time, this process reveals that a characteristic property of idioms is their place at the intersection of lexicon, syntax, and discourse.

Dobrovolskij, D. 1999. Haben transformationelle Defekte der Idiomstruktur semantische Ursachen? In Fernandez-Bravo, N., Behr, I., and Rosier, C. (Eds.) *Phraseme und typisierte Rede*. Tübingen: Stauffenburg, 25-37.  
Nunberg, J., Wasow, T., and Sag, I. 1994. *Idioms*. *Language* 71. 3:491-538.

Work supported by the Wolfgang Paul Preis of the Alexander von Humboldt Foundation/ZIP Program.



---

Gergely Pethő, Péter Csátár  
University of Edinburgh, Universitát Debrecen  
**Recognition and processing of figurative language**  
Fr 11.30-12.00

---

We believe that the investigation of the recognition and processing of figurative language in humans may lead to conclusions that could be relevant for modelling such processes by computational means. We therefore present an overview of well-documented psychological and psycholinguistic results concerning these issues. We assume that first of all two kinds of linguistic phenomena have to be distinguished: 1. Collocations, idioms and metaphorically/ metonymically motivated interpretations of simple or complex expressions which are stored in the lexicon (which is supported both by theoretical considerations and experimental results). In such cases, the issue from a computational perspective is not one of processing non-compositional language, but rather a relatively simple problem of lexical disambiguation. 2. Non-compositional phenomena in a narrow sense, in the case of which information in the lexicon does not directly lead to a contextually appropriate compositional interpretation of the expression. These include non-conventionalised cases of indirect speech acts, irony, metonymy (e.g. Nunberg's well-known examples) and especially creative uses of metaphor.

Different theses have been suggested in the psycholinguistics literature as to what triggers a non-literal interpretation in these cases and how it is derived (pragmatic and cognitive approaches), but all of these have conceptual or empirical difficulties, as will be discussed.

The crucial question, namely, what should be identified and tagged as a metaphor (or other kind of figurative language) by human annotators in a corpus is similarly unresolved. Intuitions of speakers whether an expression is a metaphor strongly diverge, and existing concepts of metaphor according to different theories are obviously not identical either extensionally or intensionally, although not explicit enough to be operationalised.

It has to be therefore decided 1) which concept of metaphor can serve as a useful basis for the processing of corpora and 2) how the identification of metaphors according to this chosen concept can be operationalised. We suggest that although the conceptual theory of metaphor of Lakoff and Johnson is relatively compatible with the usual ontological tools of current computational linguistics, its concept of metaphor is not a useful one in the long run, since this theory denies the distinction between the two categories of expression mentioned above, and therefore conflates issues of disambiguation of lexical ambiguity with ones of interpretation of non-compositional figurative language.

## Challenges in lexical semantics: Non-compositionality in frame-semantic annotation

Fr 12.00-12.30

---

The SALSA project at Saarland University [1] is annotating the German TIGER corpus with role-semantic information in the FrameNet [2] paradigm. Frame-semantic annotation aims to represent the **understood conceptual meaning**, which can be used to access the core meaning of sentences in information access tasks; at the same time, it allows to extract regularities in the syntax-semantics mapping. This works well for “compositional” cases, where the meaning is introduced lexically by a single frame evoking predicate, and the roles are filled by syntactic constituents, as in [Peter] *kauft* [einen Wagen]. However, in corpus annotation, we frequently find less compositional cases, of different types:

- **Multiword expressions:** [Ich] *bekomme* [dich] *zu Gesicht*. ≈ [Ich] *sehe* [dich].  
(Frame introduced by multi-word predicate; roles in canonical syntactic configuration)
- **Support constructions:** [Peter] *unterzog sich* [einer OP] [durch den Chefarzt].  
(Core frame and roles introduced by embedded noun; roles added by support verb)
- **Metaphors**, with literal vs. understood frame semantic meaning, e.g.:  
[Maria] *lehnte* [sich] [aus dem Fenster] (literal meaning).

What these cases have in common is that annotation of the “literal” frame semantic meaning, according to the usual syntax-semantics mapping, is *semantically inappropriate*; at the same time, these cases require *distinct* types of annotation, to reflect both their special meaning and syntax-semantics mapping properties. We have developed *special annotation schemes for the different types*, together with a set of *operational criteria* to distinguish them. In practice, however, it turns out that especially highly frequent verbs often participate in constructions which show aspects of more than one analysis, e.g. as support verb, metaphor, or simply a different reading of the base verb, as in:

- [Sie] *nahmen* [Passagiere] [als Geiseln]. vs. [Sie] *nahmen* [Passagiere] *als Geiseln*.

We present examples for the interaction of different phenomena, and discuss how our annotation scheme allows us both to identify doubtful cases and to represent the maximum amount of information, using underspecification and other means. We also show a preliminary evaluation of our first corpus release with respect to these phenomena.

[1] K. Erk, A. Kowalski, S. Pado and M. Pinkal. 2003. Towards a Resource for Lexical Semantics: A Large German Corpus with Extensive Semantic Annotation. Proceedings of ACL 2003, Sapporo, Japan.

[2] C. Fillmore, C. Johnson and M. Petruck. 2003. Background to FrameNet. Intl. Journal of Lexicography 16.

## AG 12 Adaptive Ontologien auf syntaktischen Strukturen

Koordination: Uwe Mönnich, Kai-Uwe Kühnberger

### Programm

Donnerstag 23.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
09.00-09.30	Uwe Mönnich, Kai-Uwe Kühnberger	<i>Einführung</i>
09.30-10.00	Timm Lichte, Jan-Philipp Soehn	<i>Negation als Auftretenskotext – Zur Skopuserkennung mittels syntaktischer Annotation</i>
10.00-10.30	Mihaela Vela, Thierry Declerck	<i>Automatische Extraktion von semantischen Relationen aus linguistisch annotierten Texten</i>
10.30-11.00	Helmar Gust, Petra Ludewig	<i>Aspekte der Extraktion von Ontologien aus Korpora</i>
11.00-11.30	Kaffeepause	
11.30-12.00	Steffen Staab	<i>Lernen von Ontologien</i>
12.00-12.30	Harald Huber	<i>Adaptive Ontologien in der Praxis</i>

Freitag 24.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
10.30-11.00	Alexander Mehler, Matthias Dehmer, Rüdiger Gleim	<i>Automatische Klassifikation von Hypertextgraphen</i>
11.00-11.30	Konstantin Todorov	<i>Aspects of Learning Ontologies with Support Vector Machines</i>
11.30-12.00	Peter Geibel	<i>Text Classification and Clustering using Kernels for Structured Data</i>
12.00-12.30	Sophia Katrenko	<i>Tree-based techniques for learning ontologies</i>
12.30-13.00	Willem Robert van Hage	<i>A Method to Combine Linguistic Ontology-Mapping Techniques</i>
13.00-13.30		<i>Diskussion</i>

---

Timm Lichte, Jan-Philipp Soehn  
Universität Tübingen  
**Negation als Auftretenskontext –**  
**Zur Skopuserkennung mittels syntaktischer Annotation**  
Do 09.30-10.00

---

Ausgehend von Untersuchungen zum Negationsskopos (Ladusaw, 1980 und Linebarger, 1987) soll der Frage nachgegangen werden, wie mithilfe der syntaktischen Annotation eines partiell geparsten Korpus die Skopoi von Negationsoperatoren erkannt werden können. Unser Ziel ist es, Distributionsmuster der Korpuselemente (Lemmata und Lemmaketten) relativ zu Negationsoperatoren herauszufiltern, diese statistisch auszuwerten und mit dieser Information Negative Polaritätselemente (NPIs) zu extrahieren. NPIs sind lexikalische Einheiten, denen Distributionsrestriktionen hinsichtlich semantischer Eigenschaften ihres Auftretenskontexts nachgesagt werden. In der Literatur werden als Lizenzierer u.a. negative bzw. abwärtsimplikative Kontexte (siehe Zwarts, 1997 und van der Wouden, 1997) und non-veridikale Kontexte (Giannakidou, 1997) diskutiert. Anders als für das Englische oder Niederländische gibt es für das Deutsche noch keine umfassende Sammlung von NPIs. Wir erhoffen uns zum einen, die Extraktionsmethode zur Ermittlung vielversprechender NPI-Kandidaten einzusetzen, um eine solche Sammlung zu erstellen; zum anderen eröffnet die Extraktionsmethode die Möglichkeit, das in der Literatur vornehmlich introspektive Vorgehen empirisch gegenzuprüfen.

Um die Zuverlässigkeit der Methode zu gewährleisten, scheint es unerlässlich, den Geltungsbereich eines NPI-Lenzierers (bzw. Negationsoperators) möglichst genau zu bestimmen. Wir werden feststellen, dass allenfalls eine Annäherung erreicht werden kann, dass diese Annäherung jedoch ausreicht, da durch die statistische Evaluierung eine detailliertere Auflösung keine signifikant besseren Ergebnisse zeigt.

- Giannakidou, A. 1997. *The Landscape of Polarity Items*. Dissertation Rijksuniversiteit Groningen.
- Linebarger, M., 1987. *Negative Polarity and Grammatical Representation*. *Linguistics and Philosophy* 10, S. 325–387.
- Ladusaw, W., 1980. *Polarity Sensitivity as Inherent Scope Relations*. Garland Press, New York.
- van der Wouden, T. 1997. *Negative Contexts. Collocation, Polarity and Multiple Negation*, Routledge, London.
- Zwarts, F., 1997. *Three Types of Polarity*. In: Hamm, F., Hinrichs, E. W. (Hrsg.), *Plurality and Quantification*. Kluwer Academic Publishers, Dordrecht, S. 177-237.

---

Mihaela Vela, Thierry Declerck  
Universität des Saarlandes, DFKI Saarbrücken  
**Automatische Extraktion von semantischen Relationen aus linguistisch  
annotierten Texten**  
Do: 10.00 – 10.30

---

Es werden Ergebnisse einer Diplomarbeit vorgestellt, die im Rahmen des europäischen Forschungsprojekts „Esperanto“ geschrieben wurde. Diese Arbeit stellt eine Methode und eine Implementierung für die automatische Extraktion von semantischen Relationen aus linguistisch annotierten medizinischen Texten dar. Die linguistische Analyse besteht in Abhängigkeitsstrukturen, die sowohl innerhalb von Phrasen als auch zwischen größeren linguistischen Fragmenten von einem Parser erkannt werden. Damit meinen wir, zum Beispiel, die „head-modifier“-Abhängigkeit oder Relation, die zwischen einem Adjektiv und dem Kopf einer Nominalphrase existiert, aber auch die grammatischen Relationen, die in einem Satz vorhanden sind, wie zum Beispiel die Subjektrelation zwischen einem Prädikat und einer bestimmten Nominalphrase.

Vier verschiedenen Arten von linguistischen Konstruktionen wurden berücksichtigt, um unsere Methode für die automatische Extraktion von semantischen Relationen aus größeren Textmengen zu definieren und zu testen: Pre- und Postmodifikation von nominalen Köpfen, phrasale Koordinationen und grammatische Relationen. Zusätzlich zu den strukturellen Eigenschaften werden auch lexikalische semantische Ressourcen herangezogen, wie zum Beispiel die semantische Klassifikationen von Adjektiven<sup>1</sup>, Adverbien<sup>2</sup>, Präpositionen und Verben, die die extrahierten semantischen Relationen bis zu einem gewissen Grad bestimmen können (z.B. als lokale, temporale, kausale Relationen etc.). Die extrahierten semantischen Relationen wurden, da wo es möglich war, nach UMLS<sup>3</sup> Relationen benannt. Dies erlaubt es, die automatisch extrahierten semantischen Relationen mit einem bestehenden und sehr verbreiteten Thesaurus zu vergleichen, und so unser System zu evaluieren. Wir werden auch die Ergebnisse einer ersten Evaluierung vorstellen. Als Schluss zu diesem Abstrakt sei noch darauf hingewiesen, dass unser System in Zukunft auch für die Zwecke der „literature-based scientific discovery in the bio-medical domain“ eingesetzt wird.

---

<sup>1</sup>Klassifikation nach Lee, S.1994. *Untersuchungen zur Valenz des Adjektivs in der deutschen Gegenwartssprache*. Lang.

<sup>2</sup>Klassifikation nach Lobeck, A. 2000. *Discovering Grammar: An Introduction to English Sentence Structure*. Oxford University Press.

<sup>3</sup>Unified Medical Language System.

---

Helmar Gust, Petra Ludewig  
Universität Osnabrück  
**Dynamische Ontologien im Kontext von dokumentbasiertem  
Wissensmanagement**  
Do 10.30-11.00

---

Ausgangspunkt der Betrachtungen ist die Notwendigkeit für größere internationale Unternehmen, ihr Wissensmanagement hinsichtlich elektronischer Dokumente zu verbessern. Dabei sind folgende Anforderungen zu beachten:

- Eine unternehmensweit verbindliche Ontologie muss mit einer global verbindlichen Terminologie verbunden sein, sollte aber möglichst auch Bezüge zu lokalen Varianten herstellen (s. a. Wagner & Kunze 1999).
- Die Ontologie darf nicht nur die Bezüge zwischen aktuellen Dokumenten, sondern muss auch die Einbindung zeitlich zurückliegender sowie künftiger Dokumente unterstützen.
- Die Ontologie sollte die Zusammenarbeit mit unternehmensexternen Partnern unterstützen, indem sie auch auf unternehmensexterne Quellen anwendbar ist.

Um die Akzeptanz der anvisierten Ontologie zu erhöhen und die Entwicklungskosten zu reduzieren, bietet es sich an, sie mit möglichst viel maschineller Unterstützung zu generieren, zu evaluieren (Staab & Studer 2003) und zu warten.

Im Studienprojekt ASADO (Bleichner et al. 2005) wurde von studentischen Projektmitgliedern unter Verwendung einschlägiger korpuslinguistischer Verfahren (Stemle in Bleichner et al. 2005) und einer erweiterten Konstituentenanalyse von Nominalphrasen (Möller in Bleichner et al. 2005) eine Terminologie extrahiert, die sich gut als Ausgangspunkt für eine Ontologie eignet. Bei einem Großteil der Fachtermini handelt es sich um zum Teil geschachtelte Komposita. Im Engineering spielt Wortbildung offensichtlich eine zentrale Rolle, wenn es um das Referieren auf relevante Subkonzepte geht. Auf diesem Wege entstehen vor allem viele kleinere hierarchische Strukturen, d.h. ein Wald aus Bäumen. Diese Strukturen können teilweise durch gemeinsame Oberkonzepte zusammengefügt werden und lassen sich darüber hinaus auch durch geeignete Relationen horizontal miteinander verbinden. Derartige Ontologien dürften sich sinnvoll für das Labeling von Clustern in Dokumentarten nutzen lassen (Bleichner und Stefaner in Bleichner et al. 2005). Gegenstand des Beitrags sind erste Lösungsvorschläge für einen datenbasierten inkrementellen Aufbau einer solchen Ontologie.

Bleichner, Martin G. et al. 2005. *ASADO - The Analysis and Structuring of Aviation Documents – Final report*. Institute of Cognitive Science, University Osnabrück.

Staab, S., Studer, R. 2003. *Handbook on Ontologies*. Springer-Verlag.

---

Steffen Staab  
Universität Koblenz  
**Lernen von Ontologien**  
Do 11.30-12.00

---

Wir präsentieren zwei neuartige Ansätze zur automatischen Akquisition von Begriffshierarchien aus Textkorpora. Der eine Ansatz basiert auf Formaler Begriffsanalyse (FCA), einer Methode zur Untersuchung von Verbänden, die primär zur Datenanalyse verwendet wird. Er bildet Begriffshierarchien auf der Basis von Kontextinformationen von Termen. Verwendung findet hierbei Harris' Hypothese, dass sich aus statistischen Kontextinformationen eines Terms semantische Kategorien aufbauen lassen. Basierend auf den Kontextinformationen produziert FCA einen Verband, der in eine korrespondierende Begriffshierarchie überführt wird. Zur Evaluierung dieses Ansatzes haben wir FCA mit anderen Clusteringalgorithmen auf der Basis zweier Goldstandard-Ontologien verglichen.

Der zweite Ansatz basierte ebenfalls auf Harris' Hypothese. Er verwendet als Mittel der statistischen Analyse allerdings Suchergebnisse, wie sie z.B. von Google produziert werden. Als Kondensationspunkt für die Suche dienen zum einen Begriffe, die in einem Text auftauchen und noch nicht in eine bereits gegebene Begriffshierarchie einsortiert sind (z.B. „cabriolet“), zum anderen Schlüsselaustrücke wie „such as“. Die kombinierte Suche nach diesen Begriffen und Schlüsselaustrücken führt zu Fragmenten in der Ergebnisanzeige wie „cars such as cabriolets“, die mittels einer gemischt statistisch-semantischen Analyse dem Aufbau einer Begriffshierarchie dienen können.

- Bloehdorn, S., P. Cimiano, A. Hotho, S. Staab. "An Ontology-based Framework for Text Mining". *Zeitschrift für Computerlinguistik und Sprachtechnologie* (GLDV-Journal for Computational Linguistics and Language Technology), 20(1), 2005: 87-112.
- Cimiano, P., S. Staab. „Learning by Googling“. *SIGKDD Explorations*. 6(2), 2004: pp. 24-33.
- Cimiano, P., A. Hotho, S. Staab, 2005. „Learning Concept Hierarchies from Text Corpora using Formal Concept Analysis.“ In: *Journal of Artificial Intelligence Research*. 24(2005), pp. 305-339.
- Cimiano, P., G. Ladwig, S. Staab. "Gimme' the context. Context driven automatic semantic annotation with C-Pankow". In: *Proceedings of the 14<sup>th</sup> International World Wide Web Conference, WWW 2005*, Chiba, Japan, May, 2005. ACM Press.

---

Harald Huber  
USU Software AG, Möglingen  
**Adaptive Ontologien in der Praxis**  
Do 12.00-12.30

---

"Ontologie" sind sicherlich einer der häufig gebrauchten Begriffe der letzten 24 Monate. Das Semantic Web, die OWL, TopicMaps, all das sind Themen, die in der letzten Zeit doch stark strapaziert wurden. Die zugrunde liegenden Techniken sind zwar weitaus älter, doch wurde die Aufmerksamkeit der Öffentlichkeit erst durch den Wissensmanagement-Trend der letzten Jahre darauf gelenkt. Eine Vielzahl von Softwareanbietern wie beispielsweise Moresophie, Intelligent Views, USU, Ontopia, Empolis und weitere haben die Idee der semantischen Darstellungen in ihr Produktportfolio aufgenommen. Allerdings leiden alle diese Produkte darunter, dass die Erstellung der Ontologie einen nicht unerheblichen Aufwand verursacht, ohne dass der Ontologie als solche direkt und einfach ein Mehrwert zugeordnet werden kann. Daher sind die meisten Unternehmen nicht bereit, den Aufwand für die Ontologie-Erstellung und -Wartung zu betreiben. Auch im Falle des Semantic Webs fehlt bis heute das Geschäftsmodell, das die Ontologie-Erstellung rentabel macht. Entsprechend gering ist daher bis heute die Verbreitung dieser Softwarewerkzeuge. Die Automatisierung dieses Prozesses ist somit außerordentlich attraktiv und wird mit viel Aufwand analysiert. Die USU AG ist diesen Weg mit dem USU KnowledgeMiner 4 gegangen. Die Hauptproblematik dabei ist, dass ein System keine zweifelsfrei korrekte Semantik erzeugen kann, zumindest ist dies die Überzeugung der USU AG. Der entscheidende Ansatz der USU AG ist also, eine "System-interne" Darstellung zu finden, die automatisch erstellbar ist und die andererseits vom Anwender bzw. Redakteur als Semantik interpretiert werden kann und die damit redaktionell betreubar ist. Die von der USU entwickelte Technologie wurde nun in rund 30 Fällen seit  
eineinhalb Jahren produktiv eingesetzt. Der Vortrag stellt im ersten Teil die Technologie vor. Im zweiten Teil werden dann die Erfahrungen beschrieben, die in den verschiedenen Kundenprojekten aufgetreten sind.



---

Alexander Mehler, Matthias Dehmer, Rüdiger Gleim  
Universität Bielefeld, Universität Darmstadt, Universität Bielefeld  
**Automatische Klassifikation von Hypertextgraphen**  
Fr 10.30-11.00

---

Mit dem Aufkommen der Neuen Medien rücken zunehmend Dokumentstrukturen in das Blickfeld des automatischen Lernens, die bislang weitgehend unberücksichtigt geblieben sind. Dabei handelt es sich um Graphstrukturen, welche die Nichtlinearität von Webdokumenten modellieren, ob nun im Bereich des Content, Structure oder Usage Mining (Chakrabarti 2002). Es geht genauer um attribuierte, hierarchisierte Wurzelgraphen, deren Hyperkanten Hyperlinks repräsentieren und von einer kernbildenden Hierarchie überlagert werden, welche zum Beispiel eine autoreseitig bevorzugte Inhaltsgliederung kodiert. Damit entsteht eine Klassifikationsaufgabe, welche die Integration von Graphstrukturmodellen und numerischen Verfahren der Klassifikation verlangt (Bunke et al. 2001). In diesem Zusammenhang scheiden jedoch bisherige Ansätze zur Graphähnlichkeitsmessung mit dem Ziel einer anschließenden Klassifikation aus. Dies betrifft genauer Ansätze, die auf Graphisomorphien oder Graphgrammatiken beruhen. Beide Ansätze weisen komplexitätstheoretische Defizite auf und scheiden folglich für die Zwecke einer automatischen Klassifikation großer Mengen von Graphen aus. Genau diese Aufgabe aber stellt sich im Zusammenhang der korpusbasierten Analyse webbasierter Strukturen. In diesem Beitrag wird ein Graphmodell zur Repräsentation der Dokumentstrukturen von Webdokumenten vorgestellt, das sich an einem Webgenre-Modell orientiert und folglich eine hohe strukturelle Komplexität webbasierter Dokumentstrukturen berücksichtigt. Auf dieser Grundlage wird ein Klassifikationsalgorithmus beschrieben und empirisch evaluiert, der Instanzen dieses Graphmodells automatisch klassifiziert und damit als ein Vorschlag für das automatische, unüberwachte Lernen von Webstrukturen auf der Basis großer Korpora von Webdokumenten angesehen werden kann. Der Beitrag knüpft somit – insgesamt gesehen – an jene Aufgaben an, wie sie sich mit dem Aufkommen webbasierter Dokumente der computerlinguistischen Strukturmodellierung stellen und wie sie mit herkömmlichen Methoden der Baumähnlichkeitsmessung nicht adäquat behandelt werden können.

Chakrabarti, S. (2002): *Mining the Web: Discovering Knowledge from Hypertext Data*. San Francisco: Morgan Kaufmann.

Bunke, H., S. Günter, and X. Jiang (2001): Towards bridging the gap between statistical and structural pattern recognition: Two new concepts in graph matching. In: *Proceedings of the Second International Conference on Advances in Pattern Recognition*, Berlin: Springer, 1-11.

---

Konstantin Todorov  
Universität Osnabrück  
**Aspects of Learning Ontologies With Support Vector Machines**  
Fr 11.00-11.30

---

During the last years, the need of exploring all capacities of the World Wide Web and making queries more efficient led to the creation of what is now known as the *Semantic Web*. The information has been structured in ontologies, defined in [3] as the explicit specification of a set of objects, concepts and other entities that are assumed to exist in some area of knowledge together with the relationships that hold among them. The ontology creation has been carried out manually or semi-automatically under the supervision of a knowledge engineer. Due to that and the growing usage of ontologies the problem of overlapping knowledge is encountered frequently. Application of machine learning techniques comes as a natural issue in classification of the elements of different ontologies and automatic creation of between-ontologies mappings thus structuring the Semantic Web and making queries more efficient.

The present talk aims at identifying the benefits of the application of *Support Vector Machines (SVMs)* (and other machine learning techniques) to help structuring the semantic web and establishing semantic mappings between ontologies. After a short overview of the method as introduced in [1], I will describe several levels on which the latter problem could be solved. Since we would like to be able to measure the common level of knowledge covered by two given input ontologies, I will discuss text categorization techniques mapping different terminologies used in the ontologies, classification of concepts within ontologies and definition of a similarity measure between ontologies on a concept level. Here I refer to [1], [2] and [4].

Since data on the web consist mostly of natural language documents, I will argue on the importance of finding an adequate representation of data in order to be usable from the *SVMs*, which take their inputs only from some subset of  $\mathbf{R}^n$ . Finally, I will briefly outline several reasons why the choice of *SVMs* seems to be appropriate for solving the up-listed problems.

[1] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines*, Cambridge University Press, 2000

[2] A. Doan, J. Madhavan, P. Domingos, and A. Halvey. Learning to map between ontologies on the semantic web. In *Proceedings of the 11th International Conference on World Wide Web*, pages 662--673, 2002.

[3] T. Gruber. Towards Principles for the design of Ontologies used for Knowledge Sharing. In *International Journal Human-Computer Studies*, 43, p. 907-928. 1993

[4] A. Maedche and S. Staab. Ontology learning for the semantic web. *IEEE intelligent Systems*, 16 (2), 2001.

---

Peter Geibel  
Universität Osnabrück  
**Text Classification and Clustering using Kernels for Structured Data**  
Fr 11.30-12.00

---

Support Vector Machines (SVMs) are known to be an excellent choice for classifying text data based on a TFIDF representation. SVMs are based on a linear approach that is combined with a mapping into a usually high-dimensional feature space. This mapping is done using so-called *kernels*, which correspond to the scalar product of image vectors in this feature space. Kernels can also be regarded as similarity measures having some additional properties.

Recent machine learning research in the area of kernel methods emphasizes the possibility of defining kernels for *structured data*, i.e. for sets, strings, trees, graphs, logical expressions, and the like. In the context of text processing, such complex structures are used for describing single sentences (e.g., parse trees) or whole paragraphs. Trees are also used for describing the document structure (given by title, abstract, chapters and sections, references etc. or by HTML tags), whereas graphs have to be used for specifying the link structure for hypertext documents.

It is possible to use structure kernels for a variety of tasks that are not restricted to classification. A valid kernel defines an implicit mapping of the structured data into a vector space. It is therefore possible to apply the whole range of kernel methods: support vector regression and Gaussian processes allow predicting real values for e.g. tree-structured input. Kernel-PCA can be used for constructing good features that are based on the original TFIDF features as well as on the document structure and the link structure in case of hypertext documents. Note that kernel-PCA for non-structured data already allows for detecting non-linear dependencies in the dataset. This information can be combined with the structural properties, e.g. document structure and link information. The clustering of documents can be based on such kernels, although it also allows more general structural similarity measures.

Based on own research in this field, the talk will give an overview on kernel methods in general, and specifically on kernels for structured data. I will suggest potential applications of kernel methods in the context of text processing that go beyond the application of SVMs with e.g., linear, polynomial or RBF kernels.

---

Sophia Katrenko  
Universiteit van Amsterdam  
**"Tree-based techniques for learning ontologies"**  
Fr: 12.00-12.30

---

By definition, an ontology is often seen as a hierarchical structure of concepts and relations between them (Maedche, 2002). There have been several approaches of how to learn concepts and relations from the text corpora. Most of them either deal with statistical properties or use linguistic information. Furthermore, the linguistic information has been explored in several ways. The approach that is used most often relies on so-called Hearst patterns which allow for hyperonymy discovery by means of hand-written patterns.

In this talk, I will present a more general approach to produce such patterns which makes use of dependency trees. As it has been pointed out by many researchers in the field of computational linguistics, syntax and semantics are not independent (Rastier, 2002). Thus, it is of certain interest to use more complex syntactic structures and evaluate in how much they contribute to the semantic knowledge extraction. While considering syntactic structure, several possible approaches can be used which, in turn, are dependent on a particular machine learning technique to be applied. In case of the attribute-value representation, each tree can be represented by specific tree levels chosen by an expert. However, doing so, one is limited to the certain number of features which are selected in advance. Moreover, there is no guaranty that selected features are the most informative. In as much as there is no machine learning method performing best ("no free lunch" theorem), there is no feature selection method which would yield the best representation.

Nevertheless, by considering dependency trees, it is possible to find patterns for concept and relation learning which are trees themselves (Dietterich, Th. 2002). I will show that it is possible to use the same generic approach for both, concept and relation learning.

I will discuss possible representations and approaches starting with the baseline which is sliding window to more sophisticated using the information from the dependency structures in two ways, as a set of pre-selected features and as frequent subtrees. The approaches I will discuss throughout my talk will be accompanied by results from an experiment on biomedical data.

Dietterich, Th. 2002. *Machine Learning for Sequential Data: A Review*. In Proceedings of the Joint IAPR International Workshop on Structural, Syntactic, and Statistical Pattern Recognition.

Maedche, A. 2002. *Ontology Learning for the Semantic Web*. Boston: Kluwer Academic Publishers.

Rastier, F. 2002. *Semantics for Descriptions*. From Linguistics to Computer Science. CSLI Publications.

---

Willem Robert van Hage  
Vrije Universiteit Amsterdam  
**A Method to Combine Linguistic Ontology-Mapping Techniques**  
Fr 12.30-13.00

---

Ontologies are widely used to provide access to the semantics of data. To provide integrated access to data annotated with different, yet related, ontologies, one has to relate these ontologies in some way. This is commonly done by cross-referencing concepts from these ontologies. In different contexts this practice is called ontology mapping, schema matching, or meaning negotiation.

In the literature one can find surveys of the widely varying methods of automated ontology mapping, for instance, in the surveys done by Kalfoglou and Schorlemmer and Rahm and Bernstein. Some compare concepts based on shared instances, some use relations between the concepts, and others use the terms denoting the concepts. This talk will be about automated ontology-mapping techniques in the latter category. They come from the field of information retrieval and computational linguistics. Furthermore this talk will be about ways to combine these techniques to achieve better results.

The evaluation experiments performed to measure this were done in the food domain. The mapping problem they focus on is discovering subclass relations between the AGROVOC thesaurus of the United Nations and the “US nutrient database for standard reference”.

This talk covers results of joint work with Sophia Katrenko and Guus Schreiber, presented at the ISWC 2005. The details of which can be found in the article mentioned below.

- Van Hage, W.R.; Katrenko S.; Schreiber G. 2005. *A Method to Combine Linguistic Ontology-Mapping Techniques*. In: Proceedings of the International Semantic Web Conference (ISWC) 2005
- Kalfoglou, Y.; Schorlemmer M.. *Ontology mapping: the state of the art*. The knowledge engineering review, 18(1):1-31, march 2003.
- Rahm, E.; Bernstein, P.A.. *A survey of approaches to automatic schema matching*. VLDB Journal, 10(4), 2001.

## **AG 13 Lexikalisch-semantische Ressourcen zur Sprachdokumentation und maschinellen Sprachverarbeitung**

Koordination: Claudia Kunze, Lothar Lemnitzer, Rainer Osswald

### **Programm**

Mittwoch 22.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
13.45-14.45	Claudia Kunze, Lothar Lemnitzer, Rainer Osswald	<i>Opening and introduction to the lexical semantic resources GermaNet and HaGenLex</i>
14.45-15.45	Aljoscha Burchardt, Katrin Erk, Anette Frank, Andrea Kowalski, Sebastian Pado, Manfred Pinkal	<i>Consistency and coverage: Challenges for exhaustive semantic annotation</i>
15.45-16.15	Kaffeepause	
16.15-16.45	Petra Steiner	<i>Verb classes in language use vs. FrameNet's frames</i>
16.45-17.15	Sabine Schulte im Walde, Heike Zinsmeister	<i>A statistical grammar as empirical resource for inducing lexical semantic phenomena</i>
17.15-17.45	Chris Biemann, Rainer Osswald	<i>Experiments on extracting lexical semantic features from contextual attributes</i>
17.45-18.15	Matthias Hölzner	<i>Gewichtung nominaler Argumentstellen</i>

Donnerstag 23.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
09.00-09.30	Iryna Gurevych	<i>Thinking beyond the nouns – computing semantic relatedness across parts of speech</i>
09.30-10.00	Verginica Barbu Mititelu, Eduard Barbu	<i>A theoretical framework for checking a wordnet quality</i>
10.00-10.30	Kurt Eberle	<i>The lexicon and the treatment of lexical ambiguity in translate and office dictionary</i>
10.30-11.00	Michael Beißwenger, Angelika Storrer, Sandra Wellinghoff	<i>Terminologische Wortnetze für die Informationssuche in Fachtextkorpora</i>
11.00-11.30	Kaffeepause	
11.30-12.00	Ernesto William De Luca, Andreas Nürnberger	<i>The use of lexical resources for sense folder disambiguation</i>
12.00-12.30	Jan Burse	<i>Using lexical semantic resources to extract product profiles from text catalogues</i>

Freitag 24.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
10.30-11.00	Luis Sarmiento, Ana Sofia Pinto, Luis Cabral	<i>REPINETO: A collaborative wide-scope gazetteer for entity recognition in Portuguese</i>
11.00-11.30	Gerd Unruh, Jens Baitinger, Dejan Milosavljevic, Florian Schlegel	<i>Optimising the access to large linguistic data structures</i>
11.30-12.00		<i>Discussion and closing</i>

---

Claudia Kunze, Lothar Lemnitzer, Rainer Osswald  
Universität Tübingen, Universität Tübingen, FernUniversität in Hagen  
**Opening and introduction to the lexical semantic resources GermaNet  
and HaGenLex**  
Mi 13.45-14.45

---

Large-scale lexical semantic resources are of central importance for advanced natural language processing (NLP). In our talk, we introduce GermaNet and HaGenLex, two lexical semantic resources for German, and report on the progress and perspectives of these projects.

GermaNet belongs to the wordnet family, which goes back to Princeton WordNet. Wordnets encode the most prominent lexical and conceptual relations between lexical units and concepts of a given language, like synonymy, antonymy, hyperonymy, meronymy, and the causal relation, in a net-like structure. They constitute valuable resources both for NLP applications and for theoretical investigations of lexicalization patterns. Several poly-lingual initiatives like EuroWordNet and CoreNet have been launched for establishing inter-lingual links across language-specific wordnets. Current activities within the GermaNet project include the improvement of the mono-lingual as well as the cross-lingual coverage and the acquisition of semantic roles from corpora. Other research topics focus on the terminological extension of GermaNet in the professional domain and the integration of wordnets with ontologies.

Whereas GermaNet was originally developed as a stand-alone lexical semantic resource, the initial motivation behind the HaGenLex project was to build a large computational lexicon that provides an existing semantic parser with syntactic and semantic information. The main task of the parser is to transform natural language expressions into semantic representations that are based on the so-called MultiNet knowledge representation formalism. The semantic information in HaGenLex thus conforms to the MultiNet formalism and includes, for instance, semantic roles and selectional restrictions of verbal arguments, but also conceptual relations of the wordnet type. Current research topics are the proper semantic representation of prepositional verbs in HaGenLex, the adequate treatment of derivation and alternation, and the semi-automatic extension of the lexicon by corpus-based methods.

- Hartrumpf, S., Helbig, H. and Osswald, R. 2003. The Semantically Based Computer Lexicon HaGenLex – Structure and Technological Environment. *Traitement automatique des langues*, 44(2): 81-105.
- Kunze, C. 2004. Lexikalisch-semantische Wortnetze. In: K.-U. Carstensen et.al. (Hrsg.): *Computerlinguistik und Sprachtechnologie*. 2. Auflage. Spektrum Akademischer Verlag, pp.423-431.
- Kunze, C. and Lemnitzer, L. 2005. Die Zukunft der Wortnetze oder die Wortnetze der Zukunft – ein Roadmap-Beitrag. In: B. Fisseni et.al. (Hrsg.): *Sprachtechnologie, mobile Kommunikation und linguistische Ressourcen*. Peter Lang Verlag, pp. 339-352.



---

Aljoscha Burchardt, Katrin Erk, Anette Frank, Andrea Kowalski,  
Sebastian Pado, Manfred Pinkal  
Universität des Saarlandes

**Consistency and coverage: Challenges for exhaustive semantic  
annotation**

Mi 14.45-15.45

---

The SALSA project at Saarland University is annotating the German TIGER corpus with role-semantic information in the FrameNet paradigm; a first release is in preparation. Since one of SALSA's motivations is to semi-automatically acquire a semantic lexicon from the corpus, we have to maximise both annotation *consistency* and *coverage*, which are especially difficult to reconcile for the frequent cases of *vagueness*. We apply two strategies:

To ensure consistency, annotation is done by two annotators in changing pairs. Inter-annotator agreement (on a test set) is fairly high at 89% on frames (roles 87%). We currently do double *adjudication*, too. Inter-adjudicator agreement is at 99% on frames (roles 96%). The remaining disagreements typically involve linguistically difficult phenomena.

Vagueness occurs e.g. when sentences are compatible with more than one frame (i.e. a proto-typical situation). Example (1) shows a frequent usage of *feststellen*, where the intended meaning could be either of the frames STATEMENT and COMING\_TO\_BELIEVE. We annotate such cases with both frames, labeling them as "underspecified". We also allow role underspecification, as for the roles CONTENT and TOPIC of the frame AWARENESS in (2).

(1) Zudem gibt es, so *stellte* die Prüfungskommission jüngst *fest*, viele Unregelmäßigkeiten.

(2) Rektor Svec sei [<sub>Content/Topic</sub> über die dubiosen Tätigkeiten] „nicht *informiert*“ gewesen.

A practical advantage of underspecification is that it does not force annotators to make impossible choices. A theoretical advantage is that we can identify lemmas and frames that are prone to vagueness, and encode this information in the SALSA lexicon.

In our talk, we will address the appropriateness of underspecification in practical annotation: Do annotators use it to encode real vagueness and ambiguity, or do they feel (subjectively) unsure? By comparing annotated and adjudicated data, we investigate how the use of underspecification "evolves". Preliminary figures indicate that frame underspecification is mostly confirmed in adjudication, while role underspecification shows a less regular picture.

K. Erk, A. Kowalski, S. Pado and M. Pinkal. 2003. Towards a Resource for Lexical Semantics: A Large German Corpus with Extensive Semantic Annotation. Proceedings of ACL 2003, Sapporo, Japan.

C. Fillmore, C. Johnson and M. Petruck. 2003. Background to FrameNet. Intl. Journal of Lexicography 16.

A. Kilgarriff and J. Rosenzweig. 2000. Framework and results for English SENSEVAL. Computers and the Humanities 34 (1-2).

---

Petra Steiner  
Universität Erfurt  
**Verb classes in language use vs. FrameNet's frames**  
Mi 16.15-16.45

---

This investigation compares semantic classes for lexical units with verb classes, which are induced from parsed corpora based on syntactic and distributional properties. The underlying approach is based on the assumption that there is a strong connection between the semantic and the syntactic properties of verbs, as proposed by Levin (1993) and supported by research during the last decade.

Two kinds of materials have been used: Firstly, this analysis is based on the lexicographic database of the FrameNet project, which has been carefully developed by drawing on huge corpora, secondly, it is based on syntactically-annotated corpora (Treebank-3). The information extracted from these databases comprises verbs of the lexical field of *communication* (e.g. *announce*, *claim*, *confess*), as discussed in Baker & Ruppenhofer (2002) in their comparison of Levin's and FrameNet's classes. For these verbs, the syntactic patterns (see Hunston & Francis 2000) are extracted from Treebank-3. The relevant features are the syntactic phrase types, grammatical functions, and heads of prepositional phrases within their subcategorization frames. Statistical hierarchical clustering is performed based on the similarities of these discriminating features of lexemes and yields classifications which are closer to Levin's classes than to FrameNet's.

Furthermore, the lexically overlapping word forms of verbs of noise (from the frame of COMMUNICATION\_NOISE and the frame of MOTION\_NOISE) are classified. For those verbs, which comprise lexical units such as *purr*, *screech*, or *splutter*, lexical units and features are extracted from the FrameNet annotation database and clustered by the same methods as in the preceding experiment. Clustering results show, that both frames are highly consistent with the pattern-based classifications of the data. This implies that patterns can be used to differentiate between lexical overlapping frames, provided that sufficient and correctly frame-annotated data are available. So this method produces results that are useful in the development of unsupervised or semi-supervised devices in the creation of lexical databases.

Baker, Collin F., & Josef Ruppenhofer. 2002. FrameNet's Frames vs. Levin's Verb Classes. In J. Larson and M. Paster (Eds.) *Proceedings of the 28th Annual Meeting of the Berkeley Linguistics Society*. 27-38.

Hunston, Susan, & Gill Francis. 1999. *Pattern Grammar. A corpus-driven approach to the lexical grammar of English*. Amsterdam, Philadelphia: Benjamins. (Studies in Corpus Linguistics 4).

Levin, Beth. 1993. *English verb classes and alternations*. Chicago: Chicago University Press.

---

Sabine Schulte im Walde, Heike Zinsmeister  
Universität des Saarlandes, Universität Tübingen  
**A statistical grammar as empirical resource for inducing lexical  
semantic phenomena**  
Mi 16.45-17.15

---

Our talk presents the grammar modelling framework of *Head-Lexicalised Probabilistic Context-Free Grammars (HeadLex-PCFGs)*, cf. Charniak (1997) and Carroll and Rooth (1998), as an empirical resource for inducing quantitative lexical properties at the syntax-semantics interface. The trained grammar model provides lexicalised rules and syntax-semantics head-head co-occurrences which enable the induction of semantic phenomena.

**Properties of the German HeadLex-PCFG:** The context-free grammar was developed with specific attention towards verb subcategorisation. The grammar concentrates on mass phenomena; i.e. in favour of covering linguistic properties of a large amount of lexical items, selected irregularities were disregarded. In addition, the rules deliberately contain a considerable degree of ambiguity, in order to train lexeme-specific preferences.

**Inducing lexical semantic phenomena:** We demonstrate the usage of the grammar model for inducing various lexical properties at the syntax-semantics interface. For example, Schulte im Walde (2003) defined subcategorisation properties for verbs with respect to frame types and selectional preferences, and used them to induce semantic verb classes. Zinsmeister and Heid (2003, 2004) identified idiosyncratic properties of adverbs and of noun compounds by extracting distributional data from the grammar model. Additional potential for semantic properties includes e.g. word order preferences with respect to verbs and clause types, and the induction of semantic adjective classes as based on modifiers and head-head combinations.

**Discussion of design properties:** The design and the exploitation of the German HeadLex-PCFG are quite idiosyncratic, as compared to the induction of lexical properties from treebanks, or Viterbi parses. More specifically, we would like to discuss the general HeadLex-PCFG framework as basis for lexical acquisition, the design criteria of mass vs. detailed phenomena description, the deep vs. flat analysis structures, the advantages and disadvantages of introducing ambiguity into grammar rules, and the development and training effort.

Carroll, G. & M. Rooth 1998. Valence Induction with a Head-Lexicalized PCFG. In: *Proceedings of EMNLP*.

Charniak, E. 1997. Statistical Parsing with a Context-Free Grammar and Word Statistics. In: *Proceedings of the National Conference on Artificial Intelligence*.

Schulte im Walde, S. 2003. Experiments on the Automatic Induction of German Semantic Verb Classes. PhD thesis, University of Stuttgart.

Zinsmeister, H. & U. Heid 2003. Identifying predicatively used adverbs by means of a statistical grammar model. In: *Proceedings of Corpus Linguistics*. - 2004. Collocations of Complex Nouns: Evidence for Lexicalisation. In: *Proceedings of KONVENS*.

---

Chris Biemann, Rainer Osswald  
Universität Leipzig, FernUniversität in Hagen  
**Experiments on extracting lexical semantic features from contextual  
attributes**  
Mi 17.15-17.45

---

Natural language processing tasks that rely on deep semantic analysis, such as question answering, call for large lexical semantic resources of high quality. Since constructing such resources by hand is extremely time-consuming, it is tempting to extract lexical semantic information automatically from large text corpora by statistical methods. We argue that in order for such an approach to be successful with respect to size and quality of the resulting lexicon, it is prerequisite to start with a reasonably large and reliable semantic base lexicon.

Our talk describes experiments on the automatic assignment of semantic features to nouns. The base lexicon is HaGenLex, a manually compiled semantic lexicon for German, and the co-occurrence statistics is derived from the German corpus of the Wortschatz Project. The underlying assumption is, first, that the semantic features of a lexeme can be determined by analysing appropriate contextual attributes that co-occur with the lexeme in the corpus, and, second, that the entries of the base lexicon can be used to determine as to which contextual attributes are indicators for which semantic features. For nouns, a simple contextual attribute is the attributive adjective, which is fairly easy to access in the corpus; another possible attribute is the verb complement position occupied by the noun.

We have pursued several experiments on noun classification based on attributive adjectives. The training set consists of about 7,000 HaGenLex nouns, which are classified with respect to their ontological type (*situation*, *substance*, etc) and several binary semantic markers such as ANIMATE and MOVABLE; from the corpus, we extracted over 160,000 nouns that co-occur with at least one of 23,400 adjectives. The resulting classifier has been evaluated by cross-validation and shows an overall precision of 82% at a recall of 34%.

We discuss the implications of these results and possible ways to take further contextual attributes into account.

- Biemann, C., Bordag, S., Heyer, G., Quasthoff, U. and Wolff, C. 2004. Language-independent Methods for Compiling Monolingual Lexical Data. In: *Proceedings of CicLING 2004*. LNCS 2945. Springer Verlag, pp. 215-228.
- Biemann, C. and Osswald, R. 2005. Automatische Erweiterung eines semantikbasierten Lexikons durch Bootstrapping auf großen Korpora. In: B. Fisseni et al. (Hrsg.): *Sprachtechnologie, mobile Kommunikation und linguistische Ressourcen*. Peter Lang Verlag, pp. 15-27.
- Hartrumpf, S., Helbig, H. and Osswald, R. 2003. The Semantically Based Computer Lexicon HaGenLex – Structure and Technological Environment. *Traitement automatique des langues*, 44(2): 81-105.
- Miller, G.A. and Charles, W.G. 1991. Contextual correlates of semantic similarity. *Language and Cognitive Processes*, 6(1): 1-28.

---

Matthias Hölzner  
Universität Dortmund  
**Gewichtung nominaler Argumentstellen**  
Mi 17.45-18.15

---

Die Konzeption lexikalisch-semantischer Ressourcen für den deutschen Wortschatz (wie z.B. HaGenLex o.ä., vgl. Osswald/Helbig 2004) ist in besonderem Maße von der Frage geleitet, welche Gesichtspunkte bei der Beschreibung aus Sicht der theoretischen Semantik als wichtig angesehen werden. In meinem Beitrag soll die Beschreibung der Valenzeigenschaften von Substantiven im Mittelpunkt stehen. Die allgemein vertretene These von der grundsätzlichen Fakultativität der Argumentrealisierung bei valenten Substantiven muss angesichts der in Hölzner (2005a) erhobenen Daten neu überdacht werden: Etliche intraphrastische Realisierungswerte bestimmter nominaler Argumentstellen sind sehr hoch, und zwar in der Tat ähnlich hoch wie die Realisierungswerte von obligatorischen Ergänzungen bei Verben. Es erscheint aber auch möglich zu sein, dass die Frage nach Obligatorik vs. Fakultativität nominaler Argumente grundsätzlich falsch gestellt ist. Es hat sich vielmehr der Eindruck verstärkt, dass bei Substantiven fast ausnahmslos eine mehr oder weniger stark ausgeprägte Fokussierung auf eine bestimmte Argumentstelle vorliegt, mit der dann auch unterschiedliche Grade der Weglassbarkeit einher gehen (vgl. Hölzner 2005b). Diese „fokussierten“ Argumentstellen sind bei den *ung-*Nominalisierungen in den meisten Fällen die am tiefsten in die „lexikalisch-semantische Struktur“ (LSS im Sinne von Ehrich/Rapp 2000) eingebetteten Argumente (bspw. erweist sich beim Substantiv *Lieferung* das THEME-Argument als stark fokussiert; es ist zugleich auch das rangniedrigste Argument in der LSS). Bei einer solchen Fokussierung ist offensichtlich eine Realisierung außerhalb der Nominalphrase zumindest erschwert (bei *Lieferung* wird das THEME-Argument in ca. 60 % aller Belege intraphrastisch realisiert). Umgekehrt gibt es Argumentstellen, die für die intraphrastische Argumentrealisierung eindeutig dispräferiert sind (bspw. wird das AGENT-Argument in nur 0,74 % aller *Berücksichtigung*-Belege intraphrastisch realisiert).

- Ehrich, Veronika / Rapp, Irene (2000): Sortale Bedeutung und Argumentstruktur: *-ung-*Nominalisierungen im Deutschen, in: *Zeitschrift für Sprachwissenschaft*, S.245-303.
- Hölzner, Matthias (2005a): *Substantivvalenz. Empirische Untersuchungen zu semantischen und syntaktischen Umgebungen deutscher Substantive*. Dissertation Universität Dortmund.
- Hölzner, Matthias (2005b): The syntax-semantic interface of German nouns, in: *Proceedings of the 2005 International Course and Conference on Role and Reference Grammar*, S.397-416.
- Hölzner, Matthias (2006): Empirisch fundierte Überlegungen zur Substantivvalenz, in: *Sprachwissenschaft* (1).
- Osswald, Rainer / Helbig, Hermann (2004): Derivational semantics in HaGenLex – an interim report, in: Langer, Stefan / Schnorbusch, Daniel (Hrsg.): *Semantik im Lexikon*. Tübingen, S.89-130.

---

Iryna Gurevych  
EML Research gGmbH

**Thinking beyond the nouns – computing semantic relatedness across parts of speech**  
Do 09.00-09.30

---

Semantic relatedness is any kind of lexical-functional association between two words. For example, given the words “surfen”, “Internet” and “Apfel”, human judges are likely to classify “surfen - Internet” as being more closely related than “Internet - Apfel” or “surfen - Apfel”. Information about semantic relatedness is often required in natural language processing (NLP) applications. Previous work to approximate human judgments of semantic relatedness on the basis of GermaNet (Kunze, 2004) was subject to certain limitations from the point of view of robust NLP applications. Firstly, experimental work based on a limited number of word pairs (65) does not always translate into good performance in NLP systems, which process large amounts of unrestricted discourse. Secondly, NLP applications have been confined to using information about semantic relatedness of nouns only. In this case, the discourse is represented as a set of noun concepts, which does not capture important information from the text contained in verbs or adjectives. Extending the representations to include content words beyond nouns requires that metrics of semantic relatedness are improved such that they operate for words belonging to different parts of speech. In the present contribution, we introduce a new dataset containing 350 word pairs of different parts of speech created in a corpus-based manner. Human subjects annotated the word pairs for their semantic relatedness. In the next step, we applied a number of existing metrics (Gurevych and Niederlich, 2005) and a new metric based on path lengths to compute semantic relatedness for the new dataset. We analyze the performance of individual metrics with respect to the new task of computing semantic relatedness across parts of speech. We discuss the coverage of GermaNet for 350 corpus-based word pairs and implications for pre-processing the data. Further structural analysis of GermaNet determines what kind of knowledge relevant to compute semantic relatedness is currently captured in GermaNet and how it could be extended to cover more types of semantic relatedness existing across parts of speech.

- Iryna Gurevych and Hendrik Niederlich. 2005. *Measuring Semantic Relatedness of GermaNet Concepts*. In: B. Fisseni, H.-Ch. Schmitz, B. Schröder & P. Wagner (eds.) *Sprachtechnologie, mobile Kommunikation und linguistische Ressourcen: "Applications of GermaNet II"*. Peter Lang, 462-474.
- Claudia Kunze. 2004. *Lexikalisch-semantische Wortnetze*. In: K.-U. Carstensen, C. Ebert, C. Endriss, S. Jekat, R. Klabunde & H. Langer (eds.), *Computerlinguistik und Sprachtechnologie. Eine Einführung*. Heidelberg, Germany: Spektrum Akademischer Verlag, 423-431.

---

Verginica Barbu Mititelu, Eduard Barbu  
Romanian Academy Bucharest, Graphitech  
**A theoretical framework for checking a wordnet quality**  
Do 09.30-10.00

---

In the last decade wordnets became more and more important for AI. They were successfully used for tasks ranging from WSD and summarization to machine translation and IR. However, despite PWN's success many authors outlined some of its problems. For example, Gangemi, Guarino and Oltramari criticized the confusion between IS-A and INSTANCE-OF relations. Gomez also criticized PWN based on a semantic interpretation algorithm that links verb classes with ontological categories of wordnet. With the advent of the semantic web, which makes extensive use of the existing wordnets, the wordnet quality becomes even more important and also makes the focus of our paper.

Wordnet quality<sup>1</sup> refers to two things: wordnet consistency and correctness, and wordnet external adequacy. For defining wordnet consistency and correctness we consider a wordnet to be a set of assertions about the world. The assertions take the following form:  $A = \text{"The concept } X \text{ is in Relation with the concept } Y\text{"}$ , where *Relation* is any of the conceptual relations defined by a wordnet. We use the notion of consistency as in studies of logical systems.

The problem of correctness arises when we want to see if an assertion that a wordnet makes is true or not. Many of PWN correctness errors are due to an informal specification of the conceptual relations used. Therefore we formalize the main conceptual relations in PWN and we demonstrate the efficacy of working with formally defined meaning of relations by testing a set of PWN synsets.

The problem of external adequacy arises when we want to see how well the wordnet reflects the language reality that is if it has broad (tending to extensive) coverage of the language or if its subjective component is reduced. In this context we argue that a wordnet should be data driven from its early stages, based on empirical evidence, such as balanced corpora, thesauri, monolingual dictionaries, etc. The notion of regular polysemy is also very helpful in this respect.

Gomez, F. 1998. "Linking WordNet Verb Classes to Semantic Interpretation". In *Proceedings of the COLING-ACL Workshop on the Usage of WordNet on NLP Systems. Universite de Montreal, Quebec, Canada, pp. 58-64.*

Gangemi, A., Guarino, N., Oltramari A. 2001. Conceptual Analysis of Lexical Taxonomies: The Case of WordNet Top-Level In C. Welty, B. Smith (eds.), *Proceedings of FOIS ACM press*, pp. 285-296.

---

<sup>1</sup> Wordnet quality should be kept distinct from data quality; the latter imposes conformity of wordnets to technical specifications.

---

Kurt Eberle  
Lingenio GmbH

**The lexicon and the treatment of lexical ambiguity in *translate* and *office dictionary***

Do 10.00 – 10.30

---

*translate* is a commercial Machine Translation system whose roots trace back to IBM's Logic based Machine Translation project (LMT). *office dictionaries* are machine-readable translation dictionaries which permit to look up words in the context so that the contextually adequate meaning and translation can be highlighted in the corresponding lexical entry.

These translation aids are based on dictionaries with detailed syntactic and semantic information about the words and translation relations. Considering word readings, their size is in the six-digit range each. In this talk we will sketch the lexicon formalism of these underlying dictionaries. Using examples from highly ambiguous words (*sein/make/faire*), we illustrate the formal means of specifying complements of word readings (which are: simple categorial restrictions, syntactic or semantic type information and phrase schemata for multi-word expressions and idioms). We discuss the pros and cons of different designs of entries (lemmata) and their readings, called *senses*, and translation relations between senses: On the one hand, one aims at *flat* (underspecified) descriptions of senses for obtaining economic results in parsing with only few analyses (which combine different readings in one underspecified representation). On the other hand, detailed knowledge about the senses (with precise semantic information about the admissible types of complements) often is a must for determining the correct transfer equivalent. Further diverging interests are presented by ergonomic considerations including presentation, readability, tractability of entries, etc.

We will show what kinds of problems arise from semantically deep descriptions in the presence of metonymic use of words in texts and by the fact that there are always faulty selectional restrictions in large dictionaries and which issues we see in order to optimize robustness and ergonomics of the system as a whole.

Bläser, B. 1993. „TransLexis: An Integrated Environment for Lexicon and Terminology Management” In: P. Steffens (ed) Machine Translation and the Lexicon, Third International EAMT Workshop, Heidelberg, April 26-28, 1993, Proceedings. Lecture Notes in Computer Science 898 Springer 1995.

Eberle, K. 2001. „FUHR-based MT, head switching and the lexicon.” In: Proceedings of the eighth Machine Translation Summit. Santiago de Compostela.

McCord, M. 1989. „Design of LMT.” In: Computational Linguistics 15

Storrer, A., Schwall, U. 1993. “Description and Acquisition of Multiword Lexemes In: P. Steffens (ed) Machine Translation and the Lexicon, Third International EAMT Workshop, Heidelberg, April 26-28, 1993, Proceedings. Lecture Notes in Computer Science 898 Springer 1995.

<http://www.lingenio.de>. *translate* and *office dictionary*



---

Michael Beißwenger, Angelika Storrer, Sandra Wellinghoff  
Universität Dortmund  
**Terminologische Wortnetze für die Informationssuche in  
Fachtextcorpora**  
Do 10.30-11.00

---

In unserem Beitrag möchten wir zeigen, dass der am *Princeton WordNet* entwickelte Wortnetz-Ansatz sich auch gut dafür eignet, Bezüge zwischen terminologischen Einheiten und theorie- und domänenspezifischen terminologischen Systemen zu repräsentieren und diese Repräsentation für die Informationssuche in Fachtextcorpora oder für die automatische Verlinkung von Fachtexten nutzbar zu machen. Grundeinheiten der von uns vorgeschlagenen Modellierung terminologischer Wortnetze sind *Termini* und *TermSets*. Ein *TermSet* enthält *Termini* einer Domäne, die aus verschiedenen terminologischen Systemen (z.B. von verschiedenen Autoren oder „Schulen“) stammen, die aber auf ähnliche Entitäten des Gegenstandsbereichs der Domäne referieren; es entspricht also cum grano salis dem Konstrukt des *Synsets* in allgemeinsprachlichen Wortnetzen. Das Relationeninventar des englischen *WordNet* wird um Relationen ergänzt, mit denen Bezüge zwischen verschiedenen terminologischen Systemen rekonstruierbar gehalten werden. Am Beispiel eines terminologischen Wortnetzes zu den Domänen „Texttechnologie“ und „Hypertextforschung“ werden wir diese spezifischen Relationen und deren Funktion erläutern. Der Mehrwert der vorgeschlagenen wortnetzbasierter Modellierung gegenüber terminologischen Thesauri besteht darin, dass die Suche nach *Termini* in Fachtextcorpora auf verschiedenen Granularitätsstufen erfolgen kann: Je nach Nutzerinteresse und Anwendungskontext kann nach einem *Terminus* im Rahmen einer bestimmten Theorie oder aber nach einer konzeptuellen Entität in einem wissenschaftlichen Gegenstandsbereich (also nach allen *Termini* eines *TermSets*) gesucht werden. Am Beispiel des o.g. terminologischen Wortnetzes möchten wir zeigen, wie dieser Mehrwert für die Recherche und die Informationsgewinnung in einem hypertextualisierten Fachtextcorpus genutzt werden kann. Abschließend möchten wir Überlegungen dazu anstellen, (1) wie der Aufbau terminologischer Wortnetze aus Fachtextcorpora mit sprachtechnologischen Tools unterstützt werden kann und (2) wie sich terminologische Wortnetze mit allgemeinsprachlichen Wortnetzen verknüpfen lassen.

---

Ernesto William De Luca, Andreas Nürnberger  
Otto-von-Guericke-Universität Magdeburg  
**The use of lexical resources for sense folder disambiguation**  
Do 11.30-12.00

---

In this presentation we discuss the problem of sense disambiguation using lexical resources with a focus on the application of disambiguation methods in information retrieval systems. Here, lexical resources can be used to improve the search process, for example, by categorizing search results into classes where the search terms are used in the same (or very similar) meaning. Based on the analysis of the WordNet (Fellbaum, 1998) structure we can see that a redesign of the concepts/SynSets is recommended for information retrieval purposes (see De Luca and Nürnberger 2004). Therefore, we implemented four online methods to merge SynSets based on the relations of hypernymy and hyponymy, meaning context and domain. We evaluated our approach based on a small corpus of 252 documents retrieved from web searches that had been manually annotated. We compared the manual annotated classes with the Sense Folders assigned using the approach described in (De Luca and Nürnberger 2005) extended by the above mentioned merging functions. Based on this first evaluation, the hypernym approach seems to appropriately merge Sense Folders that have similar hypernyms which even might be labelled with different domain descriptions. However, a better classification is obtained for words that have fewer meanings (SynSets) before merging starts. The second approach based on hyponyms almost never merged SynSets due to the usually very different hyponyms assigned to each sense. Using the third approach, we merge SynSets not having the same hypernyms, but similar words given from the descriptions of all relations and words together. With the fourth approach, that merges SynSets that are assigned to the same domain, we are sure to merge Sense Folders that belong to the same context, describing it in a different way. The classification is always the best, but the Factotum problem described in (De Luca and Nürnberger 2004) persists. If this merged class contains very different meanings and is used for classification, this classification is worse than before. The possibility to exclude such classes will be studied in future work, e.g. by analyzing approaches that exploits combined information from the first three merging methods.

- De Luca, E. W., and Nürnberger, A., 2004, "Improving Ontology-Based Sense Folder Classification of Document Collections with Clustering Methods", In: Proc. of 2nd Int. AMR Workshop, pp 72-86.
- De Luca, E. W., and Nürnberger, A., 2005. Meta Search Engine for User Adaptive Information Retrieval Interfaces for Desktop and Mobile Devices In: Proc. of the PIA 2005 – Workshop on New Technologies for Personalized Information Access, Edinburgh, Scotland.
- Fellbaum, C., 1998, "WordNet, an electronical lexical database", Cambridge, MIT Press.

---

Jan Burse  
WB DATA AG  
**Using lexical semantic resources to extract product profiles from text catalogues**  
Do 12.00-12.30

---

In the following we consider the use of lexical semantic resources in the extraction of product profiles from building industry text catalogues. We describe a system that was built as part of the Swiss KTI Project 6621.2. The text catalogues are available in the form of hierarchically organised text positions, as shown in the German example below:

```
104340000 Faserdämmstoffe aus Glaswolle
104341000 Isover Glaswoll-Dämmstoffe
104341100 Isover Matten ca. kg/m3 12
104341110 Isover Rollisol ca. kg/m3 12
104341111 m2Isover Rollisol mm 100
```

Profile expressions consist of parameter value pairs. The tools available before the project only provided find and replace of whole text positions by profile expressions. The new goal was to produce more accurate and efficient tools, as well as reusable resources. A further requirement was to support multiple languages.

Our approach was to start from a seed lexicon and then to develop a grammar and complete the lexicon. For the seed lexicon we used some synset matching to combine the English WordNet dictionary and thesaurus with the German, French and English Tulox dictionary. The grammar formalism was derived from definite clause grammars.

We provided our test users with semi-automatic processing. The semi-automatic processing includes a chart parser driven by the dictionary and the grammar rules and as well symbolic processing driven by the thesaurus. We manually developed some grammar whereby the test users completed the dictionary and the thesaurus for a set of German example text catalogues.

- Atserias, J. et al.: Combining Multiple Methods for the Automatic Construction of Multi-lingual WordNets, in Nicolov, N. et al.: Recent Advances in Natural Language Processing II, John Benjamins B.V., 2000
- Barnett, B., Lehmann, H. and Zoeppritz, M.: A Word Database for Natural Language Processing, in 11th International Conference on Computational Linguistics, Bonn, 1986
- Helbig, H.: Die semantische Struktur natürlicher Sprache, Wissensrepräsentation mit MultiNet, Springer Verlag, 2001
- Lehrberger, J.: Automatic Translation and the Concept of Sublanguage, in Readings in Machine Translation, ed. Nirenburg, S., Somers, H. and Wilks, Y., The MIT Press, 2003
- Miller, G.A.: Nouns in WordNet, in Fellbaum, C. et al: WordNet: An Electronic Lexical Database, MIT Press, 1998.
- Ondracek, N.: Der DIN Merkmalsserver und seine Anwendung zum semantischen Mappen von Produktinformationen. Paradine Präsentation, 2002
- Stegman, R. & Woerndl, W.: Using GermaNet to Generate Individual Customer Profiles, Workshop on Applications of GermaNet, University of Bonn, Germany, March, 2005

---

Luís Sarmento, Ana Sofia Pinto, Luís Cabral  
Linguateca, Pólo CLUP

**REPENTINO: A Collaborative wide-scope gazetteer for Named Entity  
Recognition in Portuguese**  
Fr 10.30-11.00

---

Portuguese is one language in which the lack of publicly available lexical-semantic resources has hindered the development of named entity recognition (NER) systems. Building and maintaining such resources is costly and researchers refrain to invest in their construction individually. Our answer to this problem was doing it in collaborative fashion. The Porto Node of Linguateca has thus started the development of REPENTINO (REPositório para reconhecimento de ENTIdades NOmeadas), a publicly available gazetteer organized according to a flexible hierarchy that includes 11 top categories and 97 subcategories, currently storing ca. 450,000 examples of named entities.

A significant percentage of examples has been mined from WPT03 (the 14 GB collection of the Portuguese Web available from [http://xldb.fc.ul.pt/linguateca/WPT\\_03.html](http://xldb.fc.ul.pt/linguateca/WPT_03.html)), using simple lexical patterns searches. We will discuss the advantages and limitations of such simple methods regarding the type of entities that are actually possible (or not) to mine from corpora. A great number of examples have also been manually collected from the web in content specific sites, which allowed us to greatly increase the variety of REPENTINO. Additionally, various on-line contributors have also submitted more than 5,000 examples, using a form that we developed for this specific purpose in REPENTINO's website. We hope that this number will increase after our initial bootstrapping effort. The classification hierarchy used in REPENTINO was developed from a reduced set of reasonably consensual semantic categories inspired by the HAREM evaluation contest (<http://poloxldb.linguateca.pt/harem.php>). This initial set was then refined using a top down approach, according to the type of named entities which we were actually able to find in corpora and on the web, leading to a large hierarchy with some semantic categories which are usually absent from this type of resources.

REPENTINO has already been used as the underlying gazetteer of SIEMÊS, our NER system, incidentally one of the top scoring systems in HAREM, and we attribute its good performance to the high diversity of examples in REPENTINO. REPENTINO has been manually validated to ensure its quality and is now freely available for download in XML format at <http://www.linguateca.pt/repentino/>. Interactive search functions are also provided in the site.

---

Gerd Unruh, Jens Baitinger, Dejan Milosavljevic, Florian Schlegel  
Fachhochschule Furtwangen

**Optimising the access to large linguistic data structures**

Fr 11.00-11.30

---

We are investigating fast accesses to linguistically structured data like ontologies and frames:

1) using relations, implemented in a relational data base system, e.g. MySQL

2) mapping the relational data to memory images in order to gain more speed.

The object oriented paradigm is used for programming, especially the .NET technology from Microsoft. Programming languages are C# and C++, which can be mixed. The access to the data base is carried out through „Data Sets“ of ADO .NET, which uses the „trick“ of keeping as many records as possible in the memory in order to avoid access to disks. An open source effort of Novell will make this technology also available on Unix platforms.

The view on concepts is object oriented (as ancient Indian logic does), i.e. consisting of attributes and functions. The concepts are interconnected via relations (with possible attributes), and form the different data structures for ontologies. Verb-Frames refer to concepts of ontologies, but may also refer to their syntactic information.

In order to trigger the intuitive handling of concept-relations and frames, the structures are visualized for input, navigation, and searching. A drag and drop interface into the concept- and verb - structures allows the user to drag comfortably data from other sources into his own ontology or frame, and vice versa. A programming interface to the concept- and verb – structures via an interface class allows the use of the components for applications, e.g. semantic searches in texts.

The presentation will be a combination of talk and demonstration of the software.

Baker, C.F., Ruppenhofer J. and Fillmore C.J. 2003. „FrameNet: Theory and Practice“.

Gius, E. 2003. „Vergleich maschinenlesbarer deutscher Lexika nach linguistischem Inhalt, Wertebereichen und Kodierung“. Wissenschaftliche Hausarbeit zur Erlangung des akademischen Grades eines Magister Artium der Universität Hamburg.

Hartrumpf, S., Helbig, H. and Osswald, R. 2003. The Semantically Based Computer Lexicon HaGenLex – Structure and Technological Environment. *Traitement automatique des langues*, 44(2): 81-105.

Helbig, H. 2005. „Knowledge Representation and the Semantics of Natural Language“. Springer.

Kunze, C. and Lemnitzer L. 2004. „GermaNet Software Package“.

Langer, S. 1997. „Selektionsklassen und Hyponymie im Lexikon“. Dissertation Universität München.

Miller, G.A. and Fellbaum, C. „WordNet 2.1 Software“ <http://wordnet.princeton.edu>

Osswald, R. 2003. „Die Verwendung von GermaNet zur Pflege und Erweiterung des Computerlexikons HaGenLex“. GermaNet Workshop Tübingen.

Schnorbusch, D. 1998. „Einfache deutsche Verben. Eine syntaktische und semantische Beschreibung der verbalen Simplizia für das elektronische Lexikonsystem CISLEX“. Dissertation Universität München.

## AG 14 Hybridialekte

Koordination: Anja Voeste

### Programm

Mittwoch 22.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
13.45-14.15	Anja Voeste	<i>Hybridialekte: Einführung</i>
14.15-14.45	Helmut Spiekermann	<i>Ausgleichs- und Verdrängungsprozesse im Regionalstandard Südwestdeutschlands</i>
14.45-15.15	Michael Elmentaler	<i>Zur Rolle des interregionalen Sprachkontakts bei der Genese regionaler Umgangssprachen</i>
15.15-15.45	Roland Kehrein	<i>Zum Variationsraum zwischen der deutschen Standardsprache und Dialekten</i>
15.45-16.15	Kaffeepause	
16.15-16.45	Stephen Laker	<i>Standardisierung und Hybridisierung im Norden Englands</i>
16.45-17.15	Elissa Sobotta	<i>Hybridialekte in Paris. Phonologische Ausgleichsprozesse bei Migranten aus Aveyron und Guadeloupe</i>
17.15-17.45	Doris Tophinke, Evelyn Ziegler	<i>Warum Spracheinstellungen nicht ohne Berücksichtigung des Kontextes beschrieben werden können: Aufriss einer neuen Forschungsperspektive</i>
17.45-18.15		<i>Abschlussdiskussion</i>

---

Anja Voeste  
Universität Augsburg  
**Hybridialekte**  
Mi 13.45-14.15

---

Der Wandel der gesprochenen Sprache, insbesondere der Rückgang gesprochener lokaler Varietäten impliziert keinen mit der Schriftsprachentwicklung vergleichbaren Übergang zu einer Standardvarietät. Anstelle prozessbasierter Übergangsmodele vom Basisdialekt zur Standardsprache soll in der AG *Hybridialekte* diskutiert werden, ob und inwieweit gesprochene Varietäten in Ausgleichsprozessen nicht vielmehr neu entstehen. Mein Einführungsvortrag wird die für die unterschiedliche Entwicklung der gesprochenen Sprache maßgeblichen Faktoren in den Blick nehmen:

Voraussetzung ist – so die These – der für West- und Mitteleuropa im 20. Jahrhundert, insbesondere für die Zeit nach dem Zweiten Weltkrieg charakteristische grundlegende raumstrukturelle, mediale und soziale Wandel. War zur Zeit der Entstehung von Wenkers Sprachatlas Ende des 19. Jahrhunderts der Gegensatz zwischen Städten und flachem Land noch markant, so änderte sich dies mit dem Aufbrechen der traditionellen ländlichen Strukturen, mit der Entstehung weit ins Land hinausreichender urbaner Agglomerationen, mit der Durchindustrialisierung, schließlich der Tertiarisierung der Gesellschaft. Die Stadt-Land-Unterschiede wurden nicht zuletzt durch die immer dichtere Vernetzung der Kommunikationsräume nivelliert; die Bilder der nun sich entfaltenden Konsumgesellschaft wurden durch die durchgreifende Medialisierung omnipräsent.

Diese neuen raumstrukturellen, medialen und sozialen Bedingungen reflektieren sich in der Entwicklung der gesprochenen Sprache: lokale Varietäten werden nicht einfach nur verdrängt, sondern entstehen in neuen Verdichtungsräumen vielfach neu. In Trabantenstädten etwa bilden sich lokale Koines aus; supralokale Hybridialekte werden zu regionalen Umgangssprachen. Zu diskutieren ist, ob Interpretamente wie „Akkomodation“ oder „Prestigedruck“ diese Vorgänge adäquat beschreiben können und ob die Theoriebildung in anderen Disziplinen wie Geschichte, Soziologie, Geographie/Raumplanung hier für die Sprachwissenschaft instruktiv sein kann.

---

Helmut Spiekermann  
Albert-Ludwigs-Universität Freiburg  
**Ausgleichs- und Verdrängungsprozesse im Regionalstandard  
Südwestdeutschlands**  
Mi 14.15-14.45

---

Der Regionalstandard Südwestdeutschlands zeigt auf unterschiedliche Art und Weise Ausgleichs- und Verdrängungserscheinungen. Betrachtet man den Begriff „Ausgleich“ aus der Perspektive der Dialekte, so lässt sich feststellen, dass trotz der großen strukturellen Unterschiede zwischen den in Südwestdeutschland gesprochenen (fränkischen, schwäbischen, niederalemannischen, südalemannischen) Dialekten es doch in dem Sinne zu einer Einigung zwischen ihnen kommt, dass bestimmte regionale/dialektale Merkmale als typische Kennzeichen eines gemeinsamen südwestdeutschen Standards angesehen werden können. Ein Ausgleich findet auch auf der Ebene nicht-regionaler, nicht-standardsprachlicher Merkmale (z.B. die Tilgung von /t/ in <nicht>) statt, die nicht an Dialektgrenzen gebunden und insgesamt in Südwestdeutschland gleichmäßig verbreitet sind. Eine Verdrängung sprachlicher Formen kann man in Fall der regionalen/dialektalen Merkmale annehmen, die beim Übergang vom Dialekt/von der Regionalsprache hin zum (regionalen) Standard verloren gehen.

Im Vortrag wird auf der Basis von Daten aus elf Städten in Baden-Württemberg gezeigt, welche Ausgleichs- und Verdrängungsprozesse im Regionalstandard Südwestdeutschlands erkennbar werden. Es werden dabei in einer diachronen Perspektive Daten aus den Jahren 1961 und 2000/02 miteinander kontrastiert. Der Vergleich zeigt, dass die zu beobachtenden Prozesse noch nicht abgeschlossen und der Zielpunkt zur Entwicklung eines südwestdeutschen Regionalstandards noch nicht erreicht ist.



---

Michael Elmentaler  
Christian-Albrechts-Universität Kiel  
**Zur Rolle des interregionalen Sprachkontakts bei der Genese regionaler  
Umgangssprachen**  
Mi 14.45-15.15

---

Die regionalen Umgangssprachen werden nach dem traditionellen Erklärungsmodell (Bellmann, Munske) als Produkte eines unvollständigen Erwerbs der standarddeutschen Zielsprache durch Dialektsprecher betrachtet, wobei die vom Standard abweichenden Merkmalsvarianten als Lernerinterferenzen gelten. Dieses ‚Lernersprachen‘-Modell hat jedoch, wie sich am Beispiel der niederrheinischen Regiolekte zeigen lässt, nur eine begrenzte Erklärungskraft. Am Niederrhein sind heute zwei in ihrem Variantenprofil stark voneinander abweichende Regiolekte in Gebrauch, die etwa auf der Höhe der Uerdinger Linie aufeinander treffen. Die verwendeten Varianten sind nur zum Teil als Lernerinterferenzen interpretierbar, und die Divergenzen zwischen dem nord- und dem südniederrheinischen Regiolekt lassen sich nicht auf alte Dialektunterschiede zurückführen. Daher ist eine Regiolektentstehung aufgrund ausschließlich ‚vertikaler‘ Ausgleichsprozesse zwischen den alten Basisdialekten und der neuhochdeutschen Standardsprache auszuschließen. Vielmehr deutet das spezifische Merkmalsprofil des zwischen Uerdinger und Benrather Linie gesprochenen Regiolekts darauf hin, dass es sich um eine ripuarische Lehnvarietät handelt.

---

Roland Kehrein  
Philipps-Universität Marburg  
**Zum Variationsraum zwischen der deutschen Standardsprache und  
Dialekten**  
Mi 15.15-15.45

---

Die Standardsprache mit ihren regional geprägten oralen Erscheinungsformen des Alltags bilden seit einigen Jahren einen Interessenschwerpunkt innerhalb der Variationslinguistik (horizontale Dimension). Eng damit verknüpft und ebenfalls im Fokus ist die sprachhistorische Frage nach dem Verhältnis der lokalen Dialekte und der ursprünglich rein schriftlichen Standardsprache und ihrer Aussprachenormierung. Daraus wiederum ergibt sich die Frage nach der linguistischen Struktur des Bereichs zwischen diesen Polen (vertikale Dimension).

In der hier präsentierten Studie, die Teil eines groß angelegten Projekts ist, werden beide Dimensionen untersucht: Auf der horizontalen Dimension wird die standardorientierte Sprachverwendung einer definierten sozialen Gruppe in einer flächendeckend vergleichbaren, relativ formellen Alltagssituation betrachtet (Polizeibeamte bei der Entgegennahme von Notrufen). Auf der vertikalen Dimension

werden an ausgewählten Orten des Sprachgebietes mit denselben Informanten weitere Aufnahmen – mit situativ gesteuert mehr oder weniger standard- bzw. dialektorientierter Sprechweise – durchgeführt und analysiert. Zusammen mit Erhebungen der individuellen Standard- und Dialektkompetenz kann auf diese Weise das Variationsspektrum der Sprecher beschrieben werden.

In meinem Vortrag werde ich Ergebnisse von Dialektalitätsmessungen und Variablenanalysen an einer größeren Anzahl von Orten aus den großen deutschen Dialektregionen vorstellen. Diese Analysen erlauben zunächst die Untersuchung arealer Unterschiede hinsichtlich der Quantität und der Qualität regionaler Merkmale in der Situation Notrufannahme (horizontale Dimension). Außerdem lassen sich die unterschiedlichen Strukturen der Variationsspektren beschreiben. Auf dieser Grundlage lassen sich Schlussfolgerungen hinsichtlich vertikaler und horizontaler Prozesse (z.B. Akkomodations- oder Divergenzprozesse) ableiten.

---

Stephen Laker  
Universität Leiden

**Standardisierung und Hybridisierung im Norden Englands**  
Mi 16.15-16.45

---

In meinem Vortrag möchte ich zeigen, wie sich die traditionellen Dialekte Nordenglands im Laufe der neuenglischen Periode, und vor allem im letzten Jahrhundert, radikal verändert haben. Dabei widme ich mich insbesondere phonologischen Wandelprozessen. Anhand von mindestens zwei nordenglischen Dialekten soll gezeigt werden, wie manche traditionelle Aussprachevarianten verdrängt wurden, während andere teilweise erhalten blieben. Was man heutzutage antrifft, ist eine Angleichung an die Standardsprache, letztlich eine Hybride. Sicherlich findet man zwar noch Indizien der alten Aussprache, wie sie im SED (= Survey of English Dialects) aufgezeichnet sind, jedoch sind viele Elemente quasi ausgeradiert worden. Eine zentrale Frage ist, warum bestimmte Aussprachevarianten verloren gingen und andere nicht. Meine Beantwortung dieser Frage soll nach einer Präsentation der relevanten Daten zur Diskussion gestellt werden. Zudem sind, etwa aus dem deutschen Sprachraum, Beispiele ähnlicher Standardisierungsprozesse bekannt, die zusätzlich in die Diskussion einbezogen werden können.

---

Elissa Sobotta  
Ludwig-Maximilians-Universität München  
**Hybriddialekte in Paris**  
Mi 16.45 – 17.15

---

Im *melting pot* Paris sind gebürtige Pariser in der Minderheit; der Großteil der Bewohner der französischen Hauptstadt stammt vielmehr aus der Provinz oder aus dem Ausland. So treten in Paris ständig französische Varietäten in Kontakt und beeinflussen sich gegenseitig. Dieser Koineisierungsprozess ist wohl bekannt: Bereits das sog. Französisch, die Grundlage des heutigen Französisch, soll nicht etwa auf den ursprünglichen Dialekt von Paris zurückgehen, sondern bereits auf eine Ausgleichsmundart. Doch obwohl die Rolle von Paris in der französischen Sprachgeschichte immer wieder hervorgehoben wird und aufgrund des französischen Zentralismus geradezu ein Paradebeispiel für Hybridisierungsprozesse ist, fehlt es an detaillierten empirischen Untersuchungen zur Konvergenz französischer Varietäten. Dieser Vortrag möchte einen Beitrag dazu leisten, diese Lücke zu schließen. Er befasst sich exemplarisch mit dem phonologischen Profil zweier Gruppen innerstaatlicher Migranten in Paris, nämlich Franzosen aus den *départements* Aveyron (Südfrankreich) und Guadeloupe (Karibik). Dabei handelt es sich um zwei besonders typische Migrantengruppen in Paris. Erstens aufgrund ihrer zahlenmäßigen Stärke: Es gibt mit 320 000 Personen in der *région parisienne* mehr Aveyronnais als in Aveyron selbst; die 150 000 Antillais lassen neben Guadeloupe und Martinique von einer dritten Antilleninsel sprechen, der Ile-de-France. Zweitens handelt es sich um zwei Gruppen mit besonders dichten sozialen Netzwerken (Vereine, Medien, Folkoregruppen etc.) und einer jeweils neuen Identität, die sich im Migrationskontext herausgebildet hat.

Das Korpus der hier vorgestellten Untersuchung besteht aus 20 Minuten Spontansprache von insgesamt 100 Sprechern. Die Daten wurden im Rahmen des *Projekts Phonologie du Français Contemporain* (PFC) erhoben. Die untersuchten phonologischen Variablen sind das Schwa und finale Obstruent-Liquid-Gruppen. Die Analyse zeigt, dass neben sozialen Faktoren (Dauer des Aufenthalts in Paris, Zugehörigkeitsgefühl zur neuen und alten Heimat etc.) auch sprachinterne Faktoren für die Erklärung des Übergangs zwischen den Varietäten herangezogen werden müssen, wie etwa die Morphosyntax und die Semantik. Insbesondere der Sprachgebrauch (Frequenz von Kookkurrenzen) spielt eine wichtige Rolle.

---

Doris Tophinke, Evelyn Ziegler  
Technische Universität Darmstadt, Philipps-Universität Marburg  
**Warum Spracheinstellungen nicht ohne Berücksichtigung des Kontextes  
beschrieben werden können: Aufriss einer neuen Forschungsperspektive**  
Mi 17.15-17.45

---

Spracheinstellungen, spätestens in den 80er Jahren zu einem „hot topic“ in der Soziolinguistik avanciert, interessieren vorrangig unter dem Gesichtspunkt ihrer Relevanz für den Spracherwerb, das Sprachverhalten (Varietäten- und Varietätenwahl) und die Sprachentwicklung (Dialektabbau, Ausbildung von Regionalsprachen, Destandardisierung). Dementsprechend gilt für viele SoziolinguistInnen: „Variation kann nicht ohne Einstellungen und Einstellungen können nicht ohne Bezug auf Variation beschrieben werden“ (Dittmar 1996: 18). In dieser funktionalen Perspektive bleibt das Phänomen selber allerdings unterbestimmt. Es erscheint konstruiert und opak, insbesondere da die aktuelle Genese von Spracheinstellungen in konkreten Kommunikationssituationen (z.B. Narration, Interview, Gruppendiskussion) in den Hintergrund rückt. Zwar werden die Möglichkeiten der sozialen Bedingtheit immer als gegeben vorausgesetzt, aber nicht näher analysiert. Das überrascht, denn die Klärung dieser Frage ist dringend notwendig, und zwar aus zwei Gründen. Erstens führt die Untersuchung dieser Zusammenhänge zu einem klareren Gegenstandsverständnis, umso mehr als bloße Definitionen keine Einblicke in soziokognitive Prozesse liefern. Zweitens kann eine Untersuchung der Rahmenbedingungen, d.h. der soziokulturellen, situativen und interaktionalen Faktoren sowie ihrer Zusammenhänge, mit dazu beitragen, das Verhältnis zwischen Spracheinstellung und Sprachverhalten genauer zu fassen.

Der vorliegende Beitrag versteht sich als ein Versuch, diese neue Perspektive in ihren Umrissen aufzuzeigen und ihre Bedeutung im Rahmen der Spracheinstellungsforschung herauszuarbeiten. Das Interesse gilt dabei dezidiert der Kontextproblematik, d.h. Fragen der Strukturierung und Modellierung kontextueller Einflussgrößen. Möglichkeiten und Reichweite des Ansatzes werden am Beispiel von Spracheinstellungsäußerungen in Weblogs diskutiert.

Dittmar, Norbert (1996): *Soziolinguistik*. Heidelberg: Gross.

## Studentische AG

Koordination: Corinna Handschuh, Christina Unger, Jan Wohlgemut

### Programm

Mittwoch 22.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
13.45-14.15		<i>Begrüßung</i>
14.15-14.45	Anja Moos, Jürgen Trouvain	<i>Prosodische Textstrukturierung bei Sprachsynthese</i>
14.45-15.15	Joseph T. Farquharson	<i>Lexical Transfer: How Many Linguistic Events?</i>
15.15-15.45	Corinna Handschuh , Jan Wohlgemuth	<i>Kultur Höflichkeit und Optimalität</i>
16.15-17.15	Kristina Henschke	<i>Die Sprachsituation im Victoria River District, Nordaustralien</i>

Donnerstag 23.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
10.00-10.30	Thomas Hanke	<i>Additional Remarks on the Typology of Numerals</i>
10.30-11.00	Dominik Bauer, Jürgen Trouvain	<i>Sprecherspezifische Muster von gefüllten Pausen</i>
11.00-11.30	Kaffepause	
11.30-12.00	Vitor Zimmerer	<i>Die "Macht der Wörter" über die Gefühle - empirisch untersucht</i>

Freitag 24.02.2006

<b>Zeit</b>	<b>Name</b>	<b>Titel</b>
10.30-11.30	Sara Neuhauser	<i>Phonetische und Linguistische Variation bei der Stimmverstellung unter Verwendung eines fremdsprachigen Akzents</i>
11.30-12.00	Maksym Gorshenin	<i>Merkmale des aktivischen Sprachtypus in Indianersprachen und im Chamorro</i>
12.00-12.30		<i>Abschlussdiskussion</i>

Schon vielfach wurde an Hand von Produktionsexperimenten festgestellt, dass Texte, die mehrere Absätze umfassen, vom Vorleser prosodisch strukturiert werden (Lehiste, 1975; Mayer, 1999 u.a.). So verlängert sich in der Binnenstruktur beispielweise die Pausendauer nach Sätzen am Absatzende und die Stimmtonhöhe findet sich an Absatzgrenzen in extremeren Bereichen als innerhalb des Absatzes (Deklination der Stimmtonhöhe über Absätze hinweg) (Lehiste, 1975; Mayer, 1999). Sprachsynthesysteme verwenden solche Merkmale bisher üblicherweise nicht.

Deshalb haben wir die Beurteilung der Strukturierung längerer Texte, wenn diese durch Sprachsynthese erzeugt werden, untersucht. In zwei Hörtests, in denen die Sprachsynthese "Mary" (Schröder & Trouvain, 2003) verwendet wurde, wurden unterschiedliche Varianten bezüglich Pausendauer und Stimmdeklination dargeboten. Beurteilt wurde die wahrgenommene Textgliederung und das Textverständnis unter Veränderung der prosodischen Textstrukturierung.

Die zwei folgenden *Hypothesen* wurden überprüft:

1. Absätze in einem Text werden nur dann wahrgenommen, wenn diese durch eine längere Pause getrennt sind und/oder die Deklination auf Makroebene berücksichtigt wird.
2. Die Verbesserung der Textstrukturierung durch Verlängerung der Pausen an Absatzenden und Deklination auf Makroebene führt zur besseren Verständlichkeit des Inhalts.

Dazu hörten jeweils zehn Probanden in beiden Tests Nachrichtentexte in vier Versionen:

1. die nicht veränderte Originalversion der Sprachsynthese
2. mit längerer Pause zwischen den Absätzen
3. mit veränderter Tonhöhe (letzter Satz des Absatzes 8% tiefer, erster Satz 8% höher)
4. Version 2 mit Version 3 kombiniert

Die Versuchspersonen mussten bei beiden Experimenten den Text gemäß ihres auditiven Eindrucks in Absätze gliedern. Um den Einfluss der textuellen Information auszublenden, wurden in Experiment 1 die Texte delexikalisiert. Beim zweiten Experiment musste außerdem die Strukturierung beurteilt und Fragen zum Inhalt beantwortet werden.

Die Ergebnisse zeigen, dass Hypothese 1 zwar nicht auf signifikantem Niveau bestätigt werden kann, aber eine sehr starke Tendenz erkennbar ist. Das zweite Experiment jedoch bestätigte Hypothese 1 insofern, dass die beiden Versionen mit der verlängerten Pausendauer eine signifikante Verbesserung erbrachten.

Auf die Merkfähigkeit wirkte sich die veränderte Strukturierung nicht signifikant aus, so dass Hypothese 2 nur tendenziell bestätigt werden kann. Hierbei können methodische Schwierigkeiten mit der Auswahl der Texte, der Art der inhaltlichen Abfrage sowie der Dauer der zu beurteilenden Äußerungen einen Einfluss ausgeübt haben.

Auch wenn die Rolle der "Paragraphen-Intonation" unklar bleibt, so ist hiermit gezeigt worden, dass sich eine an das natürliche Vorbild angelehnte Anpassung der Pausierung in der synthetischen Sprache lohnt.

- Lehiste, I. (1975): The phonetic structure of paragraphs. In: Cohen & Nooteboom (Hg.), *Structure and Process in Speech Perception*. Springer: Berlin. 195-203.
- Mayer, J. (1999): Prosodische Merkmale von Diskursrelationen. *Linguist. Ber.* 177, 65-86.
- Schröder, M. & Trouvain, J. (2003): The German text-to-speech synthesis system MARY: A tool for research, development and teaching. *Intern. J. of Speech Technology* 6, 365-377.
- 

Joseph T. Farquharson  
University of the West Indies, Max Planck Institute for Evolutionary  
Anthropology  
**Lexical Transfer: How Many Linguistic Events?**  
Mi 14.45-15.15

---

Dictionaries (on historical principles) normally compress all the attested meanings for a particular word into one article. This practice, while unavoidable within the field of lexicography, is often not reflective of the knowledge of most speakers taken as individuals. Most linguists will agree that no speaker has all the words/lexemes listed in dictionaries such as the *Oxford English Dictionary* or the *Deutsches Wörterbuch* in his/her active vocabulary. This notion also extends to the equally true conclusion that no speaker is familiar with all the existing word classes/meanings (sememes) for all the words (s)he knows. For example, an item might be listed in a dictionary as both a noun and a verb, but there may be some speakers who only use it as a noun but never as a verb. Also, a word may be listed as having twenty senses, but probably no speaker knows all twenty. This issue has been largely ignored owing to the general lack of an effective methodology within descriptive semantics.

The paper arises out of my general research which involves tracing the etymology of Niger-Congo words in Jamaican Creole (an English-based Creole language spoken in the Caribbean). Given the problem outlined above the paper will address what exactly it means to inherit or retain a WORD, and how issues related to the shape versus the semantic concepts impact on our approach.

The analysis is geared towards providing a more nuanced presentation of this phenomenon which touches the work of etymologists and comparative linguists. Lexemes can either be inherited with the same or slightly different semantic profiles than their etyma. The second scenario can be further subdivided into cases where semantic content is added and cases where there is loss of some content. Similarity or difference is not always a straightforward indicator of continuity; either might be due to later parallel development based on cross-linguistically common semantic features. The analysis will focus on two aspects of lexical transfer: (i) the inheritance of multiple lexical categories visible in trans-categorial lexemes, and (ii) the survival of the sememes (i.e. semantic content) from the substrate language(s). An attempt will be

made to delineate what category or semantic features represent inheritance as opposed to innovation. The paper will approach these entities from three related angles:

- **Viability:** a substrate lexeme is viable if it designates a referent or event which is intimately linked to the community of speakers engaged in the shift (e.g. aspects of religion, kinship, etc.) or if the lexifier language contains no suitable replacement.
- **Utility:** obviously the lexeme will only survive if opportunities for its usage continue to exist, but bear in mind that “presumably” dead words can be resurrected.
- **Performance:** the survival of the sememes associated with a particular lexeme, is based on whether that sememe is passed on through performance by members of the speech community. This relates to loss/gain of semantic content.

---

Corinna Handschuh, Jan Wohlgemuth  
Heinrich-Heine-Universität Düsseldorf, Max-Planck-Institut für Evolutionäre  
Anthropologie, Leipzig  
**Kultur, Höflichkeit und Optimalität**  
Mi 15.15-15.45

---

In diesem Vortrag wollen wir einen explorativen Versuch vorstellen, das Feld der sprachlichen Höflichkeit mit den Mitteln der Optimalitätstheorie (OT) zu beschreiben. Höflichkeit und ihr Ausdruck durch sprachliche Mittel sollte nicht als Universalie verstanden werden, die auf feststehenden Regeln beruht. Sie (bzw. die dafür verwendeten Strategien) wird vielmehr beeinflusst durch verletzbare Restriktionen kommunikativen Verhaltens. Konversationsregeln – oder -maximen – wie die von LAKOFF (1973) und GRICE (1975) beschriebenen sind verletzlich und liegen häufig miteinander im Widerstreit, z.B. wenn die Qualitätsmaxime einem Sprecher abverlangt würde, eine Wahrheit zu äußern, die das Gesicht von Sprecher oder Zuhörer verletzen könnte. Aufgrund dieser Natur der sprachlichen Höflichkeit scheint es ein viel versprechender Ansatz zu sein, sie mit dem Formalismus der Optimalitätstheorie zu beschreiben.

Anhand von Beispielsituation des Deutschen, Japanischen und Javanischen wollen wir zeigen, wie die Wahl unterschiedlicher Höflichkeitsstrategien sprachübergreifend von denselben Constraints beeinflusst wird, die lediglich durch ihre unterschiedliche Gewichtung zu den verschiedenen Verhaltensweisen führen. Diese Verhaltensweisen können dabei nicht nur interkulturell variieren, sondern auch innerhalb einer Sprechergemeinschaft zwischen unterschiedlichen Situationen.

Im Rahmen des Vortrages wollen wir die Vorzüge und Einschränkungen einer solchen Analyse diskutieren, die es ermöglichen könnte im Rahmen einer faktoriellen Typologie der Höflichkeitsstrategien Generalisierungen über den Bereich der sprachlichen Höflichkeit zu formulieren, welche erklären könnten, warum ein bestimmtes kommunikatives Verhalten in einigen Kontexten als unhöflich, in anderen wiederum als höflich angesehen wird.



- GRICE, H[ERBERT] P[AUL] (1975): *Logic and Conversation*. In: COLE, PETER / MORGAN, JERRY L. [Hrsg.]: *Speech acts*. New York [u.a.]: Academic press. (=Syntax and Semantics. 3). Pp. 41-58.
- LAKOFF, ROBIN (1973): *The logic of politeness; or, minding your P's and Q's*. In: CORUM, C. / SMITH-STARK, T.C. / WEISER, A. [Hrsg.]: *Papers from the Ninth Regional Meeting: April 13-15, 1973*. Chicago Linguistic Society. Chicago, Ill.: Chicago Linguistic Society, 1973. pp. 292-305.
- 

Kristina Henschke

Graz

## **Die Sprachsituation im Victoria River District, Nordaustralien**

Mi 16.15-17.15

---

Quer durch Nordaustralien verläuft die Grenze zwischen der größeren Familie der Pama-Nyunga-Sprachen und die eher als Restkategorie zu bezeichnenden Non-Pama-Nyunga-Sprachen, die vielen Unterfamilien angehören. Entlang dieser Grenze begegnen sich in der Gegend des Victoria River die jaminjungischen (Non-Pama) und die Ngumpin-Sprachen (Pama); traditionell sind Mehrsprachigkeit und enge verwandtschaftliche Beziehungen zwischen verschiedenen Sprachgruppen im Norden Australiens nicht ungewöhnlich, und so hat in dieser Gegend lange Zeit ein intensiver Sprachkontakt zwischen sehr verschiedenen Sprachen geherrscht.

Seit der Besiedelung der Gegend durch die weißen Australier spielt auch Englisch als kulturell, ökonomisch und politisch dominante Sprache eine stetig zunehmende Rolle in der multilingualen Landschaft des Victoria River District, und die Aborigine-Sprachen der Gegend sind akut vom Austerben bedroht.

Im Rahmen eines DoBeS-Projektes (Dokumentation bedrohter Sprachen – ein von der Volkswagen-Stiftung finanziertes und vom MPI für Psycholinguistik in Nijmegen technisch umgesetztes Programm zur Sammlung und Archivierung von Materialien über und in bedrohten Sprachen) habe ich die Chance, mir selber ein Bild von der Sprachsituation zu machen und bei der linguistischen Erfassung insbesondere der Sprache Ngarinyman mitzuarbeiten. Neben einem Überblick über die Sprachen der Region (insbesondere eben des Ngarinyman) und der Rolle des Englischen für die heutigen Sprachsituation möchte ich daher auch Einblicke in die Feldforschung bieten und andere Aspekte der Arbeit mit einer bedrohten Sprache, wie etwa Orthographieentwicklung, bilinguale Schulprogramme oder die Situation der Sprechergemeinden berücksichtigen.

---

Thomas Hanke  
Freie-Universität Berlin  
**Additional Remarks on the Typology of Numerals**  
Do 10.00-10.30

---

Greenberg's seminal article on numerals contains inter alia a classification of additive "links" (Greenberg 1978:264f.), accepted afterwards without further investigation. On the base of a sample of 300+ languages (Hanke 2005), I present some improvements: Comitative constructions are only an indirect source via coordination. Superessives ('on', 'upon') are a common locative source, others including movements and even 'under'. 'Left' does not necessarily suppress the expression of additive bases – cases like *twelve* < '2-left' integrate in a typology of numeral omission. Other sources like bodily concepts deserve equal attention. The main importance of serial sums (base, +1, +2, +3, ...) is correct, but other sums are more than "sporadic" marginals. Altogether, additive numeralization is a special type of conventionalization and is definitely not limited to simple "links".

GREENBERG, Joseph H. (1978). Generalizations About Numeral Systems. In GREENBERG (Ed.), *Universals of Human Language*. Vol. 3 *Word Structure* (pp. 249-295). Stanford.  
HANKE, Thomas (2005). *Bildungsweisen von Numeralia – eine typologische Untersuchung*. Berlin.

---

Dominik Bauer, Jürgen Trouvain  
Universität des Saarlandes, Phonetikbüro Trouvain, (Saarbrücken)  
**Sprecherspezifische Muster von gefüllten Pausen**  
Do 10.30-11.00

---

In dieser Studie wurden Häufigkeit, Typ und Dauer von gefüllten Pausen an einem Korpus deutscher Spontansprache (Jessen et al., 2003) untersucht. Häsitationen wie gefüllte Pausen sind ebenso wie Silbendehnungen und stille Pausen typische Marker von Spontansprache. Der Inhalt dieser "Pausen" wird oft mit "eh" und "ehm" dargestellt, welche die Kurz- und die Langform der gefüllten Pausen darstellen (Clark & Fox Tree 2002). Die Tatsache, dass Sprecher große interindividuelle Unterschiede im Bezug auf ihr Häsitationsverhalten aufweisen, macht die Analyse der gefüllten Pausen interessant für die forensische Analytik, für die idiosynkratische Muster im Sprechverhalten wertvolle Hinweise bieten.

Diese Untersuchung umfasst 72 männliche Sprecher, von denen Aufnahmen unter Normal- und Lombard-Bedingung vorlagen. Die Ergebnisse der 2014 untersuchten gefüllten Pausen zeigte, dass es typische individuelle Muster gab. Drei Sprecher produzierten unter beiden Bedingungen überhaupt keine gefüllten Pausen. Drei weitere Sprecher markierten mit über 10 gefüllten Pausen die obere Grenze für

Pausenhäufigkeit in dieser Studie und grenzten sich dadurch klar von den anderen Sprechern ab. Es wurde eine klare Tendenz der meisten Sprecher festgestellt, "eh" anstelle von "ehm" zu produzieren. In der Gruppe der 36 "Viel-Pausierer" gab es sechs Sprecher, die kein "ehm" produzierten und sechs Sprecher, die kaum "eh" produzierten. Auf der anderen Seite gab es in dieser Gruppe keine exklusiven "ehm"-Pausierer, aber sieben Sprecher produzierten mehr "ehm" als "eh". Die Erwartung, dass Sprecher unter der erhöhten kognitiven Belastung der Lombard-Bedingung mehr gefüllte Pausen produzieren, konnte nicht bestätigt werden. Die mittlere Dauer der gefüllten Pause ist abhängig von der phonetischen Realisierung: "eh" (364 ms) ist kürzer als "ehm" (508 ms). Die häufigste Form von "eh" ist in Artikulation eingebettet, während die häufigste Form von "ehm" von zwei stillen Pausen umgeben ist. Ausreichend lange Tat- und Vergleichsaufnahmen vorausgesetzt, können solche Muster helfen, einen individuellen Sprecher zu erkennen.

---

Vitor Zimmerer

Heinrich-Heine-Universität Düsseldorf

**Die „Macht der Wörter“ über die Gefühle – empirisch untersucht.**

Do 11.30-12.00

---

Wenn in der Sprachkritik von der „Macht der Wörter“ über Handlungsweisen die Rede ist, so ist, wie in Sternbergers (1957) „Aus dem Wörterbuch des Unmenschen“, oft der Einfluss der Sprache auf Emotionen und damit auf die Einstellung zu verschiedenen Konzepten gemeint. Eine manipulierende Sprache, z.B. aus der Politik, soll durch das Wecken von bestimmten Assoziationen Konzepte attraktiv oder abstoßend machen. Es handelt sich also nicht bloß um einen Einfluss der Sprache auf die Wahrnehmung und Kategorisierung von Konzepten im Sinne der Whorfhypothese, wie sie in vielen psycholinguistischen Studien (z.B. zur Farb- oder Raumwahrnehmung) untersucht wird. In diesen Studien fehlt nämlich das Element der Affektion, also die Ebene, auf der die emotionale Reaktion auf Konzepte stattfindet. Dabei ist die Affektion zumindest für die bewusste Handlung wichtig, muss doch zuerst feststehen, ob das Ergebnis einer Handlung überhaupt erwünscht ist. Dies erklärt auch, warum sich viele sprachkritische Argumente in den Medien auf die Affektion beziehen (Vgl. Wengeler 1996).

In diesem Vortrag soll zunächst der Frage nachgegangen werden, wie die emotionale Wirkung von Wörtern aussehen könnte. Dazu werden Ergebnisse aus Primingstudien, unter anderem der affektiven Primingexperimente von Storbeck und Robinson (2004) hinzugezogen. Die Interpretation dieser Studien kommt zu dem nicht überraschenden Ergebnis, dass für die affektive Wirkung die semantische Ebene aktiviert werden muss, es reicht nicht eine bloße morphologische Aktivierung.

Eine eigene Studie soll die emotionale Wirkung von Wörtern in einer Versuchssituation überprüfen, in der die bezeichneten Konzepte auch auf nichtsprachlichem Wege deutlich gemacht werden. Dazu werden Versuchsgruppen unterschiedliche Versionen des gleichen Films präsentiert, die sich lediglich in ihrem das Gesehene beschreibenden Audiokommentar unterscheiden. Ein Film verwendet dazu positiv konnotierte, ein Film negativ konnotierte Wörter. Nach Zeigen des Films werden die Versuchspersonen gebeten, das Gesehene zu bewerten und Assoziationen

zum Film niederzuschreiben. Zu den Kontrollvariablen gehören der Sprecher des Audiokommentars, Stimmlage und Betonung. Andere Variablen lassen sich nicht so einfach kontrollieren: So sollte im optimalen Falle das Wissen um den Inhalt des Filmes und die Grundeinstellung zum Inhalt bei allen Versuchspersonen gleich sein, so dass beobachtete Unterschiede in der Bewertung nur auf die emotionale Wirkung der Wortwahl zurückgeführt werden können.

Die Studie kann einen Effekt der Sprache auf die Assoziationen nachweisen, nicht aber auf die Bewertung. Dies spricht dafür, dass wenn dem Sprecher deutlich ist, welche Konzepte die Wörter genau bezeichnen, die Wortwahl keinen Einfluss auf die Bewertung nimmt.

Storbeck, Justin/Robinson, Michael D. (2004): Preferences and Inferences in Encoding Visual Objects: A Systematic Comparison of Semantic and Affective Priming. In: Personality and Social Psychology Bulletin.

Sternberger, Dolf et al. (1957): Aus dem Wörterbuch des Unmenschens. Hamburg.

Wengeler, Martin (1996): Sprachthematizierungen in argumentativer Funktion: Eine Typologie. In: Böke, Karin/Jung, Matthias/Wengeler, Martin (Hrsg.): Öffentlicher Sprachgebrauch. Opladen, S. 413-430.

---

Sara Neuhauser

Friedrich-Schiller-Universität Jena

**Phonetische und linguistische Variation bei der Stimmverstellung unter Verwendung eines fremdsprachigen Akzents**

Fr 10.30-11.30-

---

Die Möglichkeit einer Stimmverstellung steckt potenziell in jeder Sprachaufnahme, die im Zusammenhang mit einer Straftat steht. Eine Verstellungsart stellt die Stimmverstellung unter Verwendung eines fremdsprachigen Akzents dar.

Für diese Studie wurden zwei Akzente (französischer und türkischer Akzent) ausgewählt, durch die zwölf deutsche Muttersprachler ihre Stimmen verstellen sollten. Untersucht wurden die phonetischen (phonatorischen und artikulatorischen) und linguistischen Variationen, die die Sprecher für ihre Verstellungsversuche vornahmen. Die Analyse des Sprachmaterials zeigte, dass es den Sprechern gelang, zur Stimmverstellung durch einen fremdsprachigen Akzent vielfältige Variationen vorzunehmen, die sich vorwiegend auf den artikulatorischen Bereich konzentrierten. Überraschend war jedoch, dass mehr als die Hälfte aller Sprecher auch Variationen im phonatorischen Bereich zur Stimmverstellung nutzte, während sie im außerphonetischen Bereich jedoch weniger Abweichungen vornahmen als erwartet.

Insgesamt waren die Variationen überraschend vielfältig und entsprachen in vielen Fällen den Abweichungen, die tatsächlich von Deutsch sprechenden Nicht-Muttersprachlern erwartet werden können. Allerdings konnten sie nicht mit der Konsistenz durchgeführt werden, mit der sie wahrscheinlich von einem authentischen Nicht-Muttersprachler durchgeführt worden wären.

Künzel, Hermann J. (1987): Sprechererkennung. Grundzüge forensischer Sprachverarbeitung. Heidelberg: Kriminalistikverlag

- Rolffs, Songül (2003): Türkisch. In: Ursula Hirschfeld et al. (Hrsg.): Phonetik international: von Afrikaans bis Zulu; kontrastive Studien für Deutsch als Fremdsprache. Waldsteinberg: Popp
- Schmidt, Lothar (2003): Französisch. In: Ursula Hirschfeld et al. (Hrsg.): Phonetik international: von Afrikaans bis Zulu; kontrastive Studien für Deutsch als Fremdsprache. Waldsteinberg: Popp
- 

Maksym Gorshenin  
Universität Bremen

**Merkmale des aktivischen Sprachtypus in Indianersprachen und im Chamorro**

Fr 11.30-12.00

---

Der Aktiv-Stativ genannte Sprachtypus, der nicht selten als eine Mischform aus dem Ergativtypus und dem Nominativtypus beschrieben wird, weist jedoch in Wirklichkeit – hinter so „harmlos“ aussehenden oberflächlichen Phänomenen wie „diskrete“ Behandlung des zentralen (valenzgebundenen) Partizipanten intransitiver Verben – fundamental andere Prinzipien der Strukturorganisation auf als die des Nominativ- oder Ergativtypus. Besonders wichtig ist dabei die spezifische Strukturiertheit des Lexikons und die Art der Beeinflussung der Syntax durch das Lexikon.

„Echte“ Aktivsprachen sind heutzutage nur in Amerika belegt. Interessanterweise lassen sich aber viele typische Züge des Aktivaufbaus in unterschiedlichen Maßen auch in Sprachen wieder finden, die eigentlich keine (reine) Aktivsprachen sind. Im Rahmen dieses Vortrags sollen einige wichtige Eigenschaften des Aktiv-Stativ-Typus in indigenen Sprachen Amerikas (hauptsächlich im Dakota/Lakhota) exemplifiziert und mit (morpho)syntaktischen Charakteristika der austronesischen Sprache Chamorro, deren Strukturen in vielen Domänen eine aktiv-stativ-ähnliche Organisation aufweisen, verglichen werden.

Kathrin Beck, Heike Zinsmeister  
Universität Tübingen  
**Semi-automatische Erstellung eines Lexikons von  
ung-Nominalisierungen mit präpositionaler Ergänzung**  
Do 12.30-14.00

---

Für Anwendungen wie "question answering" oder maschinelle Übersetzung ist es wichtig, Zusammenhänge zwischen präpositionalen Ergänzungen von ung-Nominalisierungen und diesen Argumenten bei den Basisverben zu erkennen. z.B.

- Wanderung durch die Heide ⇔ wandern durch die Heide (direktional)
- Unterstützung durch Deutschland ⇔ Deutschland (Subj.) unterstützt

**Grundlagen:** "TüBa-D/Z" Baubank (Tübingen) mit 22.000 Sätzen aus der taz und Suchtool "TIGERSearch"

**Vorgehen:** Automatische Suche aller ung-Nominalisierungen mit Präpositionen oder Sätzen als Modifikatoren, Suche nach syntaktischen Regeln, in welchen Fällen welche Präposition verwendet wird.

**Fragen:** Welche Argumentfunktion beim Verb erfüllen die präpositionalen Ergänzungen der Nominalisierung? Z.B. Wann wird die "durch"-PP der Nominalisierung als Subjekt und wann als präpositionale Ergänzung des Verbs dargestellt? (vgl. Lauterbach 1993)

Helfen Anbindung im Satz, Determiner, Numerus oder modifizierende Adjektive dabei, Regeln zu erstellen (vgl. Aldinger 2005)?

Spielt die Art der Vorgangsnominalisierung eine Rolle bei der Desambiguierung (vgl. Ehrich & Rapp 2000)?

**Ergebnisse:** Wir stellen korpusabgeleitete Regeln für die Abbildung von Präpositionalattributen von ung-Nominalisierungen auf den verbalen Valenzrahmen vor. Z.B. können Ergänzungen mit "durch" zu 94% als Subjekt des Basisverbs dargestellt werden. Die restlichen 6% sind direktionale Argumente zu einem zugrunde liegenden Bewegungsverb.

Aldinger, N. 2005. Corpus-driven disambiguation. In: *Proceedings of CL-2005*, Birmingham.

Ehrich, V., Rapp, I. 2000. Sortale Bedeutung und Argumentstruktur: ung-Nominalisierungen im Deutschen. In: *Zeitschrift für Sprachwissenschaft* 19.2, 245-303.

Lauterbach, S. 1993. *Genitiv, Kompositum und Präpositionalattribut: zum System nominaler Referenz im Deutschen*. Iudicum Verlag.

---

Heike Bieler, Stefanie Dipper, Manfred Stede, Arthit Suriyawongkul  
Universität Potsdam  
**Automatic Text Summarization**  
Do 12.30-14.00

---

Ziel automatischer Textzusammenfassung ist es, den Inhalt eines Textes zu identifizieren und dessen relevante Teile dem Nutzer in kondensierter und an die spezifischen Anwendungsanforderungen angepasster Form zu präsentieren.

Typischerweise umfasst der Prozess des automatischen Zusammenfassens zwei Schritte:

(i) Bestimmung des “Extrakts”, d.h. Auswahl derjenigen Sätze, die die Zusammenfassung bilden sollen; (ii) “Glättung” des Extrakts, so dass ein kohärenter Text entsteht.

Schritt 1: In unserer Implementierung gehen derzeit folgende Faktoren in die Extraktbestimmung ein: Satzrelevanzen, berechnet auf der Grundlage von Termrelevanzen (Gewichtung nach  $tf*idf$ : Terme mit einer besonders hohen Frequenz im Text, aber niedriger allgemeiner Frequenz werden als relevant gewertet); Textsegmente und Segmentthemen; Dokumentstruktur (z.B. Position im Text, Layout). Als nächstes soll in unserer Implementierung auch der Textaufbau in Abhängigkeit von der Textsorte erkannt und mit berücksichtigt werden (z.B. in der Textsorte “Produktbeurteilung”: Bestimmung der Abschnitte allgemeine Produkt-Information vs. Produkt-Beurteilung).

Schritt 2: Häufig ergeben sich aus Schritt 1 nicht kohärente Texte, die z.B. unaufgelöste Pronomen enthalten. Viele Systeme kompensieren solche Probleme, indem sie Pronomen im Extrakt vermeiden, d.h. Sätze, die (einige/viele) Pronomen enthalten, werden nicht ausgewählt. Alternativ werden die Vorgängersätze zu den Sätzen mit problematischen Pronomen in den Extrakt aufgenommen. Beide Ansätze können nicht befriedigen, da im ersten Fall wichtige Sätze verloren gehen können, im zweiten Fall das Extrakt länger wird.

Im Gegensatz dazu setzen wir in unserem Verfahren auf eine intelligente Koreferenzanalyse: unaufgelöste Pronomen im Extrakt können damit in vielen Fällen durch ihr Antezedens ersetzt werden. In ähnlicher Weise wollen wir z.B. auch definite NPs oder Diskursmarker, die die Kohärenz des Extrakts beeinträchtigen, geeignet modifizieren.

Zur Veranschaulichung unseres Ansatzes werden wir ein Tool vorführen, das verschiedene Analyseschritte und –ergebnisse unserer Implementierung darstellt und den Einfluss der verschiedenen Faktoren visualisiert.

---

Felix Bildhauer  
Universität Bremen  
**Pragmatische Information in einer HPSG-Grammatik**  
Do 12.30-14.00

---

Das Poster dokumentiert die Implementierung eines HPSG-Fragments im TRALE-System. Das Fragment basiert auf Engdahl & Vallduví (1996) Ansatz zur Repräsentation von Informationsstruktur, vermeidet dabei jedoch einige Probleme der ursprünglichen Formalisierung.

Engdahl & Vallduví (1996) Vorschlag geht auf die Theorie in Vallduví (1992) zurück und wird bei Engdahl (1999) leicht verändert wieder aufgenommen. Problematisch daran ist u.a., dass die Informationsstruktur von Phrasen davon abhängig gemacht wird, ob bestimmte Merkmal-Wert-Paare der Tochterkonstituenten instantiiert sind oder nicht. Wenn man Merkmalstrukturen als Modelle für sprachliche Objekte versteht, sollten diese jedoch maximal spezifisch sein (d.h. Merkmale können nicht nicht-instantiiert bleiben). Darüber hinaus werden unterschiedliche "Verpackungen" von Information (*information packaging*) als unterschiedliche Instruktionen zur Aktualisierung des Wissenstandes (*knowledge-store*) eines Hörers interpretiert. Ohne zusätzliche Beschränkungen macht die Formalisierung jedoch falsche Voraussagen über Prosodie und mögliche Aktualisierungs-Instruktionen.

In der hier vorgestellten Implementierung treten diese und andere Probleme der Formalisierung bei Engdahl & Vallduví nicht auf. Das Grammatikfragment kann "live" getestet werden (Demo).

- Engdahl, E. & Vallduví, E. 1996. "Information packaging in HPSG", In: C. Grover & E. Vallduví (hg.), *Studies in HPSG*, Edinburgh, Centre for Cognitive Science, University of Edinburgh, 1-31.
- Engdahl, E. 1999. "Integrating pragmatics into the grammar". In: L. Mereu (hg.) *Boundaries of Morphology and Syntax*, Amsterdam, Benjamins, 175-194.
- Haji-Abdolhosseini, M. 2003. "A constraint based approach to information structure and prosody correspondance". In: St. Müller (hg.) *Proceedings of the HPSG03 Conference, Michigan State University, East Lansing*, Stanford, CSLI (<http://csli-publications.stanford.edu/HPSG/4/haji-abdolhosseini.pdf> 14.12.2005)
- Klein, E. 2000. "Prosodic constituency in HPSG". In: R. Cann, C. Grover & Ph. Miller (hg.), *Grammatical interfaces in HPSG*, Stanford, CSLI.
- Kuhn, J. 1995. *Information Packaging in German. Some motivation from HPSG-based translation*, ms. University of Edinburgh (<ftp://ftp.ims.uni-stuttgart.de/pub/papers/kuhn/Info-Pack.pdf> 13.12.2005).
- Pollard, C. & Sag, I.A. 1994. *Head-Driven Phrase Structure Grammar*, Chicago, University of Chicago Press.
- Vallduví, E. 1992. *The informational component*, New York, Garland.



---

Caroline Clemens, Thomas Hempel  
Universität Bremen, Technische Universität Berlin/Siemens AG  
**Was verraten die Log-Dateien telefonischer Sprachdialogsysteme über  
die Benutzer?**  
Do 12.30-14.00

---

Telefonische Sprachdialogsysteme ermöglichen einen mündlichen Dialog zwischen Mensch und Computer am Telefon. Der Benutzer vollzieht seine Eingaben entweder per Tastendruck im sog. Mehrfrequenz-Wahlverfahren (Touchtone) oder in gesprochener Sprache, die von einem automatischen Spracherkennungssystem verarbeitet wird. Das System spielt als akustische Ausgabe eine Audio-Aufzeichnung ab oder generiert aus einem geschriebenen Text synthetische Sprache (sog. Text to Speech).

Log-Dateien werden üblicherweise verwendet, um den Ablauf eines Computerprogramms automatisch zu dokumentieren. Bei der Programmierung wird festgelegt, bei welchen Programmschritten ein Eintrag in die Log-Datei erfolgt.

Die Log-Dateien von Sprachdialogsystemen können mit genauer Zeitangabe Ereignisse in einem Sprachdialog erfassen. Beispielsweise wird festgehalten, ob eine Eingabe vom Spracherkennungssystem erkannt wird, welchen Menüpunkt der Benutzer wählt oder wann der Benutzer durch seine Eingabe die Ausgabe des Systems vorzeitig unterbricht. Zusätzlich kann das Spracherkennungssystem Informationen liefern, welche Wörter erkannt wurden und mit welcher Konfidenz.

Nach einer Analyse der Log-Dateien kann der Benutzer einer bestimmten Benutzergruppe zugeordnet werden und das Sprachdialogsystem kann entsprechend reagieren. Im Dialogdesign können Dialog-Architektur, Audiodesign, Grammatiken etc. auf die Benutzergruppe abgestimmt werden.

Kontakt:

Caroline Clemens, Technische Universität Berlin, Zentrum Mensch-Maschine-Systeme, Jebensstrasse 1, 10623 Berlin, Graduiertenkolleg prometei, Tel.: 030-314-29634, [ccl@zmms.tu-berlin.de](mailto:ccl@zmms.tu-berlin.de).

Dr. Thomas Hempel, Siemens AG, Corporate Technology – User Interface Design, Otto-Hahn-Ring 6, 81739 München., Tel.: 089-636-48962, [thomas.hempel@siemens.com](mailto:thomas.hempel@siemens.com).

---

Thierry Declerck (mit Thorsten Trippel und Ulrich Heid)  
DFKI (Universität Bielefeld, Universität Stuttgart)  
**Standardisierungsarbeit für Sprachressourcen bei DIN (NAT AA 6)**  
Do 12.30-14.00

---

Wir berichten über internationale Normungsarbeit im Bereich von Sprachressourcen. Die Normen werden von internationalen Arbeitsgruppen im Rahmen der International Organization for Standardization (ISO), entwickelt und jeweils national von entsprechenden Gruppen, in Deutschland koordiniert vom Deutschen Institut für Normung (DIN), begleitet und diskutiert.

Für die automatische Sprachverarbeitung besteht seit Jahren zunehmend Bedarf an elektronischen Ressourcen: Lexika, Korpora, Grammatiken, Annotationskonventionen, Sprachdatensammlungen, usw. Damit solche Ressourcen über einen einzelnen Anwendungskontext hinaus wiederverwertbar sind und zwischen Arbeitsgruppen ausgetauscht werden können, wird an einer Normung ihrer Repräsentationsformate und der zur Beschreibung von Ressourceninhalten benutzbaren Vokabularien (Datenkategorien). gearbeitet

Waren in der Vergangenheit Standardisierungsbemühungen auf bestimmte Ausschnitte aus dem Spektrum der linguistischen Beschreibungen von Ressourcen beschränkt (z.B. die EU-Projekte SAM im Bereich gesprochener Sprache, EAGLES und ISLE im Bereich von Morphosyntax, Syntax, lexikalischer Semantik in Texten und Lexika und Sprachtechnologie), so ist die Zielsetzung der 2002 und 2003 gegründeten ISO (TC 37 SC 4) bzw. DIN (NAT AA 6) Arbeitsgruppen breiter: es geht zum Beispiel um Metarichtlinien für die Repräsentation und Annotation von Texten ebenso wie um Datenkategorien für Lexika, morphologische und morphosyntaktische, syntaktische Analyse.

Wir zeigen in diesem Poster den aktuellen Stand der Normungsdiskussion, wobei auch den Weg zu einer ISO Norm beschrieben wird. Das Poster soll auch als Einladung an Experten in Deutschland verstanden werden, sich in die Diskussion um Normen einzubringen.

- A. Abeillé, S. Hansen-Schirra, and H. Uszkoreit, editors. 2003. *Proceedings of the 4th International Workshop on Linguistically Interpreted Corpora (LINC-03)*, Budapest.
- Sue Atkins, Nuria Bel, Francesca Bertagna, Pierrette Bouillon, Nicoletta Calzolari, Christiane Fellbaum, Ralph Grishman, Alessandro Lenci, Catherine MacLeod, Martha Palmer, Gregor Thurmair, Marta Villegas, and Antonio Zampolli. 2002. *From resources to applications. Designing the multilingual ISLE lexical entry*. In LREC 2002, pages 687-693, Las Palmas de Gran Canaria.
- Sue Atkins, Nuria Bel, Pierrette Bouillon, Thatsanee Charoenporn, Dafydd Gibbon, Ralph Grishman, Chu-Ren Huang, Asanee Kawtrakul, Nancy Ide, Hae-Yun Lee, Paul J. K. Li, Jock McNaught, Jan Odijk, Martha Palmer, Valeria Quochi, Ruth Reeves, Dipti Misra Sharma, Virach Sornlertlamvanich, Takenobu Tokunaga, Gregor Thurmair, Marta Villegas, Antonio Zampolli, and Elizabeth Zeiton. 2003. *Standards and best*

*practice for multilingual computational lexicons and MILE (the multilingual ISLE lexical entry)*. Deliverable D2.2-D3.2 ISLE computational lexicon working group, International Standards for Language Engineering (ISLE), Pisa.

Das ISLE/EAGLES Projekt: [http://www.ilc.cnr.it/EAGLES96/isle/ISLE\\_Home\\_Page.htm](http://www.ilc.cnr.it/EAGLES96/isle/ISLE_Home_Page.htm)

ISO TC37/SC4 on Language Resources Management: <http://www.tc37sc4.org/>

DIN Normenausschuss Terminologie (NAT):

[http://www.nat.din.de/index.php?lang=de&na\\_id=nat](http://www.nat.din.de/index.php?lang=de&na_id=nat)

Das LIRICS Projekt: <http://lirics.loria.fr/>

---

Kurt Eberle  
lingenio

### Von LMT zu translate

## Entwicklungsschritte vom IBM MÜ-Projekt zum aktuellen Lingenio-Produkt

Do 12.30-14.00

---

In einer Installation der neuesten Version des Übersetzungssystems translate und einer korrespondierenden Entwicklungsumgebung, die es erlaubt, sowohl Ergebnisse der verschiedenen Analyse- und Generierungskomponenten zu illustrieren, als auch den Arbeitsplatz des Lexikographen, werden anhand einschlägiger Beispiele einige grundlegende Konzepte von LMT skizziert und Leitlinien der Weiterentwicklung über Personal Translator zur aktuellen Version von translate aufgezeigt.

Besonderes Augenmerk wird auf das Lexikon gelegt, dessen formale Grundlagen und Strukturierungsoptionen in einem begleitenden Vortrag in der AG 13 erläutert werden.

Ausgehend von den dort diskutierten Beispielen wird demonstriert, in welcher Weise alternative Strukturierungen von mehrdeutigen Einträgen mit unterschiedlichen Übersetzungen sich auf das Verhalten des Gesamtsystems auswirken.

In diesem Zusammenhang wird auch die vom Land Baden-Württemberg als herausragendes Forschungsergebnis ausgezeichnete IntelliDict-Technologie der Lingenio-office-Wörterbücher vorgestellt, die es erlaubt, Wörter in ihrer kontextuellen Bedeutung nachzuschlagen und diskontinuierliche Formen verschiedener Art zu erkennen.

Bläser, B. 1993. „TransLexis: An Integrated Environment for Lexicon and Terminology Management” In: P. Steffens (ed) *Machine Translation and the Lexicon*, Third International EAMT Workshop, Heidelberg, April 26-28, 1993, *Proceedings. Lecture Notes in Computer Science* 898 Springer 1995.

Eberle, K. 2001. „FUDR-based MT, head switching and the lexicon.” In: *Proceedings of the eighth Machine Translation Summit*. Santiago de Compostela.

McCord, M. 1989. „*Design of LMT*.” In: *Computational Linguistics* 15

Storrer, A., Schwall, U. 1993. “Description and Acquisition of Multiword Lexemes In: P. Steffens (ed) *Machine Translation and the Lexicon*, Third International EAMT Workshop, Heidelberg, April 26-28, 1993, *Proceedings. Lecture Notes in Computer Science* 898 Springer 1995.

<http://www.lingenio.de>. *translate* and *office dictionary*

---

Stefan Evert  
Universität Osnabrück  
**Empirische Kollokationsstudien mit dem UCS-Toolkit**  
Do 12.30-14.00

---

Lexikalisierte Wortverbindungen – für die hier wie in der Computerlinguistik gebräuchlich der Begriff *Kollokationen* im weitesten Sinne verwendet wird (vgl. Manning & Schütze 1999, Kap. 5) – stellen sowohl für die linguistische Forschung als auch für computerlinguistische Anwendungen eine große Herausforderung dar. Für erstere, weil sie die spätestens seit Chomsky (1957) vorherrschende strenge Trennung zwischen Syntax und Lexikon sprengen; für letztere, weil sie aufgrund ihrer nicht ohne weiteres vorhersagbaren idiosynkratischen Eigenschaften in sprachverarbeitenden Systemen gesondert behandelt werden müssen.

Ein wesentlicher Aspekt von Kollokationen ist ihre Rekurrenz (d.h. ein unerwartet häufiges Miteinander-Vorkommen der beteiligten Wörter), auch wenn phraseologische Definitionen diese in der Regel nicht mit einbeziehen, sondern sich auf die Nichtkompositionalität, eingeschränkte Modifizierbarkeit und ähnliche idiosynkratische Eigenschaften einer Wortkombination stützen. Rekurrenz ist im Gegensatz zu letzteren eine objektiv beobachtbare Größe, lässt sich durch Anwendung statistischer Assoziationsmaße auf Korpusdaten quantitativ erfassen und stellt daher ein wichtiges Hilfsmittel zur Identifikation von Kollokationen dar. Die mittels statistischer Methoden gewonnenen Kollokationskandidaten dienen als empirisches Basismaterial für linguistische Studien oder finden nach manueller Überprüfung Eingang in traditionelle und maschinenlesbare Wörterbücher.

In der Literatur findet sich eine große Anzahl von Assoziationsmaßen, die auf völlig unterschiedlichen, aber meist plausiblen mathematischen Ansätzen beruhen. Dabei ist kein Maß in Sicht, das für alle Anwendungen gleichermaßen geeignet wäre (siehe z.B. Evert & Krenn 2005). Das im Rahmen einer Dissertation (Evert 2004) entwickelte UCS-Toolkit stellt Funktionen zur Verfügung, mit denen die mathematischen und „linguistischen“ Eigenschaften von Assoziationsmaßen empirisch untersucht und Unterschiede zwischen verschiedenen Maßen herausgearbeitet werden können. Zum Einsatz kommen dabei Simulationsexperimente, Visualisierungsmethoden und die anwendungsorientierte Evaluation von Verfahren zur Kollokationsextraktion

Chomsky, N. 1957. *Syntactic Structures*. Den Haag: Mouton.

Evert, S. 2004 (veröffentlicht 2005). *The Statistics of Word Cooccurrences: Word Pairs and Collocations*. Dissertation Stuttgart. URN urn:nbn:de:bsz:93-opus-23714. Siehe <http://www.collocations.de/phd.html>.

Evert, S. & Krenn, B. 2005. „Using small random samples for the manual evaluation of statistical association measures.“ *Computer Speech and Language* 19(4), 450-466.

Manning, C. & Schütze, H. 1999. *Foundations of Statistical Natural Language Processing*. MIT Press.

---

Aurélien Giraud  
Universität Bremen  
**Exploring a New Language Using a Machine Translation System**  
Do 12.30-14.00

---

Can the theoretical information encoded in MT systems explain by itself some grammatical concepts and help in building CALL applications, provided that this information is coupled with a suitable display utility? Or do some kind of explanations also have to be encoded beside the theoretical linguistic information? As said in Reuer, 2004: "The learner is invited to explore the language he intends to learn." The use of computers as explorative devices can help learners of a new language, and it is exactly what this research focuses on: using implemented grammars and MT systems as a tool to explore some characteristics of the language. The learning situation in consideration requires an MT system, which translates from the learner's mother language to the 'target' language being explored. Miscellaneous phenomena, which typically cause trouble to second language learners are considered.

First, a visualisation tool is used. It allows to display some relevant linguistic information in a 'self-explaining' way, which does not require the more complex automatic generation of verbal explanations, that may, depending on the learners, have to be understandable even to people not familiar to grammatical concepts. The information displayed comes either only from the grammar of the language being learned, or from the comparative linguistic information between the two languages, that is encoded in the translating system.

Second, a way of browsing the information in the system is investigated. Learners are sometimes able to identify the structures or words that cause them difficulties. Therefore, it may be useful to them to be able to query the system relative to these difficulties. The query may be a word or a structure, that can be described in non theoretical ways. The answer to a query may be examples of aligned pairs of sentences in mother and target language, together with some more relevant information displayed (e.g. thematic roles or head switching). This could also for instance provide a simple way to browse idiomatic structures.

Last the pros and cons of the approach formerly described will be briefly discussed, compared for example to the use of tagged corpora.

Dorr, B. J., Jordan, P.W. and Benoit, W. 1998. „A survey of current paradigms in machine translation“. Technical Report CS-TR-3961.

Reuer, V. 2004. „Language processing and intelligent computer-assisted language learning“. Technical Report, Institute of Cognitive Science, University of Osnabrück.

Trujillo, A. 1999. *Translation Engines: Techniques for Machine Translation*. Berlin: Springer.

---

Jürgen Handke  
Marburg  
**The Virtual Linguistics Campus**  
**([www.linguistics-online.com](http://www.linguistics-online.com))**  
Do: 12.30 – 14.00

---

## Systemvorführung.

---

Laxmi Prasad Khatiwada  
MPP, Nepal  
**Experiences of Localization in Nepali**  
Do 12.30-14.00

---

“Experience of Nepalinix Development” presentation contains following criteria

### 1. Background

ICT has developed a new methodology in linguistic field. Localization aspect of language, especially local language computing has become an inspiring discipline in developing countries.

### 2. Challenges

- Lack of Localization Policy
- Lexicalization and Institutionalization
  - Voluntary work and Authenticity

These are the main challenges of Nepali localization. The convention of standardization has not built yet in central level. We have to face many problems due to this.

### 3. Problem faced in Project development

- Lack of Institutional Existence in Nepal in NLP: Nepal has a rich linguistic culture but we have no technical infrastructure.
- Arrangement of Investment: Financial resource is also a major factor for developing countries like us. This made a little bit inconvenience in the following sector, especially to hire or manage the expert HRD.

- a. Investigation
  - ✕☞ Data gathering
  - ✕✕☞ Fact finding
- b. Analysis
- c. System design

4. Language Localization Requirements: These are the main criteria to develop a human machine interface in once own language. So first we developed these sets.

5. Works Accomplished by Madan Puraskar Pustakalaya.

Bootable CD: I will show some snapshots about the production listed below.

- Input Methods
- Interface
- User Application: Web Browser

---

Mareile Knees  
Universität Jena  
**Annotation und Analyse temporal-anaphorischer  
Bezüge am Beispiel von *danach*<sup>1</sup>**  
Do 12.30-14.00

---

Pronominaladverbien wie *danach* und *dadurch* bestehen aus einem anaphorischen Element (z.B. *da-*) und einem relationalen Element (z.B. *-nach* oder *-durch*) (s. Fraurud 1992, Rüttenauer 1978). Anhand des anaphorischen Elements verweisen sie auf einen Referenten, der zuvor von einem Antezedenten ins Diskursmodell eingeführt wurde. Darüber hinaus stellen sie eine temporale, kausale oder sonstige diskursrelationale Beziehung zwischen dem Referenten des Antezedenten und dem Referenten des Satzes, in dem sie auftreten, her. Ich bezeichne sie daher als relationale Anaphern, die auch als Diskurskonnektoren fungieren können (Webber et al. 2003, Miltsakaki et al. 2003).

Um herauszufinden, wie die Auflösung der relationalen Anapher *danach* funktioniert, werden ausgehend von einer Korpusstudie möglichst viele verschiedene Varianten von *danach*-Vorkommen betrachtet und klassifiziert. Es ist anzunehmen, dass bei den einzelnen Varianten unterschiedliche Auflösungsmechanismen eine Rolle spielen. Mich interessieren die der Anapherauflösung zugrundeliegenden Mechanismen. Daher habe ich ein Annotationsschema entwickelt, das sich an bereits erprobten Schemata (vor allem Müller & Strubes (2001) und Poesios (2004)) orientiert. Mit Hilfe des Schemas werden nicht nur verschiedene Varianten von *danach* untersucht, sondern auch diejenigen Faktoren dokumentiert, die die Anapherauflösung beeinflussen. Ich gehe davon aus, dass bei der Anaphernrezeption ganz unterschiedliche Wissensressourcen (syntaktisches, semantisches, diskurs-stukturelles und auch Welt-Wissen) beteiligt sind (vgl. Schwarz 2000, Consten 2004).

Abschließend wird zu prüfen sein, ob und in welchem Umfang sich die Mechanismen, die dem hier untersuchten Typ der relationalen Anaphorik zugrunde liegen, so formalisieren lassen, dass sie sowohl für die Textrezeptionsforschung als auch für die maschinelle Sprachverarbeitung gewinnbringend nutzbar sind.

---

<sup>1</sup> Das Poster ist im Rahmen des DFG-Projekts „KomplexTex“ entstanden.

## Literatur

- Consten, M. 2004. *Anaphorisch oder deiktisch? Zu einem integrativen Modell domänengebundener Referenz*. Tübingen: Niemeyer.
- Fraurud, K. 1992. Situation Reference. What does ‚it‘ refer to?. In: K. Fraurud. *Processing Noun Phrases in Natural Discourse*. PhD thesis. Depart. of Linguistics, Stockholm University.
- Rüttenauer, M. 1978. *Vorkommen und Verwendung der adverbialen Proformen im Deutschen*. Hamburg.
- Miltsakaki, E., C. Creswell, K. Forbes, A. Joshi & B. Webber 2003. Anaphoric Arguments of Discourse Connectives: Semantic Properties of Antecedents versus Non-antecedents. *Proc. of the Computational Treatment of Anaphora Workshop, EACL 2003*, Budapest.
- Müller, C. & M. Strube 2001. Annotating Anaphoric and Bridging Relations with MMAX. *Proc. of the 2<sup>nd</sup> SIGdial Workshop on Discourse and Dialogue*. Aalborg, Denmark, September, 90-95.
- Poesio, M. 2004. The MATE/GNOME Scheme for Anaphoric Annotation, Revisited. *Proc. of SIGDIAL*, Boston, April.
- Schwarz, M., 2000. *Indirekte Anaphern in Texten. Studien zur domänengebundenen Referenz und Kohärenz im Deutschen*. Tübingen: Niemeyer.
- Webber, B., A. Joshi, M. Stone, and A. Knott 2003. Anaphora and Discourse Structure. *Computational Linguistics*, 29(4), 545-587.
- 

Birte Lönneker-Rodman  
Universität Hamburg

**Modelle erzählender Texte - Computerlinguistik und  
„Computernarratologie“**  
Do 12.30-14.00

---

Erzählende Texte bergen vielfältige Möglichkeiten der Ausgestaltung auf Diskurs-Ebene. Mit den Zusammenhängen zwischen der abstrakten Inhaltsdomäne einer Erzählung und ihrer direkt zugänglichen Erscheinungsform in einem Text beschäftigt sich die analytische Methode der Narratologie. Unter Bezug auf traditionell strukturalistische Modelle fiktionaler und nicht-fiktionaler Erzählprodukte unterschiedlicher Genres beantwortet sie Fragen wie: In welchem Zusammenhang stehen die Sequenz des Inhalts und die Sequenz seiner Darstellung im Text – werden Ereignisse chronologisch präsentiert oder gibt es zeitliche Umstellungen? In welchem Zusammenhang steht die Wiederholffrequenz von (ähnlichen) Ereignissen im dargestellten Geschehen und im Text? Wie wird die Darstellung durch einen Erzähler vermittelt und perspektiviert?

Zu den ersten Verwendern des Terminus *Computational Narratology* zählt Takashi Ogata, der narratologische Beschreibungen wie (Genette 1994) mit dem Ziel der automatischen Produktion von Erzählwerken operationalisieren möchte (vgl. Ogata 2004). Auch Cavazza et al. (2001) sehen Computernarratologie als ein Instrument an, das – obwohl aus analytischer Tradition hervorgegangen – bei der Generierung (hier von narrativen interaktiven Spielen) eingesetzt werden kann.

Eine Auswertung bestehender Systeme zur Generierung von Erzählungen (*Story Generation*) und von Texten in natürlicher Sprache (*Natural Language Generation*)



zeigt, dass in der Forschung zur Künstlichen Intelligenz und Computerlinguistik viele in der Narratologie thematisierte Besonderheiten der Diskurs-Ebene erzählender Texte bisher kaum berücksichtigt wurden. Dies legt als aktuelle Hauptaufgabe von Computernarratologie die Formalisierung narratologischer Begriffe nahe, die in generierende Systeme vor allem auf der Stufe der Diskurs-Strukturierung eingehen sollten. Am Beispiel „gerahmter“ Erzählungen mit möglicherweise wechselnden Erzählern werden derartige Formalisierungen in (Lönneker 2005) ausführlich dargestellt und in den Kontext von NLG-Systemarchitekturen integriert.

Cavazza, M., Charles, F. und Mead, S. J. (2001): „Narrative Representations and Causality in Character-Based Interactive Storytelling.“ In *Proceedings of cast01, Living in Mixed Realities, Bonn, Germany, September 2001*. 139–142.

Genette, Gérard (1994): *Die Erzählung*. München: Fink.

Lönneker, B. (2005): „Narratological Knowledge for Natural Language Generation.“ In *Proceedings of the 10th European Workshop on Natural Language Generation, Aberdeen, Scotland, August 2005*. 91–100.

Ogata, T. (2004): „A Computational Approach to Literary and Narrative Production: Toward Computational Narratology.“ In *Art and Science. Proceedings of the 18th Congress of the International Association of Empirical Aesthetics, Lisbon, September 2004*. 509–516.

---

Moritz Neugebauer  
Universität zu Köln  
**Syllable Constituent Tagging with SBS**  
Do 12.30-14.00

---

Syllabification is generally known as the assignment of syllabic structure to a string of phonological segments. A hierarchical representation of the internal structure of a particular syllable then comprises a syllable onset followed by a rhyme which itself may be subdivided into an obligatory nucleus and a facultative syllable coda. Formally speaking, we are dealing with a directed acyclic graph (or *DAG*) whose non-terminal nodes are known as syllable constituents (Coleman 1998). A procedure which automatically derives such a graph for a given input string is called a syllabification algorithm.

According to Walther (1999), it is possible to define a syllabification algorithm which assigns syllabic structure to a string, provided a hierarchical ordering of all segments of a given language with respect to differences in their respective sonority. This approach has been dubbed sonority-based syllabification (*SBS*).

The poster will present an implementation of Walther's algorithm in ALE (Penn & Carpenter 2002) which is applied to the task of syllable constituent tagging of a German pronunciation lexicon. The aim of this study is to demonstrate that while the concept of sonority may be considered problematic with respect to its phonetic source, sonority-based segment hierarchies may still function as a practical tool to derive good quality syllabifications without any training data and little linguistic effort on the side of the user.

- Coleman, J. 1998. *Phonological representations*. Cambridge: Cambridge University Press (= Cambridge Studies in Linguistics 85).
- Penn, G. & Carpenter, B. 2001. *ALE (The Attribute Logic Engine) User's Guide. Version 3.2.1*. <http://www.cs.toronto.edu/~gpenn/ale.html>.
- Walther, M. 1999. *Deklarative Prosodische Morphologie. Constraint-basierte Analysen und Computermodelle zum Finnischen und Tigrinya*. Tübingen: Niemeyer (Linguistische Arbeiten 399).
- 

Bettina Schrader  
Universität Osnabrück

**Exploiting linguistic and statistical knowledge for a text alignment  
system**

Do 12.30-14.00

---

Aligning parallel text, i.e. automatically setting the sentences or words in one text into correspondence with their equivalents in its translation, is a very useful preprocessing step for a range of applications, including but not limited to machine translation (cf. Brown et al. 1993), cross-language information retrieval (cf. Hiemstra 1996), dictionary creation (cf. Smadja et al. 1996) and induction of NLP-tools (cf. Kuhn 2004). Additionally, aligned corpora can be used for research in contrastive linguistics: Researchers can look up specific expressions or syntactic constructions in one language, and explore, via the alignment information, how these are translated.

The alignment of sentences can be done with sufficiently high quality using cues such as sentence length (cf. Gale and Church, 1993) or cognates (cf. Simard et al 1992). Word alignment, however, is almost exclusively done using word frequency and co-occurrence statistics. These approaches typically are insensitive to linguistic structures but assume that sentences are bags of words (cf. Brown et al 1993, Hiemstra 1996). Effort is spent on improving word alignment techniques (cf. Mihalcea and Pedersen 2003, Cherry and Lin 2003, Toutanova et al. 2002), but the initial assumptions are rarely changed.

The statistical word alignment models do not suffice to reliably align so-called rare events, i.e. words that occur 5 times or less. This is a considerable drawback, as rare events make up more than half of the vocabulary of any corpus. Another problem is the correct alignment of multiword units like idioms. Then, differences in word order, although they can be predicted by existing grammars, are not modelled well by the statistical algorithms. Only recent approaches like those by Cherry and Lin (2003), have begun to incorporate explicit syntactic information to guide the alignment process. Finally, current approaches are ill-suited to cope with multiple layers of linguistic information: they can hardly be extended so that they make efficient use of information on word forms, parts of speech, lemmas, and syntactic constituency.

Hence, we are developing a hybrid text alignment system that can operate on different levels of linguistic description: it aligns a linguistically annotated corpus using cues such as word category, word formation patterns, and syntactic constituency, but also takes statistic information into account.

Our system scans a parallel corpus for reliable alignment cues, e.g. within the

parts of speech annotation, indicating that two corpus units are probably translations of each other. When such cues are found, they are used to align the corresponding corpus units. These may be sentences or words and, if given by the corpus annotation, paragraphs or sub-sentential units like phrases or chunks. These beads are re-used: word beads are added to a system-internal dictionary and serve to determine further word or sentence beads. Sentence beads are used to restrict the contexts to compute word alignments. A final filtering step ensures that the resulting text alignment is consistent and that wrong beads are discarded. It also adds 1:0 or 0:1 beads for those items that were impossible to align.

Presently, we have begun to carry out an evaluation of our aligner, and we are constantly adding and testing new alignment strategies.

- Brown, P.F. et al. 1993: *The Mathematics of Machine Translation: Parameter Estimation*. Computational Linguistics 19(2), pp. 263-311
- Cherry, C. and Lin, D. 2003: *A Probability Model to Improve Word Alignment*. Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics, 88-95
- Gale, W. A. and Church, K. W. 1993: *A program for aligning sentences in bilingual corpora*. Computational Linguistics 19(1), 75-102
- Hiemstra, D. 1996: *Using Statistical Methods to Create a Bilingual Dictionary*. Master's Thesis, Universiteit Twente
- Kuhn, J. 2004: *Exploiting Parallel Corpora for monolingual Grammar Induction -- a pilot study*. Workshop proceedings of the 4th International Conference on Language Resources and Evaluation (LREC), 54-57
- Mihalcea, R. and Pedersen, T. 2003: *An Evaluation Exercise for Word Alignment*. HLT-NAACL 2003 Workshop: Building and Using parallel Texts. Data Driven Machine Translation and Beyond, pp. 1-10
- Simard, M. and Foster, G. F. and Isabelle, P. 1992: *Using cognates to align sentences in bilingual corpora*. Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine translation, 67-81
- Smadja, F. et al. 1996: *Translating Collocations for Bilingual Lexicons: A Statistical Approach*. Computational Linguistics 22(1), 1-38
- Toutanova, K. et al. 2002: *Extensions to HMM-based Statistical Word Alignment Models*. Conference on Empirical Methods in Natural Language Processing (EMNLP 2002), pp. 87-94

---

Caroline Sporleder  
Tilburg University  
**Cleaning and Enriching Textual Databases from the Natural History  
Domain**  
Do 12.30-14.00

---

Information about cultural heritage collections is often stored in textual or semi-textual databases. Such databases are a potentially very valuable resource for researchers in the field, especially if the contents can be systematically searched and queried. Unfortunately, information retrieval from a database can be adversely affected by a variety of factors, such as errors and inconsistencies in the data or a suboptimal structure of the database itself. However, methods from data mining and natural language

processing can help to detect and correct errors and inconsistencies, and to add meta-data (such as named entity tags) in order to aid information retrieval.

The poster presents research carried out as part of the MITCH project (Mining for Information in Texts from the Cultural Heritage), a joint research project between Tilburg University and Naturalis, the Dutch National Museum of Natural History. The ultimate goal of the MITCH project is to develop techniques to discover new knowledge in databases and (recently digitised) fieldbooks describing Naturalis's vast collection of animal and plant specimens.

As a first step towards this goal, we investigated various methods to automatically clean up and enrich a manually created database. The database contains over 16,000 records describing individual amphibian and reptile specimens and the circumstances of their collection (e.g., who collected them, where and when they were found etc.). Many fields are textual, others are numeric. The database is multi-lingual; mostly Dutch and English are used (and Latin for taxonomic names) but some fields also contain Portuguese, French or German. As there is no restriction on which languages can be entered into the database, we had to develop language independent methods for data clean-up and enrichment. To ensure the general applicability of our techniques, we also did not want to feed in too much background knowledge. Instead, we exploited the structure of the database itself to:

- identify information that was entered in the wrong field
- detect and (semi-)automatically correct content errors (e.g., wrong taxonomic labels)
- bootstrap a named entity tagger

(The research presented here is joint work with Marieke van Erp, Tijn Porcelijn, and Antal van den Bosch. It was funded by NWO, the Netherlands Organisation for Scientific Research, and carried out at the Naturalis Research Labs in Leiden.)

## Verzeichnis der Vortragenden

Vorname	Name	AG
Helen	Aristar-Dry	6
Maria	Averintseva	8
Jens	Baitinger	13
Dik	Bakker	1
Eduard	Barb	13
Verginica	Barbu Mititelu	13
Dominik	Bauer	Stud
Nicole	Baumgarten	8
Kathrin	Beck	CL
Michael	Beißwenger	13
Emily M.	Bender	1
Peter	Berck	6
Eva	Berlage	7
Judith	Berman	7
Hans-Jörg	Bibiko	6
Balthasar	Bickel	4
Heike	Bieler	CL
Chris	Biemann	13
Felix	Bildhauer	CL
Walter	Bisang	1
Marion	Blondel	3
Hardarik	Blühdorn	8
Christiane	Bongartz	10
Antal	Bosch, van den	CL
Raymond	Boyd	1
Angelika	Braun	9
Daan	Broeder	6
Monika	Budde	2
Aljoscha	Burchardt	13
Isolde	Burr	9
Jan	Burse	13
Heidi	Byrnes	10
Luis	Cabral	13
Marcus	Callies	10
Bert	Cappelle	7
Mary	Carroll	8
Faye	Chalcraft	7

Ralph	Christensen	9
Andreas	Claus	6
Caroline	Clemens	CL
Christelle	Cosme	8
Onno	Crasborn	3
Mily	Crevels	4
Sonia	Cristofaro	1
Berthold	Crysmann	2
Péter	Csatár	11
Michael	Cysouw	4
Franz Josef	d'Avis	5
Ernesto		
William	De Luca	13
Gert	De Sutter	7
Thierry	Declerck	12
Thierry	Declerck	6,CL
Nicole	Dehé	5
Matthias	Dehmer	12
Francesca	Del Gobbo	5
Laurence	Delort	8
Tanya	Denmark	3
Rainer	Dietrich	9
Alexis	Dimitiradis	6
Christine	Dimroth	10
Stefanie	Dipper	6, CL
Norbert	Dittmar	10
Anaid	Donabedian	1
Sandra	Döring	5
Scott	Drellishak	1
Sebastian	Drude	2
Andreas	Dufter	7
Kurt	Eberle	13,CL
Sabine	Ehrhardt	9
Michael	Elmentaler	14
Martin	Elsig	7
Nick	Enfield	4
	Engberg-	
Elisabeth	Pedersen	3
Katrin	Erk	13
Marieke	Erp, van	CL

Stefan	Evert	CL
Thórhallur	Eythórsson	7
	Fabricius-	
Cathrine	Hansen	8
Marjan	Fanaei	1
Joseph T.	Farquharson	Stud
Christiane	Fellbaum	11
J.	Fenlon	3
Anita	Fetzer	5
Jürg	Fleischer	7
Dan	Flickinger	1
Christian	Fortmann	5
Anette	Frank	13
Michael	Franke	8
Matthias	Fritz	1
Nanna	Fuhrhop	2
Dirk	Geeraerts	7
Peter	Geibel	12
Ljudmila	Geist	7
Thomas	Gergen	9
Anna	Giacalone	4
Jost	Gippert	6
Aurélien	Giraud	CL
Rüdiger	Gleim	12
Rob	Goedemans	6
Aslı	Göksel	2
Jeff	Good	1
Maksym	Gorshenin	Stud
Michael	Götze	6
Irene	Greftegreff	3
Günther	Grewendorf	9
Patrick	Grommes	10
Tom	Güldemann	4
Iryna	Gurevych	13
Helmar	Gust	12
Ulrike	Gut	10
Stefanie	Haberzettl	10
Olaf	Hackstein	8
Willem		
Robert	Hage, van	12

Jürgen	Handke	CL
Corinna	Handschuh	Stud
Thomas	Hanke	Stud
Martje	Hansen	3
Sandra	Hansen	9
Silvia	Hansen-Schirra	6
Verena	Haser	7
Martin	Haspelmath	1
Katharina	Haude	1
Ulrich	Heid	11,CL
Johannes	Helmbrecht	4
Thomas	Hempel	CL
Kristrina	Henschke	Stud
Jens	Heßmann	3
Kris	Heylen	7
Kristine	Hildebrandt	4
Julia	Hillebrand	9
Ludger	Hoffmann	9
Anke	Holler	8
Matthias	Hölzner	13
Peter E.	Hook	4
Harald	Huber	12
Dmitry	Idiatov	4
Hitoshi	Isahara	11
Michael	Jessen	9
Trevor	Johnston	3
Jóhannes		
Gísli	Jónsson	7
Georg A.	Kaiser	7
Gunther	Kaltenböck	5
Sophia	Katrenko	12
Yordanka	Kavalova	5
Roland	Kehrein	14
Marc	Kemps-Snijders	6
Katja	Keßelmeier	11
Mirjam	Kessler	6, CL
Olesya	Khanina	1
Kathrin	Kirsch	6
Tibor	Kiss	11
Tanja	Kiziak	5



Katarina	Klein	7
Gerson	Klumpp	7
Mareile	Knees	CL
Els	Kooij, van der Koptjevskaja-	3
Masha	Tamm	1
Bernd	Kortmann	7
Andrea	Kowalski	11,13
Kai-Uwe	Kühnberger	12
Claudia	Kunze	13
Kow	Kuroda	11
Stephen	Laker	14
Lothar	Lemnitzer	13
Kent	Lerch	9
Silke	Lewinski, von	9
Tim	Lichte	12
Hans- Heinrich	Lieb	2
Eva	Lippert	10
Rudy	Loock	5
	Lönneker-	
Birte	Rodman	CL
Andreas	Lötscher	9
Anke	Lüdeling	10
Rosemarie	Lühr	8
Petra	Ludewig	12
Karin	Luttermann	9
	Maden-	
Ursula	Weinberger	10
Yaron	Matras	4
Hiram H.	Maxim	10
Alexander	Mehler	12
Dejan	Milosavljevic	13
Uwe	Mönnich	12
Annie	Montaut	1
Anja	Moos	Stud
Claire	Moyse-Faurie	1
Joybrato	Mukherjee	10
Kristin	Mulroony	3
Katrin	Mutz	2

Pieter	Muysken	4
	Naghzguye	
Mehrdad	Kohan	1
Keiko	Nakamoto	11
David	Nathan	6
Martin	Neef	7
Moritz	Neugebauer	CL
Sara	Neuhauser	Stud
Stella	Neumann	9
Frederick	Newmeyer	1
Brenda	Nicodemus	3
	Noel Aziz	
Patrizia	Hanna	2
Andreas	Nolda	2
Rick	Nouwen	5
Hazim	Nozawa	11
Andreas	Nürnbergger	13
Freddy	Offenga	6
Tunde	Opeibi	9
Rainer	Osswald	13
Sebastian	Pado	13
Brigitte	Pakendorf	4
Prashant	Pardeshi	4
Jörg	Peters	5
Gergely	Pethö	11
Svetlana	Petrova	8
Christiane	Pilot-Raichoor	1
Manfred	Pinkal	11,13
Ana Sofia	Pinto	13
Shana	Poplack	7
Tijn	Porcelijn	CL
	Prasad	
Laxmi	Khatiwadaa	CL
Josep	Quer	3
Wiebke	Ramm	8
Monika	Rathert	9
Jochen	Rehbein	6
Ingo	Reich	8
Jürgen	Reischer	11
Antje	Roßdeutscher	8

Albert	Russel	6
Louisa	Sadler	2
Said	Sahel	7
Jeanette	Sakel	4
Natascha	Sakhonenko	10
Andrea	Sansò	4
Luis	Sarmento	13
Adam	Saulwick	6
Ron	Schaefer	1
Sabine	Schall	9
Adam	Schembri	3
René	Schiering	7
Florian	Schlegel	13
Natalie	Schluter	11
Monika S.	Schmid	6
Barbara	Schmiedtowá	10
Stefan	Schneider	5
Bettina	Schrader	CL
	Schulte im	
Sabine	Walde	13
Ulrike	Schultz	9
Nicole	Schumacher	10
Thomas M.	Seibert	9
Frank	Seifart	1
Guido	Seiler	7
Gary	Simons	6
Romuald	Skiba	6
Stavros	Skopeteas	1,6
Elissa	Sobotta	14
Jan-Philipp	Söhn	12
Michael	Solf	8
Kåre	Solfjeld	8
Dirk	Speelman	7
Helmut	Spiekermann	14
Caroline	Sporleder	CL
Steffen	Staab	12
Yvonne	Stakic	9
Harald	Stamm	2
Ekaterini	Stathi	11
Joyce	Stavick	8

Manfred	Stede	8, CL
Markus	Steinbach	5
Petra	Steiner	13
Barbara	Stiebels	1
Don	Stilo	4
Thomas	Stolz	4
Angelika	Storrer	13
Christiane	Stutterheim, von	8
Arthit	Suriyawongkul	CL
Josef	Szakos	1
Konrad	Szczesniak	10
Felix	Sze	3
Oliver	Teuber	2
Höskuldur	Thráinsson	7
Konstantin	Todorov	12
Doris	Tophinke	14
Paul	Trilsbeek	6
Thorsten	Trippel	6, CL
Jürgen	Trouvain	Stud
Christina	Unger	Stud
Gerd	Unruh	13
Mihaela	Vela	6,12
Josje	Verhagen	10
Marie- Hélène	Viguier	2
Alice	Vittrant	1
Anja	Voeste	14
Hein	Voort, van der	4
Uwe	Vosberg	7
Fetzer	Vries, de	5
Bernhard	Wälchli	1
Phillip	Wallage	7
Maik	Walter	10
Stephan	Walter	9
Joel	Walters	10
Heide	Wegener	10
Sandra	Wellinghoff	13
Søren	Wichmann	4
Björn	Wiemer	1
Bernd	Wiese	2

Richard	Wiese	7
Menzo	Windhouwer	6
Sara	Wischer	8
Peter	Wittenburg	6
Jan	Wohlgemuth	Stud
B.	Woll	3
Angelika	Wöllstein	8
Kai	Wörner	6
Ulrike	Wrobel	3
Arne	Zeschel	11
Ulrike	Zeshan	6
Evelyn	Ziegler	14
Vitor	Zimmerer	Stud
Heike	Zinsmeister	13,CL