

---

Markus Bader, Tanja Schmid & Josef Bayer

**An experimental perspective on coherent infinitives in German**

*Mittwoch/Wednesday: 17:30*

---

1. Introduction: Since Bech (1955/57) and Evers (1975), German (and Dutch) infinitival constructions have attracted enormous attention from syntacticians. Two questions have figured prominently over the years and across various changes in linguistic theory: (i) What are infinitival complements? (ii) What governs coherence? Answers given to (i) range from CP to IP to VP. Answers given to (ii) are usually rooted in certain assumptions about the lexical representation of verbs. Certain infinitive-taking verbs undergo a process of restructuring which results in clause union, whose core property is a verbal complex in West-Germanic. Other infinitive-taking verbs reject clause union. Since most of the verbs in question also allow extraposition, and since some even allow finite CP-complements, yet another question appears: (iii) are coherence verbs associated with multiple lexical entries? Most of the work so far rests on a somewhat fragile data base. Corpus studies as well as behavioral data could give important clues to a better understanding of the phenomenon and its place in grammar.

The present study launches an attempt in three steps: (A) A corpus study tries to determine which verbs appear in which frequency in which infinitive constructions. (B) A questionnaire study investigates a number of coherence-sensitive constructions across a wide spectrum of infinitive-taking verbs. (C) A self-paced reading study investigates one aspect of coherent constructions in detail: the scope of negation.

2. Corpus study: We conducted a corpus study based on the COSMAS System provided by the Institut für Deutsche Sprache, Mannheim, investigating all the verbs that were used in the questionnaire study reported below.

3. Questionnaire study: This study takes a close look at native speakers' intuitions with respect to various coherence-sensitive tests across 56 control verbs which can be grouped in 5 classes according to subcategorisation and control properties. 7 syntactic constructions were tested (embedded and matrix verb moved to SpecCP; "long" scrambling of a pronoun; "long-distance" passive; wide scope of negation; narrow scope of negation; extraposed; intraposed).

4. Selfpaced-Reading Study: Coherence was further investigated in a self-paced reading study of sentences requiring a wide scope construal of a negative quantifier. Subjects had to read sentences in which the infinitive contained a negative quantifier such as *keines von den Büchern* ("none of the books"), and which ended with the tag *und X auch nicht* ("... and neither did X"). This tag is only compatible with the wide scope of negation, which in turn is only possible when the embedded infinitival is constructed coherently.

Representative results of these studies are:

Coherence tests are valid as they can be shown to correlate, and intraposition but not extraposition or narrow scope of negation correlates with the coherence tests

A cluster analysis yields a separation of coherent and non-coherent verbs with a substantial but not complete overlap with the predefined verb classes

In the selfpaced reading study, reading times on the matrix verbs but not on the tag requiring wide scope correlate with the mean coherence measure derived from the questionnaire study.

In our presentation, we will discuss implications of these data, in particular concerning

- the relationship of corpus data to judgment data
- the relevance of graded judgment data for both theories of grammar and theories language processing

- the interaction of hierarchical structure, linear order, and verb-specific lexical information for determining “coherence”

---

Gisbert Fanselow & Stefan Frisch

**Effects of local ambiguity**

*Freitag/Friday: 13:30*

---

Processing difficulty is one of the factors that influences acceptability judgements. We will report the results of a series of experiments investigating the impact of a particular type of processing difficulty, viz., local ambiguities, on acceptability ratings for a wide range of data, including split noun phrases, agreement in coordinate constructions, long wh-movement, grammatical function ambiguities and particle movement. Except for domains in which the assignment of grammatical functions is directly involved, the presence of local ambiguities may make sentences appear more grammatical than they in fact are. We will relate this finding to corpus frequency data and try to account it in terms of theories of reanalysis in sentence processing.

---

Sam Featherston

**Judgements in syntax: Why they are good, how they can be better**

*Mittwoch/Wednesday: 14:00*

---

In this talk we shall report our investigations into the nature of judgements: the factors that they are sensitive to, their relationship to frequency data, and how best they are gathered, building on the work of Schütze (1996) and Keller (2000). We shall argue that judgements can be made “harder” evidence and they have certain advantages over other data types.

We have investigated this issue by looking at the same syntactic phenomena in two data types: experimentally obtained judgements and corpus frequencies. In a series of studies we have applied experimental methods (including magnitude estimation of well-formedness, Bard et al 1996) to syntactic issues but also to questions of the nature of grammaticality. We compare these results with corpus data (COSMAS, IDS Mannheim) on the same constructions in order to determine the similarities and differences between them. This leads to insights into the relationship of the datatypes, the factors measured and the nature of the constructs of well-formedness underlying them.

On the basis of these studies we shall make a number of claims. First, that judgements can be made a much “harder” data type without recourse to experimentation. Second, that with systematic manipulation two different types of judgements can be distinguished: categorical judgements and relative judgements. The different patterns obtained from relative judgements on the one hand and from both categorical judgements and frequency data on the other hand reveal two very different modes of operation, which, we argue, must be taken as evidence of two different modules in linguistic processing. Our Decathlon Model specifies a Constraint Application module which functions cumulatively, and an Output Selection module which functions competitively. Relative judgements tap into the output of the first, while categorical judgements and frequency data additionally reflect the processing of the second. It follows that relative judgements can be more informative about the formants of

grammaticality, and thus about the grammar, than either frequency data or categorical judgements.

To make the most effective use of relative judgements, we require two things: first a procedure for obtaining them, and second a standard set of reference judgements against which to measure them. This latter is anyway required to deal with the phenomenon gradient well-formedness, which is attracting increasing attention in syntax (eg Bresnan et al 2001).

We shall thus make three proposals to the syntactic community making use of judgement data: 1) that a standard set of reference judgements be adopted, in order to give an inter-subjective reference scale for gradient well-formedness judgements 2) that syntacticians should submit their judgements to a public database of judgements, where they would be open to peer-group scrutiny and comment. 3) that a standard procedure for distinguishing syntax-relevant and syntax-irrelevant effects in data be adopted.

### References

- Bard E., Robertson D. & Sorace A. 1996 Magnitude estimation of linguistic acceptability. *Language* 72 (1), 32-68
- Bresnan J., Shipra D. & Manning C. 2001 Soft constraints mirror hard constraints: Voice and person in English and Lummi. In: *Proceedings of the LFG 01 Conference*, University of Hong Kong. CSLI Publications: Stanford
- Schütze C. 1996 *The Empirical Base of Linguistics: Grammaticality Judgements and Linguistic Methodology*. Chicago: University of Chicago Press
- Keller, Frank. 2000. *Gradience in Grammar: Experimental and Computational Aspects of Degrees of Grammaticality*. PhD Thesis, University of Edinburgh.

---

Janet Dean Fodor & Yoshihisa Kitagawa

### **Prosodic influence on syntactic judgments**

*Donnerstag/Thursday: 9:00*

---

For some syntactic constructions, constraints at the PF-syntax interface require a distinctive prosodic contour. This may also imply the converse: that without the special prosody, the construction may be (mis-)perceived as syntactically ill-formed. It may happen, therefore, that syntactic judgments are uncertain or inconsistent when sentences are presented in written form with no prosody indicated, as is common. Recent work on sentence processing has shown that even in silent reading a prosodic contour is mentally projected onto written word sequences. According to the Implicit Prosody Hypothesis (IPH; Fodor 2002), it is the default prosody that tends to be projected. Hence, constructions requiring a non-default prosody are the most vulnerable to misjudgments concerning syntactic wellformedness.

We present a case study which supports this conjecture. In Japanese, the important theoretical question of whether Subjacency restricts covert (LF) movement has been a topic of dissension (see, e.g., Watanabe, 1992) due to opposing judgments of acceptability for sentences such as (1) below, in which a *wh*-phrase in situ in a subordinate *whether*-clause (a ‘*wh*-island’) is intended to have matrix clause scope. Deguchi & Kitagawa (2002) observed that “prosodic eradication” (i.e., compression of pitch and amplitude range, virtually suppressing lexical accents) is required between the surface position of a *wh*-phrase and the position of its scope marker (an interrogative complementizer, COMP<sub>WH</sub>, at the end of the relevant clause). Eradication is clearly a marked prosody. Eradication extended over many words and multiple clauses is highly dispreferred (as is known from examples in which scope ambiguity and Subjacency are not at issue). Thus, if readers tend to assign default prosody, they would eradicate only in the subordinate clause, as indicated by the underline in (1), and fail to apply

eradication across both clauses. In that case the matrix-scope interpretation will often be overlooked, resulting in judgments that seem to show Subjacency at work. Other readers may spot the possibility of extending the eradication through the matrix clause, and then judge the sentence to be acceptable. Hence the discord in syntactic theory is quite probably prosodic in origin.

An opposite tendency is found in sentence (2), in which the *wh*-phrase has been scrambled out of the embedded clause to the top of the matrix clause. For (2) the matrix-scope interpretation is easy to compute, but some readers fail to recognize the interpretation with embedded-clause scope. While there might be other, non-prosodic, explanations for (1) or for (2) alone, they offer no account of this surprising contrast between (1) and (2). For example, a semantic explanation, based on greater complexity of the discourse presuppositions of a matrix scope reading, would mis-predict a preference for embedded-clause scope in both constructions. The prosodic explanation, however, can account for (2) as well as (1).

In (2), the preference for avoiding extended eradication is outweighed by another (perhaps universal) tendency: a preference for congruence of prosodic and syntactic structure where possible (though an exact match between them cannot always be achieved, for various principled reasons; Selkirk, 2000). In (2), the scrambled *wh*-phrase precedes a constituent (*John-wa*) of the matrix clause. If embedded-scope is intended, eradication extends from the *wh* to the end of the subordinate clause, trapping the intervening matrix-clause constituent within the prosodic eradication domain. Syntax/prosody congruence then favors inclusion of the rest of the matrix clause in the eradication domain also (despite the preference otherwise for shorter eradication). But then, if readers apply eradication to the end of the matrix clause, as indicated by the dashed line in (2), they would naturally compute the associated syntactic/semantic scope (i.e., matrix scope), and would find the embedded-scope reading unacceptable.

These examples suggest that unexpected ‘blindness’ of the syntactic processor to certain analyses of sentences, leading to judgments of ungrammaticality, may be attributable to very general phonological tendencies, coupled with a preference for a simple relationship between syntactic and phonological structure. The methodological implication is that – at least for syntactic structures which require a marked prosody – judgments need to be obtained on spoken, not written, materials, so that they are not at the mercy of informants’ ability to think up unusual prosodic contours to assign to them. To the best of our knowledge there is only one report in the literature of an empirical study which compares judgments of spoken versus written materials; this is Keller & Alexopoulou (2001) on focus and word order in Greek. We advocate more widespread use of spoken (especially, recorded) sentences for obtaining syntactic wellformedness judgments. In support of this annoying but essential modification of current practice, we will present comparative judgment data on (1), (2) and other related constructions.

- (1) *Wh* in situ in subordinate clause. Matrix scope of *nani* (bound to the matrix-final  $COMP_{WH}$  *-no*) is often rejected → Disagreement about Subjacency.

[<sub>CP1</sub> John-wa [<sub>CP2</sub> Mary-ga nani-o tabeta-kadooka ] imademo siritagatteiru-no ]?  
 -TOP -NOM what-ACC ate-WHETHER even.now want.to.know-COMP<sub>WH</sub>  
 ‘What<sub>1</sub> does John still want to know [ whether Mary ate t<sub>1</sub> ]?’

- (2) *Wh* scrambled out of subordinate clause. Ambiguous, but subordinate scope of *nani* (bound to the subordinate  $COMP_{WH}$  *-ka*) is often rejected → Disagreement about ‘LF reconstruction’.

[<sub>CP1</sub> Nani-o<sub>1</sub> John-wa [<sub>CP2</sub> Mary-ga t<sub>1</sub> tabeta-ka ] imademo siritagatteiru-no ]?  
 what-ACC -TOP -NOM ate-COMP<sub>WH</sub> even.now want.to.know-Q  
 ‘Does John still want to know [ what<sub>1</sub> Mary ate t<sub>1</sub> ]?’

## References

- Deguchi, M. and Kitagawa, Y. (2002) Prosody and wh-questions. In M. Hirotani (ed.), *Proceedings of NELS-32*, 73-92.
- Fodor, J. D. (2002) Psycholinguistics cannot escape prosody. In *Proceedings of the Speech Prosody 2002 Conference*, Aix-en-Provence, pp.83-88.
- Keller, F. and Alexopoulou, T. (2001) Phonology competes with syntax: Experimental evidence for the interaction of word order and accent placement in the realization of information structure. *Cognition* 79:3, 301-372.
- Selkirk, E. (2000) In M. Horne (ed.) *Prosody, Theory and Experiment: Studies presented to Gösta Bruce*, Dordrecht:Kluwer.
- Watanabe, A. (1992) Subjacency and S-structure movement of wh-in-situ. *Journal of East Asian Linguistics* 1, 255-291.

---

Alexander Geyken & Christiane Fellbaum

### What is the optimal corpus size for the study of idioms?

*Freitag/Friday: 12:30*

---

Idioms can be considered a special class of "words," and their syntactic behavior is particularly interesting. Because idioms tend to be regarded as fixed lexical units whose constituents often do not show the full range of syntactic behavior found in the literal language, the syntactic variation of idioms is of relevance to the question of wellformedness and speakers' intuitions. Variations on the citation form of idioms are often dismissed as colloquial or substandard and are relegated to the margins of the spoken language, literature, advertising language, or humorous play with words. One reason may well be the relative infrequency of idioms.

A consensus seems to have emerged that corpus data constitute a more objective basis for research than constructed data. We investigate the syntactic behavior of German idioms on the basis of a large reference corpus of 1 billion words of running text (Klein and Geyken, 2000; Fellbaum, Kramer, and Stantcheva, 2003). Here we present the results of an experiment that addresses the following questions:

- (a) Are all corpus data equally valid or should one recognize differences in their quality and reliability? For example, are particular examples characteristic of specific genres, such as humor or advertising language?
- (b) How large does a corpus have to be to yield data matching that in a contemporary paper dictionary?

We select a test set of 50 VP idioms from the Duden 11, which includes not only the citations forms but also syntactic variations of an idiom. We partition our corpus into 100 random selections of 10 mio words each. We determine the frequency of occurrence of idioms and variations from the test set in each partial corpus. Taking the union of the partial corpora we derive a function relating the number of idiom tokens and the corpus size which lets us extrapolate the minimal size of a corpus that contains at least one token of the idiom and the syntactic variations in the test set.

We draw some conclusions on the effect of corpus size and genre on the frequency of idioms and that of syntactic variations.

## References

- Fellbaum, C., U. Kramer, and D. Stantcheva (2003), "Semantische, syntaktische und diachronische Analyse deutscher Idiome: Eins und etwas in VP-Idiomen". Eingeladener Vortrag 39. IDS Jahrestagung, Mannheim 11.-13.3.2003
- Klein, W., and A. Geyken (2000), Projekt "Digitales Wörterbuch der deutschen Sprache des 20. Jh.". In: Jahrbuch der BBAW 1999, Berlin: Akademie Verlag, S. 277-289.
- Senellart, J. (1996), "Statistique Prudence: Quelques études statistiques avec les noms-composés dans le Journal Le Monde". 3<sup>e</sup> Journées d'Intex. Paris. Jussieu.

---

Anke Lüdeling & Stefan Evert

### Adjective-noun combinations and adjective-noun compounds

Donnerstag/Thursday: 12:00

---

Productivity – the ability of a speaker to produce and understand infinitely many expressions – is one of the main foundations for generative linguistic theory. In syntax, the discussion of productivity has focused on formal methods like recursivity and coordination in combination with the qualitative description of rules or constraints. The notion that some rules are more likely to produce new instances than others has been relegated to 'performance' therefore outside the scope of linguistic interest. In word formation, on the other hand, the problem of productivity has been researched quantitatively as well as qualitatively. Several models to model different degrees of productivity using corpus data have been suggested (Baayen 2001, Bauer 2001, Lüdeling & Evert 2003).

Using the adjectival modification of nouns as an example, we argue that the quantitative productivity measures developed for morphology can be employed to measure the productivity of syntactic construction as well and that there is competition between syntax and morphology. German has two ways to productively modify a noun with an adjective: (a) syntactically, as in *roter Wein* 'red wine' and morphologically, as in *Rotwein* 'red wine'. In the literature the processes are viewed separately and it is sometimes argued that they have (incompatible) different properties – it is claimed for example that the compounds are terms while the syntactic combinations are semantically transparent. A qualitative and quantitative corpus study of adjective noun combinations reveals that the qualitative differences are gradual rather than categorical and that productivity rates of competing processes are closely interdependent.

## References

- Baayen, Harald (2001) *Word Frequency Distributions*. Kluwer, Dordrecht.
- Bauer, Laurie (2001) *Morphological Productivity*. Cambridge University Press, Cambridge.
- Lüdeling, Anke & Evert, Stefan (2003) *Linguistic experience and productivity: Corpus Evidence for fine-grained distinctions*. In: Proceedings of the 2003 Corpus Linguistics Conference, Lancaster.

---

Markus Meyer

**Problems and possibilities of modelling gradience in grammaticality/acceptability**

*Donnerstag/Thursday: 10:30*

---

In my presentation I want to show which problems are connected with the empirical foundation of syntactic modelling. Further I want to show which model theoretical and experimental possibilities currently exist to draw a ‚Grenze des Grammatischen‘ between wellformed and not-wellformed sentences in a language in such a way that gradience in judgments can be taken into consideration. From my point of view necessary standards for the empirical basis of modelling in syntax can only be defined if one has clarified which role a basic concept like ‚gradience in grammaticality/acceptability‘ plays in syntax and in which way one has to put the ‚syntax internal‘ speech about ‚grammaticality‘ in relation to ‚syntax external‘ phenomena like judgments of speakers.

Looking at Chomsky’s Generative Grammar from the beginning onwards and some versions of OT I firstly want to explain why there are mainly conceptual reasons for not being able to take gradience in grammaticality/acceptability into proper consideration (Miller and Chomsky 1963, Prince and Smolensky 1993, Keller 1998, Keller and Sorace 2002).

As gradience plays a fundamental role in empirical research I want to show secondly that one of the main problems in setting standards in linguistics is that methods of measurement are not reflected well enough (Bard et al. 1996): While in linguistic research in most of the cases judgements are collected with nominal and ordinal scales, interval and ratio scales are well applied in empirical research (Cowart 1997, Bard et al. 1996). The use of these latter scales entails that values of wellformedness can be modelled as a continuum. Research has shown that speakers are able to give very fine judgements.

The question of which scale to use is only one of the various questions of how to establish standards of syntactic research. I want to show thirdly that the way examples are used in syntactic research indicates that it is not reflected what can be shown by ‚using examples‘ (in the sense of Wittgenstein 1995) and how to use them (e. g. lack of standardization of notation of wellformedness-values).

In conclusion one has not only to distinguish carefully between ‚syntax internal‘ speech about ‚grammaticality‘ and ‚syntax external phenomena‘ but also between objects of syntactic theory and objects of empirical research (Chomsky 1964, Coward 1997, Bortz and Döring 2002). Therefore ‚gradience‘ is a complex concept which has to be made explicit in its different dimensions before starting to establish possible standards.

---

Stefan Müller & Stephan Oepen

**Example sentences and making them useful for  
theoretical and computational linguistics**

*Freitag/Friday: 11:30*

---

The systematic collection of test examples is of vital importance for both theoretical and computational linguistics. With rare exceptions, linguists use their private data collections (filing cards, plain text files, databases with complex search tools). Parts of these data collections are published in theoretical papers, but current research is to a large extent theory-driven and not data-driven. Accordingly, relevant data is not published and not discussed in

journal articles. Such data usually can be found only in monographs, since they provide more space for the discussion of problematic data. In our view databases that relate example sentences to certain phenomena are necessary for further progress in linguistics. The data shall be publically available and should be used to evaluate new theories and analyses.

A database with linguistically interesting examples is currently developed in Bremen. It contains examples from linguistic publications that are adapted to be usable for testing computational grammars. Ungrammatical examples are of particular importance. Ungrammatical examples of the kind discussed in linguistic publications can neither be found in corpora nor can they be constructed by hand in a systematic way without regarding complicated framework dependent structures that have to be excluded with respect to a suggested analysis. Depending on the theoretical framework one works in different subsets of the examples will be relevant. For the evaluation of particular analyses subsets of the examples in the database can be used. The database is an extension to the TSNLP test suite for German which, with some 4,500 items, is already quite substantial (Oepen, Netter and Klein, 1997). The new database contains examples that were collected in connection with the implementation of the Babel System (Müller, 1996). The database was systematically revised and a cross classification of phenomena was introduced. At the same time, more examples from the theoretical literature were added. When corpus data is introduced into test suites for computational linguistics problems can arise, as it is important to keep the number of lexical items that have to be introduced in a grammar for testing as small as possible. For many phenomena, however, context, intonation, and lexical material play a role, which can make examples ungrammatical after syntactic and lexical simplification.

The goal is to extend this database so that it contains all linguistically relevant phenomena that are discussed in publications or that were implemented in computational grammars. The database can then be used to evaluate various natural language processing systems with regard to their coverage. Until now the evaluation of systems was task-related, i. e. the number of (correct) analyses of utterances from a certain corpus. However, systems that achieve the same coverage with regard to corpora can be different in their adequacy from a linguistic point of view. For instance, subject-less verbs do not appear in certain kinds of text and are very rare in others. So the coverage of a grammar that does not deal with this phenomenon is not affected, but a linguistic theory that cannot explain subject-less sentences is of little interest to linguists.

The database builds on the [incr tsdb()] software package (Oepen and Callmeier, 2000), combining a simple relational database with a generic interface to computational analysis systems and a wealth of evaluation functions. The design of database and query language puts strong emphasis on ease of installation and use, aiming to make the tool usable also for inexperienced users. Test data and database software will be demonstrated during the talk

## References

- Mueller, Stefan. 1996. The Babel-System. An HPSG Prolog Implementation. In Proceedings of the Fourth International Conference on the Practical Application of Prolog, pages 263--277, London, <http://www.cl.uni-bremen.de/~stefan/Pub/babel.html>.
- Oepen, Stephan and Callmeier, Ulrich. 2000. Measure for Measure: Parser Cross-Fertilization. Towards Increased Component Comparability and Exchange. In Proceedings of the 6th International Workshop on Parsing Technologies, pages 183--194, Trento, Italy.
- Oepen, Stephan, Netter, Klaus and Klein, Judith. 1997. TSNLP. Test Suites for Natural Language Processing. In John Nerbonne (ed.), Linguistic Databases, Natural Language Processing. In John Nerbonne (ed.), Linguistic Databases, pages 13--36, Stanford, CA: CSLI Publications.



---

Andreas Nolda

**Rejection through correction: A method for collecting acceptability judgements**

*Mittwoch/Wednesday: 15:00*

---

Explicit speaker judgements on the acceptability of isolated sentences tend to differ from the speakers' actual usage. This divergence is particularly frequent when the sentences to be evaluated are considered non-standard or when their acceptability depends on accentuation and context.

In this talk I shall discuss a method, developed in co-operation with Manfred Krifka, for collecting acceptability judgements in a more indirect and integrated way. This method was applied to German 'split topicalizations' ('split NPs') in a study aiming at separating, on a per-idiolect basis, unacceptable variants of this construction from variants which are (still) acceptable in an appropriate context.

German native speakers were asked to read out sentences on a questionnaire and to cross them off in case they would not speak like that; optionally a correction could be added. For each sentence, the questionnaire provided an suitable context. Syntactic accents were specified informally. In order to increase personal involvement, the sentences were declared belonging to the screenplay of a film in which the informants should play themselves.

The data collected in this way were interpreted according to the following principles. A sentence from the questionnaire—in the specified accentuation—is considered as unacceptable for the speaker if and only if it has been crossed off or corrected in some way. A sentence is regarded as unacceptable for the speaker with respect to the 'split topicalization' variant it instantiates if and only if the sentence was crossed off or corrected in a 'relevant' way. A relevant correction yields a sentence of another variant of 'split topicalization' or no such construction at all.

---

Matthias Schlesewsky

**The emergence of linguistic judgements**

*Mittwoch/Wednesday: 16:30*

---

The present paper is concerned with the origin of acceptability judgements and will discuss consequences with respect to data acquisition and the scope of data in linguistic theory.

I will address the question "whence linguistic judgements?" with regard to the well-known contrast between subject- and object-initial structures (wh-movement, scrambling and topicalisation). A number of experimental techniques will be discussed, ranging on a temporal scale from a real-time reflection of comprehension processes (measured by event-related brain potentials and speed-accuracy trade-off) via speeded acceptability judgements and functional magnetic resonance imaging to a fully time-insensitive method (magnitude estimation). In all cases, the data provides clear evidence that on-line processing difficulty interacts with the ease of conflict resolution in determining final (off-line) acceptability. These findings indicate that, even in the absence of time pressure, linguistic intuitions are never independent of the comprehension process and therefore can never be fully liberated from performance.

Beyond this methodological comparison, I will argue that the role of the environment (frequency of the structure, inter-individual variations etc.) in the modulation of a judgement differs as a function of the method. Furthermore, external influences like the instruction of the

experimental participants as well as the choice of filler and/ or reference sentences, seem to have a considerable impact on how sentences are judged and should be therefore considered a possible source of gradience.

Based on these observations, the second and final part of the talk will focus on standards for the acquisition and documentation of acceptability judgements. Seeing that all experimental results must fulfil the replicability requirement, it is important to consider the necessary information that must be reported in addition to the experimental data. Finally and more importantly in the context of the interaction between theory and experimental data, we must ask which requirements with regard to verifiability should be applied to syntactic theories when these diverge from linguistic intuitions.

Horst Simon

**How to build a tool for the comparative syntax of German dialects**

*Donnerstag/Thursday: 11:30*

One of the fundamental issues in a discussion of grammatical variation must be the relationship between micro- and macro-variation, i.e. the relations of (dis)similarities between different ‘dialects’ of a single ‘language’ on the one hand and between entirely different ‘languages’ on the other hand. At the speaker level, this boils down to the question whether somebody speaking, say, Bavarian and Standard German can be said to be bilingual in the same sense as somebody who speaks Turkish and Standard German; additionally, it needs to be understood how the former relates to somebody speaking Thuringian and Standard German. – For an informed consideration of these issues it is essential to have a thorough understanding of the extent and the type of variation that can be found in different varieties of a ‘language system’.

In my ongoing project I work out an empirical tool that is designed to clarify dialectal variation in the syntax of German nominal phrases. In order to obtain strictly comparable data for each dialect, I develop a detailed questionnaire combining a variety of elicitation techniques as well as a narration stimulus for free text production.

In my paper I present a preliminary version of the empirical instruments and discuss the methodological rationale underlying it.

Britta Stolterfoht, Thomas Weskott, Ina Bornkessel & Matthias Schlesewsky

**The task-dependency of acceptability judgements: Processing scrambling and topicalization in German**

*Mittwoch/Wednesday: 15:30*

Word order variations are amongst the most extensively investigated phenomena in German sentence processing research. Numerous studies have revealed a robust preference for the canonical word order, which was shown with different methods (e.g., Bader & Meng, 1999; Hemforth, 1993; Friederici 1998). However, most of the studies presented the critical non-canonical sentences in isolation. Such a procedure appears questionable in view of the information structural characteristics of scrambling and topicalization, which require a specific context. It is assumed that scrambled or topicalized sentences with a transitive agentive verb and an accusative object DP are only licensed by a context which requires the

subject DP to appear as new information (focus) and the object DP as given information (background) (Rochemont, 1986; Stechow, 1991). Only a few studies have looked into the processing of word order variations in context and these have shown rather inconsistent data (Bayer & Marslen-Wilson, 1992; Meng et al., 1999; Weskott, 2002). This may be a result of the different structures and/or different methods used in these experiments.

In our study, we aimed to examine whether different types of acceptability judgement tasks are equally sensitive to contextual information. In addition, we also investigated whether different structures (topicalization vs. scrambling, sentences with ambiguous vs. unambiguous DPs) are also judged differently in the same contexts. We presented question-answer pairs and varied the variables **context**: (1) neutral, (2) focus subject, (3) focus object; **word order**: (1) subject-object, (2) object-subject; **structure**: (1) topicalization, (2) scrambling; **ambiguity**: (1) ambiguous, (2) unambiguous; and **task**: (1) How acceptable is Sentence 1?, (2) How acceptable is Sentence 2?, (3) How natural is the dialogue?, (4) How natural is Sentence 2 as a continuation of Sentence 1?.

Our results show that all factors have a significant influence on context sensitivity. Context effects were highly dependent on structural characteristics as well as on the type of judgement task. Therefore, the presence or absence of context effects in previous studies cannot be interpreted in a global way, but only in relation to structural and task specific properties.

Stefan Sudhoff, Denisa Lenertová & Anita Steube

### Empirical aspects of German bridge contours

*Donnerstag/Thursday: 10:00*

In our talk, we want to present the results of two experimental studies examining (1) the positional variability of non-referential, normally “non-moving” constituents serving as so-called I-topics in German Bridge Contours (BCs) and (2) the properties of constituents associated with the stressed focus particle *auch*, which seem to behave like I-topics.

BCs (the term is used here as referring to the whole construction with its special prosodic, semantic, syntactic, and pragmatic properties) have two separate pitch accents. The rise is a contrastive-focus accent, and the fall is a focus accent. The semantic support for a contrastive-focus accent is the additional propositional meaning that there exists an alternative to the I-topic in the set comprising the discourse topic just as in correction sentences with one contrastive accent only (cf. Steube 2003).

BCs are categorical sentences, i.e. their I-topics are moved out of the focus domain. In German, I-topics take up the following positions: the prefield; an I-topic position immediately after  $C^\circ$  (cf. Haider & Rosengren 1998, Frey 2000); the topic position before the pronouns; the positions of pronouns; the positions of scrambled XPs; an I-topic position between two sentence adverbials (cf. Haftka 1995).

Our intuition was not sensitive enough to determine which of these positions can be taken up by non-referential, normally “non-moving” elements. For that reason we carried out two perception experiments testing (A) main clauses and (B) subordinate clauses. We selected directional adverbials, modal adverbials and separable prefixes as the constituents to be examined and tested them in the following positions: for (A) I. the prefield, II. the I-topic position behind  $C^\circ$ , III. a position behind the pronouns, IV. a position behind sentence adverbials; for (B) I. the I-topic position behind  $C^\circ$ , II. a position between two pronouns, III. a position behind the pronouns, IV. a position behind sentence adverbials.

The tested material is based on sentence pairs, each consisting of a context sentence and a target sentence, both of which are syntactically parallel BCs (cf. example (1) with a

directional adverbial in position I. of condition A). The subjects were presented the material auditorily and had to give grammaticality judgements on a five point scale.

- (1) /**Zum Arzt** hat Rolf seine Frau sicher \mit dem MOTORRAD gefahren.  
/ **Ins Büro** hat er sie tatsächlich \mit dem AUTO gebracht.

From the results we drew the conclusion that the German I-topic-positions must not in general be extended to “non-moving” non-referential constituents. They are fully acceptable only in the prefield, whereas the other positions yield a greater variability in the grammaticality judgements. The experiment also demonstrates that German main and embedded clauses show parallel restrictions on I-topic positions.

The aim of the second study was to find empirical evidence for Krifka’s (1999) claim that the associated constituent of the stressed focus particle *auch* is an I-topic, which can, but need not be prosodically marked as such. Our hypothesis is that this optionality in the prosodic realisation should disappear in ambiguous cases, where stressed *auch* can associate with more than one constituent to its left (located in the prefield or middlefield), cf. (2).

- (2) **Am Dienstag** ist Ingo wahrscheinlich AUCH in der Bibliothek gewesen.

We combined a production experiment, in which the subjects had to produce sentences of this kind in contexts triggering one of the possible readings, with a perception experiment, in which other subjects were presented the produced data without context and had to disambiguate them in a sentence completion task. The findings of an acoustic analysis of the production material will be compared with the outcome of the perception experiment.

#### References:

- Frey, W. (2000): Über die syntaktische Position der Satztopiks im Deutschen. In: *ZAS Papers in Linguistics*, 20, 137 – 172.
- Haftka, B. (1995): Syntactic Positions for Topic and Contrastive Focus in the German Middlefield. In: Kohlhoff, I., S. Winkler & H.-B. Drubig (ed.): *Proceedings of the Göttingen Focus Workshop 17th DGfS*, March 1 – 3, 1995. Heidelberg.
- Haider, H. & I. Rosengren (1998): Scrambling. In: *Sprache und Pragmatik. Arbeitsberichte*, 49
- Krifka, M. (1999): Additive Particles under Stress. In: *Proceedings of SALT 8*. Cornell: CLC Publications, 111 – 128.
- Steube, A. (2003): Bridge Contours in German assertive main clauses. In Abraham, W. & L. Molnarfi (eds.): *Optionality in Syntax and Discourse Structure – Aspects of Word Order Variation in (West-)Germanic and Other Indo-European Languages*. Berlin: Mouton de Gruyter, 163 – 190.

---

Hans Uszkoreit

**Empirical methodology for grammar research:  
Treebanks, phenomena-based data sets and online experiments**

*Freitag/Friday: 13:00*

---

Currently linguistics is developing a respectable empirical branch. However there seem to be several competing paradigms of empirical language research. Among them are the utilization of raw corpora, the creation and exploration of linguistically interpreted corpora such as treebanks, the design of sophisticated online experiments and the exploitation of subjective rating data. Moreover, the validity and relevance of introspective data has become the issue of a heated debate.

We will describe and characterize the most common sources for obtaining linguistic data such as introspection, subjective acceptability judgements, online experiments and corpus research. Each of these sources exhibits certain characteristics that make it indispensable. The respective contributions of these data sources to theoretical research will be compared. An approach of integrating the different methods in a systematic research strategy will be outlined.

Furthermore we will propose a setup for linking introspective data, found examples and large scale interpreted corpora. The basis of this setup is the annotation of different sorts of data with a unified metadata scheme. The result will be an evolving dynamic database of linguistic phenomena with an ontological foundation. The approach will be exemplified by data, ontologies and annotations from our own research on word order, structural complexity and processing load.