

Acceptability judgments and linguistic competence

Abstract

Acceptability judgments are the primary source of data for linguistic theory, based on the assumption that they reliably reflect linguistic competence. However, this assumption has always been challenged by studies showing the influence of extralinguistic factors on the judgment process, leading to recent linguistic research employing better experimental methodology. This paper discusses the most significant finding of this research, the existence of gradient judgments, and how they have been accommodated within linguistic theories. The implications of the judgment-competence relation beyond theoretical linguistics (e.g., first and second language acquisition and psycholinguistics) are also discussed.

Keywords: acceptability judgments, sentence grammaticality, linguistic competence, linguistic theory, experimental syntax, gradience

1. Introduction

Sentence grammaticality plays a crucial role in linguistic theory, whose goal is to model the cognitive procedures underlying our linguistic knowledge. Knowledge of sentence grammaticality is accessed through acceptability judgments, which form the primary source of evidence for sentence grammaticality.¹ The validity of using judgment data critically rests on the assumption that a correlation exists between native speakers' tacit knowledge of sentence grammaticality (linguistic competence) and the acceptability judgments they provide (performance). Thus, grammatical sentences should elicit high acceptability ratings, and ungrammatical sentences, low ones. Without this correlation between competence and performance there would be no good reason to use acceptability judgments as linguistic evidence; as a source of data they would be as informative as speakers' intuitions about the number of words in a sentence.

This paper reviews the role of acceptability judgments in linguistic theory, beginning, in section 2, with early debates on the competence/performance relation questioning the use of judgment data. Section 3 discusses one way in which this issue has been addressed, i.e., by strengthening the empirical foundations of linguistic research. Section 4 discusses gradient judgments, one of the most significant results of improvement in methodology. Current event-related brain potential (ERP) research is introduced in section 5 as a potential alternative method of accessing knowledge of

¹ While the term "grammaticality judgment" is often used interchangeably with "acceptability judgment," the latter term is more appropriate when referring to conscious judgments on the naturalness and comprehensibility of a sentence. According to Chomsky (1965), acceptability is a concept that belongs to the study of performance, while grammaticality is associated with competence. Given that speakers are unable to consciously access their linguistic knowledge, it is actually not possible for them to make judgments on sentence grammaticality (see Schütze 1996: 26).

sentence grammaticality. Section 6 presents implications of the competence/performance relation for psycholinguistic research, followed by the conclusion in section 7.

2. The competence/performance debate

The role of acceptability judgments in defining linguistic competence has been a controversial issue since they were first introduced as linguistic evidence over fifty years ago. The main question is, To what extent can the task of making judgments, being a behavioral process, reflect the speaker's internal knowledge of sentence grammaticality? Chomsky (1965:3-4) clearly stated that there was no direct correlation between judgments and grammaticality except in the case of an "ideal speaker-hearer." In later work, he made a similar claim that "informant judgments do not reflect the structure of the language directly... [they] may fail to provide direct evidence as to grammatical status because of the intrusion of numerous other factors" (Chomsky 1986: 36). While recognizing that the relation between competence and performance was not absolute, at the same Chomsky was able to justify the use of judgments as primary data by exploiting the abundance of data in those earlier days to build grammars based only on the "clear cases" of grammaticality. Furthermore, Chomsky claimed that in some instances it was not speaker judgments but rather the theory itself that should decide on the grammaticality of a sentence (Chomsky 1957: 14). Such claims raised questions about what exactly constituted a "clear case," and when and why the theory should decide on grammaticality (see Labov 1975, Ringen 1979 and discussion in Schütze 1996).

Although it was widely acknowledged that acceptability judgments do provide insights into linguistic competence, it was also recognized that they were in fact based on a behavioral process that was subject to nonlinguistic influences (e.g., Bever 1970, Levelt 1974, Newmeyer 1983, Birdsong 1989, Schütze 1996). Over the years studies have revealed many subject-related and methodological sources of judgment variation. These include the judger's mental state, linguistic sophistication, theoretical bias, and social attitudes, as well as clarity of test instructions, frequency of sentence usage, perceptual difficulty, repeated exposure and context (e.g., Bolinger 1968, Bever 1970, Spencer 1973, Levelt 1974, Labov 1975, Greenbaum 1976, 1977, Carroll, Bever and Pollack 1981, Nagata 1988, Cowart 1997, and Snyder 2000; see Schütze 1996 for a thorough overview of this research).

Another source of criticism in the use of acceptability judgments was the informal methods of data collection followed by linguists who relied on their own introspective judgments or those of a few colleagues to support their linguistic models. Not adhering to the normal procedures of psychological experimentation, such elicitation methods were criticized as being unsystematic and unreliable (e.g., Maclay and Sleator 1960, Bever 1970, Labov 1972, 1975, Greenbaum 1973 and Derwing 1977).

Despite these objections, linguists continue to use acceptability judgments as evidence for their models of grammar because of factors such as the ease with which they are collected, their success in uncovering new facts about language, and linguists' specialized knowledge of what is required in making judgments (Newmeyer 1981). Here, then, is the paradox: linguistic theory is built on evidence from acceptability judgments, even though the judgments may not directly reflect linguistic competence. Without an operational criterion for determining grammaticality independently of

native speaker judgments, the entire enterprise of linguistic theory has until recently been built upon this inextricable link between acceptability judgments and linguistic competence.

3. Strengthening the empirical foundation

In order to legitimize the use of acceptability judgments as a data source, it is crucial that linguists understand and control for the factors that intervene in the relation between judgments and linguistic knowledge. An important step in this direction is to strengthen the empirical foundation of linguistic research by utilizing experimentally controlled methods of judgment collection that minimize the effect of confounding factors. In the early days of linguistic research little attention was paid to criticisms regarding the informal procedures used by linguists in eliciting acceptability judgments (e.g., Labov 1975, Levelt 1974, Derwing 1980, Birdsong 1989), as theoretical linguistics was still at a stage where there were “masses of evidence” and not enough theoretical tools to describe them. The basic facts of English grammar were still being discovered, and “sharpening of the data by more objective tests” was considered to be “a matter of small importance for the problems at hand” (Chomsky 1965:19-20). As noted by Levelt et al. (1977), the clear cases of grammaticality and ungrammaticality were sufficient for constructing and testing linguistic theory. Yet even as the theory was refined to accommodate unclear cases in which acceptability judgments did not correlate with predicted grammaticality, the focus was more on the contribution of the associated structures to linguistic theory rather than on how judgments on those structures were elicited (e.g., Lakoff 1973, Hindle and Sag 1975, Lasnik and Saito 1984, Ross 1987).

Back in 1965 Chomsky speculated that the day would come “when the kinds of data that we now can obtain in abundance will be insufficient to resolve deeper questions concerning the structure of language” (p. 21). Thirty years later Schütze (1996) claimed that this day had come, as linguists investigated issues that required a “methodology more systematic than reliance on everyday common sense” (p. 27). In the decade or so since then, works such as Cowart (1997), Keller (2000) and Featherston (2007) have provided linguists with the incentive and tools to pursue the challenge of developing a better methodology for linguistic research, resulting in a new field of study referred to as experimental syntax.

4. Gradient acceptability judgments

The most significant finding from recent studies in experimental syntax is the empirical reality of gradient acceptability judgments, i.e., judgments that are not clearly acceptable or unacceptable, but fall somewhere on a continuum between the two (see Bard, Robertson and Sorace 1996, Gibson and Thomas 1999, McDaniel and Cowart 1999, Keller 2000, Keller and Sorace 2003, Fanselow and Frisch 2006, Featherston 2005a,b,c, 2007, Francis and Matthews 2006, Alexopoulou and Keller 2007, Hofmeister 2007, Sprouse 2007, 2008, Fanselow, Lenertová and Weskott 2008, and Sag, Hofmeister and Sneider 2008, among others). While linguists had always acknowledged the gradient nature of acceptability judgments (e.g., Chomsky 1965, Lakoff 1973, Newmeyer 1983), few studies had seriously explored how such judgments fit into a linguistic model. Some notable exceptions were Lakoff (1973), Hindle and Sag (1975), Watt (1975) and Ross (1972, 1987), who investigated gradient judgments on a variety of structures including topicalization, positive *anymore*, strained

anaphora and pseudoclefts. As well, some early studies within the generative framework explored gradience found in *wh*-island constructions (e.g., Lasnik and Saito 1984, Aoun, Hornstein, Lightfoot and Weinberg 1987). The more recent experimental studies have compelled us to reevaluate our assumptions regarding the nature of grammaticality and the relation between judgments and linguistic competence, as we are confronted with gradient data that cannot easily be attributed to confounding factors resulting from inadequate methodology. Better experimental procedures for collecting and testing hypotheses about acceptability judgments have resulted in more reliable data and a better understanding of what a theory of linguistic competence must encompass.

Linguistic studies investigating gradient judgments fall into two categories. The first category consists of studies that maintain the direct correlation between acceptability judgments and linguistic knowledge by assuming that gradient judgments reflect gradience in the grammar. According to this view, gradience is considered to be an integral part of the grammar, manifested in the rules and constraints of the grammar, in the environment in which they apply, or in the properties of grammaticality itself (e.g., McDaniel and Cowart 1999, Keller 2000, Keller and Sorace 2003, Featherston 2005a,b,c, 2007).

Keller (2000), for example, proposes two types of constraints, hard and soft, whose violations result in differing degrees of unacceptability. Hard constraints such as inversion and agreement are immune to context effects, are developmentally stable, and induce serious unacceptability when violated. In contrast, soft constraints such as definiteness and referentiality (when extracting from picture NPs) are context-dependent, developmentally optional, and result in mild unacceptability when violated. Murasugi's (2008) Symmetric Movement Constraint favors movement to a structurally

similar position over an asymmetric one (e.g., subject-to-subject raising over subject-to-object raising), accounting for the relatively low acceptability of subject-to-object raising sentences such as *They declared Cathy to be the owner*. An example of gradience that is dependent on environment can be found in McDaniel and Cowart (1999), where it is claimed that the acceptability of resumptive pronouns depends on the grammatical status of the traces that they are associated with: when the trace is allowed by the grammar, the trace is preferred over a resumptive pronoun, but in cases where the trace is illicit, the acceptability of the pronoun increases.

A more radical approach assumes there to be no distinction between grammaticality and ungrammaticality. For Featherston (2005c), such “categoricity is ... an abstraction from the primary data,” as grammaticality is “a true continuum with no fixed endpoints... [A] structure is never absolutely ‘grammatical’ and ‘ungrammatical’ in this model of grammaticality, only ever more or less grammatical” (p. 676). Sampson (2007) makes a similar claim based on language use, suggesting that the notion of “ungrammatical” is “a delusion, based on a false conception of the kind of thing a human language is” (p. 1), as all sentences have the potential to be grammatical in certain discourse situations or at a later point in time.

In the second category are studies that preserve the categorical nature of grammaticality, and seek nonlinguistic explanations for the mismatch between discrete grammatical knowledge and gradient judgments (e.g., Sprouse 2007). With better controls on many of the methodological factors that were present in earlier studies, it has become easier to differentiate processing from linguistic factors, resulting in an increase in processing explanations for gradient structures that were traditionally accounted for linguistically (see, for example, Kluender and Kutas 1993, Francis and

Matthews 2006, Fanselow and Frisch 2006, Alexopoulou and Keller 2007, Hofmeister 2007, Fanselow, Lenertová and Weskott 2008, and Sag, et al. 2008).

In one such study, Hofmeister (2007) compared reading times on structures involving phrases with varying degrees of informational content extracted from three constructions: wh-islands (*Who/which employee* did Albert learn whether they dismissed after the annual performance review?), relative clauses (*What/which poll/which political poll* did the reporter that Scooter avoided discuss during an evening news segment?), and clefted indefinites (*It was a communist/an alleged communist/an alleged Venezuelan communist* who the members of the club banned from ever entering the premises.). In all three cases the amount of information encoded in the extracted phrase had a significant effect on processing, revealed by faster reading times for more informative NPs (e.g., *which NP*) at the gap site where the extracted phrase had to be reaccessed. While earlier studies had accounted for the *who* vs. *which NP* difference in terms of linguistic concepts such as D-linking (Pesetsky 1987) or referentiality (Rizzi 1990), Hofmeister proposed his Memory Facilitation Hypothesis, which states that elements with more informative content facilitate their subsequent retrieval from memory, thus resulting in higher degrees of acceptability.

The problem with many of the studies in both the discrete and gradient grammaticality categories is their lack of conclusive evidence for either a gradient or discrete grammar. This is due in part to the fact that the interpretation of their results is often biased toward the theoretical model that the study assumes, without explaining why the alternative view cannot also be supported. For example, Sprouse (2007) reports that there is no syntactic priming effect for certain ungrammatical sentences in an acceptability judgment task, and claims that this supports categorical rather than continuous grammaticality since a categorical grammar would predict that only

grammatical sentences have licit representations and thus will show syntactic priming effects. In other words, these results are taken as evidence of an asymmetric extra-grammatical effect on acceptability that affects grammatical but not ungrammatical structures. However, it is not clear why such effects could not equally be interpreted as affecting structures on the higher rather than lower end of grammaticality in a gradient grammaticality model. It appears that the affected structures are categorized as ungrammatical because of Sprouse's preconceived notion of categorical grammaticality; a proponent of gradient grammaticality would have claimed that syntactic priming affects only those structures within a certain range of acceptability.

It is admittedly very difficult to design an experiment that will be able to empirically distinguish discrete from gradient grammar through the use of acceptability judgments, given that gradience can appear even when judging membership in well-defined categories such as odd or even numbers (Armstrong, Gleitman and Gleitman 1983). As Schütze (1996: 69) explains, "If performance mechanisms induce graded structures by themselves, and if... they can never be circumvented because competence is not directly accessible, then it might not be possible to investigate empirically how a grammar itself classifies sentences."

5. ERP studies

An alternative approach to validating the use of judgment data is to compare the elicited judgments of native speakers with the results of another method of assessing grammaticality that does not rely on speaker judgments. The recording of event-related brain potential (ERP) in response to linguistic stimuli shows promise as an alternative method of accessing knowledge of sentence grammaticality. ERPs provide a

continuous, on-line record of the processes underlying language comprehension, without requiring a secondary task involving overt, conscious decision (Garnsey 1993, Osterhout 1994). Many studies have shown that ERPs are sensitive to the processes involved in syntactic analysis. For example, there is considerable evidence that syntactic violations elicit a large positive-going wave with an onset around 500 ms, known as the P600 or Syntactic Positive Shift (SPS) (see, for example, Osterhout and Holcomb 1992, Hagoort, et al. 1993, Osterhout, et al. 1994). While it is clear that the P600 reliably co-occurs with syntactic anomaly, Osterhout (1994) observes that we are far from understanding what cognitive-neural events underlie these ERP effects. It is not known, for example, “whether the P600 directly reflects the processes underlying syntactic analysis, or the processes that respond to syntactic anomaly, or the processes that attempt a reanalysis subsequent to an anomaly...” (p. 39). Despite our limited knowledge of the exact interpretation of ERPs, we can use them as a potential tool with which to investigate our linguistic competence.

It appears, though, that we have not yet reached the stage where we can use ERPs to confirm the validity of acceptability judgments as evidence for sentence grammaticality. ERP studies that include a separate off-line judgment task mainly use the judgments to confirm the correct categorization of stimulus items as grammatical or ungrammatical (e.g., Hagoort, et al. 1993). While there do exist studies that measure ERPs during the judgment process, in most cases they do not actually compare neurophysiological and behavioral responses (e.g., Osterhout and Holcomb 1992, Osterhout and Nicol 1999, Allen, et al. 2003, Kaan and Swaab 2003). There are, however, two exceptions: Osterhout and Mobley (1995), and Osterhout, Bersick and McLaughlin (1997). Osterhout and Mobley (1995) found that violations of pronoun gender agreement (e.g., *The aunt heard that he had won the lottery*) elicited a P600

response, but only for participants who judged those sentences to be unacceptable. Those who interpreted the *he* to refer to someone other than the aunt found such sentences to be acceptable; this was reflected in the absence of a P600 effect. Osterhout, et al. (1997) found a similar P600 effect with violations of gender stereotypes (e.g., *The nurse prepared himself of the operation*), but, in contrast to the previous study, this effect was present even when participants did not judge the violations to be unacceptable.

A similar mismatch between imaging and behavioral responses to grammaticality was found in an ERP study on second language learners. Tokowicz and MacWhinney (2005) observed that their English-speaking learners of Spanish were sensitive to violations of determiner gender agreement, a linguistic property that does not exist in their first language, but that their responses on a judgment task were only at chance. Tokowicz and MacWhinney suggested that, during sentence processing, learners have better access to implicit knowledge measured by ERPs than conscious, explicit knowledge. Clearly, more research is necessary to clarify the relation between ERP and behavioral responses to grammaticality before it can be determined whether their results are contradictory or complementary.

6. Implications beyond linguistic theory

The correlation between acceptability judgments and linguistic competence has implications beyond theoretical linguistics, as it is assumed, to a certain degree, in any experimental study that explores sentence grammaticality. First, the grammaticality of stimulus items, whether it be in a processing or acquisition study, is normally determined by the acceptability judgments of the researchers or the participants in a

norming study. Secondly, in analyzing the linguistic errors of populations such as second language learners or aphasic persons, it is again the researchers' judgments that determine the grammaticality of the structures in question. Thirdly, the native speaker controls whose judgments form the standard of comparison are assumed to provide judgments that correlate with their linguistic competence. Disciplines do differ, however, in their particular assumptions regarding the relation between sentence acceptability and grammaticality.

In second language acquisition studies, the rationale for the use of acceptability judgments is the same as that in theoretical linguistics, i.e., they provide a relatively direct window into the learner's linguistic competence (e.g., Arthur 1980). Judgment differences between native speakers and language learners can thus be explained to some extent as differences in their grammars. Researchers have questioned, though, whether the same criteria for acceptability are being used by the two groups (Birdsong 1989, White 1989, Ellis 1991). Studies have shown that learners often resort to strategies such as guessing, translation or analogy when they are unsure of the grammaticality of a sentence (e.g., Ellis 1991, Davies and Kaplan 1998). Moreover, in a recent study McDonald (2006) proposed a processing account of late second language learners' poor performance on a judgment task based on difficulties with memory, decoding, and speed. Research exploring differences in the the judgment task by native speakers and language learners could contribute to our understanding of the judgment process in all language users.

In first language acquisition studies, judging acceptability is considered to be a metalinguistic skill that demonstrates children's ability to evaluate syntactic form independently of its meaning (Hakes 1980). It is thus important that judgments be elicited on sentences that the child has already acquired, thereby ensuring that those

structures are part of the child's linguistic competence (Cairns, Schlisselberg, Waltzman and McDaniel 2006). Any discrepancies between the grammaticality of the structures and the child's judgments on their acceptability can be attributed to the child's underdeveloped ability to make correct judgments. In contrast, adults are assumed to have sufficiently developed this ability so that their acceptability judgments accurately reflect sentence grammaticality.

The traditional method of assessing linguistic knowledge in populations with language impairments is through comprehension tasks such as sentence-picture matching or object manipulation. Moreover, it has been a matter of debate whether poor performance on such tasks indicates a deficit in representing or processing syntactic information (see Caplan et al. 2007). Evidence for a processing account was presented in Linebarger, et al. (1983), whose agrammatic participants performed above chance on an acceptability judgment task despite impaired comprehension. Linebarger et al. claimed that their participants' performance on the acceptability judgment task indicated "significant sparing of syntactic knowledge in agrammatism" (p. 361). Many studies since have provided further support for intact linguistic knowledge in agrammatics demonstrated by performance on both on-line and off-line acceptability judgment tasks (e.g., Schwartz, et al. 1987, Wulfeck 1988, Lukatela, et al. 1988, Shankweiler, et al. 1989, Wulfeck and Bates 1991, Devescovi, et al. 1997).

In contrast, ERP studies with Broca's aphasics have found either no electrophysiological response (Wassenaar, et al. 2004) or a reduced and delayed P600 component (Wassenaar and Hagoort 2005) in response to ungrammatical sentences. Both these studies interpret the ERP results as reflecting a language processing deficit, contrasting with the conclusions of earlier studies that syntactic knowledge may be intact in Broca's agrammatics.

7. Conclusion

Regardless of the uncertainties surrounding the relation between judgment data and linguistic competence, and the objections to the methods used to collect speaker judgments, linguists and psycholinguists will continue to use acceptability judgments as an important source of evidence for their theories. To them, judgments are the most familiar and accessible data source, whose positive contributions far outweigh their shortcomings. It is essential, however, that linguists strive as best they can to ensure the reliability of their data, for example by improving the methods of data collection or by exploring alternative data sources (such as ERPs) that could be used to corroborate the judgment data.

Sampson (2007) is on the right track when he states that “the only way that we could ultimately know speaker intuitions to be reliable... would be to check a language-description based on intuitions against one based on empirical observation” (p. 16). However, a description based purely on empirical observation (for Sampson, corpus data) would be time-consuming to construct and necessarily incomplete, given the extremely large number and range of utterances from which generalizations must be made, in addition to the inability of corpora to contain examples of all generalizations that need to be tested (Meurers 2007). This contrasts with intuition data, which can be collected in a relatively short time by many speakers in a variety of languages (Newmeyer 1983, Phillips and Lasnik 2003). The challenge is to find empirical data from any domain that can shed light on the relation between acceptability judgments and linguistic competence.

References

- Alexopoulou, Theodora, and Frank Keller. 2007. Locality, cyclicity and resumption: At the interface between the grammar and the human sentence processor. *Language*, 83, 110-160.
- Allen, Mark, William Badecker and Lee Osterhout. 2003. Morphological analysis in sentence processing: an ERP study. *Language and Cognitive Processes*, 18, 405-430.
- Aoun, Joseph, Norbert Hornstein, David Lightfoot and Amy Weinberg. 1987. Two types of locality. *Linguistic Inquiry*, 18, 537-577.
- Armstrong, Sharon Lee, Leila R. Gleitman and Henry Gleitman. 1983. What some concepts might not be. *Cognition*, 13, 263-308.
- Arthur, B. 1980. Gauging the boundaries of second language competence: A study of learner judgments. *Language Learning*, 30, 177-194.
- Bard, Ellen Gurman, Dan Robertson and Antonella Sorace. 1996. Magnitude estimation of linguistic acceptability. *Language*, 72, 32-68.
- Bever, Thomas G. 1970. The cognitive basis for linguistic structures. In John R. Hayes, ed., *Cognition and the Development of Language*. New York: John Wiley and Sons.
- Birdsong, David. 1989. *Metalinguistic Performance and Interlinguistic Competence*. New York: Springer-Verlag.
- Bolinger, Dwight. 1968. Judgments of grammaticality. *Lingua*, 21, 34-40.
- Cairns, Helen Smith, Gloria Schlisselberg, Dava Waltzman and Dana McDaniel. 2006. Development of a metalinguistic skill: judging the grammaticality of sentences. *Communication Disorders Quarterly*, 27, 213-220.

- Caplan, David, Gloria Waters, Gayle DeDe, Jennider Michaud and Amanda Reddy. 2007. A study of syntactic processing in aphasia I: Behavioral (psycholinguistic) aspects. *Brain and Language*, 101, 103-150.
- Carroll, John M., Thomas G. Bever and Chava R. Pollack. 1981. The non-uniqueness of linguistic intuitions. *Language*, 57, 368-383.
- Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.
- Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge: MIT Press.
- Chomsky, Noam. 1986. *Knowledge of Language: Its Nature, Origin, and Use*. New York: Praeger.
- Cowart, Wayne. 1997. *Experimental Syntax: Applying Objective Methods to Sentence Judgments*. London: Sage Publications.
- Davies, William D. and Tamar I. Kaplan. 1998. Native speaker vs. L2 learner grammaticality judgements. *Applied Linguistics*, 19, 183-203
- Derwing, Bruce L. 1980. Against autonomous linguistics. In Thomas A. Perry, ed., *Evidence and Argumentation in Linguistics*, 163-189. Berlin: Walter de Gruyter.
- Devescovi, Bates, D'Amico, Hernandez, Marangolo, Pizzamiglio and Razzanos. 1997. An on-line study of grammaticality judgements in normal and aphasic speakers of Italian. *Aphasiology*, 11, 543-579.
- Ellis, Rod. 1991. Grammaticality judgments and second language acquisition. *Studies in Second Language Acquisition*, 13, 161-186.
- Fanselow, Gisbert and Stephan Frisch. 2006. Effects of processing difficulty on judgements of acceptability. In Gisbert Fanselow, Caroline Féry, Ralf Vogel and Matthias Schlesewsky, eds., *Gradience in Grammar: Generative Perspectives*, 291-316. Oxford: Oxford University Press.

- Fanselow, Gisbert, Denisa Lenertová and Thomas Weskott. 2008. Studies on the acceptability of object movement in Spec,CP. In Anita Steube, ed., *The Discourse Potential of Underspecified Structures*, 413-437. Berlin: Mouton de Gruyter.
- Featherston, Sam. 2005a. That-trace in German. *Lingua*, 115, 1277-1302.
- Featherston, Sam. 2005b. Magnitude estimation and what it can do for your syntax: Some wh-constraints in German. *Lingua*, 115: 1525-1550.
- Featherston, Sam. 2005c. Universals and grammaticality: Wh-constraints in German and English. *Linguistics*, 43, 667-711.
- Featherston, Sam. 2007. Data in generative grammar: The stick and the carrot. *Theoretical Linguistics*, 33, 269-318.
- Francis, Elaine J. and Stephen Matthews. 2006. Categoriality and object extraction in Cantonese serial verb constructions. *Natural Language and Linguistic Theory*, 24, 751-801.
- Garnsey, Susan M. 1993. Event-related brain potentials in the study of language: An introduction. *Language and Cognitive Processes*, 8, 337-356.
- Gibson, Edward and James Thomas. 1999. Memory limitations and structural forgetting: the perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes*, 14, 225-248.
- Greenbaum, Sidney. 1976. Syntactic frequency and acceptability. *Lingua*, 40, 99-113.
- Greenbaum, Sidney. 1977. The linguist as experimenter. In Fred R. Eckman, ed., *Current Themes in Linguistics: Bilingualism, Experimental Linguistics, and Language Typologies*, 125-144. New York: John Wiley and Sons.

- Hagoort, Peter, Colin Brown and Jolanda Groothusen. 1993. The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, 8, 439-484.
- Hakes, D.T. 1980. *The Development of Metalinguistic Abilities in Children*. New York: Springer.
- Hindle, Donald and Ivan Sag. 1975. Some more on anymore. In Ralph W. Fasold and Roger W. Shuy, eds., *Analyzing Variation in Language: Papers from the Second Colloquium on New Ways of Analyzing Variation*, 89-110. Washington, DC: Georgetown University Press.
- Hofmeister, Philip. 2007. Retrieval and gradience in filler-gap dependencies. In *Proceedings of the 43rd Regional Meeting of the Chicago Linguistic Society*.
- Kaan, Edith. 2007. Event-related potentials and language processing: A brief overview. *Language and Linguistics Compass*, 1/6, 571-591.
- Kaan, Edith and Tamara Y. Swaab. 2003. Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience*, 15, 90-110.
- Keller, Frank. 2000. *Gradience in Grammar: Experimental and Computational Aspects of Degrees of Grammaticality*. PhD dissertation, University of Edinburgh.
- Keller, Frank and Antonella Sorace. 2003. Gradient auxiliary selection and impersonal passivization in German: an experimental investigation. *Journal of Linguistics*, 39, 57-108.
- Kluender, Robert and Marta Kutas. 1993. Subjacency as a processing phenomenon. *Language and Cognitive Processes*, 8, 573-633.
- Lakoff, George. 1973. Fuzzy grammar and the performance/competence terminology game. In Claudia Corum, T. Cedric Smith-Stark and Ann Weiser, eds., *Papers*

from the Ninth Regional Meeting, Chicago Linguistic Society, 271-291.

Chicago: Chicago Linguistic Society.

Labov, William. 1975. Empirical foundations of linguistic theory. In Robert Austerlitz, ed., *The Scope of American Linguistics*, 77-133. Lisse: Peter de Ridder.

Lasnik, Howard and Mamoru Saito. 1984. On the nature of proper government. *Linguistic Inquiry*, 15, 235-289.

Levelt, W.J.M. 1974. *Formal Grammars in Linguistics and Psycholinguistics. Volume III: Psycholinguistic Applications*. The Hague: Mouton.

Levelt, W.J.M., J.A.W.M. Van Gent, A.F.J. Haans and A.J.A. Meijers. 1977. Grammaticality, paraphrase and imagery. In Sidney Greenbaum, ed., *Acceptability in Language*, 88-101. The Hague: Mouton.

Linebarger, M., M. Schwartz and E. Saffran. 1983. Sensitivity to grammatical structure in so-called agrammatic aphasics. *Cognition*, 13, 361-392.

Lukatela, K., S. Crain and D. Shankweiler. 1988. Sensitivity to inflectional morphology in agrammatism: investigation of a highly inflected language. *Brain and Language*, 33, 1-15.

McDaniel, Dana and Wayne Cowart. 1999. Experimental evidence for a minimalist account of English resumptive pronouns. *Cognition*, 70, B15-24.

McDonald, Janet L. 2006. Beyond the critical period: processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, 55, 381-401.

Meurers, W. Detmar. 2007. Advancing linguistics between the extremes: Some thoughts on Geoffrey R. Sampson's "Grammar with grammaticality". *Corpus Linguistics and Linguistic Theory*, 3-1, 49-55.

- Murasugi, Kumiko G. 2008. Asymmetric movement in raising-to-object structures. *Canadian Journal of Linguistics*, 53, 85-100.
- Nagata, Hiroshi. 1988. The relativity of linguistic intuition: The effect of repetition on grammaticality judgments. *Journal of Psycholinguistic Research*, 17, 1-17.
- Newmeyer, Frederick J. 1983. *Grammatical Theory: Its Limits and Its Possibilities*. Chicago: Chicago University Press.
- Osterhout, Lee. 1994. Event-related brain potentials as tools for comprehending language comprehension. In Charles Clifton, Jr., Lyn Frazier, and Keith Rayner, eds., *Perspectives on Sentence Processing*, 15-44. Hillsdale, NJ: Lawrence Erlbaum.
- Osterhout, Lee and Phillip J. Holcomb. 1992. Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31, 785-806.
- Osterhout, Lee and Janet Nicol. 1999. On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies. *Language and Cognitive Processes*, 14, 283-317.
- Osterhout, Lee and Linda A. Mobley. 1995. Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language*, 34, 739-773.
- Osterhout, Lee, M. Bersick and J. McLaughlin. 1997. Brain potentials reflect violations of gender stereotypes. *Memory and Cognition*, 25, 273-285.
- Osterhout, Lee, Phillip J. Holcomb and David A. Swinney. 1994. Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 786-803.

- Pesetsky, David. 1987. *Wh-in-situ: Movement and unselective binding*. In E. Reuland and A.G.B. ter Meulen, eds., *The Representation of (In)definiteness*, 98-128. Cambridge: MIT Press.
- Phillips, Colin and Howard Lasnik. 2003. Linguistics and empirical evidence: Reply to Edelman and Christiansen. *Trends in Cognitive Science*, 7, 61-62.
- Rizzi, Luigi. 1990. *Relativized Minimality*. Cambridge: MIT Press.
- Ross, John Robert. 1987. The category squish: Endstation Hauptwort. In Paul M. Peranteau, Judith N. Levi and Gloria C. Phares, eds., *Papers from the Eighth Regional Meeting, Chicago Linguistic Society*, 316-328. Chicago: Chicago Linguistic Society.
- Ross, John Robert. 1987. Islands and syntactic prototypes. In Barbara Need, Eric Schiller and Anna Bosch, eds., *CLS 23: Papers from the 23rd Annual Regional Meeting of the Chicago Linguistic Society*, 309-320. Chicago: Chicago Linguistic Society.
- Sag, Ivan A., Philip Hofmeister and Neal Snider. 2008. Processing complexity in subjacency violations: The Complex Noun Phrase Constraint. In *Proceedings of the 43rd Annual Meeting of the Chicago Linguistic Society*.
- Sampson, Geoffrey R. 2007. Grammar without grammaticality. *Corpus Linguistics and Linguistic Theory*, 3-1, 1-32.
- Schütze, Carson T. 1996. *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Chicago: University of Chicago Press.
- Schwartz, M.F., M.C. Linebarger, E.M. Saffran and D.S. Pate. 1987. Syntactic transparency and sentence interpretation in aphasia. *Language and Cognitive Processes*, 2, 85-113.

- Snyder, William. 2000. An experimental investigation of syntactic satiation effects. *Linguistic Inquiry*, 31, 575-582.
- Spencer, N.J. 1973. Differences between linguists and nonlinguists in intuitions of grammaticality-acceptability. *Journal of Psycholinguistic Research*, 2, 83-98.
- Sprouse, Jon. 2007. Continuous acceptability, categorical grammaticality, and experimental syntax. *Biolinguistics*, 1, 117-128.
- Sprouse, Jon. 2008. The differential sensitivity of acceptability to processing effects. *Linguistic Inquiry*, 39, 686-694.
- Tokowicz, Natasha and Brian MacWhinney. 2005. Implicit and explicit measures of sensitivity to violations in second language grammar. *Studies in Second Language Acquisition*, 27, 173-204.
- Wassenaar, Marlies and Peter Hagoort. 2005. Word-category violations in patients with Broca's aphasia: An ERP study. *Brain and Language*, 92, 117-137.
- Wassenaar, Marlies, Colin M. Brown and Peter Hagoort. 2004. ERP effects of subject-verb agreement violations in patients with Broca's aphasia. *Journal of Cognitive Neuroscience*, 16, 533-576.
- Watt, W.C. 1975. The indiscreteness with which impenetrables are penetrated. *Lingua*, 37, 95-128.
- White, Lydia. 1989. *Universal Grammar and Second Language Acquisition*. Amsterdam: Benjamins.
- Wulfeck, Beverly. 1988. Grammaticality judgments and sentence comprehension in agrammatic aphasia. *Journal of Speech and Hearing Research* 31: 72-81.
- Wulfeck, Beverly and Elizabeth Bates. 1991. Differential sensitivity to errors of agreement and word order in Broca's aphasia. *Journal of Cognitive Neuroscience*, 3, 258-272.

