RESPONSE

# Real bad grammar:
# Realistic grammatical description
# with grammaticality*

JENNIFER  FOSTER

## 1. Introduction

Sampson (this issue) argues for a concept of "realistic grammatical description" in which the distinction between grammatical and ungrammatical sentences is irrelevant. In this article I also argue for a concept of "realistic grammatical description" but one in which a binary distinction between grammatical and ungrammatical sentences is maintained. In distinguishing between the grammatical and ungrammatical, this kind of grammar differs from that proposed by Sampson, but it does share the important property that invented sentences have no role to play, either as positive or negative evidence.

  Our propensity to make mistakes, and the fact that many people are forced to speak and write in a language that is not their native one means that sentences are produced which contain grammatical errors. These naturalistic ungrammatical sentences, as opposed to the invented starred examples often used within the linguistics community, have been dismissed as uninteresting. Although I do not wish to give naturalistic ungrammatical sentences the prominence given by Carnie (2002) to invented ungrammatical sentences when he suggests that it is necessary to determine the ungrammatical sentences in a language in order to determine the grammatical ones (see Sampson, this issue), I do, however, think that naturalistic ungrammatical sentences are of interest to linguists studying language production, language loss and language learning, and that the grammatical/ungrammatical distinction cannot therefore be completely dismissed. Also, for grammar development within the field of natural language processing, the grammatical/ungrammatical distinction cannot be ignored or denied because this can lead to the development of grammars which do not accurately analyse ungrammatical sentences. This article focuses particularly on this second argument

in favour of maintaining the grammatical/ungrammatical distinction, and when I speak of grammar development, I am particularly thinking of those large scale natural language grammars which are used to automatically parse natural language.

Linguistic evidence in the form of grammaticality judgements can be used to distinguish grammatical sentences from ungrammatical ones but, crucially, these judgements should be made only on naturalistic data *in context*. Sampson (this issue) argues that Chomsky's conception of language as a set of sentences, with the role of a linguist to establish which strings are in this set, is unhelpful because it focuses undue attention on the grammatical/ungrammatical distinction: I believe that an unfortunate consequence of this definition of language is that it places too much emphasis on the sentence as an isolated unit.

## 2. Grammars for natural language processing

A fundamental debate within the linguistics community has concerned what it is a grammar is supposed to model: should a grammar model competence or performance? Should a grammar reflect a psychological reality or a social reality? Lamb (2000), for example, distinguishes between a "theory of the linguistic extension" which is a theory of the utterances produced by a speaker or community, and a "theory of the linguistic system" which is a theory of the human cognitive system which is capable of producing and understanding such utterances. In the practical domain of natural language processing, there is no such debate. The grammar of a computer parser which is to form part of a practical application must be a theory of the linguistic extension and must describe the productions of a speech community. In proposing the competence/performance distinction, Chomsky remarked that the language produced by a speech community is rife with slips and imperfections (Chomsky 1961: 130−131). Therefore, if a computer parser has to accurately parse actual language, it will have to accurately parse imperfect language, in particular the kind of imperfect language that we routinely produce *and* are capable of understanding. It will be able to do this if it is equipped with some knowledge of deviant sentence structures.

A precision grammar distinguishes between the grammatical and the ungrammatical and purposely describes only grammatical sentences. An example is the English Resource Grammar (ERG) (Copestake and Flickinger 2000), a broad coverage HPSG grammar of English. Baldwin et al. (2004) make the point that, if a grammar is to form the basis of a natural language processing system which performs not just sentence parsing but also sentence generation, it should not be able to generate ungrammatical sentences. A parser using such a grammar will reject un-

grammatical sentences outright. However, a parser which gives the response "no" or "ungrammatical" to a sentence such as (1),[1] may be capable of distinguishing between the grammatical and ungrammatical but of what practical use is this ability if it cannot hint at the meaning of an utterance whose ill-formedness is quite commonplace?

(1)  Want to *saving* money?

Of course, one could argue that robust parsing techniques (such as constraint relaxation (Fouvry 2003) or parse-fitting (Penstein Rosé and Lavie 1997)) could be employed to handle ungrammatical sentences but such techniques will be more effective if they are tailored to specific types of ungrammatical language − a natural extension of this is then to actually let the grammar describe the structure of ungrammatical sentences in the same way that it describes the structure of grammatical sentences.

A parser whose grammar is derived automatically from a treebank of naturalistic sentences is unconcerned with whether or not a sentence is grammatical. Typically, grammaticality is assumed, and this assumption will be quite accurate if the treebank sentences come from a high-quality newspaper such as *The Wall Street Journal*. The fact that such grammars do not purposely set out to exclude ungrammatical sentences together with the fact that such grammars are generally based upon a large body of data means that parsers equipped with such grammars are quite likely to return a parse for an ungrammatical sentence. However, since such a parser does not have a concept of ungrammaticality, it will not be aware
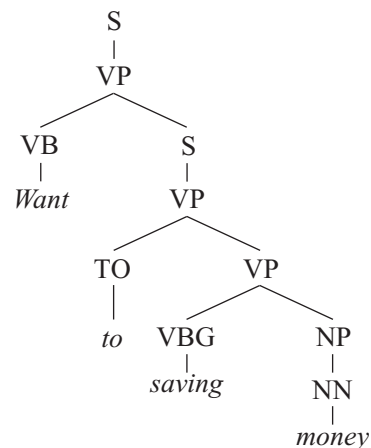


Figure 1. *Parse returned by Charniak's parser for sentence (1)*

that there is something deviant about the sentence, with the result that the parse it produces for the sentence will not necessarily be the correct one, that is, it will not necessarily reflect what the person who produced the ungrammatical sentence intended to express. For example, Charniak's most recent parser[2] (Charniak 2000) will provide the reasonable parse in Figure 1 for sentence (1) but it is less successful, for example, on the ungrammatical (2), returning the parse in Figure 2.

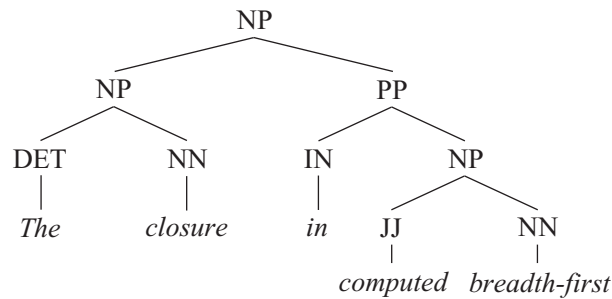(2)   The closure in computed breadth-first.



Figure 2.  *Parse returned by Charniak's parser for sentence (2)*

## 3.  Grammar requirements

The following are the requirements for the type of grammar which I believe should be developed by computational linguists and used by a parsing system:

1. The grammar should have a component which describes the structure of the grammatical sentences that occur in language.
2. The grammar should have a component which describes the structure of the ungrammatical sentences that occur in language.

Like a treebank grammar, this grammar aims to be a direct reflection of language rather than an indirect inflection via linguistic intuition. However, unlike a treebank grammar, this grammar does explicitly distinguish between the grammatical and the ungrammatical, and this distinction relies on linguistic intuition. This distinction is binary, but this does not mean that the rules in each component of the grammar cannot be probabilistic. A linguistic structure described in the first grammatical component of the grammar could be assigned a probability based on

how frequently this structure appears in grammatical data. Similarly, a linguistic structure described in the ungrammatical component of the grammar could be assigned a probability based on how frequently this structure shows up in ungrammatical data. The grammatical component of the grammar is quite similar to a precision grammar which has been tested using corpus evidence. An example is the afore-mentioned ERG which has been tested using sentences from the British National Corpus (Baldwin et al. 2004). The ungrammatical component is, of course, not implemented by a precision grammar.

What kind of evidence is needed in order to develop and test the second component of the grammar, the part of the grammar which describes ungrammatical sentences? Since this grammar is to form the basis of a parsing system, its description of ungrammaticality must reflect the kind of ungrammaticality that actually occurs in language. This means that naturalistic ungrammatical sentences will be needed as evidence rather than imagined ones. Baldwin et al. (2004) argue that naturalistic ungrammatical sentences such as (1) or (2) constitute "haphazard noise" and are useful only to test that a grammar does not overgenerate. I am arguing that a grammar that is capable of generating the kind of ungrammatical sentences that people actually produce, is not guilty of overgeneration, provided that the grammar knows that these kinds of sentences are ungrammatical. Therefore, to test the second ungrammatical component of the grammar, which is essentially a theory of real ungrammaticality, it is necessary to collect a corpus of naturalistic sentences which are considered by speakers of the language to be ungrammatical.

How does this definition of grammar relate to the one suggested by Sampson (this issue)? The two are broadly in agreement since they aim to describe language as it is actually used and both reject the need for negative evidence in grammar development. According to Sampson (2001, and footnote 3, this issue), if a grammar is constructed so that it excludes sentences whose structure has not actually been observed, then negative evidence becomes irrelevant. In order to exclude a sentence from the grammar, it is not necessary to verify that it is ungrammatical. It is enough not to have observed the sentence in practice. Once the sentence is observed, then this observation has the potential to count as a refutation of a grammar which excludes the sentence, and the grammar will need to be modified accordingly. As Sampson notes, this is Popper's view of the nature of a scientific theory: it should maximize the number of statements it makes which are refutable by observable evidence.[3]

Where the two notions of grammar differ is in their treatment of situations when an ungrammatical sentence such as (1) (repeated for convenience as (3)) is actually observed.

(3)  Want to *saving* money?

(4)  Want to *save* money?

(5)  Want to *start* saving money?

Surely, if such a string is observed, it should serve as a refutation of any grammar which prohibits it and the grammar in question would need to be modified so that it no longer prohibits this sentence? Sampson (2001) argues that the grammar should not be changed to accommodate such an observation since our knowledge that people make mistakes in language (such as omitting a word, using the wrong verb form, etc.) should allow us to relate this sentence to another sentence such as (4) or (5), both of which are accommodated by the grammar, thereby discounting the observation as a genuine refutation. The ungrammatical sentence (3) would, however, be included in the grammar described here, although it would still be recognised as a different kind of observation to a sentence such as (4) or (5) and thus would be included in the second ungrammatical component of the grammar. Recognizing it as a different kind of observation is the same thing as making a grammaticality judgement, and a method to make this kind of judgement as reliably as possible is described in the next section.

The type of grammar suggested by Sampson could actually be used as the grammatical component of the grammar advocated here. It would include rare and odd constructions (Sampson's *Dunster* constructions), and if it was a probabilistic grammar it could encode rareness, without linking this rareness in any way to grammaticality. In fact, because of the clear distinction between the two components of the grammar, the concepts of grammaticality and frequency are not conflated. This non-conflation is a positive thing, regardless of where one stands with respect to Sampson's claim that frequency data cannot be used to predict grammaticality status (see Sampson's discussion of noun phrase variability in the SUSANNE treebank, this issue).

## 4. Judging grammaticality

The use of grammaticality judgements as linguistic evidence has always been controversial. A large body of literature spanning several decades casts doubt on the validity of grammaticality judgements (see for example Labov 1972; Derwing 1979; Schütze 1996). These critiques cover various problems with the judgement process: defining grammaticality, choice of informant, the measurement scale used to measure judgements and the role of sentential context. After concluding that the grammaticality of a sentence cannot be inferred from its frequency, Sampson (this

issue) dismisses as scientifically suspect the alternative method of using grammaticality judgements. The fact that it is difficult to reliably infer grammaticality is one of the arguments he uses to support his claim that the concept of grammaticality should be more or less ignored in grammar development. I disagree to a certain extent: in building treebanks (which are now a fundamental ingredient for natural language processing), we rely on the linguistic intuition of the treebank annotators to parse sentences, and I think that it would be useful to view the grammaticality judgement process in a similar way − as a necessary evil. Grammaticality judgements, although undoubtedly problematic, can be used to effectively carve out a grammatical/ungrammatical distinction (albeit not a particularly exciting one), and I focus for the rest of the article on how this might be achieved, dealing particularly with the problems of sentential context and defining what it means for a sentence to be ungrammatical.

## 5. Defining "ungrammatical"

In order to observe instances of ungrammatical language in the language that one encounters as a part of everyday life, it is necessary to be able to tell whether something is actually ungrammatical. It is clear that a definition of the term "ungrammatical" is needed in order to be able to make consistent grammaticality judgements. The concept of "ungrammatical" is, of course, a slippery one: it can be used in an everyday and in a theoretical sense, and coming up with a definition of the term will depend on what kind of linguistic information one believes should fall under the remit of grammar. Thus, a definition such as Definition 1 is unhelpful because it merely shifts the problem onto the grammar.

**Definition 1.** Possible Definition of "Ungrammatical"
*A sentence is ungrammatical if it cannot be generated by the grammar rules of the language.*

To overcome this problem an operational definition which avoids the term "grammar" completely is needed. The definition has two conditions which must be satisfied for a sentence to be considered ungrammatical. It is given as follows:

**Definition 2.** Definition of "Ungrammatical"
*A sentence is ungrammatical if*

1. *It contains an error.*
2. *All the individual words in the sentence are well-formed words of the language in question.*

The first condition states that an ungrammatical sentence must contain an error. The word "error" is used here to mean either a mistake which occurs as an oversight or a language-learning error which occurs as a result of a lack of knowledge of the language. The second condition in Definition 2 focuses attention on the inter-word level rather than the intra-word level, which means that a sentence such as (6) will satisfy this condition for ungrammaticality whereas a sentence such as (7) will not.

(6)  Attempts for this seem to have *gone to* far.

(7)  *Atttempts* for this seem to have gone too far.

This definition of "ungrammatical" is intended to be independent of any particular linguistic theory and should be applicable by both linguists and non-linguists. Of course the decision over whether a sentence contains an error is a subjective one because it depends on a person's opinion about what constitutes an error but I would argue that some level of subjectivity is unavoidable in any linguistic task whether that be applying a grammaticality judgement or building a treebank (see Sampson and Babarczy 2003 for a related discussion on the problem of quantifying annotator variation and precision in treebank development). It is possible, for example, that a person who has been educated in the 1950s or before may genuinely consider that a clause ending in a preposition is erroneous, whereas someone who has been educated in the 1990s would probably not even be aware that such a construction was once frowned upon. The letters page of a Dublin newspaper from August 2003 contains a letter decrying as "unacceptable" the newspaper's use of the word *And* to start a sentence. To overcome this problem, which I think will exist no matter how the word "ungrammatical" is defined, it might be necessary to produce a set of guidelines to accompany the definition, in the same way that treebank annotators work with a set of bracketing guidelines (e. g., Bies et al. 1995). Such guidelines could suggest, for example, that a sentence such as (8) does not contain an error, whereas a sentence such as (9) does.

(8)  One of the clauses, which we have a big problem *with*, is that they own the intellectual property.

(9)  One of the clauses, which *for* we a big problem with, is that they own the intellectual property.

## 6. Sentential context

Chomsky's definition of a language as a set of sentences (Chomsky 1957: 13) emphasized the sentence as an individual unit of investigation, and

grammaticality judgements have usually been made on sentences appearing in isolation from a discourse context. In order to explain why sentential context cannot be ignored, it is useful to consider the information being judged when judging a sentence's grammaticality. The aspect of a sentence that is being judged is the sentence's structure, which is one aspect of the sentence's meaning. If a valid structure for the sentence can be determined, it will be accepted as grammatical. In a sentence like Chomsky's *Colorless green ideas sleep furiously*, the structure is the only part of the sentence's overall meaning which is transparent. For some sentences, the structure is less obvious and only becomes clear when other aspects of the sentence's meaning are in place. An example of such a sentence is *He me* given by Van Dijk (1976: 44). Appearing on its own this sentence may be judged ungrammatical but appearing as part of the discourse in (10), its meaning can be determined.

(10) a. *Did you hit him?*
    b. *No. He me.*

Knowing its meaning implies knowing its structure. A different sentential context can lead to the rejection of the same utterance:

(11) a. *Where is he going?*
    b. *I don't know. He me.*

In this case, it is very possible that the utterance made by a does not lead to an understanding of b's *He me* response. We could guess that what is meant is the sentence *He didn't tell me* but our linguistic knowledge tells us that, unlike *hit* in the previous context, *didn't tell* cannot be omitted from the sentence in this context. Thus, a structure which fits in with the surrounding context *and* with our linguistic knowledge cannot be found, and the utterance must be rejected.

Chomsky (1964) and Ziff (1964) distinguish between those sentences which are interpretable out of context and those which need a context in order to be interpreted. Chomsky (1964) terms the former "grammatical" and the latter "semi-grammatical". A similar position is taken by those who view grammaticality as a graded rather than a dichotomous property (Lakoff 1973; Aarts et al. 2004). It could be argued that the binary grammatical/ungrammatical distinction could be maintained if the mathematical view of language as a set of sentences was rejected and replaced by one in which sentences only exist as part of a real communication. Viewing language as a set of sentences forces us to decide upon the set membership of an utterance such as *He me* when it is clear that it belongs both to the set of sentences (when appearing in the discourse

context (10)) and to the set of non-sentences (when appearing in the discourse context (11)). According to Newmeyer (1983: 55), the unclear cases are those that "with a little imagination can be given a context in which they sound perfectly acceptable." If this is true, supplying a context along with a sentence takes the imagination element out of the judgement process and should result in fewer unclear cases. A similar point was made by Lakoff (1971), Snow (1975) and Crain and Steedman (1985). It seems obvious, therefore, that when a sentence's grammaticality is being judged, its sentential context must be available. With this in mind, the ungrammatical sentences which could be used to inform the ungrammatical component of the grammar proposed in this paper should be observed and then judged in their natural context as part of the basic linguistic process of understanding written language, i.e., reading.

Another advantage of collecting instances of ungrammaticality in context is that it allows some sentences to be included as evidence which might be excluded if they were encountered out of context. This kind of sentence needs a context to make it ungrammatical and is, therefore, the opposite of the elliptical *He me* case discussed which needed a context to be considered grammatical. An example is (12):

(12)  We can order then directly from the web.

Taken out of context, it is not difficult to imagine a context in which the verb *order* is occurring without an object and is being modified by the adverb *then*. Little room is left for imagination when the sentence is viewed in its discourse context which is an email discussion about buying a present for someone. In this context it is clear that the word *then* is really a mistyped form of the object pronoun *them*.

## 7.  Applying the method

Between July 2002 and January 2004 I applied Definition 2 to the sentences I encountered while reading, resulting in a corpus of ungrammatical written English. The corpus contained 923 sentences and approximately 20,000 words (Foster and Vogel 2004; Foster 2005). In constructing this corpus, I occasionally encountered sentences which were not ungrammatical according to Definition 2, but which I think are worthy of mention.

Despite the fact that the sentences are encountered in context, it is still possible (although rare) to find a sentence which is ambiguous between a grammatical and ungrammatical reading. An example is the following which appeared in an email discussion between friends debating restaurant choices:

(13) I hear that that Antonio's is a nice place but I'm sure Peter wants to go there to remind him of Erasmus!!

In this example it is quite probable that the second *that* is a duplication error and that the sentence is ungrammatical according to Definition 2. However, this does not represent a clear case because there is a second contextually plausible reading in which the second *that* is a demonstrative determiner and *that Antonio's* is a noun phrase.

Another example is the (in my opinion) stylistically awkward (14), which appeared in a newspaper article:

(14) We even offer the older members to mark their bingo cards for them while they go outside and have a cigarette.

(15) We even offer to mark the older members' bingo cards for them while they go outside and have a cigarette.

While I might prefer (14) to be rewritten as (15), it would be misguided to conclude that an error has occurred. There is something about this particular ditransitive use of the verb *offer* in (14) that I find not quite right, but it is not clear to me that my opinion would be shared by others. As Sampson (this issue) points out, we all have our own experiences of language which lead us to use certain constructions and avoid others. A similar case is Sentence (16) which appeared in an email written by a non-native English speaker:

(16) I am Tamara who is doing a Phd in genetics.

(17) I am Tamara and I am doing a Phd in genetics.

The problem with (16) is that it seems like an unnatural way of expressing what a native speaker would express using a sentence such as (17). It is the use of the non-restrictive relative clause which marked the sentence as odd, but as with (14), having a vague sense that a sentence is in someway abnormal is not the same thing as knowing for definite that an error has occcured. Sentences (14) and (16) are similar to the *Dunster* constructions discussed by Sampson (this issue): they should be included in the first grammatical component of the grammar proposed in this article, rather than the ungrammatical component.

Thus, the idea in applying this type of grammaticality judgement is to judge as ungrammatical only those utterances where one is certain that an error has occurred, and not sentences which a sub-editor or a language teacher might rephrase. Applying the definition to build a corpus of ungrammatical English sentences (Foster 2005) resulted in ungram-

matical sentences which could be described as mundane − approximately 90% could be corrected by just one application of a correction operator (inserting, replacing or deleting a word).[4] This concept of ungrammaticality is close to that labelled "impoverished" by Sampson (this issue). I would argue, however, that an impoverished notion is better than a confused one which results from judging sentences out of context and is better than no concept at all.

## 8. Conclusion

In this article I have argued that an empirical distinction between grammatical and ungrammatical sentences can by maintained by applying grammaticality judgements to real sentences in their real context. The distinction which results can be used as evidence for a grammar which would allow parsers to accurately parse the kind of ungrammatical sentences that people typically produce. To extend the "sentences as roads" analogy used by Sampson (this issue), the ungrammatical sentences that I think should be included in grammatical description could be seen as roads (some less travelled than others) with potholes in them.

*Dublin City University*

## Notes

  * Author's e-mail address: <jfoster@computing.dcu.ie>.
  1. Unless otherwise stated, all example sentences are taken from a corpus of ungrammatical sentences collected by the author, or abbreviated and/or anonymised versions of the corpus sentences (see Foster 2005 for a detailed description of this corpus).
  2. Downloaded as *reranking-parserJune06.tar.gz* from ftp://ftp.cs.brown.edu/pub/nlparser/ in July 2006.
  3. Stefanowitsch (2006) argues that a corpus can be a source of negative evidence, contrary to the traditional generative grammar claim that it cannot. By applying collostructional analysis to corpus data, accidentally absent constructions can be distinguished from significantly absent ones, and those which are significantly absent can be considered to be ungrammatical. This is a different position to the one taken by Sampson, who argues that negative evidence is not actually required.
  4. A notable exception is the following sentence which occurs in a message which used to pop up on my laptop screen when the laptop was either connected to or disconnected from the main power supply: *LaptopsRUs Hotkey Utility show the indicators on your display and save brightness adjustments each power supplying conditions.*

## References

Aarts, Bas, David Denison, Evelien Keizer and Gergana Popova (eds.)
  2004        *Fuzzy Grammar: A Reader*. Oxford University Press.

Baldwin, Timothy, John Beavers, Emily M. Bender, Dan Flickinger, Ara Kim and
Stephan Oepen
2004    Beauty and the Beast: What running a broad-coverage precision gram-
        mar over the BNC taught us about the grammar − and the corpus. In:
        Stephan Kepser and Marga Reis (eds.), *Pre-Proceedings of the Interna-
        tional Conference on Linguistic Evidence: Empirical, Theoretical and Com-
        putational Perspectives.* Tübingen: ●, 21−26.

Bies, Ann, Mark Ferguson, Karen Katz, and Robert MacIntyre
1995    Bracketing Guidelines for Treebank II Style, Penn Treebank Project.
        *Technical Report MS-CIS-95-06*, University of Pennsylvania, Philadel-
        phia, PA.

Carnie, Andrew
2002    *Syntax: A Generative Introduction.* Oxford: Blackwell.

Charniak, Eugene
2000    A maximum-entropy-inspired parser. *Proceedings of the Annual Meeting
        of the North American Association for Computational Linguistics
        (NAACL-00),* 132−139.

Chomsky, Noam
1957    *Syntactic Structures.* Mouton, The Hague.
1961    Formal discussion. The development of grammar in child language. Re-
        printed 1971 in J. P. B. Allen and P. van Buren (eds.), *Chomsky: Selected
        Readings*, Oxford: Oxford University Press, 129−134.
1964    Degrees of Grammaticalness. In: Jerry A. Fodor and Jerrold J. Katz,
        384−389. [Also appeared 1961 as: Some methodological remarks on gen-
        erative grammar, *Word* (17), 219−239. Reprinted in Aarts et al. (2004),
        321−325.]

Copestake, Ann and Dan Flickinger
2000    An open-source grammar development environment and broad-coverage
        English grammar using HPSG. *Proceedings of the 2nd International Con-
        ference on Language Resources and Evaluation (LREC-02),* Athens,
        Greece.

Crain, Stephen and Mark Steedman
1985    On not being led up the garden path: the use of context by the psycholog-
        ical syntax processor. In: David D. Dowty, Lauri Karttunen, and Arnold
        M. Zwicky (eds.), *Natural Language Parsing: Psychological, Computa-
        tional and Theoretical Perspectives.* Cambridge: Cambridge University
        Press, 320−358.

Derwing, Bruce L.
1979    Against autonomous linguistics. In: Thomas A. Perry (ed.), *Evidence and
        Argumentation in Linguistics.* Berlin and New York: Mouton de Gruyter,
        163−189.

van Dijk, Teun A.
1976    Acceptability in context. In: Sidney Greenbaum (ed.), *Acceptability in
        Language.* The Hague: Mouton, 39−62.

Fodor, Jerry A. and Jerrold J. Katz (eds.)
1964    *The Structure of Language: Readings in the Philosophy of Language.* ●:
        Prentice-Hall.

Foster, Jennifer
2005    Good reasons for noting bad grammar: Empirical investigations into
        the parsing of ungrammatical written English. Unpublished dissertation,
        University of Dublin, Trinity College.

Foster, Jennifer and Carl Vogel
  2004    Parsing ill-formed text using an error grammar. *Artificial Intelligence Review* 21, 269−291.
Fouvry, Frederik
  2003    Robust processing for constraint-based grammar formalisms. Unpublished dissertation, Department of Language and Linguistics, University of Essex.
Labov, William
  1972    *Sociolinguistic Patterns.* ●: University of Pennsylvania Press.
Lakoff, George
  1971    Presupposition and relative well-formedness. In: Danny D. Steinberg and Leon A. Jakobovits (eds.), *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics and Psychology.* Cambridge: Cambridge University Press, 329−340.
  1973    Fuzzy grammar and the performance/competence terminology game. *Papers from the Ninth Regional Meeting of the Chicago Linguistic Society,* 52−59.
Lamb, Sidney
  2000    Bidirectional processing in language and related cognitive systems. In: Michael Barlow and Suzanne Kemmer (eds.), *Usage-Based Models of Language.* Stanford, California: CSLI Publications, 87−120.
Newmeyer, Frederick J.
  1983    *Grammatical Theory, its Limits and its Possibilities.* Chicago: The University of Chicago Press.
Penstein Rosé, Carolyn and Alon Lavie
  1997    An efficient distribution of labor in a two stage robust interpretation process. *Proceedings of 1997 Conference on Empirical Methods in Natural Language Processing (EMNLP-97),* 1129−1135.
Sampson, Geoffrey
  2001    Objective evidence is all we need. In: *Empirical Linguistics.* New York: Continuum, 122−141.
Sampson, Geoffrey and Anna Babarczy
  2003    Limitations to annotation precision. *Proceedings of the 4th International Conference on Linguistically Interpreted Corpora,* 61−68.
Schütze, Carson T.
  1996    *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology.* Chicago: The University of Chicago Press.
Snow, Catherine E.
  1975    Linguists as Behavioral Scientists: Towards a Methodology for Testing Linguistic Intuitions. In: A. Kraak (ed.), *Linguistics in the Netherlands 1972−1973.* Assen: Van Gorcum, 271−275.
Stefanowitsch, Anatol
  2006    Negative evidence and the raw frequency fallacy. *Corpus Linguistics and Linguistic Theory,* 2(1), 61−77.
Ziff, Paul
  1964    On understanding 'Understanding Utterances'. In Fodor and Katz (1964), 390−399.