Theoretical Linguistics 2007, 33: 381-394

As a matter of facts – comments on Featherston's sticks and carrots[1]

Hubert Haider, Universität Salzburg, FB Linguistik (2006)

It is easy to agree and equally easy to disagree with the two main positions of Sam Featherston (2007), respectively, namely, that grammar theory would benefit from improving standards of data assessment, and, that syntactic well-formedness is inherently gradient rather than dichotomist. I subscribe to the first position and I see good grounds for not subscribing to the second claim. Since the two positions are dependent, denying the second has implications for accepting the first.

Data handling in Generative Grammar is far from satisfactory, both in terms of observational and descriptive adequacy. Observational adequacy is at issue when (a small set of) examples is meant to be representative for a whole class of expressions that are assumed to constitute the extension of the property the example sentence is meant to display. Descriptive adequacy is at stake when the class property is paired with a grammatical analysis that is intended to capture the grammatical causality of the given property of the class.

A failure on the observational level is in most cases the misidentification of a side effect as the main effect. In the case of deviant sentences, the source of the deviance is misidentified. Wellformedness is always a conjunction of multiple properties [$p \wedge q \wedge r \wedge ...$]. If a sentence is deviant, this amounts to the falsity of the conjunction. A conjunction is false if one or more of the conjuncts are false. A failure on the observational level happens when the source of the deviance of a given example is mistaken to be a failure of , let us say, property P_i, but in reality, it is the effect of an overlooked property P_j. It is evident that careful evaluation and testing of data is the proper measure for reducing this kind of mistakes.

Missing the level of descriptive adequacy is missing the adequate grammatical generalizations for a given set of facts. This is an everyday problem of a scientist. You hypothesize, but your hypothesis happens to be not fully adequate. There are relevant data that do not fit. This is a critical point. You may either heroically give up your dearly fostered hypothesis (unlikely), or you stick to it (very likely) and try to protect it by replacing or adding subsidiary assumptions. This is where observational adequacy reappears. Since you are desperate to find exonerating evidence, you might be inclined to call up witnesses that a fair judge in a fair trial would have to reject since the witness does not meet the criterion of observational truthfulness.[2]

The diagnosis seems to be clear in such a case, but even if the problem is identified, the adequate remedy may be not at all clear. Investing energy into the diagnosis helps avoiding a wrong diagnosis but is does not guarantee a solution of the problem. Let me illustrate this with two cases discussed in Featherston's paper. One concerns descriptive adequacy (alleged *that-t* effects in German), the other concerns observational adequacy (*nicht XP/XP nicht* tags).

Does German display a '*that-t* effect'? (see Featherston, p.6). Note that in this case we focus on the theory-to-data fit and not so much on the data-to-analysis perspective. So, we have to decide first what would constitute the primary evidence and second, we have to check whether the data we adduce are indeed representative for the property at issue. According to

[1] Peer commentary on Featherston (2007).
[2] More often, the problem is solved pragmatically. Conflicting evidence is simply ignored or repressed.

(Chomsky 1981:243), (1a) is "a typical example of the *[that-t] effect." The German counterpart of (1a) is (1b). Does (1b) suffer from a *that-t* violation?

(1) a. Who_i do you think [(*that) e_i left]?
b. Wer_i hast du gedacht [dass e_i weggegangen ist?]
who have you thought [that left has]

From a particular theoretic point of view (Haider 2003), it is predictable that German is not subject to a *that-t* constraint. This constraint applies to a trace in a spec-position that are immediately[3] preceded by a complementizer. In German, as in other OV languages, the subject may remain in its VP-internal position. So, the *that-t* constraint would not apply.

Featherston points to his study from 2003 that registered a difference between long distance subject extraction and object extraction. But, would this very difference constitute evidence for the presence of this effect in German? The absence of the *that-t* constraint in German must not be equated with the absence of any difference between long-distance subject vs. object extraction. There is a difference, but its source is not a constraint of grammar. Its source is a parsing impediment. If the subject is extracted, a stimulus sentence like (1b) starts with a potential subject. So, the parser identifies the '*wer*' (who_{Nom}) in (1b) as a subject pronoun and next, it runs into a preliminary clash with the agreement features of the finite matrix verb, if its features do not match. If they match, this just postpones the clash, since the spurious match will have to be cancelled again once the matrix subject is met and processed. So, in any case, for a subject fronted out of an embedded clause, the parser meets a temporary obstacle. Object extraction does not inflict this intermediate mismatch or spurious match problem. Hence, there is a parsing difference between long distance subject extraction and non-subject extraction. A source of evidence for the role of the subject bias[4] in on-line processing is EEG research (see Bornkessel et al. 2002, Friederici et al. 2003, Schlesewsky et al. 2003).

A measurable difference in the results (s. Featherston's figure 2) between subject and object extraction is expected, but it does not by itself warrant the conclusion that German indeed embodies a *that-t* constraint.[5] The observed difference could be a mere side effect of an independent property of subjects, namely the obligatory subject-verb agreement. What would careful data management mean in this case?

If German was indeed subject to the same constraint as English, we would predict a systematic subject-object asymmetry for all kinds of extraction constructions that leave a trace in the subject position of a C-introduced clause. Thus, the effect should not only show in wh-clauses, as in (1b, 2a), but also in relative clauses (2b), in declarative V2-clauses (2c) with long distance subject-extraction, and in comparative (2d) and equative clauses (2e).

(2) a. Wer_i wohl meint er, dass e_i ihm seine Arbeit hier bezahlen werde (Paul 1919:321)
who probably meant he that him his labour here pay will
b. Alles, was_i ich dachte, dass e_i mich aufheitern würde (Paul 1919:321)

---

[3] In English, intervening clause initial adverbials cancel the effect, as noted by Browing (1996):
i) Who do you think that *(under these circumstances) would disagree?

[4] The first potential subject XP met by the parser is taken to be the local subject for the time being.

[5] Paul (1919: 321f.) devoted a subsection of his German Grammar to long distance wh-dependencies and referred to them justly as '*Satzverschlingung'* (sentence intertwining). In his data collection (excerpts from novels by German writers) you will strike on examples of that-t violations for interrogative and relative clauses.

everything which I thought that me cheer-up would

c. Die Freimaurer$_i$ hat es geheißen, dass e$_i$ sich dafür interessiert hätten
the freemasons has it been-said that themselves for-it interested had

d. Er hat mehr Leute eingeladen, *O*$_i$ als sie glaubt, dass e$_i$ im Saal Platz finden würden
he has more people invited than she thinks that in-the hall room find would

e. Er hat so viele Leute eingeladen, *O*$_i$ wie er glaubt, dass e$_i$ im Saal Platz finden würden
he has so many people invited as she thinks that in-the hall room find would

Note that (2d,e) are the cardinal cases to be tested against (2a,b,c) in order to ascertain the agreement effect. In (2d,e), the fronted item is a phonetically void operator, hence there is no overt agreement clash. So the prediction is that the parsing difficulty in (2a-c) is absent for (2d,e) and hence that a clause with a subject gap is on a par with a clause with an object gap in the case of (2d,e).[6]

The *that-t* issue is instructive for yet another aspect of data valuation. In her dissertation, Torris (1984) anticipated what Featherston calls for.[7] She carefully tested the relevant patterns and found out that systematic extractors are systematic extractors for all cases (wh-clauses, relative clauses, comparatives). On the other hand, there are extraction *admitters* and they show an unsystematic behaviour. They rate some examples as better than others, but in an unsystematic way across the various constructions. It seems as if they have learned to tolerate extractions by others but their grammar does not admit long distance extraction out of C-introduced finite clauses. Most Northern German varieties, unlike all Southern ones, prefer to not extract out of *that*-clauses at all[8] (see Grewendorf 1995, Fanselow 1987). A Northerner is at best prepared to accept extractions by speakers of other varieties, but this is not covered by her/his grammar, whence some erratic results in tests. Gathering data in contact areas may amount to gathering contaminated data.[9]

Let us recapitulate: would a careful poll with informants on the presence/absence of the *that-t* constraint in the German grammar improve theorizing? The answer is less clear in the given case. If you compare long-distance subject extraction with object extraction, you will find differences. But this is not enough for proving a *that-t* effect. What you have to prove is that the differences are causally related to the *that-t* constraint and that the differences amount to a grammaticality defect. Unfortunately, you cannot simply read this off from the test results unless you construct a test that discriminates *that-t* effects from effects of other sources, as for instance processing based subject-object differences. Constructing this test is guided by your hypotheses. A simple inspection of informant reactions on subject object asymmetry stimuli,

---

[6] Note that in English, the that-t effect applies in all five construction types. Here is an example for the comparative/equative cases:

i)*He interviewed more people than the secretary said that *t* applied for the job

ii)*He interviewed as many people as the secretary said that *t* applied for the job

[7] And so did Wöllstein-Leisten (2001), in the case of the 'third construction', apparently unnnoticed by Featherston (see section 4.2. on the third construction). The results depicted in figure 7 match her findings.

[8] But note that the author of (2a), according to Paul (1919), is Theodor Storm from the very North of Germany. Nevertheless, long distance extraction of a subject across a 'dass' was not alien to him.

[9] Andersson & Kvam (1984) tested extractions out of *dass*-clauses at various locations in Germany and not only showed what amounts to a contrast between extractors and non-extractors, but also a kind of adaption effect for non-extractors that tolerate extractions by others.

however carefully they are gathered, will not reveal the nature of the asymmetry.[10] It is your construal.

Let us turn now to the data-to-analysis fit. In the discussion of the '*aber XP nicht*' vs. '*aber nicht XP*' construction, Featherston refers to competing analyses (clausal ellipsis vs. negated phrase). The ellipsis analysis construes the tag as an elliptic clause, the competing analysis construes it as a negated phrase without ellipsis. Featherston suggests that data gathered from controlled studies with informants will facilitate the choice of the adequate grammatical analysis.

The study Featherston reports shows a clear preference for [*aber XP nicht*], if XP is the subject. For objects, [*aber XP nicht*] and [*aber nicht XP*] are both acceptable, with a slight preference for the latter, independently of the type of contextual bias. Featherston concludes that this supports the clausal analysis because "the arguments in the tag prefer their canonical clausal positions relative to the negative: subject preceding, object following."

This account would not be persuasive, however. In German, both the subject and the objects precede the sentence negation (3). Indefinites (3a) as well as definites (3c) avoid the scope of negation. So, both the subject and the object precede the negation particle in the canonical order. Hence, there is no compelling reason for the observed pattern preference in terms of different canonical clausal positions.

(3) a. dass es was *nicht* beachtet hat
that he something not noted has
b.*dass er *nicht* was beachtet hat
that he not something noted has
c. dass er den Mann *nicht* beachtet hat
that he the man not noted has
d.?dass er *nicht* den Mann beachtet hat
that he not the man noted has

If you ask yourself whether clausal ellipsis or phrasal negation is the adequate analysis, you ask the wrong question. It is not an issue of either-or since each analysis is adequate, but not in each case.

[*Nicht XP*] is phrasal negation, [*XP nicht*] is clausal ellipsis. The sequence [*aber nicht XP*] is a right conjunct with a negated XP, as in (4a). In the conjunct, the negation particle has to precede the XP (4b), since this is the position of phrasal negation. (4c) is the extraposition variant of (4a). The right conjunct is extraposed.

(4) a. [Zeitungen *aber nicht Magazine*]$_{NP}$ liest Peter oft
newspapers but not magazines reads Peter often
b.*[Zeitungen aber Magazine nicht]$_{NP}$ liest Peter oft
newspapers but magazines not reads Peter often
c. Zeitungen liest Peter oft [*aber nicht Magazine*]
newspapers reads Peter often but not magazines

[10] This is reminiscent of the debate on the feasibility of *discovery procedures* in the middle of the past century.

The sequence [*aber XP nicht*] is a part of a reduced clause, as in (5). The negation follows the XP since this is the position of the clausal negation.

(5) a. [Peter liest oft Zeitungen] aber [Magazine ~~liest Peter~~ nicht]
Peter reads often newspapers but magazines reads Peter not
b. [Peter liest oft Zeitungen] aber [Petra ~~liest Zeitungen~~ nicht][11]
Peter reads often newspapers but Petra reads newspapers not

What is the source of the clear preference relation Featherston reports? Why is the conjunction analysis (4) avoided for subjects? Again, as in the allege *that-t* cases, agreement is the crucial factor. If the subject is a conjunct of two NPs, a decision has to be made on singular or plural agreement.

(6) a. Er und sie $\text{lesen}_{\text{pl.}}$ /*$\text{liest}_{\text{sg.}}$ (beide) gerne Zeitung
he and she read/reads (both) gladly newspapers
b. [Er aber auch sie] $\text{lesen}_{\text{pl.}}$/*$\text{liest}_{\text{sg.}}$ (beide) gerne Zeitung
he but also she read/reads (both) gladly newspapers
c.??[Er aber nicht sie] lesen_{pl} gerne Zeitung
he but not she read gladly newspapers

In (6c), the fact that the subject is a conjunct of two NPs would trigger plural agreement, as in (6b), but negation semantically removes the referent of the second conjunct from the denotation of the conjoined NP. So, the NP is grammatically plural but denotationally singular. This seems to be enough conflict potential for shunning this construction in the case of subjects. For objects, the agreement problem does not arise, so the clausal as well as the phrasal option is equally viable.

With this in mind, the results of Featherston's study are expected: for subjects, only the clausal variant is available. This predicts the clausal order [*aber XP nicht*] for subjects. For objects, both options are available. What could make the phrasal conjunct version [*aber nicht XP*] slightly preferable for objects? It is closer to the surface since it does not involve the reconstruction of ellipsis. In other words, the preference of the phrasal construction [*nicht XP*] over the clausal construction [*XP nicht*] is an economy-based preference and a disambiguating strategy.[12]

What is the advantage of careful experimental studies with appropriate test subjects? It is the established rule in the methodologically more mature branches of cognitive science, for good reasons. The situation in linguistics is highly reminiscent of the situation at the birth of psychology in the days of Wilhelm Wundt and his plea for experimental data instead of introspective ones. Linguistics and psychology, both intend to scientifically model mental capacities and both have to rely on secondary evidence for their primary target of explanation, namely the functioning of the mental capacity that produces the data. In both cases, the observable facts are only indirectly related to their primary source.

[11] An indefinite must not be in the scope of negation. Instead, the indefinite would be negated: '*keine Zeitungen'*- no newspapers.

[12] ‚Paul hat Peter geküsst, aber *Petra nicht*' is ambiguous, but ‚Paul hat Peter geküsst, aber *nicht Petra*' is not. So the choice of the latter for an object is a means of disambiguating, whence the preference for objects.

If you want to publish a paper in a psychology journal of some renown, you have to carefully explain how you gathered, organized and statistically analyzed the data before you start to develop your theoretical claims. Introspection is not considered a reliable source of evidence in cognitive science. Its role is in heuristics, but not in data assessment. Once relevant data are established in the field, you must not ignore them. In other words, you may use established data without having to retest them in each case, but when you introduce additional evidence, the onus of the proof for observational adequacy is on your side, and, you are expected not to ignore established data if they are relevant for your claim. In the linguistics reality, we see people freely digressing in both directions – data judgements tuned to the temporary theoretical desires on the one hand and eclectic data selection on the other – without sanctions. The problem is not that people would do this, the problem is that this is an accepted habit in grammar theory in the scientific discourse.

A welcome fringe benefit of a change in attitudes would be the absolute obligation to provide testable evidence for one's favourite novel theoretical ideas. Generative Grammar is not free of post-modern extravagances that praise an extravagant idea simply because of its intriguing and novel intricacies as if novelty and extravagance by itself would guarantee empirical appropriateness. In arts this may suffice, in science it does not. Contemporary papers too often enjoy a naive verificationist style[13] and seem to completely waive the need of independent evidence for non-evident assumptions. The rigorous call for testable and successfully tested independent evidence is likely to disturb many playful approaches to syntax and guide the field eventually into the direction of a serious science. At the moment we are at best in a pre-scientific phase of orientation, on the way from philology to cognitive science.

Let us turn now to the second issue: **Is grammaticality a matter of degree?**

The answer seems to be evident. 'Grammatical' is an ungradeable predicate. It denotes the characteristic function of grammatically well-formed expressions. Either an expression is a member of the set of grammatical expressions or it is a member of the complement set (ungrammatical expressions). Surely, the member status cannot be characterized in terms of a decimal between zero and one, it is either zero (not a member) or one (a member of the grammatical set). The set is not fuzzy. It is a stunning property of the human language capacity that human grammars produce discrete properties rather than fuzzy or probabilistic ones.

For instance, what would '0.69 grammaticality' mean for a given sentence? Surely, it does not mean that 69 out of 100 informants would judge this sentence grammatical. Nor would it mean that a given informant has signalled acceptance in 69 cases our of 100 when a sentence of the same kind is presented to him. In other words, it is not a probability measure.

What it seems to mean is that informants tend to rate the given expression as more/less deviant in comparison to a (grammatical) variant of this expression. '0.69' seems to mean something like this: "the expression you want me to judge is not complete gibberish, but clearly deviant." It is unlikely however, that it could mean "I judge this sentence and I apply a careful grammaticality check on all relevant dimensions and the result is that the given expression

[13] In simple words: "I can show you that in the present version of my theory, the German construction under analysis is derivable, if I assume the devices D_1, D_2, and D_3. Therefore, D_1, D_2, and D_3 are part of German grammar." Serious science would claim an obligation to provide substantive independent evidence for D_1, D_2, and D_3.

passes 69% of the grammaticality principles, but fails 31%.". You may even grant weighted principles and a computation of the weighted result, but you would nevertheless miss the point. What you are grading is deviance, not grammaticality. The degree of rejection seems to be inversely related to the ease of processing and repair, as Featherston acknowledges in section 5.4. Let me illustrate this with the examples in (7) and (8).

(7) *Acceptability as a function of the ease of repair*

a. *Das sein kein guter Satz
this be no good sentence

b. *Das nix gut Satz
this 'nothin' good sentence

c. *Satz guter kein sein das
sentence good no be this

Even without testing, you may be pretty sure that these sentences are rated deviant. If we tested them, and we asked for a magnitude estimation, we would not expect informants to judge them as equally deviant. (7a) seems to be a good candidate for a deviant but in its deviant status fairly acceptable sentence. (7c), the mirror image of (7a), on the other hand, comes close to a random sequence of words. (7a) is easy to repair. You just replace 'sein' by the appropriate finite form. (7b) is clearly deviant, but easy to understand and therefore easy to repair. (7c) contains many sites of repair before it is rendered understandable. It is the worst.

Ease of repair combines with ease of processing. Processing difficulties may reduce the acceptability of grammatical sentences to a degree similar to ungrammatical ones. Most informants (including professional syntacticians), when they are confronted for the first time in their life with (8a), rate it as unacceptable (implicitly comparing it to a_1 to a_3). But it is fully well-formed.[14]

(8) *Acceptability as a function of the ease of processing*

a. Das sind es (a_1: Das $sind_{pl}$ sie_{pl}; a_2: Das $ist_{3.sg.}$ $es_{3.sg.}$ a_3: Das $bist_{2.sg.}$ $du_{2.sg.}$).
this are it this are they this is it this art thou

b. Fest steht, dass Max die Kollegen nicht vorgestellt bekamen
certain is that Max the colleagues not presented got

c. Hätte, hätte es geregnet, alles abgesagt werden müssen?
had, had it rained, everything cancelled been ought

d. Hätte, hätte, hätte es geregnet, alles abgesagt werden müssen, jemand protestiert?
had, had, had it rained, everything cancelled been ought, someone objected

(8b) appears to be unacceptable because of the disappointed choice of ‚Max' as the subject, combined with the desperate struggle for the identification of the real subject 'die Kollegen' – in fact a disguised indirect object promoted to subject in the German construction that resembles the English get-passive[15] – and the concomitant initial confusion on the theta-mapping for the arguments of *vorgestellt* (introduced).

[14] Thanks to M.Bierwisch for reporting me this datum (credit: Ewald Lang). Informants reject it first, but they accept it immediately once you ask for instance "It this a possible answer for: Are these really 47 envelopes?"

[15] Scrambling of an object without overt case markin across the surface subject in the German bekommen (*get*)-construction produces strong feelings of deviance for informants. Thanks to M.Schlesewsky for the datum.

(8d), finally, is a specimen of the well-known trouble-making class of centre embeddings, aggravated by re-occurrences of the same word. The parser gives up at the recursion of the centre-embedded conditional in (8d), while (8c) is only seen as a cumbersome case.

What the examples in (7) and (8) have in common is processing difficulty, or, in Featherston's words, a mental effort criterion. Ungrammatical sentences (7) are difficult because they need repair while they are processed. Garden-path sentences (8) are difficult because the parser is misled and gets stuck. An informant senses the difficulty and tries to map it on whatever scale you provide. But, we must not misinterpret processing difficulty as ungrammaticality. An ungrammatical sentence is ungrammatical, regardless of whether it is easy to process or not.

Featherston is right, if he insists that what you squeeze out of informants never is categorical but graded, unless you impose a forced choice. Does this warrant the conclusion that grammaticality is gradient? No, simply because it amounts to a category mismatch. Grammaticality is an unobservable quality. It is the status of well-formedness that a grammar determines for a given expression. Acceptability is a characterization of more ore less successful processing attempts. Featherston (sect. 5.4) claims that "this position cannot be disproved but we find no evidence to support it". What kind of evidence would we have to find? For instance this: ask an informant whether one or both or none of the sentences in (9), uttered by a person, would tell that (s)he is not a native speaker of German.

(9) a. Seit neun Jahren lebe ich in Österreich
since nine years live I in Austria
b. Seit neun Jahren, ich lebe in Österreich
since nine years I live in Austria

My prediction is this. More than 95% of the test subjects will pick (9b) as an indicator, and less than 5% will choose either (9a) or both. Assume you really found this result. What would be the interpretation? Presumably something along these lines: either (9b) has been uttered inadvertently as a rare mistake in L1, or it has been uttered because the interim L2 grammar of German admits or requires it. This supposition would be sufficient for predicting the close to 100% result.[16]

If I assumed, on the other hand, that grammatical accuracy is graded and fluctuating, I would always allow for a margin of misses and would therefore be reluctant to jump to the conclusion that an ungrammatical utterance as (9b) reveals anything about an incomplete command of the grammar of an L2 learner. Our immediate readiness to jump to the conclusion, however, is evidence for the categorical nature of competence. As an informant, I shall be willing to take (9b) as representative for a class of similar violations of the V2-constraint and thereby I am willing to follow the test instructions and convince myself that (9b) is a good candidate for the identification of a non-native speaker. If Featherston insists that his preferred method, namely magnitude estimation, does not provide evidence for categorical judgements, I agree. But this does not prove or disprove anything about the categorical nature of acceptability judgements. It is just a property of the particular method.

[16] You would not expect 100% because in each test situation there are external influences. An infomant may be tired, diverted, uncooperative, etc.

No potential informant is gifted with an 'absolute pitch' sense of grammaticality. All we can introspectively ascertain is a feeling like "I prefer expression A to expression B to a higher/ lower degree". Expression B may be explicitly given as a target of comparison. If it is not given, we nevertheless compare A to what we think would be a more preferable variant of A, if A appears to be not fully acceptable.

Not only do we lack a sense for an 'absolute pitch', but the relative pitch in magnitude estimation also may differ from person to person. The variance in the aggregate data, however, should in my opinion not be mistaken as an indicator of grammatical variance. It should be acknowledged as what it seems to be, namely a test artefact. An informant is asked for a metalinguistic decision – e.g. "map your acceptability feelings for this sentence on a scale between 0 and 10" or better "guess how your neighbour would map the acceptability of this sentence on a scale between 0 and 10" – and (s)he will try her/his best on a task for which (s)he lacks everyday life experience, and like most referees, the informant will be reluctant to touch the (categorical) margins of the scale (whether it is explicitly given or not).

In general, fuzzy tasks produce fuzzy results, but as Niels Bohr once commented his dish washing activity: It's just like in physics. We use dirty water and a dirty sponge but nevertheless we end up with clean glasses. We may add: it helps if you start with clean water. The only irritating thing about Featherston's plea for double-checking on introspectively gained data is the fact that such a plea is far from obsolete yet.

References

Andersson, Sven-Gunnar & Sigmund Kvam. 1984. *Satzverschränkung im heutigen Deutsch*. Tübingen: Narr.

Bornkessel, Ina, Matthias Schlesewsky, Angela D. Friederici. 2002. Grammar overrides frequency: Evidence from the online processing of flexible word order. *Cognition* 85:21–30.

Browning, Margaret. 1996. CP-recursion and 'that-t'-Effects. *Linguistic Inquiry* 27:237-255

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.

Fanselow, Gisbert. 1987. Konfigurationalität: Untersuchungen zur Universalgrammatik am Beispiel des Deutschen. Tübingen: Narr.

Featherston, Sam. 2007. Data in Generative Grammar – the stick and the carrot. *Theoretical Linguistics 33: 269-318.*

Friederici, Angela, Matthias Schlesewsky, and Christian J. Fiebach. 2003. Wh-movement versus scrambling: The brain makes a difference. *Word order and scrambling*, ed. by Simin Karimi. Oxford: Blackwell. (325–344).

Grewendorf, Günther. 1995. German. A grammatical sketch. In. Jacobs, Joachim & A. von Stechow & W.Sternefeld & T.Vennemann eds. Syntax. Berlin: de Gruyter (1288-1319)

Haider, Hubert. 2005. How to turn German into Icelandic – and derive the VO-OV contrasts. *The Journal of Comparative Germanic Linguistics 8:1-53.*

Paul, Herman. 1919. *Deutsche Grammatik. vol. III, part IV: Syntax*. Halle an der Saale: Niemeyer

Schlesewsky, Matthias, Ina Bornkessel, Stefan Frisch. 2003. The neurophysiological basis of word order variations in German. *Brain and Language* 86.116–128.

Torris, Thérèse. 1984. *Configurations syntaxiques et dépendances discontinues en allemand contemporain.* Unpublished PhD Dissertation. Paris: Université de Paris VIII-Vincennes.

Wöllstein-Leisten, Angelika. 2001. *Die Syntax der dritten Konstruktion.* Tübingen: Stauffenburg Verlag.