The relationship between grammaticality ratings and corpus frequencies: A case study into word order variability in the midfield of German clauses

Gerard Kempen and Karin Harbusch

To be published in: Kepser, Stephan and Reis, Marga (eds.), *Linguistic Evidence – Empirical, Theoretical, and Computational Perspectives*, Berlin: Mouton De Gruyter (in press).

1 Introduction

It is almost a commonplace that word order in the midfield of German clauses is flexible. Although statements to this effect do not claim that "anything goes" (cf. Eisenberg (1994, Ch. 12); Rambow (1994); Müller (1999)), they suggest that word order variability in German clauses is considerably greater than, for example, in Dutch and English ones. The following three constraints on the ordering of argument NPs have figured prominently in the linguistic and psycholinguistic literature (cf. Uszkoreit (1987), Pechmann et al. (1996), Keller (2000)):
C1: Pronominal $\prec$ Non–pronominal
C2: Nominative $\prec$ Non–nominative, and
C3: Dative $\prec$ Accusative
(where "$\prec$" means "precedes").

Recently, intuitions about word order in German have been probed in a systematic fashion by Keller (2000). From linguistically untrained native speakers he elicited "graded grammaticality" judgments via a novel technique based on the psychophysical method of Magnitude Estimation (Bard et al. 1996). In one of his experiments, he determined the strengths of C1, C2 and C3 in ditransitive subordinate clauses. None of the three "precedence" constraints turned out to be "absolute" in the sense that their violation gave rise to extremely low acceptability/grammaticality judgments (as low as vi-

olation of the absolute verb–final constraint in subordinate clauses did). In fact, Keller found C1 and C2 to have about equal strength and both to be considerably stronger than C3, which was very weak.

If linear order constraints such as C1 through C3 and their relative strengths are "psychologically real", they are expected to affect the linearization of argument NPs in actual language performance, i.e. during speaking and writing. More precisely, they yield predictions concerning the relative frequencies of argument orderings in written and spoken texts. For example:

- Linear orders that agree with a given constraint will be more frequent than orders violating it.
- Stronger constraints will give rise to a lower proportion of violations than weak constraints.
- Orders that violate more than one constraint will be rarer than violations of single constraints.

In earlier work (Kempen and Harbusch 2003), we verified that this approach is viable in principle. We hypothesized that the strength of a precedence constraint reflects the likelihood that it would actually be respected in the course of the incremental production of a sentence. On this basis, and with a somewhat different set of precedence constraints than the above, we developed a probabilistic model that predicts relative frequencies of occurrence of argument NP orderings in real texts. We assumed, furthermore, that the predicted frequency of an argument ordering corresponds directly to its rated grammaticality. The model we thus obtained yielded a satisfactory fit with Keller's ratings.

However, this proof of concept hinges on the assumption of a close correspondence between frequency and rated grammaticality. As no frequency counts are available of argument NP orderings in clauses of the same type as those presented to the participants in Keller's experiments, we decided to supply this want. In the course of this corpus work, a systematic discrepancy emerged between the frequency counts and the grammaticality ratings. We had expected argument orderings in the middle range of the grammaticality spectrum to occur in the corpus with moderate frequencies, but they were conspicuously absent. The level of flexibility emerging from the frequency counts thus was considerably lower than grammaticality judgments suggest. The discovery of this frequency–grammaticality discrepancy spawned the investigation we report here.

Preview. The frequency counts of argument NP orderings in subordinate clauses of written as well as spoken corpora are presented in Section 2. In Section 3, we propose a rather restrictive *"production–based linearization rule"* that models the actually occurring, high–grammaticality orderings. Section 4 extends the corpus study to main clauses in the written corpora and verifies that the rule holds there as well. Subsequently, Section 5 confronts the frequency counts with Keller's (2000) graded grammaticality ratings and shows that the latter license more word order freedom than the former. Section 6 proposes a theoretical account of the observed frequency–grammaticality discrepancy. Section 7 introduces a recent argument ordering model in the linguistic literature (Müller (1999)) which is almost equally strict as our linearization rule although entirely based on grammaticality judgments, and explains why it does not square well with Keller's data. Section 8, finally, summarizes the line of reasoning.

2 The corpus study

We conducted a corpus study into the frequencies of linear orders of pronominal and non–pronominal Subject (*S*, nominative), Indirect Object (*I*, dative) and Direct Object (*O*, accusative) NPs in German subordinate clauses. These clause types include those used by Keller (2000) in his grammaticality rating studies. We needed these data in order to determine to what extent graded grammaticality ratings are mirrored by the frequency of argument ordering patterns in sentence materials generated outside the laboratory. We sought an answer to this question in written as well as spoken texts.

2.1 Written language: The NEGRA–II and TIGER treebanks

2.1.1 Method

Recently, the NEGRA–II (Skut et al. 1997) and TIGER corpora (Brants et al. 2002) have become available — German treebanks that together contain about 60,000 newspaper sentences annotated in full syntactic detail. The data we report here have been aggregated over both corpora.

Using version 2.1 of TIGERSearch (König and Lezius 2000), we extracted all finite clauses introduced by a subordinating conjunction and containing an (*S*,*I*) and/or and (*S*,*O*) pair, possibly with an additional (*I*,*O*) pair (with the members of a pair occurring in any order). As for terminology, clauses containing only an (*S*,*I*) pair are labeled *in*transitive; clauses with only an (*S*,*O*) pair are *mono*transitive; a clause with an (*S*,*O*) as well as an (*I*,*O*) pair

is *di*transitive; both latter types of clauses are called transitive. We found 2578 monotransitive, 287 intransitive, and 199 ditransitive subclauses meeting the requirements (3064 sentences in all). We distinguished six types of NPs: three pronominal (labeled (*Sp*, *Ip*, *Op*) and three non–pronominal or "full" (*S*, *I*, *O*). An NP is pronominal iff it consists of a personal or a reflexive pronoun. Because a clause contains at most one token of each of the three types of grammatical function, there are 12 possible *unordered pairs* of NPs: three combinations of grammatical functions ((*S,I*), (*S,O*) and (*I,O*)) times four combinations of NP shapes (all pronominal, one member full, the other member full, all full). For each of these, we determined the frequency of the two possible orderings (i.e., of 24 *ordered pairs*). For additional methodological details, we refer to Kempen and Harbusch (2004a).

2.1.2 Results

There were 3462 ordered pairs: one from each mono– or intransitive clause; three from every ditransitive clause (Table 1).

Table 1: Frequency of the 24 ordered pairs of argument NPs in subordinate clauses extracted from the NEGRA–II and TIGER corpora. Dark–gray cells represent syntactically inadmissible constituent pairs.

		Second NP						*Total*
		Sp	*Op*	*Ip*	*S*	*I*	*O*	
First NP	*Sp*		**146**	**49**		**72**	**654**	*921*
	Op	**0**		**5**	**302**	**23**		*330*
	Ip	**0**	**0**		**89**		**93**	*182*
	S		**195**	**41**		**182**	**1476**	*1894*
	I	**0**	**2**		**53**		**67**	*122*
	O	**0**		**0**	**4**	**9**		*13*
Total		*0*	*343*	*95*	*448*	*286*	*2290*	*3462*

What immediately strikes is the high proportion of (almost) empty cells. This suggests a level of flexibility that is not very high. The ordering of the pronominal constituents is invariably *Sp—Op—Ip*, and *S* is the only full NP that may precede a pronominal NP. Variability within full NPs is somewhat greater: While *S—I—O* is the predominant order, inversions do occur regularly. Closer inspection of the data reveals, however, that the inverted order *I—S* is restricted to clauses with intransitive verbs (more precisely, to "experiencer–object" verbs as in *dass [dem Jungen]$_I$ [etwas]$_S$ widerfährt*;

that [the boy] something happens–to; 'that something happens to the boy'), and that the sequence *O—I* only occurs as standard order licensed by special ditransitive verbs (cf. *[jemanden]$_O$ [einer Prüfung]$_I$ unterziehen*; someone [a test] subject–to; 'subject someone to a test'). See Section 5 for additional data.

Before going into the reasons for the restricted variability, we need to assess whether it generalizes to other text types.

2.2 Spoken language: The VERBMOBIL dialogue corpus

The NEGRA–II and TIGER corpora consist entirely of written newspaper texts. Probably, these texts have been heavily edited by authors and editors. In the course of this process, many argument orderings that embodied low–frequency patterns may have been eliminated and replaced by alternatives in the high–frequency range. We therefore deemed it necessary to explore a text type that has undergone no off–line editing. The VERBMOBIL corpus of spoken dialogues (Burger et al. 2000) lended itself very well to this purpose.

2.2.1 Method

Having extracted all relevant subordinate clauses from the transcribed VERBMOBIL dialogues,[1] we classified the grammatical functions (*S, I, O*) by hand.

2.2.2 Results

We found 2711 monotransitive, 296 intransitive, and 168 ditransitive subordinate clauses meeting the requirements (3175 clauses in all). See Table 2 for the 3511 ordered argument pairs. Section 5 gives additional details.

Inspection of Table 2 shows that the frequency pattern is very similar to that in Table 1. The number of empty or almost empty cells is again rather high, and they occupy the same positions in the two Tables. We conclude that limited flexibility of argument NP orderings is a rather widespread phenomenon occurring in spoken as well as written language.

In the next Section, we propose a linearization rule that fits the observed pattern rather accurately.

3 A production–based linearization rule

The frequency data can be accounted for by the rather strict *"production–based linearization rule"* in Figure 1.[2] To each individual argument NP,

Table 2: Frequency of the 24 ordered pairs of argument NPs in subordinate clauses extracted from the VERBMOBIL corpus. Dark–gray cells represent syntactically inadmissible constituent pairs.

		Second NP						*Total*
		Sp	*Op*	*Ip*	*S*	*I*	*O*	
First NP	*Sp*		**687**	**263**		**7**	**2003**	*2960*
	Op	**2**		**4**	**40**	**1**		*47*
	Ip	**0**	**2**		**174**		**151**	*327*
	S		**13**	**18**		**2**	**132**	*165*
	I	**0**	**0**		**0**		**4**	*4*
	O	**1**		**5**	**1**	**1**		*8*
Total		*3*	*702*	*290*	*215*	*11*	*2290*	*3511*

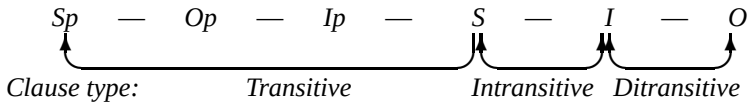

Figure 1: "Production–based linearization rule" representing the linearization options observed in the written and spoken corpora in clauses headed by a mono–, di–, or intransitive head verb. Transitive clauses include both mono– and ditransitive ones.

the rule assigns a standard ("primary") position among its clausemates. The pronominal NPs have a fixed position in the anterior region of midfield. (This region is traditionally called "Wackernagel position".) The primary position of full NPs is posterior to that of the pronouns. However, each full NP has an additional "secondary", more anterior placement option, which is indicated by the labeled arrows. This "freedom in restraint" is conditional upon mono–, di–, or intransitivity of the head verb. Mild conceptual factors such as animacy (Kempen and Harbusch 2004a), definiteness (Kurz (2000); Kempen and Harbusch (2004b)) and referential ease (yielding "light" NPs; cf. Hawkins (2004); Wasow (2002); Kempen and Harbusch (2004b)) enable full argument NPs to occupy the more leftward position.

Of the 6239 written and spoken subordinate clauses extracted from the corpora, only 17 exemplars violate the production–based rule (0.27%). We list these exceptions *in extenso* in Table 3[3].

Table 3: Exceptions to the production–based linearization rule in the NEGRA–II, TIGER, and VERBMOBIL corpora. The source of each clause has been added to the reference number (N=NEGRA–II, T=TIGER, V=VERBMOBIL).

Monotransives	
(1-N)	da [diese Aufgabe]$_O$ [der Landrat]$_S$ selbst übernehmen wolle
(2-T)	Seit [das]$_O$ [die Ärzte]$_S$ in der Heimat wissen
(3-T)	sofern [dies]$_O$ [eine der Seiten]$_S$ wünscht
(4-T)	Wenn [Freund wie Feind]$_O$ nun [eine Frage]$_S$ bewegte
(5-V)	weil [das]$_O$ ja [die Firma]$_S$ trägt, die Kosten
(6-V)	weil [den Totensonntag]$_O$ [ich]$_{Sp}$ nicht für sehr geeignet halte
(7-V)	nachdem [sich]$_{Op}$ [es]$_{Sp}$ ja nur um eine kurze Besprechung handelt
(8-V)	weil [sich]$_{Op}$ [es]$_{Sp}$ um ein fünftägiges Arbeitstreffen handelt
Ditransives	
(9-T)	Wenn [er]$_{Sp}$ [der Pest]$_I$ nicht schleunig [uns]$_{Op}$ entreißt
(10-T)	daß [es]$_{Sp}$ [den Menschen]$_O$ [sich]$_{Ip}$ selbst entfremde
(11-V)	damit [wir]$_{Sp}$ [den Flug]$_O$ [uns]$_{Ip}$ danach einrichten können
(12-V)	daß [wir]$_{Sp}$ vielleicht auch [ein paar Sehenswürdigkeiten]$_O$ [uns]$_{Ip}$ anschauen
(13-V)	daß [wir]$_{Sp}$ [die achtunddreißigste Woche]$_O$ [uns]$_{Ip}$ dafür ausgucken
(14-V)	daß [wir]$_{Sp}$ [Donnerstag vormittag]$_O$ einfach [uns]$_{Ip}$ vornehmen
(15-V)	wenn [wir]$_{Sp}$ schon [so viel]$_O$ [uns]$_{Ip}$ anschauen
(16-V)	wenn [ich]$_{Sp}$ [mir]$_{Ip}$ [es]$_{Op}$ so recht überlege
(17-V)	wenn [Sie]$_{Sp}$ [uns]$_{Ip}$ [sie]$_{Op}$ zuschicken könnten

4 Generalizing the production–based linearization rule to main clauses

If the production–based rule really determines the order of argument NPs in the midfield, it should generalize to *main clauses with inversion of Subject and finite verb*. In order to test this prediction, we checked the ordering patterns in mono– and ditransitive main clauses in the NEGRA–II and TIGER treebanks. We found 5025 ordered pairs: one from each mono– or intransitive clause; three from every ditransitive clause. Comparison of Tables 4 and 1 reveals a very similar frequency distribution over the cells.

The most salient exceptions the linearization rule involve extraposed full Subject NPs. We found 20 (out of 5025) *O–S* pairs, which is somewhat higher than in the subordinate clauses (4 out of 3462 pairs). In a high proportion of these cases, the Subject NP is extremely long. An example:

- Für die neue Konzertsaison, die am 25. September beginnt, erwarten [das Publikum]$_O$ wieder [*einige Höhepunkte, die der musikalische Leiter der Konzerte, Michael Schneider, unter genauen Vorgaben zusammengestellt hat*]$_S$.

This also applies to four rule–violating *Op–I–S* main clauses (see the corresponding bar in the bottom panel of Figure 3 below). This is one of them:

- Hier bot [sich]$_O$ [den sechs bis 12jährigen]$_I$ [*die Gelegenheit, die großen Düsenflugzeuge auf dem Rollfeld zu beobachten und die Flughafenfeuerwehr zu besuchen*]$_S$.

We assume that these orderings are due to the overriding influence of "end weight" (cf. Hawkins (2004); Wasow (2002)). Whether they are *real* exceptions is disputable, though, because extraposed constituents often end up in the endfield of a clause. In the absence of a nonfinite verb separating the midfield from the endfield ("rechte Satzklammer"), one cannot tell which field the extraposed constituent has selected.

In sum, the NEGRA–II and TIGER data justify generalization of the production–based linearization rule to argument orderings in main clauses. The fact that clauses independently selected from the three corpora appear to respect the rule, may reinforce thrust in it and weigh up against the disadvantage of the relatively small corpus sizes.

Table 4: Frequency of the 24 ordered pairs of argument NPs in the midfield of main clauses extracted from the NEGRA–II and TIGER corpora. Dark–gray cells represent syntactically inadmissible constituent pairs.

		Second NP						*Total*
		Sp	*Op*	*Ip*	*S*	*I*	*O*	
First NP	*Sp*		**153**	**59**		**118**	**937**	*1267*
	Op	**0**		**4**	**894**	**16**		*914*
	Ip	**0**	**2**		**171**		**99**	*272*
	S		**37**	**19**		**201**	**2134**	*2391*
	I	**0**	**0**		**79**		**75**	*154*
	O	**0**		**0**	**20**	**7**		*27*
Total		*0*	*192*	*82*	*1164*	*342*	*3245*	*5025*

5 Comparing grammaticality judgments and frequency data

In his Experiment 6, Keller (2000) elicited grammaticality ratings for ditransitives subordinate clauses where at most one constituent is pronominal. This yields four (what we will call) "families" of ditransitive clauses: one family with all NPs full (*S,I,O*) and three families with one pronominal NP ((*Sp,I,O*), (*S,Ip,O*) and (*S,I,Op*)). Each member of a family represents one possible ordering of the NPs. Hence, every ditransitive family comprises six members (i.e., six argument permutations). Experiment 10 dealt with monotransitive subordinate clauses, but here both *S* and *O* could be pronominal. Thus there are four monotransitive families — (*S,O*), (*Sp,O*), (*S,Op*) and (*Sp,Op*) — with two members each. It is important to keep the clause families separate because a family represents a unique combination of grammatical functions and (non–)pronominal NP shapes. This allows a view of linear order preferences uncontaminated by production factors that control the choice between a pronominal or non-pronominal NP shape.

The ratings for the ditransitive clause families are shown in Figures 2 and 3, together with the corresponding corpus frequencies; Figure 4 displays ratings and frequencies for the monotransitive families. The grammaticality ratings of the NP orderings decrease from left to right. (The exact values are listed in the third column of Tables 5 and 6.)[4]

The Figures reveal a substantial correlation between the grammaticality rating and the corpus frequency of the members of a clause family: In all mono– and ditransitive clause families, the most frequent members are the ones that received the highest ratings; and members with very low ratings are absent from the corpus. However, quite a few orderings that are rated at least

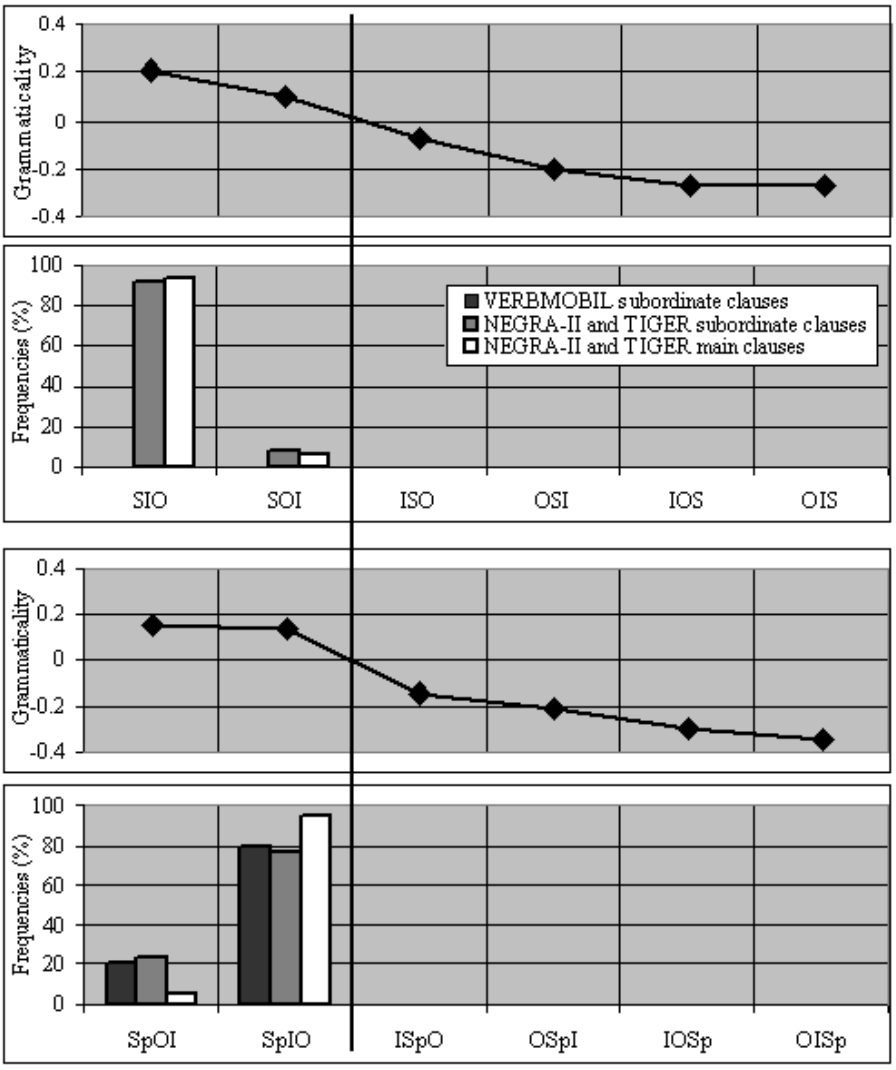

Figure 2: Grammaticality ratings for the (*S,I,O*) and (*Sp,I,O*) families in, respectively, the first and the third panel. Data from Keller's (2000) Experiment 6. The corresponding bar diagrams display the relative corpus frequencies for the argument orderings, expressed as percentages of the total number of ditransitive clauses in their family. In the topmost bar diagram, the bar for the (*S,I,O*) family in the VERBMOBIL corpus is missing due to the non–occurrence of such clauses in that corpus.

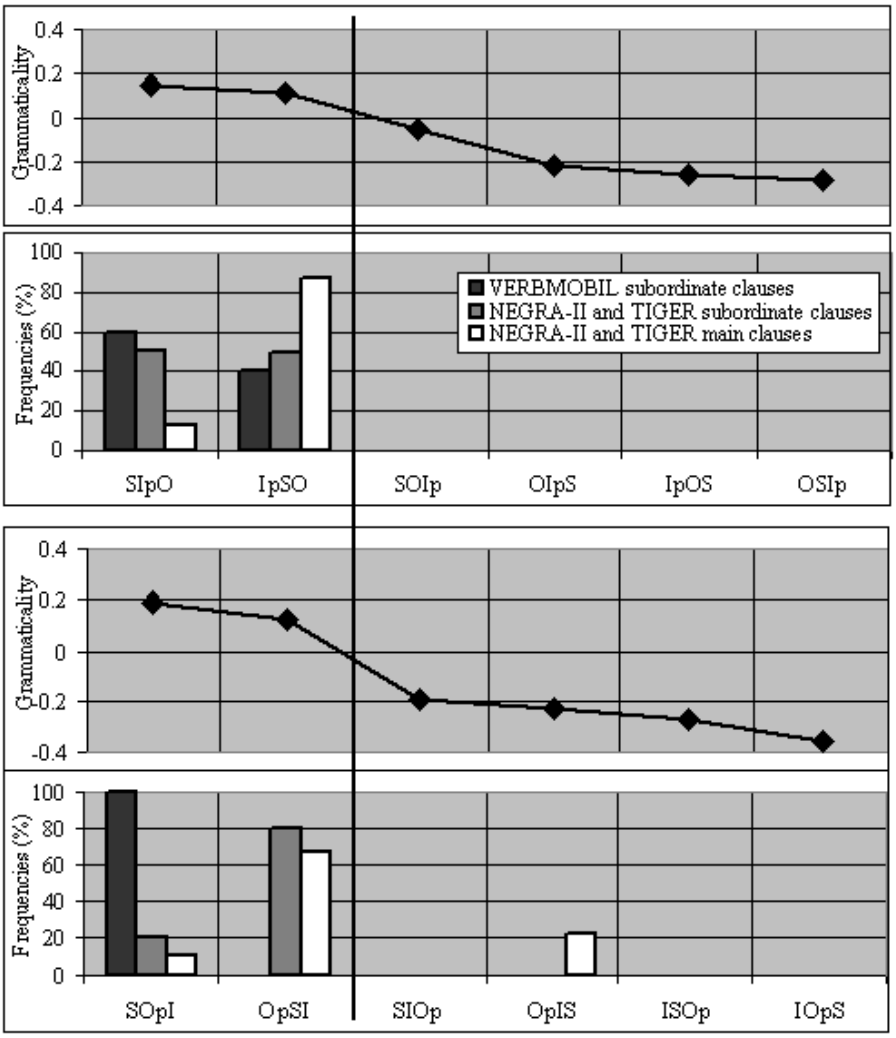

Figure 3: Grammaticality ratings for the (*S,Ip,O*) and (*S,I,Op*) families in, respectively, the first and third panel. Data from from Keller's (2000) Experiment 6. The corresponding bar diagrams display the relative corpus frequencies for the argument orderings, expressed as percentages of the total number of ditransitive clauses in their family.

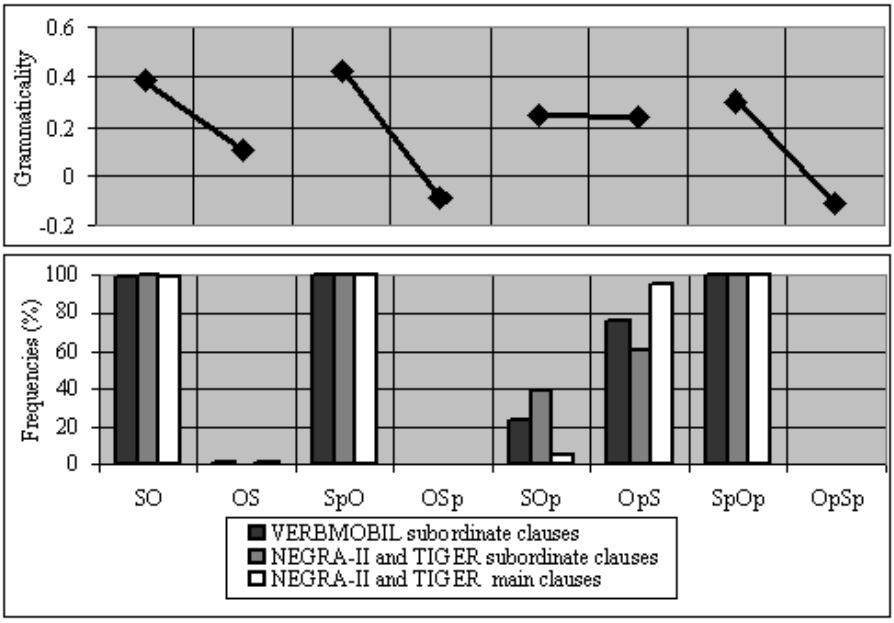

Figure 4: Upper panel: Grammaticality ratings for all families of monotransitive subordinate clauses, from Keller's (2000) Experiment 10. Lower panel: Relative corpus frequencies for the same families of clauses, expressed as percentages of the total number of monotransitive clauses in their family.

average in grammatical quality, also have zero corpus frequencies. Consider, in particular, the ditransitives to the immediate right of the thick vertical lines in Figures 2 and 3. In other words, *the grammaticality judgments tend to be more lenient than the corpus data.*[5]

In the next Section we propose an explanation for this discrepancy.

6 Severity of rule violation affects corpus frequencies and grammaticality ratings differently

Let us assume that the internal grammar of native speakers of German includes our production–based linearization rule — in other words, that the rule is "psychologically real". This implies that under normal conditions the grammatical encoding mechanisms operative in these speakers of German will function according to this rule, and sentences violating it will rarely occur. When performing the grammaticality judgment task and detecting an argument ordering that conflicts with the rule, the informants can react in two different ways:

- They reject any ill-formed linear order and assign the same low grammaticality score to all clauses embodying a deviant order.
- To ill-formed clauses they assign a grammaticality score commensurate with the severity of rule violation, i.e. with number and seriousness of the deviations.

In the latter case, the grammaticality score an argument order receives is predicted to be a function of the *similarity between this order and the order(s) licensed by the production–based linearization rule*. In order to evaluate the viability of this hypothesis, we derived from it a simple similarity metric. Arguments that prefer an early position in the clause are the Subject NP (full or pronominal) and the pronominal NPs. Mono– and ditransitive orderings where these constituents indeed occupy early positions are therefore more similar to the rule than orderings where these constituents have moved to the right. Calculation of the average ranking number of the Subject and pronominal NPs in an argument ordering yields an "anteriority" score for each member of a family of clauses. We predict high negative correlations between anteriority and grammaticality: the lower the anteriority score of an ordering (that is, Subject and pronominal NPs in early positions), the higher the grammaticality score. Anteriority values can be calculated for both the mono– and the ditransitive clauses.

Comparison of the resulting scores with the grammaticality ratings for each member of a family of clauses — see Table 5 and Table 6, columns 2 and 8 — exhibits a high rank correlation. In only one case (marked in gray), the rank orders are reversed.

Table 5: Predicting the grammaticality ratings for ditransitive clauses. Columns 2 and 3: grammaticality values and their rank order (from Keller's Experiment 6). Columns 4 and 5: corpus frequencies. Columns 6 and 7: anteriority ranks. Column 8: anteriority score (mean rank).

Ordering	Grammaticality		Frequency		Anteriority ranks		
pattern	rank	judgment	written	spoken	subject	pronoun	mean
$S-I-O$	1	.2083	54	0	1	–	1
$S-O-I$	2	.0994	5	0	1	–	1
$I-S-O$	3	-.0716	0	0	2	–	2
$O-S-I$	4	-.2038	0	0	2	–	2
$I-O-S$	5	-.2667	0	0	3	–	3
$O-I-S$	6	-.2736	0	0	3	–	3
$Sp-O-I$	1	.1519	4	1	–	1	1
$Sp-I-O$	2	.1386	13	4	–	1	1
$I-Sp-O$	3	-.1463	0	0	–	2	2
$O-Sp-I$	4	-.2081	0	0	–	2	2
$I-O-Sp$	5	-.2936	0	0	–	3	3
$O-I-Sp$	6	-.3471	0	0	–	3	3
$S-Ip-O$	1	.1471	30	6	1	2	1.5
$Ip-S-O$	2	.1144	29	4	2	1	1.5
$S-O-Ip$	3	-.0516	0	0	1	3	2
$O-Ip-S$	4	-.2164	0	0	3	2	2.5
$Ip-O-S$	5	-.2612	0	0	3	1	2
$O-S-Ip$	6	-.2810	0	0	2	3	2.5
$S-Op-I$	1	.1938	3	1	1	2	1.5
$Op-S-I$	2	.1235	12	0	2	1	1.5
$S-I-Op$	3	-.1876	0	0	1	3	2
$Op-I-S$	4	-.2247	0	0	3	1	2
$I-S-Op$	5	-.2694	0	0	2	3	2.5
$I-Op-S$	6	-.3550	0	0	3	2	2.5

We conclude that the average similarity between to–be–judged argument orderings and orderings licensed by the production–based rule is a good predictor of the grammaticality ratings. Stated differently, *the grammaticality ratings appear sensitive to the number and seriousness of violations of the rule*. Argument orderings that embody mild violations of the rule, receive

Table 6: Predicting the grammaticality ratings for monotransitive clauses. Columns 2 and 3: grammaticality values and their rank order (from Keller's Experiment 10). Column 4 and 5: corpus frequencies. Columns 6 and 7: anteriority ranks. Column 8: anteriority score (mean rank).

Ordering	Grammaticality		Frequency		Anteriority ranks		
pattern	rank	judgment	written	spoken	subject	pronoun	mean
$S-O$	1	.3818	1358	122	1	–	1
$O-S$	2	.1078	4	1	2	–	2
$Sp-O$	1	.4180	603	1852	1	1	1
$O-Sp$	2	-.0887	0	1	2	2	2
$S-Op$	1	.2482	189	12	1	2	1.5
$Op-S$	2	.2412	290	40	2	1	1.5
$Sp-Op$	1	.3024	134	681	1	1	1
$Op-Sp$	2	-.1071	0	2	2	1	1.5

medium–range grammaticality scores due to this sensitivity but are virtually absent from the corpora because the grammatical encoding mechanism in speakers/writers does not (or hardly ever) produce them.

7 Graded grammaticality versus "graded ungrammaticality"

In the foregoing we have tried to model the data we extracted from the spoken and written corpora as faithfully and concisely as possible, with the "production–based linearization rule" as the central outcome. We arrived at this rule independently from Müller's (1999) system of word order constraints which resembles ours in many respects. The following properties of his model represent points of similarity:

- The midfield includes an early region where only pronominal arguments can go — the so–called Wackernagel position.
- Pronouns in the Wackernagel position do not scramble (i.e., their order is fixed).
- Full NPs occupy a region posterior to the Wackernagel position.
- In the posterior region, scrambling is allowed.
- Of the full NPs, only the Subject is allowed to precede the Wackernagel position.

There are only a few differences:

- A full Subject that precedes one or more pronominal NPs, is placed to the left of the Wackernagel position, whereas in our rule it goes to the same position as pronominal Subjects.

- Scrambling of full NPs is slightly more liberal in Müller's model. For example, full Direct Objects may precede full Subjects in the post-Wackernagel region.

The similarity between the two rule systems is surprising because the empirical evidence Müller adduces in support of his model is not production– or corpus–based but consists entirely of grammaticality judgments. Given the frequency–grammaticality discrepancies we observed above, one would expect a rather poor fit between a model based on judgments and one based on production data. Keller (2000), too, notes that his ratings are at variance with Müller's model. So, how to account for the convergence of a "performance–based" and a "competence–based" linearization component?

Our answer presupposes that, somewhere on the grammaticality continuum ranging from "perfectly well–formed" to "seriously ill–formed", there is a critical value called the "production threshold". Syntactic structures with grammaticality values above this threshold will occur in corpora with moderate–to–high frequencies. Structures whose grammaticality scores lie slightly above or slightly below the production threshold, will have zero or, at best, very low frequencies — they are "marked". Structures with even lower grammaticality ratings are only delivered in case of a malfunctioning production mechanism or deliberate output distortion.

The argument orderings licensed by our linearization rule or by Müller's grammar presumably have grammaticality values that exceed the production threshold or are in its vicinity; they all represent unmarked or marked cases. Our performance–based rule converges with Müller's competence-based grammar because both aim to model linear orders above and around the production threshold.

Where on the grammaticality continuum should we place the clauses of Keller's experiments? While high–grammaticality clauses indubitably exceed the production threshold, those with medium–range or low grammaticality ratings most probably are subthreshold. If so, they are expected to have zero corpus frequencies. At any rate, a positive correlation between grammaticality and frequency may exist only for structures with grammaticality ratings above the production threshold.

Apparently, "graded grammaticality" should be distinguished from "graded ungrammaticality". Keller's rating data probably stem in part from the latter domain, and the model he proposes aims to account not only for graded grammaticality but also for graded ungrammaticality. In contrast, Müller's grammar, like our production–based rule, only covers graded grammaticality, i.e. the range of grammaticality ratings above and around the production threshold. The domain of grammaticality judgments that constitutes the empirical basis for Müller's grammar, therefore differs from that investigated by Keller. This, in turn, implies that Keller's data and the conclusions drawn from it (e.g., the relative strengths of constraints C1 through C3 in Section 1), should not be used directly to evaluate Müller's grammar. The two empirical domains overlap only in part.

We conclude this Section with two remarks relating to Müller's grammar. First, while our approach is purely *descriptive*, his grammar can *explain* important facts. For instance, we did no more than observe that the primary position of the full Subject NP follows rather than precedes pronouns in the Wackernagel position. Müller's grammar is superior in that it links this property to other parts of the grammar. (For details we refer to his article.) Second, we propose to follow Müller (1999; endnote 11) in the treatment of "strong" pronominal arguments, that is, those carrying sentence accent or preceded by adverbs such as *auch* 'also', *selbst* 'even', *nur* 'only', etc. They function as full NPs and can occupy positions in the post–Wackernagel region of the midfield and be subject to scrambling.

8 Summary and conclusion

We presented the results of several corpus studies into the frequencies of argument NP orders in the midfield of subordinate and main clauses of German. In both the written and the spoken corpora, the amount of argument order variation appeared to be relatively small. We devised a rather restrictive "production–based linearization rule" that describes the actually observed orderings. Comparison of the corpus frequencies and the grammaticality values from Keller's (2000) study revealed a systematic discrepancy: the assignment of medium–range grammaticality scores to quite a few argument orderings with zero frequencies in the corpora.

In order to explain the frequency–grammaticality discrepancy, we posited three hypotheses:

- The strict production–based linearization rule (or a mechanism yielding equivalent output) is causally involved in, and constrains, gram-

matical encoding during spoken and written sentence production.

- The grammaticality raters in Keller's study estimated the average similarity between the to–be–judged argument ordering and the order(s) licensed by the strict linearization rule.

- The grammaticality continuum specifies a critical value called the "production threshold". Structures with grammaticality values above this threshold will occur in corpora with moderate–to–high frequencies. Structures with grammaticality scores in the neighborhood of the production threshold, will have zero or, at best, very low frequencies — they are "marked". Structures with even lower grammaticality ratings are only generated in case of malfunction of the grammatical encoder.

We show that these assumptions, in addition to accounting for the frequency–grammaticality discrepancy, also help to resolve the seeming contradiction between strict and more lenient linguistic (judgment–based) rule/constraint systems for the linear order of argument NPs in the midfield of German main and subordinate clauses, in particular the one between Müller (1999) and Keller (2000).

Author addresses
Gerard Kempen
Psychology Department, Leiden University &
MPI for Psycholinguistics, Nijmegen
kempen@fsw.leidenuniv.nl

Karin Harbusch
Computer Science Department, University of Koblenz–Landau
harbusch@informatik.uni-koblenz.de

Notes

1 The corpus was accessed at URL
http://www.ims.uni-stuttgart.de/projekte/verbmobil/Dialogs/
and explored by means of the on–line search tool made available at the website. In writing the search patterns we heavily used the part–of–speech (PoS) tags included in the transcriptions. No other grammatical annotations are available. As the dialogues had been recorded with several different types of microphones, many dialogue turns occur more than once, although with different codes. We eliminated such duplications before estimating the corpus frequencies.

2 The rule allows only 15 out of 48 logically possible ditransitive permutations of a full or pronominal subject, a ditto direct object and a ditto indirect object. In addition, it licenses 5 monotransitive and 6 intransitive argument NP orders.

3 In order to enable the reader to judge the (un–)markedness of the rule-violating orderings in Table 3, we list here the complete sentences together with some preceding context. (In the VERBMOBIL corpus, the accessible context is restricted to the current dialogue turn.)
(1-N) In den Ferien bietet sich dazu die Gelegenheit. Die Leiterin der Limesschule, Karola Kofler, stellte auf Anfrage klar, daß sie gebeten worden sei, die Eltern nicht zu informieren, da diese Aufgabe der Landrat selbst übernehmen wolle.
(2-T) Der Andrang ist beträchtlich. Seit das die Ärzte in der Heimat wissen, ist dem einen oder anderen Befragungsbogen, den die Behinderten vor Beginn der Reise über ihren Gesundheitszustand ausfüllen lassen müssen, nicht mehr 100prozentig zu trauen.
(3-T) Die Parteien haben sich wie folgt geeinigt: 1. Die Übergangszeit dauert zwölf Monate und kann einmalig um ein weiteres Jahre verlängert werden, sofern dies eine der Seiten wünscht.
(4-T) 1967, als er im iraelisch–arabischen ‘Blitzkrieg” die palästinensische Westhälfte seines “Reiches” einschließlich der Altstadt von Jerusalem verlor, und im “Schwarzen September” 1970, während des beinahe tödlichen Sturmes der PLO-Fedajiin auf die jordanische Monarchie. Wenn Freund wie Feind nun eine Frage bewegte, dann war es die um den Gesundheitszustand des Monarchen und um das Schicksal Jordaniens, wenn er einmal abgetreten sein wird.
(5-V) ja, dem steht nichts entgegen, weil das ja die Firma trägt, die Kosten.
(6-V) da würde ich sagen, entscheiden wir uns doch für den ersten Advent, weil den Totensonntag ich nicht für sehr geeignet halte.
(7-V) gut, nachdem sich es ja nur um eine kurze Besprechung handelt, denke ich, daß uns eine Stunde reichen wird. also machen wir Montag, von elf bis zwölf sowas. Montag, der fünfte April dann.
(8-V) guten Tag, Frau Diesner. ich habe Sie vorhin angerufen, weil ich wollte, daß wir Angesicht zu Angesicht einen Termin ausmachen. und weil sich es um ein fünftägiges Arbeitstreffen in der Filiale *AGTR* in Bonn handelt, dachte ich mir, ist am besten, wir sitzen uns hier gegenüber. und ich habe meinen Terminkalender mitgebracht und vielleicht könnten Sie dann kucken, wie es bei Ihnen aussieht .
(9-T) Hatte Fabian als Textgrundlage seiner nur schwer zu dechiffrierenden Bilderwelt einen eigenen Text mit dem vielsagenden Titel Das unsichtbare Genie geschrieben, so blieben in Thomas Roths Experiment nur diese Textsplitter aus dem Original übrig: “Wenn er der Pest nicht schleunig uns entreißt, ... so steigt der Leiche eines ganzen Volkes dies Land , ein Grabeshügel, aus der See”.
(10-T) Der Megatrend der Medienentwicklung beschäftigte die meisten Ref-

erenten und Teilnehmer. Hauke Brunkhorst warnte vor kulturpessimistischer Aufgeregtheit – angefangen mit Musik und Theater bis zum Internet werde jedem Medium unterstellt, daß es den Menschen sich selbst entfremde.
(11-V) ja , sollen wir uns dann vielleicht darüber jetzt unterhalten, damit wir den Flug uns danach einrichten können?
(12-V) ja, auf jeden Fall. möchte auch so ein bißchen was von Hannover sehen, daß wir vielleicht auch ein paar Sehenswürdigkeiten uns anschauen, würde ich sagen? wenn Sie da Interesse dran haben?
(13-V) ja, also, eine Woche müßte drin sein. ich würde vorschlagen daß wir also, daß wir die achtunddreißigste Woche uns dafür ausgucken.
(14-V) daß wir Donnerstag vormittag einfach uns vornehmen. ja, das das ist prima.
(15-V) wenn wir schon so viel uns anschauen.
(16-V) ach, wenn ich mir es so recht überlege, wäre der Samstag lieber.
(17-V) obwohl, wenn Sie uns sie zuschicken könnten, wäre auch nicht so schlecht.

4 Note that only a subset of the ditransitive clauses considered in Section 2 (cf. Tables 1 and 2) were included in Figures 2 and 3: Especially in the VERBMOBIL dialogues, quite a few clauses contain more than one pronominal argument NP.

5 A second, less salient discrepancy concerns the ditransitive family members whose ordering pattern complies with the production–based linearization rule, i.e., the pairs to the left of the thick vertical lines in Figures 2 and 3: The member with the highest grammaticality rating does not always have the highest frequency. Instead, the most frequent one tends to be the "primary" ordering in the production–based rule. The only exception seems to be the (*S,Ip,O*) family, where the two orderings left of the vertical line have nearly identical frequency scores. However, in the midfield of *main* clauses in the written corpora, the highest frequency score is obtained by the ordering whose grammaticality rating is second best, i.e. by *Ip–S–O* — also the "primary" one. We will not discuss this phenomenon any further here, except for noting that this frequency pattern is a second reason, in addition to the one mentioned at the end of Section 3, why the ordering Sp–Op–Ip–S–I–O deserves the appellation "primary" for transitive clauses.
For the sake of completeness, we list here the frequencies of the ordering patterns in the main clauses from NEGRA–II and TIGER:
Ditransitives: SIO: 56; SOI: 4; SpOI: 1; SpIO: 18; SIpO 8; IpSO: 59; SOpI: 2; OpSI: 12; OpIS: 4. This sums up to 164 clauses for the four ditransitive families. In addition, we found 39 clauses with two or three pronominal arguments.
The mono-transitives yielded the following frequencies: SO: 1415; OS: 14; SpO: 487; SOp: 31; OpS: 580; SpOp: 89.

References

Bard, Ellen G., Dan Robertson, and Antonella Sorace
1996 Magnitude estimation of linguistic acceptability. *Language*, 72: 32–68.

Brants, Sabine, Stefanie Dipper, Sylvia Hansen, Wolfgang Lezius, and George Smith
2002 The TIGER Treebank. In Erhard Hinrichs and Kiril Simov, (eds.), *Proceedings of the Workshop on Treebanks and Linguistic Theories (TLT)*, pp. 24–41. Sozopol, Bulgaria.

Burger, Susanne, Karl Weilhammer, Florian Schiel, and Hans G. Tillmann
2000 Verbmobil data collection and annotation. In Wolfgang Wahlster, (ed.), *Verbmobil: Foundations of Speech–to–Speech Translation*, pp. 537–549. Springer, Berlin, Germany.

Eisenberg, Peter
1994 *Grundriss der Deutschen Grammatik (3rd revised edition)*. J.B. Metzler Verlag, Stuttgart, Germany.

Hawkins, John A.
2004 *Efficiency and complexity in grammars*. Oxford University Press, Oxford, U.K.

Keller, Frank
2000 *Gradience in grammar: Experimental and computational aspects of degrees of grammaticality*. Ph.d. thesis, University of Edinburgh, Edinburgh, UK.

Kempen, Gerard and Karin Harbusch
2003 Word order scrambling as a consequence of incremental sentence production. In Holden Härtl and Heike Tappe, (eds.), *Mediating between concepts and language — Processing structures*, pp. 141–164. Mouton De Gruyter, Berlin, Germany.
2004a A corpus study into word order variation in German subordinate clauses: Animacy affects linearization independently of grammatical function assignment. In Thomas Pechmann and Christopher Habel, (eds.), *Multidisciplinary approaches to language production*, pp. 173–181. Mouton De Gruyter, Berlin, Germany.
2004b Generating natural word orders in a semi–free word order language: Treebank–based linearization preferences for argument NPs in subordinate clauses of German. In Alexander Gelbukh, (ed.), *Proceedings of the Fifth International Conference on Intelligent Text Processing and Computational Linguistics (CICLING), Seoul, Korea*, pp. 350–354. Springer, Lecture Notes in Computer Science, Berlin, Germany.

König, Esther and Wolfgang Lezius
2000 A description language for syntactically annotated corpora. In *Proceedings of the International Conference on Computational Linguistics (COLING), Saarbrücken, Germany*, pp. 1056–1060.

Kurz, Daniela
2000 A statistical account on word order variation in German. In Anne Abeillé, Thorsten Brants, and Hans Uszkoreit, (eds.), *Proceedings of the COLING Workshop on Linguistically Interpreted Corpora (LINC), Luxembourg*.

Müller, Gereon
1999 Optimality, markedness, and word order in German. *Linguistics*, 37: 777–815.

Pechmann, Thomas, Hans Uszkoreit, Johannes Engelkamp, and Dieter Zerbst
1996 Wortstellung im deutschen Mittelfeld. Linguistische Theorie und psycholinguistische Evidenz. In *Perspektiven der Kognitiven Linguistik*, pp. 257–299. Westdeutscher Verlag, Oplaten, Germany.

Rambow, Owen
1994 *Formal and Computational Aspects of Natural Language Syntax*. Ph.d. thesis, University of Pennsylvania, Philadelphia, PA, USA.

Skut, Wojciech, Brigitte Krenn, Thorsten Brants, and Hans Uszkoreit
1997 An annotation scheme for free word order languages. In *Proceedings of the Fifth Conference on Applied Natural Language Processing (ANLP)*, pp. 27–28. Washington D.C., USA.

Uszkoreit, Hans
1987 *Word Order and Constituent Structure in German*. CSLI Publication, Stanford, CA, USA.

Wasow, Thomas
2002 *Postverbal behavior*. CSLI Publications, Lecture Notes 145, Stanford, CA, USA.