Evaluating Competition-based Models of Word Order

Frank Keller
keller@cogsci.ed.ac.uk
Institute for Communicating and Collaborative Systems
Division of Informatics, University of Edinburgh
2 Buccleuch Place, Edinburgh EH8 9LW, UK

Abstract

The ordering of constituents in semi-free word order languages has attracted considerable attention in theoretical linguistics. Three types of models have been proposed to explain word order preferences, based on (a) weighted constraints, (b) Optimality Theory (c) syntactic weight. All three models use grammatical competition to explain the interaction of word order constraints. They rely on intuitive judgments or corpus studies, but have not been evaluated against experimental data. This is the purpose of the present paper. We report the results of a magnitude estimation experiment investigating word order in German, focusing on the interaction of verb position, case marking, pronominalization, and information structure. The experimental data are compatible with models (a) and (b), indicating that relativized (ranked or weighted) constraints are essential in explaining word order preferences. Model (c), on the other hand, is not compatible with the data.

Introduction

The languages of the world differ substantially in the degree of word order variation they allow. On the one end of the spectrum, we find languages like English, which exhibit a relatively fixed word order. On the other end, there are languages like Warlpiri (an Australian language), which allow a large degree of word order variation. Many languages exhibit a semi-free word order, i.e., the word order is fixed in some respects, but variable in others.

Word order variation typically manifests itself not in binary acceptability judgments, but in the form of word order preferences, to which a diverse set of factors contribute, including syntactic, pragmatic, and phonological factors. This poses an interesting challenge for linguistic theory, which is equipped to deal with binary ungrammaticality resulting from the violation of individual linguistic constraints, rather than with preferences (degrees of acceptability) that emerge from the interaction of a diverse set of factors.

A number of approaches have been developed to deal with this challenge, all of which diverge from conventional linguistic frameworks in assuming a relative (weighted or ranked) rather than an absolute (binary) notion of linguistic constraints. Three main types of models have been proposed, based on (a) weighted constraints (Jacobs, 1988; Uszkoreit, 1987), (b) Optimality Theory (Choi, 1996; Müller, 1999), (c) syntactic weight (Hawkins, 1992). All of these models use a notion of grammatical competition to explain the interaction of the factors that influence word order.

Models (a)–(c) rely on informal, intuitive acceptability judgments (and on corpus data in the case of (c)). It is safe to assume that such judgments allow to determine binary acceptability reliably. However, their reliability is much less obvious with respect to degrees of acceptability like the ones that occur in word order data (Bard, Robertson, & Sorace, 1996; Cowart, 1997; Schütze, 1996). This makes it desirable to evaluate linguistic models of word order against experimentally collected acceptability data.

The purpose of the present paper is to provide a first step towards such an experimental evaluation.[1] The methodology we use is magnitude estimation, which has been shown to yield reliable, yet maximally delicate judgments of linguistic acceptability (Bard et al., 1996). The empirical domain of our investigation is the variation in the order of verb complements in German, a semi-free word order language. We outline the necessary linguistic background in the following section.

Word Order in German

German has a fixed verb order. Subordinate clauses are verb final, while yes/no questions require verb initial order, and declarative main clauses have the verb in second position. In the generative literature, the subordinate clause order is generally considered the basic order from which the main clause and question orders are derived by movement (e.g., Haider, 1993). The present experiment will focus on subordinate clauses (which is also customary in the processing literature on German, e.g., Bader & Meng, 1999). Using subordinate clauses avoids potential confounds from topicalization and other phenomena that can occur in verb second clauses.

While verb order is fixed in German, the order of the complements of the verb is variable, and a number of factors have been claimed to influence complement order. These factors include case marking, thematic roles, pronominalization, information structure, intonation, definiteness, and animacy (Choi, 1996; Jacobs, 1988; Müller, 1999; Uszkoreit, 1987; Scheepers, 1997). The present study focuses on the effect of case marking, pronominalization, and information structure on word order.

We elicited acceptability judgments for four subordinate clause orders, illustrated by the examples in (1). As mentioned above, subordinate clauses in German require verb final order (see (1a), (1b)). Verb initial orders (see (1c), (1d)) give rise to strong unacceptability.

[1] Previous experimental work on word order preferences in German (Pechmann, Uszkoreit, Engelkamp, & Zerbst, 1994; Scheepers, 1997) only dealt with isolated stimuli, i.e., failed to address contextual effects on order, one of the topics of the present paper

(1) a. **SOV:**
Maria glaubt, dass der Vater den Wagen kauft.
M.-NOM believes that the father-NOM the car-ACC buys
'Maria believes that the father will buy the car.'
b. **OSV:** Maria glaubt, dass den Wagen der Vater kauft.
c. **VSO:** Maria glaubt, dass kauft der Vater den Wagen.
d. **VOS:** Maria glaubt, dass kauft den Wagen der Vater.

We also examined the influence of pronominalization on word order. The experiment included sentences where none of the NPs was pronominalized (see (1)), but also sentences where the subject, object, or both the subject and the object were pronominalized (see (2)).[2]

(2) a. Maria glaubt, dass er den Wagen kauft.
Maria-NOM believes that he-NOM the car-ACC buys
'Maria believes that he will buy the car.'
b. Maria glaubt, dass der Vater ihn kauft.
Maria-NOM believes that the father-NOM it-ACC buys
'Maria believes that the father will buy it.'
c. Maria glaubt, dass er ihn kauft.
Maria-NOM believes that he-NOM it-ACC buys
'Maria believes that he will buy it.'

Information structure figures as a determinant of complement order in the accounts of Choi (1996), Jacobs (1988), Müller (1999), and Uszkoreit (1987). Information structural effects can be studied by embedding the sentence in a question context: the *wh*-phrase marks the focussed constituent, while the other constituents are non-focussed, or ground (Vallduví, 1992). The following contexts were used in the experiment:

(3) a. **All Focus:** Was gibt's neues?
'What's new?'
b. **S Focus:** Wer kauft den Wagen?
'Who will buy the car?'
c. **O Focus:** Was kauft der Vater?
'What will the father buy?'

A null context condition was included as a control, allowing us to study how subjects react in the absence of any contextual information.

Models of Word Order

Weighted Constraints

Uszkoreit (1987) models word order preferences using weighted constraints. In such a setting, linguistic constraints are annotated with a numeric weight that reflects their importance in determining grammaticality (for a similar proposal, see Jacobs, 1988). Uszkoreit assumes constraint competition, i.e., not all constraints are necessarily satisfiable in a given linguistic structure. This entails that grammaticality is a gradient notion; the degree of grammaticality of a linguistic structure is computed as the sum of the of the weights of the constraint violations the structure incurs.

Uszkoreit (1987, p. 114) proposes the following constraints on word order in German (constraints irrelevant to the data under consideration are omitted and constraint names are provided):

(4) a. VERB: X $\prec$ V[−MC]
b. NOM: [+NOM] $\prec$ [+ACC]
c. FOC: [−FOCUS] $\prec$ [+FOCUS]
d. PRO: [+PRO] $\prec$ [−PRO]

These constraints are constituent order constraints, with '$\prec$' denoting linear precedence. The constraint VERB relies on the feature MC (main clause) to specify verb order; if this feature is negative (i.e., in a subordinate clause), then the verb has to succeed any other constituent. The constraint NOM requires that nominative precedes accusative. The information structural requirement FOC specifies that ground constituents (marked [−FOCUS]) precede focused constituents. The constraint PRO requires pronouns to precede full NPs.

Uszkoreit does not provide weights for the constraints in (4).[3] Intuitively, however, we expect a violation of VERB to lead to serious unacceptability, i.e., VERB should receive a higher weight than the other constraints.

Optimality Theory

Standard Optimality Theory (OT; Prince & Smolensky, 1993) assumes a binary notion of grammaticality; a linguistic structure is either optimal (and thus grammatical) or suboptimal (and thus ungrammatical). However, OT can be extended to model gradient grammaticality; Müller (1999) puts forward a modified version of OT based on the distinction between grammaticality (manifested in binary judgments) and markedness (associated with word order preferences). Grammaticality is handled in terms of conventional OT-style constraint competition. This competition can yield several grammatical candidates, among which further competition takes place based on markedness constraints. The markedness competition then induces a preference order on the candidates that predicts their relative acceptability. (Note that the grammaticality/markedness dichotomy is reminiscent of the distinction of hard and soft constraints proposed by Keller (1998).)

In Müller's account, the constraints on pronoun order belong to the realm of grammaticality, while the constraints on case order and focus-ground order (among others) belong to the realm of markedness. We omit technical details and only state constraints relevant to the present data set:

(5) a. NOM: [+NOM] $\prec$ [−NOM]
b. FOC: [−FOCUS] $\prec$ [+FOCUS]
c. AN: [+ANIMATE] $\prec$ [−ANIMATE]

Note that the constraints NOM and FOC are similar to Uszkoreit's constraints in (4). AN is an additional constraint that requires animate NPs to precede inanimate ones. In contrast to Uszkoreit, Müller postulates an explicit constraint ranking:

(6) NOM $\gg$ AN $\gg$ FOC

In addition to the markedness constraints in (5), a set of grammaticality constraints is postulated (omitted here). These constraints deal with pronoun order and ensure that pronouns occur at the left periphery of the clause. All candidates that fail to meet this requirement are predicted to be (categorically) ungrammatical. In contrast to Uszkoreit, Müller does not include constraints on verb order.

[2] Note that only masculine NPs were used, as these are unambiguous in their case marking, both as full NPs and as pronouns (while the case morphology of feminine and neuter NPs exhibits syncretism).

[3] Pechmann et al. (1994) tentatively assume that all constraints have equal weights, which entails that the degree of unacceptability only depends on the number of violations.

Syntactic Weight

Hawkins (1992) proposes an approach to word order preferences that also relies on grammatical competition, but makes very different assumptions concerning the source of this competition. Hawkins assumes that constituent order is determined by the syntactic weight of the constituents, a notion that is supposed to reflect how easily the constituents can be recognized by the human parser. According to Hawkins (1992, p. 200), relative syntactic weight explains word orders frequencies in corpora, as well as the relative acceptability of different orders in native speaker's judgments.

Hawkins proposes Immediate Constituent to Word Ratio (ICR) as a metric for syntactic weight. Intuitively, ICR measures the length of a constituent relative to its position in the clause (see Hawkins, 1992, for details). If two sentences differ in average ICR, the one with the higher average ICR is predicted to be more acceptable. The ICR for a given word is calculated as n/m, where n is the number of the constituent, while m is the number the word, counted from left to right. The average ICR for a sentence is obtained by averaging the ICRs of its words. As an example, consider the ICRs for the sentences in (1):

(7) a. M. glaubt, dass [[der Vater] [den Wagen] kauft.] ICR
1/1 1/2 2/3 2/4 3/5 .65

b. M. glaubt, dass [kauft [der Vater] [den Wagen.]] ICR
1/1 2/2 2/3 3/4 .86

Provided that subject and object have the same length, SO and OS orders receive the same ICR, i.e., examples (1a) and (1b) both have an ICR of .65 (see (7a)), while examples (1c) and (1d) both receive an ICR of .86 (see (7b)). For pronominalized NPs, the following ICRs are predicted:

(8) a. Maria glaubt, dass [[er] [den Wagen] kauft.] ICR
1/1 2/2 2/3 3/4 .86

b. Maria glaubt, dass [[der Vater] [ihn] kauft.] ICR
1/1 1/2 2/3 3/4 .73

c. Maria glaubt, dass [[er] [ihn] kauft.] ICR
1/1 2/2 3/3 1.0

This means that Hawkins's account predicts that pronouns have to precede full NPs (if they are longer than a single word). However, if both the subject and the object are pronouns, then both SO and OS receive an ICR of 1.0, i.e., they should be equally acceptable.

Note that Hawkins predicts that information structure (focus and ground) should not play a role in determining word order preferences, contrary to claims by Müller (1999) and Uszkoreit (1987), among others.

Experiment

Method

Subjects Fifty-one native speakers of German participated in the experiment. All participants were naive to syntactic theory.

Materials A factorial design was used that crossed the factors verb order (*Vord*), complement order (*Cord*), pronominalization (*Pro*), and context (*Con*). The factor *Con* had four levels: null context, all focus, S focus, and O focus, as illustrated in (3). The factor *Vord* had four two levels: verb final (see (1a), (1b)) and verb initial (see (1c), (1d)). The two levels of *Cord* were subject before object and object before subject, as in (1a), (1c) and (1b), (1d). In the null context condition, the factor *Pro* had four levels, viz., both S and O full NPs, S pronoun and O full NP, S full NP and O pronoun, and both S and O pronouns (see (2)). In the context condition, *Pro* only had two levels, viz., no pronoun and pronoun. In the all focus and S focus contexts, the object was pronominalized, while in the O focus context, the subject was pronominalized. This design ensures that the pronoun is interpreted as ground and hence is unstressed (as the sentential stress has to fall on the focussed constituent). We are only interested in the syntactic behavior of weak (i.e., unstressed) pronouns; strong (i.e., stressed) pronouns are subject to different syntactic constraints.

This yielded a total of $\mathit{Vord} \times \mathit{Cord} \times \mathit{Pro} = 2 \times 2 \times 4 = 16$ cells for the null context condition, and $\mathit{Vord} \times \mathit{Cord} \times \mathit{Pro} \times \mathit{Con} = 2 \times 2 \times 2 \times 3 = 24$ cells for the context condition. Eight lexicalizations per cell were used, which resulted in a total of 320 stimuli. A set of 24 fillers was used in the null context condition; 16 fillers were employed in the context condition. The fillers were designed to cover the whole acceptability range.

To control for possible effects from lexical frequency, the lexicalizations for subject, object, and verb were matched for frequency. Frequency counts for the verbs and the head nouns were obtained from a lemmatized version of the Frankfurter Rundschau corpus (40 million words of newspaper text) and the average frequencies were computed for subject, object, and verb lexicalizations. An ANOVA confirmed that these average frequencies were not significantly different from each other.

Procedure The method used was magnitude estimation as proposed by Stevens (1975) for psychophysics and extended to linguistic stimuli by Bard et al. (1996) and Cowart (1997).

Subjects first saw a set of instructions that explained the concept of numerical magnitude estimation using line length. Subjects were instructed to make length estimates relative to the first line they would see, the reference line. They were told to give the reference line an arbitrary number, and then assign a number to each following line so that it represented how long the line was in proportion to the reference line. Several example lines and corresponding numerical estimates were provided to illustrate the concept of proportionality. Then subjects were told that linguistic acceptability could be judged in the same way as line length. The concept of linguistic acceptability was not defined, but examples of acceptable and unacceptable sentences were provided.

The experiment started with a training phase designed to familiarize subjects with the magnitude estimation task. Subjects had to estimate the length of a set of lines. Then, a set of practice items (similar to the experimental items) were administered to familiarize subjects with applying magnitude estimation to linguistic stimuli. Finally, subjects had to judge the experimental items. A between subjects design was used to administer the factor CON: subjects in Group A judged non-contextualized stimuli, while subjects in Group B judged contextualized stimuli. The factors *Vord*, *Cord*, and *Pro* were administered within subjects. Using a Latin square design, eight lexicalizations were created for each group. The lexi-

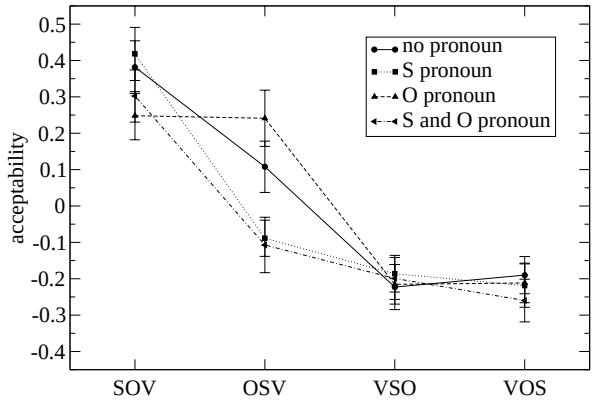

Figure 1: Interaction for word order and pronominalization, null context

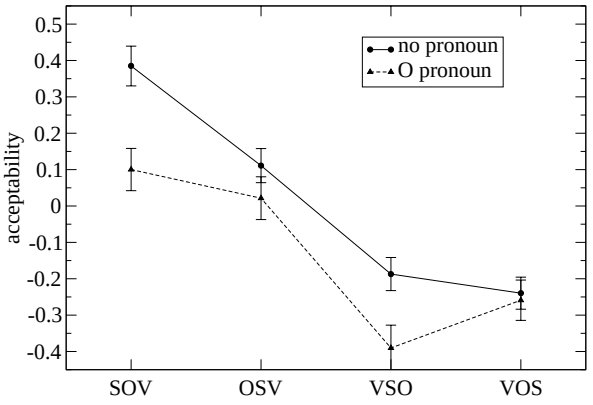

Figure 2: Interaction for word order and pronominalization, all focus context

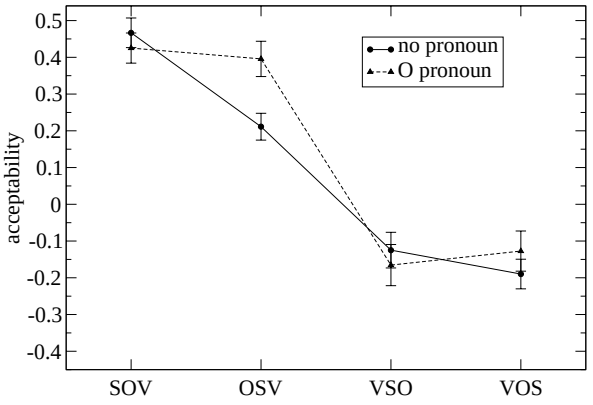

Figure 3: Interaction for word order and pronominalization, subject focus context

calizations for Group A contained 16 items, while the ones for Group B contained 24 items.

Each subject saw one lexicalization and 24 fillers in Group A or one lexicalization and 16 fillers in Group B, i.e., a total of 40 items per group. Each subject was randomly assigned to a group and a lexicalization: 20 subjects were assigned to Group A, and 31 to Group B. Instructions, examples, training items, and fillers were adapted for Group B to take context into account.

Results

The data were normalized by dividing each numerical judgment by the modulus value that the subject had assigned to the reference sentence. This operation creates a common scale for all subjects. All analyses were carried out on the geometric means of the normalized judgments, as is standard for magnitude estimation data (Bard et al., 1996; Cowart, 1997).

In discussing the results, we make use of the following abbreviations: SO for subject before object, OS for object before subject, XV for verb final, VX for verb initial. The indices 'pro' and 'full' indicate pronouns and full NPs, respectively. For instance, $VS_{full}O_{pro}$ stands for an VSO order where the subject is a full NP and the object is a pronoun.

Null Context Condition Figure 1 graphs the average judgments for each word order. An ANOVA for the null context condition revealed a highly significant main effect of *Vord* (verb order) ($F_1(1,19) = 56.911$, $p < .0005$; $F_2(1,7) = 621.924$, $p < .0005$): XV orders (mean $= .1879$) were more acceptable than VX orders (mean $= -.2129$). A highly significant main effect of *Cord* (complement order) was also obtained ($F_1(1,19) = 26.966$, $p < .0005$; $F_2(1,7) = 72.610$, $p < .0005$): SO orders (mean $= .0659$) were more acceptable than OS orders (mean $= -.0909$). The main effect of *Pro* (pronominalization) was significant by subjects only ($F_1(3,57) = 5.150$, $p = .003$; $F_2(3,21) = 0.647$, $p = .593$).

The ANOVA also revealed a significant interaction of *Cord* and *Pro* ($F_1(3,57) = 13.026$, $p < .0005$; $F_2(3,21) = 4.663$, $p = .012$). This indicates that pronominalization has an influence on complement order preference. We also found interactions of *Cord* and *Vord* ($F_1(1,19) = 47.437$, $p < .0005$; $F_2(1,7) = 17.148$, $p = .004$) and of *Vord* and *Pro* (significant by subjects only, $F_1(3,57) = 4.223$, $p = .009$; $F_2(3,21) = 1.107$, $p = .368$). A three-way interaction *Vord/Cord/Pro* was also present (significant by subjects only, $F_1(3,57) = 7.415$, $p = .009$; $F_2(3,21) = 1.900$, $p = .161$).

The meaning of the interactions involving *Vord* becomes clear from Figure 1: the effect of pronominalization on complement order is limited to verb final orders; all verb initial orders are equally unacceptable, independent of complement order and pronominalization.

Context Condition Figures 2–4 graph the average judgments for each context. An ANOVA for the context condition confirmed the main effect of verb order found in the null context condition ($F_1(1,30) = 121.507$, $p < .0005$; $F_2(1,7) = 225.903$, $p < .0005$): XV orders (mean $= .2519$) were more acceptable than VX orders (mean $= -.1973$). The main effect of complement order could also be replicated ($F_1(1,30) = 40.275$, $p < .0005$; $F_2(1,7) = 15.359$, $p = .006$): SO orders (mean $= .0785$) were more acceptable than OS orders (mean $= -.0239$). A highly significant main effect of *Con* (context) was also present ($F_1(2,60) = 28.953$, $p < .0005$; $F_2(2,14) = 54.056$, $p < .0005$), as well as a weak effect of *Pro* ($F_1(2,60) = 5.564$, $p = .025$; $F_2(2,14) = 1.511$, $p = .259$).

The ANOVA uncovered an interaction of *Cord* and context,

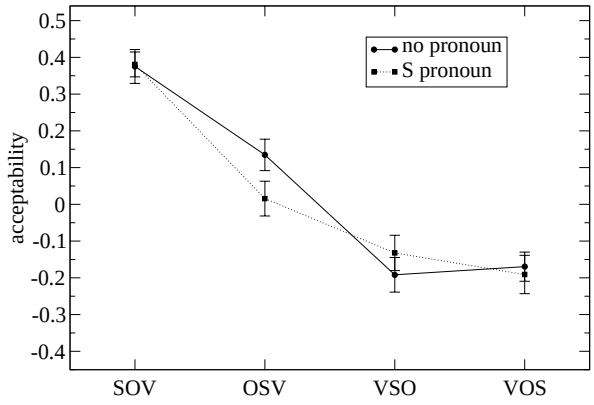

Figure 4: Interaction for word order and pronominalization, object focus context

significant by subjects and marginal by items ($F_1(2,60) = 6.016$, $p = .004$; $F_2(2,14) = 3.076$, $p = .078$), which confirms that information structure (manipulated by context) has an influence on complement order preferences. We also found a marginal interaction of *Cord* and *Pro* ($F_1(1,30) = 4.025$, $p = .054$; $F_2(1,7) = 3.634$, $p = .098$) and a highly significant interaction of *Pro* and context ($F_1(2,60) = 11.864$, $p < .0005$; $F_2(2,14) = 16.07$, $p < .0005$). Recall that our materials were designed such that in all focus and S focus contexts, the object was pronominalized, while in an O focus context, the subject was pronominalized. This means that the *Cord/Pro* and *Pro/Con* interactions are only meaningful with respect to the three-way interaction *Cord/Pro/Con* which was also significant ($F_1(2,60) = 19.718$, $p < .0005$; $F_2(2,14) = 7.73$, $p = .005$). This interaction demonstrates that the ordering of pronouns is subject to contextual effects (which will be discussed in the next section). The ANOVA also showed an interaction of *Vord* and *Cord* ($F_1(1,30) = 50.960$, $p < .0005$; $F_2(1,7) = 7.221$, $p = .031$) and of *Vord* and context ($F_1(2,60) = 10.589$, $p < .0005$; $F_2(2,14) = 11.945$, $p = .001$). The meaning of these interactions becomes clear from Figures 2–4: the interaction between complement order and context is limited to verb final orders; all verb initial orders are equally unacceptable, independent of context.

Discussion

All differences referred to in the following were significant according to post-hoc Tukey tests (space limitations prevent the inclusion of the full set of Tukey results).

Weighted Constraints The experimental findings for the null context condition provided broad support for the ordering constraints in (4), initially proposed by Uszkoreit (1987). There was a clear preference for XV over VX, in line with the predictions of the verb ordering constraint VERB. The NOM constraint, which requires nominative to precede accusative, received support from the fact that SO orders were more acceptable than OS orders. Finally, the constraint PRO, which requires that pronouns precede full NPs, explains why $S_{full}O_{pro}$ is less acceptable than $S_{pro}O_{full}$, while $O_{full}S_{pro}$ is less acceptable than both $O_{full}S_{full}$ and $O_{pro}S_{full}$ (see Figure 1).

The interactions involving the factor *Vord* demonstrated that the effects of NOM and PRO disappear if the constraint VERB is violated. This indicates that a violation of VERB is more serious than violations of PRO or NOM, which in Uszkoreit's framework means that VERB receives a higher weight than both PRO and NOM.

The behavior of VERB was replicated in the context condition. VERB leads to serious unacceptability in all contexts and blocks out all other constraint effects. Note, however, that we found an interaction of PRO and context that does not readily follow from Uszkoreit's account. The prediction that pronouns have to precede full NPs is only born out in the all focus context. In the S focus and O focus contexts, the effect of PRO disappears, which might indicate that PRO is only valid if the context fails to provide an antecedent for the pronoun.

S focus and O focus contexts showed evidence for FOC, the constraint that requires ground elements to precede focus elements. In both contexts, SO was the preferred order, even though it violates FOC in the S focus context; in both contexts, the acceptability of OS was reduced compared to SO. However, this reduction was significantly higher in the O focus context, where OS violates FOC. The overall SO preference (even if it is disfavored by the context) indicates that the effect of FOC is weak compared to the influence of NOM, i.e., NOM should receive a higher weight than FOC. Only for OS orders, i.e., when NOM is violated, the influence of FOC becomes visible. No effects of context were found for VX orders, which indicates that FOC has a lower weight than VERB, just like NOM and PRO.

Overall, we have established the following facts about constraint weights: VERB has a higher weight than PRO, NOM, and FOC. NOM, on the other hand, has a greater weight than FOC. This is compatible with the following weight assignments:

(9) $w(\text{VERB}) = 3, \quad w(\text{PRO}) = w(\text{NOM}) = 2, \quad w(\text{FOC}) = 1$

To conclude, our results provide support for Uszkoreit set of word order constraints and show that his weighted constraint model is able to account for the experimental data.

Optimality Theory Note that the weights in (9) can also be interpreted as a set of OT-style constraint ranks:

(10) $\text{VERB} \gg \{\text{PRO}, \text{NOM}\} \gg \text{FOC}$

This ranking is compatible with Müller's ranking in (6). Note that the effect of the AN (animacy) could not be tested in the present experiment: all nominative NPs were animate, while all accusative NPs were inanimate, hence a violation of NOM also entails a violation of AN.

Müller distinguishes between grammaticality and markedness, and predicts that ungrammatical candidates are categorically unacceptable, while marked structures are only dispreferred. PRO is a classified as a grammatical constraint, and hence should induce categorical unacceptability. Our data provides counterevidence to this prediction: in the null context, $S_{full}O_{pro}V$ and $O_{pro}S_{full}V$ are equally acceptable (see Figure 1), even though $S_{full}O_{pro}V$ violates PRO while $O_{pro}S_{full}V$ does not (the same pattern occurs in the all focus and S focus contexts). This is unexpected under the assumption that PRO is grammatical constraint; the data suggest that it should be reclassified as a markedness constraints.

On the other hand, VERB seems to be a genuine grammatical constraint. Its violation leads to strong ungrammaticality in all contexts, independently of which other constraints are violated (see Figures 1–4). This indicates that VERB (not explicitly dealt with by Müller) is a grammatical constraint.

Hence our data provides counterevidence for some aspects of Müller's particular account of word order in German. However, the main feature of his model, viz., the distinction between grammaticality and markedness, is supported by our experimental results.

Syntactic Weight Several of the order preferences found in this experiment are incompatible with Hawkins's account in terms of ICR. Most strikingly, we found that $S_{full}O_{full}V$ is more acceptable than $O_{full}S_{full}V$ (see Figures 1–4), even though both have the same ICR (see (7)).

As far as pronominalization is concerned, we found that in the null context, $S_{pro}O_{full}V$ is more acceptable than $S_{full}O_{pro}V$ and $O_{pro}S_{full}V$ is more acceptable than $O_{full}S_{pro}V$ (see Figure 1), consistent with the ICR predictions (see (8)). However, the predictions with respect to double pronouns were not born out: these receive the maximum ICR score of 1.0, but we found that the orders $S_{pro}O_{pro}V$ and $O_{pro}S_{pro}V$ are as unacceptable as $S_{full}O_{pro}V$ and $O_{full}S_{pro}V$, respectively, even though these orders only have an ICR of .73 (see (8)). Also, the fact that $S_{full}O_{pro}V$ and $O_{pro}S_{full}V$ are equally acceptable is unexpected as these orders differ in ICR (see (8)). This observation holds across contexts, see Figures 1–4.

Also the focus effects we found are unexpected under a syntactic weight account: the acceptability of OSV is increased in an O focus context (compared to an S focus context, see Figures 3, 4), even though the ICR remains constant.[4] Finally, the fact that VX structures are severely unacceptable across the board does not follow from syntactic weight—in fact VX orders have a higher ICR than XV orders (see (7)).[5]

To summarize, while corpus data seems to support a syntactic weight account (see Hawkins, 1992, for details), the acceptability judgments in our experiment are largely incompatible with Hawkins's predictions.

Conclusions

We reported the results of a study of word order variation in German that investigated the interaction of syntactic (complement order and verb order) and information structural constraints (pronominalization and focus). The data were used to evaluate a set of competition-based models of word order, including (a) Uszkoreit's (1987) weighted constraint model, (b) Müller's (1999) optimality theoretic account, and (c) Hawkins's (1992) syntactic weight model.

The experimental data are broadly compatible with models (a) and (b), indicating that a relativized (ranked or weighted) notion of linguistic constraints is essential for explaining word order preferences. Model (c), however, was not well-supported by the data. While this model may be suitable for describing word order distributions in corpora, it does not seem to be directly applicable to contextualized acceptability judgments such as the ones reported in the present paper.

On the other hand, we found that some of the individual linguistic assumptions made by Uszkoreit and Müller were not born out in our data. This highlights the fact that informal acceptability judgments are not sufficient to clarify the intricate preference patterns that emerge from the interaction of syntactic, pragmatic, and phonological constraints on word order. Experimentally collected judgments are necessary to obtain reliable, delicate data that can inform detailed models of word order preferences.

The results of the present study have been replicated for a free word order language (Greek) and for spoken stimuli (Keller & Alexopoulou, 2000).

[4] Note that Hawkins (1992, p. 196) concedes that informational concepts like focus play a limited role in 'structures for which syntactic weight makes either no predictions or weak predictions'.

[5] However, Hawkins argues that languages can grammaticalize word orders, which then are no longer subject to syntactic weight. This would explain the general unacceptability of VX in subordinate clauses in German.

References

Bader, M., & Meng, M. (1999). Subject-object ambiguities in German embedded clauses: An across-the-board comparison. *Journal of Psycholinguistic Research*, *28*(2), 121–143.

Bard, E. G., Robertson, D., & Sorace, A. (1996). Magnitude estimation of linguistic acceptability. *Language*, *72*(1), 32–68.

Choi, H.-W. (1996). *Optimizing structure in context: Scrambling and information structure.* Unpublished doctoral dissertation, Stanford University.

Cowart, W. (1997). *Experimental syntax: Applying objective methods to sentence judgments.* Thousand Oaks, CA: Sage Publications.

Haider, H. (1993). *Deutsche Syntax generativ: Vorstudien zur Theorie einer projektiven Grammatik.* Tübingen: Gunter Narr.

Hawkins, J. A. (1992). Syntactic weight versus information structure in word order variation. In J. Jacobs (Ed.), *Informationsstruktur und Grammatik* (pp. 196–219). Opladen: Westdeutscher Verlag.

Jacobs, J. (1988). Probleme der freien Wortstellung im Deutschen. In I. Rosengren (Ed.), *Sprache und Pragmatik* (Vol. 5, pp. 8–37). Department of German, Lund University.

Keller, F. (1998). Gradient grammaticality as an effect of selective constraint re-ranking. In M. C. Gruber, D. Higgins, K. S. Olson, & T. Wysocki (Eds.), *Papers from the 34th Meeting of the Chicago Linguistic Society* (Vol. 2: The Panels, pp. 95–109). Chicago.

Keller, F., & Alexopoulou, T. (2000). *Phonology competes with syntax: Experimental evidence for the interaction of word order and accent placement in the realization of information structure* (Rutgers Optimality Archive No. ROA-351-1099).

Müller, G. (1999). Optimality, markedness, and word order in German. *Linguistics*, *37*(5), 777–818.

Pechmann, T., Uszkoreit, H., Engelkamp, J., & Zerbst, D. (1994). *Word order in the German middle field: Linguistic theory and psycholinguistic evidence* (CLAUS Report No. 43). Department of Computational Linguistics, Saarland University.

Prince, A., & Smolensky, P. (1993). *Optimality Theory: Constraint interaction in generative grammar* (Technical Report No. 2). Center for Cognitive Science, Rutgers University.

Scheepers, C. (1997). *Menschliche Satzverarbeitung: syntaktische und thematische Aspekte der Wortstellung im Deutschen.* Unpublished doctoral dissertation, University of Freiburg.

Schütze, C. T. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology.* Chicago: University of Chicago Press.

Stevens, S. S. (Ed.). (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects.* New York: John Wiley.

Uszkoreit, H. (1987). *Word order and constituent structure in German.* Stanford, CA: CSLI Publications.

Vallduví, E. (1992). *The informational component.* New York: Garland.