The Object Order in the German Middle Field through the Lens of Information Theory: A Diachronic Study

Sophia Voigtmann (University of Kassel)
sophia.voigtmann@uni-kassel.de

The relative order of dative (Dat) and accusative (Acc) objects in German is variable. Dat>Acc (1a) is the canonical order, but Acc>Dat (1b) can be found when the Acc is given (Lenerz, 1977; Rauth, 2020; Speyer, 2011, 2015, 2016).

1. a) Ich gebe [einem Athleten]$_{Dat}$ [den Ball]$_{Acc}$.
   I give [an athlete] [the ball].
   'I gave an athlete the ball.'
   b) Ich gebe [den Ball]$_{Acc}$ [einem Athleten]$_{Dat}$.
   I give [the ball] [an athlete].
   'I give the ball to an athlete.'

This study proposes two previously unconsidered factors to influence the object order, i.e. prediction of the constituents in a sentence based on the position of the full verb (FVP) and the clauses' information profile. Thus, we test these two hypotheses:

1) Acc>Dat is more likely when the lexical verb precedes the objects.
2) Dat>Acc is more likely when the clause's lexical information profile is uneven.

When the lexical verb containing the valency information follows the objects (FV-VL), Dat>Acc is preferable because recipients discard sentence continuations with a transitive verb earlier and rank those with a di- or intransitive verb higher (Levy, 2008). If the full verb is presented first (FV-V2), Acc>Dat is possible as the necessity of *both* objects is *known*. A certain object order is less crucial to reduce uncertainties.

The second hypothesis refers to the Uniform Information Density (UID) (Levy & Jaeger, 2007): In lexically uneven clauses, it is better to use the more common Dat>Acc as familiarity with a certain construction can facilitate processing even under disadvantageous conditions, i.e.an uneven information profile (e.g. Futrell et al., 2021).

We also want to test the stability of these assumptions over time. Thus, we conducted a corpus study using the Anselm (~16th century), RIDGES (16th/17th century), GerManC corpus (17th/18th century), the Tiger and TüBa-D/Z corpus for modern German. The objects were found automatically. We analyze 1733 clauses here, 76% of them are from modern data. Acc>Dat occurs in 8% of the modern data, 10% of the 17th and 18th century and to 15% in the 16th century. Each clause was annotated for the FVP, DORM (Cuskley et al., 2021), a measure for UID based on unigram-lemma-surprisal[1] of each word in each corpus, the object's length ratio and their givenness status as well as the publication century of each text. A general logistic regression analysis (glm, (R Core Team, 2023))[2] with backward model selection was conducted. We found (among others, Table 1) that the FV-V2 is connected to the Acc>Dat-order in sentences with an uneven information profile in historical but not in modern data where Dat>Acc is generally more frequent (Figure 1). This result is interpreted as an interplay between grammatical and lexical processing difficulties which is used to keep the general processing effort constant. Lexically harder to process sentences have the more common order and vice versa. As more variation was possible in the past, speakers were more sensitive to means of facilitating processing.

Given accusative objects are linked to the Acc>Dat-order, but the likelihood of Dat>Acc increases when the full verb is in the RSB even for given accusative objects

[1] Using lemma unigram surprisal neutralizes any grammatical information and is preferable for DORM.
[2] glm(Object order ~(DORM+Length Ratio+Acc givenness (sum-coded) + Dat givenness (sum-coded) +FVP (sum-coded) + period)^3, data=data, family = "binomial").

(Figure 2). Thus, we find further (and period stable) evidence for the influence of the position of the full verb in line with our prediction.

	Est.	Std. Error	z-value	p-value			Est.	Std. Error	z-value	p-value	
Intercept	1.26	0.25	4.95	<0.001	***	**DORM: Period**	0.07	0.04	1.73	0.08	.
DORM	-0.06	0.07	-0.86	0.39		**Length ratio: FVP**	0.18	0.09	2.03	<0.05	*
Length ratio	-0.83	0.04	-1.88	0.06	.	**$Acc_{Info\text{-}Status}$: FVP**	1.18	0.55	2.15	<0.05	*
$Acc_{Info\text{-}Status}$	-2.11	0.44	-4.78	<0.001	***	**$Acc_{Info\text{-}Status}$: Period**	-0.47	0.25	-1.87	0.06	.
$Dat_{Info\text{-}Status}$	0.08	0.41	0.19	0.84		**$Dat_{Info\text{-}Status}$: Period**	0.65	0.24	2.66	<0.01	**
FVP	0.96	0.49	1.98	0.047	*	**FVP: Period**	-0.63	0.29	-2.19	<0.05	*
Period	0.38	0.16	2.42	<0.05	*	**DORM: FVP**	0.35	0.13	2.67	<0.01	**
DORM: Length ratio	0.02	0.01	2.09	<0.05	*	**DORM: FVP: Period**	-0.16	0.08	-2.05	<0.05	*

Table 1 Results of the regression analysis.

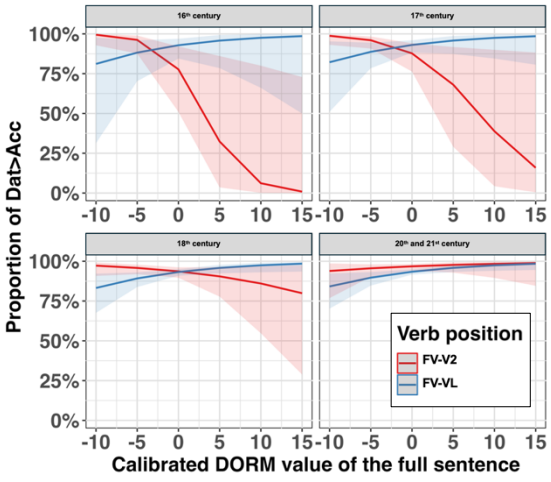

Figure 1 Interaction plot of the variables DORM, position of the full verb and period. A higher DORM indicates less uniformity.

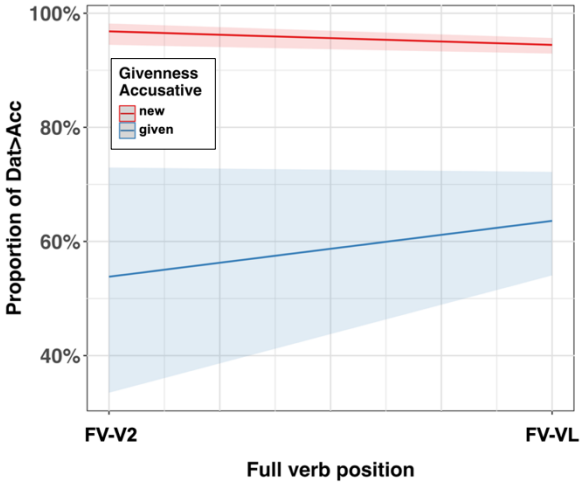

Figure 2 Interaction plot of the position of the full verb and the information status of the accusative object.

References

Cuskley, C., Bailes, R., & Wallenberg, J. (2021). Noise resistance in communication: Quantifying uniformity and optimality. *Cognition*, *214*, 104754. https://doi.org/10.1016/j.cognition.2021.104754

Futrell, R., Gibson, E., & Levy, R. (2021). Lossy-Context Surprisal: An Information-Theoretic Model of Memory Effects in Sentence Processing. *Cognitive Science*, *44*(3), e12814. https://doi.org/0.1111/cogs.12814

Lenerz, J. (1977). *Zur Abfolge nominaler Satzglieder im Deutschen*. Narr.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schölkopf, J. Platt, & T. Hofmann (Eds.), *Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference* (pp. 849–856). The MIT Press. https://doi.org/10.7551/mitpress/7503.003.0111

R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rauth, P. (2020). *Ditransitive Konstruktionen im Deutschen. Geschichte und Steuerung der Objektabfolge im Mittelfeld*. Stauffenburg.

Speyer, A. (2011). Die Freiheit der Mittelfeldabfolge im Deutschen – ein modernes Phänomen. *Beiträge Zur Geschichte Der Deutschen Sprache Und Literatur*, *133*, 14–31.

Speyer, A. (2015). Object order and the Thematic Hierarchy in Older German. In J. Gippert & R. Gehrke (Eds.), *Historical Corpora: Challenges and Perspectives* (pp. 101–124). Narr.

Speyer, A. (2016). The relative object order in High and Low German. In S. Featherston & Y. Vearsley (Eds.), *Quantitative Approaches to Grammar and Grammatical Change. Perspectives from Germanic.* (Vol. 290, pp. 143–164). De Gruyter.