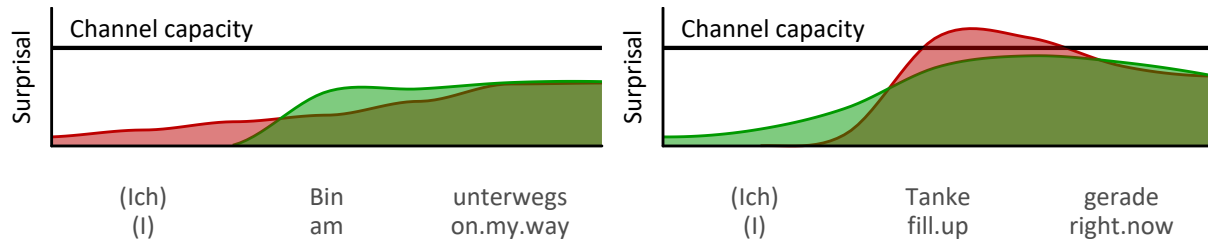


## UID constrains the usage of topic drop in German: experimental and corpus linguistic findings

Lisa Schäfer, Robin Lemke, Ingo Reich and Heiner Drenhaus (Saarland University)

lisa.schaefer@uni-saarland.de

**Background.** German allows for *topic drop* (TD) (fig. 1 & 2), the omission of a preverbal constituent from a declarative V2 sentence (Ross, 1982). We investigate the question of when TD is *used*, i.e. speaker and hearer preferences beyond grammatical properties that license TD (Fries, 1988) with a corpus study and two experiments. We provide a systematic empirical investigation of claims from the theoretical literature and combine previously isolated observations into a unifying account of the usage of TD.



**Fig. 1:** Hypothetical ID profile: *ich* ('I') creates **surprisal minimum**  $\Rightarrow$  **topic drop** more uniform  
**Fig. 2:** Hypothetical ID profile: *tanke* ('fill up') creates **surprisal maximum**  $\Rightarrow$  **full form** more uniform

**Account.** We propose an information-theoretic explanation for the usage of topic drop based on the Uniform Information Density (UID) hypothesis (Levy & Jaeger, 2007). According to UID speakers distribute surprisal ( $-\log_2 p(\text{word}|\text{context})$ ) (Shannon, 1948) uniformly across the utterance avoiding minima and maxima in the information density (ID) profile. This makes the most efficient use of the processing capacities of the hearer (surprisal  $\propto$  processing effort (Hale, 2001)). The UID account predicts that TD is more felicitous when the preverbal constituent is predictable, i.e. carries few surprisal such as *ich* in fig. 1 (*avoid surprisal minima*), and when no unpredictable verb, like *tanke* in fig. 2, is left in the sentence-initial position (*avoid surprisal maxima*). We investigate 4 factors that could constrain the usage of TD: 1) TOPICALITY (Helmer, 2017): A preverbal constituent is more predictable when it is the topic, thus omitting a topic avoids a surprisal minimum. 2) PERSON (Auer, 1993; Imo, 2014): Since the 1SG is more frequent and more salient, omitting a 1SG pronoun also avoids surprisal minima. 3) INFLECTION (Auer, 1993): An omitted constituent is more easily recoverable, i.e. TD causes less processing effort, when it occurs before a verb with distinct INFLECTION (*ich tanke* / *sie tankt* vs. *ich kann* $\emptyset$  ('can') / *sie kann* $\emptyset$ ). 4) SURPRISAL: We predict a lower ratio of TD before unpredictable verbs with high SURPRISAL because inserting a constituent before an unpredictable word smooths the ID profile and reduces processing effort.

**Corpus study.** We performed logistic regressions on 280 TDs and 162 full forms from the text messages subcorpus of FraC (Horch & Reich, 2017) to predict TD from PERSON (1SG vs. 3SG), INFLECTION (marked vs. syncretic) and unigram SURPRISAL per verb lemma. TD with 1SG is more frequent than with 3SG ( $\chi^2 = 27.63$ ,  $p < .001$ ). 1SG is more predictable as it occurs more often in the prefield in our corpus ( $n_{1SG} = 343$  vs.  $n_{3SG} = 99$ ), so omitting it avoids a surprisal minimum. The preverbal element is more likely to be realized when the verb SURPRISAL is higher ( $\chi^2 = 14.21$ ,  $p < .001$ ), because inserting a constituent before the unpredictable verb smooths the surprisal maximum on the verb.

INFLECTION does not generally favor TD but in interaction with higher SURPRISAL ( $\chi^2 = 4.86$ ,  $p < .05$ ). Recovering the omitted preverbal element causes additional processing effort to that required for processing the verb. A distinct inflection provides a cue towards the omitted expression and thus reduces the overall processing effort on the verb.

**Experiments.** As the corpus is not annotated for TOPICALITY, we investigate this factor experimentally by varying the subject of the preceding sentence, since the subject is the unmarked topic (Lambrecht, 1994). We also look at PERSON and INFLECTION which we varied between **full verbs** with inflectional marking in exp. 1 and **syncretic modals** in exp. 2. In our 2 within-subjects acceptability rating studies crossing the factors OMISSION (realized vs. omitted), PERSON (1SG vs. 3SG) and TOPICALITY (topic continuity vs. topic shift) we used 24 items like (1), 60 fillers and a 7-point Likert scale (7 = completely natural).

- (1) A: 'What's new?'
- a. B: Am Samstag geht Julia mit mir schick essen. [(Sie) **lädt** mich diesmal ein. | (Sie) **möchte** mich diesmal einladen.]  
 'B: On Saturday **Julia** dines out well with me. [(**She**) invites me this time. | (**She**) wants to invite me this time.]' [topic continuity | 3SG | omitted (realized)]
- b. 'B: On Saturday **I** dine out well with Julia. [(**She**) invites me this time. | (**She**) wants to invite me this time.]' [topic shift | 3SG | omitted (realized)]
- c. 'B: On Saturday **Julia** dines out well with me. [(**I**) invite her this time. | (**I**) want to invite her this time.]' [topic shift | 1SG | omitted (realized)]
- d. 'B: 'On Saturday **I** dine out well with Julia. [(**I**) invite her this time. | (**I**) want to invite her this time.]' [topic continuity | 1SG | omitted (realized)]

**Exp. 1** (n = 43). Utterances with TD were rated significantly better when the omitted element was 1SG ( $\chi^2 = 20.74$ ,  $p < .001$ ) and in case of topic continuity ( $\chi^2 = 7.97$ ,  $p < .01$ ). The result for PERSON is in line with the corpus study: 1SG TD is more frequent and more acceptable. The effect of TOPICALITY indicates that TD is more felicitous when the omitted constituent is a topic as it is more predictable and thus more likely to be omitted.

**Exp. 2** (n = 48). TD was again rated significantly better for 1SG than for 3SG ( $\chi^2 = 18.72$ ,  $p < .001$ ) but unlike in exp. 1 not in case of topic continuity ( $\chi^2 = 1.82$ ,  $p = 0.18$ ). The missing effect of topicality suggests that in exp. 1 inflection and topicality in combination help to recover the omitted constituent which is evidenced by the significant interaction between OMISSION and TOPICALITY. In exp. 2, we do not find such a significant interaction, so topic continuity alone seems to be too weak to facilitate recoverability without inflection.

**Discussion.** Taken together, the data support our information-theoretic account: Speakers prefer TD when this distributes processing effort more uniformly across the utterance. The corpus study shows that 1SG TD is more frequent than 3SG TD and that TD occurs less often before unpredictable verbs. The experiments confirm the result for grammatical person and show that topic continuity and inflectional marking favor TD in interaction. In sum, we provide a unifying account to the usage of TD that allows us to combine previously isolated findings and to firstly account for an effect of verb surprisal.

**Selected references.** Auer, P. (1993). Zur Verbspitzenstellung im gesprochenen Deutsch. *Deutsche Sprache*, 23, 193–222. • Fries, N. (1988). Über das Null-Topik im Deutschen. *Sprache & Pragmatik*, 3, 19–49. • Horch, E. & Reich, I. (2017). The Fragment Corpus (FraC). *CL2017*. Birmingham (UK), 392–393. • Lambrecht, K. (1994). *Information structure and sentence form: topic, focus and the mental representations of discourse referents*. Cambridge: Cambridge University Press. • Ross, J. R. (1982). Pronoun Deleting Processes in German. *LSA*, San Diego (CA), 43.