Investigating fragment usage with a gamified utterance selection task

Why do we use fragments? Fragments like (1a) (Morgan, 1973) can often be used to perform the same speech act as the corresponding sentence (1b).

(1) [Passenger to conductor before entering the train:]
   a. To Paris?
   b. Does this train go to Paris?

The syntax of fragments is relatively well researched, but the question of why and when speakers use fragments is not. Some syntactic accounts propose licensing conditions on fragments (e.g. Merchant, 2004; Barton and Progovac, 2005) based on information structure or recoverability, but fragments are not always used when they are licensed, as the acceptability of (1b) in this context shows. Intuitively, the advantage of fragments is that they allow the speaker to get a message across with less production effort. However, fragments can be enriched in different ways (see e.g. (2) for (1a)) and thus increase the risk of being misunderstood.

(2) a. How long does it take to travel to Paris?
   b. Have you ever been to Paris?

The choice between a fragment and a sentence probably consists in a trade-off between a gain in efficiency and the risk of communication failure. In what follows, I present a game-theoretic formalization of this reasoning and an pseudo-interactive experiment testing its predictions.

A game-theoretic account of fragment usage The model I propose is based on Franke's (2009) account of implicature: There is (i) a set of messages $m \in M$ that a speaker can to communicate and (ii) a set of utterances $u \in U$ which can used for this purpose. The speaker selects the utterance which is most optimal; the hearer receives it and figures out which message the speaker had in mind. The hearer computes $p(m|u)$ based on the prior likelihood of m and a denotation function $[[\cdot]]$, which returns 1 if u can be derived by grammatically licensed omission from m and 0 otherwise (see equation 1). The speaker in turn tries to maximize $L_0(u, m_i)$ for their intended m_i while keeping the production cost for u as low as possible.

$$L_0(u, m) = \frac{Pr(m) \times [[u]]_m}{\sum_{m'} Pr(m') \times [[u]]_{m'}} \quad (1)$$

Empirically founded model parameters In order to compute L_0 posterior probabilities with equation 1, I estimated M, $Pr(M)$, U and $[[u]]_m$ for all $m \in M$, $u \in U$ from a data set collected by Lemke (2021) with a production study. The data set contains about 100 utterances for each of 24 context stories (4) based on the DeScript corpus of script knowledge (Wanzare et al., 2016). The utterances were transformed into simplified representations like (3a) (pooling synonyms and to excluding ungrammatical omissions of function words, see Lemke (2021) for details), each of these representing a message like (3a). Its relative frequency is used as $Pr(m)$ in the model. Since all of the "words" in representations like (3a) can be freely omitted, this yields the set of utterances in (3b), for which $[[u]]_{(3a)}$ = 1.

(3) "Pour the pasta into the pot"
   a. `pour pasta pot.GOAL¡`

Experiment design Since the game-theoretic account is inherently interactive, I test its predictions with an interactive utterance selection design (similar to Rohde et al. (2012) for referring expressions). The production cost for utterances is implemented by an explicit cost term. Currently, the participant plays the speaker role and the listener role is simulated by the computer, who – in a initial step – behaves maximally rationally, i.e. as predicted by the

model. In each trial ($n = 15$), the participant is presented a context story and an message to communicate with one out of six utterances (see fig. 1, showing the German implementation). Their task consists in selecting one of the utterances to communicate the message. In order to model utterance cost, subjects are assigned an account of virtual coins they can spend for sending utterances (starting with 500 coins): Sentences (cost: 100) are more expensive than fragments (cost: 30) and successful communication is rewarded with 120 coins. In the experiment, there are three conditions (i) the "target utterance" (most likely given the fragment) is highlighted, (ii) the competitor (less likely, but possible), (iii) the distractor is highlighted. According to model predictions, subjects should use fragments more often in the target than in the competitor condition, and most often in the unambiguous distractor condition,

Preliminary results and dicussion Data collection is ongoing, but the results of the first list of a pilot study indicate that – as expected – the rate of fragment choice is highest in the unambiguous distractor condition (46%). Furthermore, fragments are used more often (20%) when they refer to a predictable message than when they refer to an unpredictable one (14%). The analysis of the further data currently being collected will show whether this pattern is consistent. If it were, it would provide empirical support for a rational and game-theoretic account of fragment usage. Interestingly, the data collected so far also indicate a strong overall bias for using sentences, even in the unambiguous distractor condition, which will be also subject to further research.

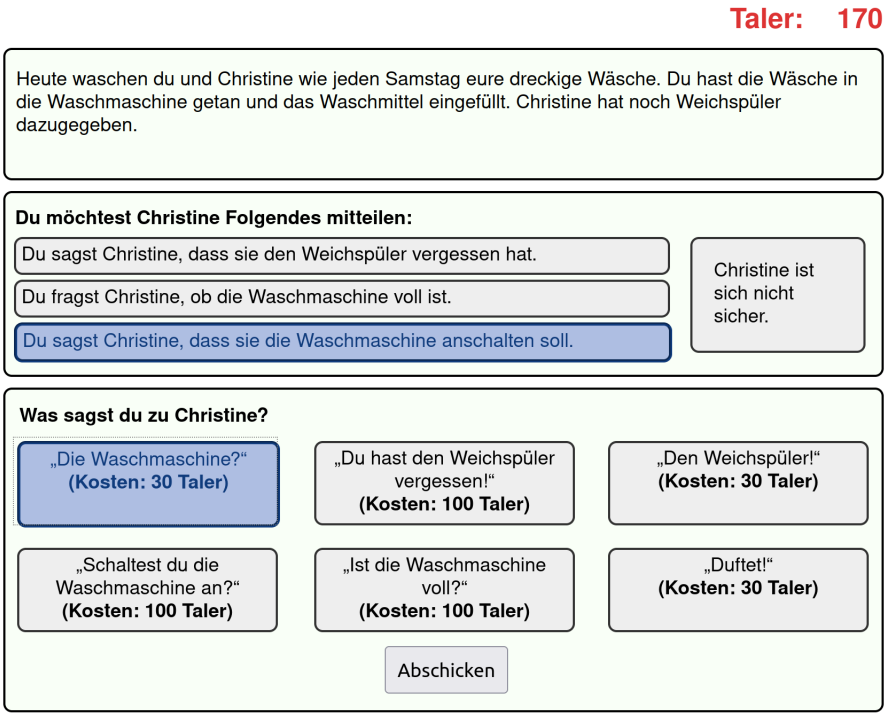

Figure 1 Sample utterance selection display (German) showing the context story, three messages and six utterances.

References •Barton, E. and Progovac, L. (2005). Nonsententials in Minimalism. In Elugardo, R. and Stainton, R. J., editors, Ellipsis and Nonsentential Speech, 71–93. Springer, Dordrecht. •Franke, M. (2009). Signal to Act: Game Theory in Pragmatics. PhD thesis, Universiteit van Amsterdam. •Lemke, R. (2021). Experimental Investigations on the Syntax and Usage of Fragments. Number 1 in Open Germanic Linguistics. Language Science Press, Berlin. • Merchant, J. (2004). Fragments and ellipsis. Linguistics and Philosophy, 27(6):661–738. •Morgan, J. (1973). Sentence fragments and the notion 'sentence'. In Kachru, B. B., Lees, R., Malkiel, Y., Pietrangeli, A., and Saporta, S., editors, Issues in Linguistics. Papers in Honor of Henry and Renée Kahane, 719–751. University of Illionois Press, Urbana. • Rohde, H., Seyfarth, S., Clark, B., Jaeger, G., and Kaufmann, S. (2012). Communicating with cost-based implicature: A game-theoretic approach to ambiguity. In Proceedings of the 16th Workshop on the Semantics and Pragmatics of Dialogue, 107–116. •Wanzare, L. D. A., Zarcone, A., Thater, S., and Pinkal, M. (2016). DeScript: A crowdsourced corpus for the acqui¸sition of high-quality script knowledge. In Proceedings of LREC 2016, 3494–3501, Portoroz, Slovenia.¸