

Testing a Rational Account of Fragment Usage with Crowd-sourced Production Data

Robin Lemke (Saarland University) robin.lemke@uni-saarland.de

Game-theoretic models have been applied to a range of pragmatic phenomena (Franke, 2009; Frank and Goodman, 2012), but their predictions have been tested at restricted and balanced sets of meanings. At the example of ellipsis, I test a rational account with a much more diverse and unbalanced data set collected with a crowd-sourced production task. I focus on fragments (1a) (Morgan, 1973), nonsentential utterances which are meaning-equivalent to sentences (1b) in an appropriate context. Previous research focused on the syntax of fragments, but why speakers actually use them is underexplored.

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

Account I hypothesize that speakers trade off the lower production cost for fragments (compared to sentences) with the risk of being misunderstood ((1a) could also mean *How long does it take to travel to Paris?*) and prefer fragments when the former outweighs the latter. To formally model this idea, following Franke (2009), I assume that the speaker sends a message $m \in M$ to the listener and selects the an utterance $u \in U$ to do so. The listener infers the meaning of u and if speaker and listener coordinate, both receive a reward. Therefore, the listener goes for the most likely interpretation (maximize $p(m|u)$), calculated as shown in equation 1. Sentences are unambiguous, but their cost is higher, so the speaker will prefer fragments when $p(m|u)$ is relatively high.

Method I evaluate the model with 3 pseudo-interactive utterance selection experiments. In each study, 60 subjects read a context story ($n = 15$) and select one out of 6 utterances to communicate one out of 3 messages (Fig. 1). The materials are based on a corpus of production data by Lemke (2021), from which M , U and the prior over messages $Pr(M)$ were estimated. The listener is simulated according to model predictions. In each trial, there is a fragment ambiguous between two messages: the *target* having a higher $p(m|u)$ than the *competitor*. There are 3 experimental conditions, which differ in whether the target, the competitor, or the third message (which is *unambiguously* encoded by the second fragment) is to be communicated. Utterances cost virtual coins and sentences are more expensive than fragments. Given the increasing $p(m|u)$ across the conditions (see Table 1), subjects should use fragments most frequently in the *unambiguous* condition than in the *target* and least often in the *competitor* condition.

Experiments and results Fig. 2 summarizes the data, which were analyzed with mixed effects logistic regressions (Bates et al., 2015). In exp. 1, $p(m|u)$ increased fragment ratio ($\chi^2 = 6.13, p < .05$), but some subjects produced only sentences, which yielded a net benefit given the cost structure. Therefore, in exp. 2 sentences were more costly, which increased fragment ratio further ($\chi^2 = 6.24, p < 0.05$). However, in exp. 1 and 2 there was no significant effect of $p(m|u)$ in the ambiguous conditions, so the effects found could be either evidence rational reasoning given the higher $p(m|u)$ in the unambiguous condition or that subjects just avoid ambiguity. Exp. 3 tested this by replacing the ambiguous *competitor* condition by a further unambiguous one. This increases fragment ratio ($\chi^2 = 17.52, p < 0.001$), but the effect of $p(m|u)$ was replicated, too. Taken together, this supports the expected cost-accuracy tradeoff and shows that game-theoretic reasoning can also be applied to an unbalanced and diverse data set based on production data.

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

| Condition | Lowest $p(m u)$ | Highest $p(m u)$ | Mean $p(m u)$ |
|-------------|-----------------|------------------|---------------|
| critical | 0.12 | 0.69 | 0.36 |
| competitor | 0.03 | 0.17 | 0.08 |
| unambiguous | 0.15 | 1.0 | 0.76 |

Table 1 Range of $L_0(m|u)$ probabilities and means by conditions

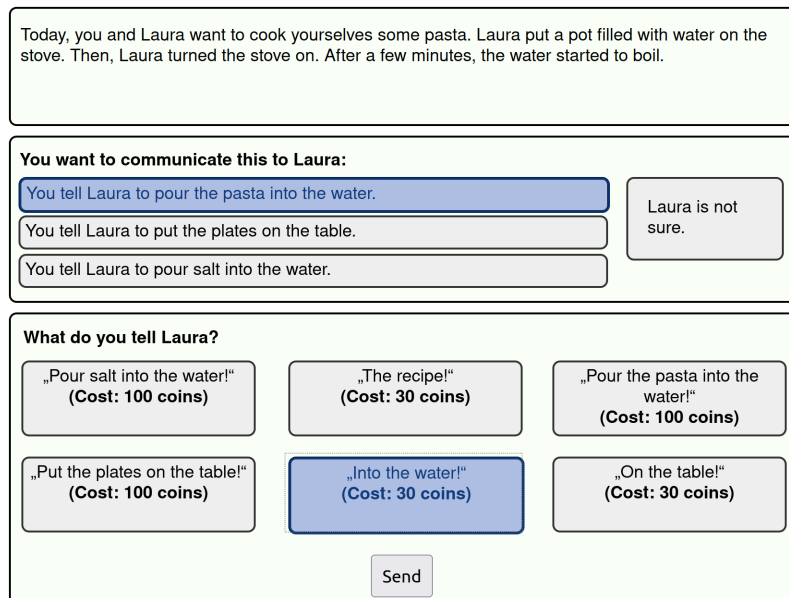


Figure 1 Screenshot of the experiment, translated to English for convenience.

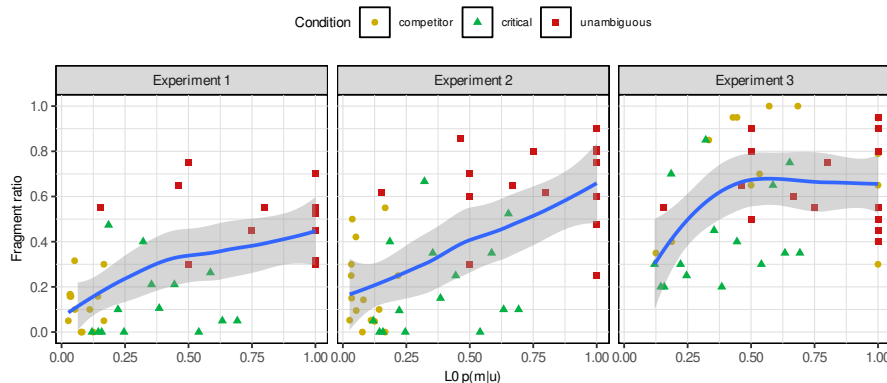


Figure 2 Ratio of fragments and sentences across the experiments and conditions.

Selected references •Lemke, R. (2021). Experimental Investigations on the Syntax and Usage of Fragments. Language Science Press. •Morgan, J. (1973). Sentence fragments and the notion 'sentence'. In Kachru, B. et al., eds, Issues in Linguistics. 719–751. University of Illionois Press.