Communicative efficiency and grammatical encoding in speech: Predicting case-marker omission in Japanese

Chigusa Kurumada (Stanford University) & T. Florian Jaeger (University of Rochester)

kurumada@stanford.edu

Optional case-marker use; Sentence recall; Speech production; Japanese

Recent work suggests that language production reflects speakers' bias to achieve efficient information transmission [1]. Speakers tend to provide more linguistic signal for information that is difficult to recover while reducing contextually inferable elements [e.g. 1,2]. However, much of this evidence has been claimed to be compatible with availability-based accounts [3], which attribute additional signal to upcoming retrieval difficulty. To discern these two theories, we assess the predictions of communicative efficiency-based accounts against a psycholinguistically understudied phenomenon -- *optional case-marking*.

Cross-linguistically, grammatical systems are sensitive to the plausibility of grammatical function (GF) assignment. In some languages, only inanimate subjects and animate objects require case-marking, whereas the more plausible animacy-to-GF assignments (i.e., animate subjects and inanimate objects) remain unmarked [4]. Case-markers hence categorically signal implausible (*less expected*) GF-assignments. In languages like Korean and Japanese, on the other hand, object case-marking has been claimed to be optional. The current study provides the first evidence that, as predicted by the communicative efficiency-based accounts, Japanese speakers tend to omit the case-marker when GF-assignment conforms to an expected pattern.

Experiment 1 (20 participants, 24 items, 48 fillers) tests whether native speakers of Japanese are less likely to mark inanimate objects, compared to animate objects, as would also be predicted based on the fact that inanimate NPs tend to be grammatical objects [5]. Spoken recall experiment (cf. [3]): Japanese speakers heard and reproduced sentences with human subjects and either a human or an inanimate object (1) (50/50 presented with or without the case-marker; subjects were always case-marked). *Speakers were less likely to produce object-markers for inanimate (plausible) object referents* ($p<.05$; no interactions; mixed logit regression, full factorial; maximum random effects, cf. [6]; same for Exp2). Similar effects have previously been observed for animacy and definiteness in Korean [7] but not in Japanese [5].

Experiment 2 (32 participants, 24 items, 48 fillers) puts communicative efficiency-based accounts to a stronger test: if animacy effects are indeed due to GF-assignment plausibility, we should see plausibility effects even when animacy, definiteness, and similar factors are held constant. Norming judgments were obtained (80 participants, non-overlapping with Exp1/2, on 100-point scales) to estimate GF-assignment plausibility for sentence pairs (2). The recall experiment manipulated whether the NP that was rated to be the more plausible grammatical object was indeed the grammatical object or not. *As predicted, speakers were less likely to produce the case-marker for object NPs which were normed to be plausible objects* ($p <.05$ for factorial effect of plausibility (high vs. low); $p <.03$ for gradient effect of plausibility norms).

These results suggest that Japanese speakers tend to case-mark GF-assignment when it is less inferable (plausible) based on the properties of the argument as well as co-occurring elements in a sentence. While it might be possible to recast this result in terms of accessibility, it is not *predicted* by any existing availability-based account. Beyond the relevance for psycholinguistics, these results suggest that speakers' bias toward encoding unpredictable information with more linguistic signal might drive cross-linguistically attested grammatical patterns [4,7].

Experiment 1: example item (English translation); [ANI= Animate, INA = Inanimate]
(1) a. A teacher saw <u>a policeman</u> [ANI] on a street. b. A teacher saw <u>a fire-engine</u> [INA] on a street

Experiment 2: norming study items (English translation): Each of the 24 Animate-Animate noun pairs was combined with 2 kinds of verbs. Participants distributed 100 points to the two sentences according to the plausibility of the object GF assignment.
(2) a. A doctor treated a patient in a hospital room. vs. A patient treated a doctor in a hospital room.
b. A doctor waited for a patient in a hospital room. vs. A patient waited for a doctor in a hospital room.

[1] Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. [2] Genzel, D. & Charniack, E. (2002). Entropy rate constancy in text. [3] Ferreira, V. & Dell, G. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. [4] Aissen, J. (2003). Differential object marking: Iconicity vs. economy. [5] Fry, J. (2003). *Ellipsis and topic-marking in Japanese*. [6] Jaeger, T.F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. [7] Lee, H-J. (2006). Parallel optimization in case systems: Evidence from case ellipsis in Korean.