The processing of grammatical functions in Swedish is expectation-based

Thomas Hörberg, Department of Linguistics, Stockholm University

Background. In order to facilitate information transfer during communication, natural language processing is assumed to be expectation-based, drawing on statistical regularities in the input (e.g., Jaeger 2013; MacDonald 2013; Venhuizen et al. 2019). During comprehension, linguistic and extra-linguistic information in the previous discourse set up expectations that facilitate interpretation. During production, information in the discourse is used to balance the upcoming utterance in a way that limits production costs, and that ensures the message is informative enough. Central to communication is the processing of grammatical functions (GFs) of subject and direct object. GFs express how participants are related to events or states (e.g., *who is doing what to whom*), and involve information structural properties (e.g. topichood, Foley 2011). In many languages, there are many ways to encode GFs morphosyntactically (e.g. word order and case). Speakers' GF encoding preferences depend on an interplay between NP properties (e.g, animacy) and verb semantic properties (e.g., volitionality) (Hörberg 2016). This creates complex statistical patterns in the distribution of these *GF information* types that can be utilized during on-line GF processing. This poster presents 1) findings showing that GF processing in transitive sentences in written Swedish is influenced by sentence-level expectations as based on such statistical patterns, and 2) an on-going project, extending on these findings, that investigates how GF processing is affected by discourse-level expectations.

GFs in Swedish. In Swedish, the preferred word order is SVO. However, when the direct object is topical or contrastive (Hörberg 2018), OVS word order is also used. Since lexical NPs lack case marking, OVS sentences with a lexical initial NP as 1) *Läraren gillade vi inte* ('the teacher we didn't like') are potentially ambiguous with respect to GFs, and can be costly to comprehend. Using event-related brain potentials, Hörberg et al. (2013) found pronominal subject NPs in OVS sentences (e.g., *vi* in 1) to engender a "reanalysis N400" effect (a right-parietal negativity in the 375-550 ms time window), reflecting a revision of a tentative subject-first GF assignment upon encountering the disambiguating post-verbal subject NP.

Influences of sentence-level expectations in GF processing. This poster presents results indicating that language users are sensitive to statistical regularities in the distribution of GF information, and can utilize them to overcome potential GF ambiguity problems during GF encoding and decoding in written language processing:

Writers, on the one hand, tend to balance their use of OVS word order with respect to the availability of GF information in a manner that accommodates comprehension. A corpus-study of 14500 transitive sentences of written Swedish showed that OVS sentences are less frequently used when morphosyntactic or animacy information about GFs are unavailable. In such cases, writers more frequently use an unambiguous passive construction, thereby avoiding ambiguity.

Readers, on the other hand, utilize statistical regularities in the distribution of GF information during on-line GF assignment. Such distributions in written corpora was used to train an expectation-based model of incremental GF assignment in transitive sentences. Using regression-based estimates of the *objective* probability of an object-initial GF assignment at each constituent (NP1, verb, and NP2), the model quantifies the change in the expectation of a GF assignment in terms of *Bayesian surprise* (i.e., relative entropy over GF assignments before and after seeing the constituent, cf. Kuperberg & Jaeger, 2016). A self-paced reading experiment where 45 participants read 64 transitive sentences varying with respect to word order (SVO vs. OVS), NP1 animacy (animate vs. inanimate) and verb class (volitional vs. experiencer) confirmed the most prominent model predictions. By-region reading times on the verb and NP2 converged with region-specific Bayesian surprise predicted by the model. For example, NP2 reading times in ambiguous OVS sentences were mitigated when NP1 animacy and its interaction with verb class bias towards an object-initial word order.

Influences of discourse-level expectations on GF processing. Although these findings indicate that sentence-level GF processing is sensitive to statistical regularities in the distribution of GF information, it does not address whether such regularities causally influence GF processing in spoken discourse contexts. Since sentence-level ambiguities often can be resolved in context, the utility of GF information during communication might be limited in discourse contexts (Rahkonen 2006). Indeed, sentence-level ambiguities have even been suggested to be beneficial for communication when they can be resolved in context (Piantadosi et al. 2012). The poster also presents an on-going research

project addressing this question. This project investigates how a specific type of GF information - animacy - causally affects GF processing in transitive sentences in short story discourse contexts, and whether an appropriate discourse context can facilitate on-line GF comprehension. It consists of three experimental studies with the same stimulus material; spoken short stories accompanied with cartooned images, making up scenes that are presented sequentially. Crucially, the stories set up discourse contexts which licence OVS word order; two critical scenes are accompanied with OVS sentences in which the direct object is contrastive (Hörberg 2018).

Study 1 is a picture-description task (similar to Prat-Sala & Branigan 2000), investigating whether speakers' propensity for using OVS word order is causally influenced by animacy in spoken discourse contexts, i.e., whether speakers balance their productions in a way that accomodates comprehension. Participants are exposed to the OVS sentence in the first critical scene (functioning as a prime sentence), but are asked to describe the second critical scene themselves. Crucially, the direct object of the target sentence is either animate or inanimate. If speakers are sensitive to the availibility of the animacy information and balance their use of OVS word order accordingly - in order to avoid a potential ambiguity - OVS word order descriptions should be more frequent when the direct object is inanimate.

Study 2 and 3 are comprehension studies, investigating whether a discourse context, animacy information, and/or their interaction causally affect listeners' processing and anticipations regarding GF assignments during OVS sentence comprehension. Participants are exposed to the critical sentences, either with an animate or inanimate direct object, and either with or without the discourse context. Study 1 investigates the ERP response to post-verbal subject NPs in the critical OVS sentences, and - crucially - whether the previously observed reanalysis N400 (Hörberg et al. 2013) is reduced or absent when animacy and/or the discourse context provides support for an OVS interpretation. Using the visual-world eye-trackning paradigm (e.g. Altmann & Kamide 2009), study 3 investigates whether comprehenders can predict (as indexed by the timing of eye fixations) unexpected GF assignments when animacy and/or the discourse context is provided.

The project will provide novel insights about the processing of GFs in spoken discourse contexts and to what extent animacy, discourse information, and their interaction guide these processes.

Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition, 111*(1), 55–71.

Foley, W. A. (2011). A typology of information packaging in the clause. In T. Shopen (Ed.), *Language Typology and Syntactic Description*: *Vol. Volume 1:Clause Structure* (pp. 362–446).

Hörberg, T. (2016). *Probabilistic and Prominence-driven Incremental Argument Interpretation in Swedish* (PhD thesis, Stockholm University).

Hörberg, T. (2018). Functional motivations behind direct object fronting in written Swedish: A corpus-distributional account. *Glossa: A Journal of General Linguistics*, *3*(1), 81.

Hörberg, T., Koptjevskaja-Tamm, M., & Kallioinen, P. (2013). The neurophysiological correlate to grammatical function reanalysis in Swedish. *Language and Cognitive Processes*, *28*(3), 388–416.

Jaeger, T. F. (2013). Production preferences cannot be understood without reference to communication. *Frontiers in Psychology, 4*.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology, 4*.

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, *122*(3), 280–291.

Prat-Sala, M., & Branigan, H. P. (2000). Discourse Constraints on Syntactic Processing in Language Production: A Cross-Linguistic Study in English and Spanish,. *Journal of Memory and Language*, *42*(2), 168–182.

Rahkonen, M. (2006). Some aspects of topicalization in Swedish declaratives. *Linguistics*, *44*(1), 23–55.

Venhuizen, N. J., Crocker, M. W., & Brouwer, H. (2019). Expectation-based Comprehension: Modeling the Interaction of World Knowledge and Linguistic Experience. *Discourse Processes*, *56*(3), 229–255.