Language comprehension is expectation-based (e.g. Venhuizen et al. 2019). Statistical regularities in the linguistic input set up expectations that are utilized during incremental interpretation. A central part of language comprehension involves assigning grammatical functions (GFs) to NPs, thereby determining how participants are related to events or states. In many languages, speakers have many ways to encode GFs morphosyntactically (e.g. word order, case), and their 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. In this talk, I present evidence indicating that GF assignment in transitive sentences in written Swedish is expectation-based, drawing upon such statistical patterns. I will present a corpus-based probabilistic model of incremental GF assignment in Swedish transitive sentences, together with results from a self-paced reading experiment, showing that the model’s strongest predictions are confirmed by human processing preferences.

The model is based upon 16552 transitive sentences, extracted from a corpus of written Swedish, that were annotated for word order (SVO vs. OVS), GF information (e.g., animacy, definiteness, case), and verb semantic properties (e.g. volitionality, sentence). Based on the distribution of these features, estimates of the probability for SVO vs. OVS GF assignment at each sentence region (NP1, verb, NP2) were calculated, using logistic mixed effects regression modeling. In the model, these estimates are used to predict incremental processing costs related to the change in the expectation for a GF assignment at each sentence region. This is done in terms of Bayesian surprise - the relative entropy over the two possible GF assignments before and after seeing the constituent at hand (Kuperberg & Jaeger 2016). Bayesian surprise (over syntactic trees) has also been argued to underlie the correlation between word surprisal and both processing times (Smith & Levy 2013) and certain neural responses (e.g., the N400 effect, Frank et al. 2015).

In the self-paced reading experiment, 45 participants read 64 transitive sentences that varied with respect to word order (SVO vs. OVS), NP1 animacy (animate vs. inanimate) and verb class (volitional vs. experiencer). By-region reading times on NP1, the verb, and NP2 were well-described by the region-specific Bayesian surprise predicted by the model. For example, reading times in the NP2 region observed in locally ambiguous, object-initial sentences were mitigated when the animacy of NP1 and its interaction with the verb class bias towards an object-initial word order.

These findings indicate that on-line GF assignment draws upon statistical regularities in the previous language input, as predicted by expectation-based accounts.