Embedded V2 is Anti-licensed by Discourse Familiarity

A current question in the syntactic and semantic literature concerns the conditions licensing embedded Main Clause Phenomena [**MCP**]—syntactic structures typically confined to matrix clauses (Hooper and Thompson 1973; et seq). We present new data from a quantitative corpus study looking at the distribution of Swedish embedded V2 [**EV2**]—a classic MCP (Andersson 1975; et seq), and demonstrate that EV2 is anti-licensed by **discourse familiarity**.

Background: Whereas all declarative main clauses in Mainland Scandinavian are obligatorily V2 (1a), subject-initial embedded clauses are 'optionally' V2 (1b).

(1a) Han (gillar) inte (*gillar) kor.
he likes not (likes) cows.
'He doesn't like cows.'

(1b) Han sa att han (gillar) inte (gillar) kor.
he said that he (likes) not (likes) cows.
'He said that he doesn't like cows.'

Regarding the licensing of EV2 (and MCP more broadly) it has been proposed that EV2 is licensed (but not obligatory) in the complements of certain classes of attitude predicates (exemplified by *say, believe, deny, regret*, and *know*), distinguished in terms of their LEXICAL SEMANTICS: Hooper & Thompson's classes A-E, shown in 1.

	say (class A)	*believe* (B)	*know* (E)	*resent* (D)	*doubt, deny* (C)
Factive	✗	✗	✓	✓	✗
Speech Act	✓	✗	✗	✗	✓/✗
Doxastic State	✗	✓	✓	✓	✓/✗
Emotive	✗	✗	✗	✓	✗
Discourse Familiarly	✗	✗	✗	✓	✓
EV2	**Yes**	**Yes**	**Yes**	**No**	**No**

Table 1: Verb classes identified to differ with respect to EV2 (and other Main Clause Phenomena).

Such accounts differ however, with respect to whether the FACTIVE predicates (e.g. *know, regret*) are predicted to license EV2. For Wiklund et al. (2009), *know* patterns with *say, believe*, in allowing EV2-complements, unlike *deny* and *regret*. For Haegeman and Ürögdi (2010); Haegeman (2014), and Kastner (2015) however, *all* factives select a clause headed by a D-layer, which blocks the derivation of MCP. Hence, all factives are predicted to disallow EV2. Here, we build on this work, in arguing that EV2 is indeed blocked by discourse familiarity. However, we show empirically that factivity does *not entail* discourse familiarity.

Current Study: We extract a broad range of statistical information (code available and open source) from Swedish corpora (Borin et al., 2012) representing a diverse set of genres. From this we address claims from previous literature on factors governing the realization of EV2. Contra Haegeman and Ürögdi (2010); Haegeman (2014), and Kastner (2015) we do not consistently find a significant difference in rates of EV2 by factivity (see 3). Hooper & Thompson's classes of lexical predicates however, do have a real effect on the realization of EV2, although this contribution varies widely across corpora representing different genres (see 2). In looking at the rates of EV2 by predicates within a unified semantic class, we find high degrees of variance. This is counter to a semantic selection account based purely on lexical class membership (see 3, middle-plot).

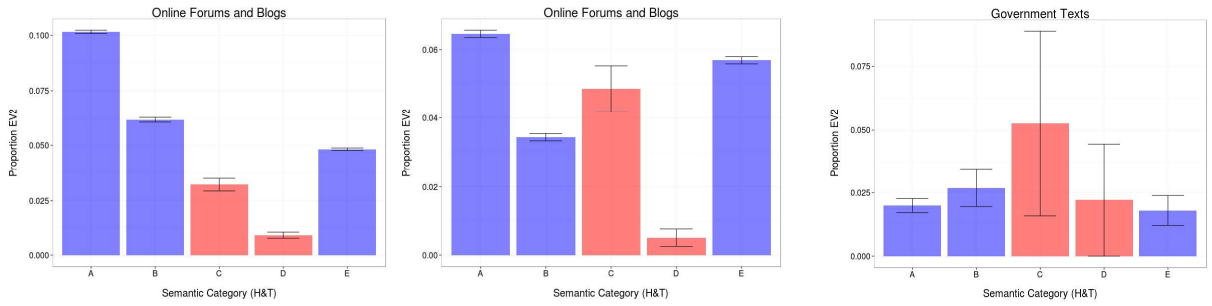

Figure 2: Distribution of EV2 across genre and predicate type.

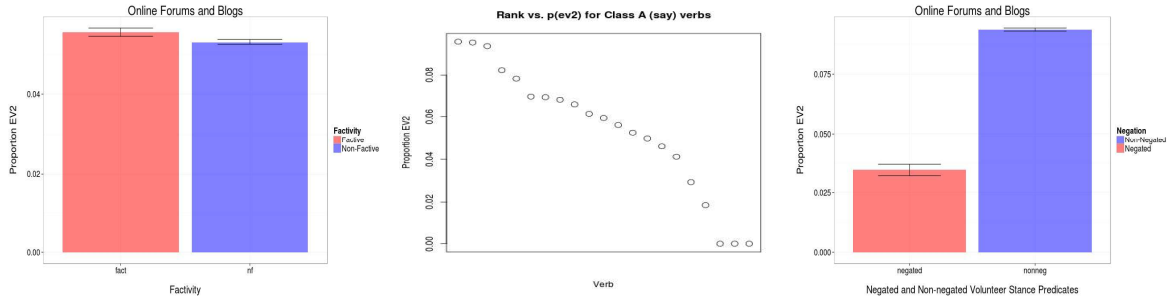

Figure 3: **Left:** Plot of probability of ev2 from factive vs. non-factive main predicates. **Middle:** Rates of EV2 for verbs within the 'say' class **Right:** Rates of EV2 for volunteer stance predicates under negation (or not)

Current proposal: Looking at the semantic properties listed in 1, we observe that the one property that differentiates the classes previously predicted to allow EV2 from those that do not is **discourse familiarity**. That is, both the emotive factives and the classic Response Stance predicates (*deny, doubt*, Cattell 1978) require that the embedded proposition is already common ground in the discourse. This is unlike the Volunteer Stance predicates (*say, believe*, Cattell 1978), and the cognitive factives: "Guess what! I found out/#I'm happy that John got accepted to Harvard.".

In addition to this observation, we give new data demonstrating that the volunteer stance predicates, when negated, take one the property of requiring the embedded proposition to be discourse familiar: "Guess what! John told me/#didn't tell me that he got accepted to Harvard." Given this observation, and the proposal that EV2 is blocked by discourse familiarity, we are now able to make a new prediction, namely that the volunteer stance predicates should show much lower rates of EV2 under negation. Indeed, this prediction is borne out; a Wilcoxon signed-rank test shows a significant difference in rates of EV2 for volunteer-stance predicates compared with their non-negated counterparts. This effect is consistent across every corpus analyzed. Mean rates of these distributions are shown in the rightmost plot of 3.

Selected References: Borin et al. (2012). Korp-the corpus infrastructure of språkbanken Haegeman, L. (2014). Locality and the distribution of main clause phenomena. Heycock, C. and J. Wallenberg (2013). How variational acquisition drives syntactic change. Jensen, T. J. and T. K. Christensen (2013). The distribution and semantics of "main clause word order" in spoken Danish complement clauses. Julien, M. (2009). Embedded clauses with main clause word order in Mainland Scandinavian. Kastner, I. (2015). Factivity mirrors interpretation. Wiklund, et al. (2009). On the distribution and illocution of V2 in Scandinavian that-clauses.