

Embedded V2 is anti-licensed by discourse familiarity:  
Quantitative and statistical evidence

Spencer Caplan & Kajsa Djärv

University of Pennsylvania

Formal Ways of Analyzing Variation  
University of York

June 30, 2017

# Introduction

## Quantitative corpus study of Swedish Embedded V2 [EV2]:

Main Clauses: obligatory V2

- (1) Han **gillar** inte (\***gillar**) hundar.  
 he likes not (likes) hundar.  
 'He doesn't like dogs.'

Subordinate Clauses: optional V2

- (2) Han sa att han (**gillar**) inte (**gillar**) hundar.  
 he said that he (**likes**) not (**likes**) hundar.  
 'He said that he doesn't like dogs.'

# Introduction

## Theoretical and empirical challenges raised by EV2:

- EV2 involves ('optional') movement of the finite verb to C;

# Introduction

## Theoretical and empirical challenges raised by EV2:

- EV2 involves ('optional') movement of the finite verb to C;
- and is a classic Main Clause Phenomenon.

# Introduction

## Theoretical and empirical challenges raised by EV2:

- EV2 involves ('optional') movement of the finite verb to C;
- and is a classic Main Clause Phenomenon.

**Question:** What drives the (variable) realization of EV2, and of embedded Main Clause Phenomena more generally?

# Introduction

## Theoretical and empirical challenges raised by EV2:

- EV2 involves ('optional') movement of the finite verb to C;
- and is a classic Main Clause Phenomenon.

**Question:** What drives the (variable) realization of EV2, and of embedded Main Clause Phenomena more generally?

- Subtle and variable judgments has made it difficult to provide a solid theoretical account;

# Introduction

## Theoretical and empirical challenges raised by EV2:

- EV2 involves ('optional') movement of the finite verb to C;
- and is a classic Main Clause Phenomenon.

**Question:** What drives the (variable) realization of EV2, and of embedded Main Clause Phenomena more generally?

- Subtle and variable judgments has made it difficult to provide a solid theoretical account;
- Theoretical and experimental work suggest a complex interaction of different (linguistic and extra-linguistic) factors;

# Introduction

## Theoretical and empirical challenges raised by EV2:

- EV2 involves ('optional') movement of the finite verb to C;
- and is a classic Main Clause Phenomenon.

**Question:** What drives the (variable) realization of EV2, and of embedded Main Clause Phenomena more generally?

- Subtle and variable judgments has made it difficult to provide a solid theoretical account;
- Theoretical and experimental work suggest a complex interaction of different (linguistic and extra-linguistic) factors;
- Corpus data has been helpful, but limited in scope.



# Introduction

**In what contexts is EV2 licensed?**

# Introduction

**In what contexts is EV2 licensed?**

**Formal Factors:** Two broad perspectives

# Introduction

In what contexts is EV2 licensed?

**Formal Factors:** Two broad perspectives

- **Lexically encoded** property of certain predicates
  - Defined in terms of different lexical classes

# Introduction

In what contexts is EV2 licensed?

**Formal Factors:** Two broad perspectives

- **Lexically encoded** property of certain predicates
  - Defined in terms of different lexical classes
- **Pragmatically derived** effect in a given context
  - Defined in terms of Assertion, or Question Under Discussion

# Introduction

In what contexts is EV2 licensed?

**Formal Factors:** Two broad perspectives

- **Lexically encoded** property of certain predicates
  - Defined in terms of different lexical classes
- **Pragmatically derived** effect in a given context
  - Defined in terms of Assertion, or Question Under Discussion

**Stylistic Factors:**

There is a prescriptive rule in Swedish against V>Neg order in embedded clauses.

→ Expect EV2 to correlate with formality.

# Introduction

## Big picture questions

- How to account for syntactic variability (or 'optionality')?
- How can theoretical linguistic questions be addressed using corpus-based and statistical methodologies?

# Introduction

**Main goal of talk:** Test 3 theoretical claims about the licensing of EV2.

- 1 **Lexical licensing 1:** Factivity blocks the derivation of main clause syntax ([Haegeman and Ürögdi 2010](#); [Haegeman 2014](#); [Kastner 2015](#))
- 2 **Lexical licensing 2:** EV2 is possible (but not obligatory) under a subset of Hooper & Thompson's (1973) predicate classes ([Wiklund et al. 2009](#); [Djävrv et al. 2017](#))
- 3 **Pragmatic licensing:** EV2 is driven by Main Point/at-issue status of the embedded proposition ([Julien 2009](#); [Jensen and Christensen 2013](#))

**Proposal:** Corpus data supports hypothesis that EV2 is blocked by discourse familiarity (building on [Djävrv 2017](#)).

# Road map

- 1 **Methods**
- 2 Experiments
- 3 Proposal
- 4 Conclusion



# Methods

How to investigate the problem computationally:

# Methods

How to investigate the problem computationally:

Given a text we'd like to:

- Automatically identify sentences which contain embedded clauses —our domain of potential application
- Algorithmically classify the status of such sentences with respect to EV2 status

# Methods

How to investigate the problem computationally:

Given a text we'd like to:

- Automatically identify sentences which contain embedded clauses —our domain of potential application
- Algorithmically classify the status of such sentences with respect to EV2 status

**Data Limitations:**

However without large-scale *parsed* data such a task is non-trivial.

# Corpora

Data was extracted from [Språkbanken](#) —a series of large-scale Swedish text corpora with automatically assigned part-of-speech tag information ([Borin et al. 2012](#)).

Genres and styles range from blogs and online forums, to newspapers, to government and academic texts.

# Coding a Diagnostic I

## Utilizing Part of Speech-tagged data:

- Find instances of the overt complementizer *att* to identify embedded clauses

# Coding a Diagnostic I

## Utilizing Part of Speech-tagged data:

- Find instances of the overt complementizer *att* to identify embedded clauses
- Ignore cases such as future-marking (*kommer att*), control structures (*att + VB*), and adverbial clauses (*för att*)

# Coding a Diagnostic I

## Utilizing Part of Speech-tagged data:

- Find instances of the overt complementizer *att* to identify embedded clauses
- Ignore cases such as future-marking (*kommer att*), control structures (*att + VB*), and adverbial clauses (*för att*)
- After removing those, only consider sentences with a single complementizer —no multiple embedding

# Coding a Diagnostic I

## Utilizing Part of Speech-tagged data:

- Find instances of the overt complementizer *att* to identify embedded clauses
- Ignore cases such as future-marking (*kommer att*), control structures (*att + VB*), and adverbial clauses (*för att*)
- After removing those, only consider sentences with a single complementizer —no multiple embedding
- The matrix verb will be the lowest (right-most) verbal element occurring before *att*. The embedded verb is the highest (left-most) verbal element following *att*.



# Coding a Diagnostic I

## Utilizing Part of Speech-tagged data:

- Find instances of the overt complementizer *att* to identify embedded clauses
- Ignore cases such as future-marking (*kommer att*), control structures (*att + VB*), and adverbial clauses (*för att*)
- After removing those, only consider sentences with a single complementizer —no multiple embedding
- The matrix verb will be the lowest (right-most) verbal element occurring before *att*. The embedded verb is the highest (left-most) verbal element following *att*.
- Verbs are lemmatized (stored without inflectional information).

# Coding a Diagnostic I

## Utilizing Part of Speech-tagged data:

- Find instances of the overt complementizer *att* to identify embedded clauses
- Ignore cases such as future-marking (*kommer att*), control structures (*att + VB*), and adverbial clauses (*för att*)
- After removing those, only consider sentences with a single complementizer —no multiple embedding
- The matrix verb will be the lowest (right-most) verbal element occurring before *att*. The embedded verb is the highest (left-most) verbal element following *att*.
- Verbs are lemmatized (stored without inflectional information).

## Coding a Diagnostic II

Considering only sentences which contain adverbial *inte* directly preceding or following embedded verb:

- If *inte* precedes the embedded verb, the structure is *V in situ*
- If *inte* follows the embedded verb, the structure is EV2

(3) Han sa att han (**inte**) gillar (**inte**) hundar.  
 he said that he **not** likes **not** dogs  
 'He said that he doesn't like dogs.'

## Coding a Diagnostic II

Considering only sentences which contain adverbial *inte* directly preceding or following embedded verb:

- If *inte* precedes the embedded verb, the structure is *V in situ*
- If *inte* follows the embedded verb, the structure is EV2

(3) Han sa att han (**inte**) gillar (**inte**) hundar.  
he said that he **not** likes **not** dogs  
'He said that he doesn't like dogs.'

Along the way we can extract a wide-range of statistical information:

- Conditional probabilities
- Matrix verb modification.
- Clause length
- etc.

# Implementation

**From this we output statistics for each lemma:**

- Number of cases which show EV2 or *in situ* order
- Control for factors such as frequency (overall, matrix, embedded), number of embedded clause cases, etc.

# Implementation

**From this we output statistics for each lemma:**

- Number of cases which show EV2 or *in situ* order
- Control for factors such as frequency (overall, matrix, embedded), number of embedded clause cases, etc.

**All code is available on Github:**

<https://github.com/scaplan/ev2-optionality>

(Documentation is on-going, so please feel free to contact us if you'd like to use or modify the code-base!)

# Verb Classification (Semantic Categories)

Rates of EV2 by corpus can be computed over all verb lemmas.

## Verb Classification (Semantic Categories)

Rates of EV2 by corpus can be computed over all verb lemmas.

However to examine effects by semantic feature/category some tagging is required:

- Approx. 25 verbs were tagged for each semantic class a la [Hooper and Thompson \(1973\)](#)
- This was done largely on the basis frequency (of taking embedded clause)
- Additional semantic properties (e.g. factivity) are provided from such a categorization
- (Adjectival predicates, e.g., *be worried*, *be aware*, *be sad*, are excluded for the time being)

(As with other data, the verbs and categorization used are available on [Github](#))



# Road map

- 1 Methods
- 2 Experiments**
- 3 Proposal
- 4 Conclusion

# Experiment 1: Effect of Style

## Previous Findings:

- Heycock and Wallenberg (2013) find rates of 31% EV2 in blog-data compared to <1% in published novels

# Experiment 1: Effect of Style

## Previous Findings:

- Heycock and Wallenberg (2013) find rates of 31% EV2 in blog-data compared to <1% in published novels

## First Goals:

- Replicate direction of previous work and explore basic distributional facts.
- Does EV2 correlate with formality?
- Compare the overall rates of EV2 across different corpora (more formal—less formal writing).

# Experiment 1: Effect of Style

| Genre           | Corpus              | Sentences | Proportion<br>p(ev2) |
|-----------------|---------------------|-----------|----------------------|
| Blogs/Forums    | Familjeliv-känsliga | 5971907   | <b>0.0636</b>        |
|                 | Familjeliv-nöje     | 458699    | <b>0.0555</b>        |
|                 | Familjeliv-adoption | 77008     | <b>0.0545</b>        |
|                 | Familjeliv-expert   | 57478     | <b>0.0522</b>        |
|                 | Bloggmix            | 2713376   | <b>0.0502</b>        |
|                 | Flashback-Politik   | 2841872   | <b>0.0457</b>        |
| Academic        | Sweacsam            | 52678     | <b>0.0375</b>        |
|                 | Academy-humanities  | 60931     | <b>0.0283</b>        |
| Government      | Rd-bet              | 372054    | <b>0.0163</b>        |
|                 | Rd-ds               | 172657    | <b>0.0141</b>        |
|                 | Rd-fpm              | 5259      | <b>0.0138</b>        |
|                 | Rd-skfr             | 81800     | <b>0.0098</b>        |
| Accessible news | Attasidor           | 8059      | <b>0.0081</b>        |

**Table:** Rates of embedded V2 across corpora of varying formality.

# Experiment 1: Effect of Style

- Does EV2 to correlate with formality?
- Yes; replicates previous pattern from [Heycock and Wallenberg \(2013\)](#).

# Experiments

**Main goal:** Test 3 theoretical claims about the licensing of EV2.

## Experiment 2: EV2 and Factivity

Lexical licensing account á la [Haegeman \(2014\)](#); [Kastner \(2015\)](#) predicts that EV2 should be blocked under [factive verbs](#).

## Experiment 2: EV2 and Factivity

Lexical licensing account á la [Haegeman \(2014\)](#); [Kastner \(2015\)](#) predicts that EV2 should be blocked under **factive verbs**.

**Factivity:** Presupposition of certain predicates (e.g. *know*, *realize*, *discover*, *regret*, *resent*) that the embedded proposition is true:

| John <b>verbs</b> that it's raining | $\rightsquigarrow$ it's raining |
|-------------------------------------|---------------------------------|
| knows                               | ✓                               |
| regrets                             | ✓                               |
| says                                | ✗                               |
| believes                            | ✗                               |
| denies                              | ✗                               |

Table: **Factive** and **Non-factive** verbs.



## Experiment 2: EV2 and Factivity

Lexical licensing account á la [Haegeman \(2014\)](#); [Kastner \(2015\)](#) predicts that EV2 should be blocked under **factive verbs**.

**Factivity:** Presupposition of certain predicates (e.g. *know*, *realize*, *discover*, *regret*, *resent*) that the embedded proposition is true:

| John <b>verbs</b> that it's raining | $\rightsquigarrow$ it's raining |
|-------------------------------------|---------------------------------|
| knows                               | ✓                               |
| regrets                             | ✓                               |
| says                                | ✗                               |
| believes                            | ✗                               |
| denies                              | ✗                               |

Table: **Factive** and **Non-factive** verbs.

- **Claim:** Factive predicates select a complement (with DP-layer) that blocks the derivation of Main Clause Phenomena.

## Experiment 2: EV2 and Factivity

Lexical licensing account á la Haegeman (2014); Kastner (2015) predicts that EV2 should be blocked under **factive verbs**.

**Factivity:** Presupposition of certain predicates (e.g. *know*, *realize*, *discover*, *regret*, *resent*) that the embedded proposition is true:

| John <b>verbs</b> that it's raining | $\rightsquigarrow$ it's raining |
|-------------------------------------|---------------------------------|
| knows                               | ✓                               |
| regrets                             | ✓                               |
| says                                | ✗                               |
| believes                            | ✗                               |
| denies                              | ✗                               |

Table: **Factive** and **Non-factive** verbs.

- **Claim:** Factive predicates select a complement (with DP-layer) that blocks the derivation of Main Clause Phenomena.
- **Prediction:** Factives should disallow EV2.

# Experiment 2: EV2 and Factivity

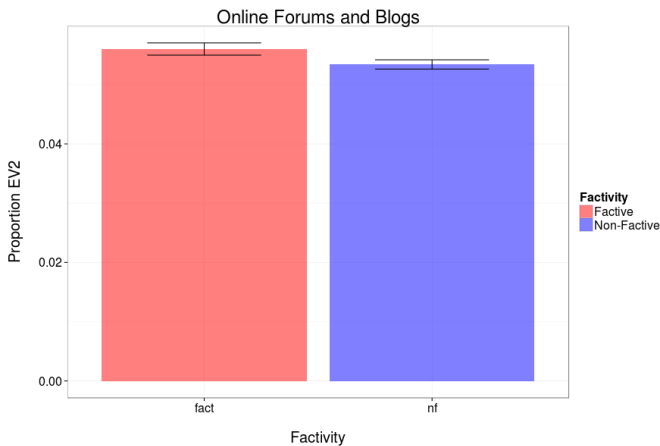


Figure: EV2 under Factive vs. Non-factive verbs.

## Experiment 2: EV2 and Factivity

### Wilcoxon rank sum test:

- $W = 748$
- $p\text{-value} = 0.6949$

|              | Types | $p(\text{ev2})$ |
|--------------|-------|-----------------|
| Factives     | 35    | 0.0337          |
| Non-Factives | 45    | 0.0356          |

**Table:** Rates of EV2 under factive vs. non-factive verbs in Flashback-Politik

## Experiment 3: Selection by certain predicate classes

Hooper and Thompson (1973) identifies five predicate classes—distinguished in terms of their lexical semantics, that are relevant to the licensing of Main Clause Phenomena.

Example from English, VP-preposing:

- (4) Mary plans for John to marry her, and marry her he will.

Possible under *say*, *believe*, *know*, but not under *deny*, *resent*:

- (5) Mary plans for John to marry her, and. . .
- I {**say**, **believe**, **know**} that marry her he will.
  - \*I {**deny**, **resent**} that marry her he will.

## Experiment 3: Selection by certain predicate classes

Wiklund et al. (2009): EV2 is licensed in the complements of the predicate classes exemplified by *say*, *believe*, *know* only.

|                      | <i>say</i> | <i>believe</i> | <i>know</i> | <i>resent</i> | <i>doubt, deny</i> |
|----------------------|------------|----------------|-------------|---------------|--------------------|
| Factive              | ✗          | ✗              | ✓           | ✓             | ✗                  |
| Speech Act           | ✓          | ✗              | ✗           | ✗             | ✓/✗                |
| Doxastic State       | ✗          | ✓              | ✓           | ✓             | ✓/✗                |
| Emotive              | ✗          | ✗              | ✗           | ✓             | ✗                  |
| Discourse Familiarly | ✗          | ✗              | ✗           | ✓             | ✓                  |
| <b>EV2</b>           | <b>Yes</b> | <b>Yes</b>     | <b>Yes</b>  | <b>No</b>     | <b>No</b>          |

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

## Experiment 3: Selection by certain predicate classes

Wiklund et al. (2009): EV2 is licensed in the complements of the predicate classes exemplified by *say*, *believe*, *know* only.

|                      | <i>say</i> | <i>believe</i> | <i>know</i> | <i>resent</i> | <i>doubt, deny</i> |
|----------------------|------------|----------------|-------------|---------------|--------------------|
| Factive              | ✗          | ✗              | ✓           | ✓             | ✗                  |
| Speech Act           | ✓          | ✗              | ✗           | ✗             | ✓/✗                |
| Doxastic State       | ✗          | ✓              | ✓           | ✓             | ✓/✗                |
| Emotive              | ✗          | ✗              | ✗           | ✓             | ✗                  |
| Discourse Familiarly | ✗          | ✗              | ✗           | ✓             | ✓                  |
| <b>EV2</b>           | <b>Yes</b> | <b>Yes</b>     | <b>Yes</b>  | <b>No</b>     | <b>No</b>          |

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

Support from experimental data from Djärv et al. (2017)  
(judgment data, likert scale)

## Experiment 3: Selection by certain predicate classes

Wiklund et al. (2009): EV2 is licensed in the complements of the predicate classes exemplified by *say*, *believe*, *know* only.

|                      | <i>say</i> | <i>believe</i> | <i>know</i> | <i>resent</i> | <i>doubt, deny</i> |
|----------------------|------------|----------------|-------------|---------------|--------------------|
| Factive              | ✗          | ✗              | ✓           | ✓             | ✗                  |
| Speech Act           | ✓          | ✗              | ✗           | ✗             | ✓/✗                |
| Doxastic State       | ✗          | ✓              | ✓           | ✓             | ✓/✗                |
| Emotive              | ✗          | ✗              | ✗           | ✓             | ✗                  |
| Discourse Familiarly | ✗          | ✗              | ✗           | ✓             | ✓                  |
| <b>EV2</b>           | <b>Yes</b> | <b>Yes</b>     | <b>Yes</b>  | <b>No</b>     | <b>No</b>          |

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

Support from experimental data from Djärv et al. (2017)  
(judgment data, likert scale)

Our data shows mixed evidence in favour of this account.



## Experiment 3: Selection by certain predicate classes

Wiklund et al. (2009): EV2 is licensed in the complements of the predicate classes exemplified by *say*, *believe*, *know* only.

|                      | <i>say</i> | <i>believe</i> | <i>know</i> | <i>resent</i> | <i>doubt, deny</i> |
|----------------------|------------|----------------|-------------|---------------|--------------------|
| Factive              | ✗          | ✗              | ✓           | ✓             | ✗                  |
| Speech Act           | ✓          | ✗              | ✗           | ✗             | ✓/✗                |
| Doxastic State       | ✗          | ✓              | ✓           | ✓             | ✓/✗                |
| Emotive              | ✗          | ✗              | ✗           | ✓             | ✗                  |
| Discourse Familiarly | ✗          | ✗              | ✗           | ✓             | ✓                  |
| <b>EV2</b>           | <b>Yes</b> | <b>Yes</b>     | <b>Yes</b>  | <b>No</b>     | <b>No</b>          |

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

Support from experimental data from Djärv et al. (2017)  
(judgment data, likert scale)

Our data shows mixed evidence in favour of this account.

## Experiment 3: Selection by certain predicate classes

Looking at the rates of EV2 across these verb classes in *one* corpus;  
→ Strong support **in favour** of such a lexical licensing account

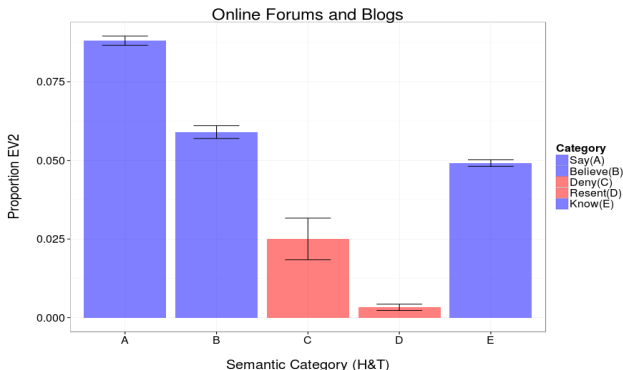


Figure: EV2 by predicate class (Blog texts).

## Experiment 3: Selection by certain predicate classes

Looking at EV2 in a different corpus (representing a different genre);

→ Strong support **against** such a lexical licensing account.

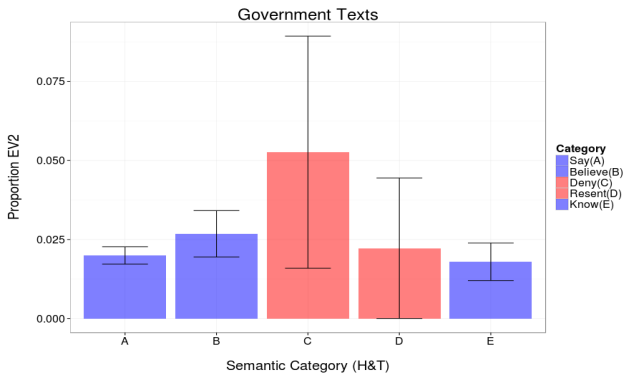


Figure: EV2 under H&T's verb types (government texts).

# Experiment 3: Selection by certain predicate classes

**Genre Effect:** overall effect of lexical class, but distribution of EV2 by verb class varies across corpora (different discourse types).

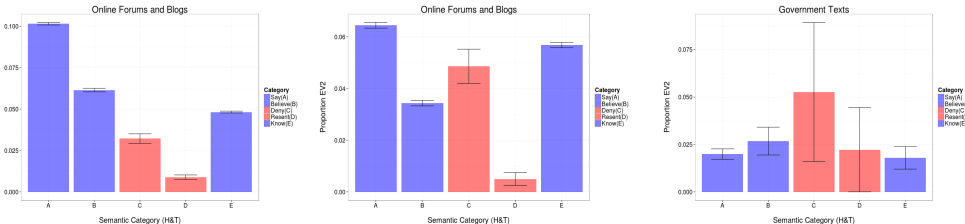
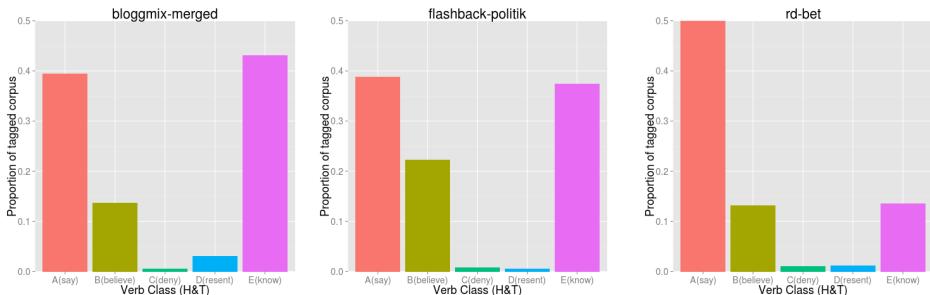


Figure: EV2 across corpora and predicate type.

# Experiment 3: Selection by certain predicate classes

## Distributional properties of verb classes:



**Figure:** Frequency of lexical class (as proportion of total tagged verbs) across corpora.

## Experiment 3: Selection by certain predicate classes

- We find variability of EV2 by predicate class across discourse type
  - This is contra a strong version of the lexical licensing account —whereby verbs of a given class either allows or disallows EV2 (regardless of discourse-context).
- Moreover, variability *within* the different classes provides further evidence against such an account. . .  
—whereby all verbs in a given class should either allow or disallow EV2.

## Experiment 3: Selection by certain predicate classes

- We find variability of EV2 by predicate class across discourse type
  - This is contra a strong version of the lexical licensing account —whereby verbs of a given class either allows or disallows EV2 (regardless of discourse-context).
- Moreover, variability *within* the different classes provides further evidence against such an account. . .  
—whereby all verbs in a given class should either allow or disallow EV2.

# Experiment 3: Selection by certain predicate classes

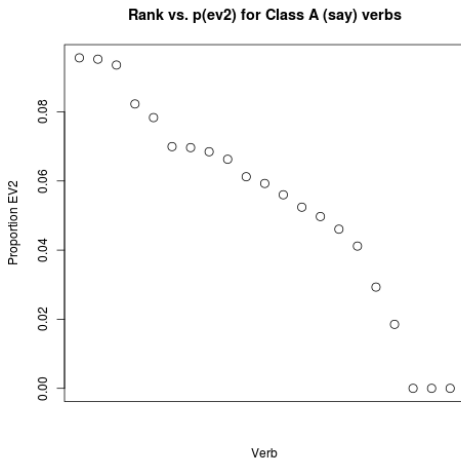


Figure: EV2 by predicate (within unified semantic class).



# Road map

- 1 Methods
- 2 Experiments
- 3 **Proposal**
- 4 Conclusion

# Accounting for the Genre Effect

**Hypothesis:** The genre effect is driven by pragmatic factors

- EV2 is associated with some pragmatic meaning ( $P_M$ )
- The overall meaning of a sentence with EV2 ( $S_M$ ) depends on  $P_M$  + the semantics of the matrix predicate
- The extent to which different  $S_M$  are **used** will vary across different types of discourses.

# Accounting for the Genre Effect

**Hypothesis:** The genre effect is driven by pragmatic factors

- EV2 is associated with some pragmatic meaning ( $P_M$ )
- The overall meaning of a sentence with EV2 ( $S_M$ ) depends on  $P_M$  + the semantics of the matrix predicate
- The extent to which different  $S_M$  are **used** will vary across different types of discourses.

**Question:** What is the specific nature of  $P_M$ ?

## The Pragmatics of EV2: Previous Work

Julien (2009); Jensen and Christensen (2013): EV2 is obligatory if the embedded clause provides the *at-issue* content, or the Main Point of the Utterance [MPU] (e.g. Simons 2007; Roberts 2012)

- (6) Q Where's John?  
A He said he can't make it today.

## The Pragmatics of EV2: Previous Work

Julien (2009); Jensen and Christensen (2013): EV2 is obligatory if the embedded clause provides the *at-issue* content, or the Main Point of the Utterance [MPU] (e.g. Simons 2007; Roberts 2012)

- (6) Q Where's John?  
A He said he can't make it today.

### Problems with the QUD-based analysis:

- Experimental work by Djärv et al. (2017) manipulated the discourse as in (6), and found no effect on EV2.
- Judgment data supports these findings (Wiklund et al. 2009).
- Moreover, if *at-issue*/MPU status was in fact what's driving the distribution across predicate classes, then we expect that the predicates that block EV2-complements should also block embedded *at-issue*/MPU content.

# The Pragmatics of EV2: Previous Work

This prediction appears to be borne out for the *resent*-class:

(7) Q Where's John?

A #I resent that he can't make it today.

\*EV2

# The Pragmatics of EV2: Previous Work

This prediction appears to be borne out for the *resent*-class:

- (7) Q Where's John?  
 A #I resent that he can't make it today. \*EV2

But it fails for the *deny*-class:

- (8) Q Where's John?  
 A I doubt he can make it today. \*EV2

# The Pragmatics of EV2: Proposal

The predicate classes that allow vs. disallow EV2 are distinguished by a different property: **discourse familiarity**



## The Pragmatics of EV2: Proposal

The predicate classes that allow vs. disallow EV2 are distinguished by a different property: **discourse familiarity**

- (9) [Uttered out of the blue:]  
*Guess what — / You know what —*
- a. ✓ They **said** on the radio that Trump resigned.
  - b. ✓ I **think** Trump resigned.
  - c. ✓ I just **found out** that Trump resigned.

## The Pragmatics of EV2: Proposal

The predicate classes that allow vs. disallow EV2 are distinguished by a different property: **discourse familiarity**

- (9) [Uttered out of the blue:]  
*Guess what — / You know what —*
- a. ✓ They **said** on the radio that Trump resigned.
  - b. ✓ I **think** Trump resigned.
  - c. ✓ I just **found out** that Trump resigned.  
→ *Don't require discourse familiarity*

## The Pragmatics of EV2: Proposal

The predicate classes that allow vs. disallow EV2 are distinguished by a different property: **discourse familiarity**

- (9) [Uttered out of the blue:]  
*Guess what — / You know what —*
- a. ✓ They **said** on the radio that Trump resigned.
  - b. ✓ I **think** Trump resigned.
  - c. ✓ I just **found out** that Trump resigned.  
→ *Don't require discourse familiarity*
  - d. # I **appreciate** that Trump resigned.
  - e. # I **doubt** Trump resigned.

## The Pragmatics of EV2: Proposal

The predicate classes that allow vs. disallow EV2 are distinguished by a different property: **discourse familiarity**

- (9) [Uttered out of the blue:]  
*Guess what — / You know what —*
- a. ✓ They **said** on the radio that Trump resigned.
  - b. ✓ I **think** Trump resigned.
  - c. ✓ I just **found out** that Trump resigned.  
→ *Don't require discourse familiarity*
  - d. # I **appreciate** that Trump resigned.
  - e. # I **doubt** Trump resigned.  
→ *Require discourse familiarity*

## The Pragmatics of EV2: Proposal

The predicate classes that allow vs. disallow EV2 are distinguished by a different property: **discourse familiarity**

- (9) [Uttered out of the blue:]  
*Guess what — / You know what —*
- ✓ They **said** on the radio that Trump resigned.
  - ✓ I **think** Trump resigned.
  - ✓ I just **found out** that Trump resigned.  
 → *Don't require discourse familiarity*
  - # I **appreciate** that Trump resigned.
  - # I **doubt** Trump resigned.  
 → *Require discourse familiarity*

→ The predicates that disallow EV2 independently require that their complements are discourse familiar.

## The Pragmatics of EV2: Proposal

The predicate classes that allow vs. disallow EV2 are distinguished by a different property: **discourse familiarity**

- (9) [Uttered out of the blue:]  
*Guess what — / You know what —*
- a. ✓ They **said** on the radio that Trump resigned.
  - b. ✓ I **think** Trump resigned.
  - c. ✓ I just **found out** that Trump resigned.  
→ *Don't require discourse familiarity*
  - d. # I **appreciate** that Trump resigned.
  - e. # I **doubt** Trump resigned.  
→ *Require discourse familiarity*

→ The predicates that disallow EV2 independently require that their complements are discourse familiar.

**Proposal:** EV2 is anti-licensed by discourse familiarity.

# Evidence that Discourse Familiarity blocks EV2

**Problem:** Can't measure the common ground in a corpus.

## Evidence that Discourse Familiarity blocks EV2

**Problem:** Can't measure the common ground in a corpus.

**Solution:** Matrix Negation– the volunteer stance verbs (*say, think*) take on the property of requiring discourse familiarity:



## Evidence that Discourse Familiarity blocks EV2

**Problem:** Can't measure the common ground in a corpus.

**Solution:** Matrix Negation— the volunteer stance verbs (*say, think*) take on the property of requiring discourse familiarity:

(10) [Uttered out of the blue:]

*Guess what — / You know what —*

- #They **didn't say** on the radio that Trump resigned.
- #I **don't think** Trump resigned.
- #I **appreciate** that Trump resigned.
- #I **doubt** that Trump resigned.

## Evidence that Discourse Familiarity blocks EV2

**Problem:** Can't measure the common ground in a corpus.

**Solution:** Matrix Negation— the volunteer stance verbs (*say, think*) take on the property of requiring discourse familiarity:

(10) [Uttered out of the blue:]

*Guess what — / You know what —*

- #They **didn't say** on the radio that Trump resigned.
- #I **don't think** Trump resigned.
- #I **appreciate** that Trump resigned.
- #I **doubt** that Trump resigned.

→ *Require discourse familiarity*

## Evidence that Discourse Familiarity blocks EV2

**Problem:** Can't measure the common ground in a corpus.

**Solution:** Matrix Negation— the volunteer stance verbs (*say*, *think*) take on the property of requiring discourse familiarity:

(10) [Uttered out of the blue:]

*Guess what — / You know what —*

- #They **didn't say** on the radio that Trump resigned.
- #I **don't think** Trump resigned.
- #I **appreciate** that Trump resigned.
- #I **doubt** that Trump resigned.

→ *Require discourse familiarity*

**Prediction:** Rate of EV2 should be notably lower for *not say*, *not think* etc. than their non-negated counterparts.

# Evidence that Discourse Familiarity Inhibits EV2

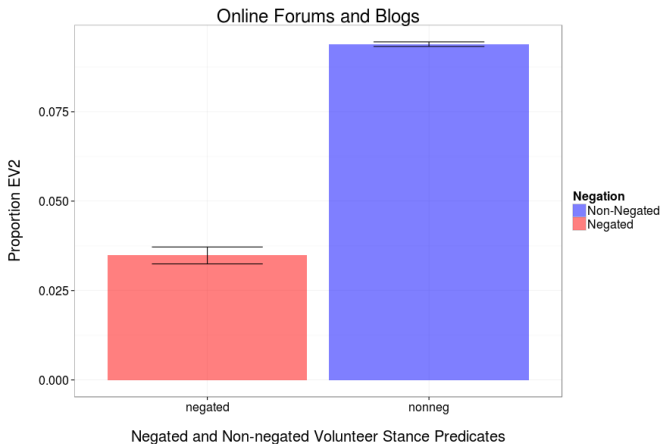


Figure: EV2 for volunteer stance predicates under negation (or not)

# Evidence that Discourse Familiarity Inhibits EV2

## Effect of negation on EV2 for Volunteer Stance predicates (Wilcoxon rank sum test):

Corpus of online forums:

- $W = 749$
- $p\text{-value} = 0.007677$

This trend holds across all corpora we looked at.

# Road map

- 1 Methods
- 2 Experiments
- 3 Proposal
- 4 **Conclusion**

# Conclusions I

**Empirical contributions:** Based on our distributional and statistical findings, any theory of EV2 needs to account for the following facts:

# Conclusions I

**Empirical contributions:** Based on our distributional and statistical findings, any theory of EV2 needs to account for the following facts:

- Rates of EV2 are graded by discourse: formality and genre-effect (but EV2 is never totally blocked);



# Conclusions I

**Empirical contributions:** Based on our distributional and statistical findings, any theory of EV2 needs to account for the following facts:

- Rates of EV2 are graded by discourse: formality and genre-effect (but EV2 is never totally blocked);
- Semantic classes may be correlatory, but are insufficient to capture the variable rates of EV2;

# Conclusions I

**Empirical contributions:** Based on our distributional and statistical findings, any theory of EV2 needs to account for the following facts:

- Rates of EV2 are graded by discourse: formality and genre-effect (but EV2 is never totally blocked);
- Semantic classes may be correlatory, but are insufficient to capture the variable rates of EV2;
- There is a significant interaction under negation—volunteer stance predicates like *say*, *believe* licence EV2; but this is largely blocked under negation.

# Conclusions II

## Theoretical contributions:

- Discourse familiarity  $\neq$  *at-issue*/Main Point status:  
→ The former is what drives the distribution of EV2  
(contra Julien 2009; Jensen and Christensen 2013);

# Conclusions II

## Theoretical contributions:

- Discourse familiarity  $\neq$  *at-issue*/Main Point status:  
→ The former is what drives the distribution of EV2  
(contra Julien 2009; Jensen and Christensen 2013);
- Discourse familiarity is a pragmatic notion—i.e., a property of an utterance in a given context. It is constrained, but not determined by the lexical semantics of specific predicates:

# Conclusions II

## Theoretical contributions:

- Discourse familiarity ≠ *at-issue*/Main Point status:  
→ The former is what drives the distribution of EV2 (contra Julien 2009; Jensen and Christensen 2013);
- Discourse familiarity is a pragmatic notion—i.e., a property of an utterance in a given context. It is constrained, but not determined by the lexical semantics of specific predicates:
  - ① *doubt, resent*—but not **all** factive predicates (e.g. *realize, find out*), require discourse familiarity (contra Haegeman and Ürögdi 2010; Haegeman 2014; Kastner 2015);

# Conclusions II

## Theoretical contributions:

- Discourse familiarity  $\neq$  *at-issue*/Main Point status:  
→ The former is what drives the distribution of EV2 (contra Julien 2009; Jensen and Christensen 2013);
- Discourse familiarity is a pragmatic notion—i.e., a property of an utterance in a given context. It is constrained, but not determined by the lexical semantics of specific predicates:
  - 1 *doubt, resent*—but not **all** factive predicates (e.g. *realize, find out*), require discourse familiarity (contra Haegeman and Ürögdi 2010; Haegeman 2014; Kastner 2015);
  - 2 Discourse familiarity arises in the interaction of certain predicates and embedding operators (e.g., negation).

# Conclusions II

## Theoretical contributions:

- Discourse familiarity ≠ *at-issue*/Main Point status:
  - The former is what drives the distribution of EV2 (contra Julien 2009; Jensen and Christensen 2013);
- Discourse familiarity is a pragmatic notion—i.e., a property of an utterance in a given context. It is constrained, but not determined by the lexical semantics of specific predicates:
  - ① *doubt, resent*—but not **all** factive predicates (e.g. *realize, find out*), require discourse familiarity (contra Haegeman and Ürögdi 2010; Haegeman 2014; Kastner 2015);
  - ② Discourse familiarity arises in the interaction of certain predicates and embedding operators (e.g., negation).
- Next step: experimentally manipulate the conversational context to further test the role of discourse familiarity.

# Acknowledgments

Thank you!

And thanks to...

- Anthony Kroch, and participants in his Syntax seminar at the University of Pennsylvania
- Florian Schwarz, Mitch Marcus, and Betsy Sneller

for helpful feedback and comments!



# References I

- Borin, L., Forsberg, M., Roxendal, J., 2012. Korp-the corpus infrastructure of språkbanken. In: LREC. pp. 474–478.
- Djärv, K., 2017. Facts and Claims: Clausal Complementation Revisited.
- Djärv, K., Heycock, C., Rohde, H., 2017. Assertion and factivity: Towards explaining restrictions on embedded V2 in Scandinavian. Language Science Press.
- Haegeman, L., 2014. Locality and the distribution of main clause phenomena. In: Aboh, E. O., Guasti, M. T., Roberts, I. (Eds.), Locality (Oxford Studies in Comparative Syntax). Oxford University Press, Oxford, Ch. 8, pp. 186–222.
- Haegeman, L., Ürögdi, B., 2010. Referential CPs and DPs: An operator movement account. Theoretical Linguistics 36, 111–152.

## References II

- Heycock, C., Wallenberg, J., 2013. How variational acquisition drives syntactic change: The loss of verb movement in scandinavian. *Journal of Comparative Germanic Linguistics* 16, 127–157.
- Hooper, J., Thompson, S., 1973. On the applicability of root transformations. *Linguistic Inquiry* 4.4, 465–497.
- Jensen, T. J., Christensen, T. K., 2013. Promoting the demoted: The distribution and semantics of “main clause word order” in spoken Danish complement clauses. *Lingua* 137, 38–58.
- Julien, M., 2009. Embedded clauses with main clause word order in Mainland Scandinavian, published on *LingBuzz*: (<http://ling.auf.net/lingBuzz/000475>).
- Kastner, I., 2015. Factivity mirrors interpretation: The selectional requirements of presuppositional verbs. *Lingua* 164, 156–188.

## References III

- Moulton, K., 2015. CPs: Copies and compositionality. *Linguistic Inquiry* 46 (2), 305–342.
- Özyildiz, D., 2016. Knowledge reports without truth. In: *Proceedings of the European Summer School in Logic, Language & Information ESSLLI 2016 Student Session*. Vol. 28. pp. 184–195.
- Roberts, C., 2012. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5, 1–69.
- Simons, M., 2007. Observations on embedding verbs, evidentiality, and presupposition. *Lingua* 117, 1034–1056.
- Wiklund, A.-L., Bentzen, K., Hrafnbjargarson, G. H., Hróarsdóttir, Þ., 2009. On the distribution and illocution of V2 in Scandinavian *that*-clauses. *Lingua* 119 (12), 1914–1938.

# Appendix I

**Experiment 1: Effect of Style** Possible explanation: age grading / change in progress / socially conditioned variation...?

Corpus data from the late 19th century suggests not:

| Genre | Corpus | Sentences | Proportion    |               |
|-------|--------|-----------|---------------|---------------|
|       |        |           | Non-ambiguous | p(ev2)        |
| News  | 1870   | 17084     | 0.0598        | <b>0.0598</b> |
|       | 1860   | 58839     | 0.0620        | <b>0.0620</b> |

**Table:** Rates of embedded V2 in newspapers from 1860 and 1870.



# Appendix I

## Interaction with other types of verbs and embedding operators

- *accept*: a response stance verb; typically disallows EV2

- (14) a. kan du inte bara slappna av och **acceptera** att  
can you not just chill out and accept that  
socialisterna **kan inte** vinna alla gånger ?  
socialists.DEF can not win every time ?  
'Why can't you not just relax and accept that the socialists aren't  
going to win every time?'
- b. **acceptera** att du kan inte älska alla men du **kan**  
accept that you cannot love everyone but you can not  
**inte** hata alla heller  
hate everyone either  
'Accept that you can't love everyone, but you can't hate everyone  
either.'

## Appendix II

- The embedded proposition  $p$  is in CG (discourse familiar);
- The speaker asserts  $\neg p$  (not discourse familiar).

# Appendix I

Djävrv (2017) argue that attitude predicates vary wrt. transitivity:

- **Transitive predicates:** select discourse familiar complement clause; encoded on a definite D-head in the complement clause (similarly to Haegeman and Ürögdi 2010; Haegeman 2014; Kastner 2015; see also Moulton 2015)
- **Ditransitive predicates:** select both a CP that encodes the propositional content of the attitude, and a pronominal argument, anaphoric to the *res* or topic situation that the attitude is about (see Özyildiz 2016).

Two theoretical options to account for the restrictions on EV2:

- **Intervention:** the derivation of EV2 is blocked by the definite D-head (similarly to previous accounts);
- **Selection:** different C-heads in the two different structures have different sets of features relevant to V2.



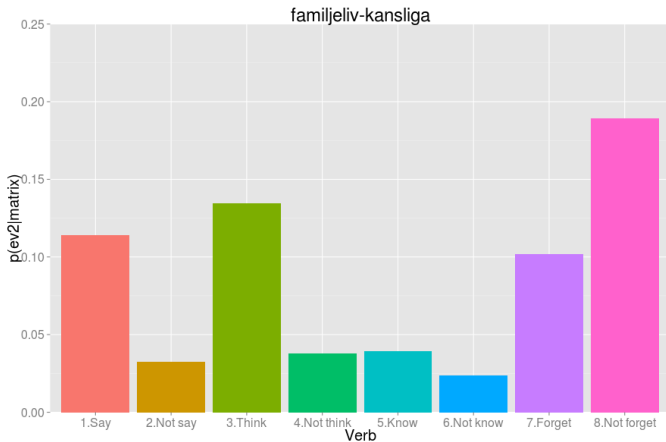
## Appendix: CP-Recursion I

| Intervening Length | Non-ambiguous Cases | $p(\text{ev2} \text{length})$ |
|--------------------|---------------------|-------------------------------|
| 0                  | 128460              | 0.0504                        |
| 1                  | 63774               | 0.0427                        |
| 2                  | 33795               | 0.0417                        |
| 3                  | 19304               | 0.0404                        |
| 4                  | 11757               | 0.0406                        |
| 5                  | 6858                | 0.0383                        |
| 6                  | 4273                | 0.0438                        |
| 7                  | 2699                | 0.0463                        |
| 8                  | 1813                | 0.0408                        |
| 9                  | 1142                | 0.0394                        |

## Appendix: CP-Recursion II

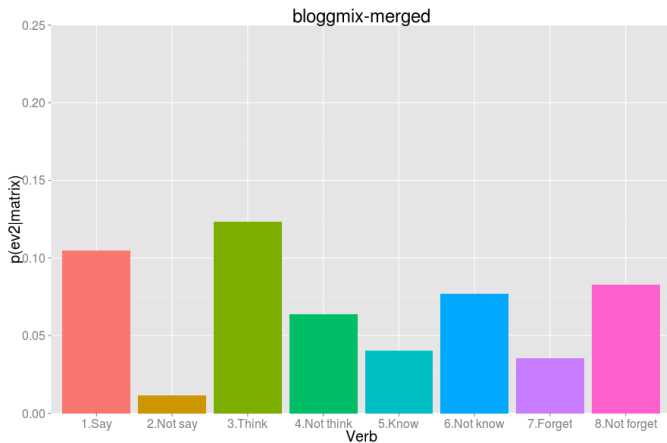
**Table:** Probability of ev2 conditioned on the amount of material (counted in words) intervening between the matrix verb and the complementizer. Data from Flashback-Politik

## Effect of Negation



**Figure:** EV2 for individual verbs under negation (vs. not)

## Appendix



**Figure:** EV2 for individual verbs under negation (vs. not)