Creating a Constructicon from the COBUILD Grammar Patterns and FrameNet
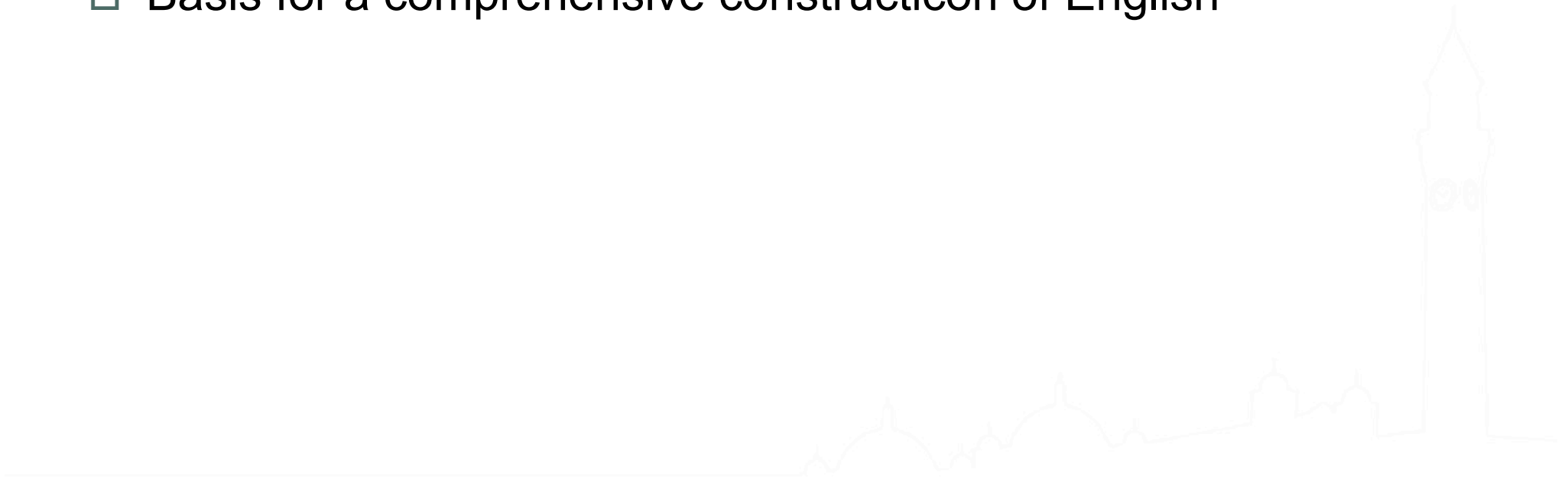
Florent Perek

University of Birmingham
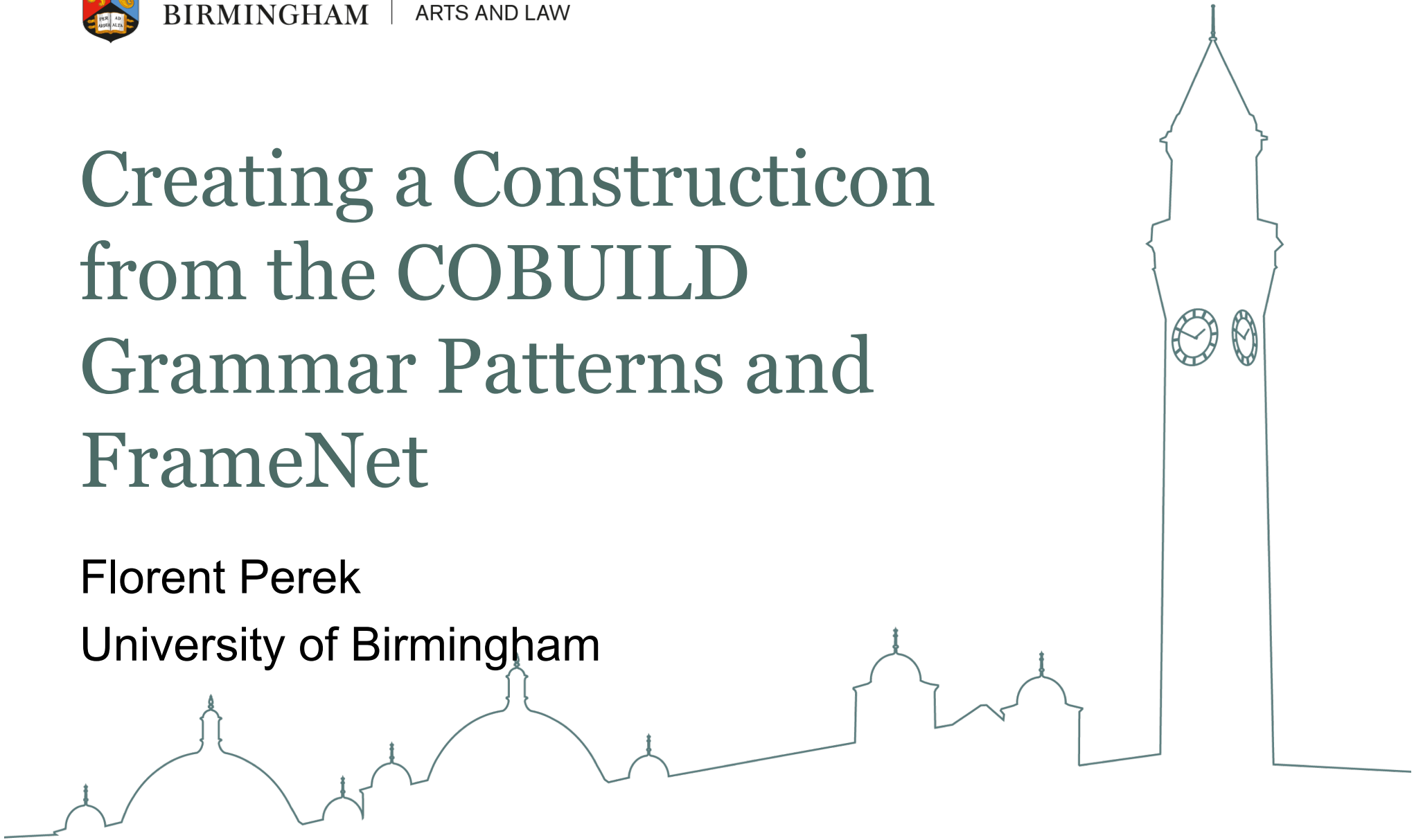
Overview

- ☐ Outline and first results of a new project (with Amanda Patten)

- ☐ Proposal: merge two resources, the COBUILD grammar patterns and FrameNet

- ☐ Basis for a comprehensive constructicon of English

COBUILD

- Lexicographic project started in the 1980s by John Sinclair with Collins publishers in Birmingham

- Design dictionaries entirely from authentic corpus data

- One key insight in particular

  - A word is better described in terms of its typical uses

  - This notably includes the syntactic frames or "patterns" it can occur in
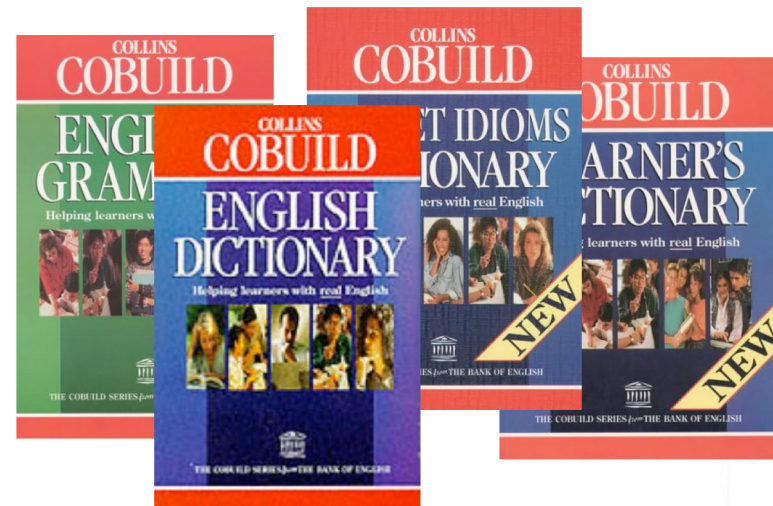
The COBUILD Grammar Patterns

☐ Proposals for compiling a pattern grammar of English (Francis 1993, Hunston & Francis 2000)

   → The COBUILD Grammar Patterns series

☐ Compilation of all the patterns mentioned in the COBUILD entries

   – Volume 1: verbs (Francis et al. 1996)

   – Volume 2: nouns and adjectives (Francis et al. 1998)

☐ Lists of all lexical items attested in these patterns
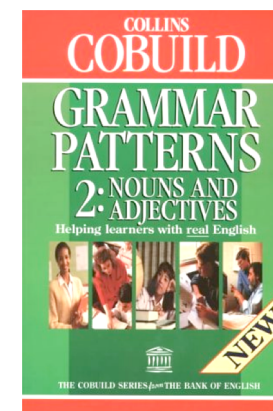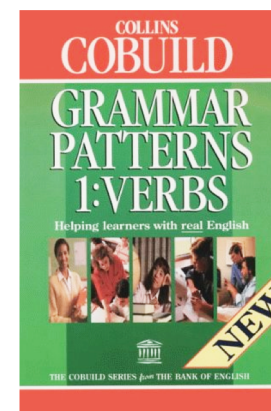
Francis, G. (1993). A corpus-driven approach to grammar – principles, methods and examples. In Baker, M., Francis, G. & Tognini-Bonelli, E. (eds). *Text and Technology: in Honour of John Sinclair*. Amsterdam: Benjamins, pp. 137–156.
Francis, G., Hunston, S. & Manning, E. (1996). *Collins COBUILD Grammar Patterns 1: Verbs*. London: HarperCollins.
Francis, G., Hunston, S. & Manning, E. (1998). *Collins COBUILD Grammar Patterns 2: Nouns and Adjectives*. London: HarperCollins.
Hunston, S. & Francis, G. (2000). Pattern Grammar: A corpus-driven approach to the lexical grammar of English. Amsterdam: Benjamins.

The COBUILD Grammar Patterns

- [ ] 124 patterns for verbs in Francis et al. (1996)

- [ ] 10,522 verbs listed under the patterns

- [ ] Verbs are grouped into meaning groups in each pattern (816 in total, avg. 6.6 groups per pattern)
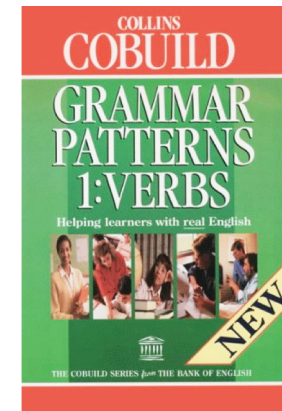
(figures calculated from the XML version provided by HarperCollins)
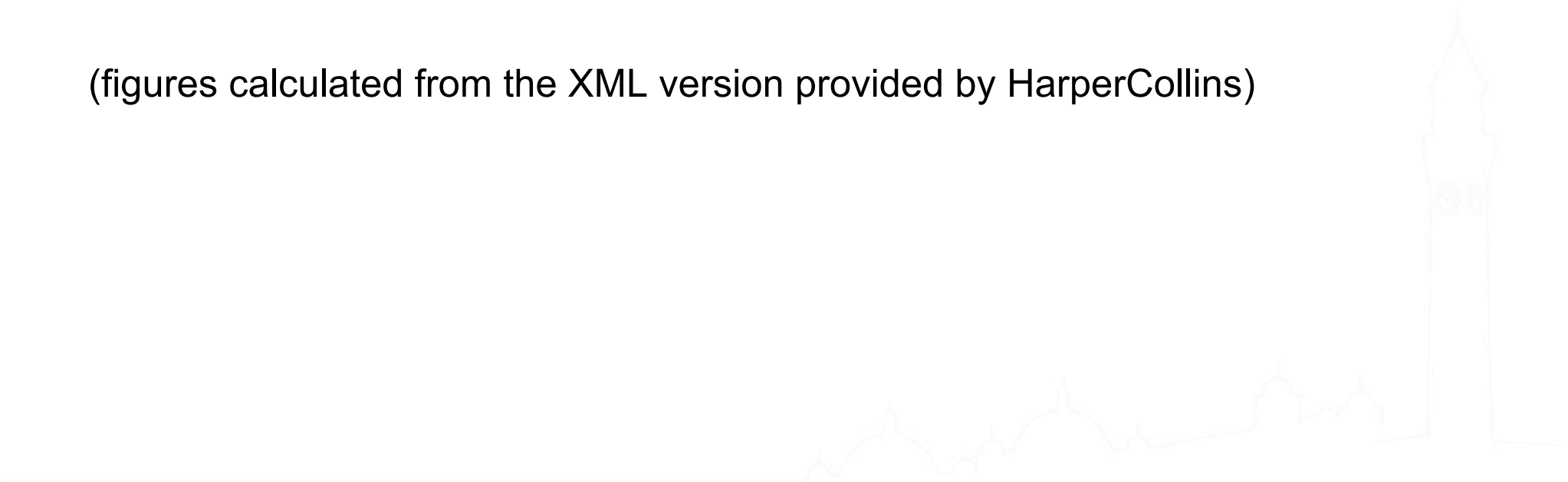
The COBUILD grammar patterns

Example: **V n *of* n**

☐ Verb followed by NP and *of*-PP

☐ Three meaning groups

- The 'rob' and 'free' group: … ***cure*** *her of a disease,*
  … ***robbed*** *them of their watches* (24 verbs)

- The 'inform' group: … ***assured*** *us of their help* (11 verbs)

- The 'acquit' and 'convict' group: … ***clear*** *him of attempting to murder,* … ***suspected*** *him of perjury* (5 verbs)

- 11 other verbs

The COBUILD grammar patterns

- ☐ COBUILD patterns ~ constructions
  - – Single coherent grammatical units
  - – Fixed parts and open slots
- ☐ However, little semantic information: only lexical senses and "meaning groups"
- ☐ To be turned into constructions, they must be paired with meaning and with semantic role information
- ☐ Idea: use FrameNet as a semantic component for patterns

FrameNet

https://framenet.icsi.berkeley.edu
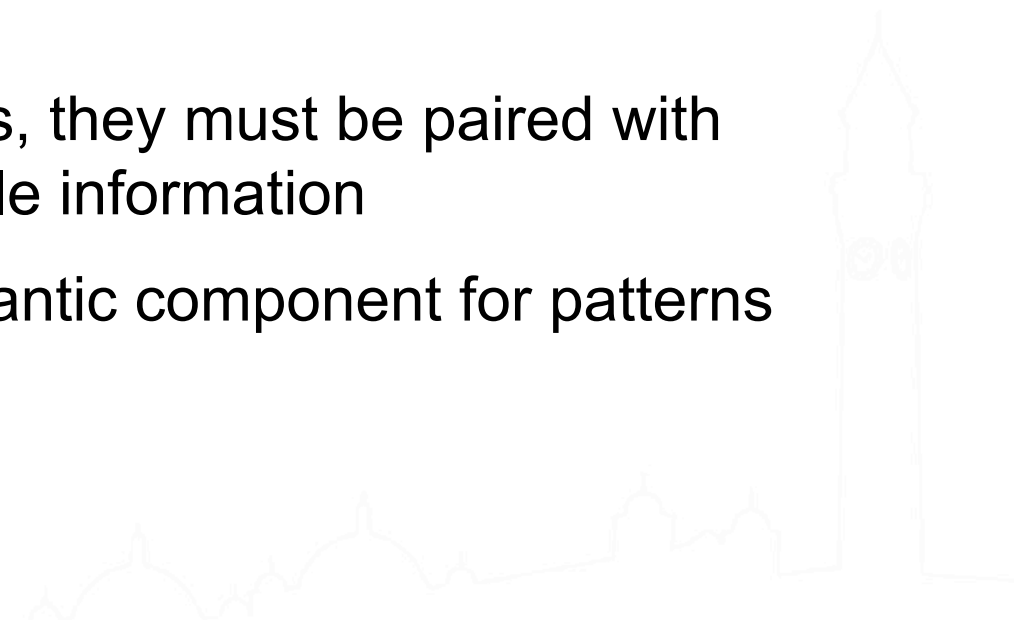
- Aims to describe the lexicon of English in terms of **semantic frames**

- Frames describe basic scenarios or situations that underlie word meanings

- Contain actors and props, called **frame elements** (FEs)

- Often viewed as the semantic component of CxG
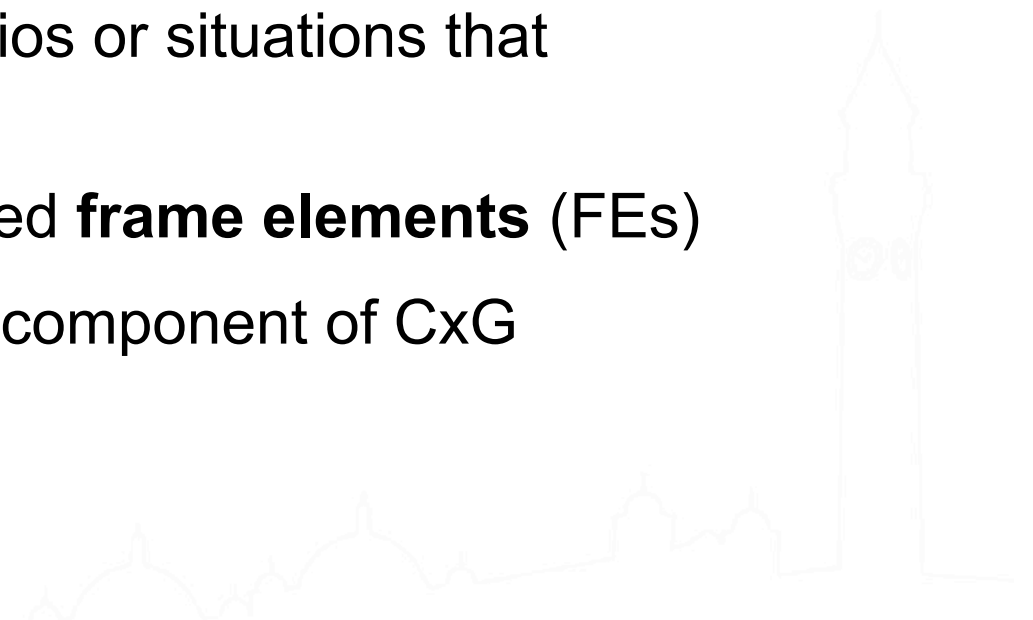
Lending

Definition:

The [Lender] gives the [Theme] to the [Borrower] with the expectation that the [Borrower] will return the [Theme] to the [Lender] after a [Duration] of time. This frame differs from the Borrowing frame in that this frame profiles the Lender in active sentences, whereas the Borrowing frame profiles the Borrower.

For the time being, this frame includes both the transfer event itself and the end state of the event, as shown in the first two examples following:

[I] am **LENDING** [a student of mine] [my book on German modal particles].

The student has not yet given back [the book] [I] **LENT** [to him].

Not only did I have to borrow a gown, but [I] was also **LENT** [a jacket]! [CNI]

[I] **LENT** [my girlfriend] [my car] [for the weekend].

[Jerry] **LOANED** [his skateboard] [to his little brother].

FEs:

Core:

|	
Borrower [Borr]	The person or institution who receives the [Theme] from the [Lender] for a [Duration]. I **LENT** the book [to him] for a whole month, and he still hasn't read it.
Lender [Lend]	The person or institution who gives the [Theme] to the [Borrower] for a [Duration]. [I] **LENT** the book to him for a whole month, and he still hasn't read it.
Theme [Th]	The object that is transferred from the [Lender] to the [Borrower] for a [Duration]. I **LENT** [the book] to him for a whole month, and he still hasn't read it.

FrameNet

☐ A word can belong to more than one frame

☐ Frame + lemma = Lexical Unit (LU)

☐ FrameNet also contains frame-to-frame relations

E.g., inheritance, perspective, use
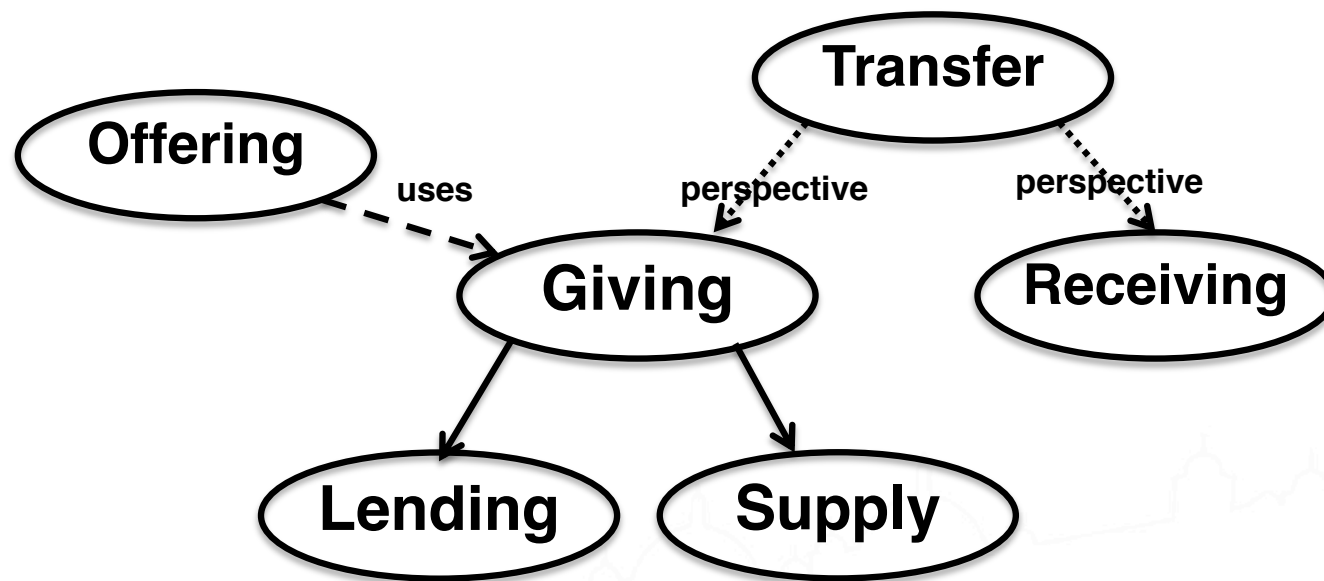
COBUILD + FrameNet

□ Proposal: match the verbs in the COBUILD patterns entries to FrameNet lexical units

- FrameNet includes valency information describing how frame elements are encoded in BNC corpus examples

- Semi-automatically matched with the COBUILD patterns

□ Potential to turn the patterns into a constructicon

- Form = pattern

- Meaning = generalization over frames used in the pattern

- Likely more than one construction for the same pattern
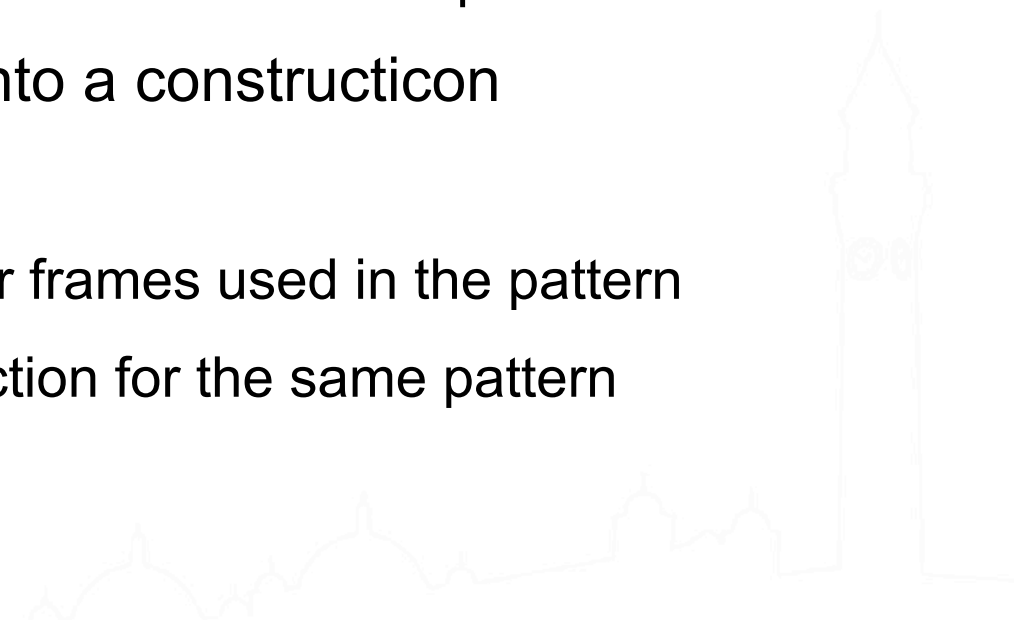
Method

□ Automatic procedure using the XML version of FrameNet and the COBUILD patterns (provided by HarperCollins)

□ Every verb listed in each pattern is looked up in FrameNet

    – If found, this returns one or more LUs

    – For each lexical unit, the annotated examples are consulted (if any)

    – If the valency realization of the frame elements matches the pattern, the LU is mapped onto the COBUILD entry

   – NB: only core frame elements are considered

Method

☐ Phrasal verbs were ignored

☐ Some patterns could not be matched to FrameNet

   – Patterns with 'dummy' *it*

     e.g., V *it* adj that

   – Missing grammatical distinctions in FrameNet

     e.g., V n-pl   (NP number not coded in FrameNet)

☐ 78 patterns matched to FrameNet
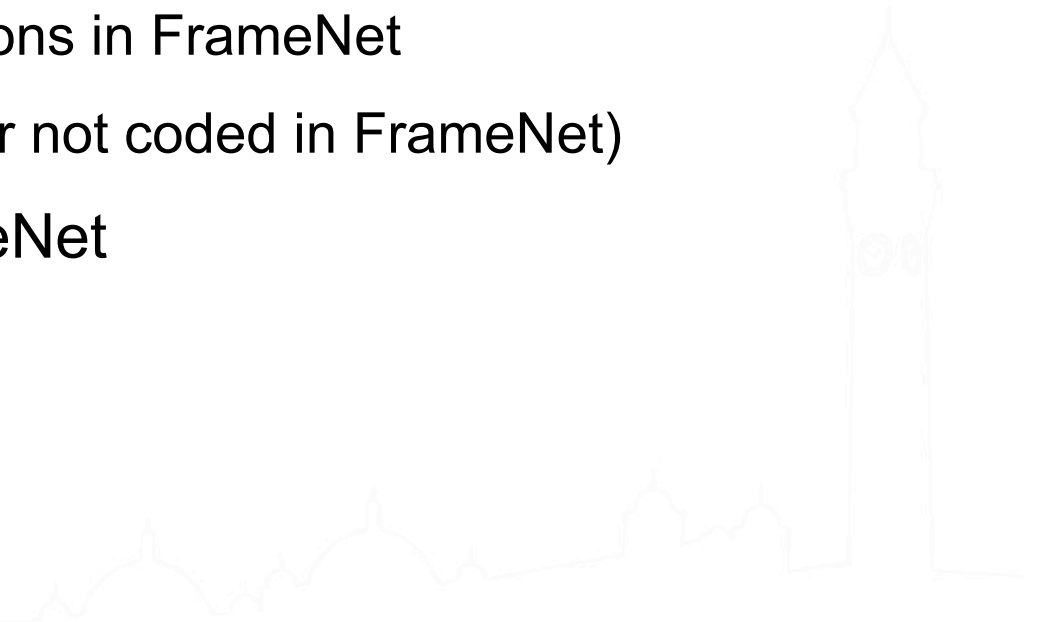
Results

Only **40.5%** of the entries in the COBUILD verb patterns matched to at least one LU in FrameNet (3063 out of 7572)
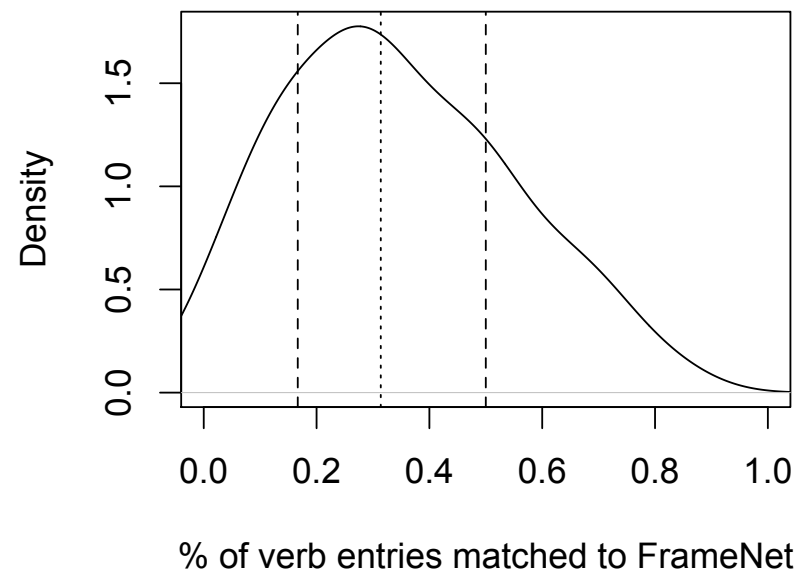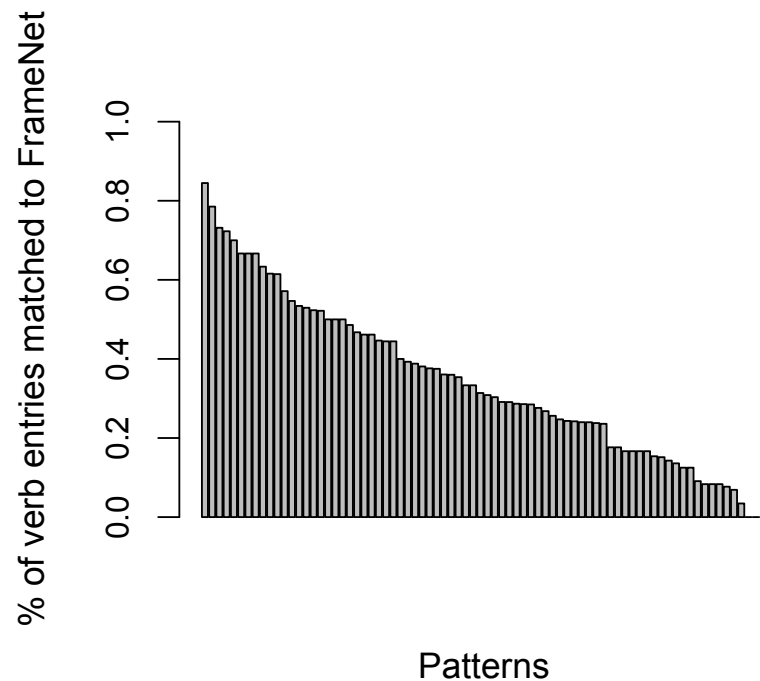

Only about 25% patterns have 50% or more matches

50% have between 17 and 50% matches

25% have less than 17% matches

Results

- ☐ Still insufficient coverage in FrameNet

- ☐ Problems with non-core frame elements

  - – E.g., Addressee for Communication, Explanation for Death

  - – Prevents these frames from being matched to "V n to n" and "V of" (for instance)

- ☐ Annotation errors and inconsistencies

- ☐ Matching the patterns to FrameNet will necessitate a lot of manual intervention
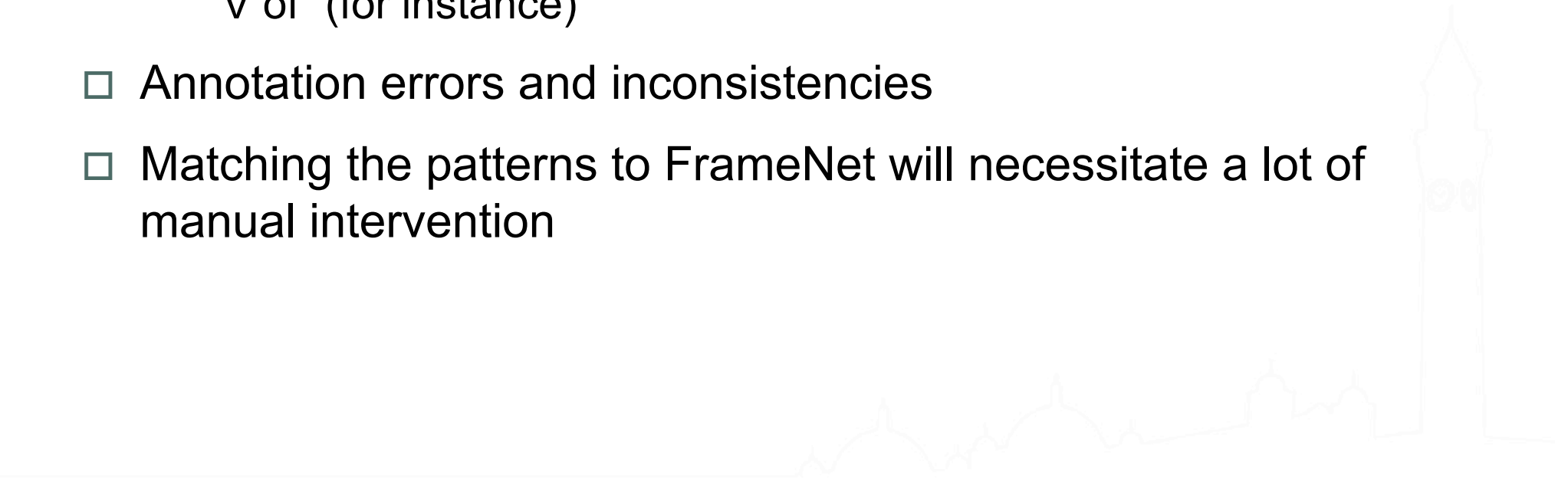
From patterns to frames

- What frames do we get when we look at a particular pattern?

- How are they related?

- How can we use this information to describe the pattern in terms of constructions?
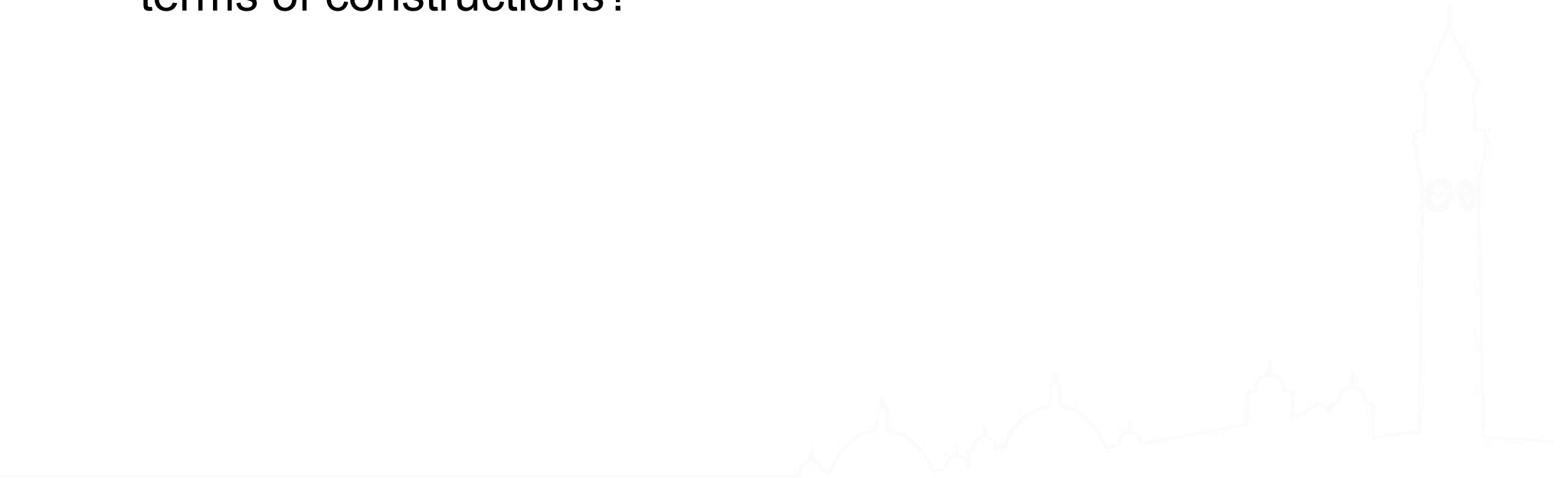
From patterns to frames

- ☐ Example: "V that"

- ☐ 255 verbs (w/o phrasal verbs)

- ☐ 10 meaning groups, for instance:
  - – The 'say' group: *claim, complain, insist, report, say, …*
  - – The 'think' group: *assume, know, think, understand, …*
  - – The 'show' group: *confirm, demonstrate, reveal, show, …*

- ☐ 62% were matched to at least one lexical unit

- ☐ Further annotation work was carried out to provide a better picture

From patterns to frames

- ☐ Lexical Units in the "V that" pattern form different "clouds" of frames (with frame-to-frame relations)

- ☐ One cloud ≈ one construction

- ☐ Usually more than one meaning group per cloud

- ☐ One meaning group ≠ one construction
  (*contra* Hunston & Su 2017)

- ☐ Two examples: communication, mental activity/emotions

A tight network: the 'say', 'add', and 'scream' groups (172 LUs)
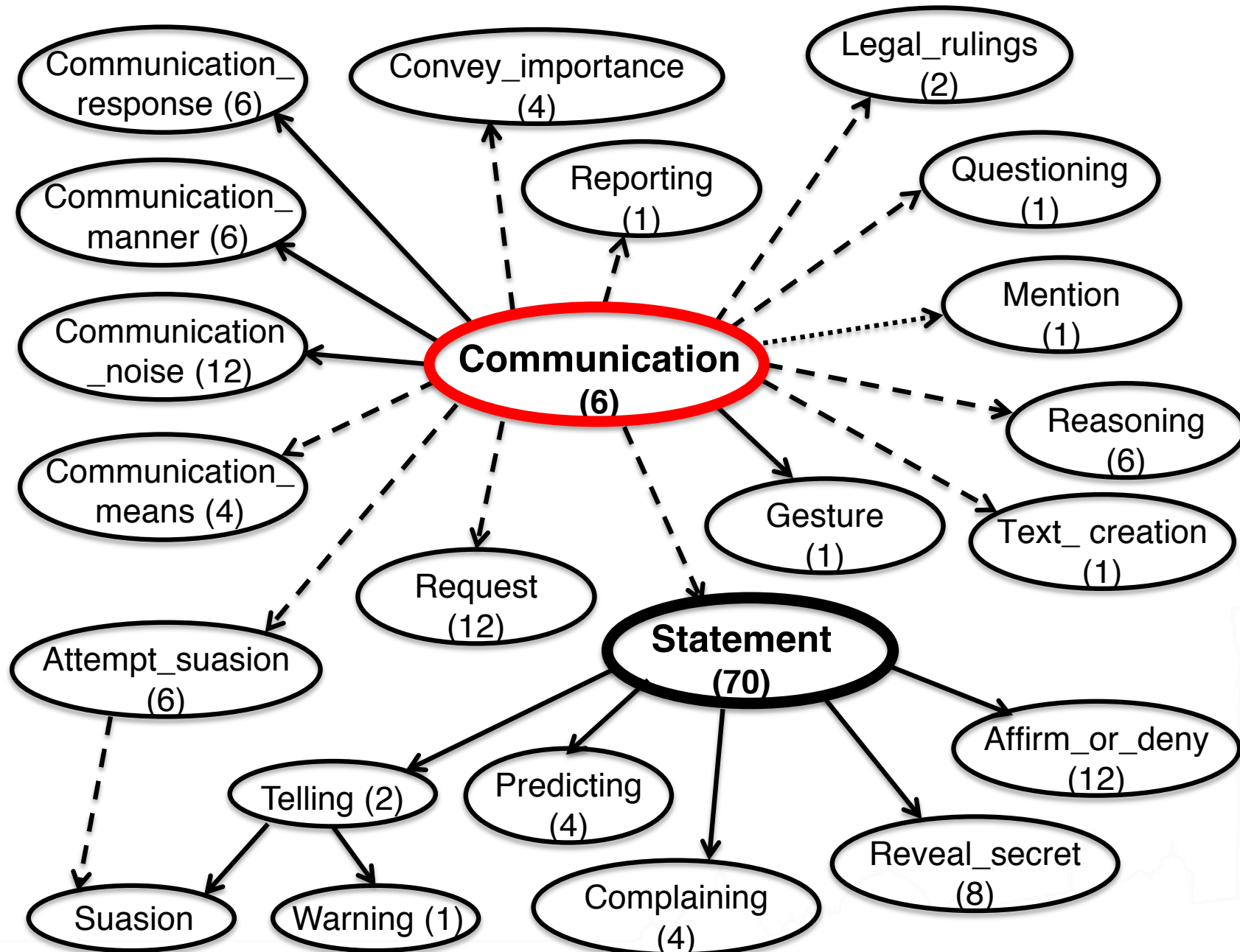
The "V that" Communication construction

- ❑ Communication frame
  - – The one frame that unifies them all
  - – Can be seen as the 'schema' shared by all uses
- ❑ Many subconstructions: different uses or forms of communication: make a statement, request, persuade, etc.
- ❑ Statement frame (verbal communication to make a claim)
  - – The most typical use: 70 LUs (101 with subframes)
 - – Can be seen as prototype, or 'core' constructional meaning
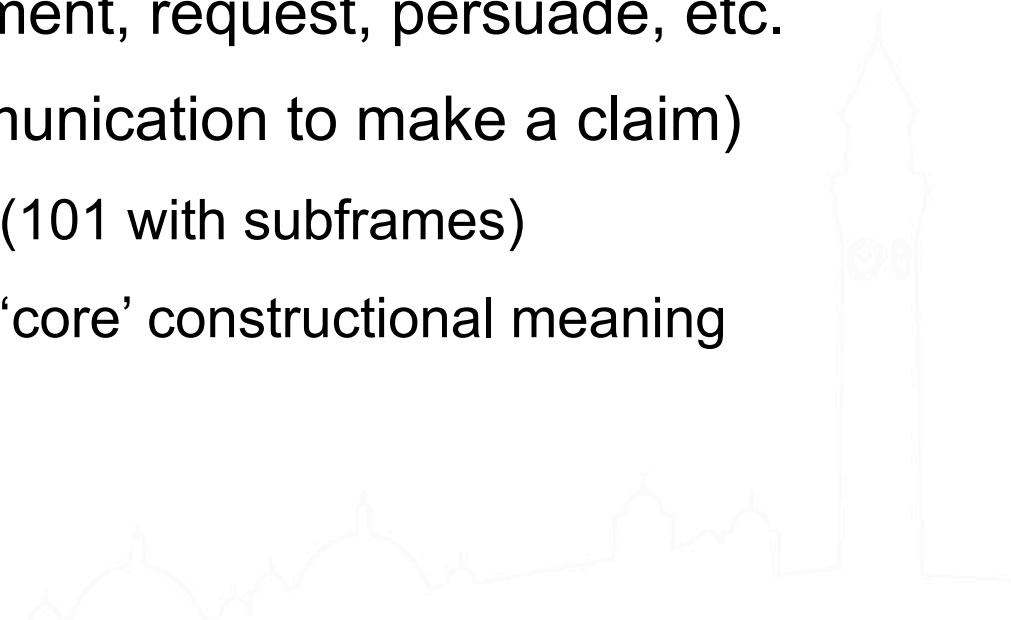
A looser network: the 'think', 'discover', and 'love' and 'hate groups (110 LUs)
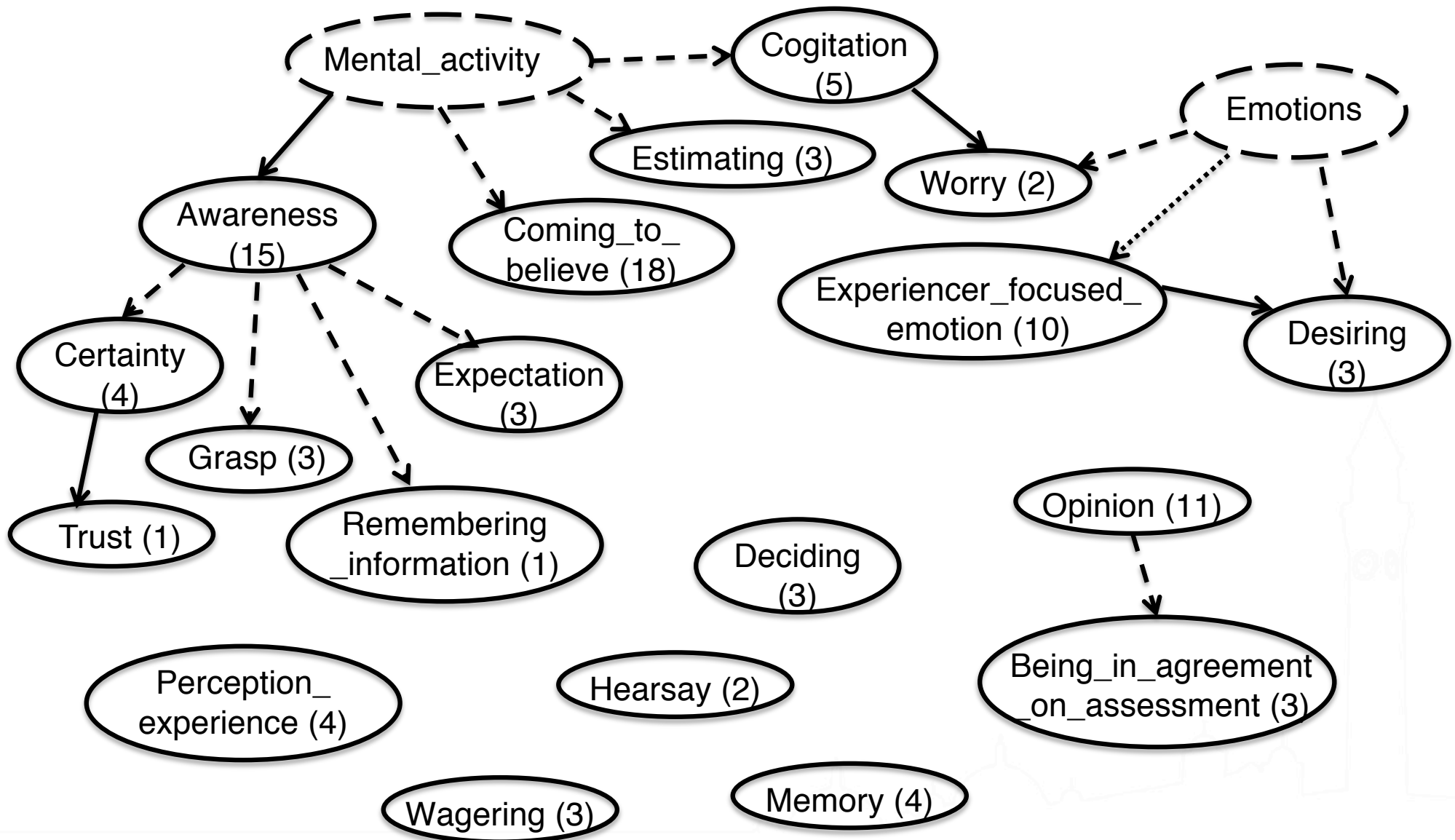
The "V that" Mental_activity & Emotions construction(s)

- ☐ Two partially overlapping networks centered on Mental_activity and Emotions

- ☐ A lot of orphans: Deciding, Memory, Opinion, …

- ☐ Highlights frame relations that are not recorded in FN

- ☐ Awareness (*know*), Opinion (*believe*), Experiencer_focused_emotion (*fear*), and Coming_to_believe (*realize*) among the most prominent

- ☐ Cluster of related constructions rather than single generalization
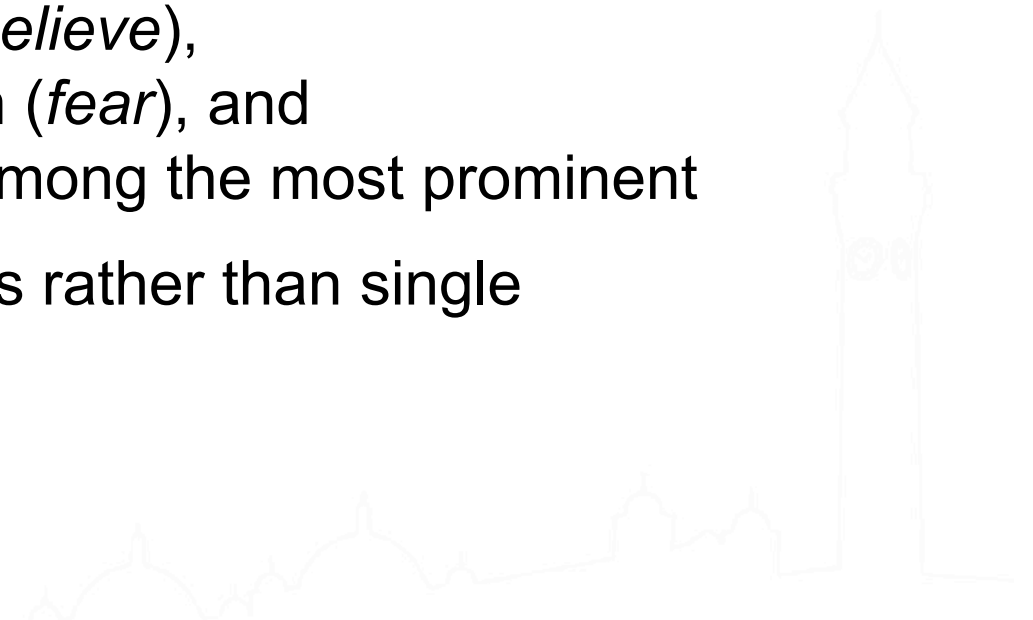
Summary

- The COBUILD Grammar Patterns and FrameNet complement each other well

- Frames can be used to turn patterns into constructions

- A lot of manual processing still necessary to merge the two resources

UNIVERSITY OF BIRMINGHAM | COLLEGE OF ARTS AND LAW

Thanks for your attention!

f.b.perek@bham.ac.uk

www.fperek.net