Identifying aboutgrams in engineering texts

Introduction

- n-grams, bundles, chunks, clusters
- “idiom principle”
- Phraseological studies - concgraming
- Clusters -> as an alternative to KW -> phraseology

Concgrams

- N-grams – contiguous
  - e.g. *take part*
- Skipgrams
  - AB, A*B e.g. *take no part, take an active part*
  - Not positional variation (AB, BA)
- “phrase-frames”
- Concgrams
  - Irrespective of any constituency and/or positional variation
  - Starting point for quantifying the extend of phraseology

Concgrams

1 (or computer) modelling of a building, framing **design**, **structural** analysis, component **design**, design
2 rocess framing plan (d) represents the final **design** of **structural** framing. One may further carry out
3 and cost-effectiveness. The **structural** **design** of a tall building involves several rather
4 instance, considering the preliminary **structural** **design** of a building, after the structure model is
5 engineers to achieve not only a safe **structural** **design**, but also a cost-effective design in terns of
6 has actually been applied to the **structural** **design** of more than 25 building projects with the
7 this model, one may perform **structural** framing **design** with assigned initial dimensions for all the
8 objectives of the preliminary **structural** framing **design** are: 1) To find which partitions are efficie
9 reduced occupied **structural** space, and shorter **design** time, have been realised. Acknowledgements T
10 **structural** analysis, optimisation, automated **design** check, and cost analysis, one may easily

Figure 1. Sample concordances for the two-word concgram *design/structural*

Aboutgrams

- Aboutness is a product of the global patternings of a text = “macrostructure”
- “aboutgrams” = word associations which are specific to a text
- Aboutness
  - Highest freq. occuring lexical phrases comprise provisional aboutgram list
  - Equally or more freq. in corpus are removed
  - Specialized (HKEC) and general corpus (BNC)

Analysis of data

Table 1.4 Most frequent lexical words in Article A

Lexical word	Frequency
design(s)	133
structural	116
model(s)	64
building(s)	63
architectural	34
optimisation	32
CAD	29
member(s)	29
analysis	28
layer	22

Table 2.5 Most frequent two-word aboutgrams in Article A

Aboutgram	Article A	HKEC	BNC
design(s)/structural	37	71	13
structural model(s)	34	38	15
building(s)/design	26	70	115
architectural/model(s)	14	14	2
structural analysis	12	16	53
design/tall	11	12	2
data capture	11	16	28
structural optimisation	10	14	0
analysis/design	9	20	101
form/structural	8	9	22

Analysis of data

- Intercollocation of collocates
  - Establish unique words (i.e. “*types*”)
  - Search for word associated with them -> 2 word concgrams
  - 2 word concgrams then become new origin
  - This process disambiguates words and a group gives a strong sense of the content, scope and argument of the document (*design [of a] [for] tall building(s)* – 11 instances)

Conclusion

- Meaning arise from words in particular combination
- Not rely entirely on KW
  - Aboutness, phraseology is not utilized
- Create tentative aboutgrams
  - Confirmed with reference to both, spec. and general reference corpus
- Human intervention in creating concgrams
- Can be used in learning and teaching context with students of English to raise their awareness of the centrality of phraseology