

# Negative Polarity Items

## Corpus Linguistics, Semantics, and Psycholinguistics

Day 3: Corpus Linguistics: Quantitative Research

Frank Richter<sup>1</sup>, Janina Radó<sup>1</sup>, Manfred Sailer<sup>2</sup>

<sup>1</sup>Universität Tübingen

<sup>2</sup>Universität Göttingen

ESLLI 2008, Hamburg

# Summary of day 2

Qualitative research is useful to

- falsify empirical claims on the distribution of NPIs.
- find natural occurrences of NPIs in theoretically interesting constructions
- compile qualitative profiles of NPIs
- systematic documentation of NPIs such as CoDII

# Corpus-based, qualitative-quantitative research on NPIs

- since frequency data are given in the corpus, we can use them!
- some corpora provide automatic collocation analysis.
- Application:
  - ▶ items with similar qualitative profiles may show distinct quantitative profiles (different types of NPIs, quasi-NPIs)
  - ▶ historical development (grammaticalization) of NPIs.
  - ▶ for frequent items: low percentage of occurrence can be taken as weak negative evidence.

# Quantitative profile of *budge* (Sinclair, 2004)

- collocational measures

- ▶ taken from idiom research
- ▶ percentage of occurrence in an NPI-licensing environment

item ( <i>N</i> )	clausem. <i>not</i> CM	n-const CN	non-cl. neg. nCM	neg.pred. NegV	others Exception
<i>budge</i> (23)	10 (43%)	2 (9%)	1 (4%)	9 (39%)	1 (4%)

- Compare the collocational behaviour of synonyms  
*budge* vs. *move*; *care/VERB* vs. *take care*

# Comparing quantitative profiles of similar words (Hoeksema, 1994)

Distributional profiles of four English verbs of indifference:

TABLE 2. FOUR VERBS OF INDIFFERENCE: DISTRIBUTION

Environments	CARE N=792	MATTER N=406	MIND N=341	BOTHER N=377
negative	53%	57%	72%	35%
other neg	12%	7%	20%	7%
interrogative	15%	13%	7%	11%
affirmative	20%	20%	1%	48%

# Historical development (Hoeksema, 1999)

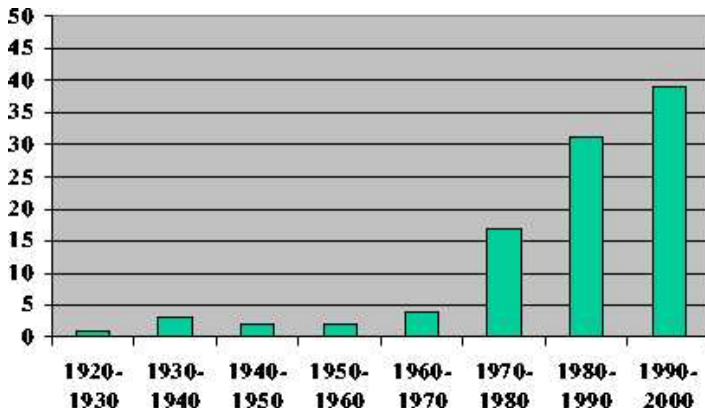
## Method:

- corpora from different historical stages
- pick relatively frequent item
- classify all occurrences according to their NPI licensing potential.
- calculate distribution (relative occurrence) in contexts over time.

## Historical development (Hoeksema, 1999)

Result: Dutch *ooit* (*ever*) turned from an NPI into a less-NPI like element over time.

Percentage of on-NPI uses of *ooit*



# Evaluation

- Method shows fine differences among NPIs, synonyms, . . .
- Preferences to NPI-licensing contexts can change without a change in meaning.
- Enough data is required; method cannot be applied to low-frequency items
- Do usage differences reflect grammatical differences?
- Most theories cannot handle quantitative data.



# Evaluation

- Method shows fine differences among NPIs, synonyms, . . .
- Preferences to NPI-licensing contexts can change without a change in meaning.
- Enough data is required; method cannot be applied to low-frequency items
- Do usage differences reflect grammatical differences?
- Most theories cannot handle quantitative data.

# Types of NPIs

- Strong NPIs:

- ▶ minimizers (*a wink, a drop, ...*)
- ▶ ok with *not*, n-word  
out with *few, not every*
- ▶ ok with *I don't think*  
out with *I don't claim*

- Weak NPIs:

- ▶ *any, ever, ...*
- ▶ ok with *not*, n-word  
ok with *few, not every*
- ▶ ok with *I don't think*  
ok with *I don't claim*

- Verbal NPIs:

- ▶ *need* (also German and Dutch equivalents), *care, bother, ...*
- ▶ ok with *not*, n-word  
ok with *few, not every*
- ▶ ok with *I don't think*  
out with *I don't claim*, inside universally quantified NPs

# Using the web as a corpus for NPI research

- Good chances to find occurrences of most NPIs.
- useful for qualitative research
- quantitative research: Is an item an NPI?
  - ▶ compare: hits for the item alone with hits for the item with negation
  - ▶ *inkling*: approx. 2,800,000 hits  
not *inkling*: approx. 2,300,000 hits
  - ▶ *candle*: approx. 63,700,000 hits  
not *candle*: approx. 546,000 hits

# Limitations of corpus approaches (Hoeksema, 1997)

- polysemy and ambiguity
  - ▶ the verb *matter* is an NPI, the noun is not.
  - ▶ *bother* (= *annoy*) is an NPI, *bother* (= *take the trouble*) is not
  - ▶ German *brauche* (*need*) is an NPI if it selects an infinitival complement, but not as a transitive verb.
- reading-dependent licensing
  - ▶ You say anything, and I'll kill you.
  - ▶ \* You said anything, and I killed you.
- inherently negative predicates (circularity!)
  - ▶ The proposal lacks any plausibility.
  - ▶ It was impossible to ever return.

# Back to the four questions

- Licenser: indirect evidence for licensing potential; hard to find constructional licensors.
- Licensee: categorization based on distributional profiles
- Relation: qualitative testing of individual constellations
- Status: attested/non-attested, frequency

# Corpus-driven research

- Automatic extraction of NPIs
- Are Hoeksema's reservations still valid?

Hoeksema, Jack (1994). On the grammaticalization of negative polarity items. In *Proceedings of the 20th Meeting of the Berkeley Linguistic Society*, pp. 273–282.

Hoeksema, Jack (1997). Corpus Study of Negative Polarity Items. Html version of a paper which appeared in the *IV-V Jornades de corpus linguistics 1996-1997*, Universitat Pompeu Fabre, Barcelona. URL:

<http://odur.let.rug.nl/~hoeksema/docs/barcelona.html>

Hoeksema, Jack (1999). Aantekeningen bij *ooit*, deel 2: de opkomst van niet-polaier *ooit*. *TABU*.

Sinclair, John (2004). *Trust the Text. Language, Corpus and Discourse*. London and New York: Routledge.