Working out Statistical Profiles of Negative Polarity Items

Timm Lichte

Methods and Statistics for linguistic Research

Rijksuniversiteit Groningen, 2003-2004

23.04.2004

Outline

- Introduction:
  - Polarity Items
  - Degrees of Negation - Classes of NPI's
- Extraction of NPI candidates
- Classification of NPI's

Introduction: Polarity Items

Negative Polarity Item (NPI):

- He hasn't seen any student.

  * He has seen any student.

- Niemand von uns war jemals im Jemen.

  * Jeder von uns war jemals im Jemen.

Introduction: Polarity Items

Negative Polarity Item (NPI):

- Niemand von uns war jemals im Jemen.

  * Jeder von uns war jemals im Jemen.

Positive Polarity Item (PPI):

- Bill has already arrived in Munich.

  * Bill has not already arrived in Munich.

- Bill ist bereits in München angekommen.

  * Bill ist nicht bereits in München angekommen.

Introduction: Polarity Items

- A NPI must be within the scope of a trigger for negation.
- A PPI must **not** be within the scope of a trigger for negation.

$\Longrightarrow$ Collocational behaviour

State of documentation

- Rich linguistical literature (formal semantics)
- No systematic documentation of PI's in German (Welte 75,Kürschner 83)

Degrees of negation

- Negation (antimorphic contexts):
  - $f(X \cap Y) = f(X) \cup f(Y)$
    $f(X \cup Y) = f(X) \cap f(Y)$
  - Deutschland habe nicht die Absicht, jemals Belgien oder Frankreich anzugreifen.
  - Mehrmals war es ihr schon so ergangen, ohne dass sie . . . jemals davon gesprochen hätte.
  - Aus diesem Dokument kann man keinesfalls ersehen, . . . daß sie es jemals gesehen hat.

Degrees of negation

- Negation (antimorphic c.): *nicht*, *ohne*
- N-words (anti-additive contexts):
  - $f(X \cup Y) = f(X) \cap f(Y)$
  - . . . und niemand von der CDU/CSU-Fraktion hat jemals erklärt, daß diese Verträge nicht existent seien.
  - Nein, diese Gedichte darf kein Fremder jemals erblicken.

Degrees of negation

- Negation (antimorphic c.): *nicht*, *ohne*
- N-words (anti-additive c.): *niemand*, *kein-*
- Monotone decreasing contexts:
  - $f(X \subseteq Y) \rightarrow f(Y) \subseteq f(X)$
  - Sehr wenige Männer oder Frauen haben jemals die Gelegenheit, so viel Gutes zu tun wie James Grant.
  - In seiner Heimat sei kaum jemals ein großer dramatischer Stil entstanden, . . .
  - Aber wie jeder weiß, der jemals einem hölzernen Kuckuck gelauscht hat, . . .

Degrees of negation

- Negation (antimorphic c.): *nicht*, *ohne*
- N-words (anti-additive c.): *niemand*, *kein-*
- Monotone decreasing c.: *wenige*, *kaum*
- Further contexts

Further contexts for *jemals*

- Adversative predicates:

  Bis zum vergangenen Sonntag weigerte sich sein Vater vehement,

  jemals CDU zu wählen.

Further contexts for *jemals*

- Adversative predicates: *s. weigern*
- *if*-clauses:

  wenn ich jemals einen Anflug von Moralität . . . verspürte, jetzt ist jeder Gedanke daran verschwunden.
- Questions:

  Hat Cäsar jemals germanische Kohorten gehabt?

Further contexts for *jemals*

- Adversative predicates: *s. weigern*
- *if*-clauses, questions
- Comparative/superlative degrees:
  - er war so in Angst, dass er sich toter stellte als jemals vorher.
  - Es soll die größte . . . Ausstellung werden, die jemals abgehalten wurde.
  - er ist zu weit fort, um ihn jemals zu finden.

Further contexts for *jemals*

- Adversative predicates: *s. weigern*
- *if*-clauses, questions
- Comparative/superlative degrees
- *nur*:
  - In ganz Schleswig-Hostein gibt es nur einen Mann, der jemals die Meisterprüfung abgelegt hat.
  - Nur ein Drittel von ihnen nimmt jemals psychiatrische Hilfe in Anspruch.

Further contexts for *jemals*

- Adversative predicates: *s. weigern*
- *if*-clauses, questions
- Comparative/superlative degrees
- *nur*
- Certain temporal conjunctions:
  - Er hatte . . . den Titel geholt, noch bevor er jemals australischer Meister war.

NPI classes

- Superstrong: only antimorphic c.
- Strong: antimorphic und anti-additive c.
- Weak: monotone decreasing c.
- Class of the other contexts unclear

Superstrong NPI's

- Example: *mals* (ndl.)
- Distribution:
  - Antimorphic (AM):

    Zijn oordelen waren vaak niet mals.
  - Anit-additive (AA):

    * Niet één van zijn oordelen was mals.
  - Monotone decreasing (DE):

    * Weinig van zijn oordelen waren mals.
- German: no superstrong NPI?

Strong NPI's

- Example: *sonderlich*, *beileibe*, *auch nur irgendw-*
- Distribution:
  - AM: Der Lehrer war nicht sonderlich erfreut.
  - AA: Kein einziger Lehrer ist sonderlich erfreut gewesen.
  - Monoton decreasing:

    * Nur wenige Kaufleute sind sonderlich zufrieden gewesen.
  - Further contexts:

    * Sind die Kaufleute sonderlich zufrieden gewesen?

Weak NPI's

- Example: *brauchen+zu*-Inf., *ausstehen können*, *ein Auge zutun*, *viel davon halten*
- Distribution:
  - Monoton decreasing:

    Höchstens eine Frau wird sich zu verantworten brauchen.

Summary

- Classification of licensing contexts and of NPIs available.
- No systematic documentation of the NPI inventory for German.

Extraction of NPI candidates

- Corpus: TüPP-D/Z (Tübingen Partially Parsed Corpus of Written German)
- taz, 2.9.’86–7.5.’99; ca. 200 mio words.
- Annotation: lemmatisation, part of speech, morphology, clause boundaries, topological fields, chunks.

Extraction of NPI candidates

- Corpus: TüPP-D/Z (Tübingen Partially Parsed Corpus of Written German)
- taz, 2.9.’86–7.5.’99; ca. 200 mio words.
- Annotation: lemmatisation, part of speech, morphology, clause boundaries, topological fields, chunks.
- Cutout: 1998 (ca. 1,2 mio sentences)
  - Context marking
  - Extraction of lemmata
  - Extraction of candidates

Context marking

- Based on part of speech and lemmatisation.
- Sentence-wise context marking:

NEG	129,027	11.0%	*nich(t)*
AM	133,338	11.4%	*weder*, *, ohne*
AA	176,540	15.1%	*niemand*, *keiner*, . . .
DE	245,464	21.0%	*wenige*, *selten*, *bevor*
DE?	308,156	26.3%	*ob*, *wer*, *?*
total	1,170,618		

Extraction of lemmata

- Lemmata with an occurrence of min. 50 : 17,157

context	#	ratio
NEG	2,065,400	14.3%
AM	2,145,195	14.9%
AA	2,765,828	19.1%
DE	3,910,076	27.1%
DE?	4,681,993	32.3%
total	14,445,709	

Extraction of candidates

- z-scores of frequency-ratios
  - frequency-ratio: $\frac{NegContext(w)}{Total(w)}$
  - $z = \frac{x-\mu}{sd}$

    (Deviation of observation and mean, scaled by the standard deviation)

Extraction of candidates

- z-scores of frequency-ratios
- Significance level at 1.96 (p=0.05)?
  - Distribution not normal, but skewed to the right.

	Minimum	Maximum	Mean	Std.Dev.	$\geq$ 1.96
AM	-2.14607	12.35070	.0000000	1.00000000	554
DE	-2.78318	8.31527	.0000000	1.00000000	596

Extraction of candidates

- z-scores of frequency-ratios
- Significance level at 1.96 (p=0.05)?
  - Distribution not normal, but skewed to the right.
  - low z-values of NPI's beileibe and Hehl:

	NEG	AM	AA	DE	DE?
lemma	beileibe	beileibe	Hehl	Hehl	Hehl
z	8.2	7.9	6.8	4.9	3.9

$\Longrightarrow$ higher significance level?

Extraction of candidates

- z-scores of frequency-ratios
- Significance level at 1.96 (p=0.05)?
- Considering the ranking by NEG z-score:

> *hinwegtäuschen, hinnehmbar, verhehlen, notwendigerweise, sondern, sonderlich, zimperlich, antasten, wahrhaben, zurückschrecken, beileibe*

$\Longrightarrow$ all reasonable candidates for NPI's !

Application of associative measures

- Mutual Information:
  - $I(w_1, w_2) = \log_2 \frac{P(w_1,w_2)}{P(w_1)P(w_2)}$
  - Ranking by I-values for NEG:

> *hinwegtäuschen, verhehlen, hinnehmbar, sondern, notwendigerweise, sonderlich, zimperlich, antasten, wahrhaben, zurückschrecken, durchsetzbar*

$\Longrightarrow$ just slightly different from ranking by NEG z-scores!

Application of associative measures

- Mutual Information
- Log-likelihood score:
  - H_1: $P(w|c) = P(w|\neg c)$
    H_2: $P(w|c) \neq P(w|\neg c)$
  - Ranking by LL-score for NEG:

    verhehlen, hinnehmbar, fechten, unerheblich, verderben, zurückschrecken, zimperlich, durchsetzbar, antasten, wahrhaben, unumstritten

    $\Longrightarrow$ quite reasonable NPI candidates!

Application of associative measures

- Mutual Information
- Log-likelihood score
- Pearson's χ^2 test:
  - $$\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$
  - Ranking by χ^2-score for NEG:

    sondern, mehr, sein, die, gar, können, daß, aber, der, es, auch

    $\Longrightarrow$ no reasonable NPI candidates!

    $\Longrightarrow$ scale by occurrence of lemma!

Application of associative measures

- Mutual Information
- Log-likelihood score
- Pearson's χ^2 test
  - $$\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$
  - Ranking by χ^2-score for NEG, scaled by occurrence of lemma:

    hinwegtäuschen, hinnehmbar, verhehlen, notwendigerweise, sondern, sonderlich, zimperlich, antasten, wahrhaben, zurückschrecken, beileibe

$\Longrightarrow$ reasonable NPI candidates!

Extraction of NPI's: Summary

- Simple z-score of frequecy-ratios seems to suffice.
- List of NPI candidates looks promising.
- But:
  - Lists of candidates without compound NPI's (*wahrhaben wollen*)
  - Class definitions of negation triggers are holey (*jemals* at place 5196)
  - Small corpus

Classification of NPI's

How can we decide the subclass of a NPI (superstrong, strong, weak)?

Classification of NPI's

How can we decide the subclass of a NPI (superstrong, strong, weak)?

- NEG $\subseteq$ AM $\subseteq$ AA $\subseteq$ DE $\subseteq$ DE?
- In general we expect increasing frequencies along the enlargement of the contexts of negation.
- Find: Mean and standard deviation of increase.
- Find: z-score for steps of increase for every Lemma.

Degree of increase

	NEG→AM	AM→AA	AA→DE	DE→DE?
mean	5.6	43	79	53
sd	8	28	39	30
z max	24.2	23.9	14.4	20.9
z min	-0.72	-1.56	-2.1	-1.8

Degree of increase

	NEG⟶AM	AM⟶AA	AA⟶DE	DE⟶DE?
mean	5.6	43	79	53
sd	8	28	39	30
z max	24.2	23.9	14.4	20.9
z min	-0.72	-1.56	-2.1	-1.8

H1 Superstrong NPI's: Low z-scores for AM⟶AA, AA⟶DE and DE⟶DE?.

Degree of increase

	NEG→AM	AM→AA	AA→DE	DE→DE?
mean	5.6	43	79	53
sd	8	28	39	30
z max	24.2	23.9	14.4	20.9
z min	-0.72	-1.56	-2.1	-1.8

H2 Strong NPI's: Only low z-scores for AA→DE and DE→DE?.

Degree of increase

	NEG→AM	AM→AA	AA→DE	DE→DE?
mean	5.6	43	79	53
sd	8	28	39	30
z max	24.2	23.9	14.4	20.9
z min	-0.72	-1.56	-2.1	-1.8

H3 Weak NPI's: z-scores at steps arbitrary.

Degree of increase

	NEG→AM	AM→AA	AA→DE	DE→DE?
mean	5.6	43	79	53
sd	8	28	39	30
z max	24.2	23.9	14.4	20.9
z min	-0.72	-1.56	-2.1	-1.8

H4 Nominal NPI's: extreme high z-scores at AM→AA.

H1: superstrong NPI's

	NEG→AM	AM→AA	AA→DE	DE→DE?
hinwegtäuschen	0.68	-0.77	-1.49	-1.40
hinnehmbar	-0.71	-0.96	-2.01	-0.68
zimperlich	1.80	-0.85	-1.04	-0.47
verkneifen	-0.71	-1.56	0.33	-1.33

H1: superstrong NPI's

	NEG→AM	AM→AA	AA→DE	DE→DE?
hinwegtäuschen	0.68	-0.77	-1.49	-1.40
hinnehmbar	-0.71	-0.96	-2.01	-0.68
zimperlich	1.80	-0.85	-1.04	-0.47
verkneifen	-0.71	-1.56	0.33	-1.33

$\Longrightarrow$ All candidates are weak NPI's.

H2: strong NPI's

	NEG→AM	AM→AA	AA→DE	DE→DE?
beileibe	-0.71	5.42	-2.06	-1.76
sondern	1.32	1.40	-1.82	-1.63
sonderlich	1.87	1.72	-1.66	-1.60
verhehlen	3.0	0.54	-0.93	-1.26

H2: strong NPI's

	NEG→AM	AM→AA	AA→DE	DE→DE?
beileibe	-0.71	5.42	-2.06	-1.76
sondern	1.32	1.40	-1.82	-1.63
sonderlich	1.87	1.72	-1.66	-1.60
verhehlen	3.0	0.54	-0.93	-1.26

$\Longrightarrow$ Candidates seem to be strong NPI's.

H3: weak NPI's

	NEG→AM	AM→AA	AA→DE	DE→DE?
scheren	5.01	3.32	4.27	0.99
unerheblich	0.45	-0.58	-0.42	2.10

H3: weak NPI's

	NEG$\rightarrow$AM	AM$\rightarrow$AA	AA$\rightarrow$DE	DE$\rightarrow$DE?
scheren	5.01	3.32	4.27	0.99
unerheblich	0.45	-0.58	-0.42	2.10

$\Longrightarrow$ Weak NPI's show diverse patterns of increase.

H4: nominal NPEs

	NEG→AM	AM→AA	AA→DE	DE→DE?
Seltenheit	-0.71	23.91	-2.10	-1.76
Ahnung	0.31	18.13	-0.26	-1.05
Hehl	0.21	19.70	-1.68	-1.53
Handhabe	1.75	16.59	-1.06	-1.13

H4: nominal NPEs

	NEG→AM	AM→AA	AA→DE	DE→DE?
Seltenheit	-0.71	23.91	-2.10	-1.76
Ahnung	0.31	18.13	-0.26	-1.05
Hehl	0.21	19.70	-1.68	-1.53
Handhabe	1.75	16.59	-1.06	-1.13

$\Longrightarrow$ Negative cohesion is responsible for this pattern of increase.

Classification of NPI's: Summary

- Promising method for strong and nominal NPI's ...
- but again 'just' candidates.
- No superstrong NPI found.
- In preparation: Application of the mentioned associative measures.

Outlook

- Refinement of the classes of negation triggers
- Enlargement of corpus cutout
- Implementation of collocational modules to detect multi-word NPI's
- PPI's