Lecture 4: Empirical Approaches in Cognitive Linguistics: Methods of Corpus Analysis and Statistics

Yoshikata Shibuya
Kyoto University of Foreign Studies

November 12, 2012

Empirical methods in cognitive linguistics

- corpus analysis
- psycholinguistic experimentation
- statistics

Contents

1. Collostructional analysis
2. Case studies: Shibuya (2011, 2012)

Topic 1

Collostructional analysis

References

- Gries, Stefan Th. & Anatol Stefanowitsch. 2004a. Extending collostructional analysis: A corpus-based perspectives on 'alternations'. *International Journal of Corpus Linguistics* 9.1:97-129.
- Gries, Stefan Th. & Anatol Stefanowitsch. 2004b. Co-varying collexemes in the into-causative. In: Achard, Michel & Suzanne Kemmer (eds.). *Language, Culture, and Mind*. Stanford, CA: CSLI, p. 225-36.
- Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics* 8.2:209-43.
- Stefanowitsch, Anatol & Stefan Th. Gries. 2005. Co-varying collexemes. *Corpus Linguistics and Linguistic Theory* 1.1:1-43.

Introduction

➢ Collostructional analysis:

- By Anatol Stefanowitsch and Stefan Th. Gries (2003)
- An extension of collocational analysis
- Specifically geared to investigating the interaction of words and the grammatical constructions associated with them
- Framed in a construction-based approach to language (form-meaning pairs, lexicon-syntax continuum)

Introduction

- Two main points:

1. To increase the adequacy of grammatical description
2. To provide data for linguistic theory-building

Theoretical background

- Definition of *construction*:

A construction is . . . a pairing of form with meaning/use such that some aspect of the form or some aspect of the meaning/use is not strictly predictable from the component parts or from other constructions already established to exist in the language (Goldberg 1996: 68).

- Construcition is directly associated with a particular meaning/function.
- Its form/meaning cannot be compositionally derived.

Theoretical background

➢ The English ditransitive construction:

- The subcategorization frame: [S V Oi Od]

e.g. *John gave Mary a book*.

Theoretical background

- A “transfer” meaning

→ The referent of the subject transfers the referent of the direct object to the referent of the indirect object.

- The construction assigns the meaning to all expressions instantiating it.

→ Irrespective of the particular verb occurring there

Theoretical background

- *Pat hit Chris the ball.*

- *Hit* is a two-place verb.

→ Meaning: "(some part of) X comes into forceful contact with (some part of) Y"

- Nothing in its meaning points to a transfer of Y to some third participant

→ However, the sentence receives the interpretation "Pat transferred the ball to Chris by coming into forceful contact with it" (cf. Goldberg 1995:34–35).

Theoretical background

- The syntactic configuration [S V Oi Od] is directly associated with the meaning "X transfer Y to Z" (semantic roles Agent, Recipient, and Theme).

- The meaning is not strictly predictable from its components or from other constructions of English.

→ The ditransitive subcategorization frame must be seen as a construction.

Theoretical background

- Semantic compatibility: Construction Grammar

  A word may occur in a construction if it is *semantically compatible* with the meaning of the construction.

- The verb *give* may occur in the ditransitive construction because the verb and the construction have the same meaning ("sb transfers sth to sb").

- The verb *deprive:* not compatible

→ Almost an antonym of it.

Methodology

➢ Collostructional analysis:

- An extension of traditional collocational analysis

- It starts with a particular construction and investigates which lexemes are strongly attracted or repelled by a particular slot in the construction (i.e. occur more frequently or less frequently than expected).

Methodology

- Lexemes attracted to a particular construction are referred to as *collexemes* of this construction.

- A construction associated with a particular lexeme may be referred to as a *collostruct*.

- The combination of a collexeme and a collostruct is referred to as a *collostruction*.

Methodology

- Issue of a suitable measure of association:

- Problems with the measures of association strength proposed in the literature:

Methodology

1. Many of the proposed statistics involve distributional assumptions that are not justified.

- Normal distribution and homogeneity of variances are just two such assumptions which are hardly ever met when dealing with natural language data.

→ It renders suspicious any statistical method based on them (e.g. Berry-Rogghe's (1974) *z-score*, Church et al.'s (1991) *t-score).*

Methodology

2. Some statistics are particularly prone to strongly overestimating association strengths and/or underestimating the probability of error when extremely rare collocations are investigated (e.g. MI).

- They are unreliable given the kind of extremely sparse data frequently encountered in corpus-linguistic tasks .

Methodology

➢ The statistic that is not subject to such theoretical and/or distributional shortcomings:

- The Fisher exact test:

1. It does not make any distributional assumptions.
2. It does not require any particular sample size.

→ Hence, this test is adopted in calculating collostruction strength in collostructional analysis.

Methodology

- How to calculate the collostruction strength:

1. The frequency of L in C
2. The frequency of L in all other constructions
3. The frequency of C with lexemes other than L
4. The frequency of all other constructions with lexemes other than L

→ The frequencies are entered in a 4-by-4 table and submitted to the Fisher exact test.

Methodology

Table 3. Crosstabulation of *accident* and the [N *waiting to happen*] construction

	accident	¬accident	Row totals
[N *waiting to happen*]	*14*	21	*35*
¬[N *waiting to happen*]	8,606	10,197,659	10,206,265
Column totals	*8,620*	10,197,680	*10,206,300*

- The figures in italics are derived directly from the corpus data.
- The remaining ones result from subtractions.
- The total number of constructions was arrived at by counting the total number of verb tags in the BNC.

Methodology

Table 4. Collexemes most strongly attracted to the [N *waiting to happen*] construction[8]

Collexeme (n)	$p_{\text{Fisher exact}}$ (collostruction strength)
accident (14)	2.12E-34
disaster (12)	1.36E-33
welkom (1)	4.46E-05
earthquake (1)	2.46E-03
invasion (1)	7.10E-03
recovery (1)	1.32E-02
revolution (1)	1.68E-02
crisis (1)	2.21E-02
dream (1)	2.45E-02
it (sex) (1)	2.83E-02
event (1)	6.92E-02

Methodology

➢ Interpretation of the results:

- The p-value (2.1216E-34) indicates that the association between *accident* and the [N *waiting to happen*] construction is very strong.

Case study

- The ditransitive construction:

a) Mary gave John a book.
b) Chris promised Pat a car.
c) John told Mary a story.

- Corpus:
  The British component of the International Corpus of English (ICE-GB)

Case study

- Semantic extensions of the construction:
  The basic sense: “X causes Y to have/receive Z” (cf. Goldberg 1995: 38; Pinker 1989: 73).

a) Mary gave John a book. ← the basic sense

Case study

b) Chris promised Pat a car. ← extension

Linked to the basic sense by virtue of the fact that the satisfaction conditions of the speech-act verb *promise* imply a transfer.

c) John told Mary a story. ← extension

A metaphorical extension based on the idea that communication is the exchange of objects (cf. Reddy 1979).

cf. Goldberg (1995)

Case study

Table 9. The ditransitive construction: basic sense and extensions

Sense	Sample verbs
Basic sense:	
Agent causes recipient to receive theme	*give, pass, hand, …* *throw, kick, …* *bring, send, take, …*
Extensions on the basis of general semantic processes (Goldberg 1995: 38):	
A. *Satisfaction conditions imply that* agent causes recipient to receive theme	*guarantee, promise, owe, …*
B. *Agent enables* recipient to receive theme	*permit, allow, …*
C. Agent causes recipient *not* to receive theme	*refuse, deny, …*
D. *Agent acts to* cause recipient to receive theme *in the future*	*leave, bequeath, grant, …*
E. *Agent intends to* cause recipient to receive theme	*bake, make, build, …* *get, grab, earn, …*
Extensions on the basis of metaphor (Goldberg 1995: 147–150):	
F. Communication as transfer, e.g. *She told Joe a fairy tale.*	*tell, teach, fax, …*
G. Perceiving as receiving, e.g. *He showed Bob the view.*	*show, give a glimpse, …*
H. Directed action as transfer, e.g. *She blew him a kiss.*	*blow (a kiss), give (a wink), …*
Exceptions based on individual verbs (Goldberg 1995: 131–136):	*cost, charge, envy, forgive …*

Case study

Table 10. Collexemes most strongly attracted to the ditransitive construction

Collexeme	Collostruction strength	Collexeme	Collostruction strength
give (461)	0	allocate (4)	2.91E-06
tell (128)	1.6E-127	wish (9)	3.11E-06
send (64)	7.26E-68	accord (3)	8.15E-06
offer (43)	3.31E-49	pay (13)	2.34E-05
show (49)	2.23E-33	hand (5)	3.01E-05
cost (20)	1.12E-22	guarantee (4)	4.72E-05
teach (15)	4.32E-16	buy (9)	6.35E-05
award (7)	1.36E-11	assign (3)	2.61E-04
allow (18)	1.12E-10	charge (4)	3.02E-04
lend (7)	2.85E-09	cause (8)	5.56E-04
deny (8)	4.5E-09	ask (12)	6.28E-04
owe (6)	2.67E-08	afford (4)	1.08E-03
promise (7)	3.23E-08	cook (3)	3.34E-03
earn (7)	2.13E-07	spare (2)	3.5E-03
grant (5)	1.33E-06	drop (3)	2.16E-02

Case study

➢ Results of the analysis:

- Demonstrates that there are associations between the ditransitive and specific verbs
- Shows that these can be ranked
- Yields results that bear on analyses of the ditransitive such as that presented by Goldberg

Case study

- The strongest collocate is *give*.

→ Collostruction strength confirms the importance of semantic compatibility.

→ Strong collexemes of a construction provide a good indicator of its meaning.

- The next strongest collocates after *give* mainly instantiate extended senses.

→ Eight of the nine extensions listed in Table 10 are instantiated by one or more of the fifteen strongest collocates.

Case study

→ Extension A by *offer*, *owe*, and *promise*
→ Extension B by *allow*
→ Extension C by *deny*
→ Extension D by *grant*
→ Extension E by *earn*
→ Extension F by *tell* and *teach*
→ Extension G by *show*,
→ Exceptional uses by *cost.*

Conclusion

- Advantages of the method:

1. The descriptive adequacy of grammatical description is strongly increased.

→ Collostructional analysis with its emphasis on (i) the grammatical structures in which collexemes are embedded and (ii) the quantification of the degree of attraction/repulsion has more precise results and more rewarding perspectives to offer.

Conclusion

2. The results presented above have implications for linguistic theorizing and model-building.

→ The very fact that there are any dependencies at all between particular words and particular grammatical structures provides strong support for theories that view grammatical structures as signs, specifically for theories that view language as a repository of linguistic units of various degrees of specificity.

Conclusion

→ If syntactic structures served as meaningless templates waiting for the insertion of lexical material, no significant associations between these templates and specific verbs would be expected in the first place.

Further reading

- Glynn, D. & Fischer, K. (eds, 2010). *Quantitative Methods in Cognitive Semantics: Corpus-driven Approaches*. Berlin/New York: Mouton de Gryuter.
- González-Márquez, M., Mittelberg, I., Coulson, S. and Spivey, M.J., (eds, 2007). *Methods in Cognitive Linguistics*. Amsterdam: Benjamins.
- Stefanowitsch, A. & Gries, Stefan Th. (eds, 2006). *Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis* (Trends in Linguistics: Studies and Monographs), Berlin: Mouton de Gruyter.