of-Genitive versus *s*-Genitive: A Corpus-Based Analysis of Possessive Constructions in 20th Century American English

Melanie Röthlisberger
Gerold Schneider

English Department, University of Zurich, Switzerland

New Methods in Historical Corpora
2011, Manchester

Contents

Contents

- Introduction
- Conclusions

Automatization

- What can we automatize?
  - Parsing : resolves ambiguity, relatively few errors
    - surface form of s-genitives is identical to contracted forms of *be* in the 3rd person singular (e.g. *Peter's painting is large* vs. *Peter's painting a house*)
    - attachment site of prepositions:
      We accused the man of robbery (of-PP attaches to verb)
      the state of emergency of the nation (2nd of-PP attaches to *state* and not to *emergency)*

    ➜ each text has been parsed with the syntactic parser Pro3Gres (Schneider 2008)
  - Envelope of Application: difficult. Many restrictions are semantic
    - Use raw counts?
    - Invent an approximation to semantic restrictions?

Raw Counts

- No restrictions, no envelope

corp	*cat*	*Sax/of*	*Saxon #*	*%*	*of #*	*%*
[1] w/o envelope						
BBrown	J	**0.05**	347	4.72%	6998	95.28%
Brown	J	**0.06**	411	6.07%	6356	93.93%
Frown	J	**0.12**	716	10.90%	5853	89.10%

→ Saxon seems to increase, of-PP seems to decrease

1930-60	**+18.44%**	**-9.17%**
1960-90	**+74.21%**	**-7.91%**

… but the envelope of application is for the Genitive alternation is not considered

Proper Nouns / Animacy

- Proper Nouns only, can be seen as stopgap to animacy

corp	*cat*	*Sax/of*	*Saxon #*	*%*	*of #*	*%*	*% PN / ALL*	*% PN / ALL*
PROPER NOUNS ONLY							Saxon	modpp-of
BBrown	J	**0.40**	255	28.68%	634	71.32%	73.49%	**9.06%**
Brown	J	**0.55**	245	35.40%	447	64.60%	59.61%	**7.03%**
Frown	J	**1.16**	468	53.67%	404	46.33%	65.36%	**6.90%**

- Saxon genitive gets more restricted to Proper Nouns
- Readiness for Proper Noun in of-PP decreases
- Again, no envelope

Alternations & Envelope

- Alternations (dative shift, saxon genitive, ...) are a research focus
- investigations using large amounts of quantitative data and statistical techniques (e.g. Bresnan and Nikitina 2009 on dative shift)
- Envelope of application (Labov 1969) = choice context (Rosenbach 2003)
  - Lexical Equality:
    - give a book to the student <=> give the student a book
    - Peter's friend <=> friend of Peter
  - Used successfully for dative shift (Lehmann and Schneider 2010)
  - Filters e.g. adjunct:
    - drive the car to London <≠> *drive London the car
  - e.g. idioms
    - Point of view <≠> *view's point
    - *view of bird <≠> bird's view
  - e.g. creators
    - Spielberg's film <≠> *film of Spielberg

Alternations & Envelope II

- Envelope of application (Labov 1969) = choice context (Rosenbach 2003)
  Filters e.g. fixed nominal expressions / proper names:
    Noah's arch <=> ?Arch of Noah
    Newton's comet <=> ?Comet of Newton
    Earth's crust <=> ?crust of the earth
    Institute of Archeology <=> *Archeology's Institute
  e.g. measures / quality
    Tin of soup <≠> *soup's tin
    King of honour <≠> *honour's king
    Half of century <≠> *century's half
  e.g. semantic restrictions:
    One's recovery <=> *recovery of one
    God's creation <≠> ?creation of god
  ... many other expressions that are not in the alternation, e.g.:
    Concentration of oxygen <=> ?oxygen's concentration
    Image of power <≠> ?power's image
    faculty of reason <≠> ?reason's faculty
  … and often sparse data problems (alternation possible, but not found)
  ← Language use, parole
- Corpus-driven :-) systematic, automatic

Lexis Envelope

- If *A's B* and *B of A* occur in corpus → valid alternation

[2c] lex envelope		*Sax/ of*	*Saxon #*	*%*	*of #*	*%*
BBrown	J	1.14	24		21	
Brown	J	0.8	28		35	
Frown	J	1.18	19		16	

→ extremely sparse, no trends, probably just random fluctuations

How can we approximate or better employ the lexis envelope?

- Restrict only one side: if *A's B* and *B of C* occur → valid alternation
  "limp on one leg"
- Use semantic classes of words

Limping Lexis Envelope

- If *A's B* and *B of C* occur in corpus → valid alternation [B = head]

corp	*cat*	*Sax/of*	*Saxon #*	*% Sax/all*	*of #*	*% of/all*
[2c] lex envelope		Dep=Dep2				
BBrown	J	**0.18**	210	**14.99%**	1191	85.01%
Brown	J	**0.29**	212	**22.22%**	742	77.78%
Frown	J	**0.59**	462	**36.93%**	789	63.07%
		incr		incr		decr?
1930-60				**0.95%**		**-37.70%**
1960-90				**117.92%**		**6.33%**
[2c] lex envelope		Head=Head2				
BBrown	J	**0.08**	236	**7.72%**	2821	92.28%
Brown	J	**0.15**	309	**12.76%**	2113	87.24%
Frown	J	**0.22**	534	**17.80%**	2466	82.20%
		incr		incr		decr?
1930-60				**30.93%**		**-25.10%**
1960-90				**72.82%**		**16.71%**

Same trends as in raw data, less clear decrease for of-PP

WordNet

Semantic Lexis Envelope

corp	*cat*	*Sax/of*	*Saxon #*	*% Sax/all*	*of #*	*% of/all*
[3] WordNet Class envelope						
BBrown	J	0.070	240	**6.59%**	3404	93.41%
Brown	J	0.104	319	**9.50%**	3039	90.50%
Frown	J	0.181	564	**15.39%**	3101	84.61%
				incr		decr

- Results from raw counts are confirmed
- Significant difference between 1930-60-90:
  Chi-square contingency table, $p < 2.2E-16$
- Similar to manual method

Statistical Significance

A. Significant difference between 1930-60-90:
   Chi-square contingency table, $p < 2.2E-16$
B. Non-signi diff manu/auto

Adding indefinite filter

- ***the*** nest of the bird <=> Bird's nest
- ***A*** nest of the bird <≠> Bird's nest
- Indefiniteness filter was easy to add
- Results change little, trends even clearer

corp	*cat*	*Sax/of*	*Saxon #*	*% Sax/all*	*of #*	*% of/all*
[3] WordNet Class envelope						
BBrown	J	0.07	240	**9.27%**	2349	90.73%
Brown	J	0.10	319	**14.63%**	1861	85.37%
Frown	J	0.18	564	**24.92%**	1699	75.08%
				incr		decr
1930-60				**+32.92%**		**-20.77%**
1960-90				**+76.80%**		**-8.70%**

Scaling: Category K

- ***the***

Scaling: BLOB, LOB, FLOB

-

corp	*cat*	*Sax/of*	*Saxon #*	*%*	*of #*	*%*
[1] w/o envelope						
BLOB	J	0.052	362	4.97%	6929	95.03%
LOB	J	0.072	425	6.72%	5897	93.28%
FLOB	J	0.093	251	8.55%	2686	91.45%

[3] WordNet						
BLOB	J	0.100	243	12.64%	1680	87.36%
LOB	J	0.123	333	16.13%	1731	83.87%
FLOB	J	0.128	169	18.31%	754	81.69%

Principle of end weight

		N	p(DepMod)	p(HeadMod)	Factor=H/D	N	p(DepMod)	p(HeadMod)	Factor =H/D
BBrown	J	240	11.67%	27.08%	2.321	3404	43.83%	33.46%	0.763
Brown	J	319	13.48%	32.92%	2.442	3039	46.03%	38.40%	0.834
Frown	J	563	11.37%	35.17%	3.094	3101	47.66%	36.96%	0.775

		N	p(DepMod)	p(HeadMod)	Factor=H/D	N	p(DepMod)	p(HeadMod)	Factor =H/D
BLOB	J	243	18.52%	16.87%	0.911	2407	38.97%	31.57%	0.810
LOB	J	333	18.02%	22.82%	1.267	2700	44.33%	34.15%	0.770
FLOB	J	169	10.06%	18.34%	1.824	1319	44.43%	32.15%	0.724

Conclusions

- **Saxon Genitive has increased** from 1930 to 60 to 90 in AmE
- of-PP has decreased, relative to frequency of Saxon
- **Saxon Genitive** gets more restricted to **Proper Nouns**
- We have presented an an approach to the **automatic detection** of pairs in the **Genitive alternation**
- Differences over the time periods are significant
- **British English** shows the same trends
- The **principle of end weight** has become stronger

References

Arppe, Antti, Gaëtanelle Gilquin, Dylan Glynn, Martin Hilpert and Arne Zeschel, 2011. “Cognitive Corpus Linguistics: Five points of debate on current theory and methodology”. To appear in Corpora 5/2.

Baker, Collin F. and Josef Ruppenhofer (2002) FrameNet's Frames vs. Levin's Verb Classes. In J. Larson and M. Paster (Eds.) In Proceedings of the 28th Annual Meeting of the Berkeley Linguistics Society. 27-38.

Bresnan, Joan and Tatiana Nikitina, 2009. “The Gradience of the Dative Alternation”. In Reality Exploration and Discovery: Pattern Interaction in Language and Life, edited by Linda Uyechi and Lian Hee Wee. Stanford: CSLI Publications.161-184.

Fillmore, Charles J., Christopher R. Johnson and Miriam R.L. Petruck, 2003. “Background to FrameNet”. International Journal of Lexicography, 16:235–250.

Labov, 1969

Leitner, Florian, Scott A. Mardis, Martin Krallinger, Gianni Cesareni, Lynette A. Hirschman, Alfonso Valencia, 2010. "An Overview of BioCreative II.5," IEEE/ACM Transactions on Computational Biology and Bioinformatics, pp. 385-399, July-September

Rosenbach, 2003.

Tse, Grace Y. W., 2003. “Validating the Logistic Model of Article Usage Preceding Multi-word Organization Names with the Aid of Computer Corpora”. Literary and Linguistic Computing 18 (3): 287-313.