Cognitive Routinization in Language Comprehension
The Case of English Finite Non-Subject Relative Clauses

Daniel Wiechmann — Friedrich Schiller University, Jena

seit 1558

INTRODUCTION

Drawing on ideas from usage-based cognitive construction grammars (Langacker 2008) and exemplar-/memory-based models of language processing (Daelemans and van den Bosch 2005). this paper proposes a corpus-based methodology that is geared to predict the processing demand of complex linguistic structures.

Assumptions:

I. All linguistic experiences (~exemplars) are *stored* in memory (from MBL)
   a. Extensionally, linguistic knowledge is the set of all stored exemplars
II. Subsets may form coherent groups (*exemplar clusters*), which are *structured* on the basis of *similarity* (from MBL)
III. To comprehend a linguistic form is either
   a. to *look it up in memory*
   b. or look up nearest neighbor & use analogical reasoning (from MBL)
IV. Look-up is fastest for *recurrent* (*sequences of) forms* (MBL & *entrenchment* in CxG)
V. Relevant units in this process are be signs, but these can be complex and abstract (CxG)

Hypothesis

The processing demand of a complex patterns is a function of its degree of entrenchment (and the degree of entrenchment of its *k*-nearest neighbors).

Test case phenomenon:

English finite non-subject relative clause constructions (RCC-types)

Goals of this paper:

I. Detect deeply entrenched RCC-types
II. Build constructional network of these RCC-types
III. Predict processing demand from network position (and check corpus-based findings against experimental results)

DATA

Corpus: British Component of International Corpus of English (ICE-GB R2)

№ = 329 finite non-subject RCC

i. Lohmi drank the only beer $_{RC}$[that he brought __ to the party].
ii. The beer $_{RC}$[that Lohmi brought __] was the first to go.

Variables used in description (theory-driven)

Factor	Levels
1. medium/register	(spoken/written)
2. syntactic type of head	(+/- lexical)
3. "uniqueness" adjective	(+/- present)
4. contentfulness of head	(+/- low)
5. animacy of head	(+/- animate)
6. definiteness of head	(+/- definite)
7. syn. type of RC-subject	(+/- lexical)
8. relativizer	(present/absent)
9. type of embedding	(right /center)

→2^9 = 512 possible types of RCCs

METHODS & RESULTS

Goal I: Detect deeply entrenched RCCs

Technique: (Hierarchical) Configural Frequency Analysis (hCFA, von Eye 1990)

A RCC-type is considered to be deeply entrenched, iff its observed frequency is statistically significantly higher than its expected frequency

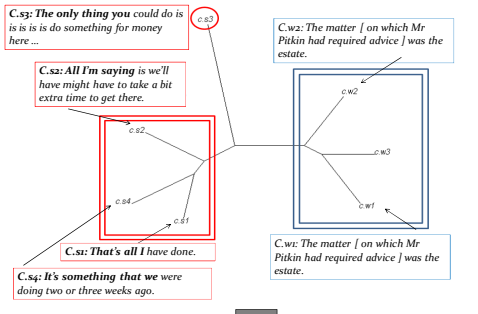
Results: Detection of 237 significant types ($p < .01$)
- **7 fully specified types** (CFA)
- **230 types w/ exactly 1 unfilled slot** (hCFA))

Example: *Cs3: **The only thing you could do** is is is is is do something for money here ...*

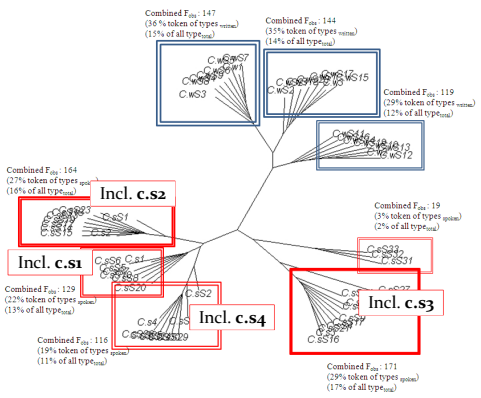
Name	c.s3	medium	spoken
		head type	lexical
Freq	15	unique A	present
Exp	1.1079	content head	low
Cont.chisq	174.1948	animacy head	inanimate
Obs-exp	>	definiteness head	definite
P.adj.bin	4.89E-10	SRC type	pronominal
Dec	***	relativizer	absent
Q	0.042	embedding	center

Goal II: Structure RCC-types on the basis of similarity

Technique: Hierarchical Agglomerative Clustering
similarity* = *distance in ***Euclidean*** space
amalgamation via ***neighbor joining tree estimation*** (Saitou and Nei 1987)
results in ***unrooted tree structures***



Elaborating the network: Inclusion of hCFA types



CONCLUSIONS

Predictions from network position to expected processing demand are fully compatible w/ **robust experimental findings**.

Spoken types exhibit all properties that have been identified in experimental studies to facilitate processing, such as

a. Animacy of RC subject facilitates processing (Mak et al. 2004)
b. Presence of prenominal adjectives makes RC more predictable & thus easier to process (Jaeger and Wasow 2008)
c. Pronominal RC subject indicates high accessibility of referents & that facilitates processing (Reali & Christiansen 2007)
d. Morphosyntactic dissimilarity of head & RC subject facilitates processing (Gordon et al. 2004)
e. Semantic indeterminacy of head (generic head noun) facilitates processing (Gennari & MacDonald 2008)

Results **encourage** the further pursuit of the idea that a **combination of constructionist and memory-based processing approaches** is beneficial to both programmes.

LITERATURE CITED

Daelemans, W., and van den Bosch, A, 2005. *Memory-Based Language Processing*, Cambridge University Press.
Jaeger, F., and Wasow, T. 2008. Processing as a Source of Accessibility Effects on Variation. *Proceedings of the 31st BLS*. 169-180. 2008.
Gennari, S. P., and MacDonald, M. C. 2008. Semantic indeterminacy in object relative clauses. *Journal of Memory and Language*, 58, 161-187.
Gordon, P.C., Hendrick, R., and Johnson, M. 2004. Effects of noun phrase type on sentence complexity. *Journal of memory and language*, 51, 97-114.
Langacker, R. 2008. *Cognitive Grammar: A Basic Introduction*. New York: Oxford University Press.
Reali, F., and Christiansen, M. 2007. Word-chunk frequencies affect the processing of pronominal object-relative clauses. *Quarterly Journal of Experimental Psychology*, 60, 161-170
Saitou, N., and Nei M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4:406-425.

ACKNOWLEDGEMENTS

I would like to thank the R Development Core Team and all contributors, in particular

Stefan Gries for providing his neat hcfa-script