Cognitive Routinization in Language Comprehension
The case of English finite non-subject relative clauses

Daniel Wiechmann

University of Jena

Introduction

This paper investigates the role of schematic constructions for processes of sentence comprehension from the viewpoint of usage-based cognitive construction grammars (Goldberg 2006, Langacker 2008).

Guiding Hypothesis:

Processing of a sentence level linguistic structure essentially is accessing complex templates (constructional schemas) from memory, such that

> ➢ for all linguistic expressions E, the greater E's degree of entrenchment (= cognitive routinization), the easier its processing

Goals:

1. Detect firmly entrenched patterns
   - Domain of interest: English finite non-subject relative clause constructions (RCC)
2. Assess constructional network of relevant domain
3. Predict processing demand from network position

Data

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

- Extraction of all bi-clausal relative clause constructions
- Manual inspection of data
- Nº "clean" bi-clausal RCCs = 1088
- Nº finite non-subject RCC = 329 ($Ratio_{spoken/written}$ ~ 2:1)

Manual coding of data:

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)

Step1: Detect entrenched patterns

- **Technique 1: Detect fuzzy association rules**
  - *k*-optimal pattern discovery (Webb and Zhang 2005)
    - Association Rule: $\{PROPERTY_A, PROP_B \dots, PROP_K\}$ $\{PROPERTY_L, PROP_M \dots, PROP_N\}$
  - **Result**: Detection of 14 significant rules ($p < .01$)
    - Rules disclose distinctive characteristics for each register
      - spoken rules invariably exhibit "simpler" factor levels
      - auditory channel → no external representation, only short term buffering, fast decay of memory traces
    - But detected rules are quite simple (→ sparse data)

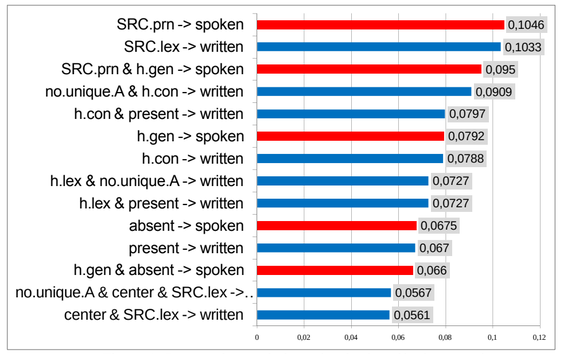


Figure 1: Detected association rules (ordered by leverage)

- **Technique 2: Detect significant configurations (types)**
  - *Hierarchical Configural Frequency Analysis* (hCFA, von Eye 1990)
    - $F_{obs} > F_{exp}$ → *"type"* (~ entrenched pattern)
  - **Results:** Detection of 237 significant types ($p < .01$)
    - 7 fully specified types (CFA)
    - 230 types w/ exactly 1 unfilled slot (hCFA)

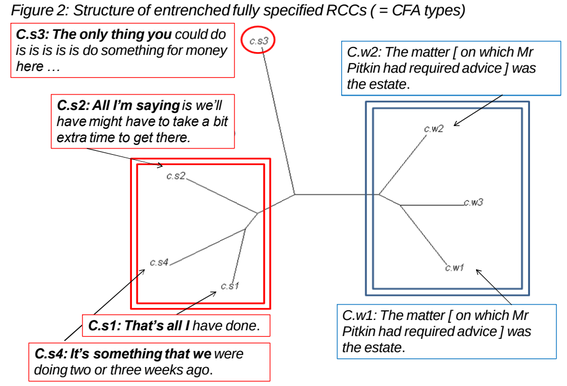
Example: fully specified $type_{CFA}$ - C.S3
Instance: *The only thing he can do is ...* [S2A-004 #138]

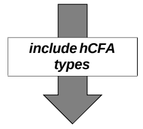
Name	c.s3	medium	spoken
		head type	lexical
Freq	15	unique A	present
Exp	1.1079	content head	low
Cont.chisq	174.1948	animacy head	inanimate
Obs-exp	>	definiteness head	definite
P.adj.bin	4.89E-10	SRC type	pronominal
Dec	***	relativizer	absent
Q	0.042	embedding	center

Step 2: Structure detected patterns

- **Hierarchical Agglomerative Clustering**
  - similarity = *distance in Euclidean space*
  - amalgamation via *neighbor joining tree estimation* (Saitou and Nei 1987)
  - ➢ gives rise to unrooted tree structures

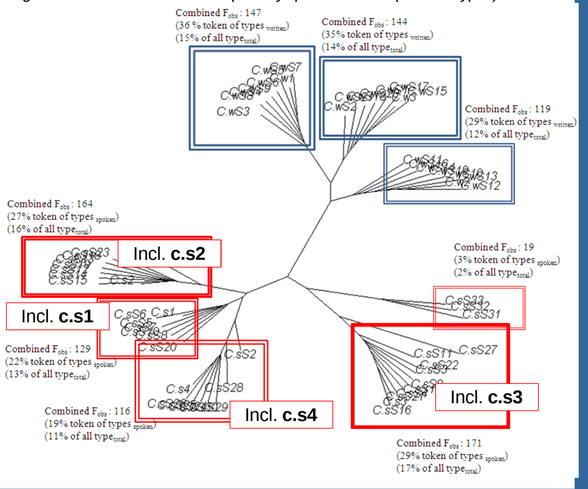
Figure 2: Structure of entrenched fully specified RCCs (= CFA types)





Predicting processing demand:
Processing demand of an expression E is a function of E's distance to a cluster from spoken L (red boxes)

Figure 3: Structure of entrenched partially specified RCCs (= hCFA types)



Conclusion

Results are captured naturally by theoretical constructs of cognitive construction grammars

- *Entrenchment* determines constructional access
- *Constructional access* is identified with processing demand

Results corroborate exemplar-based views on language representation and processing

- CFA focuses on *individuals* and not *variables*
- Detected $types_{CFA}$ show lexical specificity

Results reflect robust experimental findings

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

- Animacy of RC subject helps processing (Mak et al. 2004)
- Presence of prenominal adjectives (Jaeger and Wasow 2008)
- Accessibility of RC subject (Reali & Christiansen 2007)
- Morphosyntactic dissimilarity of head & RC subject (Gordon et al. 2004)
- Semantic indeterminacy of head (Gennari & MacDonald 2008)

References:
•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.
•Goldberg, A. 2003. Constructions: A New Theoretical Approach to Language. *Trends in Cognitive Science* 7. 5:219-224.
•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.
•Webb, G. I., and Zhang, S. 2005. k-Optimal-Rule-Discovery. *Data Mining and Knowledge Discovery,* 10,(1). 39-79.

Contact Info & Acknowledgements

For more information feel free to contact me :

Daniel Wiechmann

FSU Jena, Germany

Email: daniel.wiechmann@uni-jena.de
Web: http://daniel-wiechmann.net

I thank *Stefan Gries* for providing his neat *hcfa*-script and *GI Webb and associates* for providing an *Magnus Opus* demo