

*First Triennial Conference of the International Society for the Linguistics of English*  
**ISLE 1**

*Freiburg, October 8-11 2008*



# *Towards an explanation of relativizer omission*



Daniel Wiechmann  
Friedrich Schiller University, Jena



Phenomenon:

Relativizer omission in Non-Subject RC

### Relative Clause Construction (RCC)

Peter hates [ the car ]<sub>i</sub> ( that ) he bought \_\_\_<sub>i</sub> from his friend

                  |                  |                  |

                  head  relativizer                  gap

**What are the conditions under which relativizers are omitted?**





## Processing explanations:

Optional relativizer is **dropped** if RCC is **easy to process**  
& **produced** if RCC is **hard to process**

**But what does it mean for a pattern to be easy to process?**

Predictability	?	(e.g. Jaeger, Fox & Thompson)
Complexity	?	(e.g. Hawkins)
Memory load	?	(e.g. Gibson)
...	?	

- ▶ overt *that* makes processing easier ✓
- ▶ **R-omission** is a function of degree of **entrenchment RCC pattern**



## ► Entrenchment



**AUTOMATIZATION** is the process observed in learning to tie a shoe or recite the alphabet: through repetition or rehearsal, a complex structure is thoroughly mastered to the point that using it is virtually automatic and requires little conscious monitoring. In CG parlance, a structure undergoes progressive **ENTRENCHMENT** and eventually becomes established as a unit“

(Langacker 2008:16)



## ► Entrenchment



**AUTOMATIZATION** is the **process** observed in learning to tie a shoe or recite the alphabet: through **repetition** or rehearsal, a **complex structure** is thoroughly mastered to the point that using it is virtually automatic and requires little conscious monitoring. In CG parlance, a structure undergoes progressive **ENTRENCHMENT** and eventually becomes **established as a unit**“

(Langacker 2008:16)



## Entrenchment & language processing:



▶ For all linguistic expressions E,  
the more entrenched E,  
the easier to process E

▶ **The more entrenched a RCC pattern,  
the easier is its processing**  
(no *that* needed)



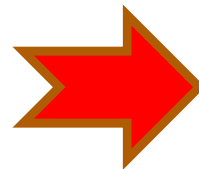


## *Detecting entrenched patterns*

simple units	(words)	<input checked="" type="checkbox"/>
bi-grams	(collocations)	<input checked="" type="checkbox"/>
n-grams w/ $n > 2$	(complex patterns)	<input type="checkbox"/>

### **configural frequency analysis**

Lienert (1969), von Eye (1990)



Corpus data:

**ICE-GB**

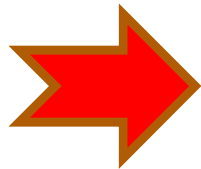
only bi-clausal RCC

$n = 329$





# *Detecting entrenched patterns:* Steps in the analysis



1

## **configural frequency analysis**

Lienert (1969), von Eye (1990)

2

## ***k*-optimal pattern discovery**

G.I. Webb and S. Zhang (2005)

3

## **binary logistic regression**

Corpus data:

**ICE-GB [test set]**

spoken (less restricted)

N = 400





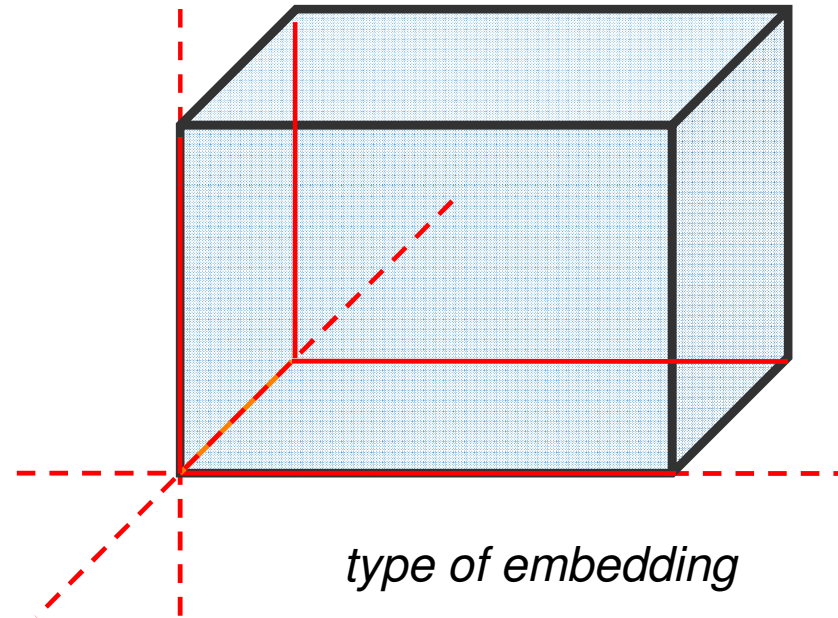
Complex constructions are highly variable patterns...

I love the papers<sub>RC</sub> [ that Steve hates so much ].

State space

*animacy head*

- [ + ANIMATE ]
- [ - ANIMATE ]



*Presence of R-element*

- [ + PRESENT ]
- [ - PRESENT ]

*type of embedding*

- Right embedding
- Center embeddiing...





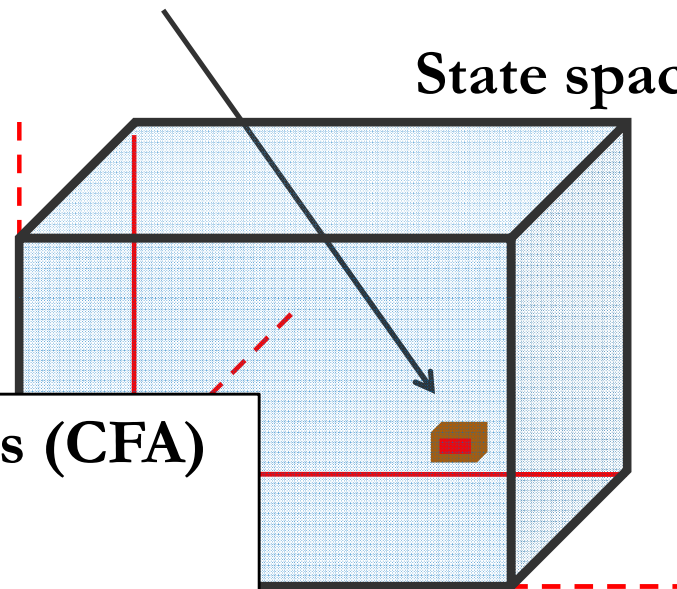
Complex constructions are highly variable patterns...

I love the papers<sub>RC</sub> [ that Steve hates so much ].

*animacy head*

- [+ ANIMATE]
- [- ANIMATE]

State space



### Configurational frequency analysis (CFA)

1. *search through state space*
2. *evaluate all configurations*
3. *identify configurations that are special*

Observed configurational frequency  $>$  expected frequency

⇒ **TYPE** (= entrenched patterns)





I love the papers<sub>RC</sub> that Steve hates so much].

	<i>variables</i>	<i>levels</i>	
1	embedding	center	right
2	uniqueness adj	present	absent
3	contentfulness head	general	specific
4	animacy head	animate	inanimate
5	definiteness head	definite	indefinite
6	type of RC-subject	lexical	pronominal
7	medium	spoken	written
8	relativizer	present	absent

Presence of R-element

- \* [ + PRESENT ]
- \* [ - PRESENT ]

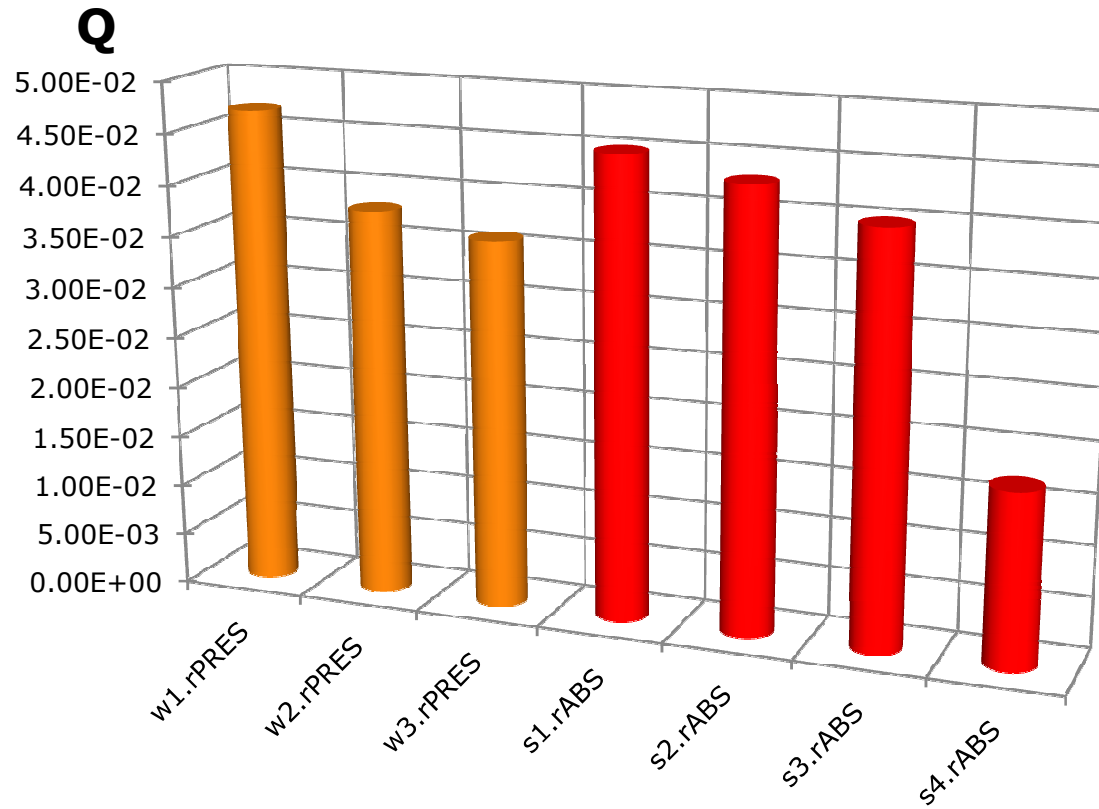
$2^8 = 256$  possible configurations





$2^8 = 256$  possible configurations

7 configurations are significant at  $\alpha = 0.001$  (binomial test w/ Bonferroni correction)



► all entrenched patterns exhibit *that*-omission

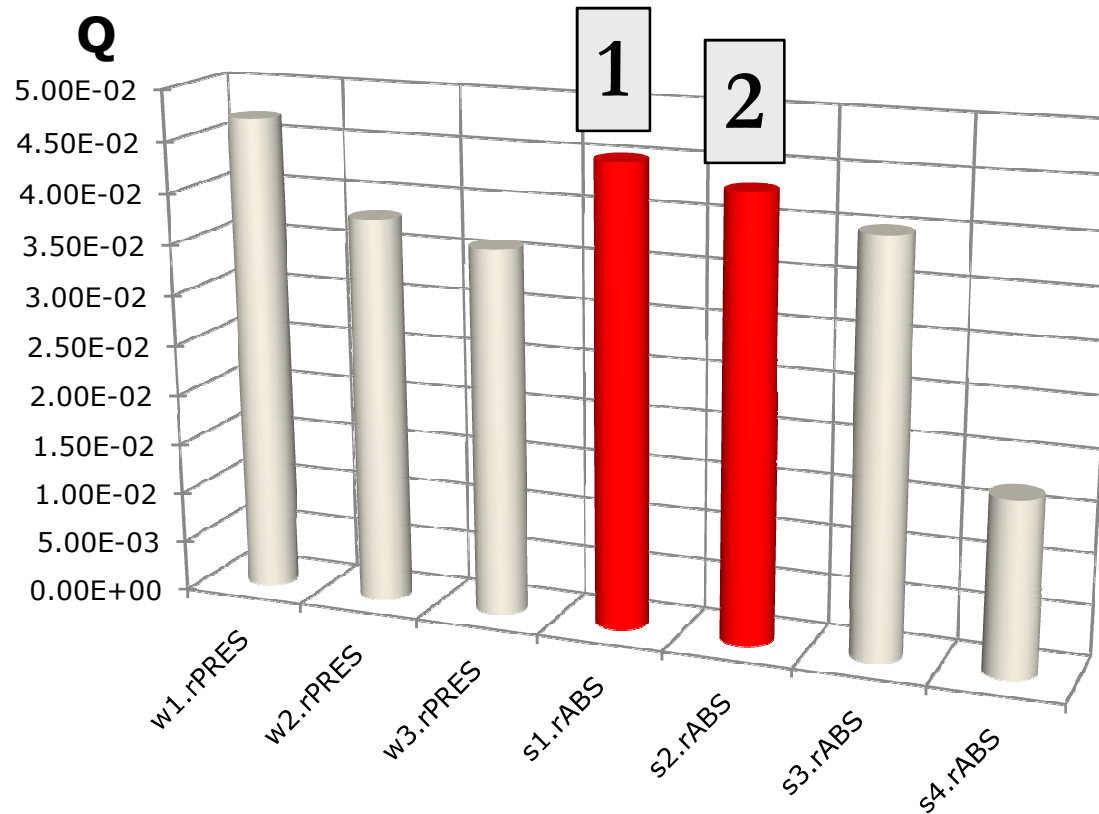


scit 1558



$2^8 = 256$  possible configurations

7 configurations are significant at  $\alpha = 0.001$  (binomial test w/ Bonferroni correction)



► all entrenched patterns exhibit *that*-omission



scit 1558



uniqueness Adj

SEM general  
inanimate  
definite

center embedding

pronominal Subj<sub>RC</sub>

1

RCC[ *The only thing* RC[ I VP<sub>RC</sub> ] VP<sub>MC</sub> ]





uniqueness Adj

SEM general  
inanimate  
definite

center embedding

pronominal Subj<sub>RC</sub>

1

RCC [ *The only thing* RC [ I VP<sub>RC</sub> ] VP<sub>MC</sub> ]

2

RCC [ SUBJ V NP [ *something* ] RC [ I VP<sub>RC</sub> ] ]

no uniqueness Adj

SEM general  
inanimate  
indefinite

right embedding

pronominal Subj<sub>RC</sub>





1

$RCC[ \textit{The only thing}_{RC} [ I \ VP_{RC} ] \ VP_{MC} ]$

2

$RCC [ \textit{This is}_{NP} [ \textit{something} ]_{RC} [ I \ VP_{RC} ] ]$



- Copula/predicative constructions dominant
- propositional content of  $p_{\text{main}}$  clause is minimal

Subj<sub>RC</sub>



These patterns are predicted by CFA to be easy...

1

$RCC [ \textit{The only thing} RC [ I VP_{RC} ] VP_{MC} ]$

2

$RCC [ \textit{This is} NP [ \textit{something} ] RC [ I VP_{RC} ] ]$





These patterns are predicted by CFA to be easy...

**1**  $RCC [ \textit{The only thing}_{RC} [ I \ VP_{RC} ] \ VP_{MC} ]$

Type of subjectRC (~givenness)



(Reali & Christiansen in press)

Animacy effects



(Mak et al. 2001, 2004)

Morphosyntactically similarity



(Gordon et al. 2001, 2004)

Semantic indeterminacy



(Gennari & MacDonald 2007)

**2**  $RCC [ \textit{This is}_{NP} [ \textit{something} ]_{RC} [ I \ VP_{RC} ] ]$





# In conclusion...

- I. Procedure**, Configural Frequency Analysis, identifies configurations (=types) that are statistically speaking special

(CFA results are quite **robust** -> results highly correlate w/ association rule mining techniques & log regression results)

- II. Type characteristics** are **in agreement w/** factors that have been shown to modulate processing difficulty to patterns that have been observed to be predicted to be easy by **experimental research**

- III. Types are linked to theory** (cognitive grammar/usage-based model) via the concept of **entrenchment**.





*Thank you for your attention!*

Presentation (and all references) available from  
[www.daniel-wiechmann.net](http://www.daniel-wiechmann.net)

I thank Stefan Th. Gries for sharing  
*hcfa 3.2* – a script for R for Windows  
[www.linguistics.ucsb.edu/faculty/stgries](http://www.linguistics.ucsb.edu/faculty/stgries)

