

Language, Communication & Cognition  
Brighton, August 4<sup>th</sup> – 7<sup>th</sup> 2008

# *Looking for the right type*

## *A Quantitative Corpus Linguistic Approach to the Processing of English Relative Clause Constructions*



seit 1558

Daniel Wiechmann  
Friedrich-Schiller-University. Jena  
[www.daniel-wiechmann.net](http://www.daniel-wiechmann.net)

*Introduction*

Data & Methods

Result & Discussion

# *Frequency & processing difficulty*



seit 1558

Language consists of a body of form-function pairings  
(signs, constructions)

- specific - schematic
- simple - complex

*Introduction*

Data & Methods

Result & Discussion

*Frequency &  
processing difficulty*



seit 1558

Language consists of a body of form-function pairings  
(signs, constructions)

- ❑ specific - schematic
- ❑ simple - complex

*Frequencies matter:*

- ❑ Learning (/development)
- ❑ Change over historical time
- ❑ On-line processing (comprehension & production)

Daniel Wiechmann

*Introduction*

Data & Methods

Result & Discussion

*Frequency &  
processing difficulty*



seit 1558

Language consists of a body of form-function pairings  
(signs, constructions)

- specific - schematic
- simple - complex
- infrequent - frequent

*Frequencies matter:*

- Learning (/development)
- Change over historical time
- On-line processing (comprehension & production)

Daniel Wiechmann

*Introduction*

Data & Methods

Result & Discussion

## *Frequency & processing difficulty*



seit 1558

Language consists of a body of form-function pairings  
(signs, constructions)

- specific - schematic
- simple - complex
- infrequent - frequent

*Relative clause constructions  
(min: main clause & relative clause)*

*The man <sub>RC</sub>[ that John hates \_ ] is actually quite nice.*

Daniel Wiechmann



**subject relatives**

the man [ who hates John ]

vs.

**non subject relatives**

the man [ who John hates \_ ]

**EXPERIMENTAL FINDINGS:**

Some non-SUBj-RC are harder than others

Pronominal subjectRC (~givenness)

(Reali & Christiansen in press)

Animacy effects

(Mak et al. 2001, 2004)

Head & Subject of RC are morphosyntactically similar

(Gordon et al. 2001, 2004)

Semantic indeterminacy

(Gennari & MacDonald 2007)

**PROPOSED EXPLANATION:**

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

Question:

**Which non-subject RCC schema is most entrenched?**

Introduction  
**Data & Methods**  
Result & Discussion

*Analysing  
Relative Clause Constructions (RCC):  
Step 1: Description*



seit 1558

**Case:**

The man <sub>RC</sub> [ that John hates ] is actually quite nice.

Daniel Wiechmann

Introduction  
**Data & Methods**  
Result & Discussion

## Analysing Relative Clause Constructions (RCC): Step 1: Description



seit 1558

**Case:**

The man<sub>RC</sub> [ that John hates ] is actually quite nice.

---

### *internal syntax*

- Subject relative (SRC)
- **Object<sub>DIRECT</sub> relative (ORC)**
- Object<sub>INDIRECT</sub> relatives
- ...

Daniel Wiechmann



**Case:** The man<sub>RC</sub> [ that John hates ] is actually quite nice.

*external syntax*

- NP [Det N RC] VP
- NP VP [ V NP [Det N RC]]
- ...

*internal syntax*

- Subject relative (SRC)
- **Object<sub>DIRECT</sub> relative (ORC)**
- Object<sub>INDIRECT</sub> relatives
- ...



**Case:** The man<sub>RC</sub> [ that John hates ] is actually quite nice.

*external syntax*

- NP [Det N RC] VP
- NP VP [ V NP [Det N RC]]
- ...

*finiteness*

- + **finite**
- - finite (reduced passive)  
*the horse raced past the barn*
- - to-infinitival  
*the right thing to do*

*internal syntax*

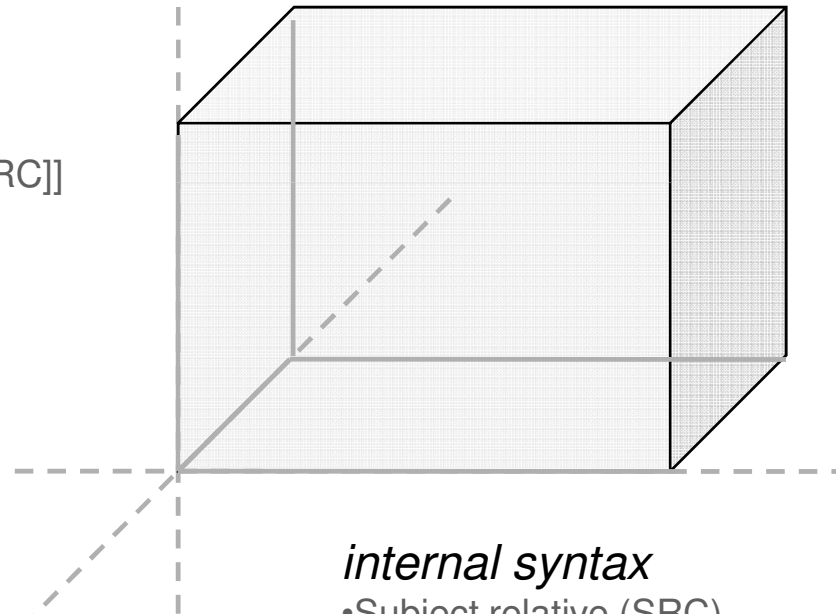
- Subject relative (SRC)
- **Object<sub>DIRECT</sub> relative (ORC)**
- Object<sub>INDIRECT</sub> relatives
- ...



**Case:** The man<sub>RC</sub> [ that John hates ] is actually quite nice.

*external syntax*

- NP [Det N RC] VP
- NP VP [ V NP [Det N RC]]
- ...



*finiteness*

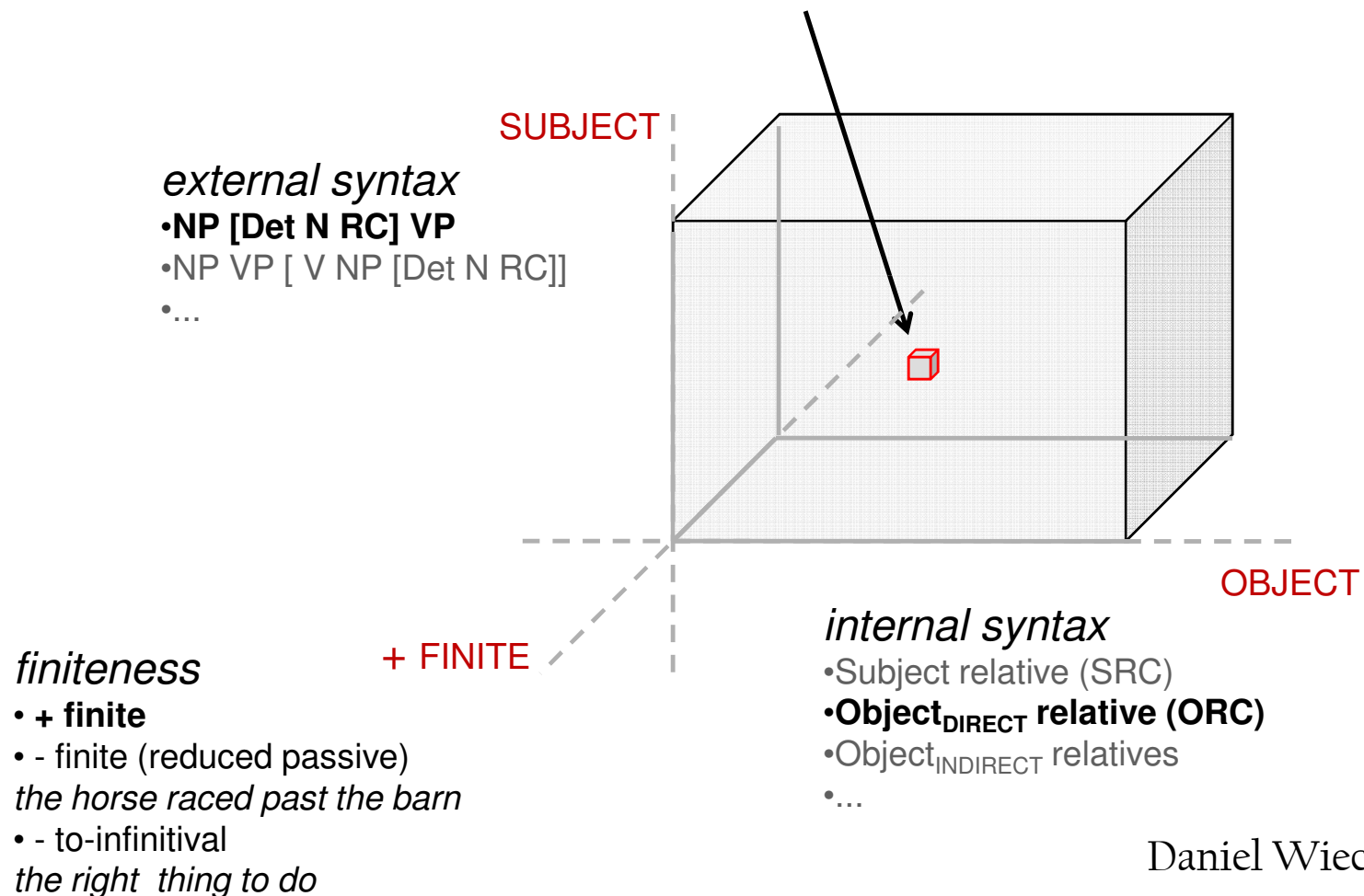
- + finite
- - finite (reduced passive)  
*the horse raced past the barn*
- - to-infinitival  
*the right thing to do*

*internal syntax*

- Subject relative (SRC)
- **Object<sub>DIRECT</sub> relative (ORC)**
- Object<sub>INDIRECT</sub> relatives
- ...

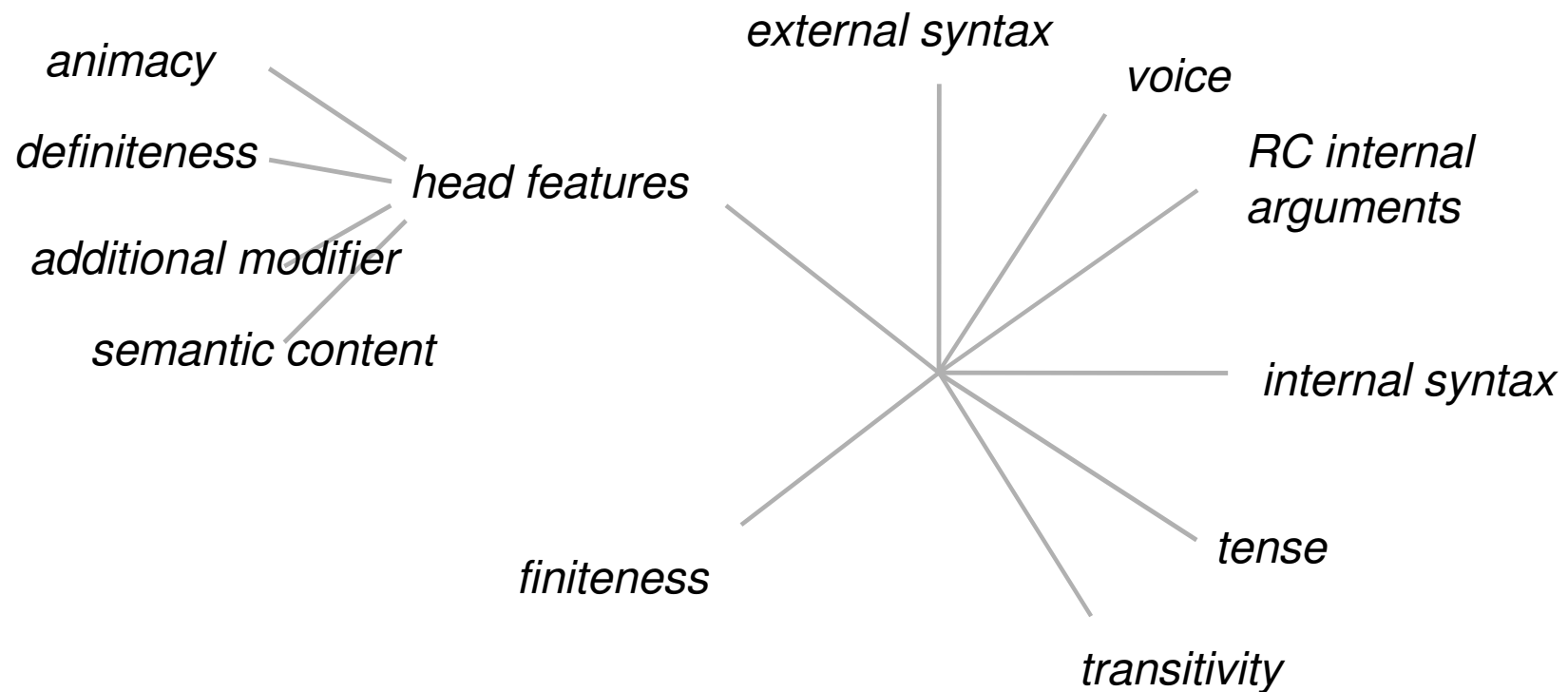


**Case:** The man<sub>RC</sub> [ that John hates ] is actually quite nice.





**Case:** The man<sub>RC</sub> [ that John hates ] is actually quite nice.



We end up with some **m-dimensional state space**

Binary variable:  $2^{\text{NUMBER OF VARIABLES}}$  possible patterns

( $2^{45}$  ~ age of universe in sec)



***TASK:***

**Configural frequency analysis (CFA)**

- *search through state space*
  - *evaluate all configurations*
  - *identify configurations that are special*



**TASK:**

## Configural frequency analysis (CFA)

- *search through state space*
- *evaluate all configurations*
- *identify configurations that are special*

**Observed** configural frequency  $>$  **expected** frequency

⇒ **TYPE** (=entrenched pattern)

Observed configural frequency  $<$  expected frequency

⇒ Anti-TYPE (=repelled pattern)



## Variables used in description

- Head features:
1. syntactic type
  2. animacy head
  3. definiteness head
  4. theta role in MC
  5. theta role in RC
- Subject of RC
6. syntactic type
  7. theta role
  8. animacy subj RC
  9. definite subj RC
10. type of embedding
  11. modality

## Data

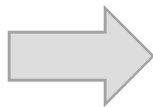
1000 English Relative Clause  
Constructions (RCC) from ICE-GB  
(500 spoken & 500 written)

Used here:

**finite, non-subject RCC**

$n_{\text{total}} = 318$  (  $n_{\text{spoken}} = 281$  )

center/right  
spoken/written



**11 dimensional state space**



INPUT

Given the factor level distinctions, there are (  $2^9 \times 3 \times 3 =$  )  
**4608 configurations** to be evaluated

OUTPUT

after poc hoc significance testing,  
**10 configurations** are considered **TYPES \***

\*

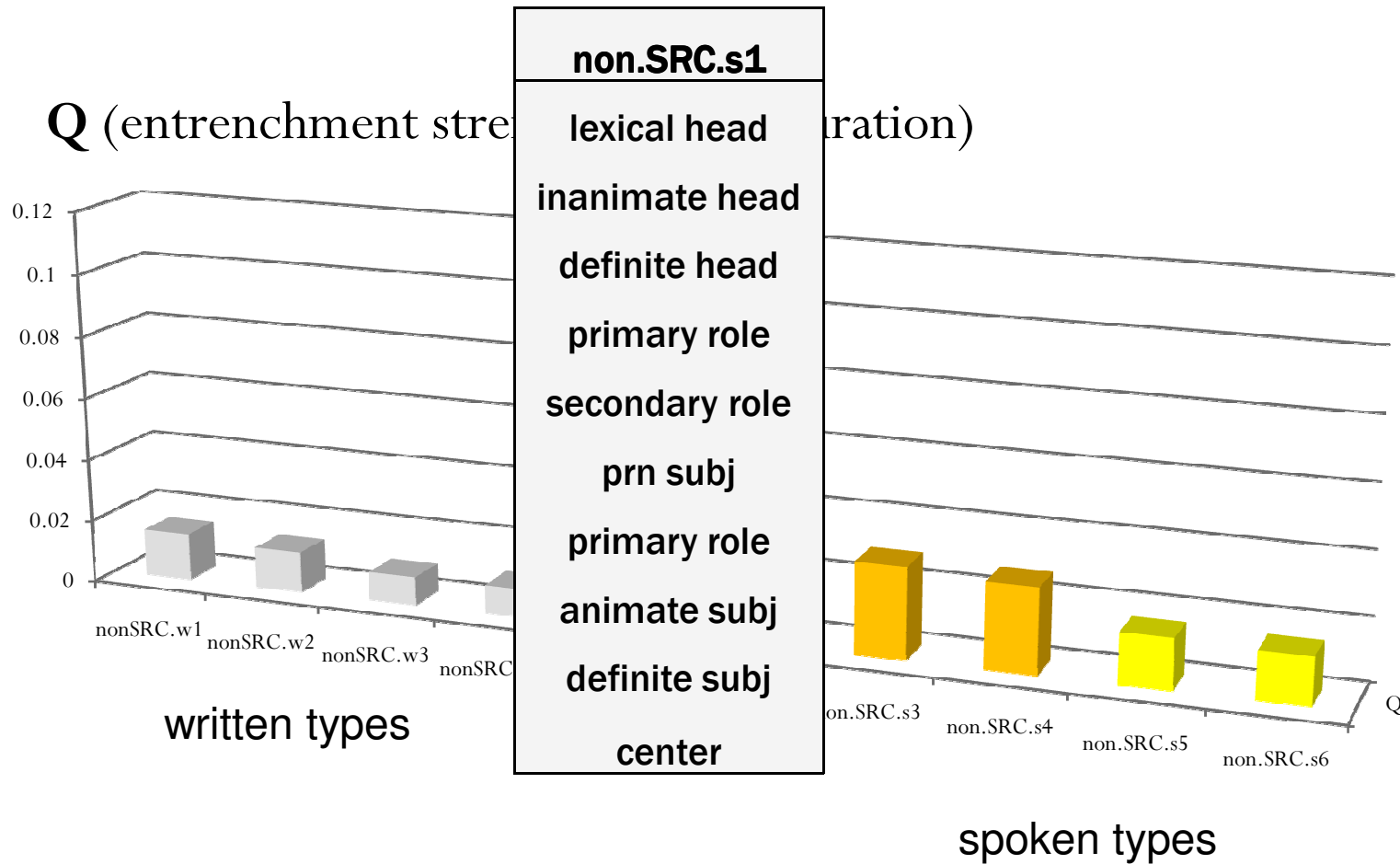
For 9 types:  
 $p < 0.001$   
& for 1 type:  
 $0.001 < p < 0.01$ )



	<b>nonSRC.w1</b>	<b>nonSRC.w2</b>	<b>nonSRC.w3</b>	<b>nonSRC.w4</b>	<b>non_SRC.s1</b>	<b>non_SRC.s2</b>	<b>non_SRC.s3</b>	<b>non_SRC.s4</b>	<b>non_SRC.s5</b>	<b>non_SRC.s6</b>	
<b>attribute values</b>	<i>head.type.bi</i>	lexical	lexical	lexical	lexical	lexical	pronominal	pronominal	lexical	pronominal	pronominal
	<i>animacy.head</i>	inanimate.head	inanimate.head	inanimate.head	inanimate.head	inanimate.head	inanimate.head	inanimate.head	animate.head	inanimate.head	inanimate.head
	<i>definiteness.head</i>	definite.head	indefinite.head	definite.head	indefinite.head	definite.head	indefinite.head	indefinite.head	definite.head	indefinite.head	indefinite.head
	<i>theta.ext.simple</i>	primary.role	secondary.role	primary.role	secondary.role	primary.role	secondary.role	primary.role	secondary.role	primary.role	secondary.role
	<i>theta.int.simple</i>	ternary.role	ternary.role	ternary.role	ternary.role	secondary.role	secondary.role	secondary.role	secondary.role	primary.role	primary.role
	<i>SRC.type.simple</i>	lexical	lexical	lexical	lexical	prn	prn	prn	prn	prn	prn
	<i>theta.SRC.simple</i>	primary.role	secondary.role	secondary.role	secondary.role	primary.role	primary.role	primary.role	primary.role	primary.role	primary.role
	<i>animacy.SRC</i>	n	n	n	n	y	y	y	y	y	y
	<i>definite.SRC</i>	y	y	n	n	y	y	y	y	y	y
	<i>embedding</i>	center	right	center	right	center	right	center	right	center	right
<b>Stats</b>	<i>Freq</i>	5	4	3	3	38	20	10	9	5	5
	<i>Exp</i>	0.1055	0.0112	7.00E-04	0.001	5.2375	1.5499	0.808	0.5713	0.0855	0.164
	<i>Cont.chisq</i>	227.0723	1420.5826	12851.1436	8994.001	204.9416	219.6311	104.5704	124.3532	282.4832	142.603
	<i>P.adj.Holm</i>	0.000444239	2.91E-06	3.14E-07	6.72E-07	2.01E-17	1.64E-12	6.40E-05	4.45E-05	0.000158199	0.003852804
	<i>Dec</i>	***	***	***	***	***	***	***	***	***	**
	<i>Q</i>	0.015	0.013	0.009	0.009	0.105	0.059	0.029	0.027	0.016	0.015

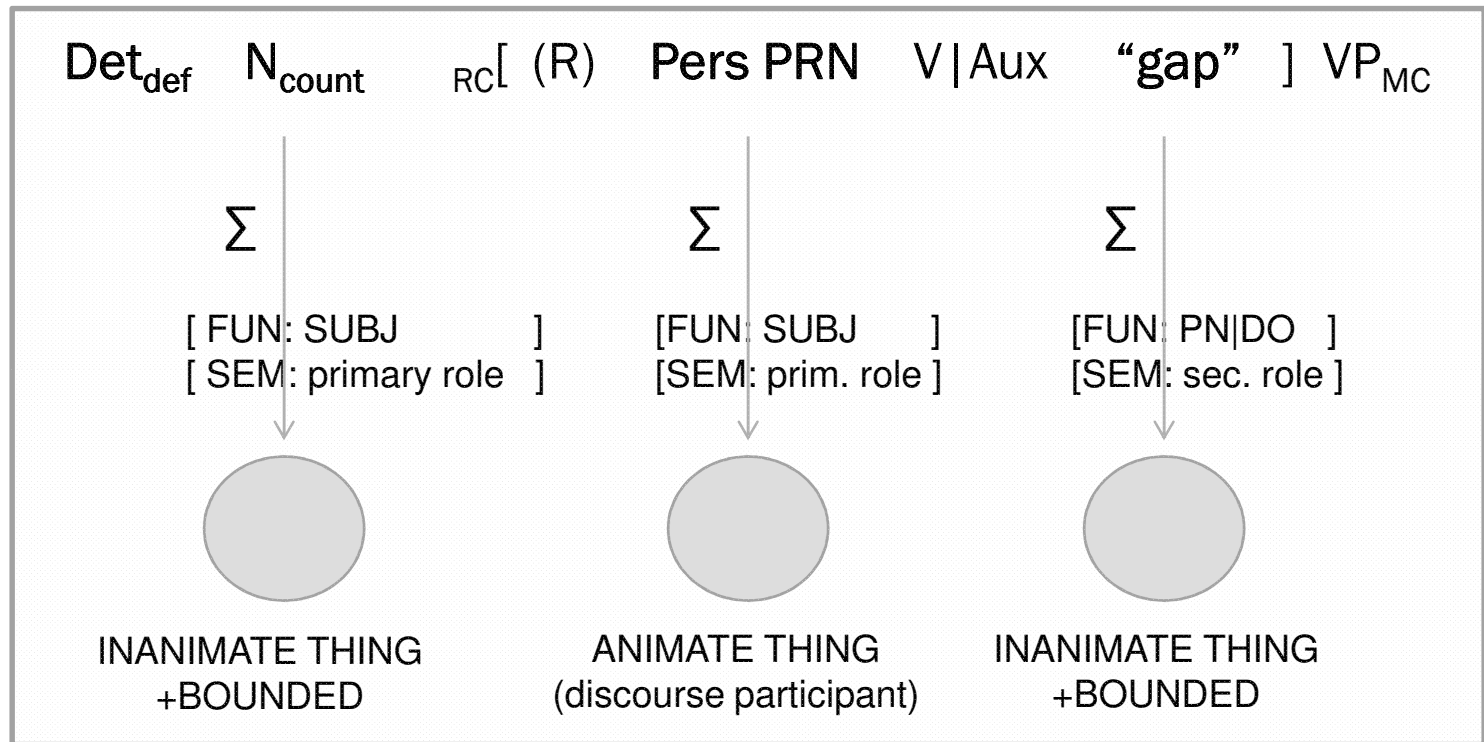
written types

spoken types





<b>non.SRC.s1</b>
lexical head
inanimate head
definite head
primary role
secondary role
prn subj
primary role
animate subj
definite subj
center



*The things* RC[ (that) I'm worried about \_ ] VP



*The things* RC [ (that) I'm worried about \_ ] VP

Det<sub>def</sub> N<sub>count</sub> RC [ (R) Pers PRN V|Aux "gap" ] VP<sub>MC</sub>

Σ  
[ FUN: SUBJ ]  
[ SEM: primary role ]

INANIMATE THING  
+BOUNDED

Σ  
[ FUN: SUBJ ]  
[ SEM: prim. role ]

ANIMATE THING  
(discourse participant)

Σ  
[ FUN: PN|DO ]  
[ SEM: sec. role ]

INANIMATE THING  
+BOUNDED

**Experimental Findings:**

Parallel function

Pronominal subjectRC (~givenness)

Animacy effects

Head & Subject of RC are morphosyntactically similar

Semantic indeterminacy

Discourse function

(Sheldon 1974)

(Real & Christiansen in press)

(Mak et al. 2001, 2004)

(Gordon et al. 2001, 2004)

(Gennari & MacDonald 2007)

(Fox & Thompson 1990)



*handle with care –  
some tentative conclusions*

## THEORY

- processing difficulty of E is a function of entrenchment of E
  - processing difficulty is distance to prototype
  - compatible w/ information theoretic & expectation-based frameworks:
    - surprisal accounts
    - entropy reduction accounts

## METHODOLOGY

- Quantitative corpus linguistic methodologies can guide and inform psycholinguistic theorizing
- CFA is a promising tool for *schema entrenchment* measurement/detection

Introduction  
Data & Methods  
*Result & Discussion*

*Concluding*



seit 1558

Thank you for your attention.

And special thanks to Stefan Th. Gries (UCSB)

All CFA-computations done with “*HCFA 3.2*”  
*R-script for Windows*

Daniel Wiechmann