Workshop on "Pattern Finding"
Tilburg University
May 11, 2010

On how to ***measure*** *degrees of* ***entrenchment*** *of* ***schematic constructions***

Daniel Wiechmann
Friedrich Schiller University, Jena

▶ 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 and online processing

Hypothesis:

► Processing demand of a given construction C is a function of C‘s degree of entrenchment

Task:

► Measure degrees of *entrenchment of constructions*

1. Query corpus for target patterns

`n ~ 1000`

ICE-GB:S1A-001 #030:1:B	*This this is a dance group which doesnot exclude people*
ICE-GB:S1A-001 #032:1:B	*I enjoyed the time that I was given to to study and [...]*
ICE-GB:S1A-001 #038:1:B	*and the the opportunity that has arisen through the [...]*
ICE-GB:S1A-001 #039:1:B	*Uhm the movement language that 's beingdeveloped is [...]*

2. Describe data points

add	text.type	embedding	head	definiteness.head	concreteness.head
ICE.GB:S1A.014#129:1:C	DIRECT.CONV	CENTER	ALL	INDEFINITE.HEAD	ABSTRACT.HEAD
ICE.GB:S1A.020#290:1:C	DIRECT.CONV	CENTER	ALL	INDEFINITE.HEAD	ABSTRACT.HEAD
ICE.GB:S1A.037#139:1:B	DIRECT.CONV	CENTER	ALL	INDEFINITE.HEAD	ABSTRACT.HEAD
ICE.GB:S1A.015#237:1:A	DIRECT.CONV	CENTER	ALL	INDEFINITE.HEAD	CONCRETE.HEAD

Address: $\{\text{FEATURE}_1, \text{FEATURE}_2, \text{FEATURE}_3, \ldots, \text{FEATURE}_K\}$

3. Search for patterns in these descriptions

Address: $\{FEATURE_1, FEATURE_2, FEATURE_3, \ldots, FEATURE_K\}$

`n ~ 1000`

Task 1: Detecting entrenched patterns

- *Methods*
  - *association rule mining …*
  - *hierarchical configural frequency analysis*
  - *…*

Task 2: Structure detected patterns based on similarity (constructional network)

- *Methods*
  - *hierarchical agglomerative clustering*

Processing predictions can now be derived from network position and degrees of entrenchment

Search for patterns in these data

Address: $\{\text{FEATURE}_1, \text{FEATURE}_2, \text{FEATURE}_3, \ldots, \text{FEATURE}_K\}$

`n ~ 1000`

Hierarchical configural frequency analysis

e.g. `hcfa(cfa)`

- evaluates complex contingency tables (usual caveats apply)
- searches for ***types***, i.e. factor level combinations that occur with above chance frequencies

Configural Frequency Analysis

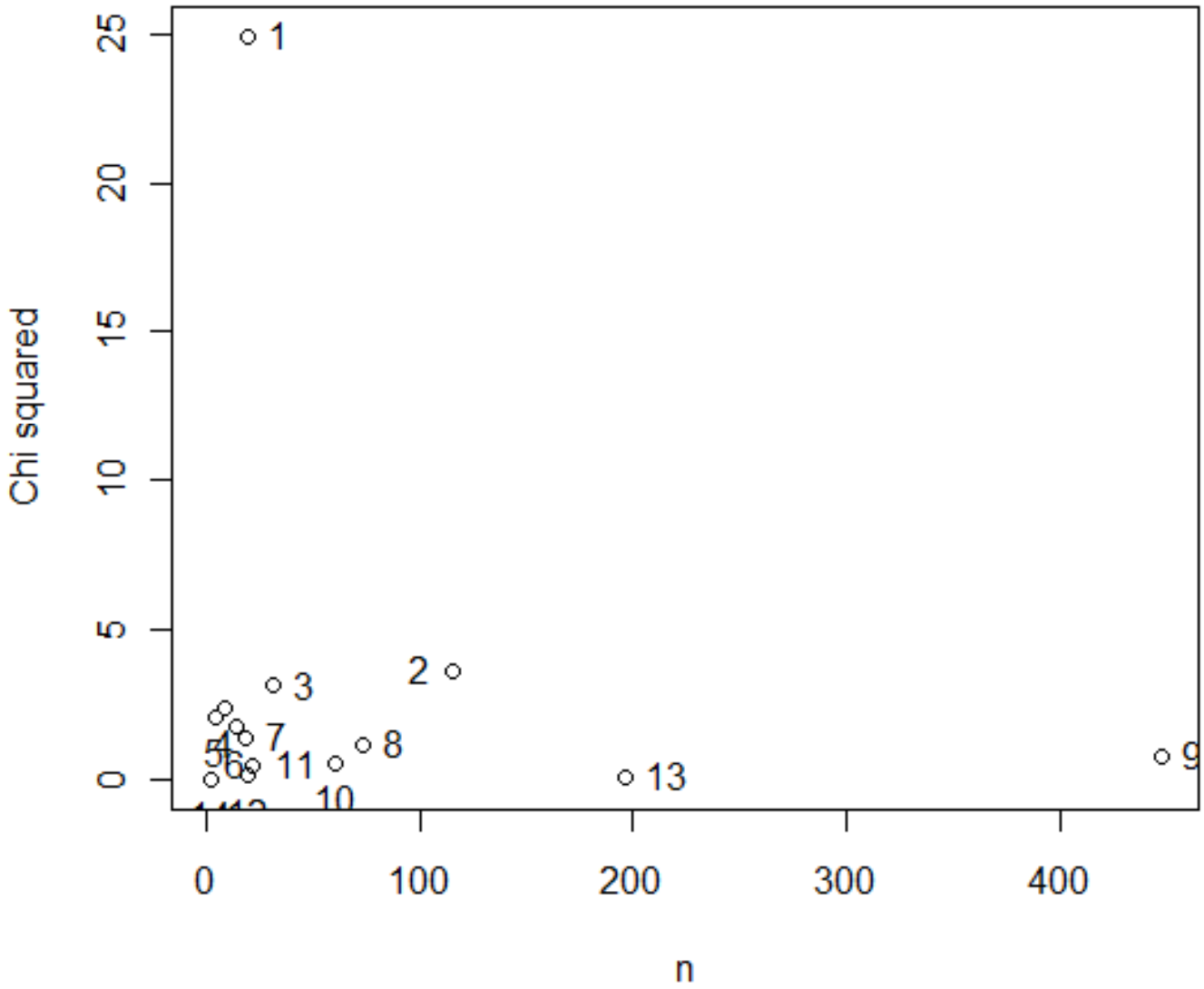
```
*** Analysis of configuration frequencies (CFA) ***

     label   n   expected           Q      chisq     p.chisq sig.chisq          z          p.z sig.z
1  B C E G 237 194.551042 0.0444282844 9.26190894 0.002339690      TRUE  3.3388386 0.0004206471  TRUE
2  A C E G 399 455.837616 0.0818794235 7.08698550 0.007764574     FALSE -3.4264875 0.9996942791  TRUE
3  A D F G  33  21.054705 0.0105809337 6.77711144 0.009233420     FALSE  2.6274520 0.0043013494 FALSE
4  B C F G  61  81.663400 0.0193416573 5.22848806 0.022219839     FALSE -2.3723728 0.9911628750 FALSE
5  A C E H  70  55.589953 0.0131669541 3.73537732 0.053271634     FALSE  1.9811900 0.0237849861 FALSE
6  A C F G 216 191.339246 0.0257241717 3.17840063 0.074617734     FALSE  1.9526336 0.0254315116 FALSE
7  A D E G  62  50.159738 0.0107654382 2.79490721 0.094564240     FALSE  1.7094949 0.0436796575 FALSE
8  A D E H   2   6.117041 0.0035991804 2.77095209 0.095989051     FALSE -1.6690626 0.9524475160 FALSE
9  B C F H   6   9.958951 0.0034726392 1.57378972 0.209657507     FALSE -1.2599753 0.8961608615 FALSE
10 B D E G  17  21.408126 0.0039058639 0.90767294 0.340732610     FALSE -0.9617123 0.8319029208 FALSE
11 B C E H  23  23.725737 0.0006443695 0.02219927 0.881558224     FALSE -0.1505553 0.5598367513 FALSE
12 A C F H  24  23.334054 0.0005910763 0.01900585 0.890349717     FALSE  0.1392820 0.4446136583 FALSE


Summary statistics:

Total Chi squared         =  43.3568
Total degrees of freedom  =  11
p                         =  4.561485e-11
Sum of counts             =  1150
```

Configural Frequency Analysis



Example:
English Relative Clause Constructions
2 clausal constituents (1 MC, 1 RC)

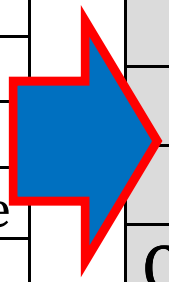
Attribute	Value	
medium	**spoken**	written
head type	**lexical**	pronominal
unique A	**present**	absent
content head	high	**low**
animacy head	animate	**inanimate**
definiteness head	**definite**	indefinite
SRC type	lexical	**pronominal**
relativizer	present	**absent**
embedding	right	**center**

e.g. ICE-GB:S1A-001 #039:1:B ***The only thing*** *[* ***you*** *could do] is [...]*

→ CFA detects a total of 7 types $(\text{Freq}_{obs} >_{sig} \text{Freq}_{exp})$ **in the data (n = 1000)**

Pattern

medium	spoken
head type	lexical
unique A	present
content head	low
animacy head	inanimate
definiteness head	definite
SRC type	pronominal
relativizer	absent
embedding	center



Stats for Pattern

Name	c.s3
Observed Freq	15
Expected Freq	1.1079
Contribution to Chisq	174.1948
Obs-exp	>
P.adj.bin	4.89E-10
Dec	***
Q	0.042

Deriving processing predictions from a similarity-based constructional network

Task 2: Build Constructional Network

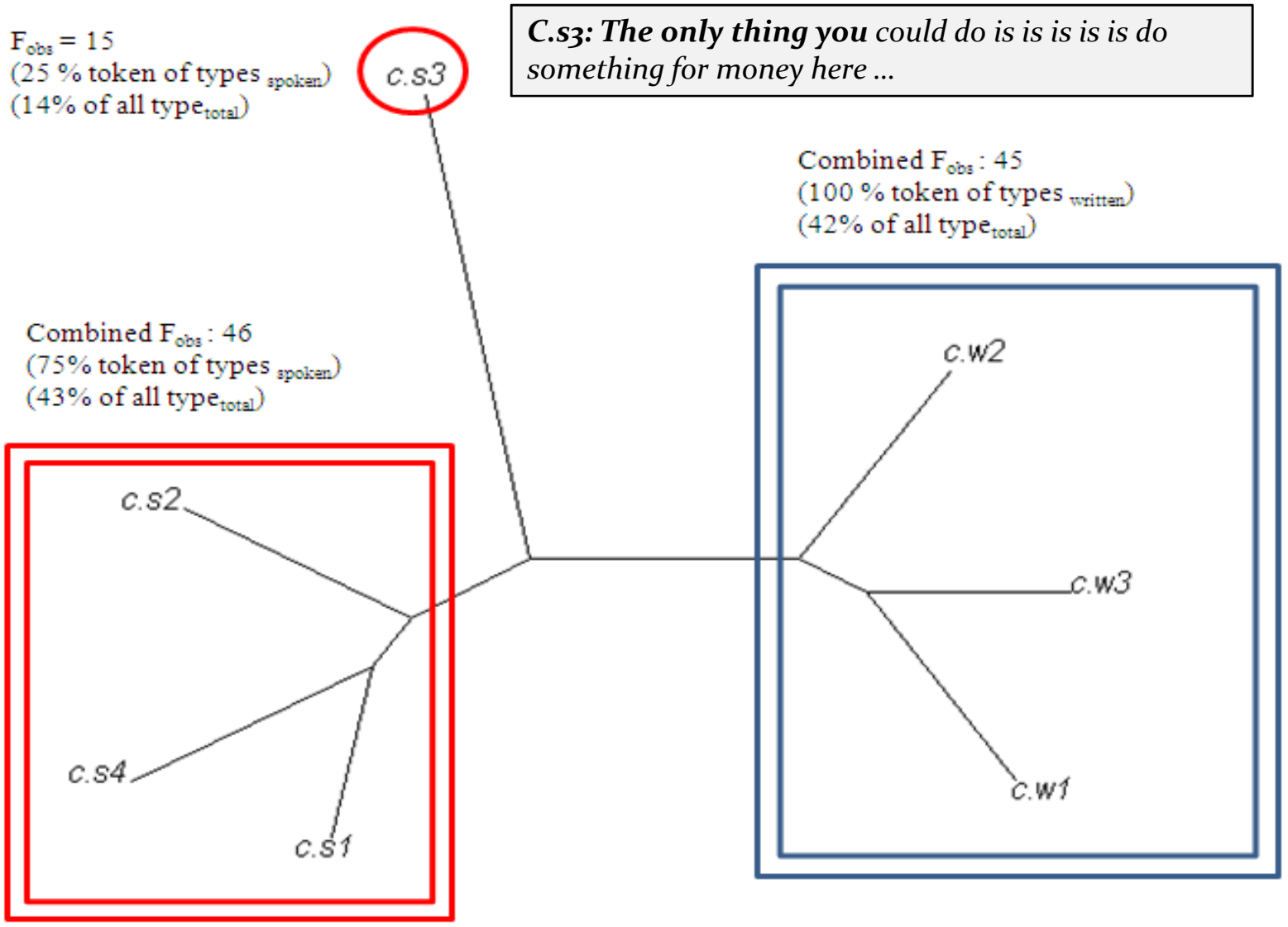
→Relate all detected types based on similarity

Method: hierarchical agglomerative clustering

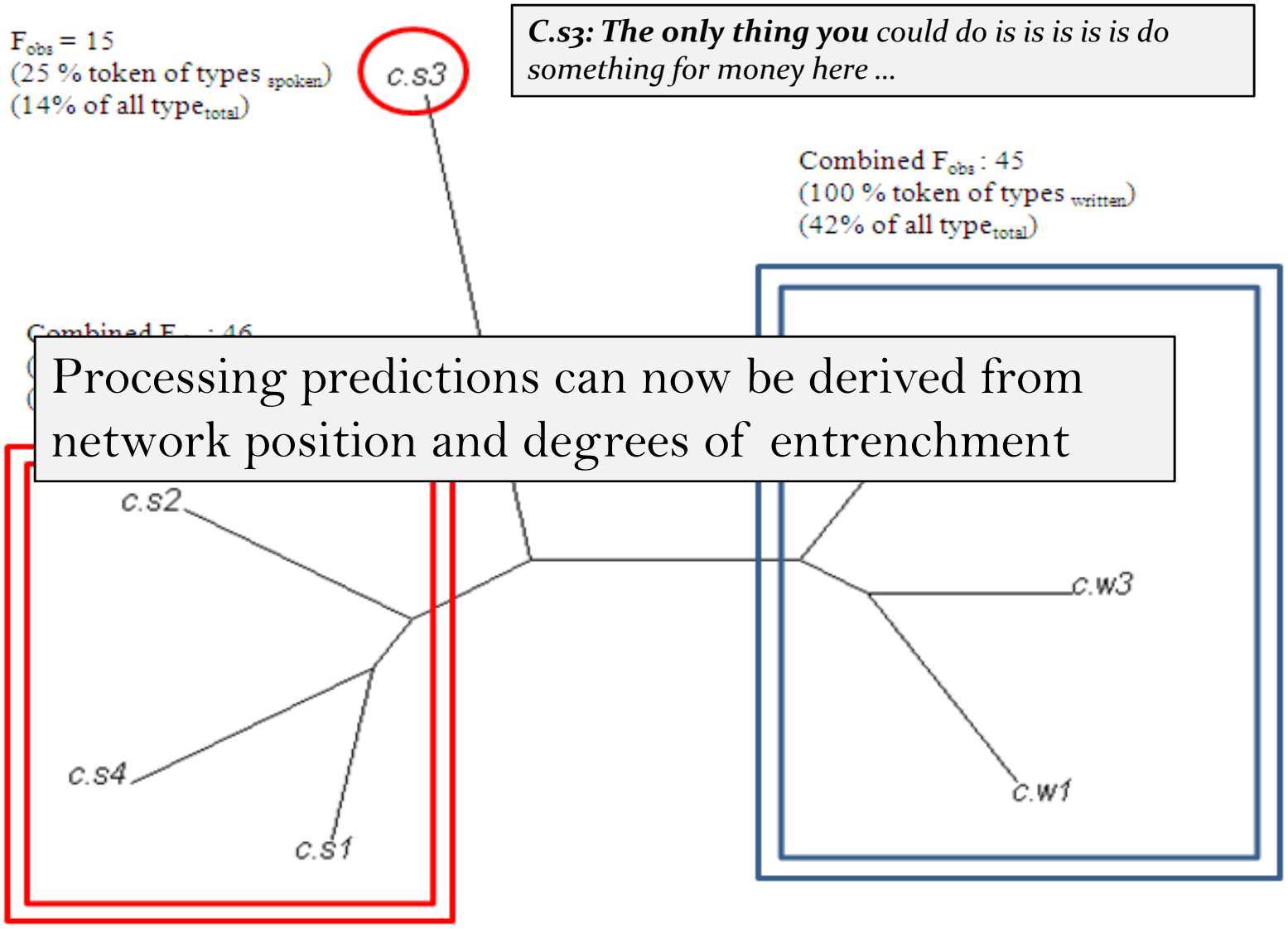
- similarity: Euclidean distance in *n*-dimensional space
- amalgamation: nearest neighbor (single linkage)

- output as unrooted tree (e.g. `nj(ape)`)

Similarity-based constructional network



Similarity-based constructional network



(Hierarchical) Configural Frequency Analysis

Attribute	Value
medium	spoken
head type	lexical
unique A	present
content head	*unspecified*
animacy head	inanimate
definiteness head	*unspecified*
SRC type	pronominal
relativizer	absent
embedding	center

We can look for higher level configurations (~ **more schematic constructions**) as well...

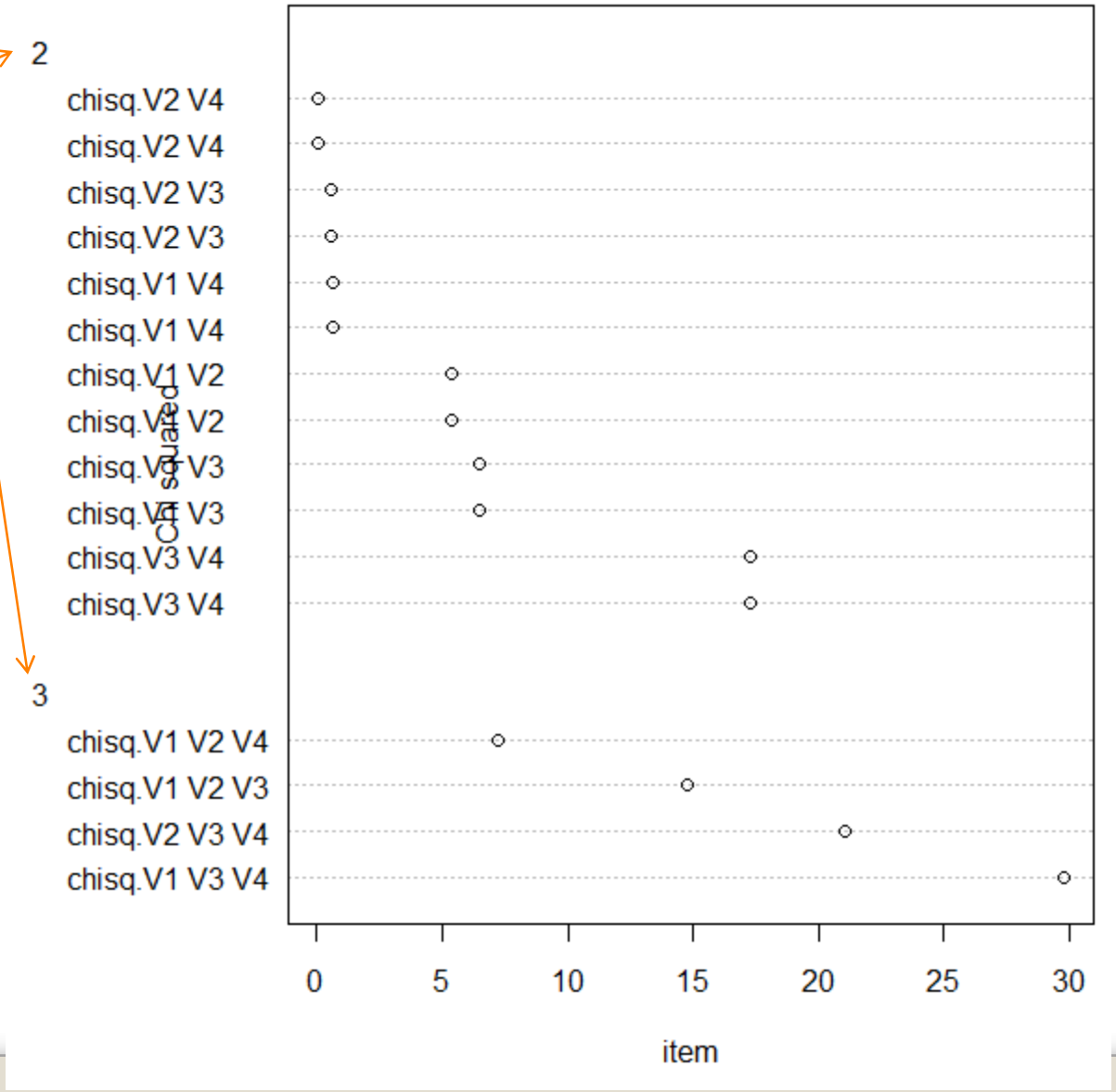
Hierarchical Configural Frequency Analysis

*** Hierarchical CFA ***

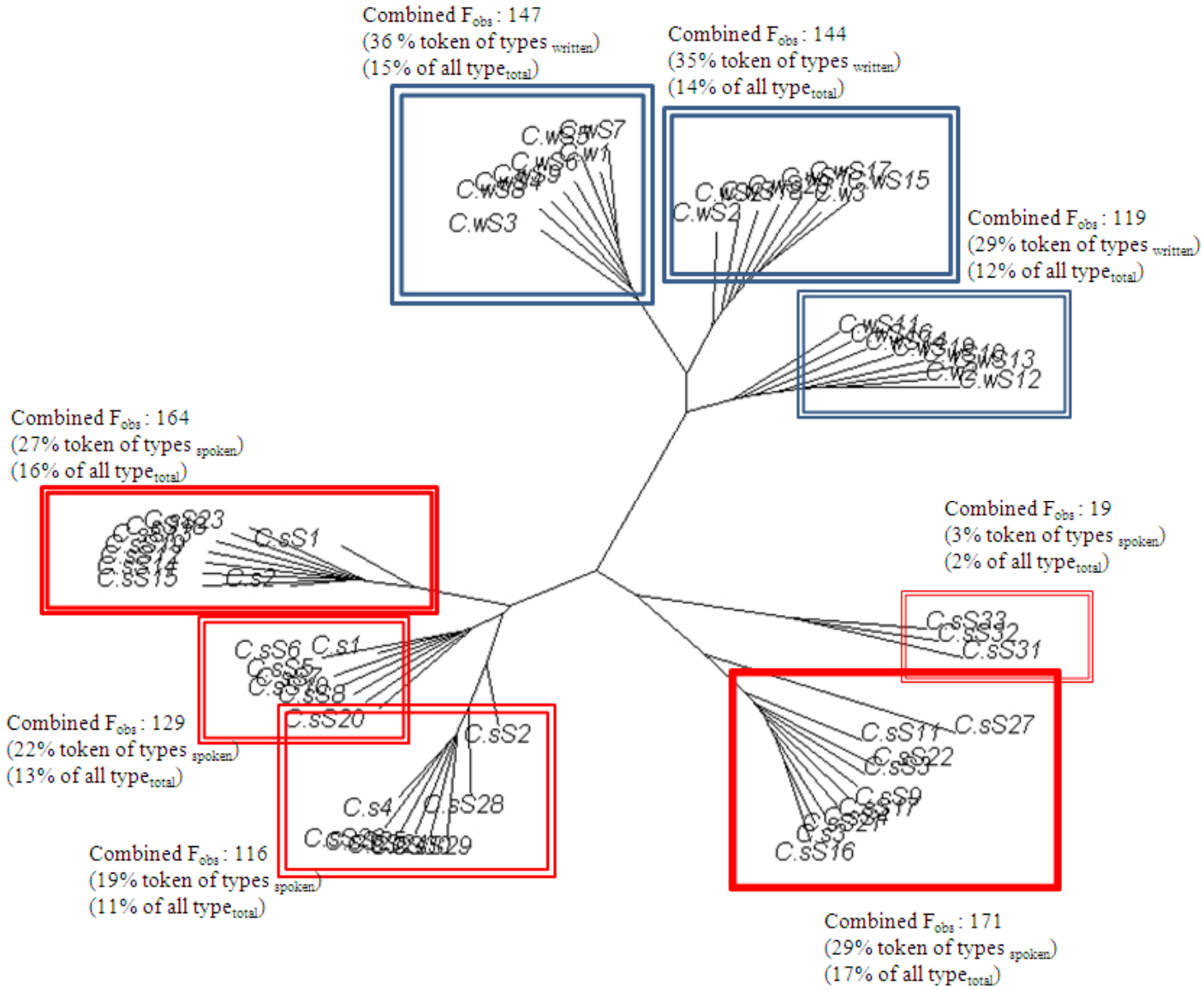
	Overall chi squared	df	p	order
V1 V3 V4	45.4701741	4	3.174446e-09	3
V1 V2 V3	28.0039946	4	1.244970e-05	3
V2 V3 V4	25.7175291	4	3.607730e-05	3
V2 V3	18.2073010	1	1.981179e-05	2
V2 V3	18.2073010	1	1.981179e-05	2
V1 V2 V4	13.6988207	4	8.321007e-03	3
V1 V2	6.8254186	1	8.986949e-03	2
V1 V2	6.8254186	1	8.986949e-03	2
V3 V4	4.3866035	1	3.622241e-02	2
V3 V4	4.3866035	1	3.622241e-02	2
V1 V4	2.5484019	1	1.104059e-01	2
V1 V4	2.5484019	1	1.104059e-01	2
V2 V4	2.3643996	1	1.241317e-01	2
V2 V4	2.3643996	1	1.241317e-01	2
V1 V3	0.8500972	1	3.565249e-01	2
V1 V3	0.8500972	1	3.565249e-01	2

Hierarchical Configural Frequency Analysis

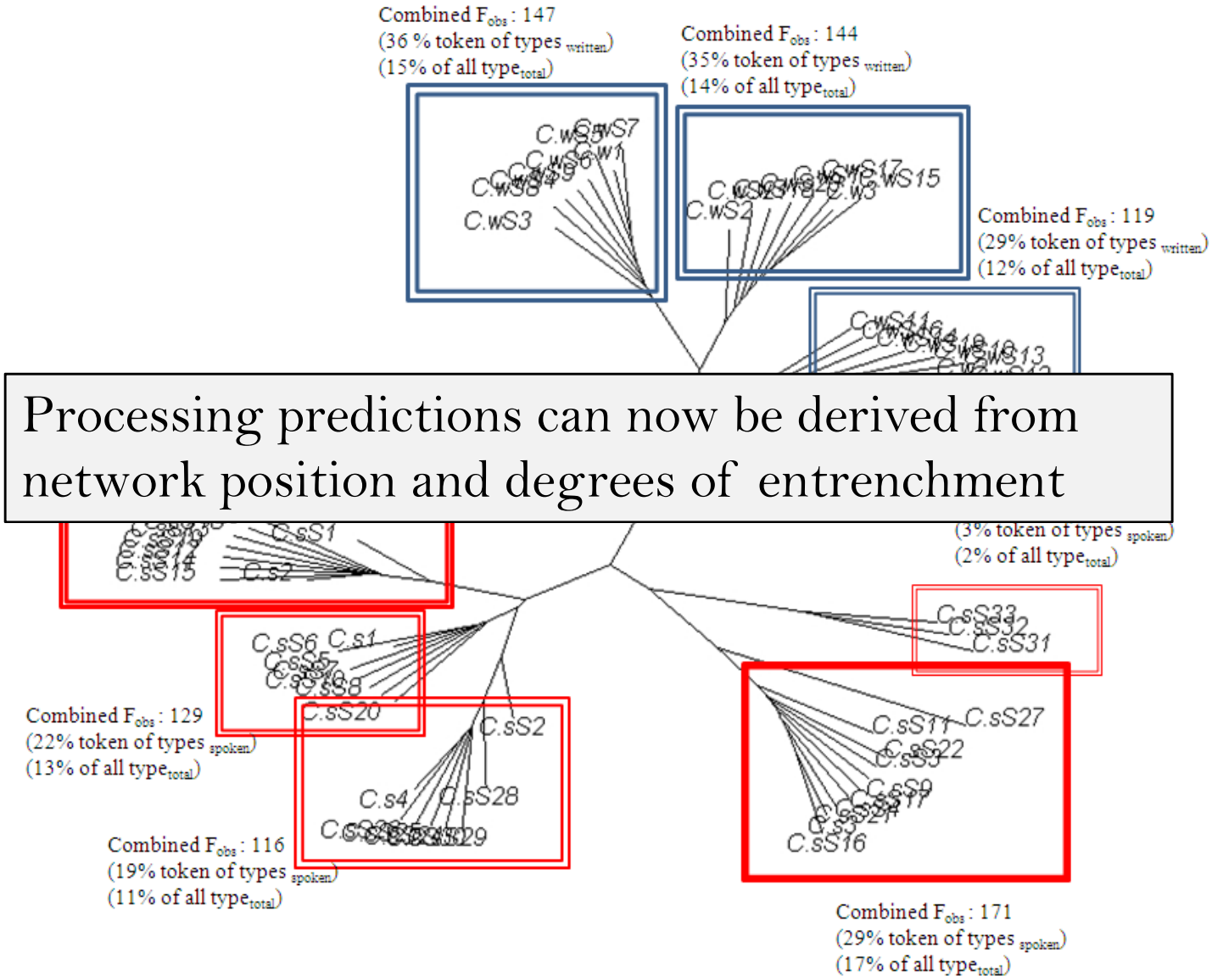
order



Similarity-based constructional network w/ higher level types



Similarity-based constructional network w/ higher level types



Thank you for your attention!

Presentation available from
www.daniel-wiechmann.net