FACULTY OF ARTS AND PHILOSOPHY

DEPARTMENT OF LINGUISTICS
RESEARCH GROUP: GLIMS

LANGUAGE PRODUCTIVITY @ WORK

THE PRODUCTIVITY OF MINIMIZING CONSTRUCTIONS IN PRESENT-DAY NETHERLANDIC DUTCH

*DAT BOEIT ME GEEN ****

Margot Van den Heede & Peter Lauwers

GHENT UNIVERSITY

BELGISCHE KRING VOOR LINGUÏSTIEK

THE MINIMIZING CONSTRUCTION

Het interesseert me geen zier.

‘It doesn’t interest me a ZIER’

Ik begrijp er geen hol van.

‘I don’t understand an arse of it’

Hij hoefde er geen seconde over na te denken.

‘He didn’t have to think a second about it’

GHENT UNIVERSITY

THE GOAL OF THIS TALK

On the basis of my dataset of minimizing nouns, we want to learn more about…

- the productivity of the different minimizing constructions in present-day Dutch
- productivity in general, namely the correlations between the different productivity measures

GHENT UNIVERSITY

OVERVIEW

1. The minimizing construction
2. General overview of the dataset
3. Productivity measures
4. Principal Components Analysis (PCA)
    - 4.1 Two macro-dimensions of productivity
    - 4.2 Productivity and semantics
5. Conclusion

GHENT UNIVERSITY

1. THE MINIMIZING CONSTRUCTION

A construction with different slots:

[geen] [MINIMIZING NOUN] [PREDICATE]

↓

the element of negation (quantifier)

GHENT UNIVERSITY

1. THE MINIMIZING CONSTRUCTION

A construction with different slots:

[geen] **[MINIMIZING NOUN]** [PREDICATE]

↓

a noun that denotes a small quantity

and that is used to reinforce sentential negation

GHENT UNIVERSITY

1. THE MINIMIZING CONSTRUCTION

A construction with different slots:

[geen] [MINIMIZING NOUN] **[PREDICATE]**

GHENT UNIVERSITY

1. THE MINIMIZING CONSTRUCTION

A construction with different slots:

[geen] [MINIMIZING NOUN] [PREDICATE]

NOT included (if the minimizer has a lower-level scope)

*U betaalt geen cent **extra*** ‘You don’t pay a penny extra’

*Hij heeft geen greintje **geduld*** ‘He doesn’t have a grain (of) patience’

GHENT UNIVERSITY

2. General overview of the dataset

Synchronic corpus research: Dutch Web 2014 (nlTenTen14), available on Sketch Engine

↓

Netherlandic Dutch subcorpus: 1.9 billion tokens

[lemma="geen|gene|genen|geene|geenen"] [] {0,2}

[word="bal"] within <s/>

GHENT UNIVERSITY

2. General overview of the dataset

Synchronic corpus research: Dutch Web 2014 (nlTenTen14), available on Sketch Engine

Selection based on literature (e.g. Hoeksema 2002), exploratory searches in the corpus (“*geen* N”) and searches based on frequent predicates

GHENT UNIVERSITY

2. General overview of the dataset

MINIMIZER

Type frequency	244
Hapax frequency	83

Combined with 641 different predicates

GHENT UNIVERSITY

2. General overview of the dataset

MINIMIZER

Type frequency	244
Hapax frequency	83

Combined with 641 different predicates

Entire dataset:

- 6224 tokens
- annotated a maximum of 100 tokens per minimizer, but a lot of them are less frequent

Focus of this talk:

- 4600 tokens
- the 46 minimizers with 100 tokens

GHENT UNIVERSITY

3. PRODUCTIVITY MEASURES

Productivity: a multi-faceted concept

- **Lexical scope**: “Productivity refers to the range of lexical items that may fill the slots of constructions” (Perek 2016: 66) TYPE FREQUENCY
- **Extensibility**: “the extensibility of a pattern to new types” (Barðdal 2008: 29) HAPAX FREQUENCY

GHENT UNIVERSITY

3. PRODUCTIVITY MEASURES

Productivity: a multi-faceted concept

- **Lexical scope**: “Productivity refers to the range of lexical items that may fill the slots of constructions” (Perek 2016: 66) TYPE FREQUENCY
- **Extensibility**: “the extensibility of a pattern to new types” (Barðdal 2008: 29) HAPAX FREQUENCY
- “Syntactic productivity is multidimensional – different aspects of productivity are not necessarily correlated (though they often are)” (Zeldes 2012: 135)

3. PRODUCTIVITY MEASURES

PRODUCTIVITY MEASURES

1. Type token ratio

2. Hapax token ratio

3. Hapax type ratio

GHENT UNIVERSITY

3. PRODUCTIVITY MEASURES

PRODUCTIVITY MEASURES

1. Type token ratio
21/100 = 0,21

2. Hapax token ratio
10/100 = 0,10

3. Hapax type ratio
10/21 = 0,48

SNARS	
begrijpen	28
snappen	20
geloven	17
interesseren	5
uitmaken	5
weten	5
geven om	2
kloppen	2
te maken hebben	2
terechtkomen	2
veranderen	2
bakken van	1
deugen	1
hebben aan	1
helpen	1
opschieten	1
terugzien	1
uitvoeren	1
vertrouwen	1
vinden aan	1
zien	1
Types	21
Hapaxes	10

GHENT UNIVERSITY

3. PRODUCTIVITY MEASURES

"ANTI-PRODUCTIVITY" MEASURES

4. Frequency of the most frequent predicate

5. Mean frequency of the 3 most frequent predicates

6. Standard deviation of the 3 most frequent predicates

GHENT UNIVERSITY

3. PRODUCTIVITY MEASURES

These are “measures of conventionalization, which aim to capture different facets of the distribution of high token frequency types”.

Van Wettere 2021: 405

(Bybee 2007; Clausner & Croft 1997; Hilpert 2015)

“ANTI-PRODUCTIVITY” MEASURES

4. Frequency of the most frequent predicate

5. Mean frequency of the 3 most frequent predicates

6. Standard deviation of the 3 most frequent predicates

GHENT UNIVERSITY

3. PRODUCTIVITY MEASURES

SNARS	
begrijpen	28
snappen	20
geloven	17
interesseren	5
uitmaken	5
weten	5
geven om	2
kloppen	2
te maken hebben	2
terechtkomen	2
veranderen	2
bakken van	1
deugen	1
hebben aan	1
helpen	1
opschieten	1
terugzien	1
uitvoeren	1
vertrouwen	1
vinden aan	1
zien	1
Types	21
Hapaxes	10

"ANTI-PRODUCTIVITY" MEASURES

4. Frequency of the most frequent predicate
28

5. Mean frequency of the 3 most frequent predicates
(28+20+17)/3 = 21,7

6. Standard deviation of the 3 most frequent predicates
5,7

$$\text{Standard Deviation} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}}$$

GHENT UNIVERSITY

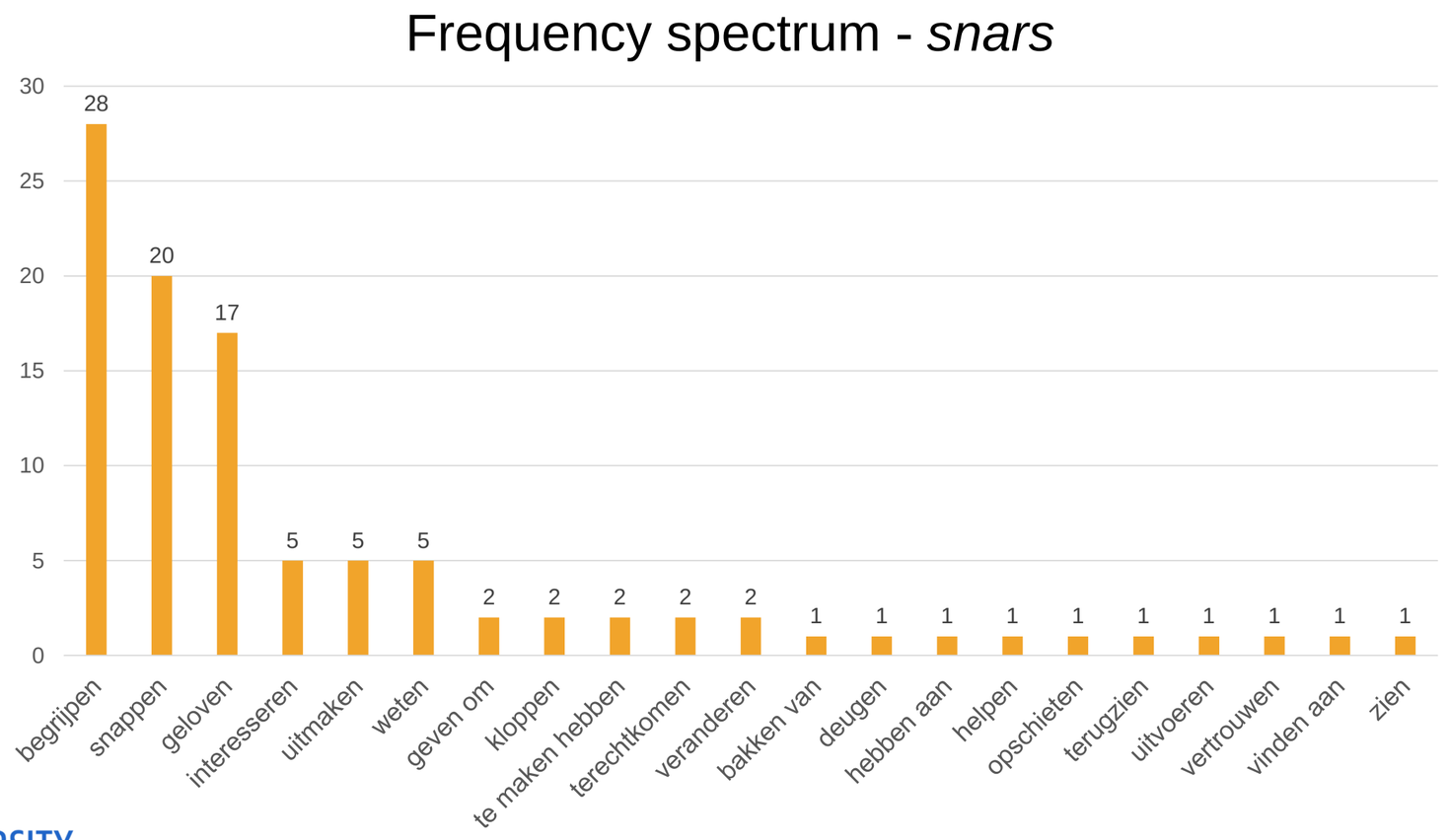
3. PRODUCTIVITY MEASURES

7. ALPHA
= the slope of the regression line

GHENT UNIVERSITY

3. PRODUCTIVITY MEASURES

7. ALPHA
= the slope of the regression line

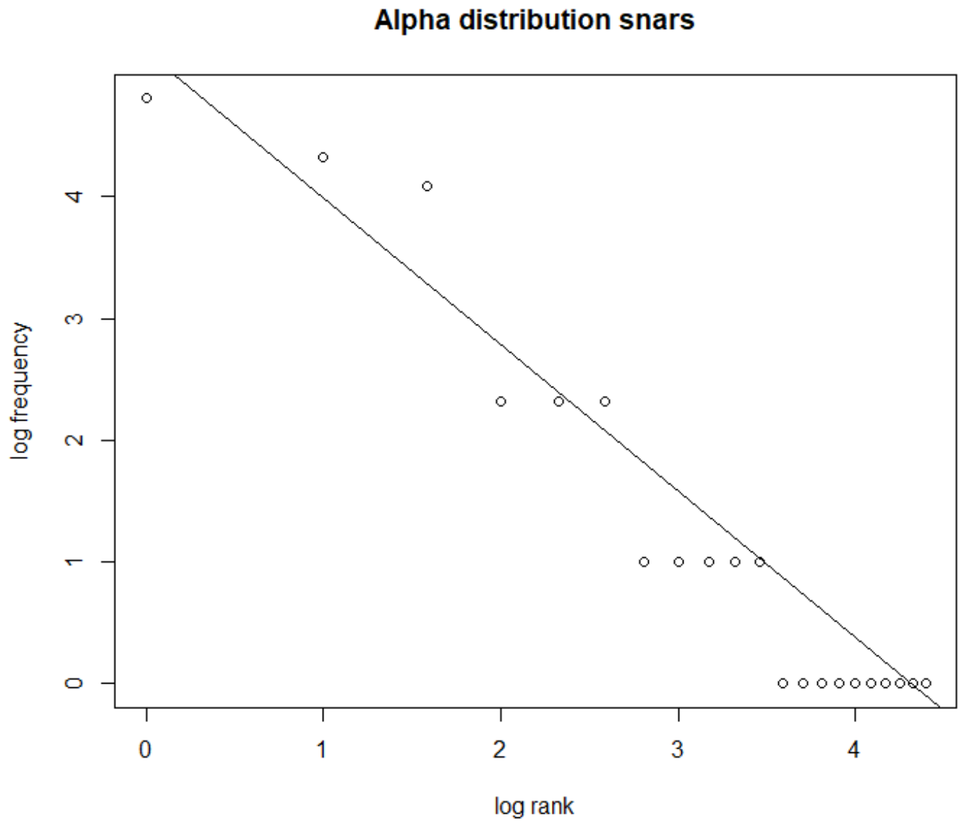


1) Frequency spectrum

GHENT UNIVERSITY

3. PRODUCTIVITY MEASURES

7. ALPHA
= the slope of the regression line



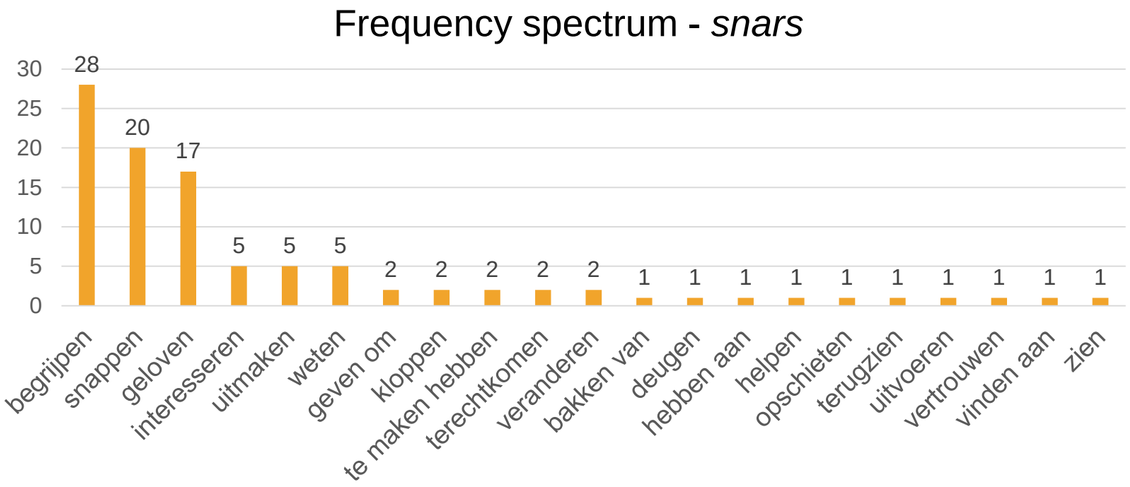
1) Frequency spectrum
2) Alpha distribution

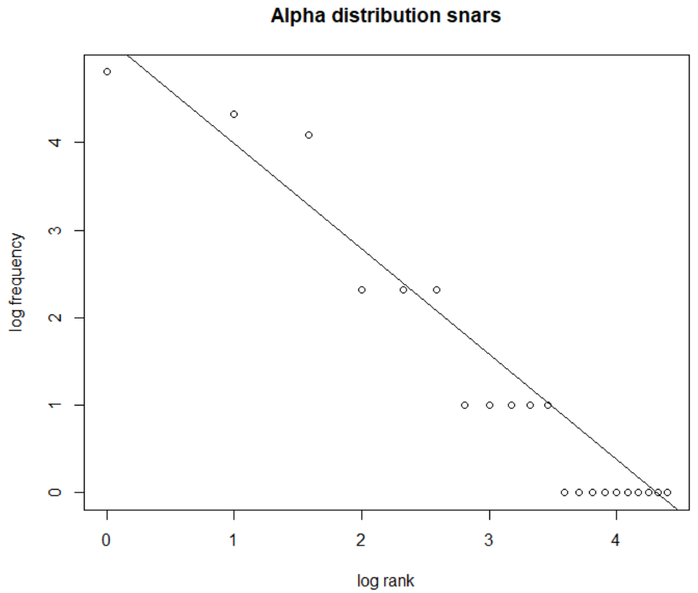
(van Egmond 2013)

→ Alpha captures the relation between the logtransformed frequencies and the ranks of the Zipfian curve

GHENT UNIVERSITY

3. PRODUCTIVITY MEASURES

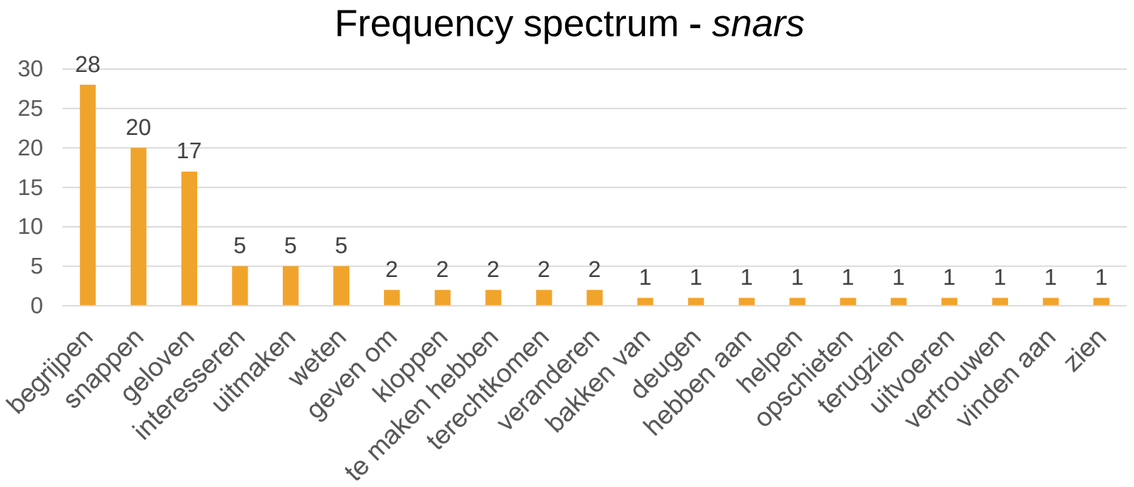


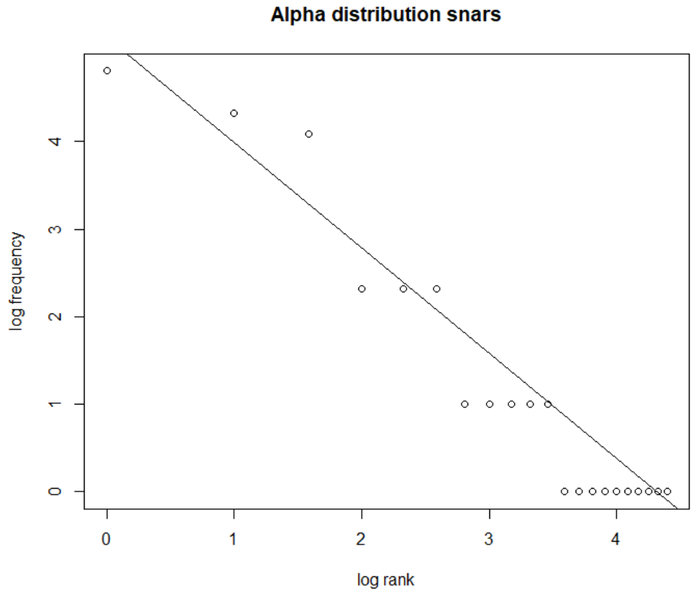


Alpha = 1,2

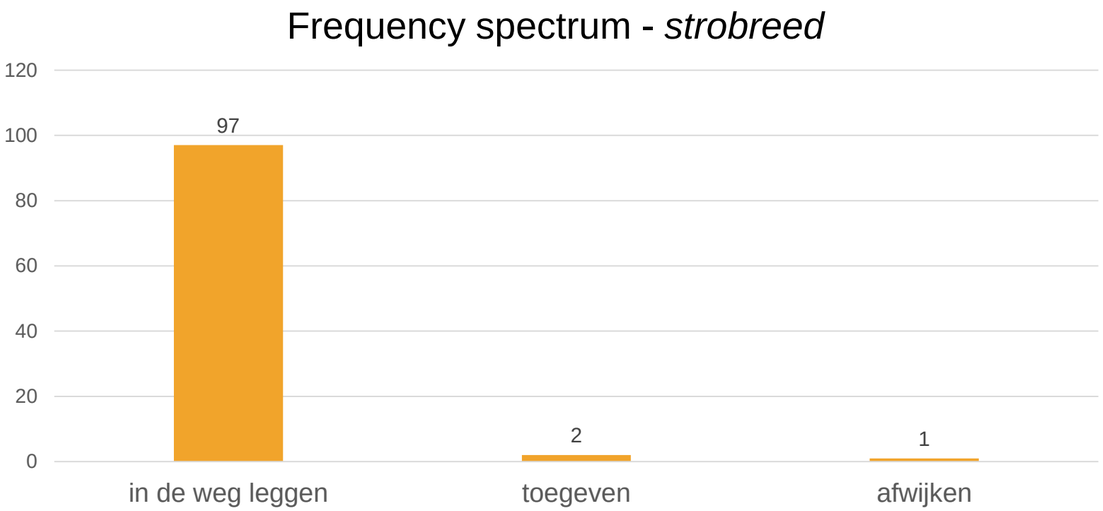
GHENT UNIVERSITY

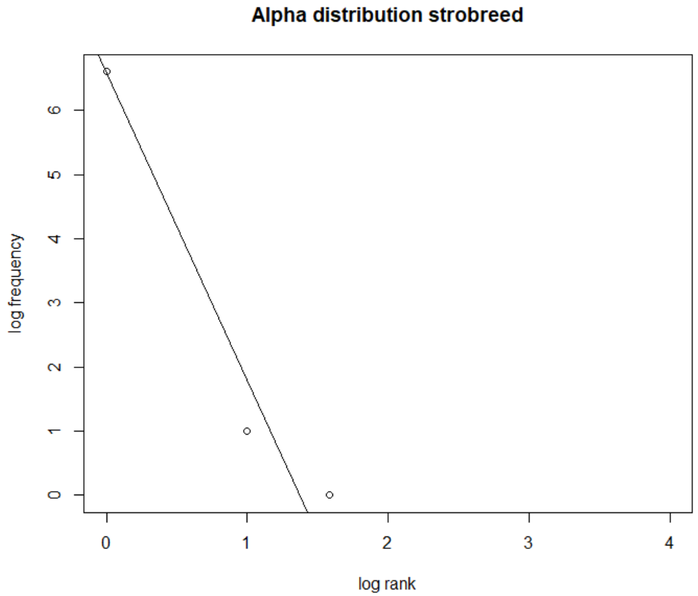
3. PRODUCTIVITY MEASURES





Alpha = 1,2





Alpha = 4,8

GHENT UNIVERSITY

3. PRODUCTIVITY MEASURES

Correlations between the productivity measures (Pearson correlation)

Positive correlation
No correlation
Negative correlation

	ZipfAlpha	MeanFrTop3	FrTop1	SDTop3	HapaxTypeRatio	TypeTokenRatio
HapaxTokenRatio	-0,5	-0,6	-0,53	-0,48	0,73	0,93
TypeTokenRatio	-0,69	-0,83	-0,77	-0,71	0,46	1
HapaxTypeRatio	-0,14	-0,06	-0,02	-0,01	1	0,46
SDTop3	0,93	0,93	0,98	1	-0,01	-0,71
FrTop1	0,88	0,94	1	0,98	-0,02	-0,77
MeanFrTop3	0,91	1	0,94	0,93	-0,06	-0,83

GHENT UNIVERSITY

R package: “corrplot”

3. PRODUCTIVITY MEASURES

Correlations between the productivity measures (Pearson correlation)

Productivity measures

	ZipfAlpha	MeanFrTop3	FrTop1	SDTop3	HapaxTypeRatio	TypeTokenRatio
HapaxTokenRatio	-0,5	-0,6	-0,53	-0,48	0,73	0,93
TypeTokenRatio	-0,69	-0,83	-0,77	-0,71	0,46	1
HapaxTypeRatio	-0,14	-0,06	-0,02	-0,01	1	0,46
SDTop3	0,93	0,93	0,98	1	-0,01	-0,71
FrTop1	0,88	0,94	1	0,98	-0,02	-0,77
MeanFrTop3	0,91	1	0,94	0,93	-0,06	-0,83

Positive correlation
No correlation
Negative correlation

"Anti-productivity" measures

GHENT UNIVERSITY

R package: "corrplot"

3. PRODUCTIVITY MEASURES

Correlations between the productivity measures (Pearson correlation)

Positive correlation
No correlation
Negative correlation

	ZipfAlpha	MeanFrTop3	FrTop1	SDTop3	HapaxTypeRatio	TypeTokenRatio
HapaxTokenRatio	-0,5	-0,6	-0,53	-0,48	0,73	0,93
TypeTokenRatio	-0,69	-0,83	-0,77	-0,71	0,46	1
HapaxTypeRatio	-0,14	-0,06	-0,02	-0,01	1	0,46
SDTop3	0,93	0,93	0,98	1	-0,01	-0,71
FrTop1	0,88	0,94	1	0,98	-0,02	-0,77
MeanFrTop3	0,91	1	0,94	0,93	-0,06	-0,83

GHENT UNIVERSITY

R package: “corrplot”

3. PRODUCTIVITY MEASURES

Correlations between the productivity measures (Pearson correlation)

Positive correlation
No correlation
Negative correlation

	ZipfAlpha	MeanFrTop3	FrTop1	SDTop3	HapaxTypeRatio	TypeTokenRatio
HapaxTokenRatio	-0,5	-0,6	-0,53	-0,48	0,73	0,93
TypeTokenRatio	-0,69	-0,83	-0,77	-0,71	0,46	1
HapaxTypeRatio	-0,14	-0,06	-0,02	-0,01	1	0,46
SDTop3	0,93	0,93	0,98	1	-0,01	-0,71
FrTop1	0,88	0,94	1	0,98	-0,02	-0,77
MeanFrTop3	0,91	1	0,94	0,93	-0,06	-0,83

GHENT UNIVERSITY

R package: “corrplot”

OVERVIEW

1. The minimizing construction
2. General overview of the dataset
3. Productivity measures
4. Principal Components Analysis (PCA)
    - 4.1 Two macro-dimensions of productivity
    - 4.2 Productivity and semantics
5. Conclusion

GHENT UNIVERSITY

4. Principal Components Analysis (PCA)

- a dimensionality reduction method, which allows to reorient the data so that the first few dimensions or principal components account for as much of the available information as possible

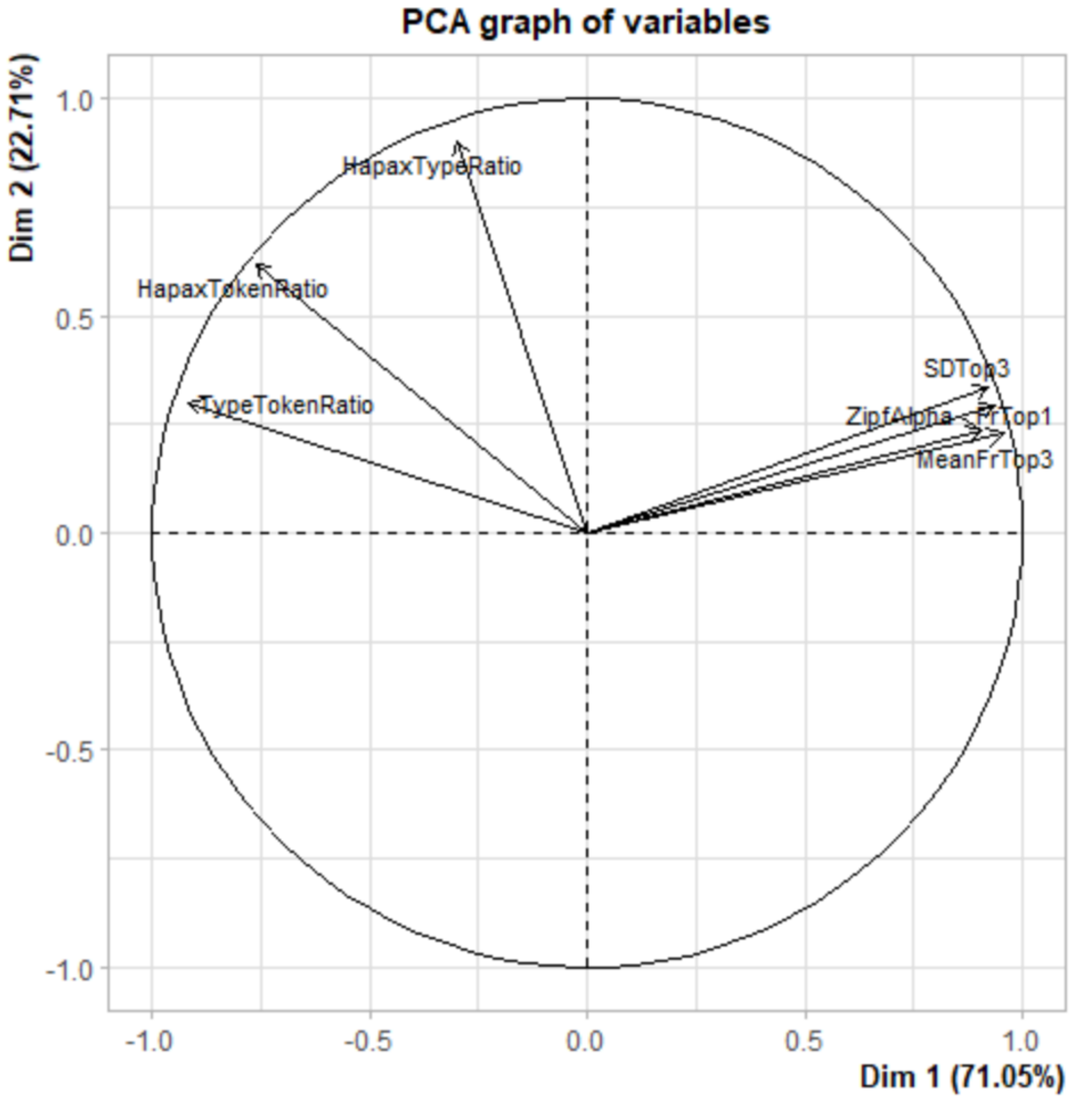
- it allows to detect and to have a more global view on the correlations between the productivity measures

4. Principal Components Analysis (PCA)

GHENT UNIVERSITY

R Package: “FactoMineR”, “factoextra”

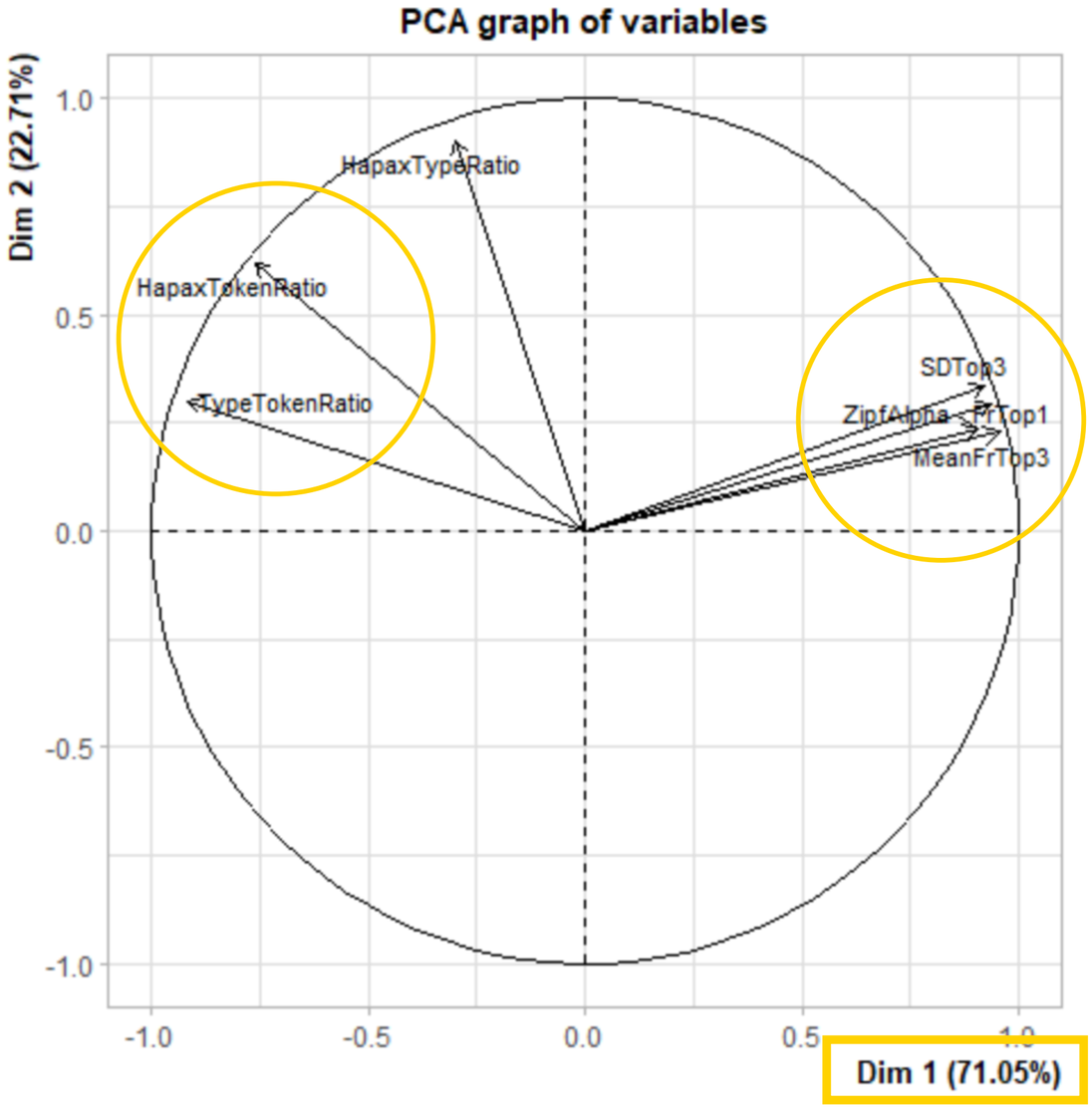
4.1 TWO MACRO-DIMENSIONS OF PRODUCTIVITY



GHENT UNIVERSITY

R Package: “FactoMineR”, “factoextra”

4.1 TWO MACRO-DIMENSIONS OF PRODUCTIVITY



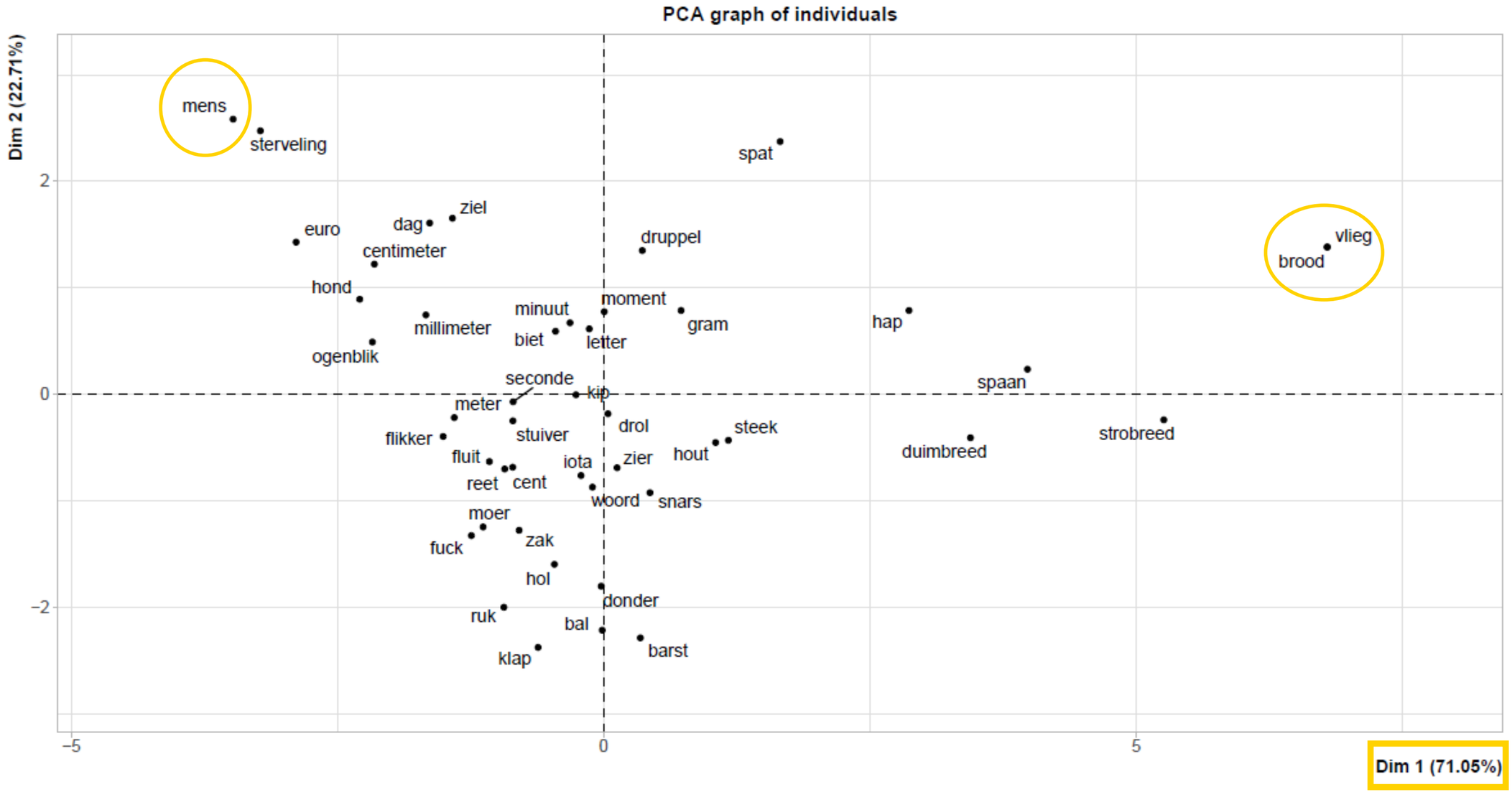
Dimension 1 correlations

MeanFrTop3	0,96
FrTop1	0,94
SDTop3	0,92
ZipfAlpha	0,91
VS.	
HapaxTokenRatio	-0.76
TypeTokenRatio	-0,92

GHENT UNIVERSITY

R Package: “FactoMineR”, “factoextra”

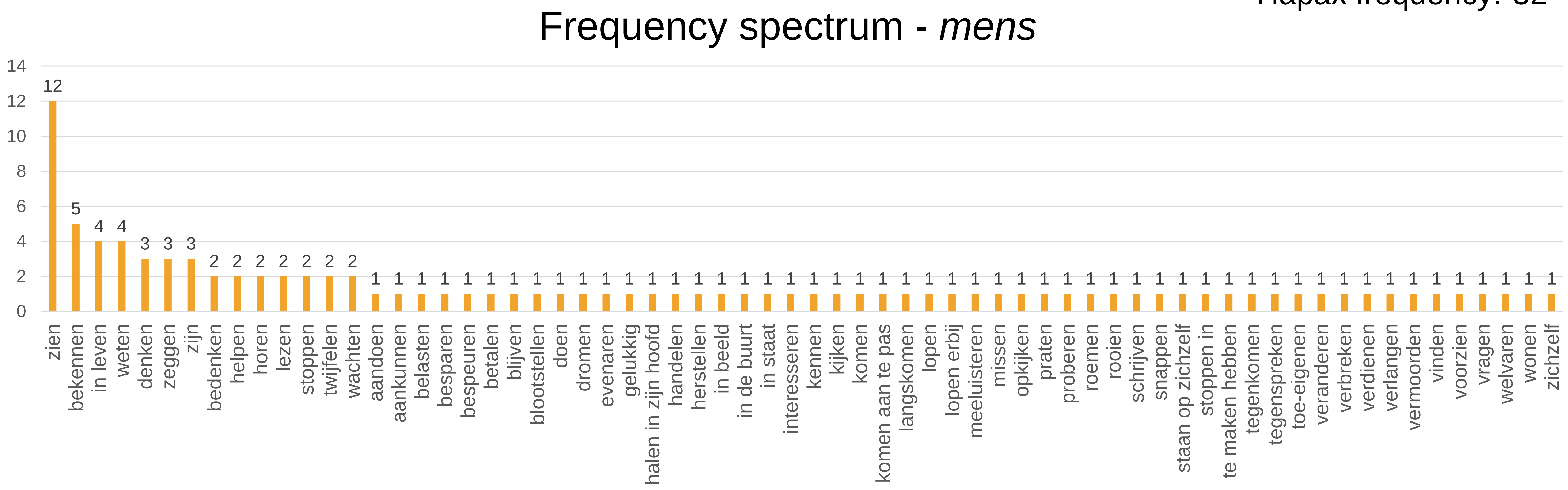
4.1 TWO MACRO-DIMENSIONS OF PRODUCTIVITY



GHENT UNIVERSITY

4.1 Two macro-dimensions of productivity

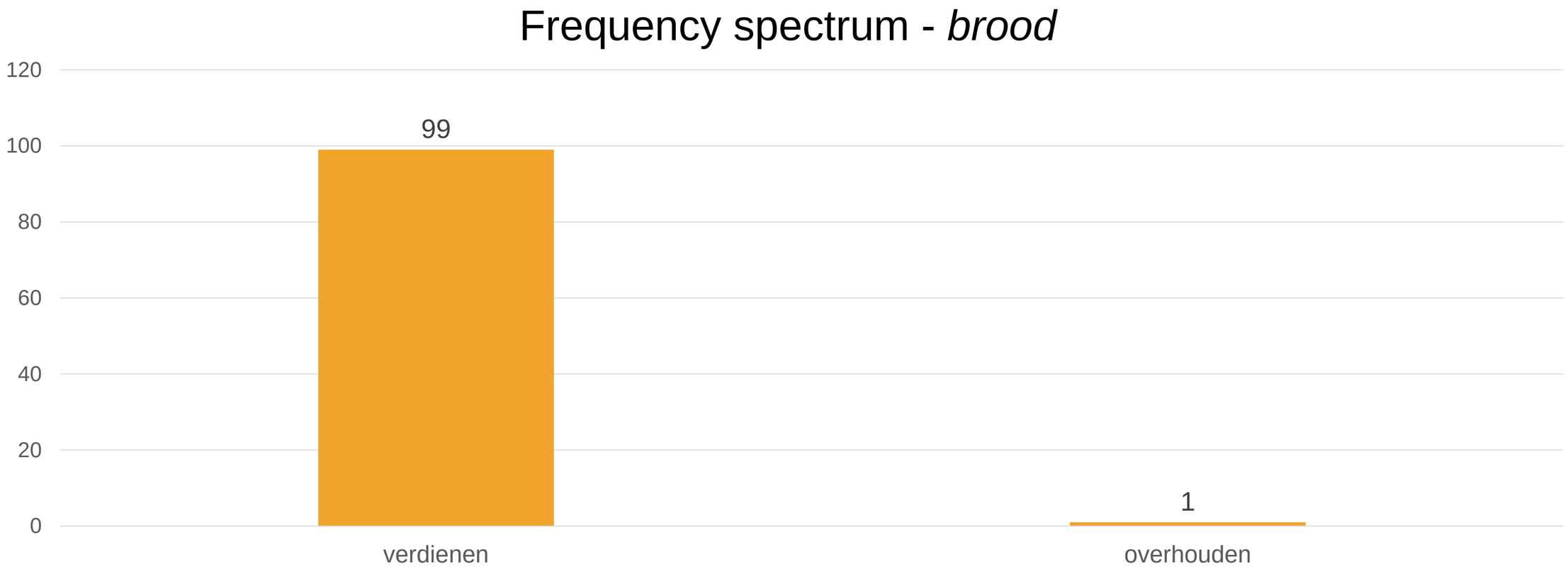
Type frequency: 66
Hapax frequency: 52



GHENT UNIVERSITY

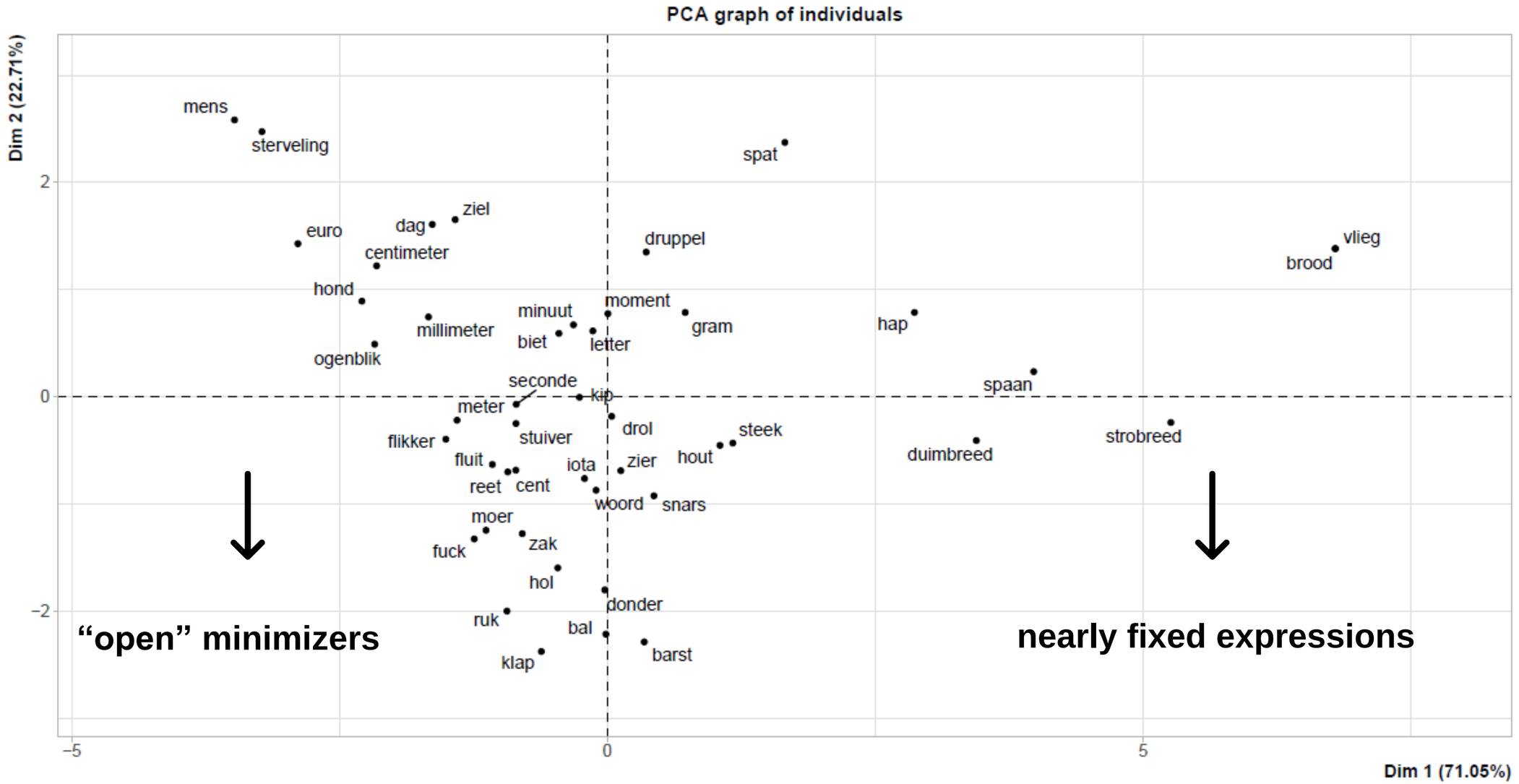
4.1 TWO MACRO-DIMENSIONS OF PRODUCTIVITY

Type frequency: 2
Hapax frequency: 1



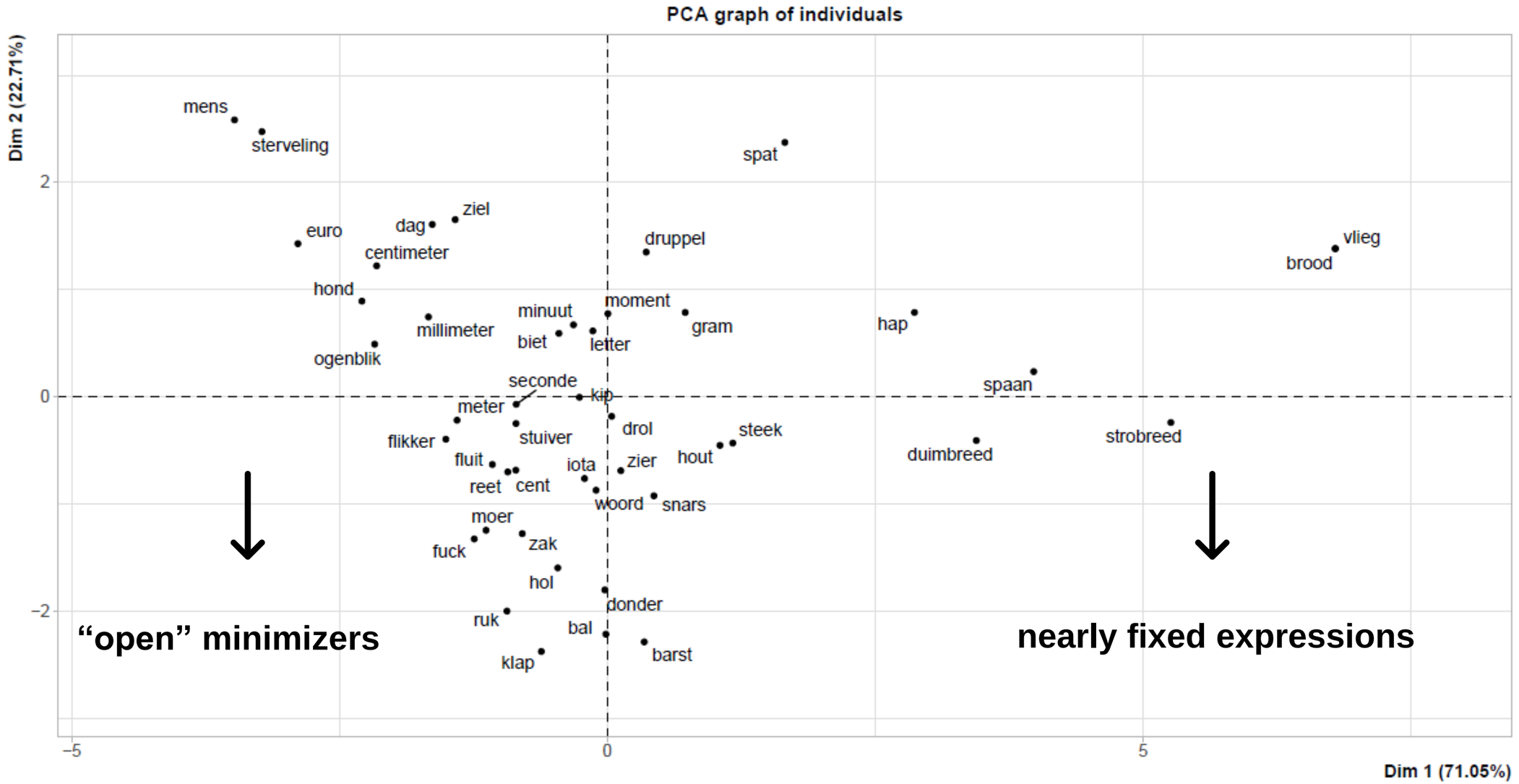
GHENT UNIVERSITY

4.1 TWO MACRO-DIMENSIONS OF PRODUCTIVITY



GHENT UNIVERSITY

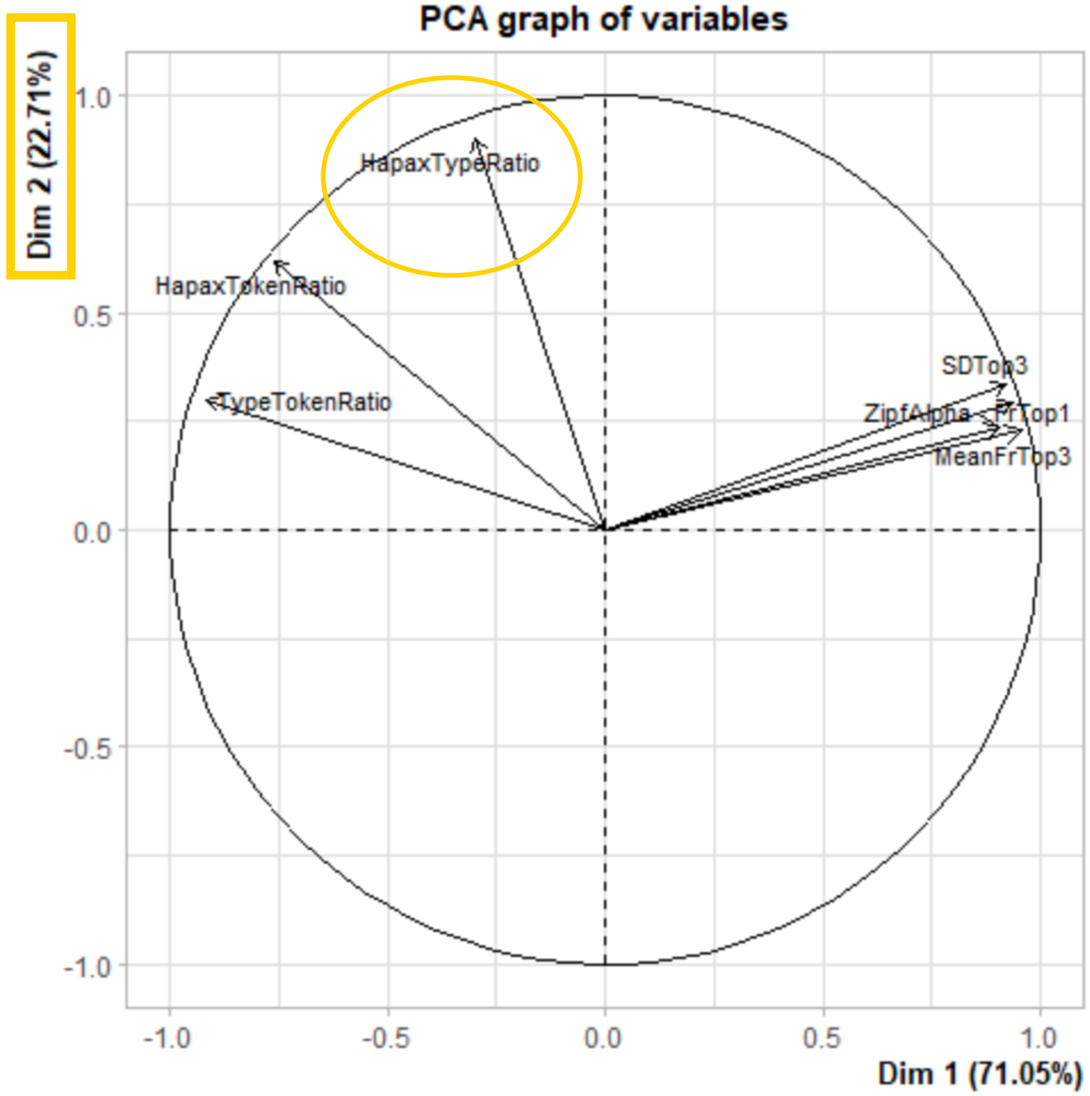
4.1 TWO MACRO-DIMENSIONS OF PRODUCTIVITY



→ very token frequent predicates can detract from productivity

(Barðdal 2008: 49; Van Wettere 2018: 610; De Smet 2020)

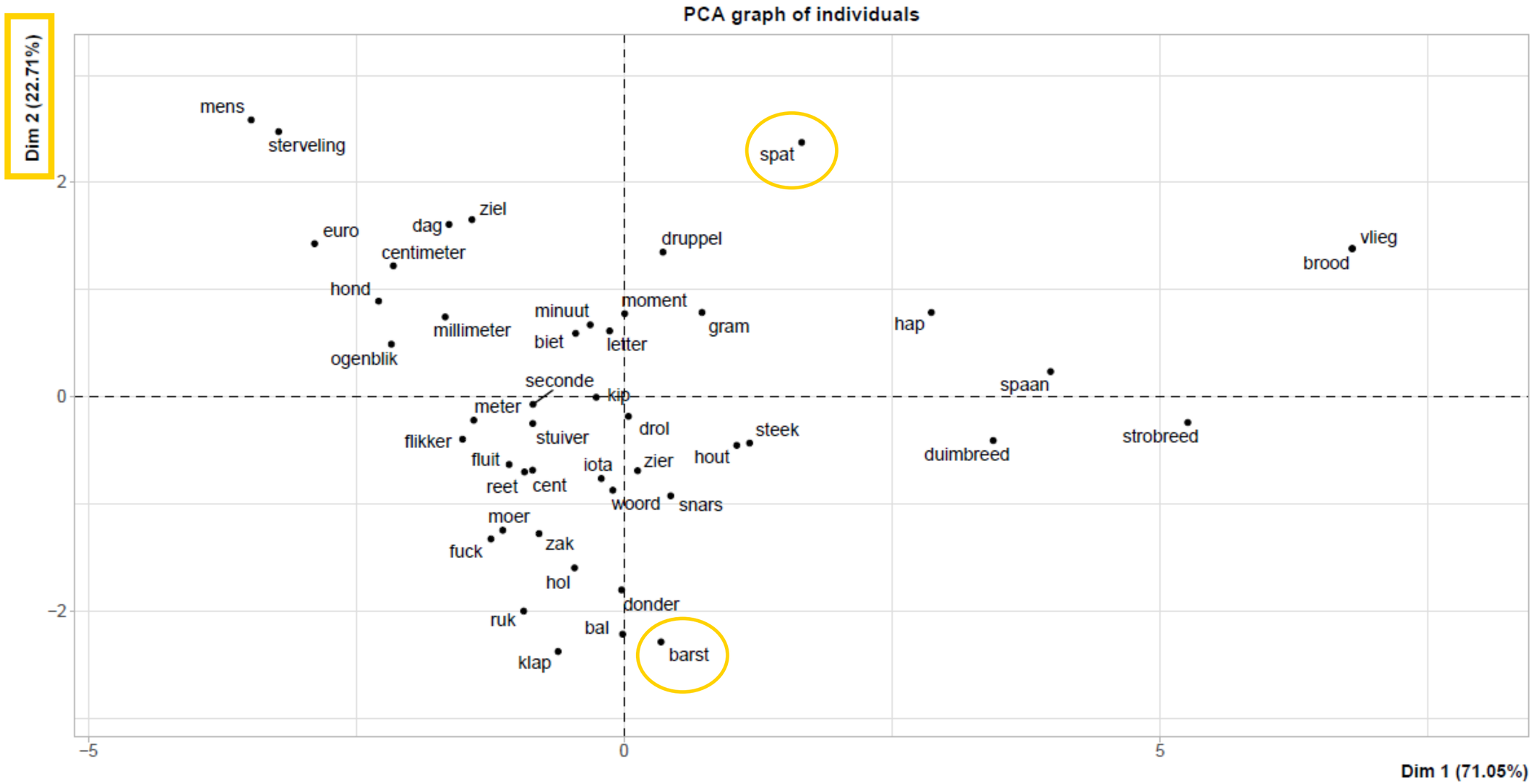
4.1 TWO MACRO-DIMENSIONS OF PRODUCTIVITY



Dimension 2 correlations

HapaxTypeRatio	0,90

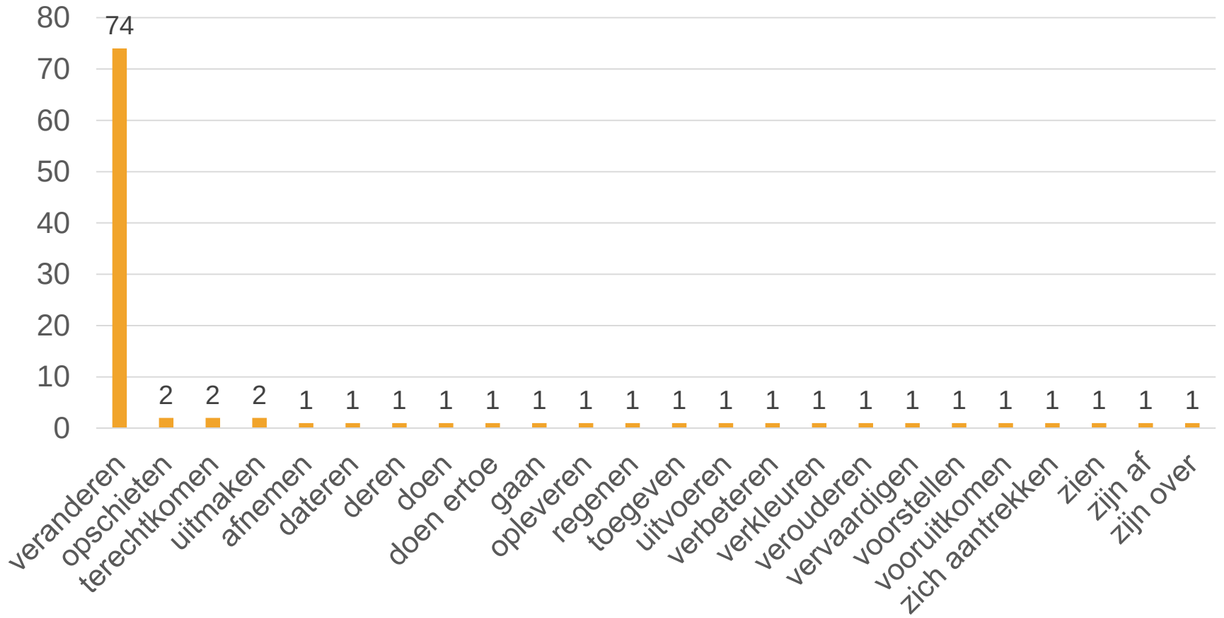
4.1 TWO MACRO-DIMENSIONS OF PRODUCTIVITY



GHENT UNIVERSITY

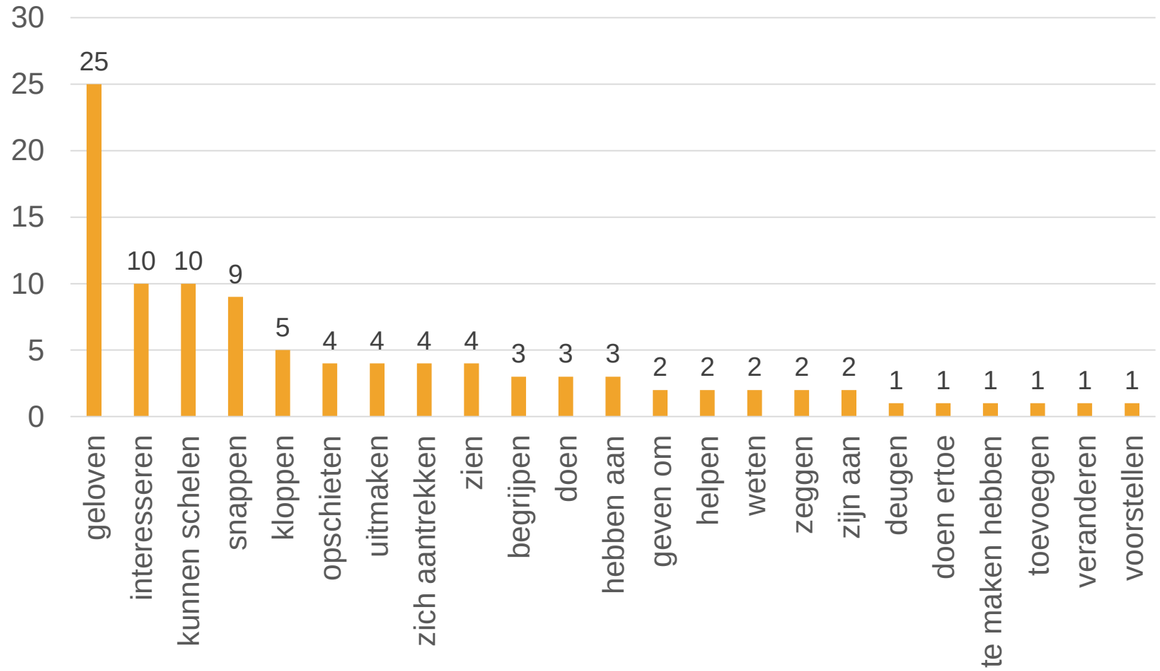
4.1 TWO MACRO-DIMENSIONS OF PRODUCTIVITY

Frequency spectrum - *spat*



Type frequency:	24
Hapax frequency:	20
Hapax type ratio:	0,83

Frequency spectrum - *barst*



Type frequency:	23
Hapax frequency:	6
Hapax type ratio:	0,26

GHENT UNIVERSITY

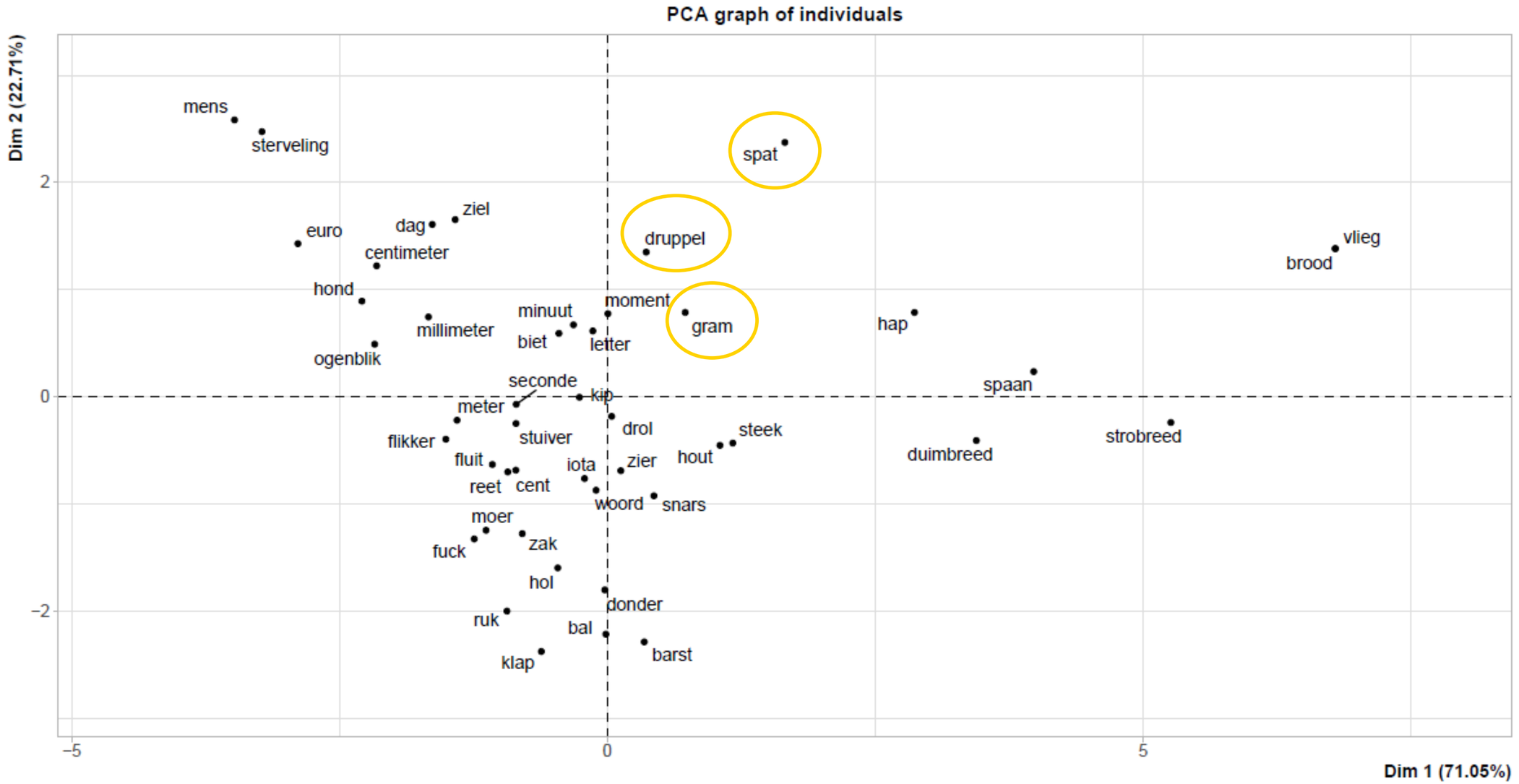
4.1 Two macro-dimensions of productivity

"Since the hapax type ratio focalizes the dichotomy between hapax and non-hapax, this measure downplays the impact of high token frequencies among the non-hapax. In other words, whereas type frequency and hapax frequency […] are very negatively impacted by a couple of high token frequency types, the hapax type ratio is more capable of detecting productivity in certain subdomains in spite of a (limited) number of high token frequency types. Therefore it is less in direct opposition with the anti-productivity measures than type and hapax frequency."

(Van Wettere 2021: 410-411)

GHENT UNIVERSITY

4.1 TWO MACRO-DIMENSIONS OF PRODUCTIVITY



GHENT UNIVERSITY

4.1 TWO MACRO-DIMENSIONS OF PRODUCTIVITY

Minimizer	**Hapax type ratio**	**Most frequent predicate(s)**	**Hapaxes belonging to the same semantic field**
geen spat 'not a splash'	0,83	*veranderen* 'to change' (frequency: 74/100)	*verbeteren* 'to improve' *verkleuren* 'to colour' *verouderen* 'to age'
geen druppel 'not a drop'	0,71	*drinken* 'to drink' (frequency: 49/100)	*innemen* 'to take' *inslikken* 'to swallow' *nuttigen* 'to consume' *ophebben* 'to have drunk' *proeven* 'to taste'
geen gram 'not a gram'	0,71	*aankomen* 'to gain weight' (frequency: 42/100) *afvallen* 'to lose weight' (frequency: 21/100)	*aankrijgen* *bijkrijgen* *erbij zijn* *afkrijgen* *kwijt zijn* *kwijtraken*

GHENT UNIVERSITY

4.1 TWO MACRO-DIMENSIONS OF PRODUCTIVITY

Minimizer	Hapax type ratio	Most frequent predicate(s)	Hapaxes belonging to the same semantic field
geen spat 'not a splash'	0,83	*veranderen* 'to change' (frequency: 74/100)	*verbeteren* 'to improve' *verkleuren* 'to colour' *verouderen* 'to age'
geen druppel 'not a drop'	0,71	*drinken* 'to drink' (frequency: 49/100)	*innemen* 'to take' *inslikken* 'to swallow' *nuttigen* 'to consume' *ophebben* 'to have drunk' *proeven* 'to taste'
geen gram 'not a gram'	0,71	*aankomen* 'to gain weight' (frequency: 42/100) *afvallen* 'to lose weight' (frequency: 21/100)	*aankrijgen* *bijkrijgen* *erbij zijn* *afkrijgen* *kwijt zijn* *kwijtraken*

GHENT UNIVERSITY

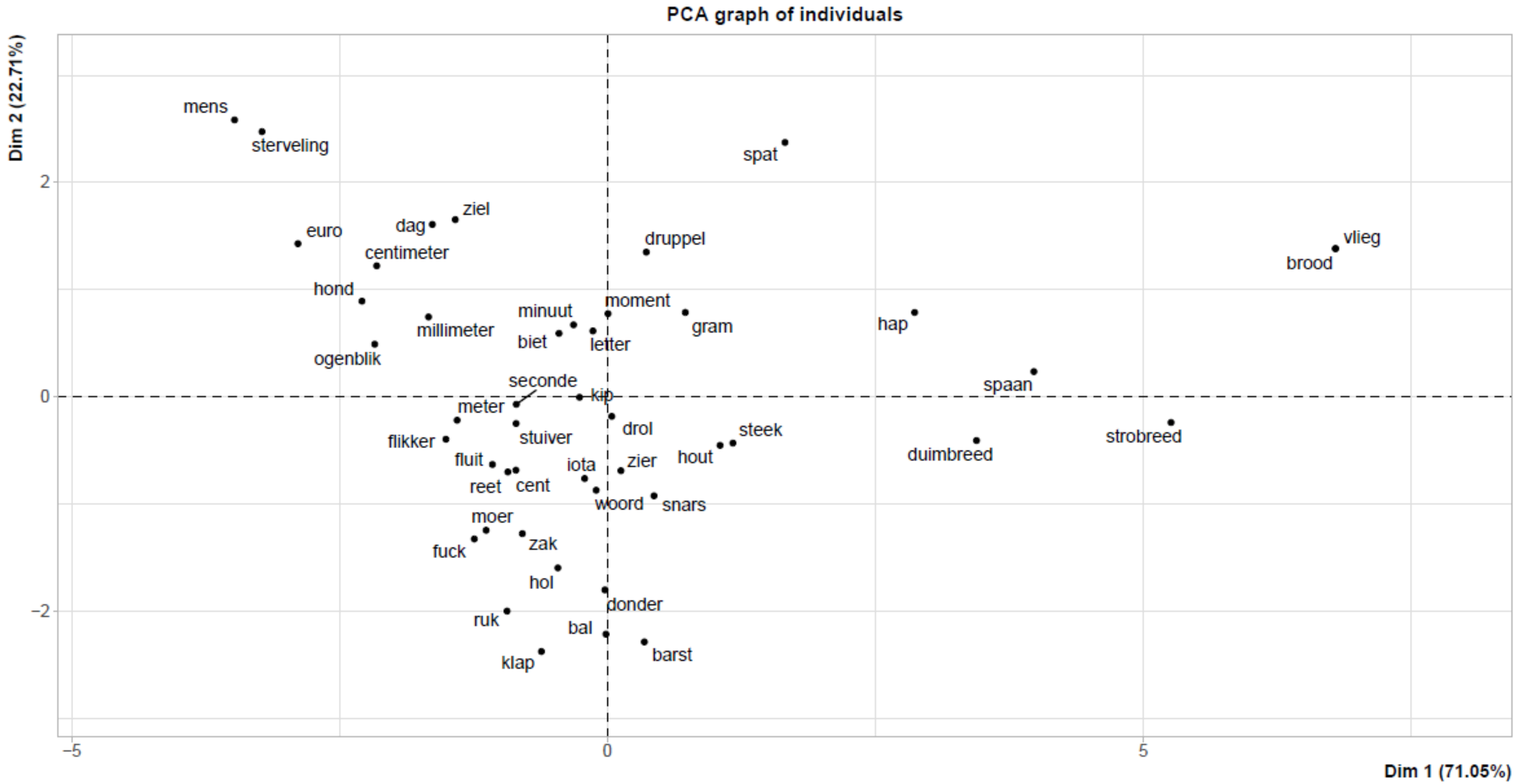
→ It is not always the case that token frequent predicates detract from productivity → They can function as attractors, but more detailed semantic analysis is needed!

OVERVIEW

1. The minimizing construction
2. General overview of the dataset
3. Productivity measures
4. Principal Components Analysis (PCA)
    - 4.1 Two macro-dimensions of productivity
    - 4.2 Productivity and semantics
5. Conclusion

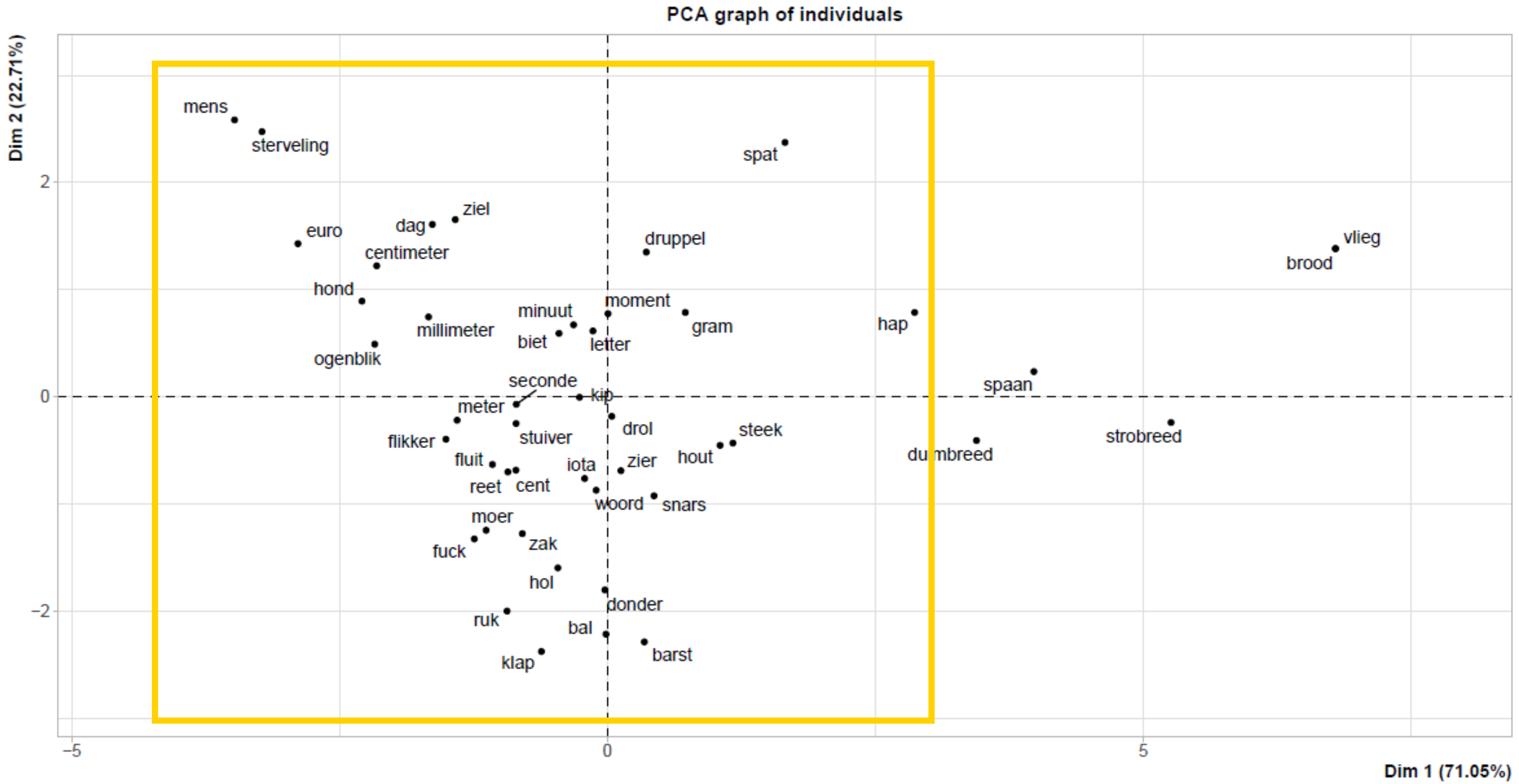
GHENT UNIVERSITY

4.2 PRODUCTIVITY AND SEMANTICS



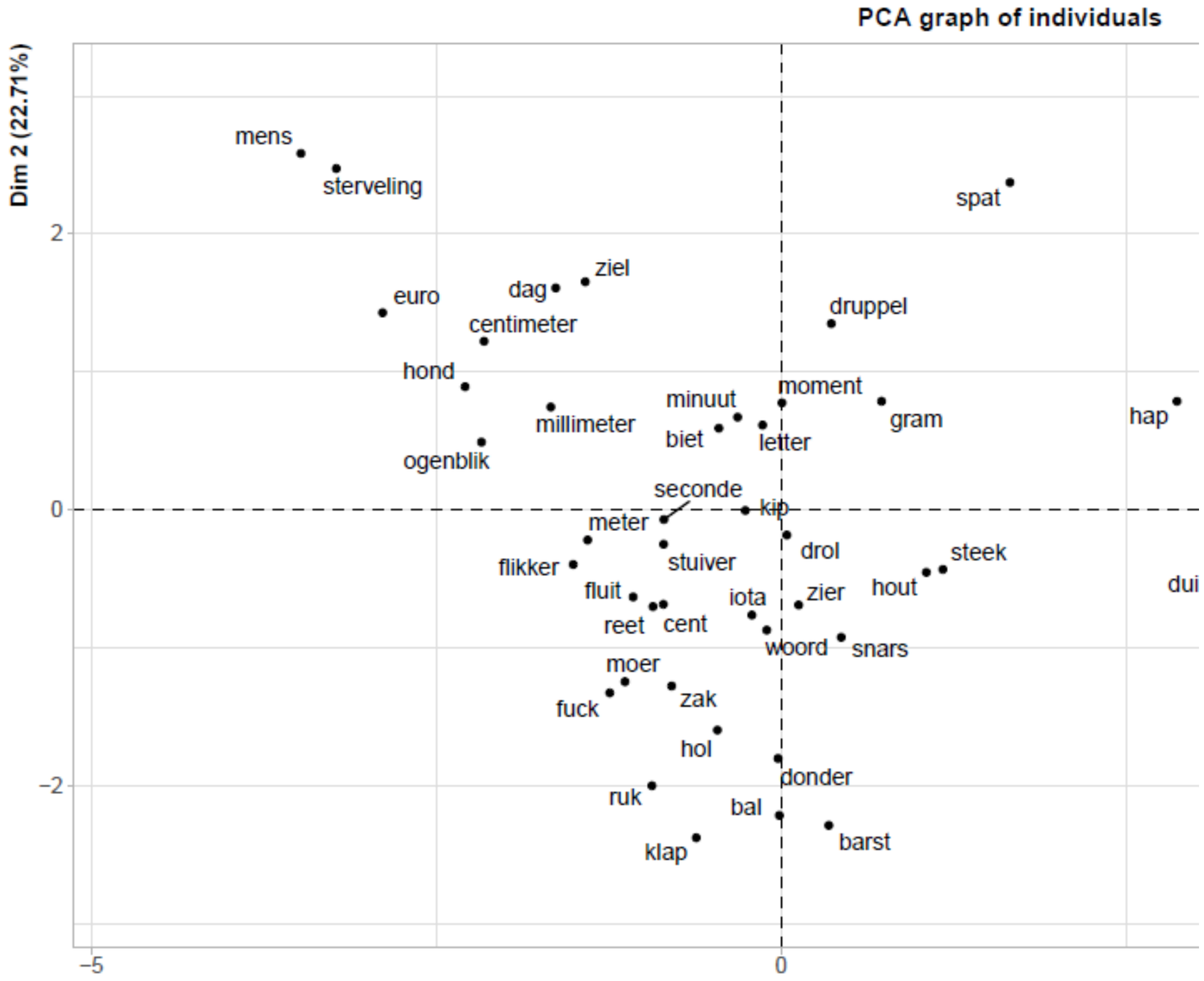
GHENT UNIVERSITY

4.2 PRODUCTIVITY AND SEMANTICS



GHENT UNIVERSITY

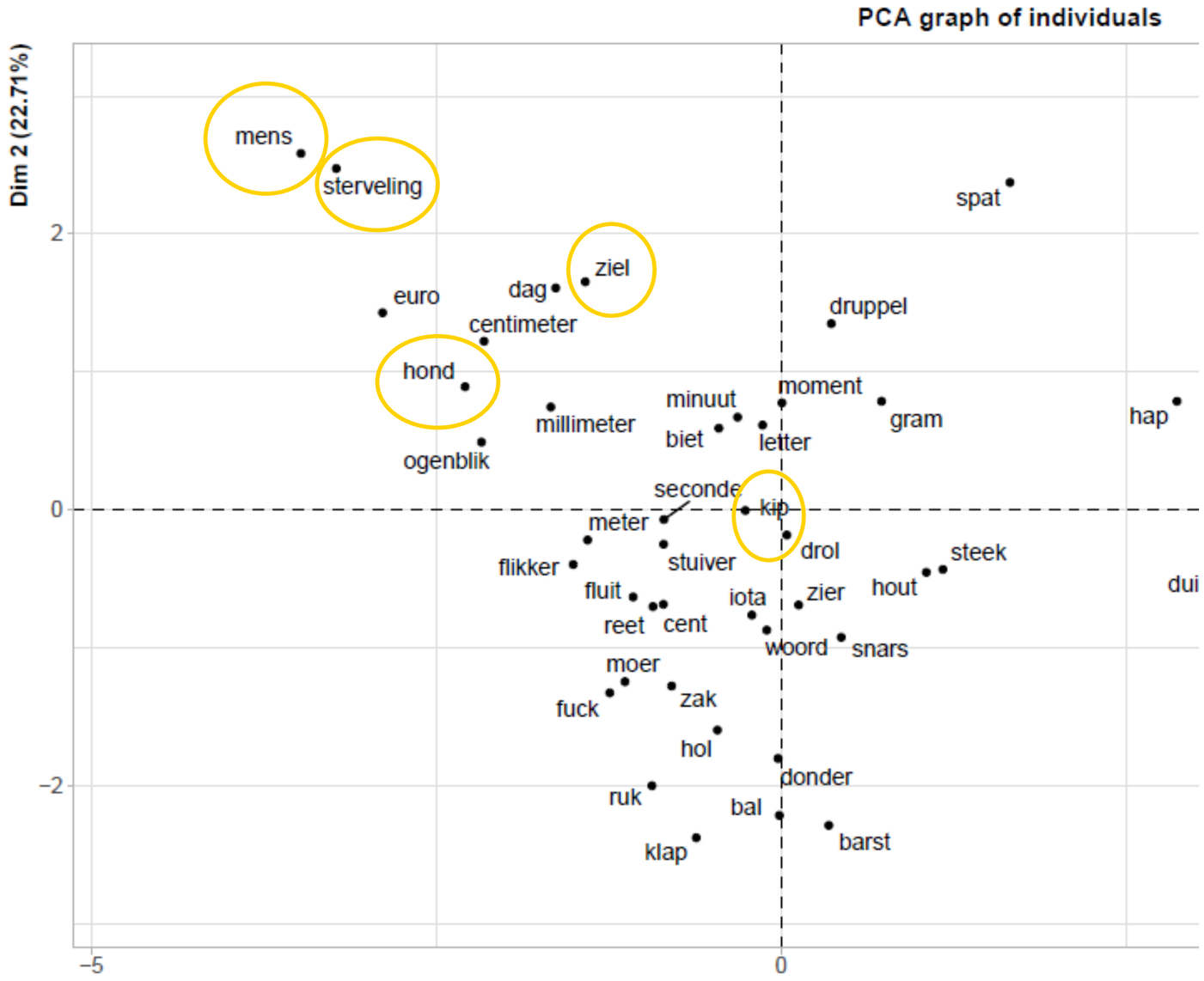
4.2 PRODUCTIVITY AND SEMANTICS



4 semantic clusters

GHENT UNIVERSITY

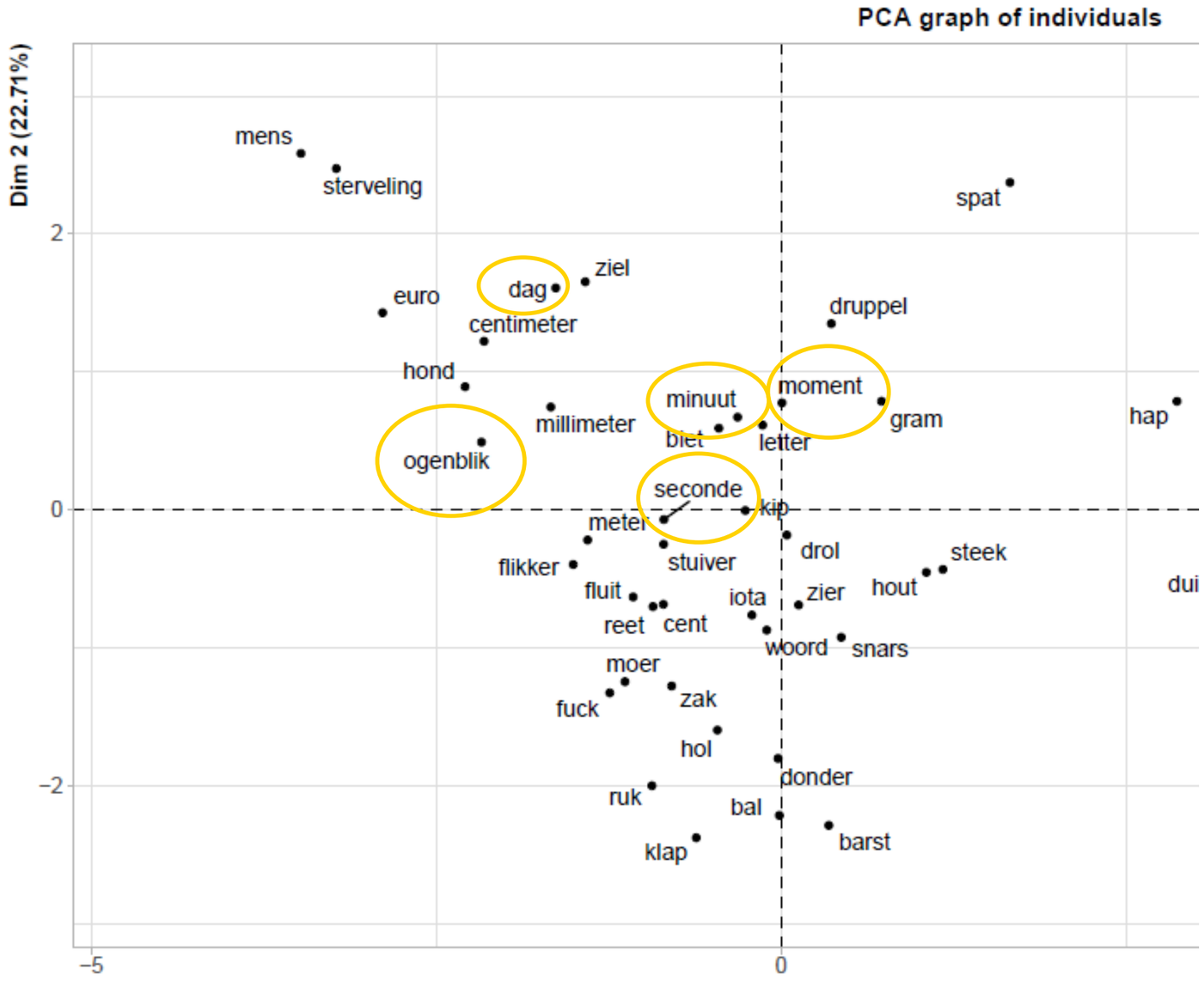
4.2 PRODUCTIVITY AND SEMANTICS



1. Minimizers referring to **people and animals**

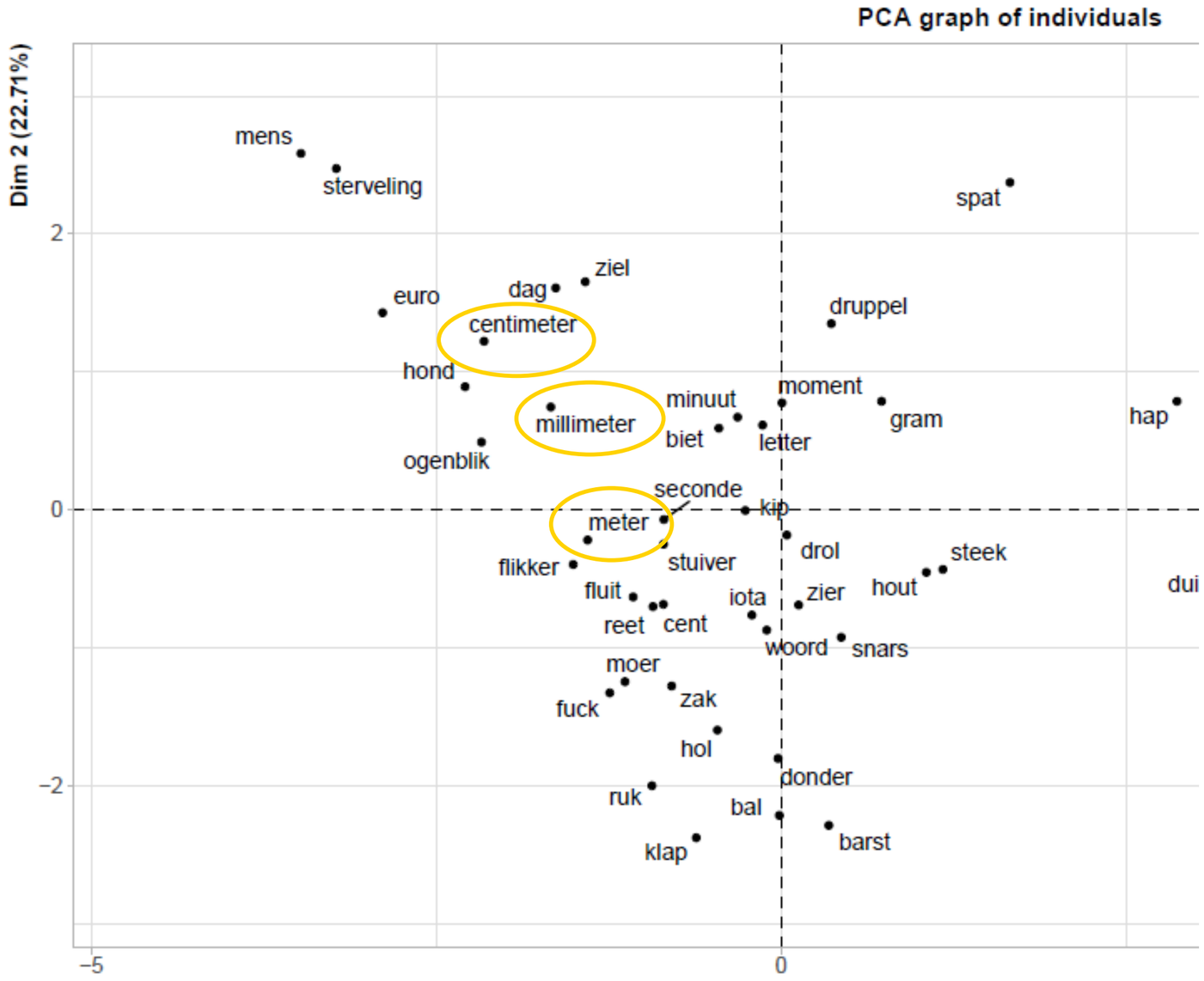
GHENT UNIVERSITY

4.2 PRODUCTIVITY AND SEMANTICS



2. Minimizers referring to **time**

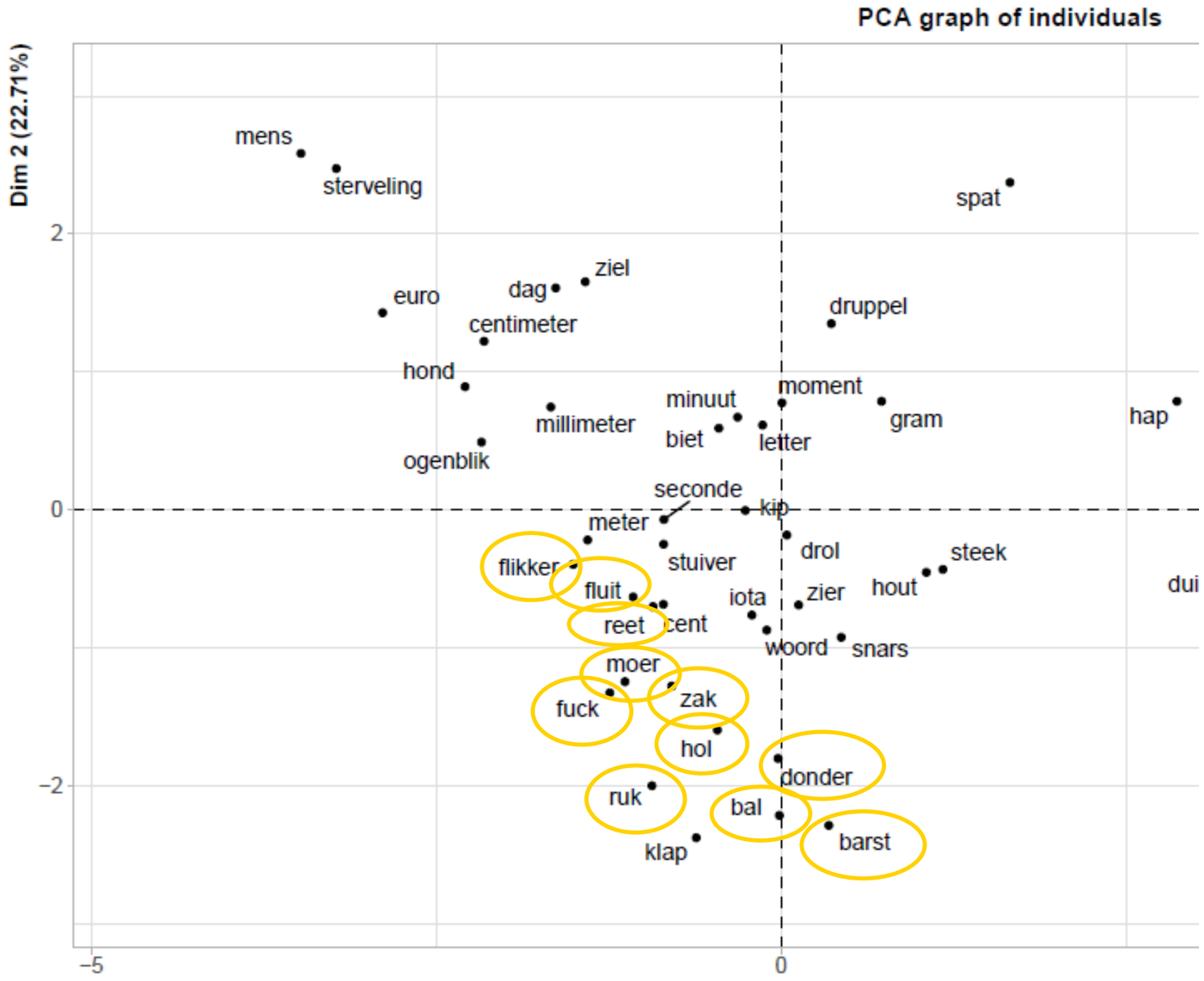
4.2 PRODUCTIVITY AND SEMANTICS



3. Minimizers referring to **distance**

GHENT UNIVERSITY

4.2 PRODUCTIVITY AND SEMANTICS



4.
Taboo
minimizers

#*!

GHENT UNIVERSITY

4.2 PRODUCTIVITY AND SEMANTICS

Semantic category	Mean type frequency	Mean hapax frequency	Mean hapax type ratio
People and animals 5 minimizers 	50,6	37	0,71
Distance 3 minimizers 	42	27,7	0,65
Time 5 minimizers 	37,8	24,8	0,65
Taboo 11 minimizers 	31,8	13,6	0,42

GHENT UNIVERSITY

4.2 PRODUCTIVITY AND SEMANTICS

Semantic category	Number of different predicates	Average number of predicates	Shared predicates
People and animals 5 minimizers 	169	33,8	5 (used in 29% of the tokens)
Distance 3 minimizers 	97	32,3	5 (used in 24,3% of the tokens)
Time 5 minimizers 	122	24,4	5 (used in 32,4% of the tokens)
Taboo 11 minimizers 	86	7,8	9 (used in 49,6% of the tokens)

GHENT UNIVERSITY

4.2 PRODUCTIVITY AND SEMANTICS

geen

bal
barst
drol
flikker
fluit
fuck
hol
moer
reet
ruk
zak

begrijpen
doen
geloven
hebben aan
interesseren
snappen
te maken hebben met
uitmaken
zijn aan

GHENT UNIVERSITY

5. CONCLUSION

- productivity in general & productivity of the minimizing constructions

5. Conclusion

- productivity in general & productivity of the minimizing constructions

→ two facets of productivity, corresponding to the macro-dimensions of the PCA

(a) dimension 1: openness vs. conventionalization
high type and hapax frequency (no token frequent predicates)

(b) dimension 2: lower-level local productivity
high hapax type ratio (the slot is extensible, despite the presence of high token frequent predicates)

5. CONCLUSION

- productivity in general & productivity of the minimizing constructions

→ two facets of productivity, corresponding to the macro-dimensions of the PCA

(a) dimension 1: openness vs. conventionalization
high type and hapax frequency (no token frequent predicates)

(b) dimension 2: lower-level local productivity
high hapax type ratio (the slot is extensible, despite the presence of high token frequent predicates)

→ importance of semantics: analogical attraction

(a) high token frequent predicates (such as *geen spat veranderen*)

(b) among taboo minimizers

GHENT UNIVERSITY

FACULTY OF ARTS AND PHILOSOPHY

DEPARTMENT OF LINGUISTICS
RESEARCH GROUP: GLIMS

Margot Van den Heede
mcvdnhee.vandenheede@ugent.be

Peter Lauwers
peter.lauwers@ugent.be

Thank you for your attention!

Part of the GOA project (BOF UGent)
Language Productivity at Work
https://www.languageproductivity.ugent.be/

Supervisors: P. Lauwers (PI), J. Barðdal, R. Enghels, T. Colleman, R. Hartsuiker, M. Taverniers, L. De Cuypere, A. Ghyselen

GHENT UNIVERSITY

REFERENCES

Barðdal, J. 2008. *Productivity: Evidence from Case and Argument Structure in Icelandic*. Amsterdam: Benjamins.

Bybee, J. 2007. *Frequency of use and the organization of language*. Oxford: Oxford University Press.

Clausner, T. C., & W. Croft. 1997. Productivity and schematicity in metaphors. *Cognitive science 21*(3), 247-282.

De Smet, H. 2020 (aop). What predicts productivity: From theory to individuals. *Cognitive Linguistics*.

Hilpert, M. 2015. From hand-carved to computer-based: Noun-participle compounding and the upward strengthening hypothesis. *Cognitive Linguistics 26*(1), 113-147.

Hoeksema, J. 2002. Minimaliseerders in het Standaardnederlands. *Tabu 32* (3-4), 105-174.

Perek, F. 2016. Recent change in the productivity and schematicity of the *way*-construction: a distributional semantic analysis. *Corpus Linguistics and Linguistic Theory 14*(1), 65-97.

van Egmond, M. 2013. Calculating Zipf's Law (and building growth curves).

Van Wettere, N. 2021. Productivity of French and Dutch (semi-)copular constructions and the adverse impact of high token frequency. *International Journal of Corpus Linguistics 26*(3), p. 396-428.

Zeldes, A. 2012. *Productivity in argument selection: From morphology to syntax*. Berlin: De Gruyter Mouton.

Corpus:

Jakubíček, M., A. Kilgarriff, V. Kovář, P. Rychlý, & V. Suchomel. 2013. The TenTen corpus family. In *7th International Corpus Linguistics Conference CL*, Lancaster, 125-127.

GHENT UNIVERSITY