Explaining analogy in word-formation: the role of lexical network structure

Sabine Arndt-Lappe, Trier University

Presenting joint work with Aaron Seiler & Tammy Ganster

Analogy

- In usage-based theories: a key mechanism of linguistic generalisation

- a linguistic form or meaning that is determined by the properties of similar forms or meanings in the language user's Mental Lexicon

- operate 'on the fly', on the basis of individual exemplars

Explaining analogy

- The Mental Lexicon
  - highly interconnected
  - shared structure

- Analogy arises from co-activation

=> The structure of the network is interesting, as it explains which analogies arise

This paper

- Two case studies exploring the relation between analogies / linguistic generalization and the network structure of the Mental Lexicon

- Testbed: Morpho-phonology, the position of stress in English complex words

- **Study 1:** How does the apparent sensitivity of English verb stress to (opaque) morphological structure emerge from network structure in the Mental Lexicon?

- **Study 2:** How do individual differences in network structure have an influence on paradigmatic co-activation, resulting in variation in stress position in complex adjectives?

Study 1: opaque morphologies

Joint work with Aaron Seiler.

English word-level stress

English stress is ...

- different for nouns and verbs

- influenced by phonological and morphological structure
    - focus in this paper: opaque (!) morphology

- semi-regular: there are predictable (and productive) patterns, but not all forms adhere to these patterns

(e.g. Chomsky & Halle 1968, Sherman 1975, Liberman & Prince 1977, Guierre 1979, Hayes 1982, Burzio 1994, Hammond 1999, Raffelsiefen 1999, Pater 2000, Fournier 2007, Zamma 2012, Dabouis & Fournier 2023)

Phonology: verbs

➢ Stressed on their final syllable if it is heavy (e.g. cajóle, eráse, collápse)

➢ Otherwise stressed on the penultimate (e.g. bóther, devélop, édit, vómit)

Opaque morphology: verbs

Disyllables

- Stress on the root of **prefixed** verbs (=final syllable)
- **Non-prefixed** verbs: Stress on final syllable, unless it is L

Prefixed		Non-prefixed	
Ultima L	**Ultima H**	**Ultima L**	**Ultima H**
a#tt**á**<ch>	a#rr**í**<ve>	c**á**nce<l>	torm**é**n<t>
co#mm**í**<t>	con#cl**ú**<de>	l**í**ste<n>	mol**é**s<t>
re#p**é**<l>	re#m**ái**<n>	v**ó**mi<t>	us**ú**r<p>

Polysyllables

- **Prefixed:** Stress on root (comprehénd, apprehénd, intervéne, introspéct, introdúce, remémber, inhábit)
- **Non-prefixed** verbs: Stress on antepenult if final is H (éxercise, cómplicate, clárify), otherwise stress on penult (e.g. manóeuvre, sequéster, solícit). Final H is often (part of) a suffix or suffix-like.

(see Dabouis & Fournier 2024)   8

The nature of ‚opaque‘ morphological effects

Hypothesis in this study:

Stress in English verbs is assigned by analogy with existing forms.

Explaining analogy:

Effects of ‚opaque morphology‘ are based on recurrence.

‚Morphological effects‘ in stress position arise from the fact that the network of shared sounds among words in the lexicon is sparse and organized in communities (‚hubs‘). Such hubs like to be stressed the same.

(esp. Skousen 1989, Skousen et al. 2013, Daelemans & van den Bosch 2005, Nosofsky 1986 et seq., Baayen et al. 2011, 2019)

Computational analogical models

- can operationalize the notions of 'similarity' and analogical sets'

- algorithm:
  - **AM(L)** (Skousen 1989 et seq., Skousen et al. 2013)

  - **TrAML** (GUI for AML that provides easy access to model measures, Arndt-Lappe et al. 2018 et seq)

  - A faster version that can handle more features: https://github.com/garfieldnate/Weka_AnalogicalModeling (Glenn 2021, functionality for extracting model measures is currently under development, please contact Sabine Arndt-Lappe if you are interested)

(esp. Skousen 1989, Skousen et al. 2013, Daelemans & van den Bosch 2005, Nosofsky 1986 et seq., Baayen et al. 2011, 2019)

AML - the general architecture

Exemplars in the Lexicon

item	onset -s_{pen}	nuc- s_{pen}	coda- s_{pen}	onset -s_{fin}	nuc- s_{fin}	coda- s_{fin}	Stress
digest	d	i		-	e	st	fin
digress	d	i		gr	e	s	fin
dilate	d	i		l	a	t	fin
dilute	d	i		l	u	t	fin
vomit	v	o		m	i	t	pen
submit	s	u	b	m	i	t	fin
intermix	t	e		m	i	ks	fin

Analogical Set:
a set of similar exemplars from the lexicon

item	onset -s_{pen}	nuc- s_{pen}	coda- s_{pen}	onset -s_{fin}	nuc- s_{fin}	coda- s_{fin}	Stress
vomit	v	o		**m**	**i**	**t**	pen
submit	s	u	b	**m**	**i**	**t**	fin
intermix	**t**	**e**		**m**	**i**	ks	fin

2x fin, 1x pen

predicted probability of fin: > 60%

Word to be stressed: *intermit*

item	onset -s_{pen}	nuc- s_{pen}	coda- s_{pen}	onset -s_{fin}	nuc- s_{fin}	coda- s_{fin}
intermit	t	e		m	i	t

11

How does AML find out which exemplars are ‚relevant'?

the similarity space

Different degrees

AML starts with the most similar exemplars.
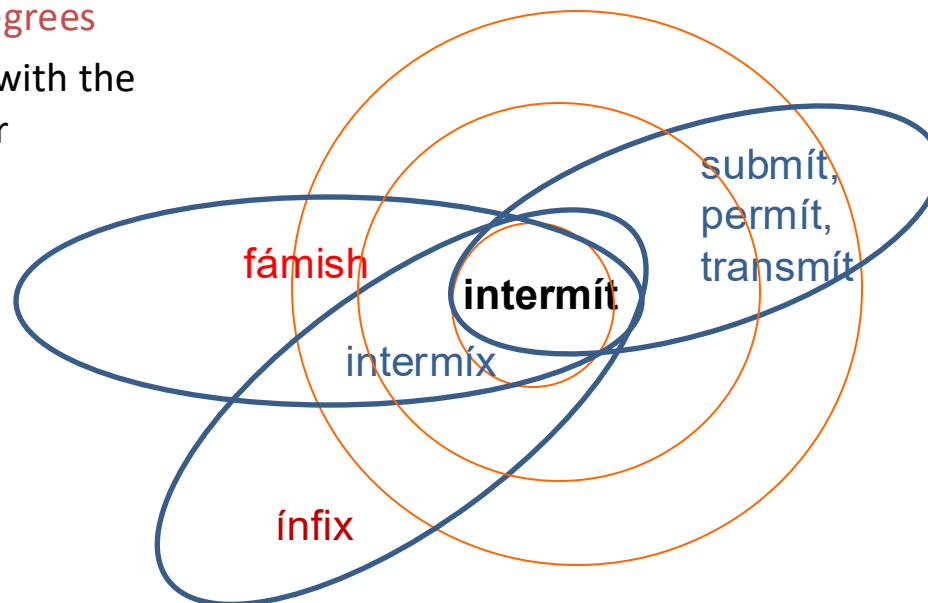
different dimensions

Along all dimensions, AML tries to include more distant exemplars.

fámish

submít, permít, transmít

intermít

intermíx

ínfix

It does so if the more distant exemplars behave like the more similar group w.r.t. to stress assignment. => minimised uncertainty

Modelling English Verbs

Data from Fournier & Dabouis (2023)

All verbs (> 1 syllable) from the *Cambridge English Pronouncing Dictionary* (Jones 2006)

Very rare verbs excluded

No variants:
- only the main pronunciation for British English
- some further variants excluded

N = 3,033

Simulation experiment

Lexicon
All verbs extracted from Jones (2006, N = 3,033)
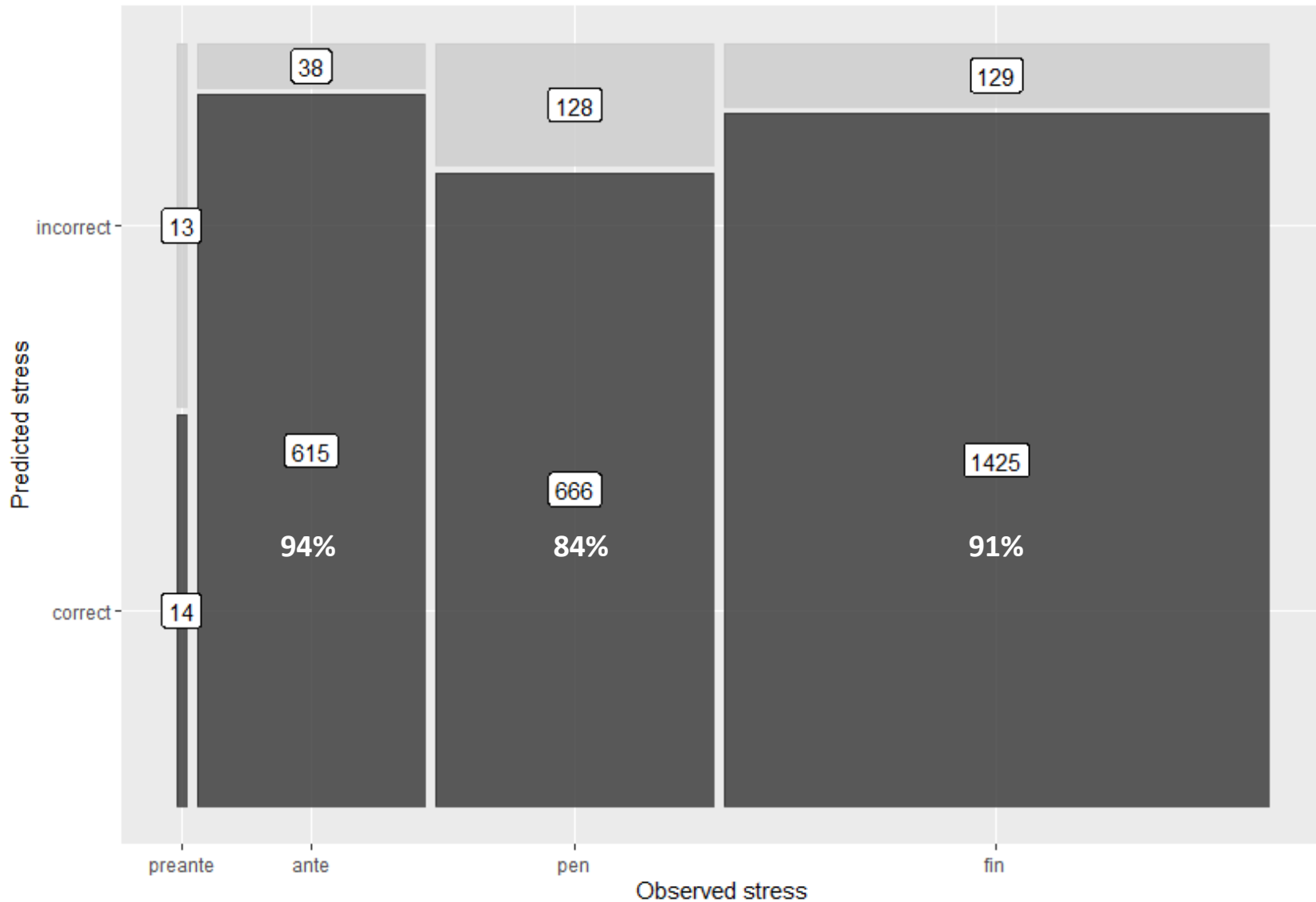
Test set
the same as the lexicon file

‚leave-one-out‘ mode

Distribution of stresses in the lexicon file

e.g. peregrinate, denaturalise

e.g. abbreviate, administer, analyse, objectify

e.g. burgle, stargaze, injure, borrow

e.g. update, increase, maintain, deplete

preante	ante	pen	fin
27	653	795	1,558
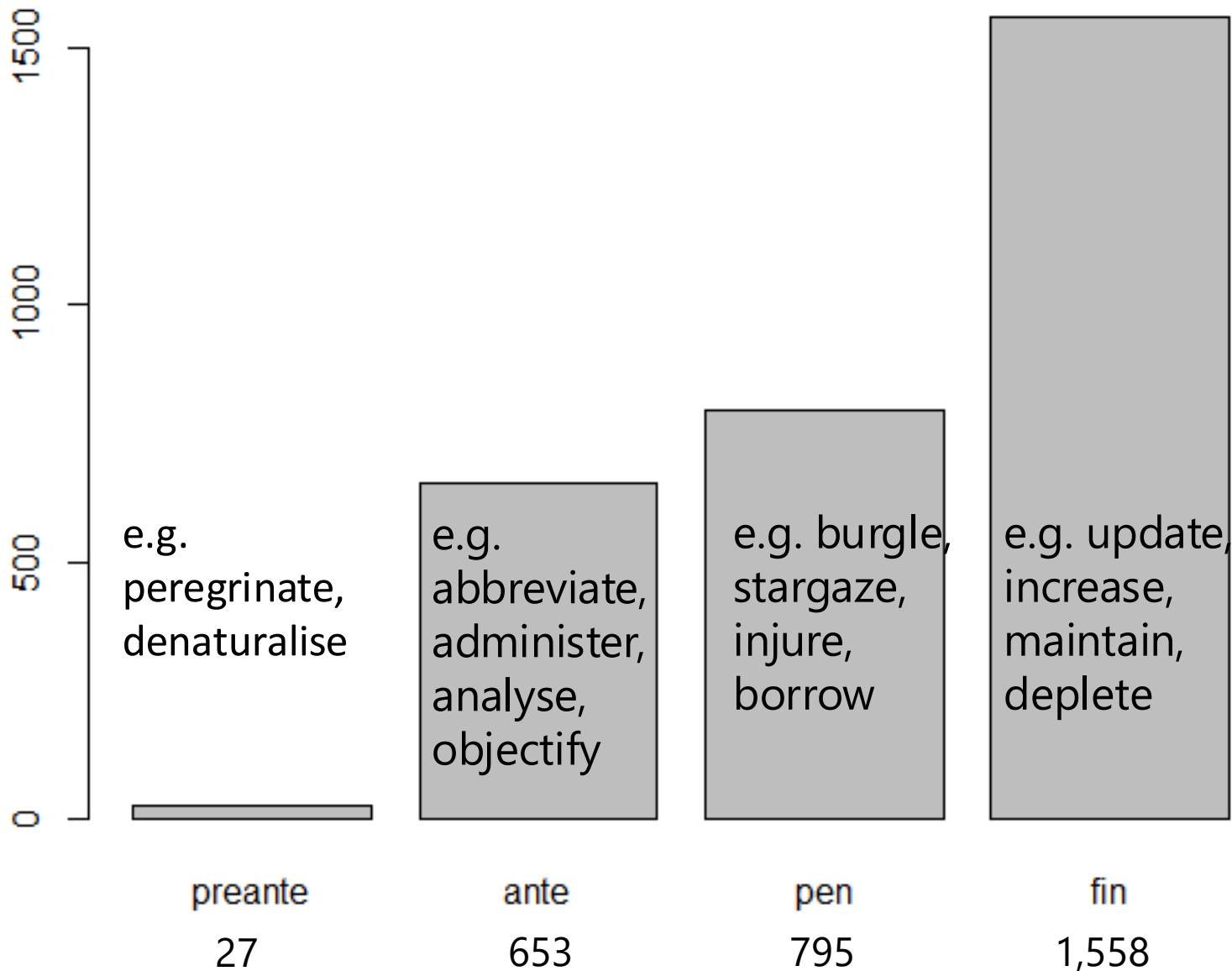
N = 3,033

Predictive accuracy, by majority vote

(ties excluded, N = 3,028)

16

Beyond the ‚black box'

Similarity and exemplar relevance in classification:

Examining the structure of Analogical Sets

Exploring the structure of Analogical Sets

The **analogical set** of a test item is the set of items from the lexicon file that the algorithm has decided to be relevant for classification of the test item.

Rationale: Find out how patterns of English verb stress emerge from properties of the distribution of verbs in the Mental Lexicon.

Analysing analogical sets: Network analysis

- Networks help to „visualize and estimate relationships among agents" in a complex system (Holster, 2020, Chapter 7)

- **Agents:** exemplars that are influential for the classification of novel items

- **Networks:**
  - **Nodes**: exemplars in the lexicon
  - **Edges**: links between exemplars
    - test word -> analogues
    - analogues -> test word

(e.g., Barabasi 2015)

Analysing analogical sets: Network analysis

- allows us to quantify the degree of connectivity between different (hubs of) analogues

Key metrics

- **Degree Centrality** (the degree of a node represents the number of links it has to other nodes)

- **Community structure** (i.e., different groupings after the split of the network)

The „verb stress" network – top influential actors

Top 5 Members per Community (Top 15 Communities by Size)
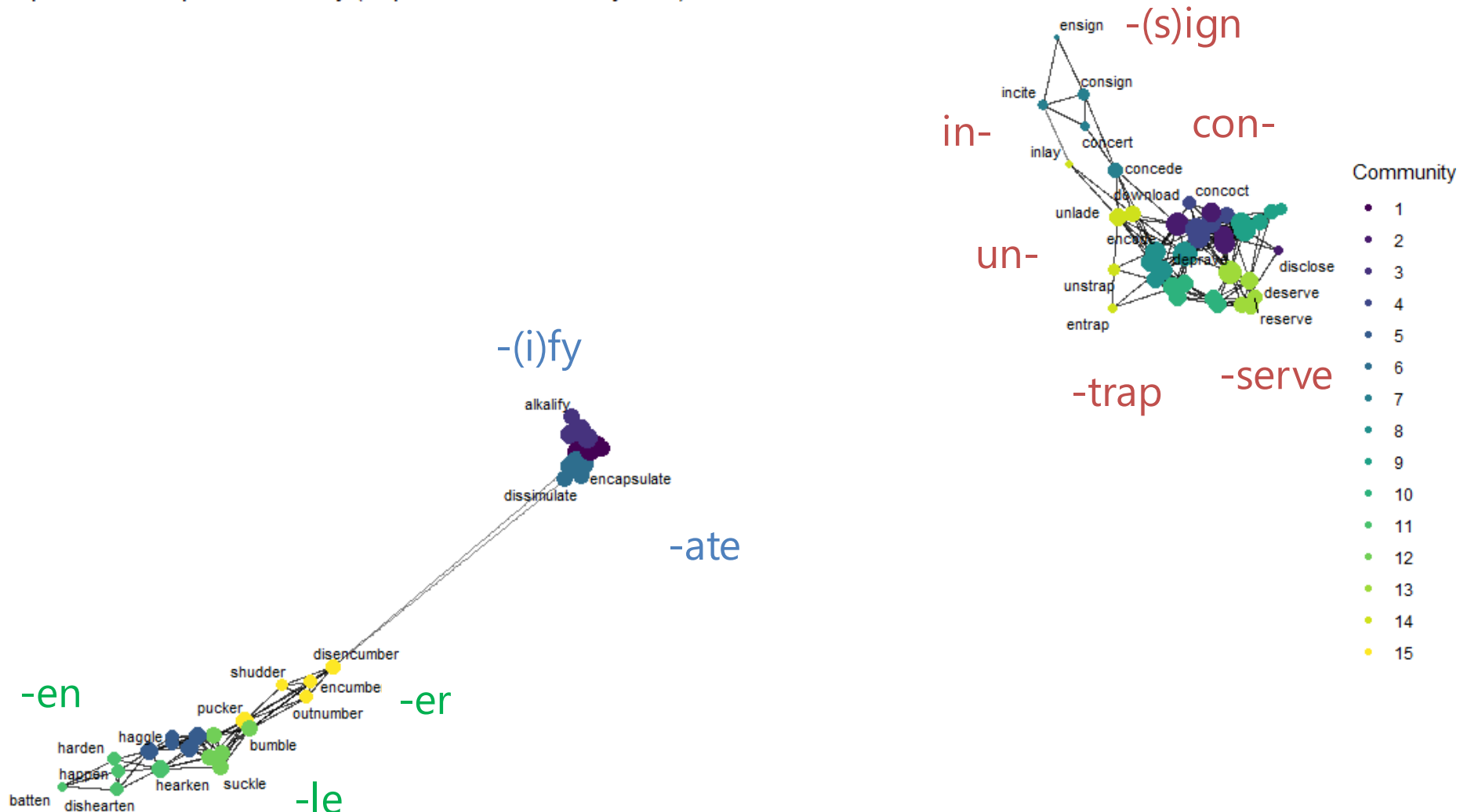
Community structure in a verb stress network

Community A	Community B
degrade	disclose
extrude	declare
retard	declaim
delude	exclaim
retread	reclaim
overcloud	disincline
preside	enclose
exceed	disclaim
regard	proclaim
becloud	declutch

(using the Infomap algorithm to detect communites, Bohlin et al. 2014)

Zooming in on subgraphs

- Split the verb stress network, according to stress type ("gante", "gpen", "gfin")

Question:

What structural differences emerge among the three networks, and what do they tell us about lexical distribution across stress types?

Degree Centrality Distributions
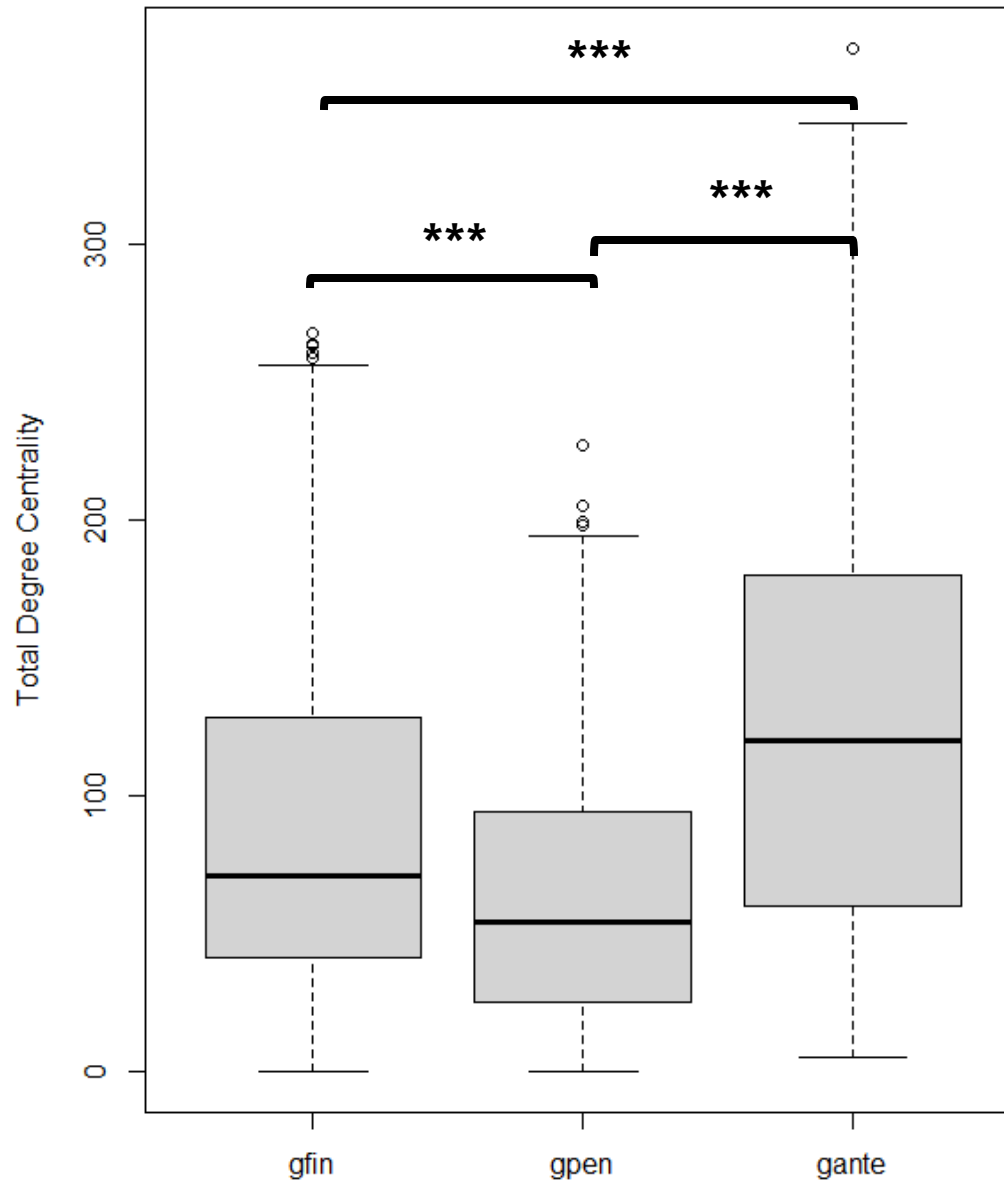
Significance based on pairwise Wilcoxon Rank-Sum Tests

Reminder: the **degree (centrality)** of a node represents the **number of links** it has to other nodes

Network Density Comparison — Modularity Scores of Networks
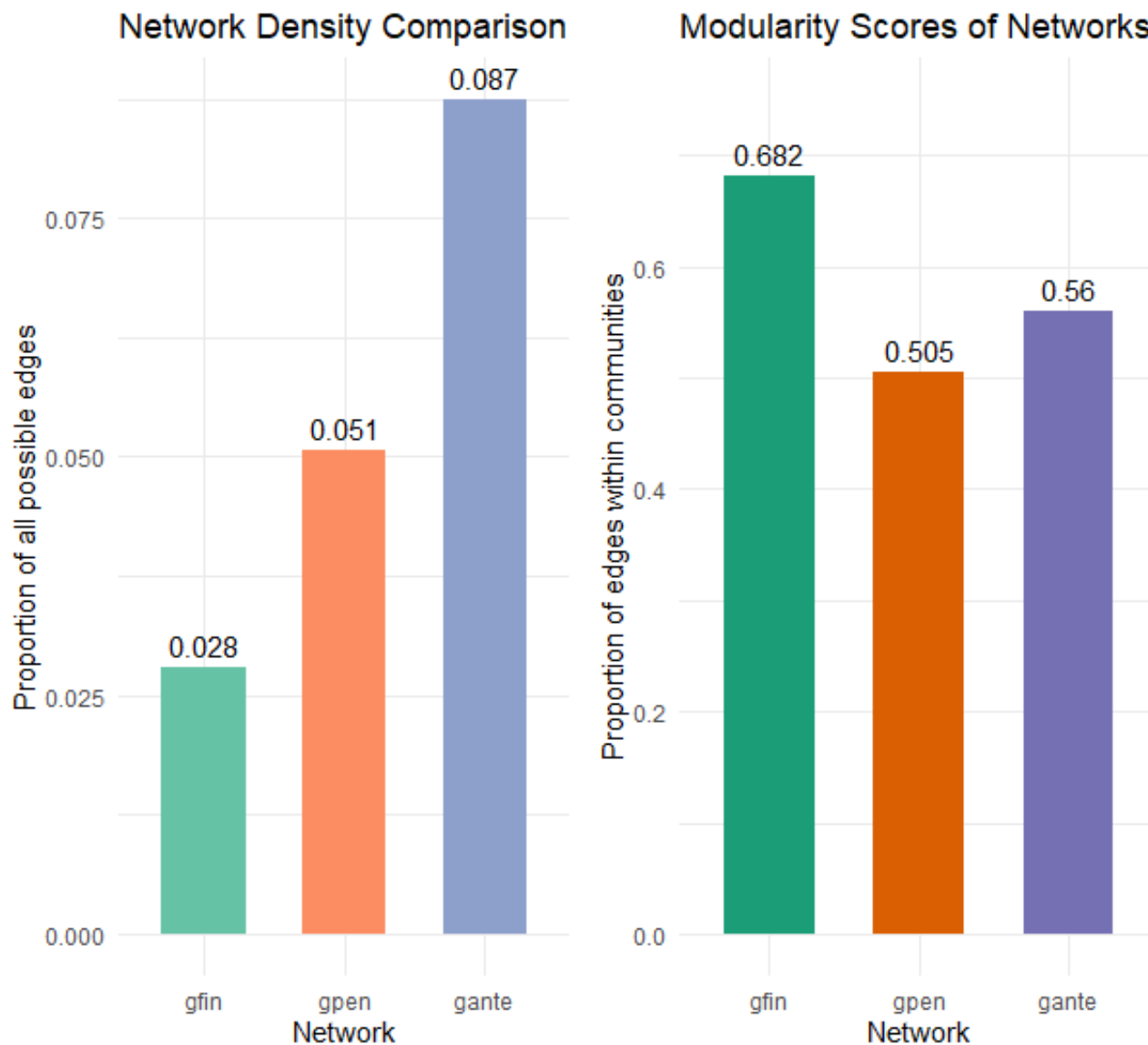
gfin: low overall connectivity, but high modularity
gante: high overall connectivity, but low modularity
gpen: in-between

What does all this mean linguistically?

 → structural differences between analogical sets for each stress type

 → different types and degrees of similarity

Final Stress

- generally smaller, distinct, well-separated clusters
- little to no overlap between groups („islands")

Community 6	Community 9
degrade	redress
decide	undress
explode	possess
extrude	caress

Antepenult stress

- flatter structural profile: long verbs (3+ syllables) ending in *-ate, -(i)fy* or *-ize*

- Presence of „super-connectors" („hubs")

- Considerable overlap, weak group boundaries

- more homogeneous group as a whole

- more abstract generalizations

Community 9	Community 13
dedicate	derogate
medicate	congregate
triplicate	interrogate
predicate	relegate

Penult stress

- in-between: groups are sufficiently different from each other, but a few shared endings blur otherwise clear community boundaries (mostly *-le, -er, -en*, to a lesser extent also *-y* and *-ish*)

Community 2	Community 8
grabble	juggle
babble	snuggle
scramble	struggle
scrabble	smuggle

Intermediate summary

- English verb stress can be accounted for very well by the analogical model

- Specifically: effects of 'opaque morphology'

Explaining analogy

- The way verbs are stressed can be directly related to the way the lexical network (and its sub-networks) are organised.

Study 2: individual differences

Joint work with Tammy Ganster.

Individual differences

- If morpho-phonology is based on analogy, i.e. on co-activation of words in the Mental Lexicon

- Individual differences relating to the structure of the Mental Lexicon should correspond to individual differences in the phonological realisation of complex words.

Generalisations about stress in English complex words

Stress preserving suffixes

- – the derived word has the same main stress as its base
- – traditionally interpreted as the effect of default stress rules <u>not</u> operating across morphological boundaries

e.g. –*ness*: *háppy ~ háppiness*

Stress ‚shifting' suffixes

- – the derived word has a different main stress from its base
- – traditionally interpreted as the effect of stress rules / markedness constraints

e.g. –*al*: *párent ~ paréntal*

Testbed:
variable stress with *-able, -ant, -(at)ive, -(at)ory*

	PRESERVATION		NON-PRESERVATION
idéntify	idéntifiable	~	identifíable
tríumph	tríumphant	~	triúmphant
eváluate	eváluative	~	evaluátive
discríminate	discríminatory	~	discminátory
resíst	resístant	~	∅
contradíct	contradíctory (mostly)	~	cóntradictory (rarely)
spéculate	spéculative (mostly)	~	specúlative (rarely), speculátive (rarely)
partícipate	partícipatory (50%)	~	participátory (50%)

PRESERVATION   NON-PRESERVATION

*all pronunciations observed in the experiment

Methodology

Remote online experiment (recruitment: Prolific)

- **153 native speakers** of British English from different backgrounds (age: **18-77** yrs**, 93** females, **60** males)
- **3** different tasks

1. Stress

– reading task

2. Individual differences in lexicon structure

– size: vocabulary size test (Nation & Beglar 2007)

– Interconnectedness: morphological processing task (masked priming with lexical decision)

– Experience: print exposure test (author recognition test)

3. Sociodemographic questionnaire

Reading task

We're trying to do something a little more **innovative**.

Well, actually, this move was **anticipatory**.

Is there an address or a name or something else that's **identifiable**?

- test sentences from Corpus of American Soap Operas (Davies 2011)
- each participant read out **30 test sentences**
- each recording **assessed by three trained raters** (raters **agree in 77% of cases**, **only agreement cases taken into account** in analyses)
- → **over 3,400** observations

Stress productions in a random sample of 50 participants
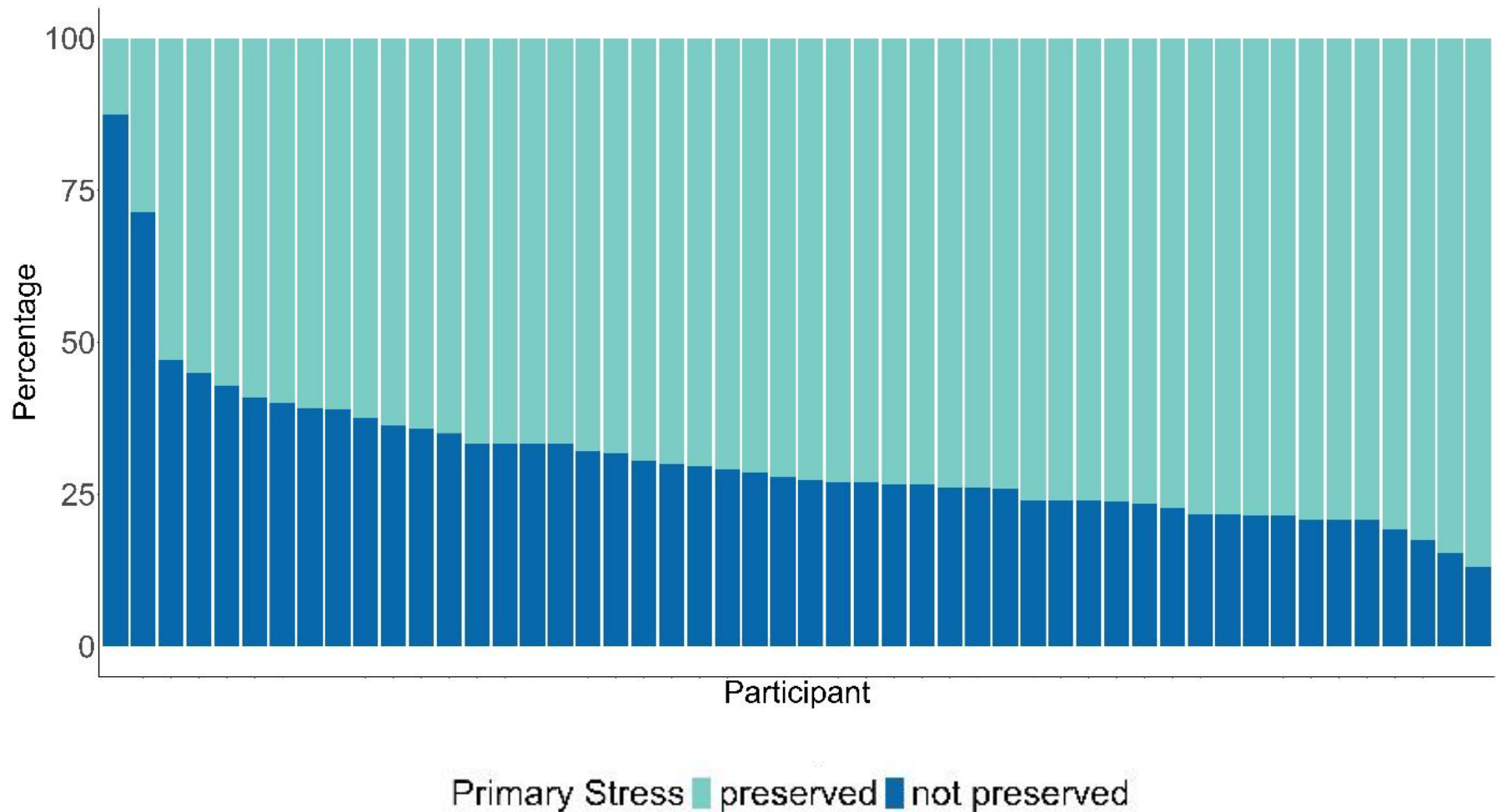
Correlates of Individuality in the Mental Lexicon

morphological processing task (N = 3,467)

- masked priming experiment with lexical decision task
- measured reaction time to three different priming conditions (complex words primes - simplex words targets)
  - morphologically related prime-target pairs
    subversion - subvert
  - orthographically related prime-target pairs
    chargeable - charisma
  - unrelated prime-target pairs
    inventive - remorse

priming effect:
MeanRT Unrelated – MeanRT Morphologically Related

Individual Differences in Morphological Processing
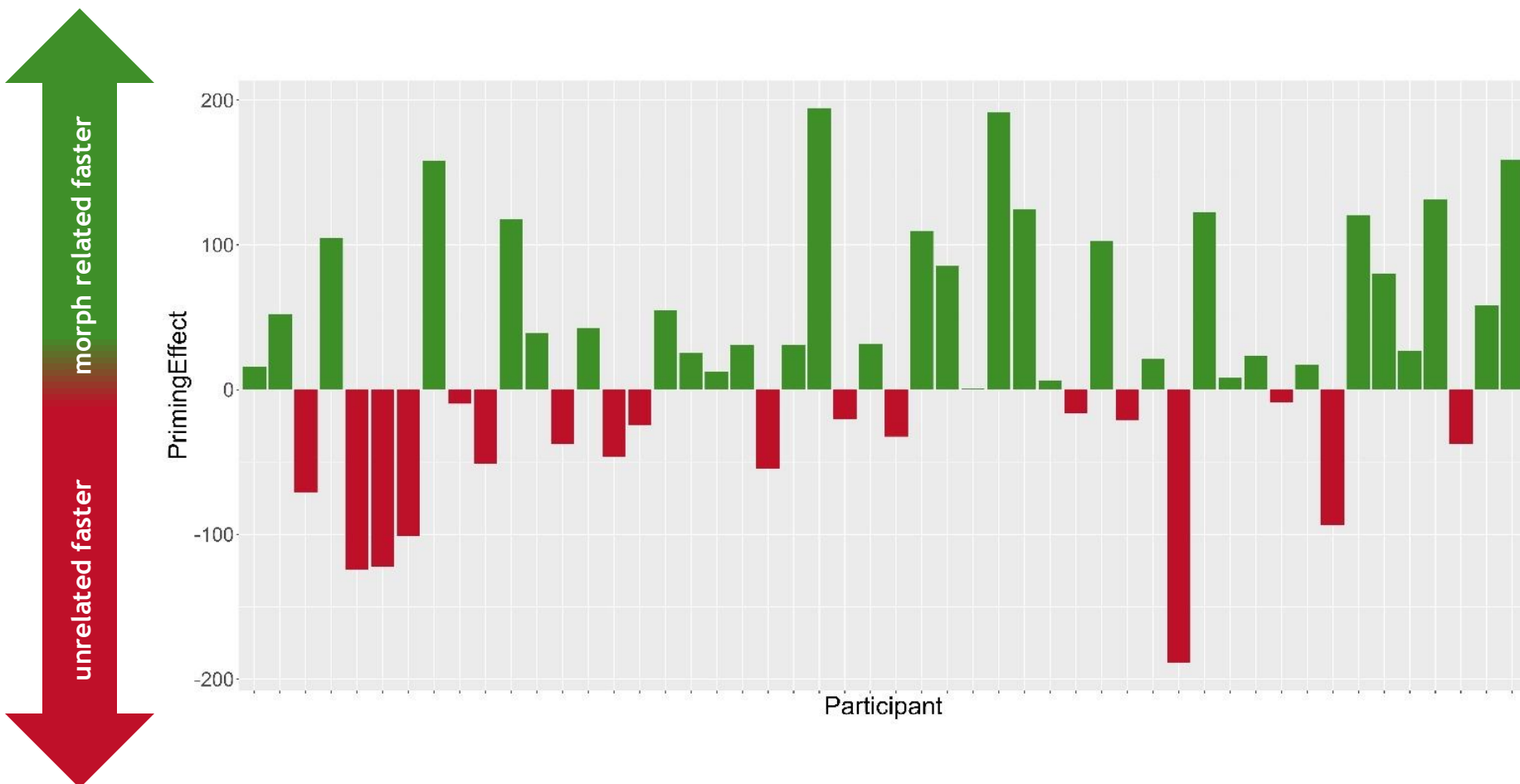
Random sample of 50 speakers

Correlates of Individuality in the Mental Lexicon

Vocabulary size test (15,300 observations)

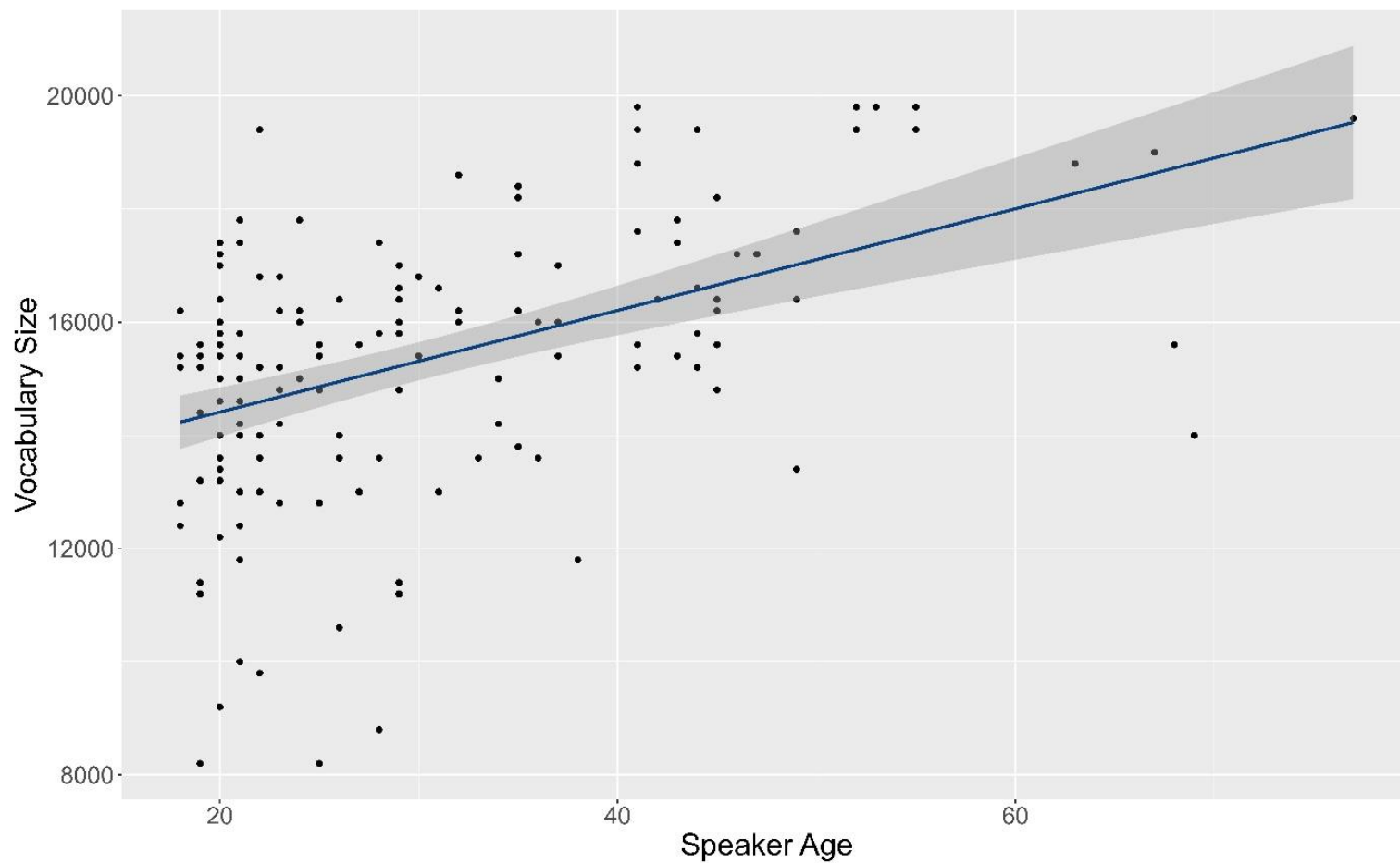- standardized and multiply validated forced choice test (Nation & Beglar 2007)
- 100 questions

Example

see: They SAW it.
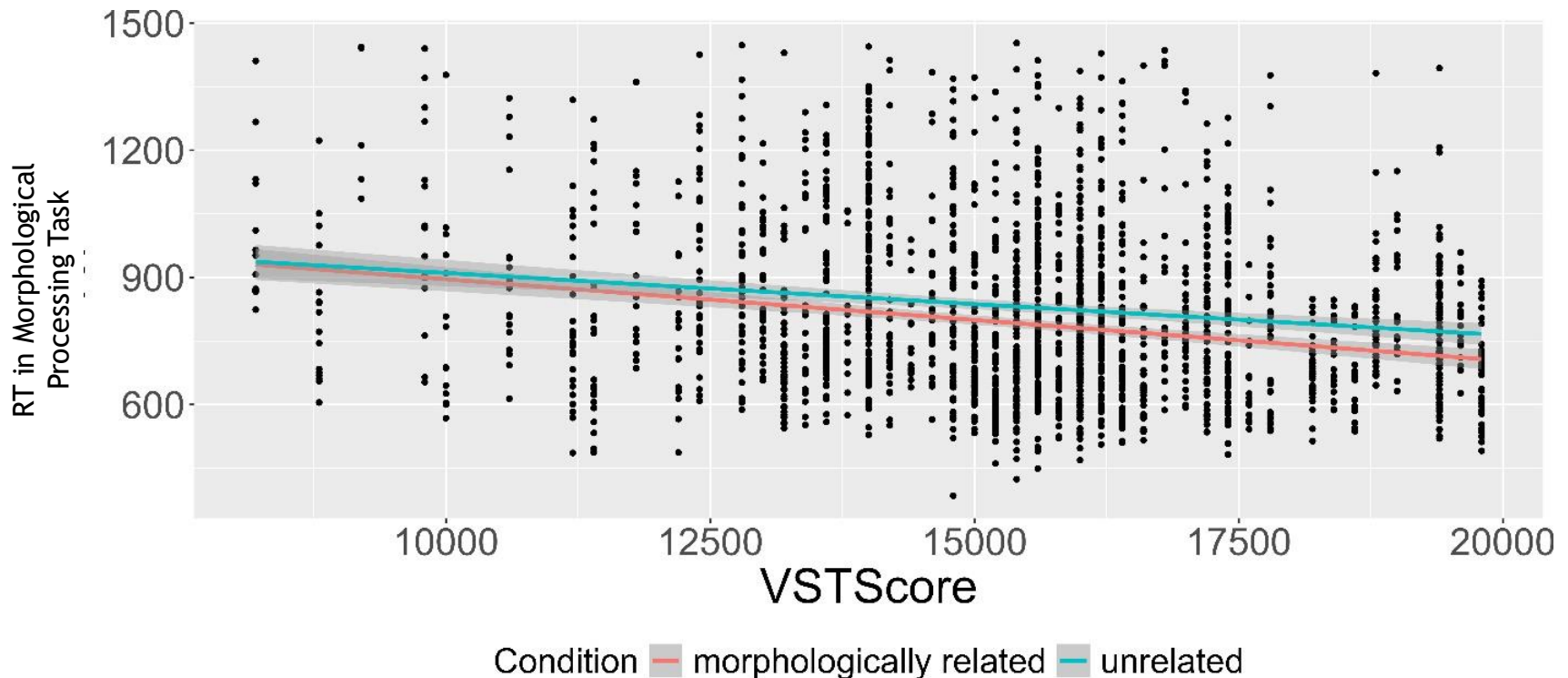a) closed it tightly
b) waited for it
c) looked at it

score from 0 - 20,000 (estimates number of known word families)

Individual Differences in Vocabulary Size

N = 153 speakers

# Vocabulary size and processing

- Vocabulary size and RT in the morphological processing task are **correlated**
- **greater vocabulary size facilitates RT** in both priming conditions
- **BUT facilitative effect of vocabulary size** is **even stronger for morphologically related** prime-target pairs compared to unrelated pairs

Putting it all together:
Individual differences in stress production

	Estimate	SE	z-value	Pr(>|z|)
(Intercept)	1.69	0.54	3.1	0.001 **
VSTScore_scaled	0.36	0.06	5.79	p < 0.001 ***
FirstLangBinary [non-native]	-0.75	0.18	-4.15	p < 0.001 ***
Marginal R^2	0.02			
Conditional R^2	0.73			
N	2870			
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
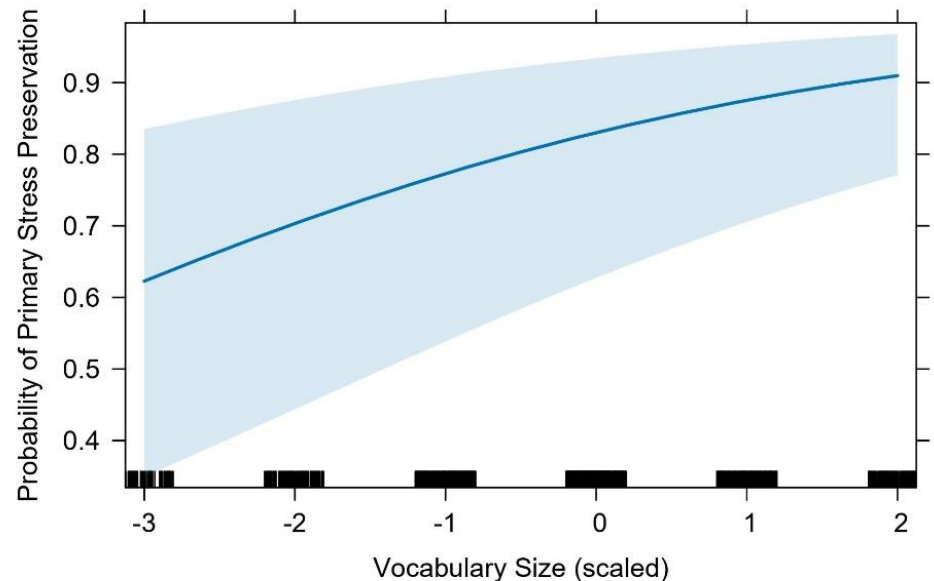
Model Formula: PriStrPreservation ~ VSTScore_scaled + FirstLangBinary + (1| Testword), N = 2870, reference level = "not preserved"

anticipátory
ánticipatory
→ antícipatory

(Variable selection by means of a Random Forest model; also considered: Lexical frequency, morphological sensitivity, author recognition)

43

What does this mean, linguistically?

- Speakers with larger vocabularies also show greater morphological sensitivity

- Speakers with larger vocabularies preserve stress in complex words more than speakers with smaller vocabularies

- Speakers with smaller vocabularies, instead, apply more phonological stress rules

What does this mean, linguistically?

anticipatory

antícipatory — speaker with large vocabulary

anticipátory — speaker with small vocabulary
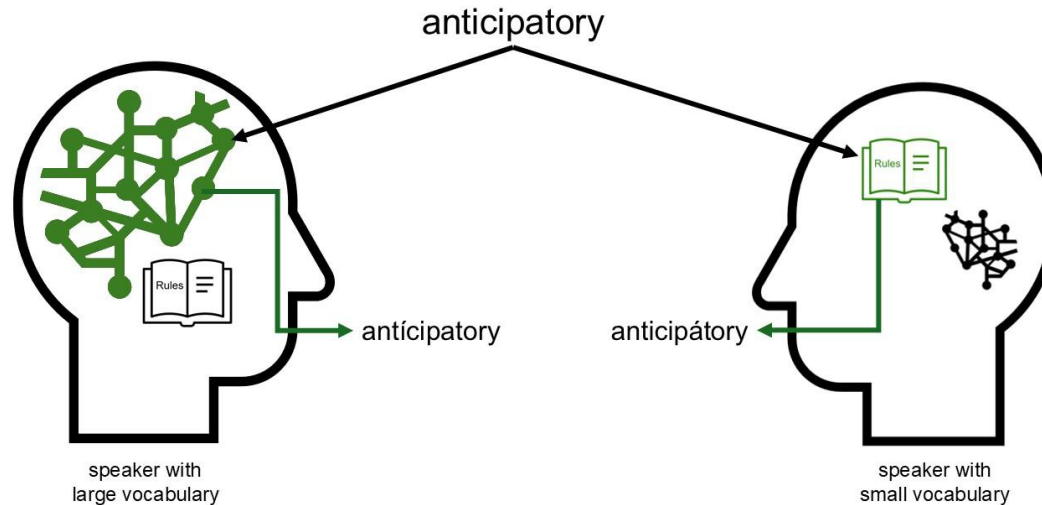
(Ganster 2025: 201, Fig. 70)

Stress preservation supports
- morphological transparency
- semantic connections between words

Stress shift supports
- stress uniformity within morph. categories

Summary & conclusion

- Two case studies about how morpho-phonological stress is related to the network structure of the Mental Lexicon

- Study 1: Opaque morphological structure emerges from the network structure of the Mental Lexicon
  - Computational Modeling (AML), Network Analysis

- Study 2: Stress variation reflects differences between speakers. Size and connectivity of the network predict the degree of stress preservation.
  - Vocabulary size, morphological sensitivity (, author recognition)

Summary & conclusion

- Findings are compatible with an analogy-based view of morpho-phonology in derivation

  - Relevance of co-activation of individual lexemes in the Mental Lexicon
  - Low level of abstraction / schematization

- More work is needed.

Thank you very much for your attention!

.... and thanks to my co-authors ☺

References

Acheson, Daniel J., Justine B. Wells & Maryellen C. MacDonald. 2008. New and updated tests of print exposure and reading abilities in college students. *Behavior Research Methods* 40(1). 278–289. https://doi.org/10.3758/BRM.40.1.278.

Arndt-Lappe, Sabine. 2014. Analogy in suffix rivalry - the case of -ity and -ness. *English Language and Linguistics* 18(3). 497–547.

Arndt-Lappe, Sabine, Ingo Plag, Kai Koch & Mikalai Krott. 2018. Transparent Analogical Modelling (TrAML). GitHub, https://github.com/SabineArndtLappe/TrAML.

Arndt-Lappe, Sabine, Robin Schrecklinger & Fabian Tomaschek. 2023. Stratification effects without morphological strata, syllable counting effects without counts – modelling English stress assignment with Naive Discriminative Learning. *Morphology* 33(4). 433–457. https://doi.org/10.1007/s11525-022-09399-9 .

Bauer, Laurie, Rochelle Lieber & Ingo Plag. 2013. *The Oxford Reference Guide to English Morphology*. Oxford: OUP.

BNC Consortium, Bordleian Libraries, Oxford. 2007. The British National Corpus, version 3 (BNC XML Edition). http://www.natcorp.ox.ac.uk/.

Bohlin, Ludvig, Daniel Edler, Andrea Lancichinetti & Martin Rosvall. 2014. Community Detection and Visualization of Networks with the Map Equation Framework. In Ying Ding, Ronald Rousseau & Dietmar Wolfram (eds.), *Measuring Scholarly Impact*, 3–34. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-10377-8_1.

Bonami, Olivier & Jana Strnadová. 2019. Paradigm structure and predictability in derivational morphology. *Morphology* 29(2). 167–197. https://doi.org/10.1007/s11525-018-9322-6.

Breiss, Canaan. 2024. When bases compete: A voting model of lexical conservatism. *Language* 100(2). 308–358. https://doi.org/10.1353/lan.2024.a929738.

Collie, Sarah. 2008. English stress preservation: the case for "fake cyclicity." *English Language and Linguistics* 12(3). 505–532.

Dabouis, Quentin & Jean-Michel Fournier. 2024. Opaque morphology and phonology: Historical prefixes in English. *Journal of Linguistics* 1–33. https://doi.org/10.1017/S002222672400015X.

Davies, Mark. 2011. *Corpus of American Soap Operas*. Available from https://www.english-corpora.org/soap/.

Eddington, David. 2000. Spanish stress assignment within the analogical modeling of language. *Language* 76(1). 92–109.

Fournier, Jean-Michel. 2007. From a Latin syllable-driven stress system to a Romance versus Germanic morphology-driven dynamics. *Language Sciences* 29. 218–236.

Ganster, Tammy. 2025. *Variable stress patterns in English complex adjectives: inter-individual differences and the nature of the phonology-morphology interface*. Trier: Ph.D. dissertation, Trier University.

Guierre, Lionel. 1979. *Essai sur l'accentuation en anglais contemporain: elèments pour une synthèse*. Université Paris-VII.

Guierre, Lionel. 2000. Pourquoi la morpho-phonologie. *Publications de l'Université de Pau* 32–46.

Hasenäcker, Jana, Elisabeth Beyersmann & Sascha Schroeder. 2020. Morphological Priming in Children: Disentangling the Effects of School-grade

Nation, Paul & David Beglar. 2007. A vocabulary size test. *The Language Teacher* (31). 9–13.

Pierrehumbert, Janet B. 2006. The statistical basis of an unnatural alternation. In Louis Goldstein, D.H Whalen & Catherine T. Best (eds.), *Laboratory Phonology 8*, 81–106. Berlin: de Gruyter.

Rainer, Franz. 2018. Patterns and niches in diachronic word formation. *Morphology* 18(3). 497. https://doi.org/10.1007/s11525-018-9333-3.

Rebrus, Péter, Péter Szigetvári & Miklós Törkenczy. 2024. No lowering, only paradigms: A paradigm-based account of linking vowels in Hungarian. *Acta Linguistica Academica*. Akademiai Kiado Zrt. 71(1–2). 137–170. https://doi.org/10.1556/2062.2023.00674.

Sano, Shin-Ichiro. 2015. The Role of Exemplars and Lexical Frequency in Rendaku. *Open Linguistics*. Walter de Gruyter GmbH 1(1). https://doi.org/10.1515/opli-2015-0005.

Seiler, Aaron. in prep. *Variability in English Verb Stress: Lexical Distributions, Analogy, and the Nature of Lexical Representations*. Ph.D. dissertation, Trier: Trier University.

Skousen, Royal. 1989. *Analogical Modeling of Language*. Dordrecht: Kluwer.

Skousen, Royal, Thereon Stanford & Nathan Glenn. 2013. *Algorithm::AM*. https://github.com/garfieldnate/Algorithm-AM.

Trevian, Ives. 2003. *Morphoaccentologie et processus d'affixation de l'anglais*. Bern: Peter Lang.

Trevian, Ives. 2007. Stress-neutral endings in contemporary British English: an updated overview. *Language Sciences* 29(2–3). 426–450. https://doi.org/10.1016/j.langsci.2006.12.016.

Trevian, Ives. 2015. *English Suffixes: Stress-assignment Properties, Productivity, Selection and Combinatorial Processes*. Bern: Peter Lang.