Does frequency solve it all?

The interface of frequency and morphological embeddedness

Annika Schebesta – Universität Siegen
Principal Investigators: Gero Kunter, Ingo Plag
Project: EMB – Morphological Embedding and Phonetic Reduction
Associated project to the DFG Research Unit FOR2373: *Spoken Morphology*

Does frequency solve it all?

Do frequency and morphological embeddedness affect the acoustic signal of NNN compounds?

Annika Schebesta – Universität Siegen
Principal Investigators: Gero Kunter, Ingo Plag
Project: EMB – Morphological Embedding and Phonetic Reduction
Associated project to the DFG Research Unit FOR2373: *Spoken Morphology*

What affects variation?

speechrate

higher speechrate leads to more phonetic reduction
(Uhmann 1992, Smith 2002, Raymond et al. 2006)

number of phonological segments

more phonological segments lead to more phonetic reduction
(Lehiste 1972, Turk & Shattuck-Hufnagel 2000)

accentuation

accented units have longer durations
(Turk & Sawusch 1996, de Jong 2004, Kunter 2011)

What affects variation?

lexical frequency

more frequent units undergo more phonetic reduction
(Pluymaekers et al. 2005, Bell et al. 2009, Arnon & Cohen Priva 2013)

phonological neighborhood

more phonological neighbors lead to less reduction (distinctness)
(Wright 2004, Munson & Solomon 2004)

What affects variation?

prosodic boundary strength

boundaries at higher prosodic domains, i.e. intonation phrase (IP) or prosodic utterance (U), affect durations of units closest to the boundary

(Turk & Shattuck-Hufnagel 2000, Turk & Shattuck-Hufnagel 2007, Bergmann 2017)

morphological boundary strength

units at weaker boundaries undergo more phonetic reduction than units at stronger boundaries

(Lehiste 1972, Sproat & Fujimura 1993, Kunter & Plag 2016)

What affects variation?

1 corpus study

2 experimental studies

	corpus study	experiment 1	experiment 2
morphological embeddedness	x interaction	x	
lexical bigram frequency			x
constituent duration	x	x	x
plosive reduction / plosive deletion		x	x

What affects variation?

1 corpus study

2 experimental studies

	corpus study	experiment 1	experiment 2
morphological embeddedness	x interaction	x	
lexical bigram frequency			x
constituent duration	x	x	x
plosive reduction / plosive deletion		x	x

What affects variation?

1 corpus study

2 experimental studies

	corpus study	experiment 1	experiment 2
morphological embeddedness	x interaction	x	
lexical bigram frequency			x
constituent duration	x	x	x
plosive reduction / plosive deletion		x	x

What affects variation?

1 corpus study

2 experimental studies

	corpus study	experiment 1	experiment 2
morphological embeddedness	x interaction	x	
lexical bigram frequency			x
constituent duration	x	x	x
plosive reduction / plosive deletion		x	x

What affects variation?

1 corpus study

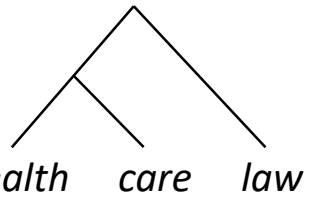
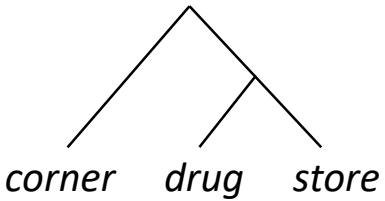
2 experimental studies

production: reading tasks (U of Alberta)

English NNN compounds

healthcare law, corner drugstore

Internal organization of NNN

	LEFT-BRANCHING	RIGHT-BRANCHING
	[health$_{N1}$ care$_{N2}$] law$_{N3}$	corner$_{N1}$ [drug$_{N2}$ store$_{N3}$]
morphological embeddedness		
lexical bigram frequency	30745 224	6 1616

Experiment 1

Data

[guest$_{N1}$ account$_{N2}$] service$_{N3}$ [account$_{N1}$ service$_{N2}$] assistant$_{N3}$

guest$_{N1}$ [account$_{N2}$ service$_{N3}$] account$_{N1}$ [service$_{N2}$ assistant$_{N3}$]

all N1N2/N2N3 bigram frequencies < 20 (COCA)

Data

[guest$_{N1}$ account$_{N2}$] service$_{N3}$ [account$_{N1}$ service$_{N2}$] assistant$_{N3}$

guest$_{N1}$ [account$_{N2}$ service$_{N3}$] account$_{N1}$ [service$_{N2}$ assistant$_{N3}$]

reading task:

The service for accounts is installed for guests.

The guest account service makes their stay more comfortable.

Data

[guest$_{N1}$ accou**nt**$_{N2}$] **s**ervice$_{N3}$ [accou**nt**$_{N1}$ **s**ervice$_{N2}$] assistant$_{N3}$

guest$_{N1}$ [accou**nt**$_{N2}$ **s**ervice$_{N3}$] accou**nt**$_{N1}$ [**s**ervice$_{N2}$ assistant$_{N3}$]

consonant sequences: nasal/fricative + PLOSIVE + fricative/nasal

(nts, ntf, stn, stm, ftn, ftm)

Data

[guest$_{N1}$ accou**nt**$_{N2}$] **s**ervice$_{N3}$ [accou**nt**$_{N1}$ **s**ervice$_{N2}$] assistant$_{N3}$

guest$_{N1}$ [accou**nt**$_{N2}$ **s**ervice$_{N3}$] accou**nt**$_{N1}$ [**s**ervice$_{N2}$ assistant$_{N3}$]

Prediction 1: more plosive deletion within embedded compound

Data

[guest$_{N1}$ accou**nt**$_{N2}$] **s**ervice$_{N3}$ [accou**nt**$_{N1}$ **s**ervice$_{N2}$] assistant$_{N3}$

guest$_{N1}$ [accou**nt**$_{N2}$ **s**ervice$_{N3}$] accou**nt**$_{N1}$ [**s**ervice$_{N2}$ assistant$_{N3}$]

Prediction 1: more plosive deletion within embedded compound

Prediction 2: less plosive deletion between embedded compound and free constituent

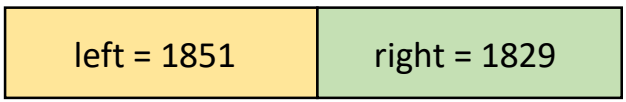
Data

25 **account service** pairs in 4 conditions = 100 compounds per speaker

41 speakers of North American English

3680 NNN compounds

(excluded items: misreadings, pauses, sound quality…)



Statistical analysis: plosive deletion

statistical analysis: glmer modelling

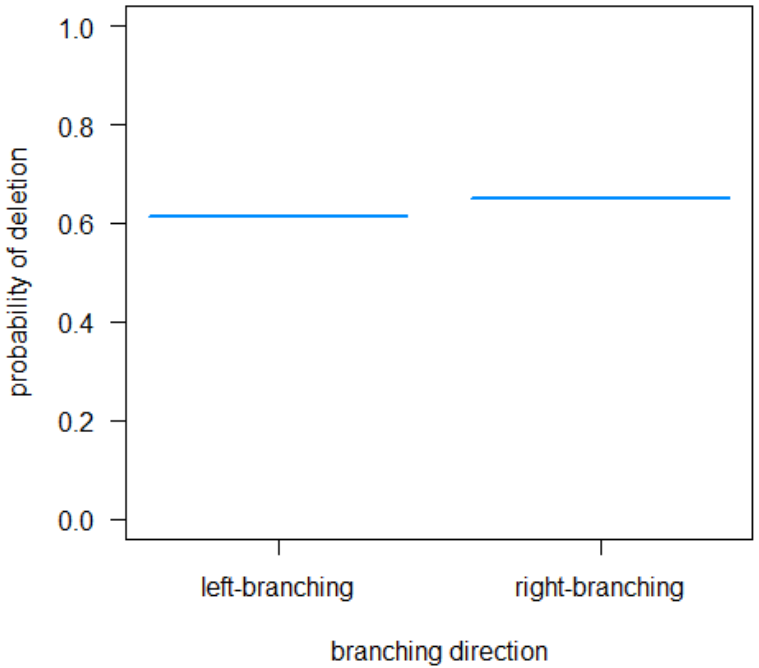
dependent variable:	plosive deletion
central interaction:	boundary * branching
predictors:	frequencies of constituent$_{plosive}$ bigram frequencies N1N2, N2N3 no. of phonological segments constituent$_{plosive}$ no. of phonological segments compound local speechrate consonant sequence pitch range phonological neighborhood constituent$_{plosive}$
random effect:	speaker, constituent

Statistical analysis: plosive deletion

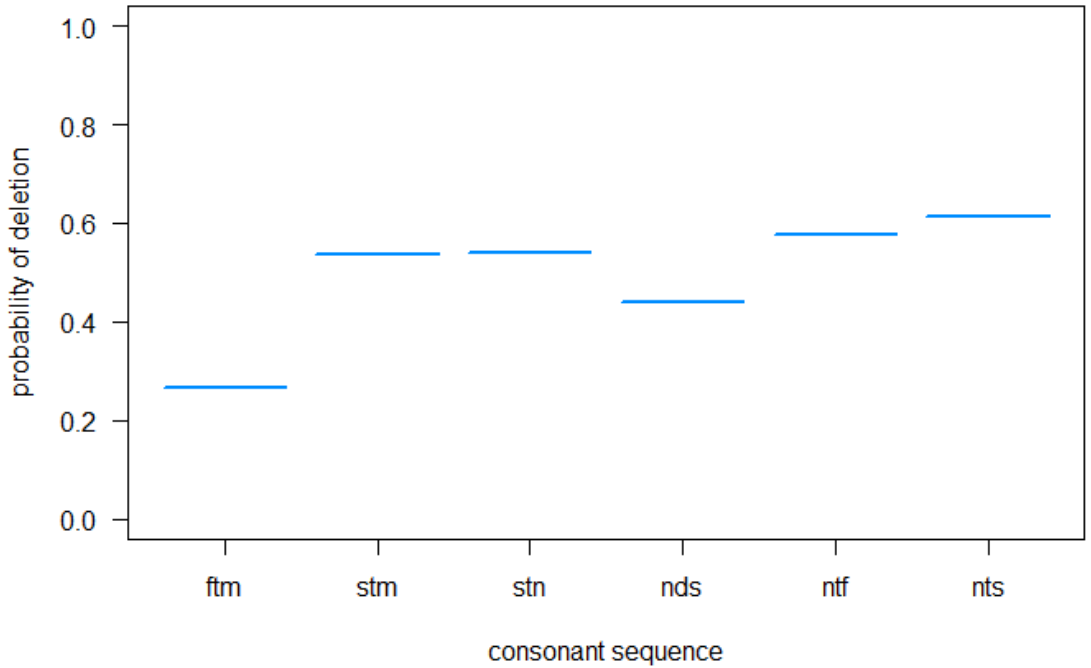
statistical analysis: glmer modelling

dependent variable:	plosive deletion
central interaction:	boundary
	branching
predictors:	frequencies of constituent$_{plosive}$
	bigram frequencies N1N2, N2N3
	no. of phonological segments constituent$_{plosive}$
	no. of phonological segments compound
	local speechrate
	consonant sequence
	pitch range
	phonological neighborhood constituent$_{plosive}$
random effect:	speaker, constituent

Results: plosive deletion reference level: left-branching



Results: plosive deletion reference level: ftm



Results: plosive deletion

[guest$_{N1}$ accou**nt**$_{N2}$] **s**ervice$_{N3}$ [accou**nt**$_{N1}$ **s**ervice$_{N2}$] assistant$_{N3}$

guest$_{N1}$ [accou**nt**$_{N2}$ **s**ervice$_{N3}$] accou**nt**$_{N1}$ [**s**ervice$_{N2}$ assistant$_{N3}$]

Prediction 1: more plosive deletion within embedded compound

not confirmed.

Prediction 2: less plosive deletion between embedded compound and free constituent

not confirmed.

Summary

boundary: no systematic plosive deletion

branching direction: more plosive deletion in right-branching NNN

consonant sequence: plosive deletion related to certain consonant clusters

effect of morphological embeddedness?

Summary

boundary: no systematic plosive deletion

branching direction: more plosive deletion in right-branching NNN

consonant sequence: plosive deletion related to certain consonant clusters

effect of morphological embeddedness? **No.**

Experiment 2

Data

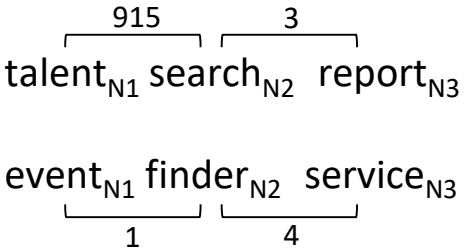
$talent_{N1}$ $search_{N2}$ $report_{N3}$

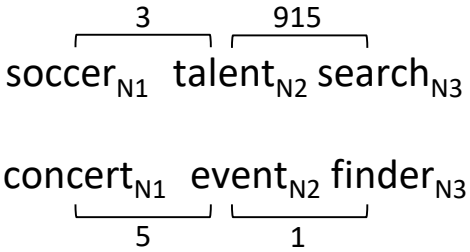
$soccer_{N1}$ $talent_{N2}$ $search_{N3}$

$event_{N1}$ $finder_{N2}$ $service_{N3}$

$concert_{N1}$ $event_{N2}$ $finder_{N3}$

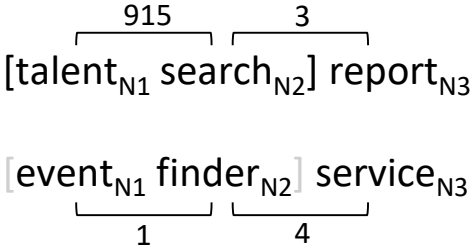
Data

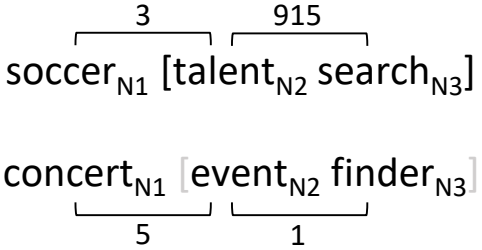




branching direction ← affected by — bigram frequency

Data





branching direction ← affected by bigram frequency

Data

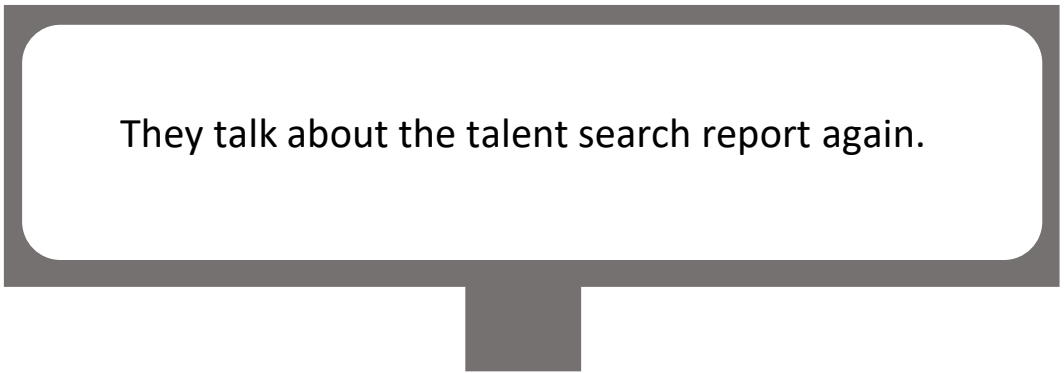
[talent$_{N1}$ search$_{N2}$] report$_{N3}$

soccer$_{N1}$ [talent$_{N2}$ search$_{N3}$]

[event$_{N1}$ finder$_{N2}$] service$_{N3}$

concert$_{N1}$ [event$_{N2}$ finder$_{N3}$]

reading task:



Data

[talent$_{N1}$ search$_{N2}$] report$_{N3}$

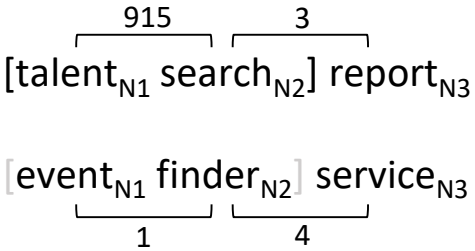
soccer$_{N1}$ [talent$_{N2}$ search$_{N3}$]

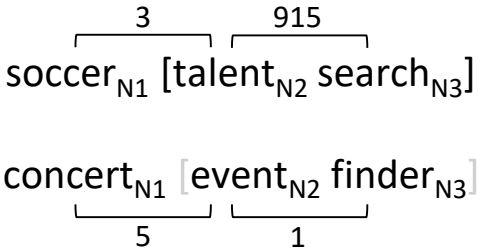
[event$_{N1}$ finder$_{N2}$] service$_{N3}$

concert$_{N1}$ [event$_{N2}$ finder$_{N3}$]

reading task:

Data





consonant sequences: nasal/fricative + PLOSIVE + fricative/nasal

(ftn, ndf, nds, ntf, nts, stn)

Data

[talent$_{N1}$ search$_{N2}$] report$_{N3}$

soccer$_{N1}$ [talent$_{N2}$ search$_{N3}$]

[event$_{N1}$ finder$_{N2}$] service$_{N3}$

concert$_{N1}$ [event$_{N2}$ finder$_{N3}$]

Prediction 1: more plosive deletion in high-frequent bigrams

Data

[talent$_{N1}$ search$_{N2}$] report$_{N3}$

soccer$_{N1}$ [talent$_{N2}$ search$_{N3}$]

[event$_{N1}$ finder$_{N2}$] service$_{N3}$

concert$_{N1}$ [event$_{N2}$ finder$_{N3}$]

Prediction 1: more plosive deletion in high-frequent bigrams

Prediction 2: less plosive deletion in low-frequent bigrams

Data

10 **talent search** / **event finder** pairs in 2 conditions = 40 compounds per speaker

43 speakers of North American English

1172 NNN compounds

(excluded items: misreadings, pauses, sound quality…)

Statistical analysis: plosive deletion

statistical analysis: glmer modelling

dependent variable:	plosive deletion
predictors:	frequency condition (high-frequent / low-frequent) position of bigram in NNN frequencies of constituent$_{plosive}$ bigram frequencies N1N2, N2N3 no. of phonological segments constituent$_{plosive}$ no. of phonological segments compound local speechrate consonant sequence pitch range phonological neighborhood constituent$_{plosive}$
random effect:	speaker, constituent

Statistical analysis: plosive deletion

statistical analysis: glmer modelling

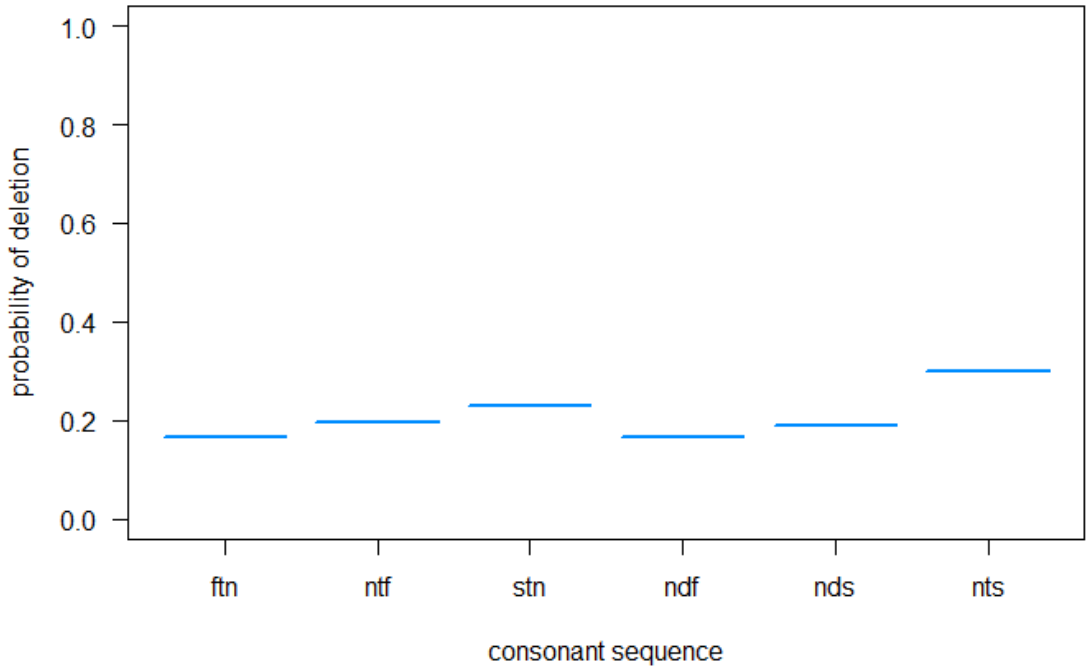
dependent variable: plosive deletion

predictors:

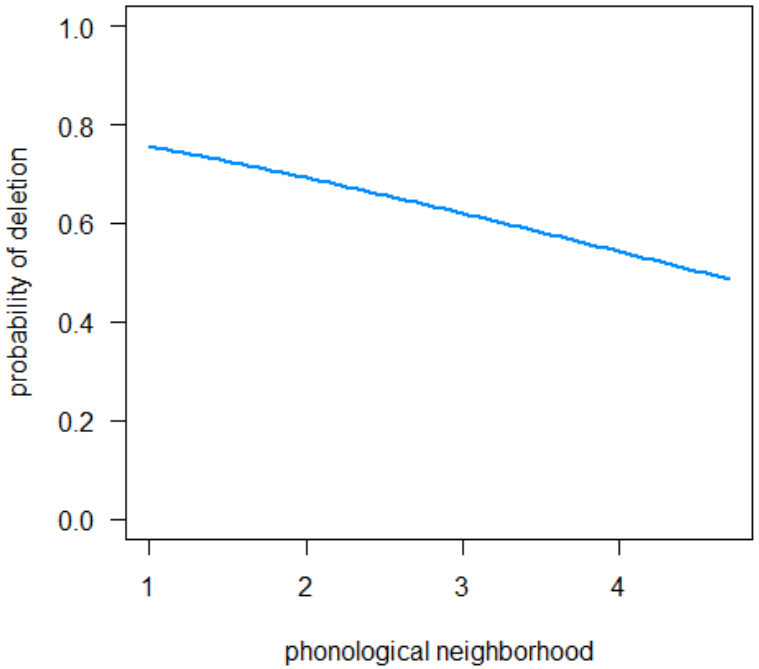
frequency condition (high-frequent / low-frequent)
position of bigram in NNN
frequencies of constituent$_{plosive}$
bigram frequencies N1N2, N2N3
no. of phonological segments constituent$_{plosive}$
no. of phonological segments compound
local speechrate
consonant sequence
pitch range
phonological neighborhood constituent$_{plosive}$

random effect: speaker, constituent

Results: plosive deletion reference level: ftn



Results: plosive deletion



Results: plosive deletion

[talent$_{N1}$ search$_{N2}$] report$_{N3}$ soccer$_{N1}$ [talent$_{N2}$ search$_{N3}$]

[event$_{N1}$ finder$_{N2}$] service$_{N3}$ concert$_{N1}$ [event$_{N2}$ finder$_{N3}$]

Prediction 1: more plosive deletion in high-frequent bigrams

not confirmed.

Prediction 2: less plosive deletion in low-frequent bigrams

not confirmed.

Summary

frequency: as much plosive deletion in high-frequent as in low-frequent bigrams

consonant sequence: plosive deletion related to certain consonant clusters

phonological neighborhood: the more neighbors, the less plosive deletion

effect of bigram frequency?

Summary

frequency: as much plosive deletion in high-frequent as in low-frequent bigrams

consonant sequence: plosive deletion related to certain consonant clusters

phonological neighborhood: the more neighbors, the less plosive deletion

effect of bigram frequency? **No.**

Rating of branching
(preliminary analysis!)

Rating

online experiment: rating task

students from University of Alberta

220 NNN compounds in their original carrier sentences

→ 100 in context sentences (experiment 1)

→ 40 in carrier sentences (experiment 2)

→ 80 in context sentences (yet another reading task…)

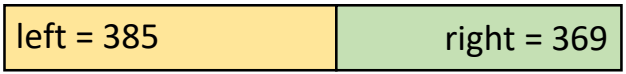
HANDLE WITH CARE, PRELIMINARY RESULTS!

Is branching really branching?

experiment 1 assumed:

native speaker judgment:

experiment 2 assumed:



high-frequent bigrams

native speaker judgment:

experiment 2 assumed:

low-frequent bigrams

native speaker judgment:

Thank you, Ben Tucker,

for having me in Edmonton and support during the experiments!

Thank you, Ingo Plag,

for supporting this dissertation project (in many ways)!

Thank you all for listening!

References

Arnon, Inbal and Uriel Cohen Priva (2013). More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and speech*, 56(Pt 3):349-371.

Bell, Alan, Jason M. Brenier, Michelle Gregory, Cynthia Girand and Dan Jurafsky. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60 (1), 92-111.

Bergmann, Pia. 2017. *Morphologisch komplexe Wörter im Deutschen. Prosodische Struktur und phonetische Realisierung.* Berlin: Language Science Press.

de Jong, Kenneth. 2004. Stress, lexical focus and segmental focus in English. Patterns of variation in vowel duration. *Journal of Phonetics* 32 (4), 493-516.

Kunter, Gero. 2011. *Compound stress in English. The phonetics and phonology of prosodic prominence*. Berlin: De Gruyter.

Kunter, Gero and Ingo Plag. 2016. Morphological embedding and phonetic reduction. The case of triconstituent compounds. *Morphology* 26 (2), 201-227.

Lehiste, Ilse. 1972. The timing of utterances and linguistic boundaries. *The Journal of the Acoustical Society of America* 51 (6B), 2018-2024.

Munson, Benjamin and Nancy Pearl Solomon. 2004. The effect of phonological neighborhood density on vowel articulation. *Journal of speech, language, and hearing research : JSLHR*, *47*(5), 1048–1058.

Pluymaekers, Mark, Mirjam Ernestus and R. Harald Baayen. 2005. Lexical frequency and acoustic reduction in Dutch. *The Journal of the Acoustical Society of America* 118 (4), 2561-2569.

Sproat, Richard and Osamu Fujimura. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21, 291-311.

Turk, Alice E. and James R. Sawusch. 1996. The processing of duration and intensity cues to prominence. *The Journal of the Acoustical Society of America* 99 (6), 3782-3790.

Turk, Alice E. and Stephanie Shattuck-Hufnagel. 2000. Word-boundary related duration patterns in English. *Journal of Phonetics* 28 (4), 397-440.

Turk, Alice E. and Stephanie Shattuck-Hufnagel. 2007. Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics* 35 (4), 445-472.

Uhmann, S. 1992. Contextualizing relevance: On some forms and functions of speech rate changes in everyday conversation. In Auer, P. and Di Luzio, A., editors, *The contextualization of language*, pages 297-336. John Benjamins, Amsterdam.

Wright R. 2004. Factors of lexical competition in vowel articulation. In: Local J, Ogden R, Temple R, editors. *Papers in Laboratory Phonology VI*. Cambridge: Cambridge University Press, pp. 26–50.

Exp. 1

[guest$_{N1}$ accou**nt**$_{N2}$] **s**ervice$_{N3}$

[accou**nt**$_{N1}$ **s**ervice$_{N2}$] assistant$_{N3}$

guest$_{N1}$ [accou**nt**$_{N2}$ **s**ervice$_{N3}$]

accou**nt**$_{N1}$ [**s**ervice$_{N2}$ assistant$_{N3}$]

Prediction 1: more plosive deletion within embedded compound

not confirmed.

Prediction 2: less plosive deletion between embedded compound and free constituent

not confirmed.

Exp. 2

[tale**nt**$_{N1}$ **s**earch$_{N2}$] report$_{N3}$

soccer$_{N1}$ [tale**nt**$_{N2}$ **s**earch$_{N3}$]

[eve**nt**$_{N1}$ **f**inder$_{N2}$] service$_{N3}$

concert$_{N1}$ [eve**nt**$_{N2}$ **f**inder$_{N3}$]

Prediction 1: more plosive deletion in high-frequent bigrams

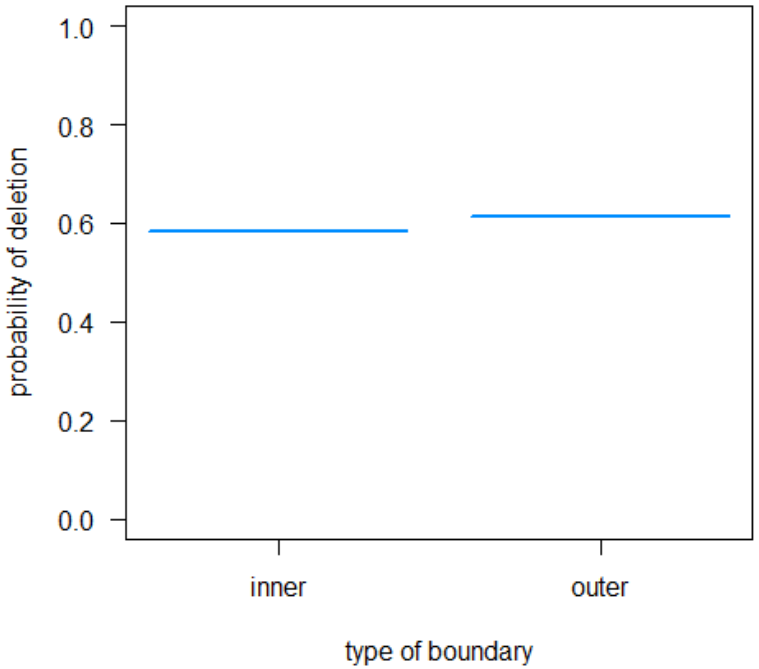
not confirmed.

Prediction 2: less plosive deletion in low-frequent bigrams

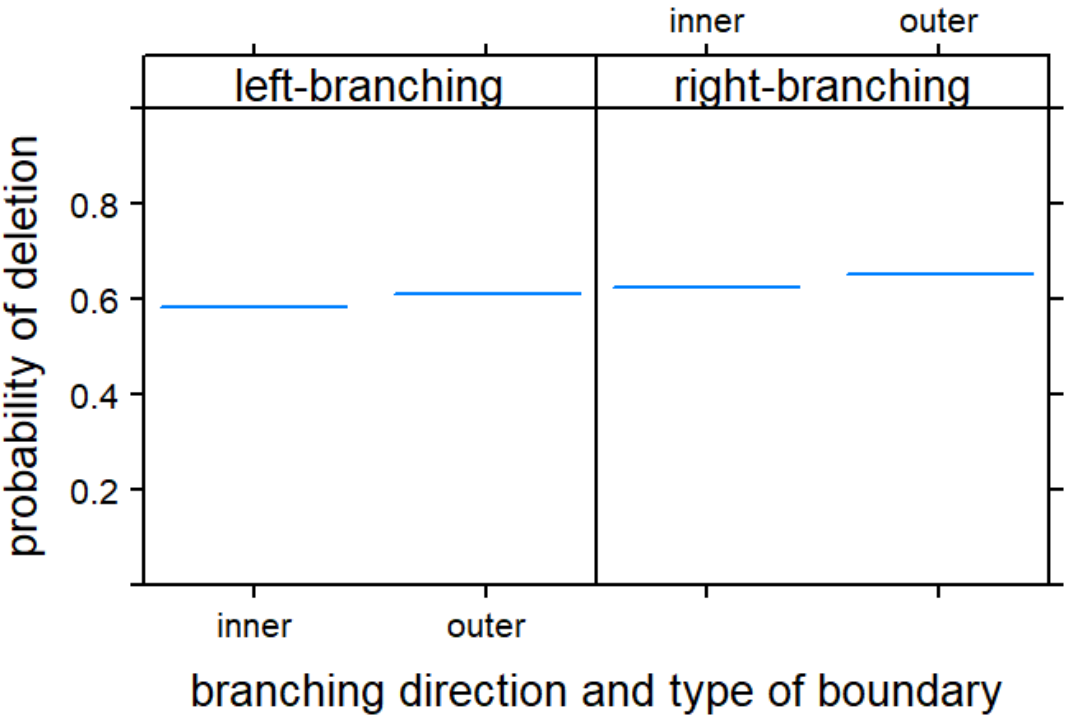
not confirmed.

Appendix

Results Exp1:
plosive deletion by boundary



Results Exp1:
plosive deletion by boundary * branching



Results Exp2:
plosive deletion by frequency condition