Prefixal Gemination in English: An experimental study on *un-* and *in-*

Sonia Ben Hedia

Heinrich-Heine-Universität Düsseldorf

Cambridge Colloquium
October 25, 2016

(De-)Gemination in English

- Sequence of two identical consonants across a morphological boundary

un-	*un-natural*
in-	*in-numerous, im-material, il-logical, ir-resistable*
dis-	*dis-satisfied*
-ly	*sole-ly, technical-ly*

- Phonetic correlates
  - Gemination: Longer duration than a singleton
  - Degemination: Same duration as a singleton

Overarching research questions

- What is the pattern of germination in English affixation?
- Which factors influence the duration of consonant length on affix boundaries?

- What are the theoretic implications? Which theories are supported, which are falsified?

Methodology

Corpus study	**Experimental study**
Prefixation *in* as in *innumerous* *un* as in *unnatural* *dis* as in *dissolve*	Prefixation *in* as in *innumerous* *un* as in *unnatural* *dis* as in *dissolve*
Suffixation *ly* as in *really*	Suffixation *ly* as in *really*

Methodology

Corpus study	**Experimental study**
Prefixation *in* as in *innumerous* *un* as in *unnatural* *dis* as in *dissolve*	Prefixation *in* as in *innumerous* *un* as in *unnatural* *dis* as in *dissolve*
Suffixation *ly* as in *really*	Suffixation *ly* as in *really*

Methodology

Corpus study	**Experimental study**
Prefixation *in* as in *innumerous* *un* as in *unnatural* *dis* as in *dissolve*	Prefixation *in* as in *innumerous* *un* as in *unnatural* *dis* as in *dissolve*
Suffixation *ly* as in *really*	Suffixation *ly* as in *really*

(De-)Gemination in English

- Sequence of two identical consonants across a morphological boundary

un-	*un-natural*
in-	*in-numerous, im-material, il-logical, ir-resistable*
dis-	*dis-satisfied*
-ly	*sole-ly, technical-ly*

- Phonetic correlates
  - Gemination: Longer duration than a singleton
  - Degemination: Same duration as a singleton
- Theoretical assumption: Degemination is affix- or stratum-dependent

Predictions: Lexical Phonology

	Level 1	Level 2
Morphological Process	in + numerous	
Phonological Process	i/n/umerous	
Phonetic Outcome	i[n]umerous	

Degemination

Predictions: Lexical Phonology

	Level 1	Level 2
Morphological Process	in + numerous	un + natural
Phonological Process	i/n/umerous	u/nn/atural
Phonetic Outcome	i[n]umerous	u[n:]atural

Degemination **Gemination**

Predictions: Morphological Separability

- Phonetic realization is dependent on morphological separability
- more separable → less reduction

  (e.g. Hay 2003, Smith et al. 2012, Ben Hedia & Plag 2016, Plag 2016)

More separable complex words geminate.

Less separable complex words degeminate.

- Separability:
  - Semantic Transparency: opaque vs. transparent
  - Type of Root: bound root vs. word
  - Relative Frequency: relative frequency of base and derivative

Empirical evidence?

- Only few studies empirically investigated gemination in English

- Corpus Study (Ben Hedia & Plag 2016)
  - *dis-* geminates
  - *-ly* degeminates

- *un-* geminates (Kaye 2005, Oh and Redford 2013, Ben Hedia & Plag 2016)

- *in-* can geminate
  - Type-dependent (Oh and Redford 2013)
  - speaker-dependent (Kaye 2005)
  - *in-* geminates (Ben Hedia and Plag 2016)

- Problems: Only very small set of types
  Contradictory results

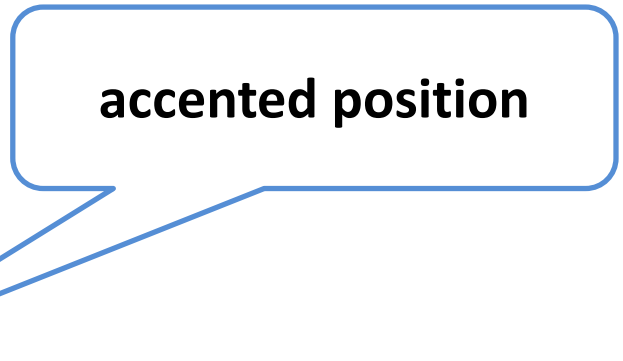
This study

- Reading experiment
- 2 affixes : *un-* and *in-*
- Comparison of nasal duration in 3 different environments

	Morphological geminate	**Singleton in base**	**Singleton in complex word**	
un	unnatural (n#n)	natural (#n)	uneven (n#V)	
im	immature (m#m)	mature (#m)	impossible (m#p)	
in	innumerous (n#n)	numerous (#n)	intolerant (n#t)	inexplicit (n#V)

Experiment

- 183 types
- Items are put in carrier sentences



John said *UNNATURAL* again.

John tells me *NATURAL* again.

It is John who said *unnatural* again, NOT HENRY.

Experiment

- 183 types
- Items are put in carrier sentences

John said *UNNATURAL* again.

John tells me *NATURAL* again.

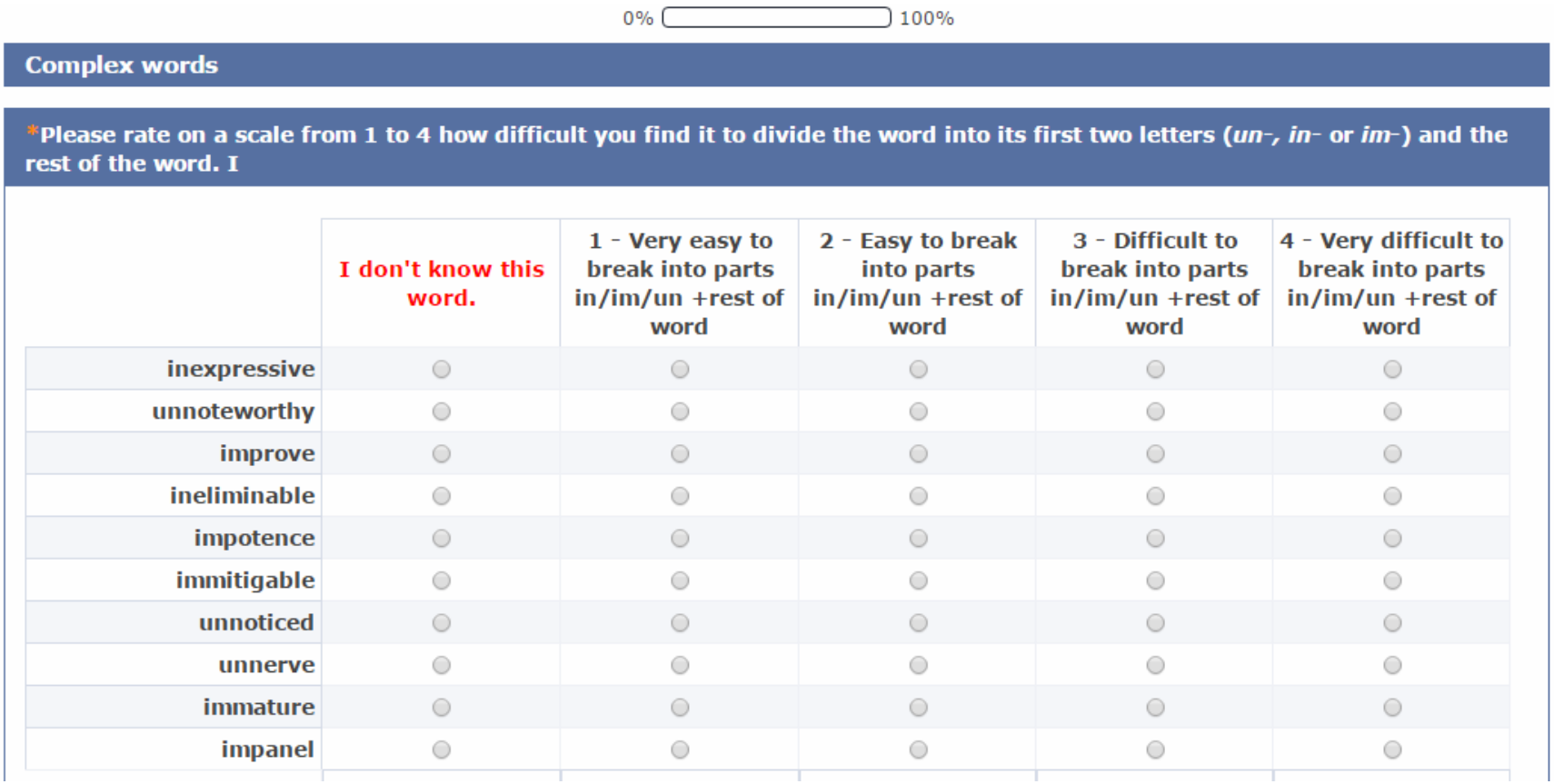
It is John who said *unnatural* again, NOT HENRY.

unaccented position

Experiment

- 29 participants (native speakers of British English) read the sentences
- Separability Rating

Separability Rating



0% 100%

Complex words

*Please rate on a scale from 1 to 4 how difficult you find it to divide the word into its first two letters (*un-*, *in-* or *im-*) and the rest of the word. I

	I don't know this word.	1 - Very easy to break into parts in/im/un +rest of word	2 - Easy to break into parts in/im/un +rest of word	3 - Difficult to break into parts in/im/un +rest of word	4 - Very difficult to break into parts in/im/un +rest of word
inexpressive	○	○	○	○	○
unnoteworthy	○	○	○	○	○
improve	○	○	○	○	○
ineliminable	○	○	○	○	○
impotence	○	○	○	○	○
immitigable	○	○	○	○	○
unnoticed	○	○	○	○	○
unnerve	○	○	○	○	○
immature	○	○	○	○	○
impanel	○	○	○	○	○

Experiments

- 29 participants (native speakers of British English) read the sentences
- Separability Rating
- Items were manually segmented
- Acoustic measurements
- Items were coded (frequencies, stress….)

Data Overview

	Morphological geminate	Singleton in base	Singleton in complex word	
un	535	549	676	
im	490	458	610	
in	88	77	422	614

Statistical Modelling

- Multiple regression with **nasal duration** as dependent variable
- Speaker and Item as random effects

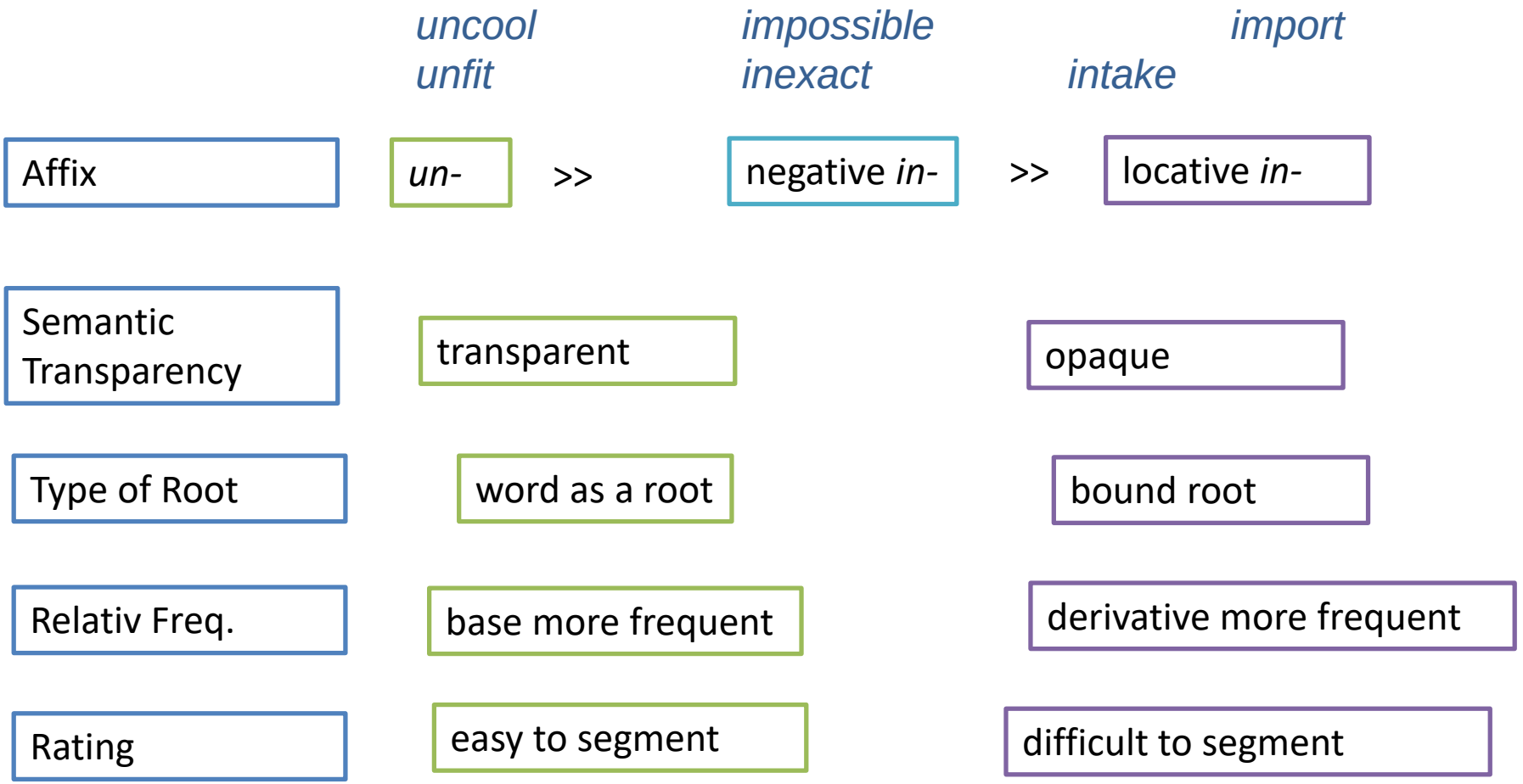
Statistical Modelling

- Variables :
  - Environment
  - Preceding Segment Duration
  - Following Segment
  - Following Segment Duraon
  - Local Speech Rate
  - Global Speech Rate
  - Stress
  - Accentuation
  - Word Form Frequency
  - Order
  - Affix
  - Semantic Transparency
  - Rating
  - Relative Frequency
  - Type of Root

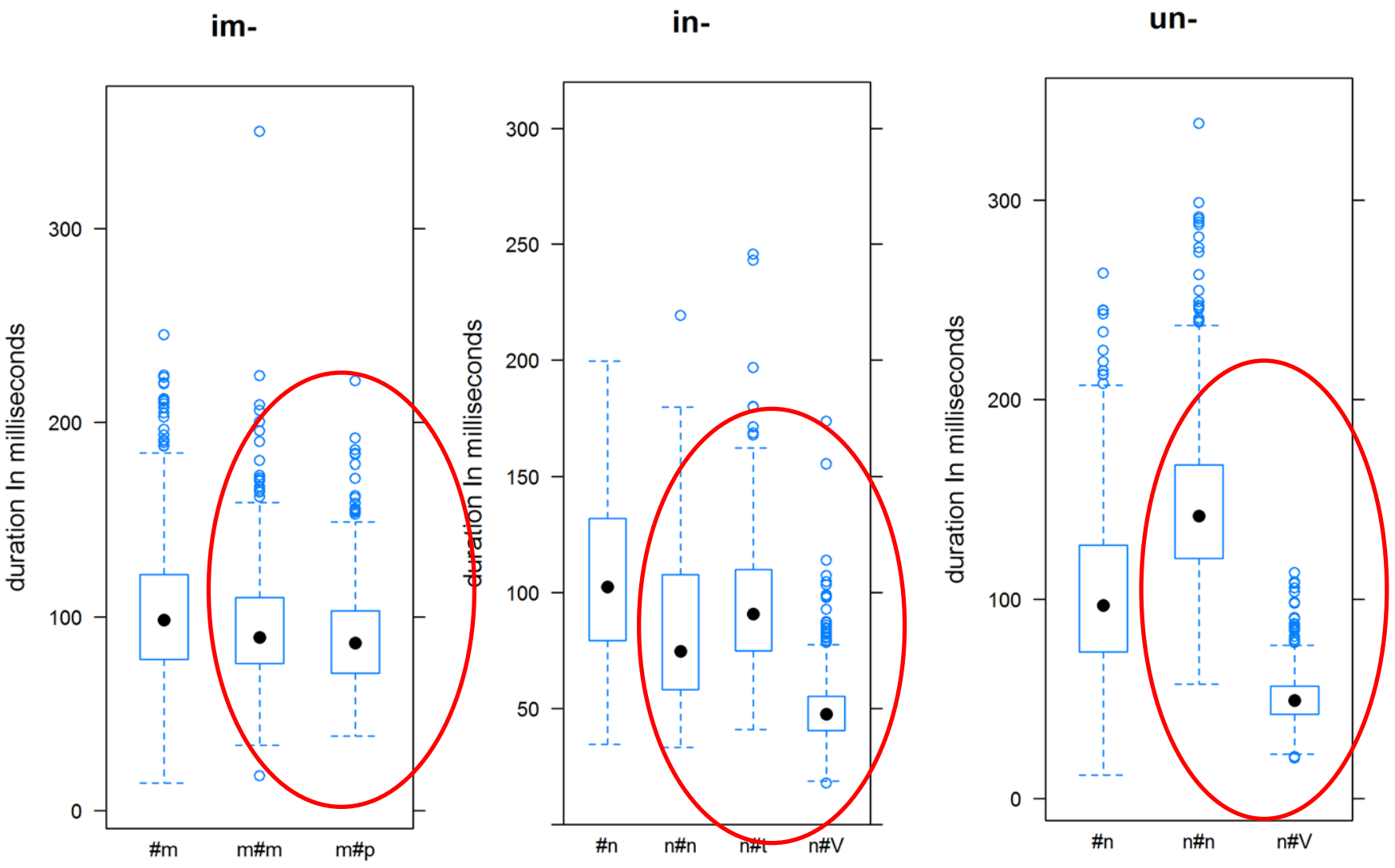
Statistical Modelling

- Variables :
  - Environment
  - Preceding Segment Duration
  - Following Segment
  - Following Segment Duraon
  - Local Speech Rate
  - Global Speech Rate
  - Stress
  - Accentuation
  - Word Form Frequency
  - Order
  - Affix
  - Semantic Transparency
  - Rating
  - Relative Frequency
  - Type of Root

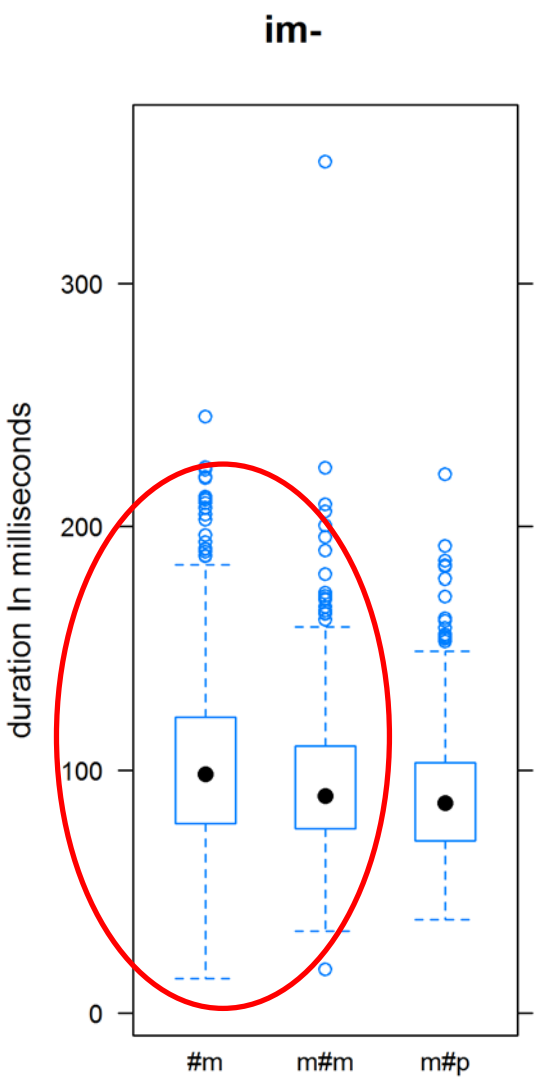
Morphological Separability

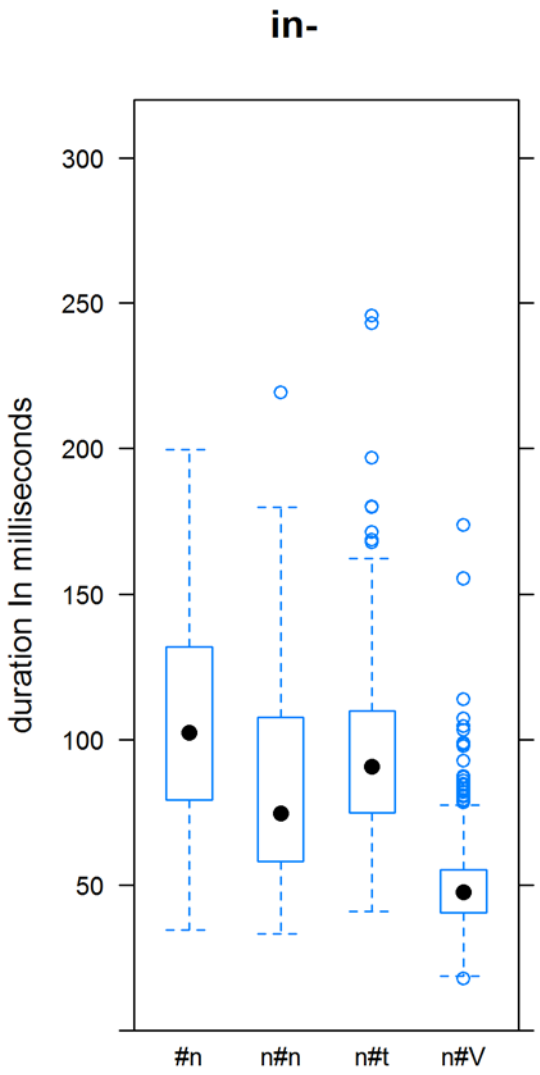


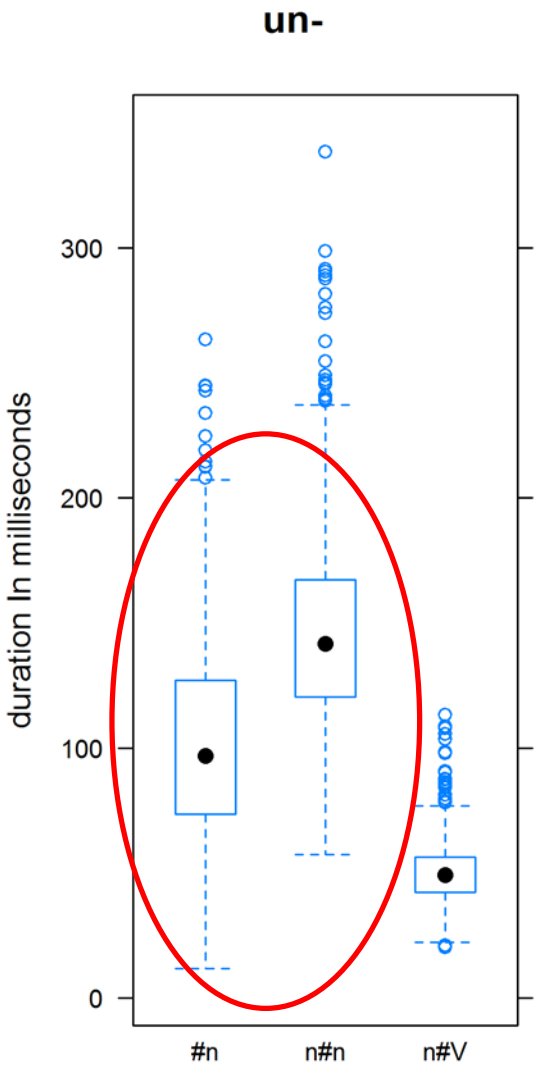
Results: Overview



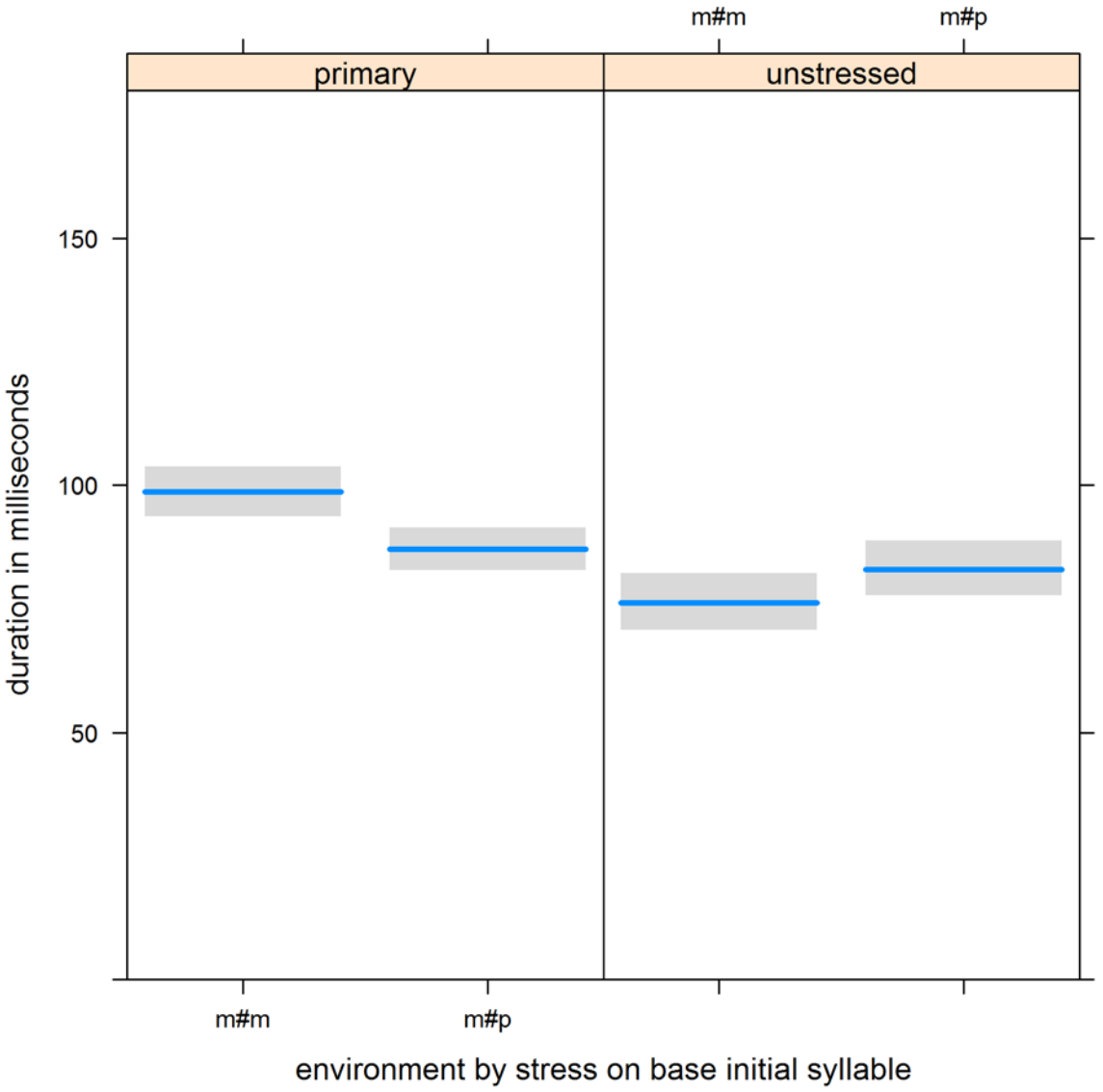
Results: Overview





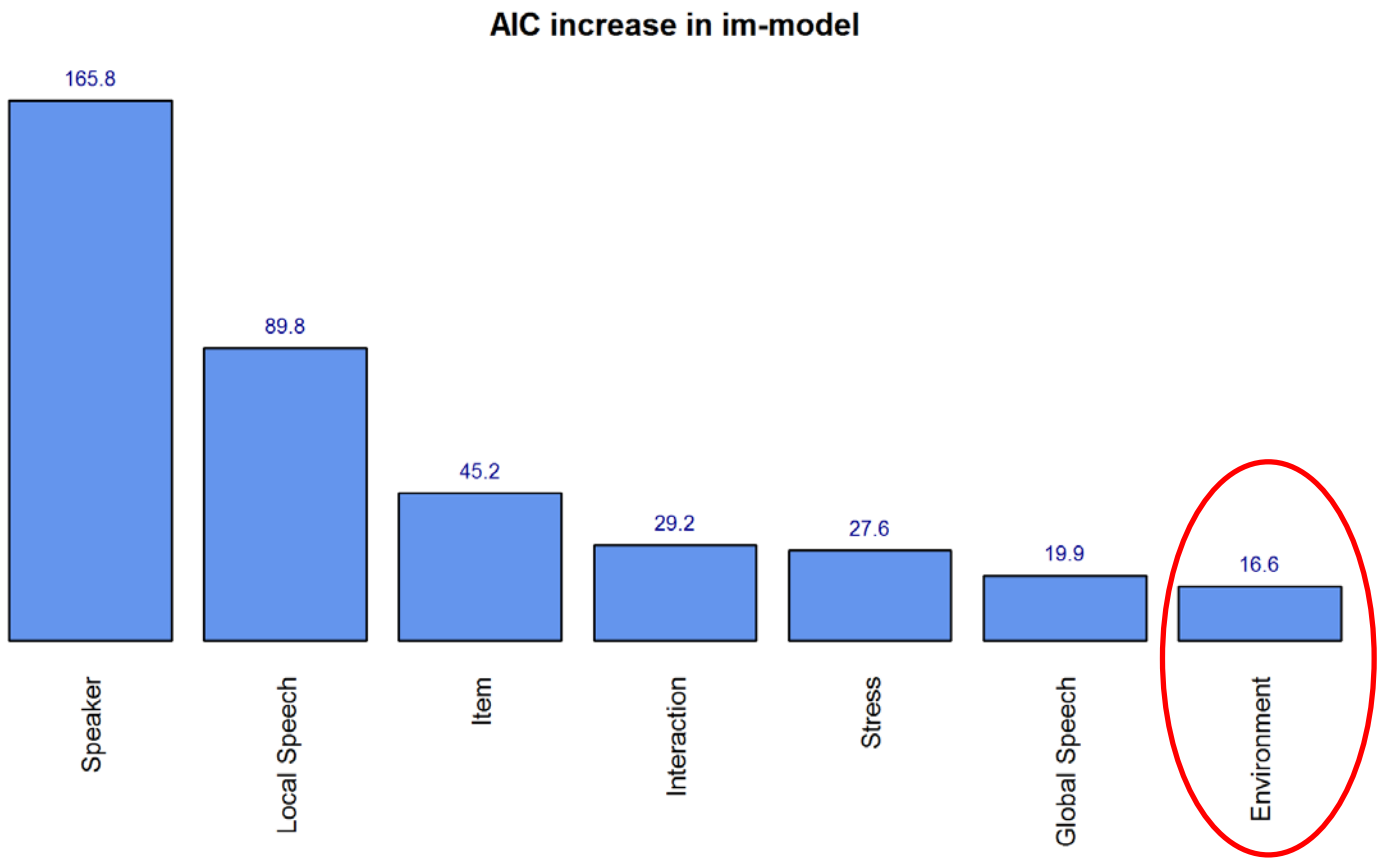


Results *in-*: m#m vs. m#p

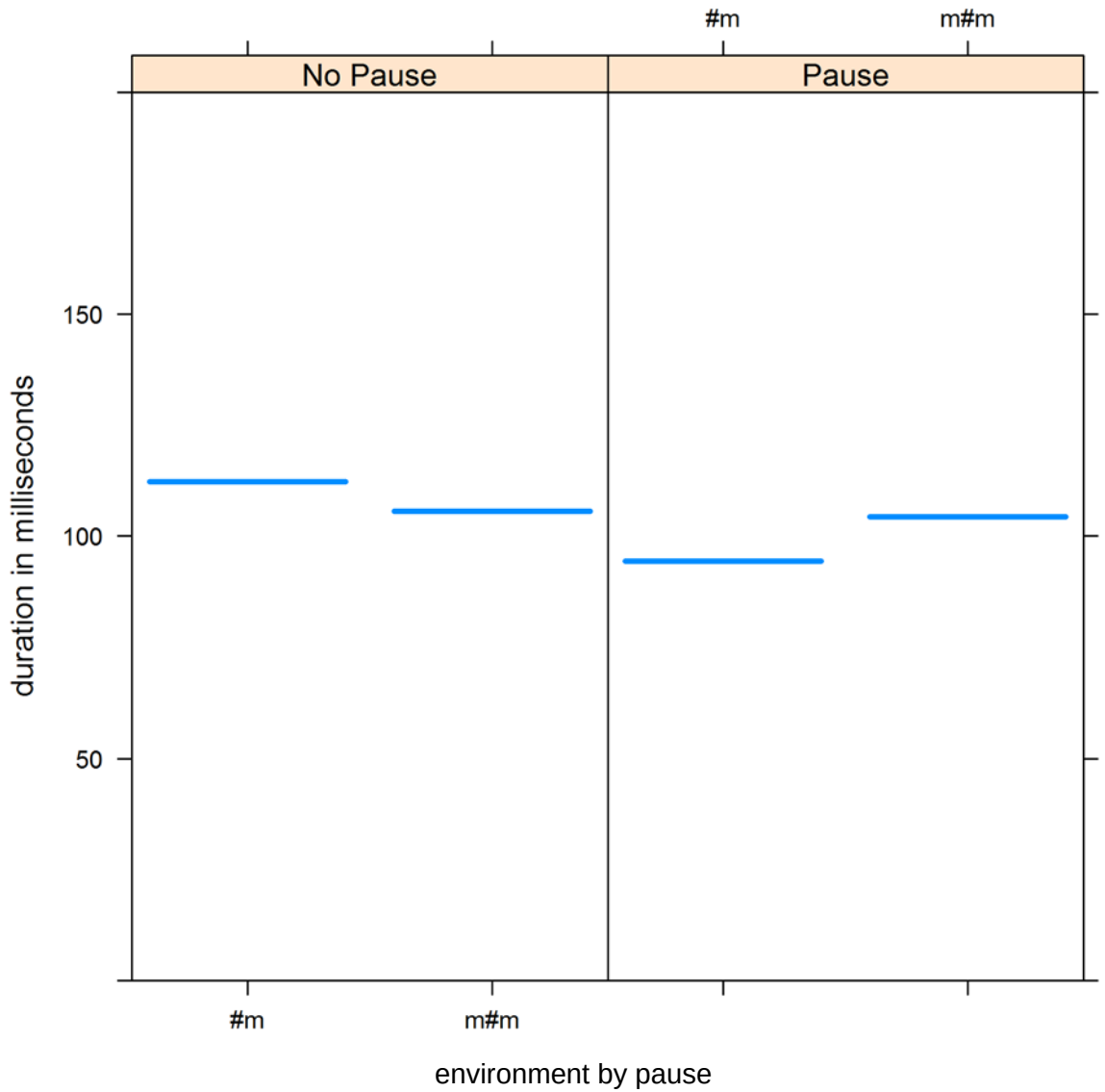


- $R^2 = 0.56$
- Covariates show expected effects
- Primary stress on base intial syllable: Doubles are 11 ms longer than singles
- Unstressed base intial syllable: Doubles are as long as singles

Results *in-*: m#m vs. m#p

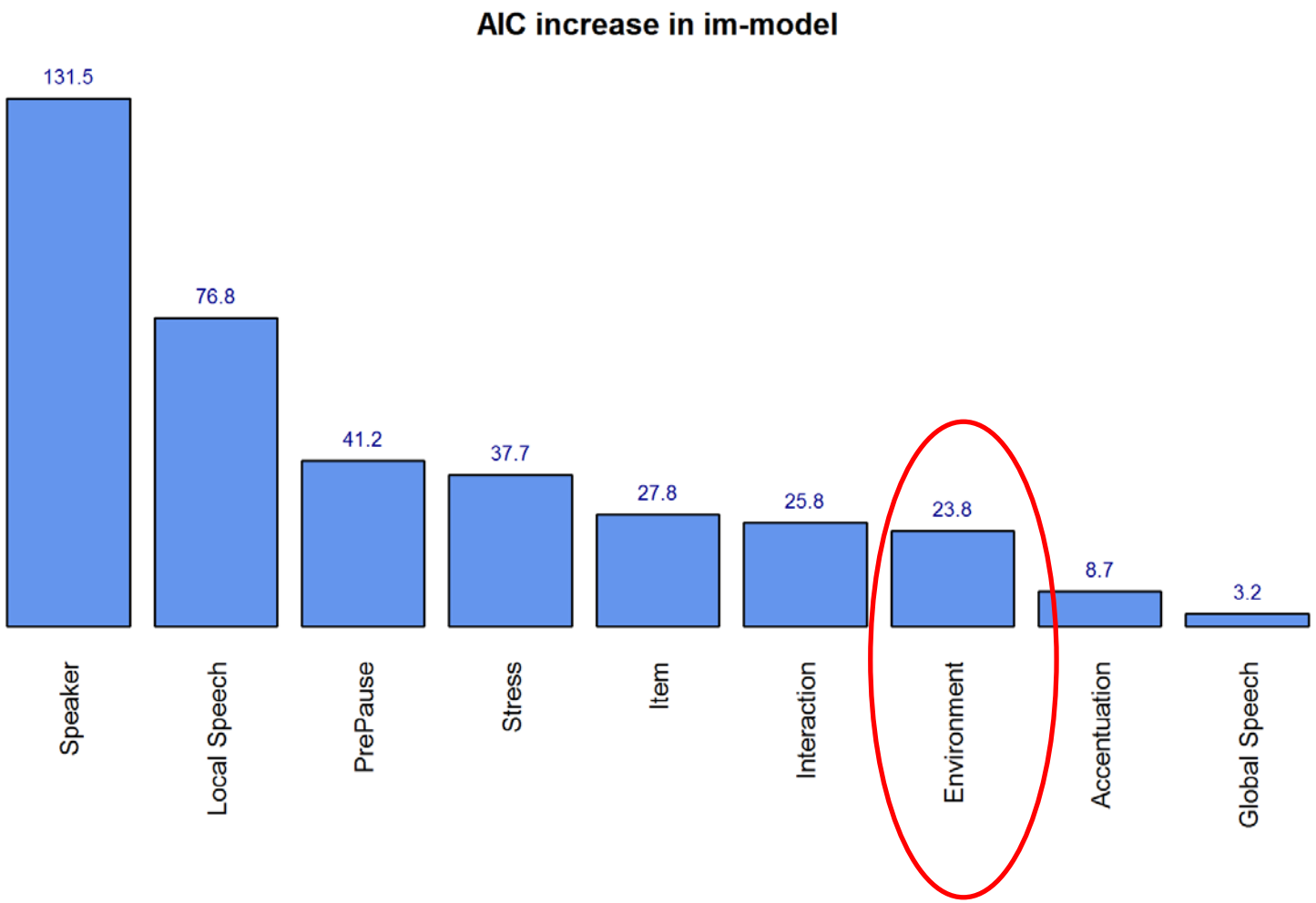


Results *in*-: m#m vs. #m

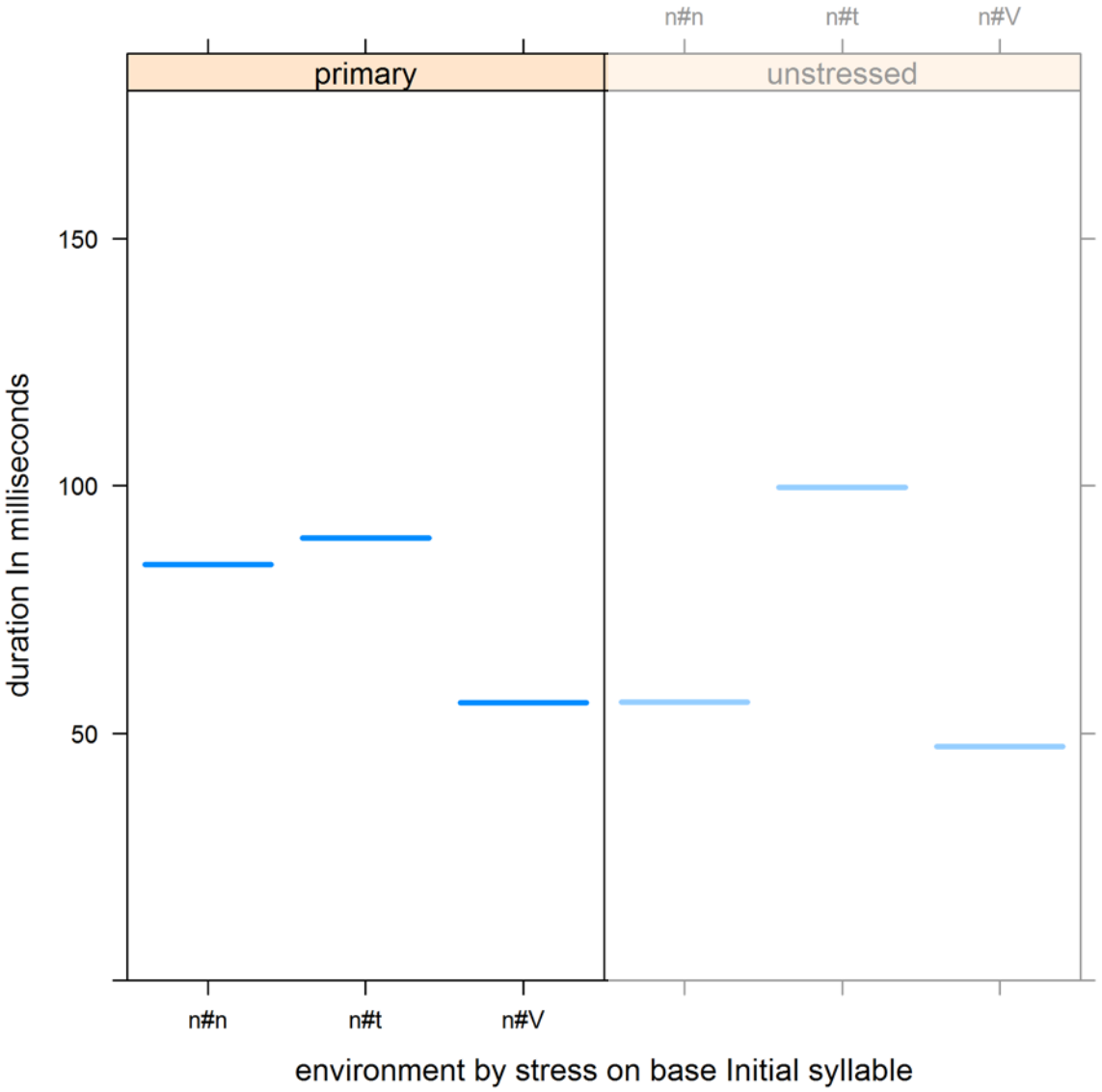


- R^2 =0.58
- Covariates show expected effects
- NoPause: Doubles and singles are of the same duration
- Pause: Doubles are 10 ms longer than singles

Results *in-*: m#m vs. #m

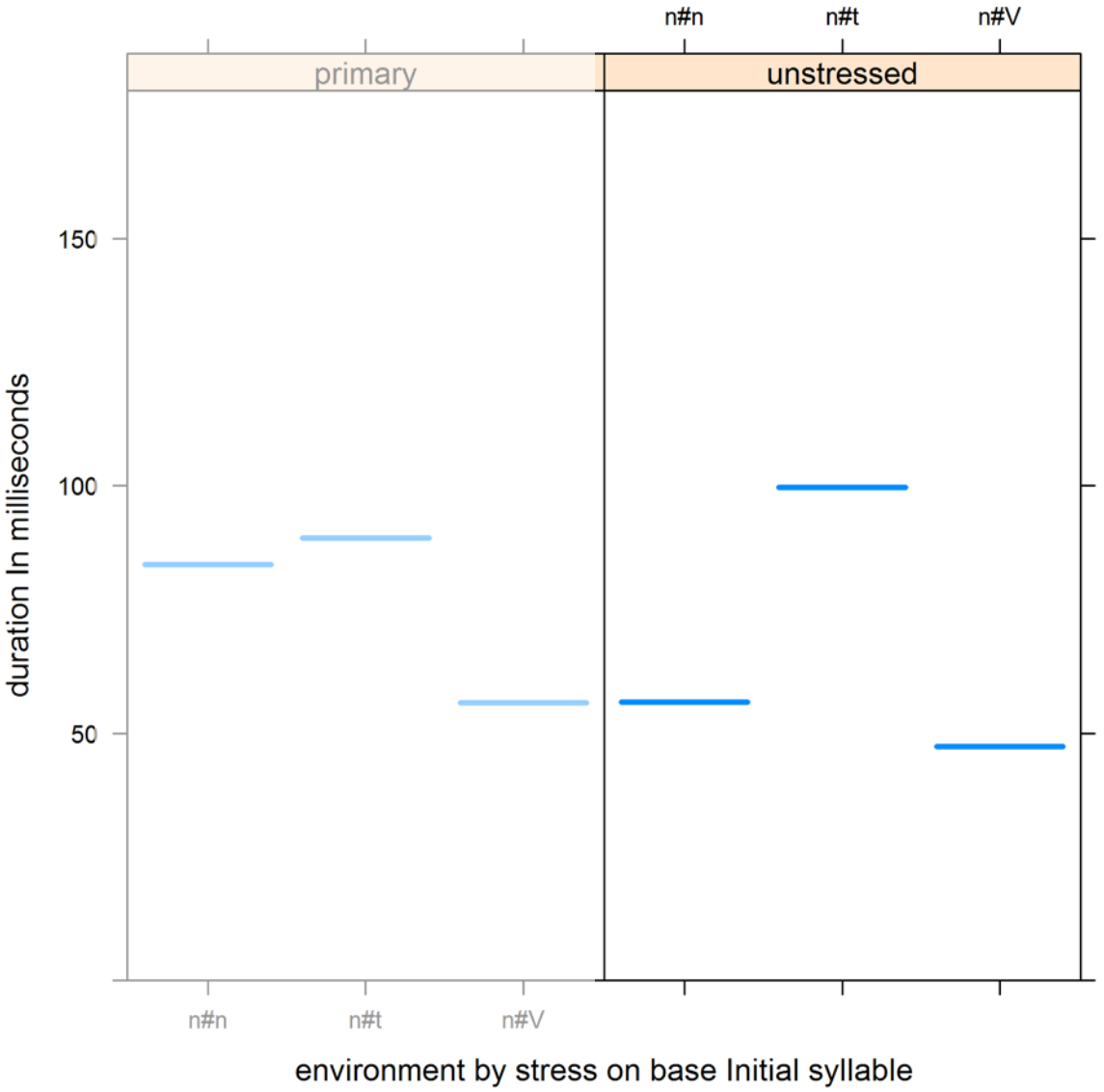


Results *in-*: n#n vs. n#t vs. n#V



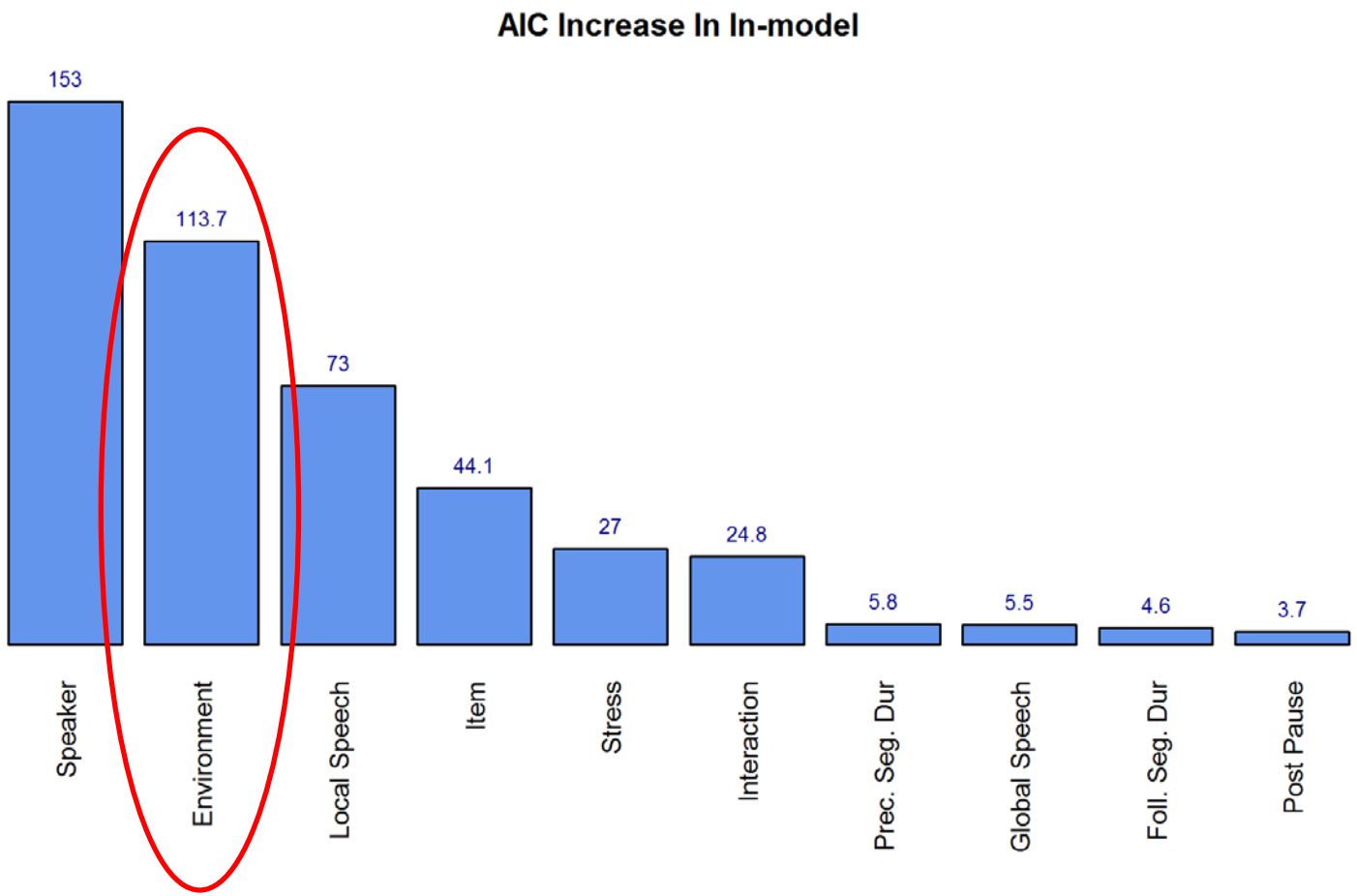
- R^2 =0.74
- Covariates show expected effects
- Doubles are as long as singles with a following stop
- Doubles are 27 ms longer than singles with a following vowel

Results *in-*: n#n vs. n#t vs. n#V



- R^2 =0.74
- Covariates show expected effects
- Doubles are 53 ms shorter than singles with a following stop
- Doubles are 8 ms longer than singles with a following vowel

Results *in-*: n#n vs. n#t vs. n#V



Summary: *in-*

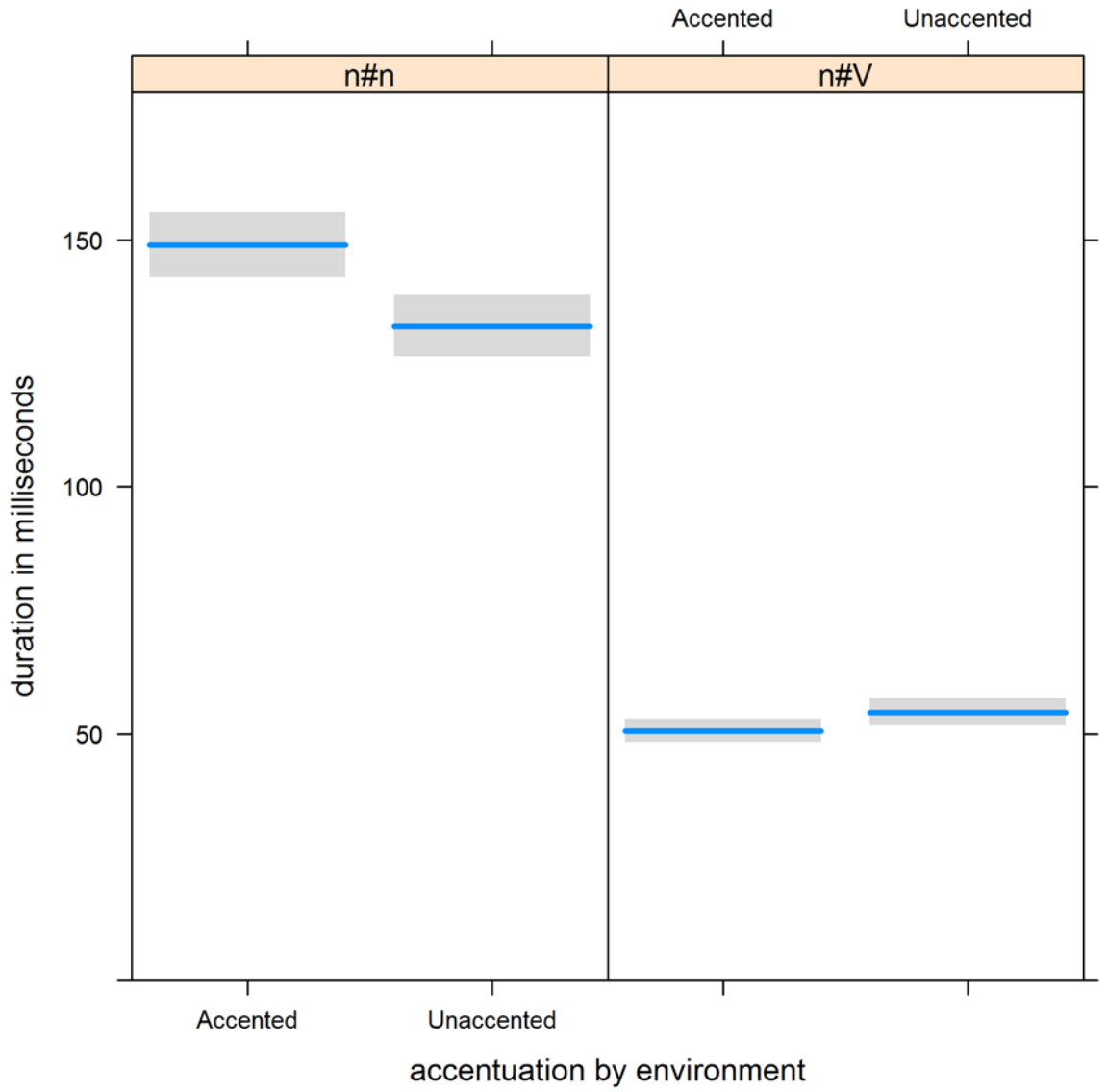
For *im-*:

- Only if there is stress on base intial syllable: Doubles are slighlty longer than singles with a following stop
- Only after a pause: Doubles are slightly longer than singles in base words
- Environment is not a powerful predictor

For *in-*:

- Stressed base intial syllable: Doubles are longer than singles with a following vowel
- Unstressed base intial syllable: Doubles are slightly longer than singles with a following vowel
- Doubles are never longer than singles with a following stop
- Environment is a powerful predictor

Results *un-*: n#n vs. n#V



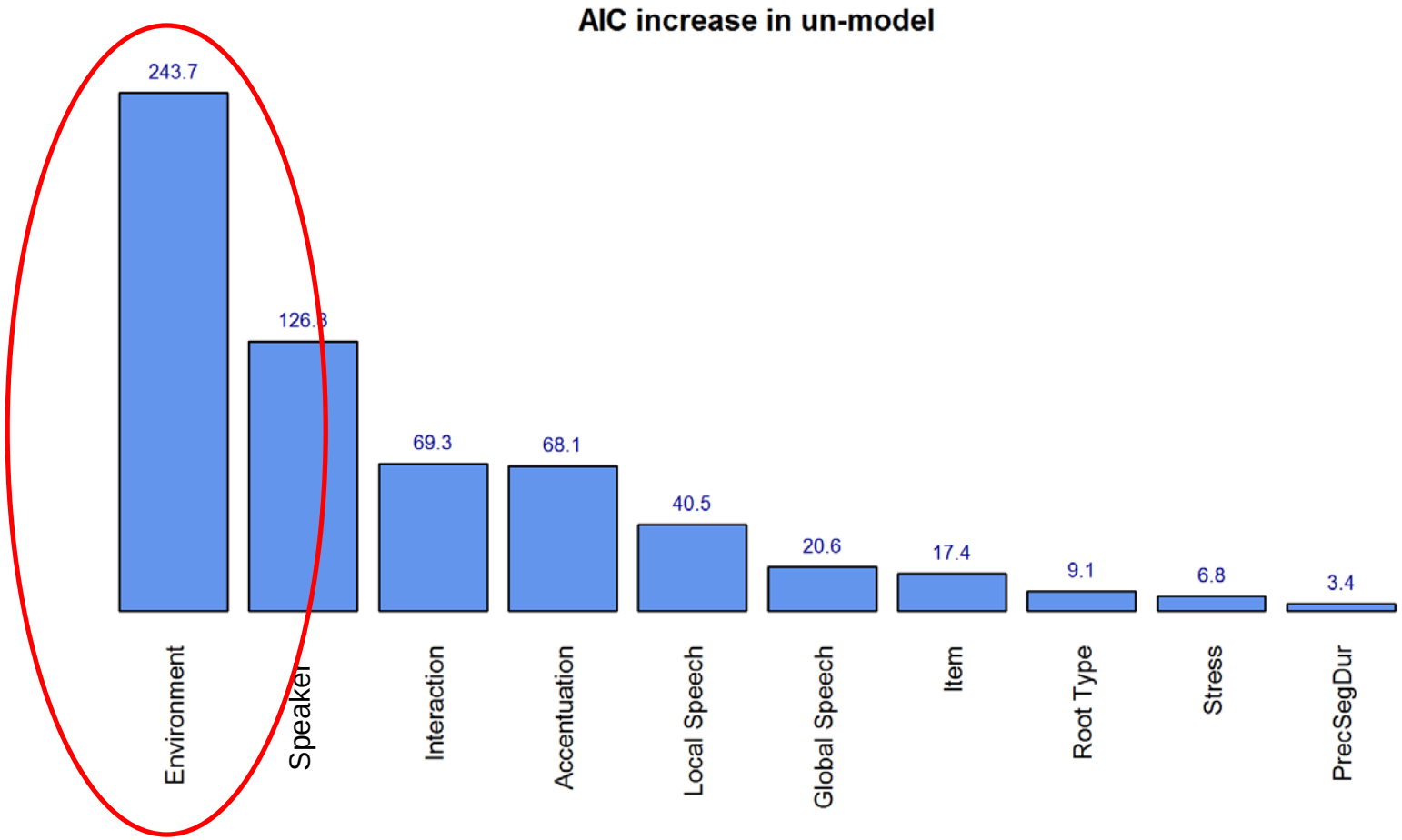
$R^2 = 0.90$

Covariates show expected effects

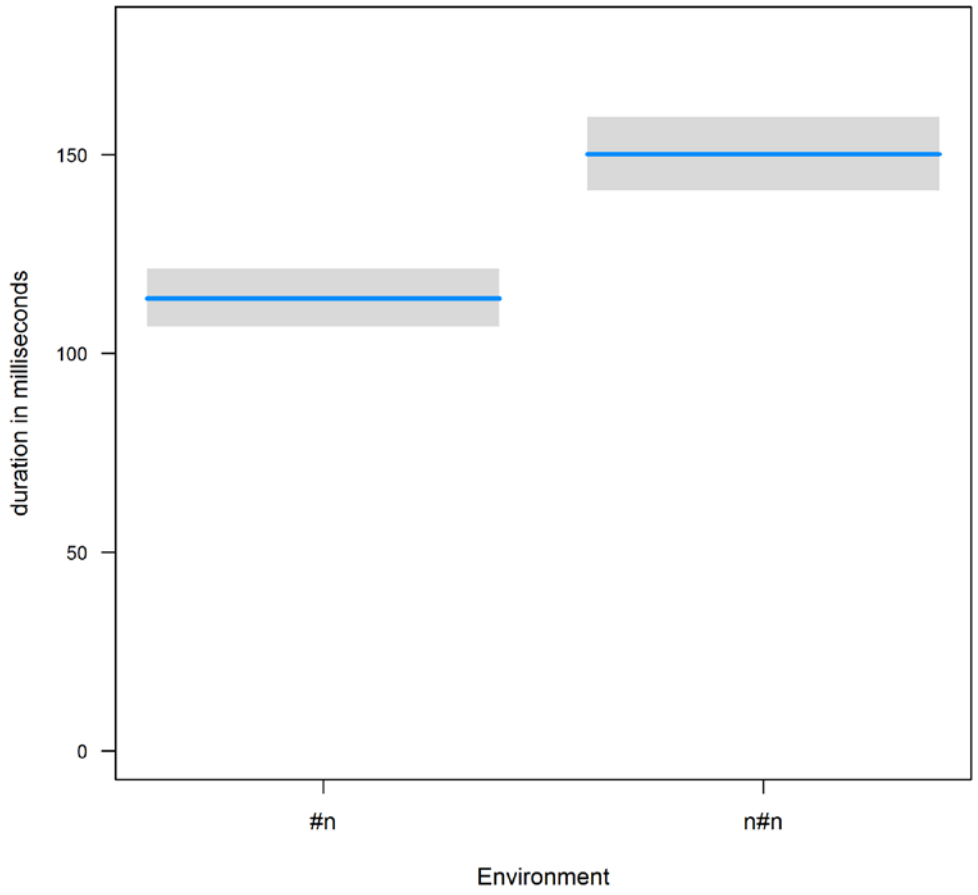
Accented doubles are 98 ms longer than accented singles

Unaccented doubles are 78 ms longer than unaccented singles

Results *un-*: n#n vs. n#V

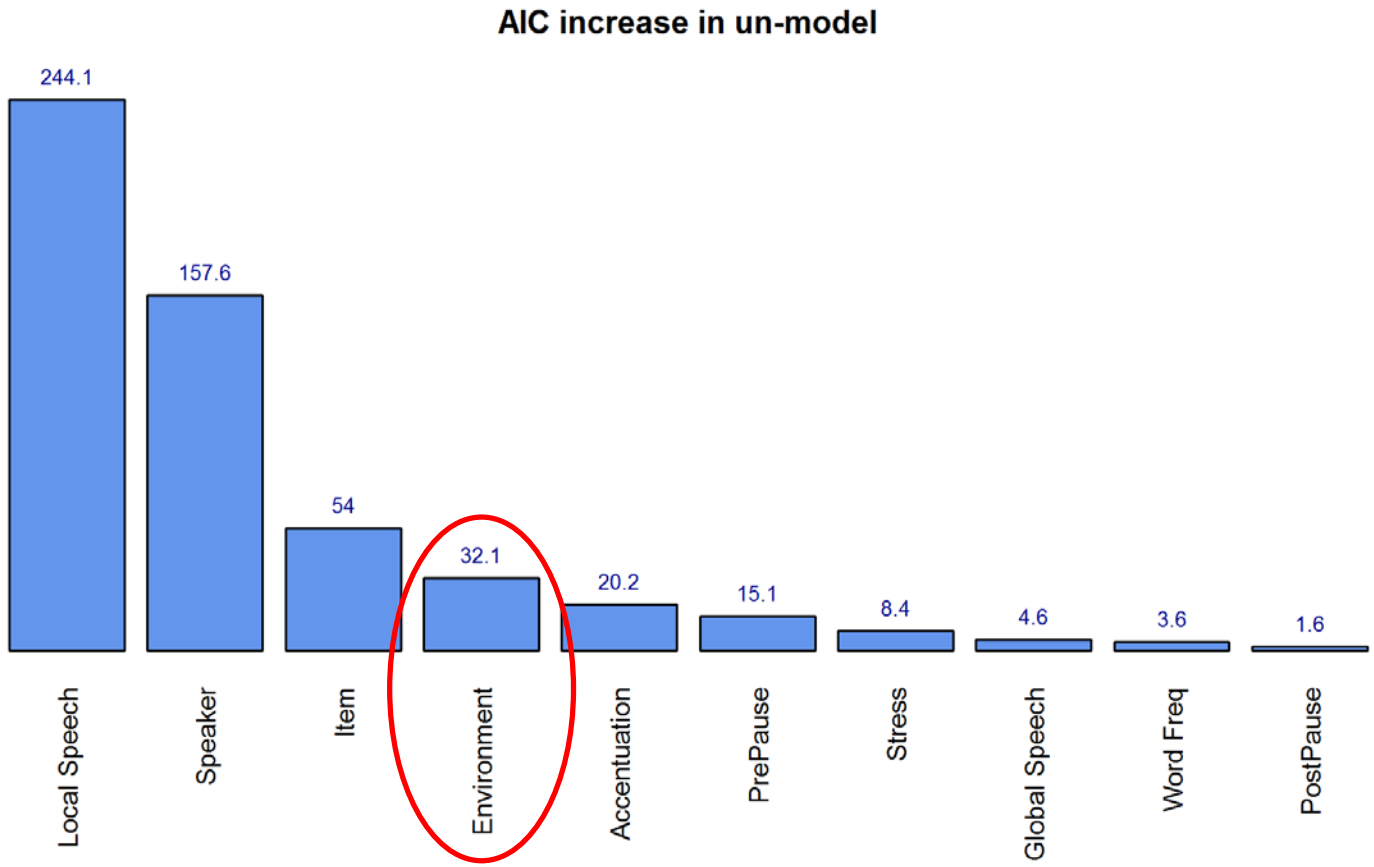


Results *un-*: n#n vs. #n



- $R^2 = 0.74$
- Covariates show expected effects
- Doubles are 36 ms longer than singles

Results *un-*: n#n vs. #n



Summary

- No effect of separability
- *un-* clearly geminates:
  - doubles are always longer than singles
- *in-* does not clearly geminate
  - For *im-*: doubles are never clearly longer than singles
  - For *in-*: only when base intial syllable is stressed, doubles are clearly longer than singles with a following vowel

un- and *in-* differ in their gemination pattern

Discussion

Does *in-* geminate? Does stress play an important role?

im-

- Experiment:
  - Stressed base intial syllable: doubles slightly longer than singletons
  - Unstressed base intial syllable: doubles as long as singletons

- Corpus: Doubles always longer than singletons

in-

- Stressed base intial syllable:
  - doubles longer than singletons with following vowel
  - doubles as long as singletons with following stop
- Unstressed base intial syllable:
  - doubles slightly longer than singletons with following vowel
  - doubles shorter than singletons with following stop

Discussion

Why do the experimental results deviate from the results of the corpus study?

Corpus	Experiment
• *un-* geminates (n#n= 90, n#V= 43)	• *un-* geminates (n#n=148/132, n#V= 51/54)
• *in-* geminates (m#m= 96, m#p= 69)	• *in-* does not clearly geminate (m#m= 99/ 76, m#p= 87/83)
• Difference in duration between negative and locative *in-*	• No difference in duration between negative and locative *in-*
• Natural conversational speech • American English • Less types	• Read speech • British English • More types

Discussion

Does separability play a role?

- No direct effect of separability on gemination
- *un-* more separable than *in-*
- Could the different gemination behavior of *un-* and *in-* be explained with a “categorical” difference in their separability?

Thank you very much for your attention!

Special thanks to…
- Melanie Bell, Anglia Ruskin University
- Phonetics Lab, Department of Theoretical and Applied Linguistics, University of Cambridge

Funding
Deutsche Forschungsgemeinschaft:
Forschergruppe 2372
- Grant PL151/8-1 ‘Morpho-phonetic Variation in English‘
- Grant PL151/7-1 ‘FOR 2737 Spoken Morphology: Central Project’

References

Ben Hedia, Sonia; Plag, Ingo (2016): Gemination and Degemination in English Affixation. Lexical Strata, Variability, and Phonetic Evidence. 1 Paper presented at “17th International Morphology Meeting”, 18.02.2016.

Boersma, P. & Weenink, D. (2014). Praat: doing phonetics by computer. Retrieved from http://www.praat.org/

Hay, Jennifer (2003): Causes and consequences of word structure. New York: Routledge (Outstanding dissertations in linguistics). Online verfügbar unter http://www.loc.gov/catdir/enhancements/fy0646/2004297538-d.html.

Kaye, Alan S. (2005): Gemination in English. In: *English Today* 21 (2), S. 43-55.

Kiparsky, Paul (1982): Lexical morphology and phonology. In: Linguistics in the morning calm. Selected papers from SICOL-1981. Unter Mitarbeit von The linguistic society of Korea. Seoul, Korea: Hanshin Pub. Co.

Mohanan, K. P. (1986). The theory of lexical phonology. Studies in natural language and linguistic theory: [v. 6]. Dordrecht, Boston, Norwell, MA: D. Reidel Pub. Co.; Sold and distributed in the U.S.A. and Canada by Kluwer Academic.

Oh, Grace E.; Redford, Melissa A. (2012): The production and phonetic representation of fake geminates in English. In: *Journal of Phonetics 40* (1), S. 82–91.

Plag, Ingo (2016) The phonetics of new words. Paper presented at the Workshop 'Expanding the lexicon', Universität Trier, November 17-18, 2016

R Development Core Team. (2014). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org

Smith, Rachel; Baker, Rachel; Hawkins, Sarah (2012): Phonetic detail that distinguishes prefixed from pseudo-prefixed words. In: Journal of Phonetics 40 (5), S. 689–705. DOI: 10.1016/j.wocn.2012.04.002.

Discussion

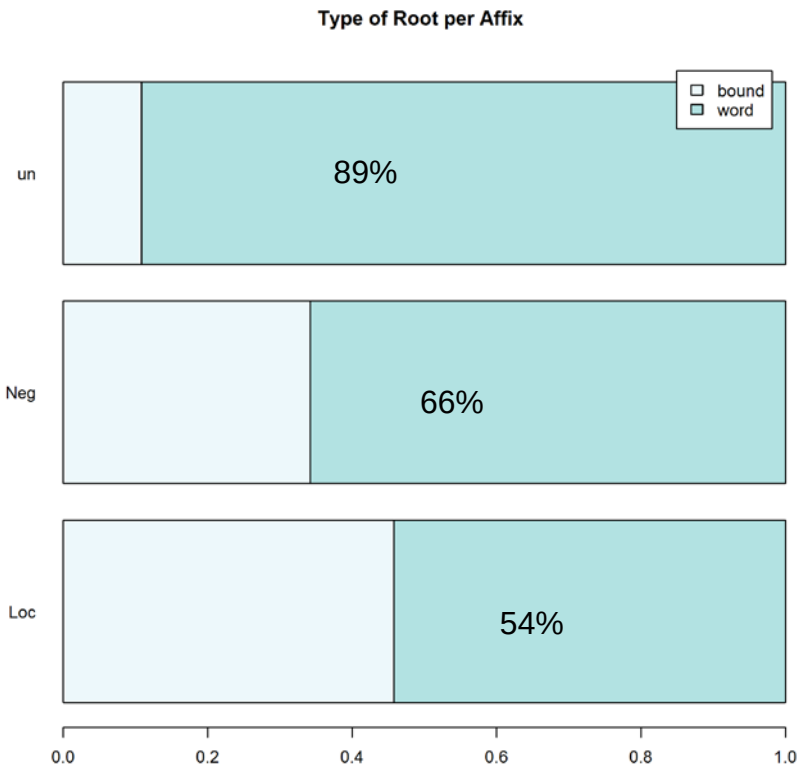
Does *in-* geminate?

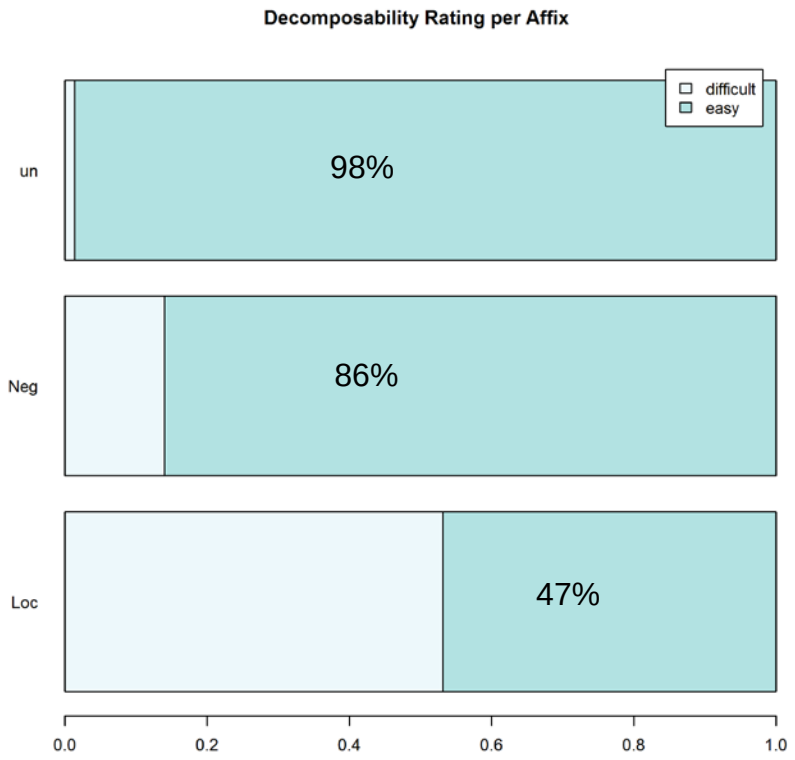
	im **m#m vs m#p**	**im** **m#m vs.#m**	**in** **n#n vs. n#t**	**in** **n#n vs. n#V**
Experiment	unstressed base: double = single	Pause before word: double = single	unstressed base: double < single 53 ms	unstressed base: double > single 8 ms
	stressed base: double > single 11 ms	No pause before word: double > single 10 ms	stressed base: double = single	stressed base: double > single 27 ms
Corpus	double > single 27 ms			

Data Overview: types

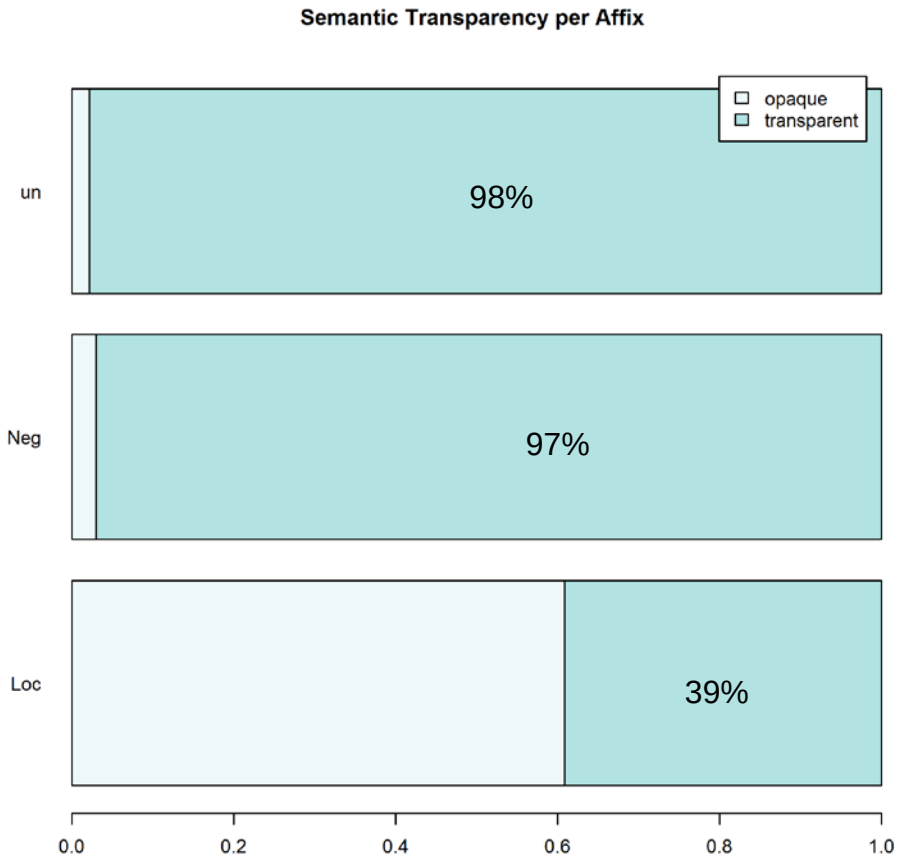
prefix	Morphological geminate (n#n)	Base (Base)	Singletons (n#V)	
un	20	20	26	
im	19	25	25	
in	4	3	19	27

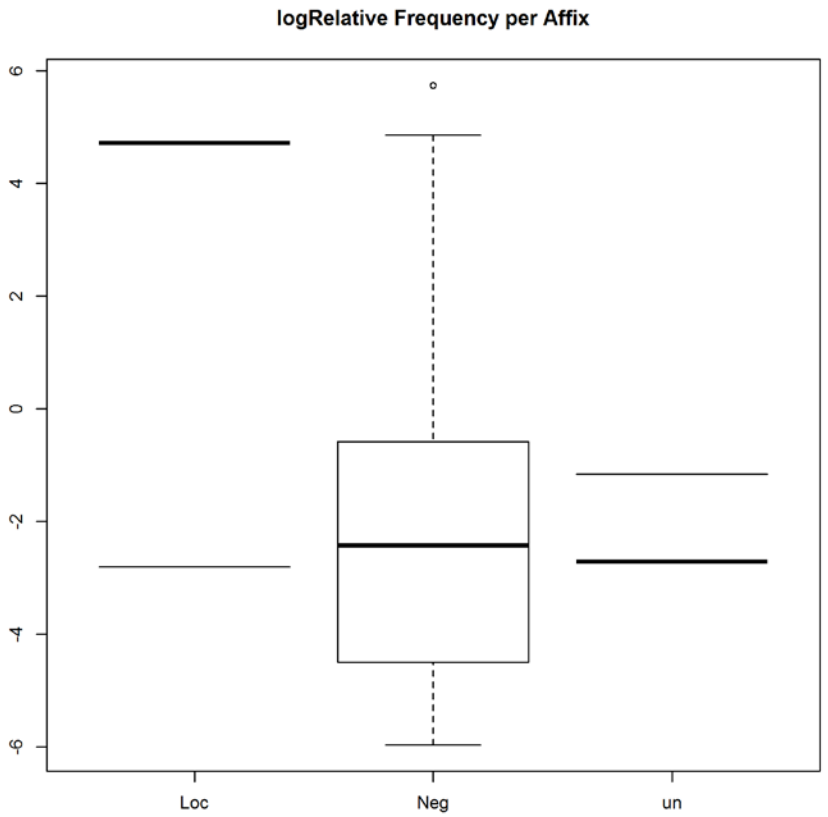
Decomposability of affixes





Decomposability of affixes



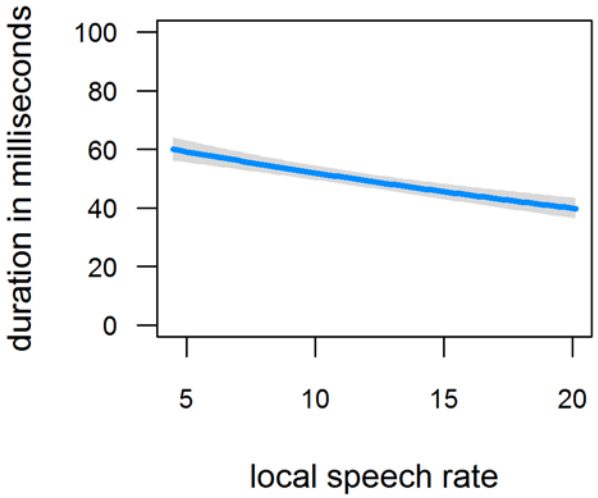


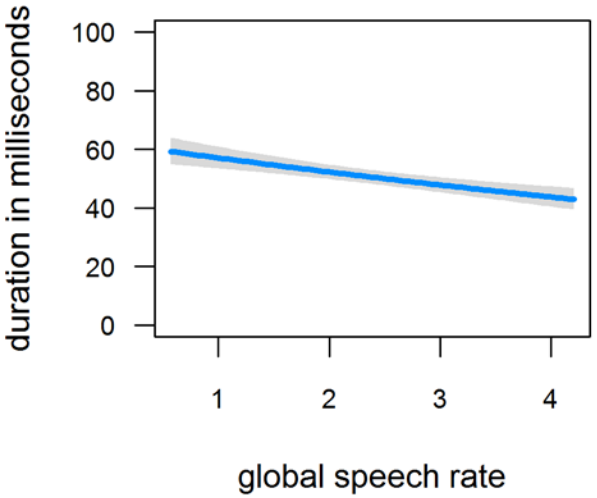
un-model: *unV* vs. *unn*

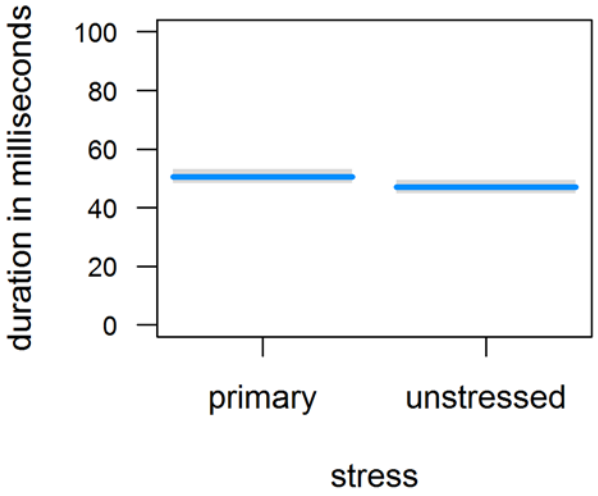
Fixed effects:

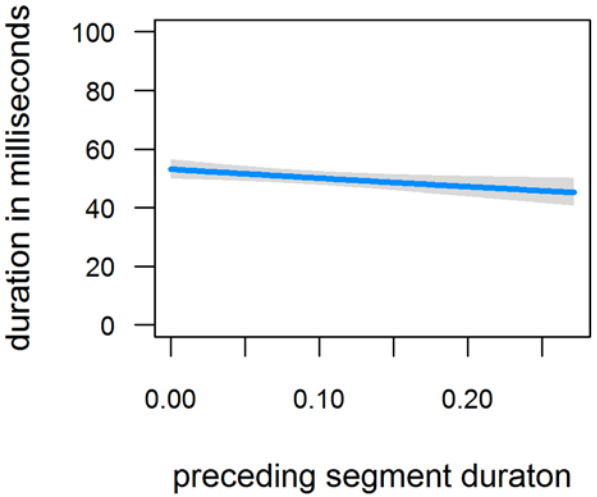
	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	9.154e-01	3.055e-03	4.524e+02	299.662	< 2e-16	***
FirstSyllBaseStressunstressed	-3.628e-03	1.202e-03	4.580e+01	-3.018	0.00414	**
LocSpeech	-1.326e-03	1.992e-04	5.746e+02	-6.655	6.62e-11	***
GlobalSpeechRate	-4.481e-03	9.369e-04	8.215e+02	-4.782	2.06e-06	***
PrecSegDur	-2.997e-02	1.289e-02	1.150e+03	-2.325	0.02024	*
TypeOfRootword	3.191e-03	1.612e-03	3.880e+01	1.981	0.05476	.
CategoryunV	-5.645e-02	1.278e-03	5.800e+01	-44.164	< 2e-16	***
AccentuationUnaccented	-6.328e-03	1.049e-03	1.144e+03	-6.033	2.17e-09	***
CategoryunV:AccentuationUnaccented	9.942e-03	1.160e-03	1.119e+03	8.571	< 2e-16	***

Results 1: un#n vs. un#V

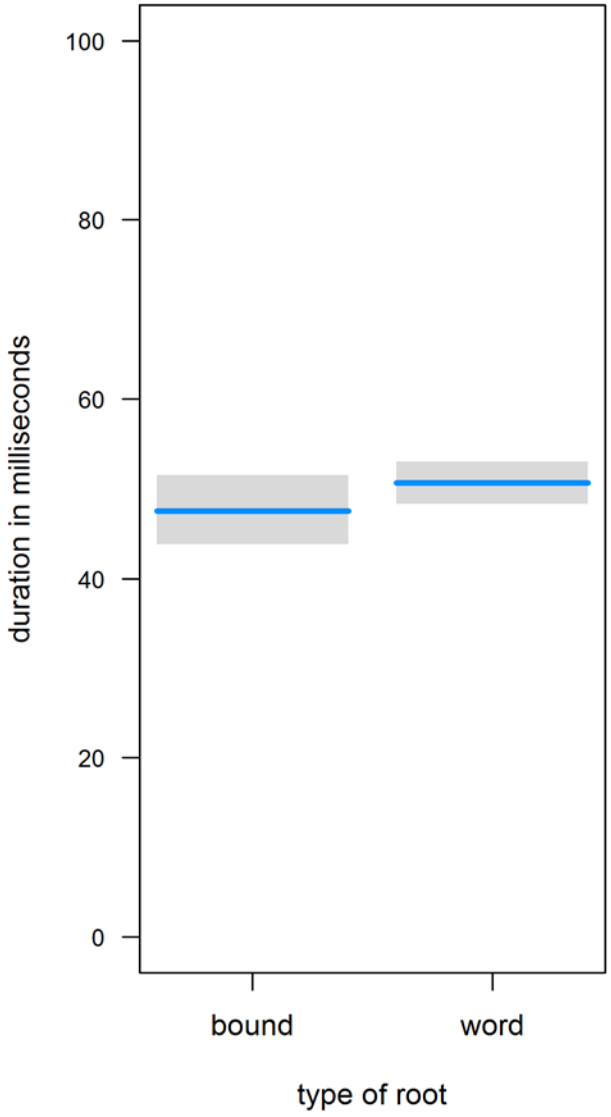








Results 1: un#n vs. un#V

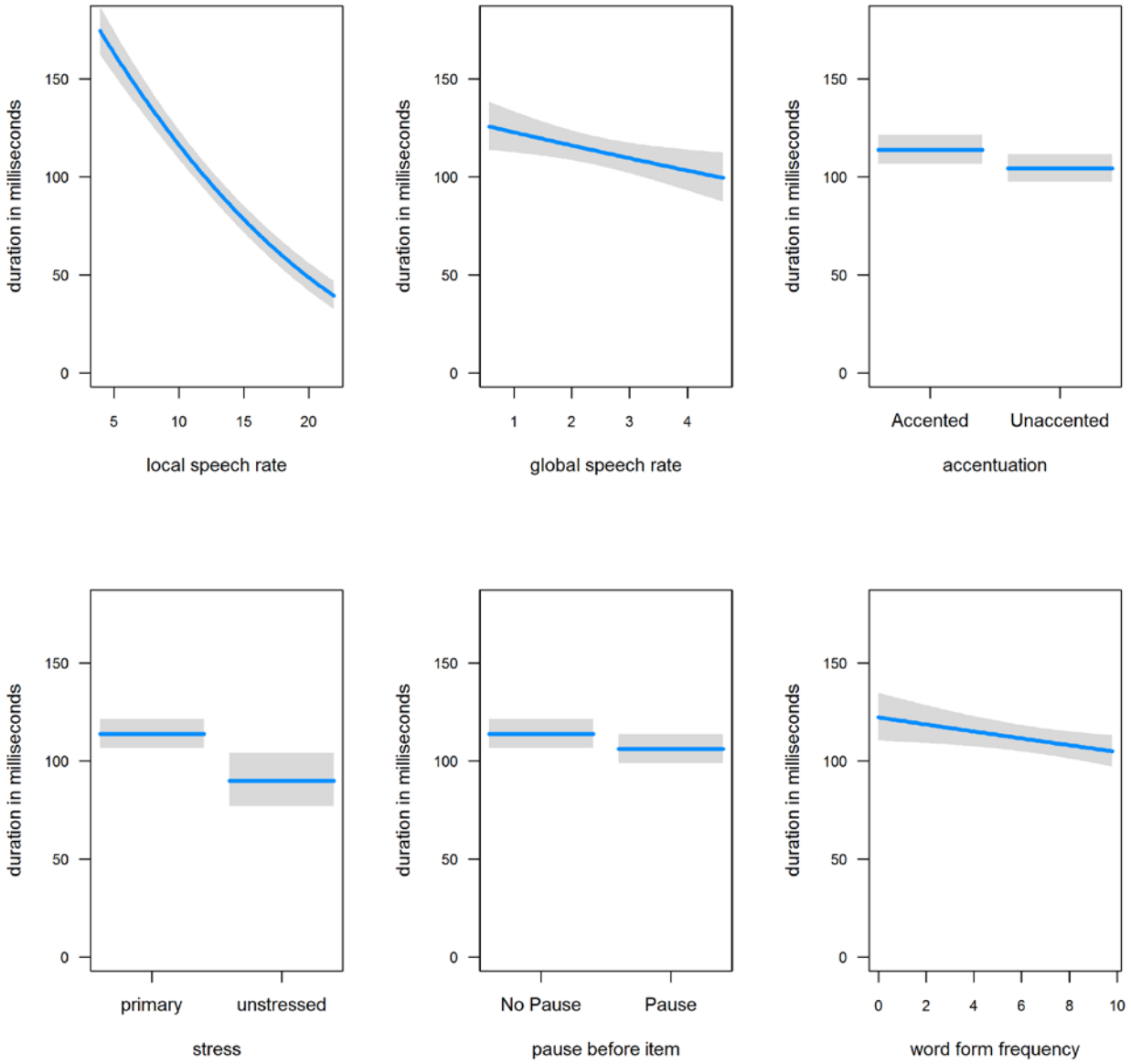


un-model: Base vs. unn

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	5.601e-01	1.306e-02	1.814e+02	42.874	< 2e-16	***
Categoryunn	4.953e-02	6.868e-03	3.170e+01	7.211	3.64e-08	***
AccentuationUnaccented	-1.445e-02	3.072e-03	1.016e+03	-4.706	2.88e-06	***
logWordFormFreq	-2.615e-03	1.117e-03	3.060e+01	-2.341	0.02591	*
FirstSyllBaseStressunstressed	-3.798e-02	1.141e-02	3.120e+01	-3.327	0.00226	**
LocSpeech	-1.237e-02	7.235e-04	7.480e+02	-17.099	< 2e-16	***
GlobalSpeechRate	-9.715e-03	3.833e-03	9.771e+02	-2.535	0.01141	*
PrePausePause	-1.160e-02	2.798e-03	1.029e+03	-4.145	3.68e-05	***
PostPausePause	-6.070e-03	3.111e-03	1.029e+03	-1.951	0.05130	.

un-model: Base vs. unn

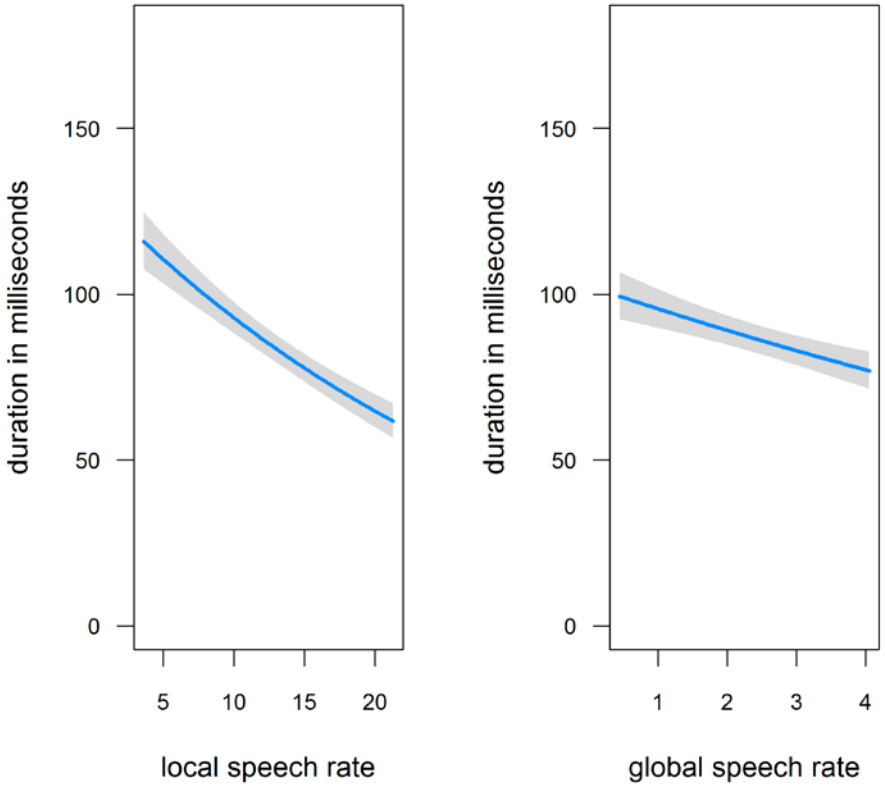


im-model: m#m vs. m#p

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	7.792e-01	4.280e-03	2.870e+02	182.045	< 2e-16	***
Categorymp	-1.251e-02	2.695e-03	3.780e+01	-4.644	4.06e-05	***
FirstSyllBaseStressunstressed	-2.563e-02	3.915e-03	3.690e+01	-6.548	1.15e-07	***
LocSpeech	-3.561e-03	3.576e-04	6.775e+02	-9.956	< 2e-16	***
GlobalSpeechRate	-7.102e-03	1.507e-03	9.530e+02	-4.713	2.81e-06	***
Categorymp:FirstSyllBaseStressunstressed	2.090e-02	5.145e-03	3.690e+01	4.063	0.000243	***

im-model: m#m vs. m#p

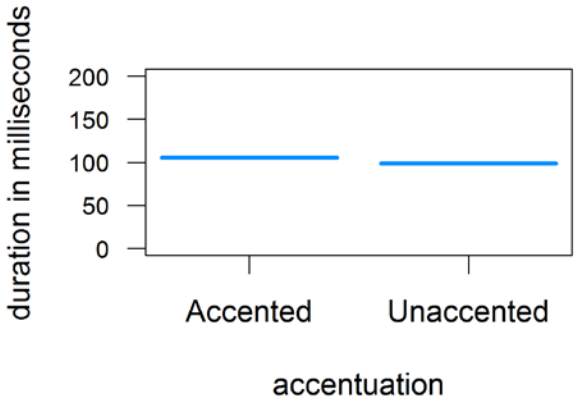


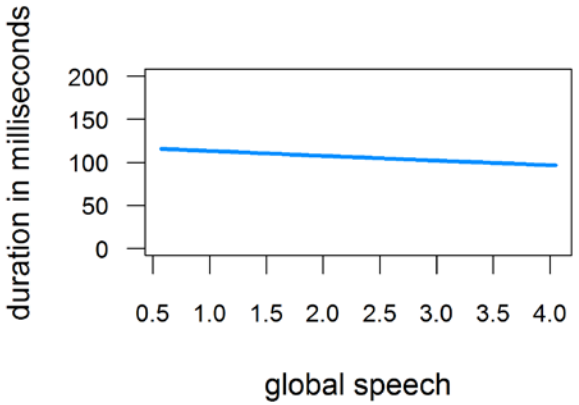
im-model: m#m vs. base

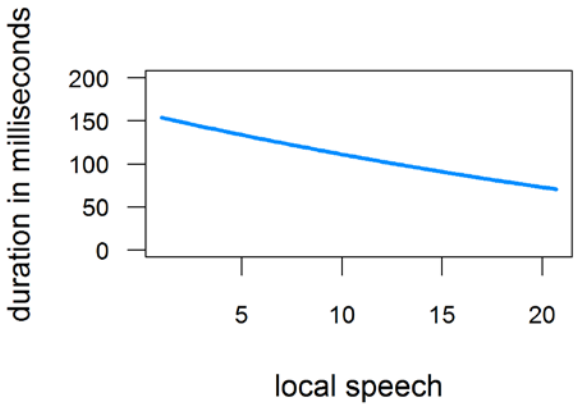
Fixed effects:

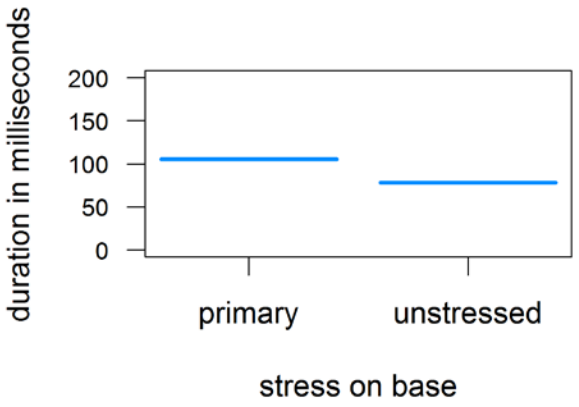
	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	4.884e-01	1.014e-02	4.097e+02	48.141	< 2e-16	***
Environmentm#m	-1.001e-02	5.021e-03	4.520e+01	-1.994	0.05223	.
PrePausePause	-2.783e-02	4.145e-03	9.004e+02	-6.713	3.38e-11	***
AccentuationUnaccented	-1.052e-02	3.207e-03	8.888e+02	-3.279	0.00108	**
FirstSyllBaseStressunstressed	-4.624e-02	5.390e-03	2.900e+01	-8.578	1.88e-09	***
LocSpeech	-6.439e-03	6.833e-04	4.939e+02	-9.423	< 2e-16	***
GlobalSpeechRate	-8.561e-03	3.781e-03	8.283e+02	-2.264	0.02381	*
Environmentm#m:PrePausePause	2.590e-02	4.903e-03	8.799e+02	5.283	1.61e-07	***

im-model: m#m vs. base







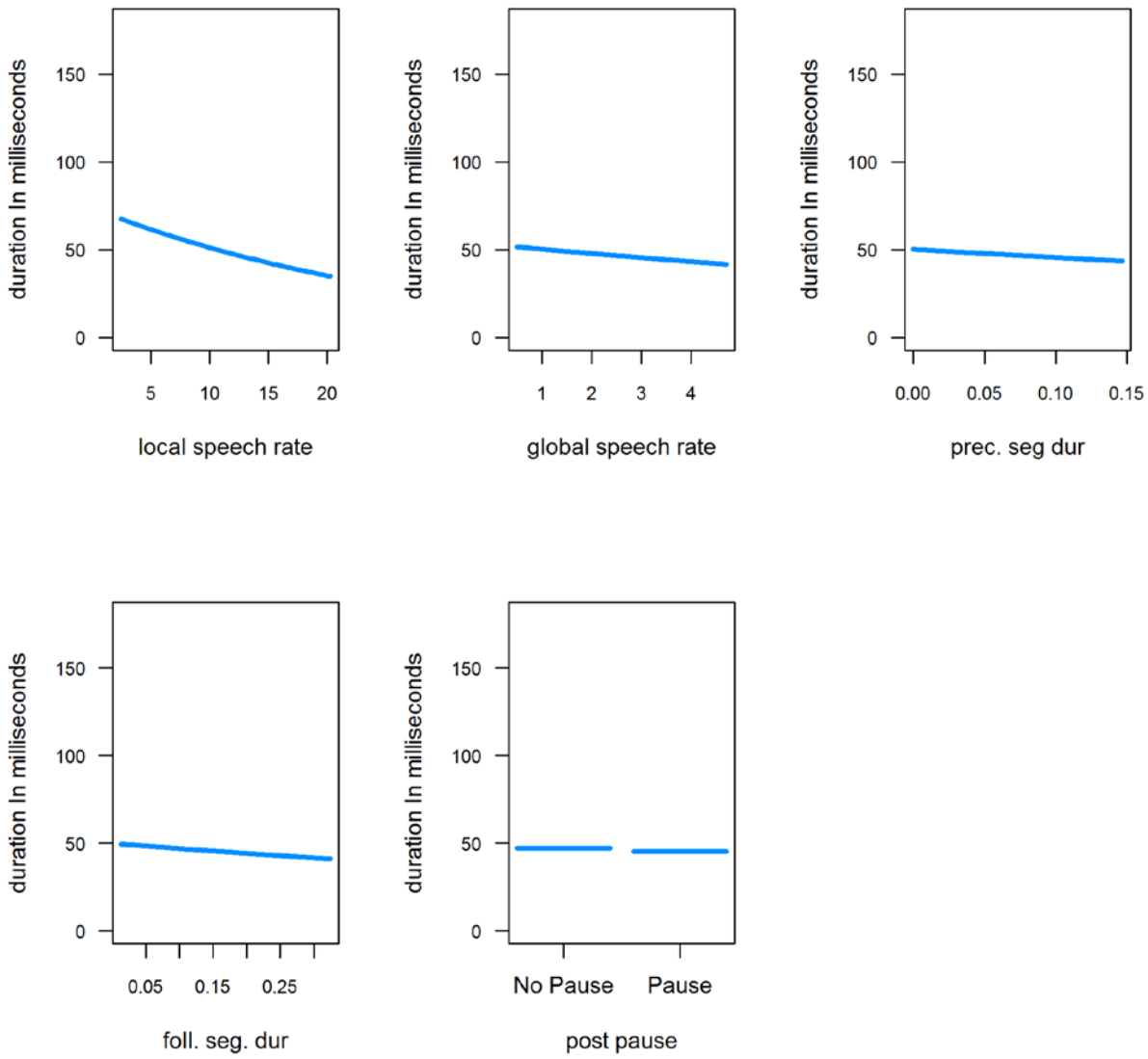


in-model: n#n vs. n#V vs. n#t

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	8.951e-01	4.690e-03	3.342e+02	190.844	< 2e-16	***
Categoryn#t	3.165e-03	2.827e-03	4.270e+01	1.120	0.269147	
Categoryn#V	-2.085e-02	3.047e-03	4.270e+01	-6.841	2.26e-08	***
FirstSyllBaseStressunstressed	-2.068e-02	5.220e-03	4.710e+01	-3.961	0.000252	***
LocSpeech	-1.872e-03	2.136e-04	8.757e+02	-8.763	< 2e-16	***
GlobalSpeechRate	-2.559e-03	9.416e-04	1.020e+03	-2.718	0.006683	**
PostPausePause	-2.138e-03	8.926e-04	1.083e+03	-2.395	0.016799	*
PrecSegDur	-4.885e-02	1.743e-02	1.079e+03	-2.803	0.005157	**
FollSegDur	-3.029e-02	1.175e-02	1.075e+03	-2.579	0.010047	*
Categoryn#t:FirstSyllBaseStressunstressed	2.637e-02	5.668e-03	4.660e+01	4.652	2.73e-05	***
Categoryn#V:FirstSyllBaseStressunstressed	1.194e-02	5.599e-03	4.740e+01	2.133	0.038118	*

in-model: n#n vs. n#V vs. n#t



Oh and Redford: difference between *un-* and *in-*

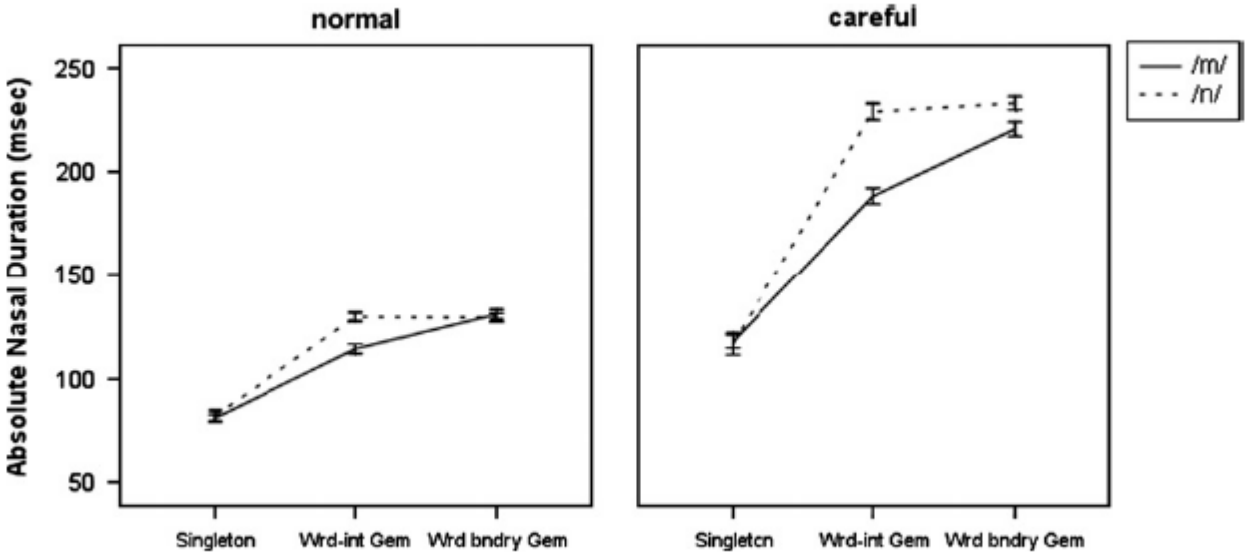
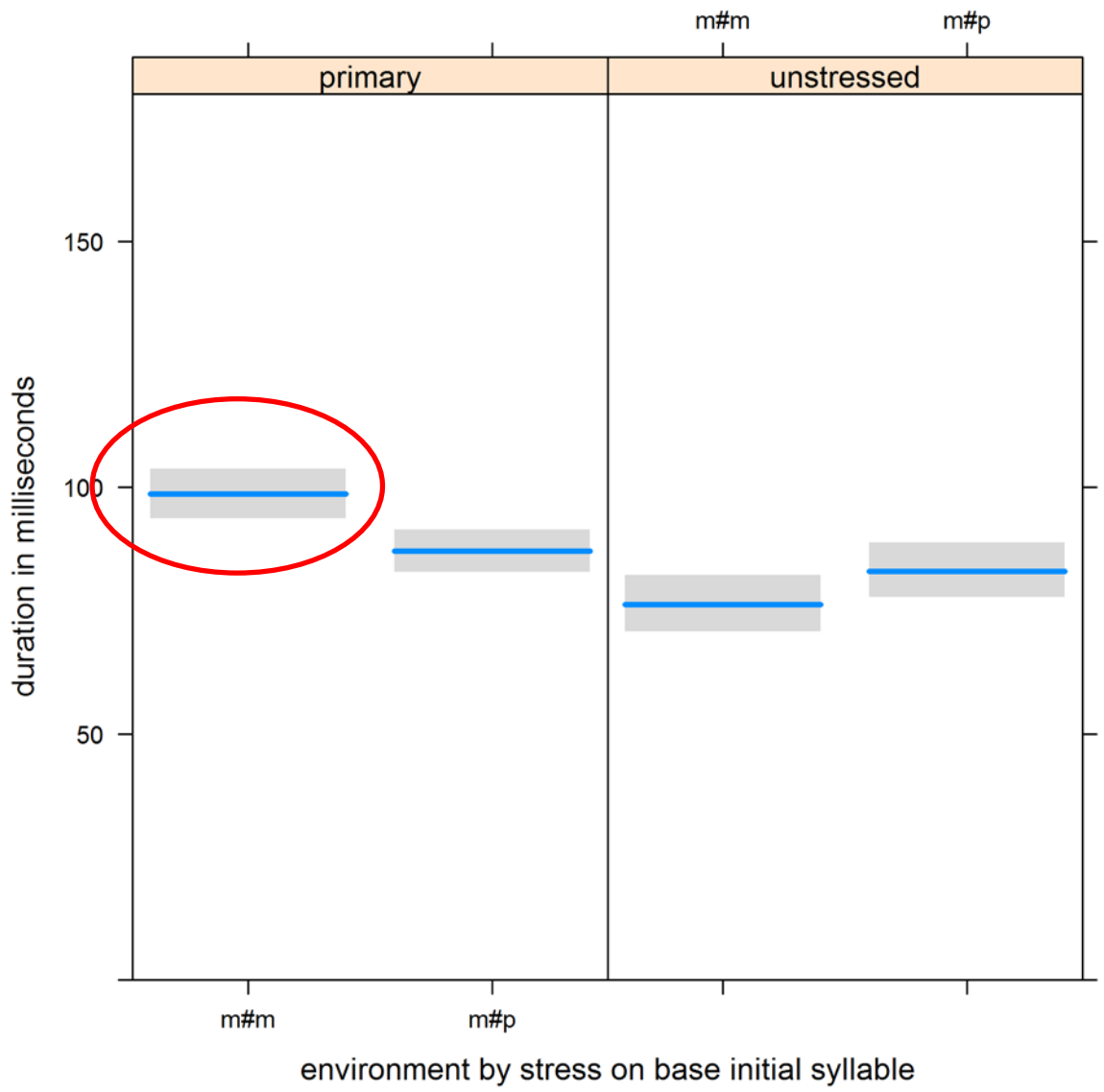
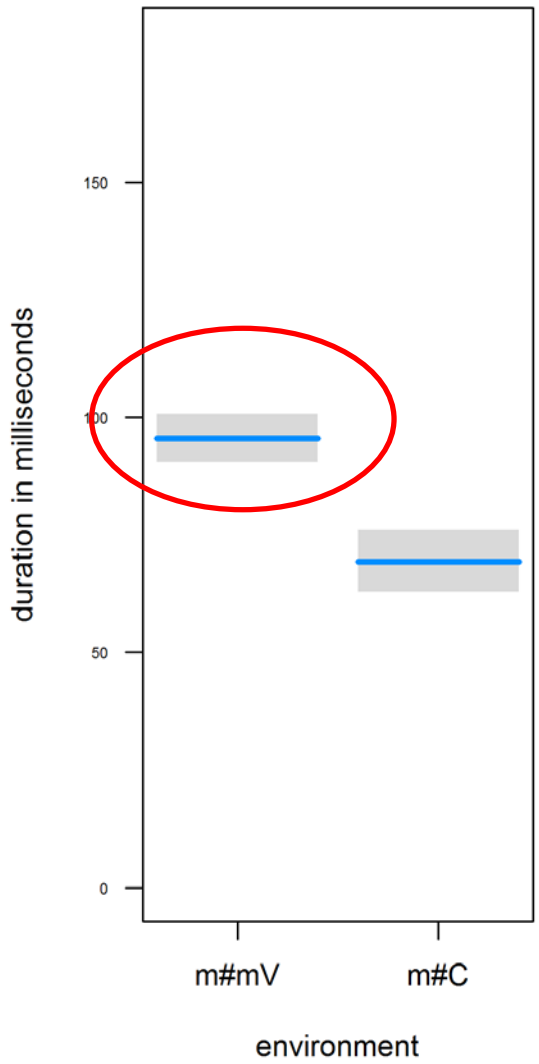


Fig. 2. Absolute nasal duration for word-internal singletons, word-internal geminates and word boundary geminates produced in normal and careful speech.

un- and *in- :* Corpus vs. Experiment





un- and *in-* : Corpus vs. Experiment