Cui bono?
English morphosyntactic variation as a compensatory strategy

Gero Kunter
Heinrich-Heine-Universität Düsseldorf
gero.kunter@uni-duesseldorf.de

Research supported by
*Deutsche Forschungsgemeinschaft (*grant KU 2896/1-1)

Johannes Gutenberg-Universität Mainz
December 8, 2014

English morphosyntactic variation:

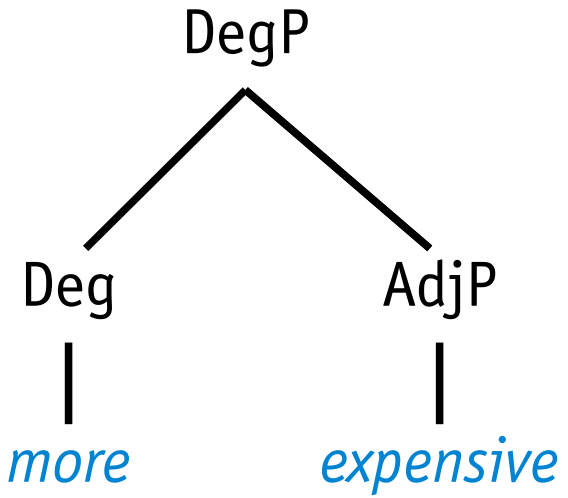
Comparatives

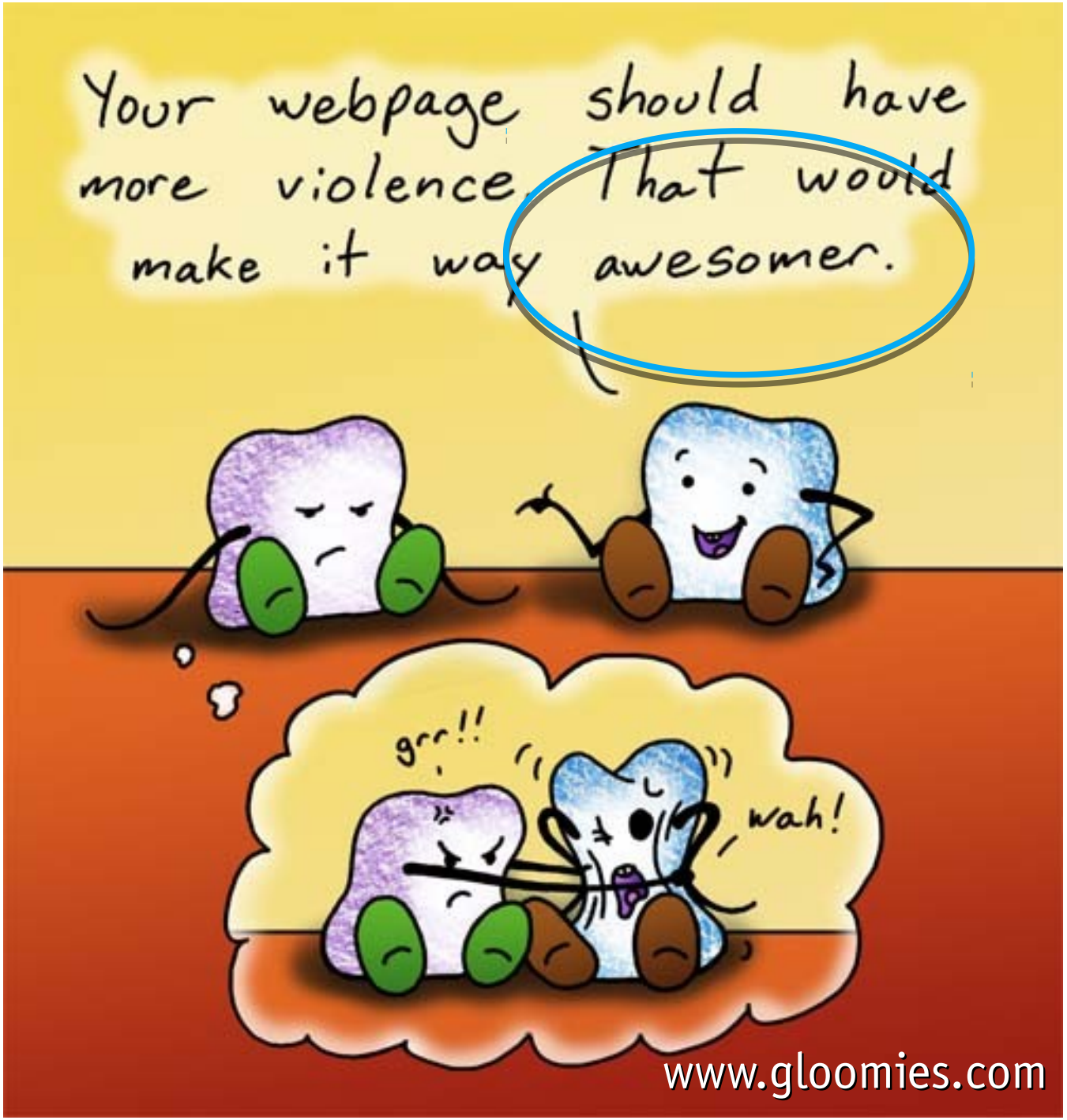
English adjective paradigms

Synthetic
(word)

{high} +	{er} →	*higher*
[Adj]	[Deg]	[Adj, Deg]

Analytic
(phrase)




Your webpage should have more violence. That would make it way awesomer.
grr!!
wah!
www.gloomies.com

They say happier times are even more happy because you've had some hard times.

(COCA SPOK_2010: NBC_Dateline)

So Stephon Marbury though is back with the Knicks, $180,000 more poor - poorer - more poor. I don't know.

(COCA SPOK_2007: NPR_Park)

Some factors determining comparative alternation

(e.g. Hilpert 2008, Mondorf 2003, 2009)

Phonological	Number of syllables, final elements of base, stress pattern
Morphological	Number of morphemes, compound adjectives
Lexical	Frequency of adjective, comparative/positive ratio
Syntactic	*to*-infinitive complementation, following *than*, premodification, predicative vs. attributive position
Semantic	Abstract vs. concrete meaning
Pragmatics	End-weight

More-support

in cognitively more demanding environments which require an increased processing load, language users [...] *tend to compensate for the additional effort by resorting to the analytic form*

(Mondorf 2009: 6)

the more-*variant makes lower demands on processing energy and hence is preferred in more complex environments.*

(Mondorf 2009: 196)

Advantages of analytic comparatives
(cf. Mondorf 2009: 7)

- *more* is unambiguous marker of degree phrase
- higher explicitness (each word has one function)
- early signal of upcoming complex structure

***More*-support apparently used by speakers for the benefit of the listener.**

Some issues

If analytic comparatives are easier to process...

... why do synthetic comparatives exist anyway?

... why is there alternation for some comparatives?

[*When*] *processing demands imposed by the linguistic environment are low, it is more economical to invoke the less explicit and shorter* -er *variant. Though this variant entails a higher processing effort in terms of dependency relations, it is easily affordable in simple environments.*

(Mondorf 2009: 196)

Some central concepts

Linguistic processing
Mapping of linguistic inputs onto internal cognitive structures, or mapping of internal cognitive structures onto linguistic output

Processing complexity
Degree of difficulty of linguistic processing

External correlate of processing complexity
Time required to process linguistic input

More-support from a psycholinguistic perspective

Hypothesis
More-support acts as a compensatory strategy for increased processing complexity.

Underlying assumption
Analytic comparatives are easier to process.

Prediction
If type- or context-dependent variables increase processing time, probability of analytic comparatives will increase.

Aims of this talk

- Which psycholinguistic evidence exists for the assumption that analytic comparatives are easier to process?
- Is this assumption valid for production as well as perception?
- Who really benefits from choosing analytic comparatives in cognitively more demanding environments?

Empirical evidence

Perception experiment (lexical decision task)

- Comparison of processing times for synthetic and analytic comparatives

Corpus study

- Effect of processing times on probability to encounter analytic comparatives in COCA

Production experiment (sentence completion)

- Effect of processing times on probability to use analytic comparatives

Perceptual processing

Indirect evidence for *more*-support in perception

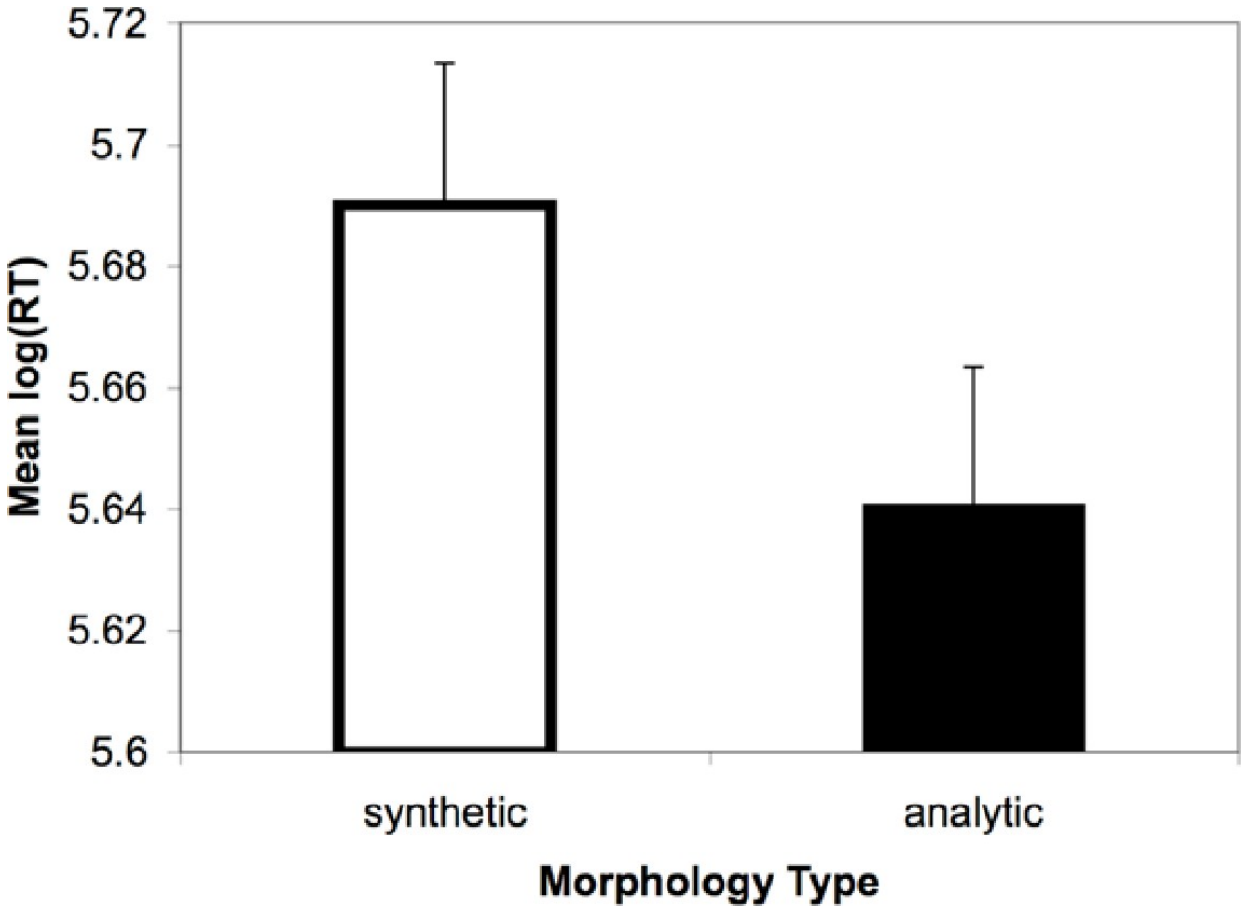
Self-paced reading experiment (Boyd 2007):

(a) *Kenny G is more mellow* ***to*** *listen to than Yanni.*
(b) *Kenny G is mellower* ***to*** *listen to than Yanni.*

Dependent variable: processing time of *to*

Results: Self-paced reading

(Boyd 2007:31)



Some factors determining comparative alternation
(e.g. Hilpert 2008, Mondorf 2003, 2009)

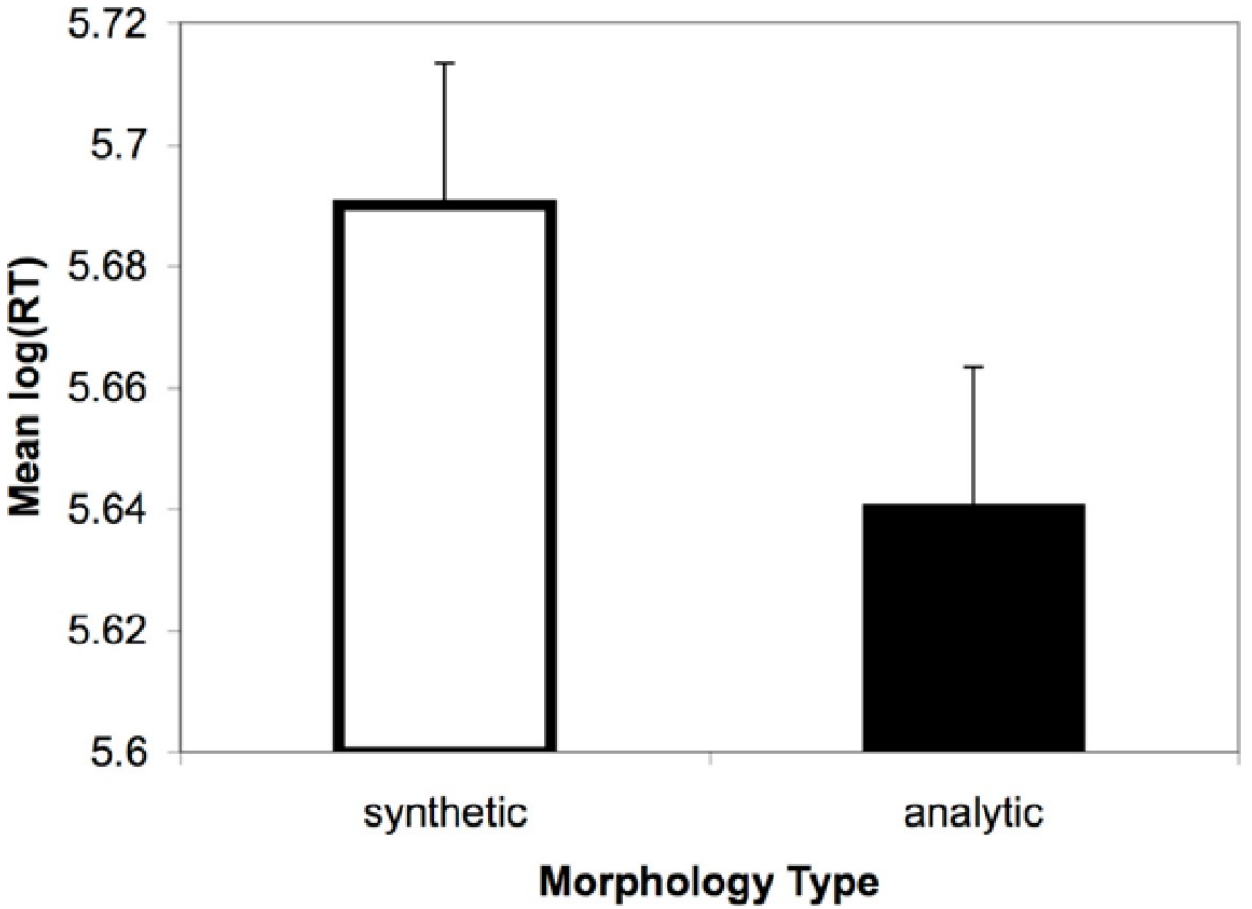
Phonological	Number of syllables, final elements of base, stress pattern
Morphological	Number of morphemes, compound adjectives
Lexical	Frequency of adjective, comparative/positive ratio
Syntactic	*to*-infinitive complementation, following *than*, premodification, predicative vs. attributive position
Semantic	Abstract vs. concrete meaning
Pragmatics	End-weight

Indirect evidence for *more*-support in perception

Self-paced reading experiment (Boyd 2007):

(a) *Kenny G is more mellow* ***to*** *listen to than Yanni.*
(b) *Kenny G is mellower* ***to*** *listen to than Yanni.*

Results: Self-paced reading
(Boyd 2007:31)



"spillover effect" from preceding synthetic comparative – or an effect of unexpected structure?

Experiment: Auditory lexical decision task

Hypothesis	Analytic comparatives show shorter reaction times than synthetic comparatives.
Adjectives	60 adjectives at least 5 analytic comparatives **and** at least 5 synthetic comparatives in Corpus of Contemporary American English (COCA, Davies 2008-)
Recordings	produced by male speaker of Canadian English
Participants	31 undergraduate students from University of Alberta, Edmonton

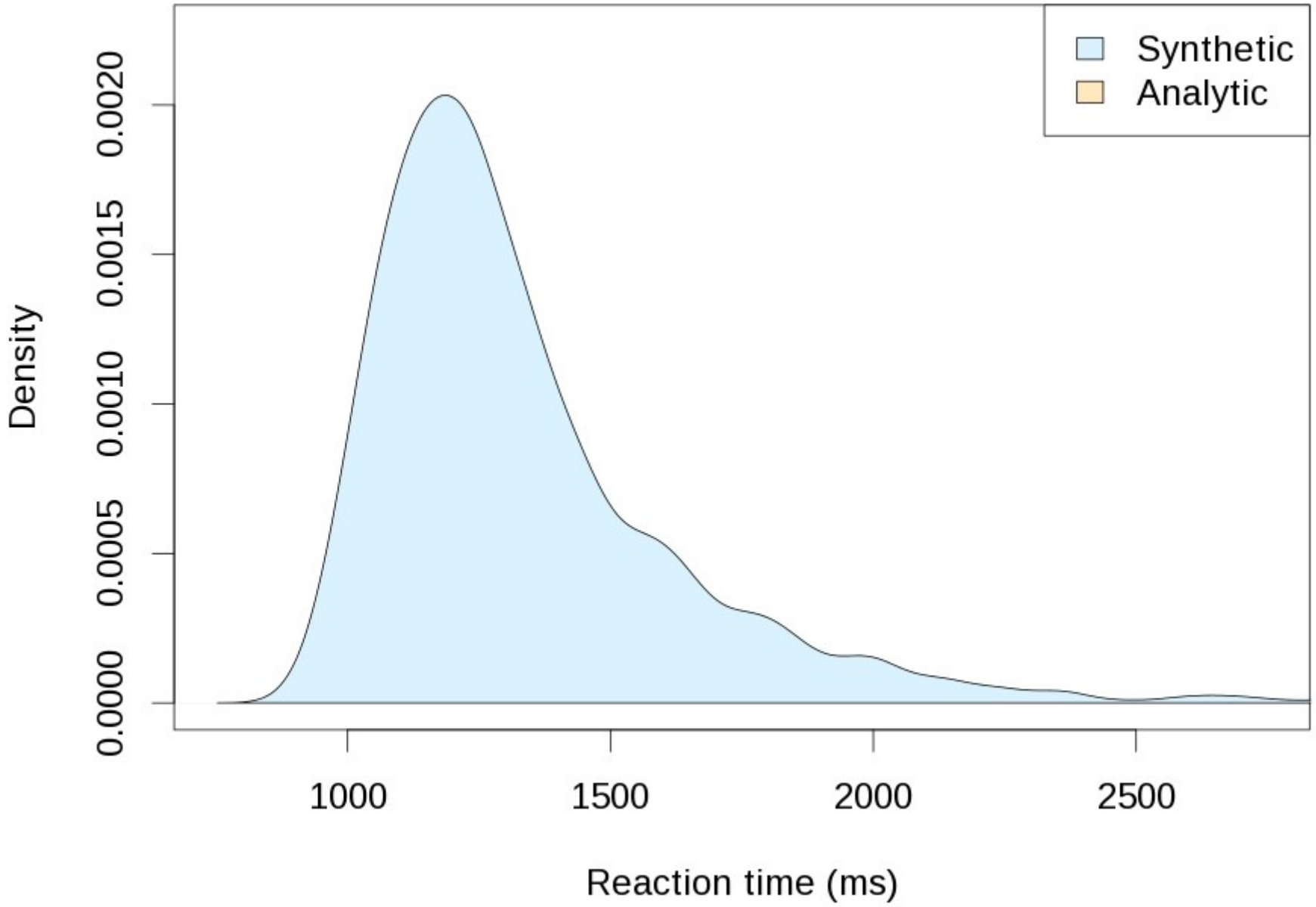
Experimental design

Stimuli	Synthetic	*colder*
	Analytic	*more cold*
Controls	Fake-synthetic	**coldic*
	Fake-analytic	*more* **gorsty*
Distractors	Existing phrases	*on wire*
	Non-existing words with *-er*	**rilker*
	Existing complex words	*chasting*

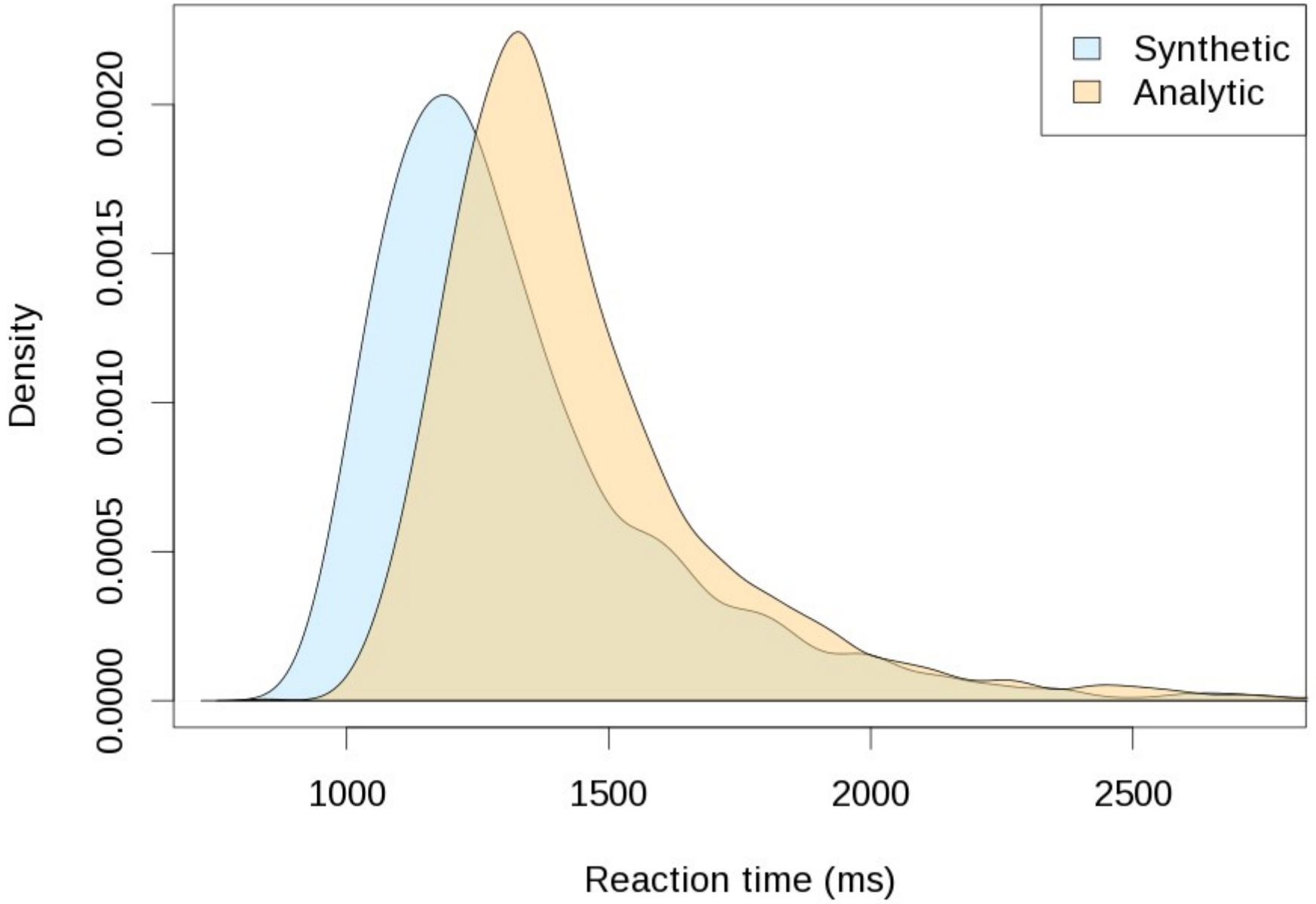
2 x 60 = 120 stimuli, 2 x 60 = 120 controls, 3 x 80 = 240 distractors → **500 trials per participant**

Perceptual processing
Analysis and results

Overall distribution of reaction times



Overall distribution of reaction times

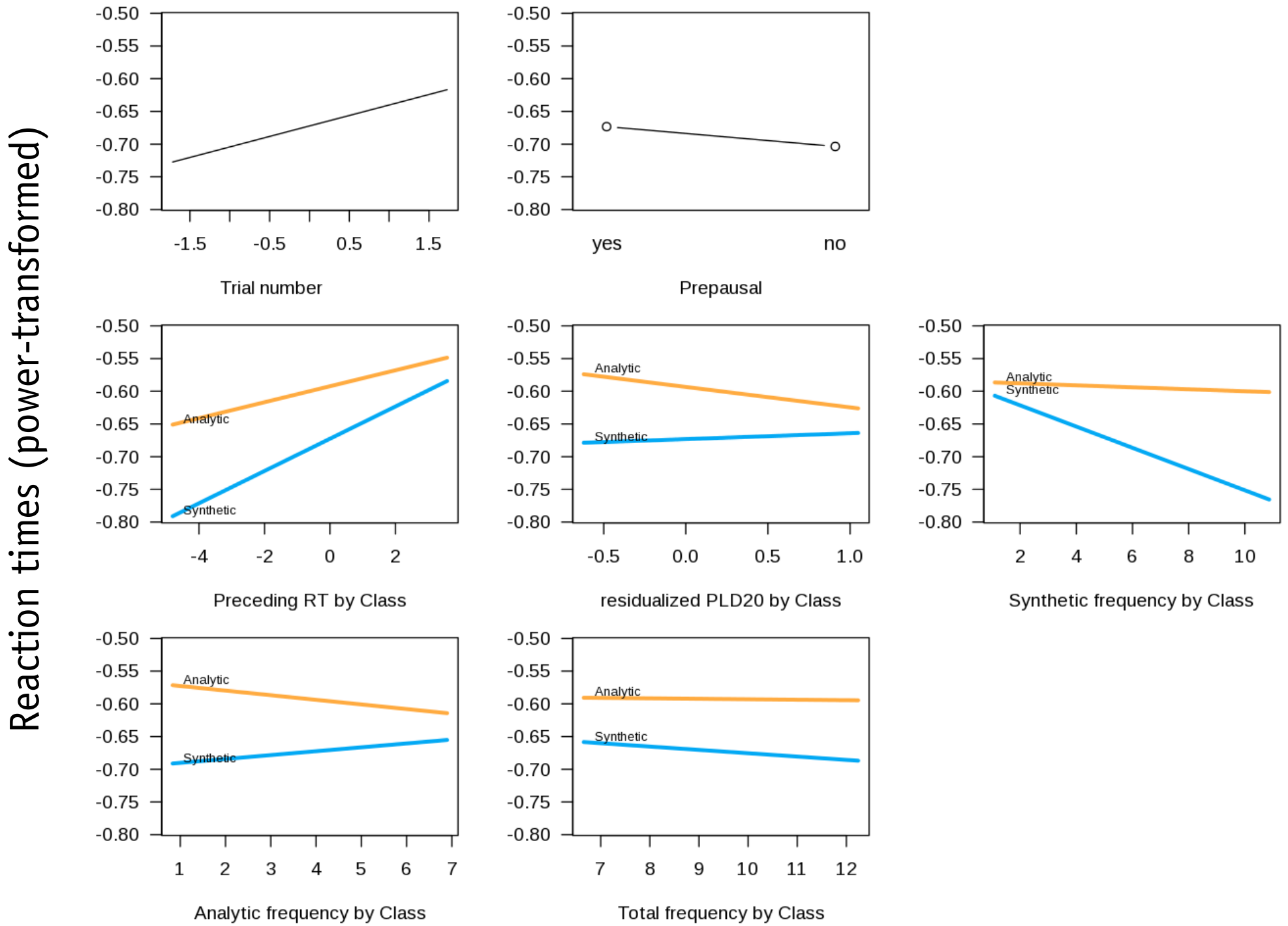


Statistical model

Analysis	Multivariate mixed-effects model
Dependent variable	Reaction time (power-transformed from milliseconds)
Random effects	Participant, Base adjective

Co-variates in mixed-effects model

Experimental	Experimental booth, Trial by Class, Trial by Prepause, Previous RT by Class
Subject	Handedness, Sex, Age
Phonological	Metrical structure of base by Class, Number of phonemes by Class
Lexical	Frequencies of base, analytic comparative, synthetic comparative all by Class (from COCA)
	Number of phonological neighbours, Mean RT, Phonological Levenshtein distance all by Class (from English Lexicon Project, Balota et al. 2007)
	Age of Acquisition by Class (from Kuperman et al. 2012)
	Inflectional Entropy by Class (cf. Moscoso del Prado Martín et al. 2004)


Reaction times (power-transformed)
-0.50
-0.55
-0.60
-0.65
-0.70
-0.75
-0.80
-1.5
-0.5
0.5
1.5
Trial number
yes
no
Prepausal
Analytic
Synthetic
-4
-2
0
2
Preceding RT by Class
-0.5
0.0
0.5
1.0
residualized PLD20 by Class
2
4
6
8
10
Synthetic frequency by Class
1
2
3
4
5
6
7
Analytic frequency by Class
7
8
9
10
11
12
Total frequency by Class

Perceptual processing of comparative variants

- Processing times for analytic comparatives are consistently longer than for synthetic comparatives
- Processing disadvantage of analytic comparatives

Listeners do not benefit from *more*-support: Synthetic comparatives are processed faster.

→ What about **speakers?**

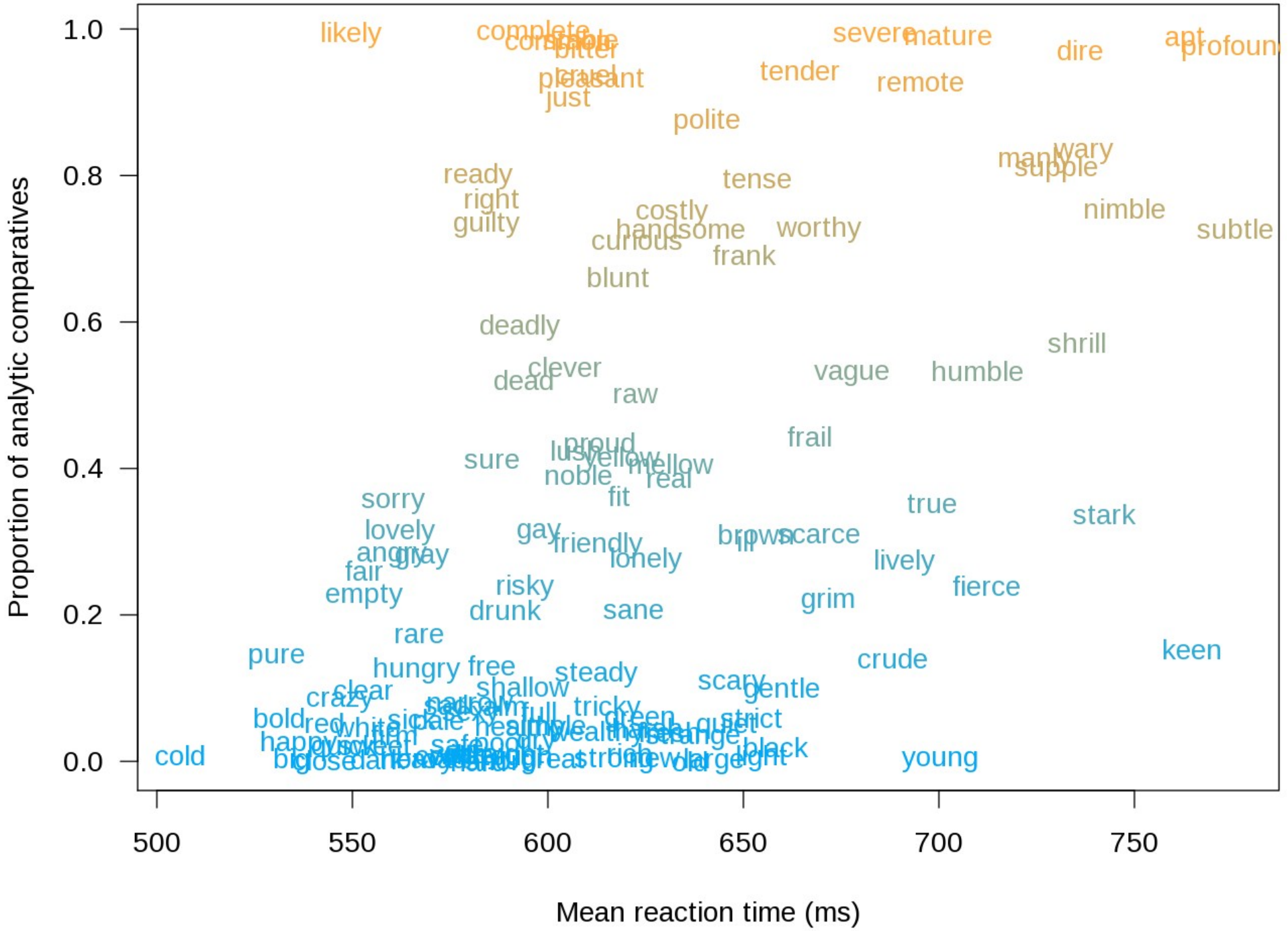
Corpus study

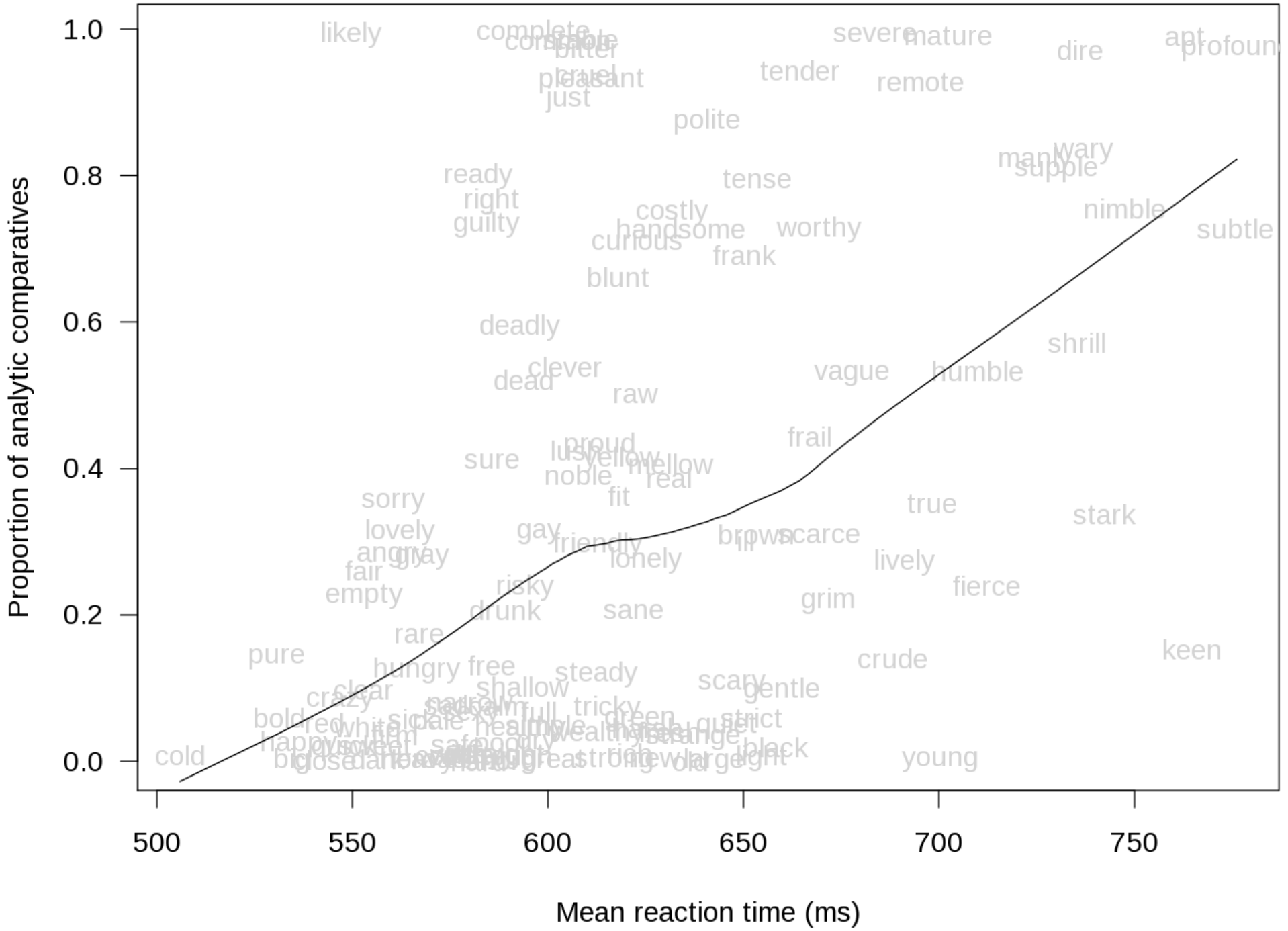
Corpus study

Hypothesis	Cognitively complex adjectives occur with a higher proportion of analytic comparatives.
Corpus	Contemporary Corpus of American English (COCA), 450 million written and spoken words, 1990-2012
Data	128 adjectives, at least 3 analytic comparatives **and** at least 3 synthetic comparatives in COCA
Complexity	Mean lexical decision times from English Lexicon Project (Balota et al. 2007)

Examples

Adjective	Base frequency	Synthetic comparatives	Analytic comparatives	Proportion of analytic
new	491932	3709	7	0.00
risky	5558	552	172	0.24
deadly	8411	128	186	0.59
nimble	767	27	83	0.75
remote	12076	31	398	0.93


Proportion of analytic comparatives
1.0
0.8
0.6
0.4
0.2
0.0
500
550
600
650
700
750
Mean reaction time (ms)
likely
complete
severe
mature
dire
apt
profoun
bitter
tender
pleasant
remote
just
polite
wary
manly
supple
ready
tense
right
costly
nimble
guilty
handsome
worthy
subtle
curious
frank
blunt
deadly
shrill
clever
vague
humble
dead
raw
frail
proud
sure
noble
real
fit
sorry
true
stark
lovely
gay
friendly
scarce
lonely
lively
fair
risky
fierce
empty
drunk
sane
grim
rare
keen
pure
crude
hungry
free
steady
scary
gentle
shallow
crazy
tricky
bold
green
strict
cold
young
black


Proportion of analytic comparatives
1.0
0.8
0.6
0.4
0.2
0.0
500
550
600
650
700
750
Mean reaction time (ms)
likely
complete
severe
mature
dire
apt
profoun
bitter
tender
remote
pleasant
just
polite
wary
manly
supple
ready
tense
right
costly
nimble
guilty
handsome
worthy
subtle
curious
frank
blunt
deadly
shrill
clever
vague
humble
dead
raw
frail
proud
sure
yellow
mellow
noble
real
fit
sorry
true
stark
lovely
gay
friendly
brown
scarce
lonely
lively
fair
empty
risky
fierce
drunk
sane
grim
rare
keen
pure
crude
hungry
free
steady
scary
shallow
gentle
crazy
tricky
bold
green
strict
cold
young
black

Statistical model

Analysis	Beta regression (proportion data)
Dependent variable	Proportion of analytic comparatives
Main predictor	Mean lexical decision times
Other co-variates	Final segment of base Metrical structure Number of phonemes Base frequency Positive-comparative ratio

Results: Beta regression

- Probability of analytic comparatives is higher for adjectives with increased cognitive complexity (i.e. higher RTs)
- Co-variates show expected effects
- Effect of cognitive complexity is independent of other co-variates

Cognitively complex adjectives occur with a higher proportion of analytic comparatives in a corpus.

Production experiment

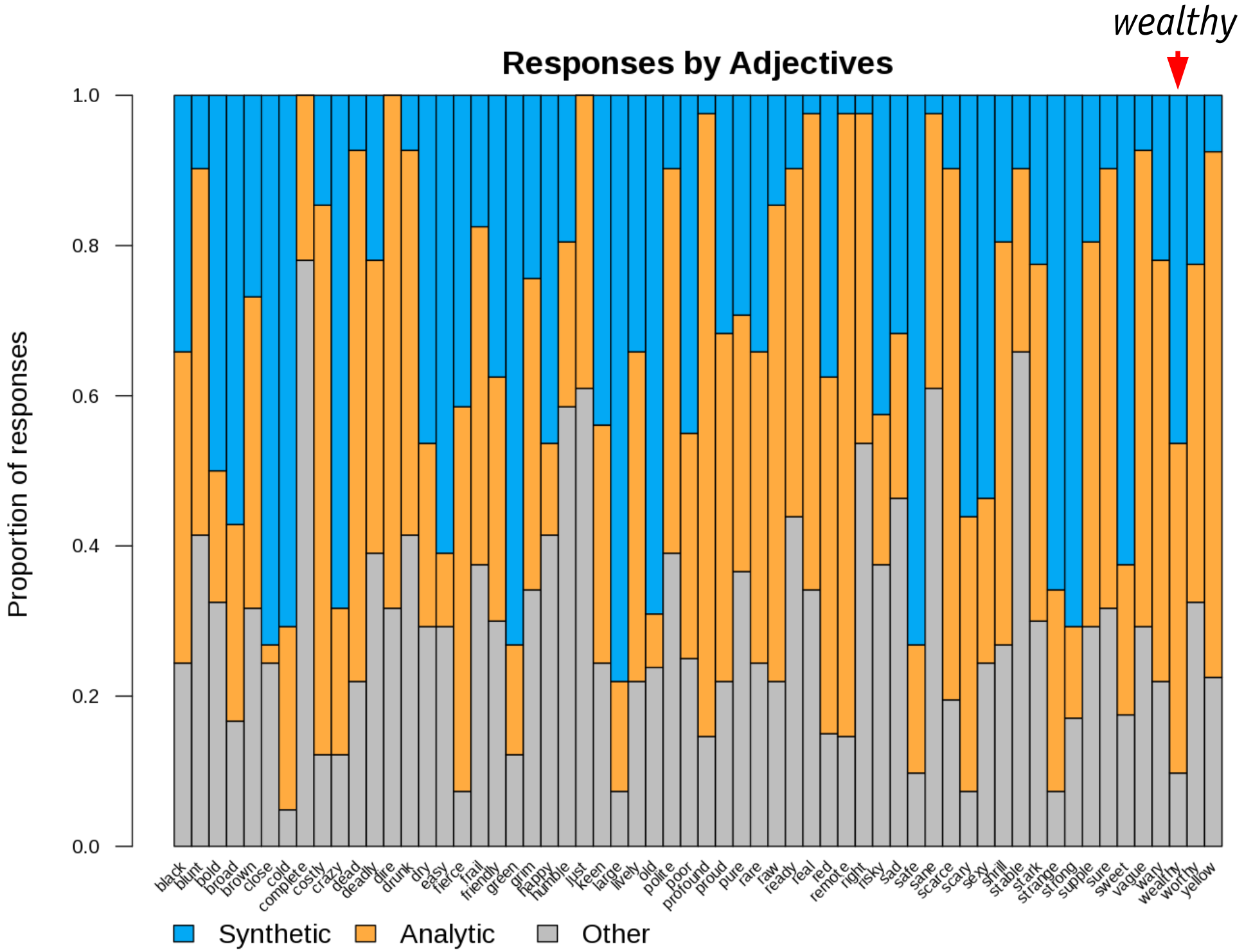
Production experiment

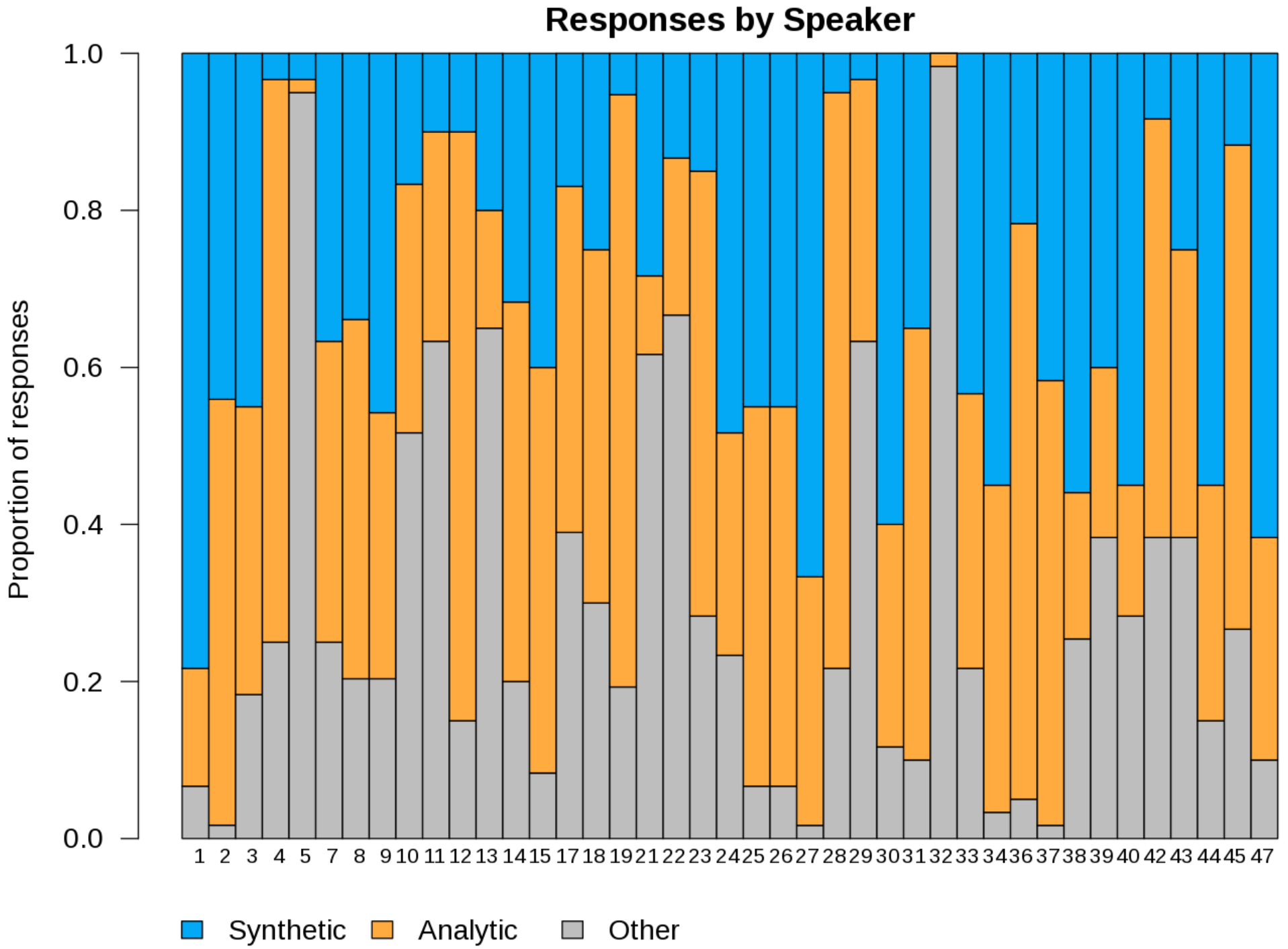
Hypothesis	Speakers are more likely to produce an analytic comparative if the adjective is cognitively complex.
Adjectives	60 adjectives (same as before) from COCA
Speakers	41 undergraduate students from University of Alberta, Edmonton
Task	Sentence completion
Complexity	Median reaction times from separate lexical decision task (same speakers, at least 14 days before production experiment)

Experimental Task

The duke is wealthy.
Yet, the king is _____.

WEALTHY


wealthy
Responses by Adjectives
Proportion of responses
1.0
0.8
0.6
0.4
0.2
0.0
black
blunt
bold
broad
brown
close
cold
complete
costly
crazy
dead
deadly
dire
drunk
dry
easy
fierce
frail
friendly
green
grim
happy
humble
just
keen
large
lively
old
polite
poor
profound
proud
pure
rare
raw
ready
real
red
remote
right
risky
sad
safe
sane
scarce
scary
sexy
shrill
stable
stark
strange
strong
supple
sure
sweet
vague
wary
wealthy
worthy
yellow
Synthetic
Analytic
Other


Responses by Speaker
Proportion of responses
1.0
0.8
0.6
0.4
0.2
0.0
1 2 3 4 5 7 8 9 10 11 12 13 14 15 17 18 19 21 22 23 24 25 26 27 28 29 30 31 32 33 34 36 37 38 39 40 42 43 44 45 47
Synthetic
Analytic
Other

Final data set

Removed:

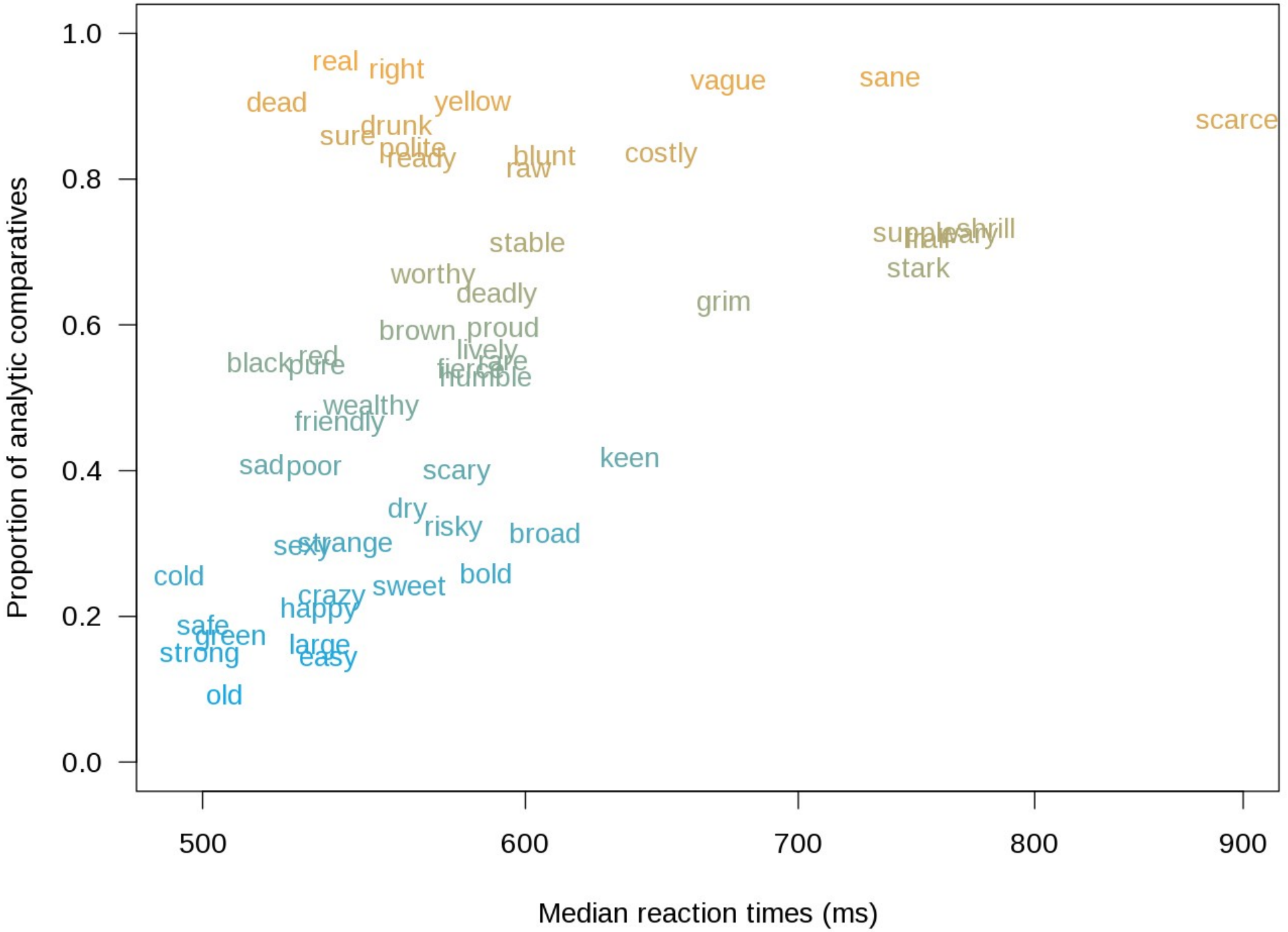
… responses that were neither analytic nor synthetic

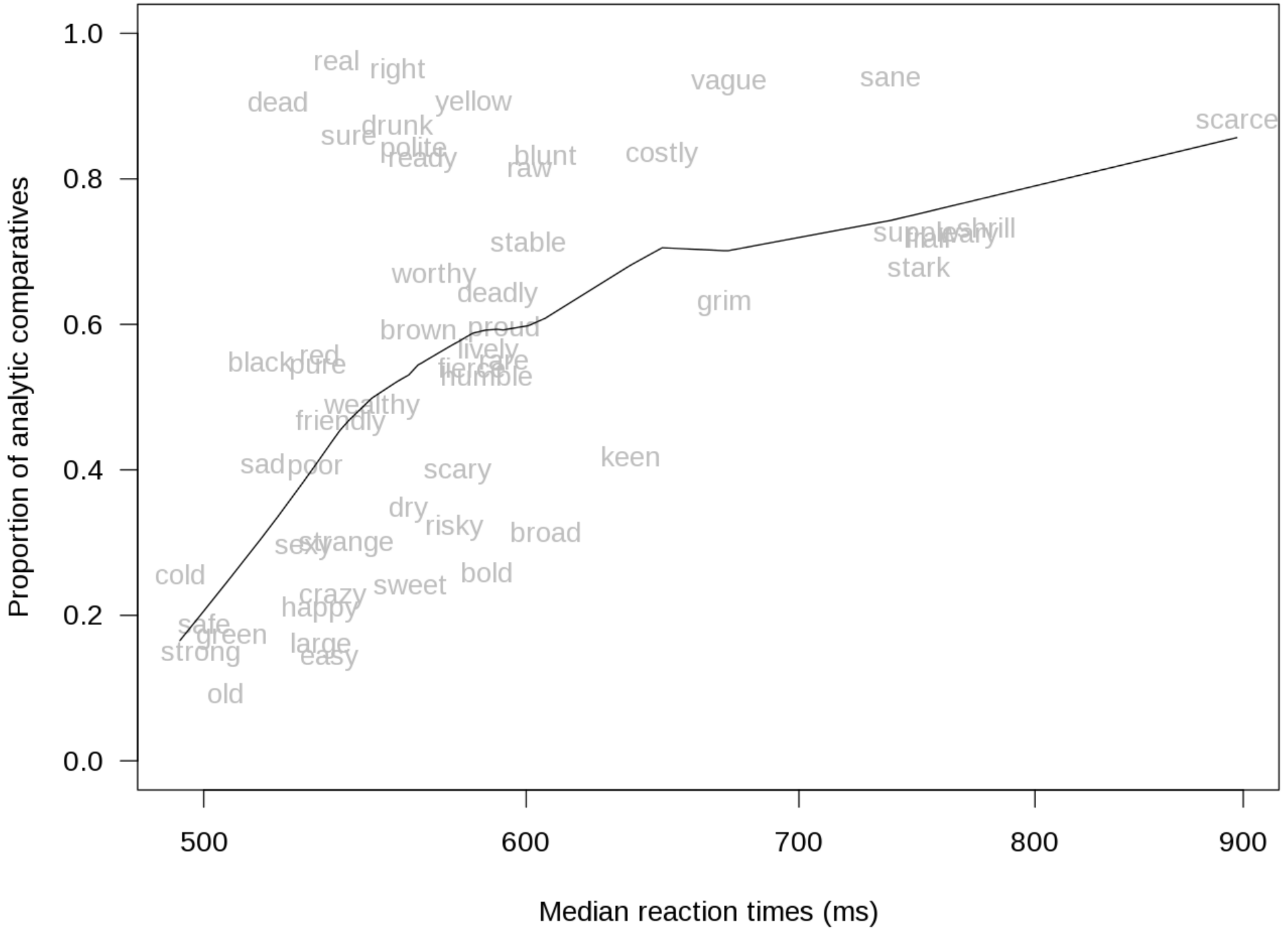
… 2 speakers who produced too few comparatives

… 6 adjectives with no variation or too few comparatives

1590 usable responses

- 747 synthetic comparatives
- 843 analytic comparatives


1.0
0.8
0.6
0.4
0.2
0.0
Proportion of analytic comparatives
500
600
700
800
900
Median reaction times (ms)
real
right
vague
sane
dead
yellow
scarce
sure
drunk
polite
ready
blunt
costly
raw
stable
shrill
stark
worthy
deadly
grim
brown
proud
lively
black
red
pure
rare
fierce
humble
wealthy
friendly
keen
sad
poor
scary
dry
risky
broad
sexy
strange
cold
bold
sweet
crazy
happy
safe
green
large
strong
easy
old

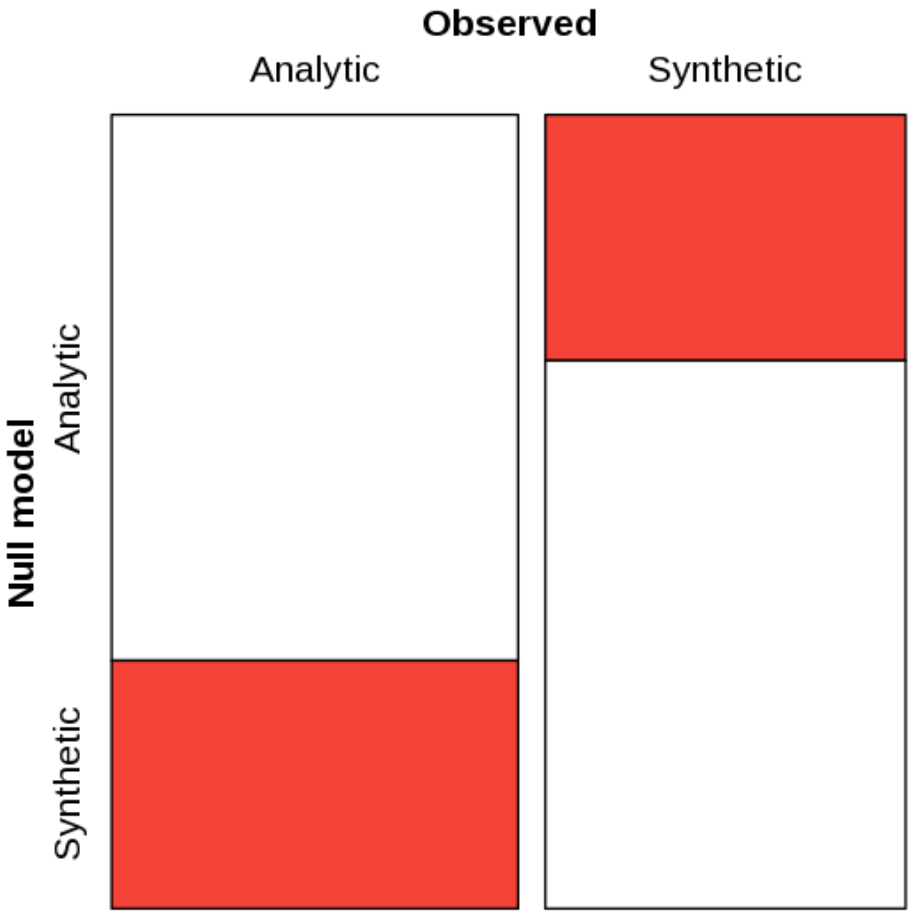

Proportion of analytic comparatives
1.0
0.8
0.6
0.4
0.2
0.0
500
600
700
800
900
Median reaction times (ms)
real
right
vague
sane
dead
yellow
scarce
drunk
sure
polite
ready
blunt
costly
raw
stable
worthy
deadly
grim
brown
proud
lively
black
red
pure
fierce
rare
humble
wealthy
friendly
keen
sad
poor
scary
dry
risky
broad
sexy
strange
cold
bold
sweet
crazy
happy
safe
green
large
strong
easy
old
stark

Statistical model

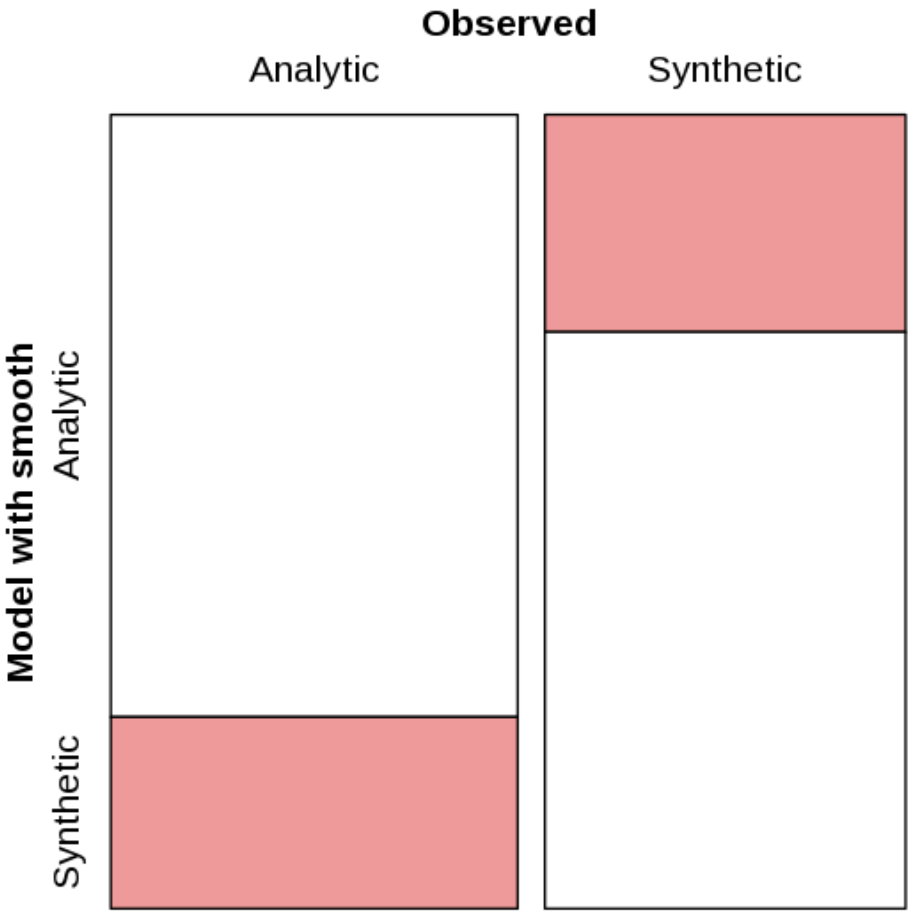
Analysis	General additive mixed regression (logistic)
Dependent variable	Type of response: Analytic/Synthetic
Smooth term	Median reaction times
Random effect	Speaker

Does a model with a smooth term for RT make better predictions than a null model (with only a random effect for Speaker)?

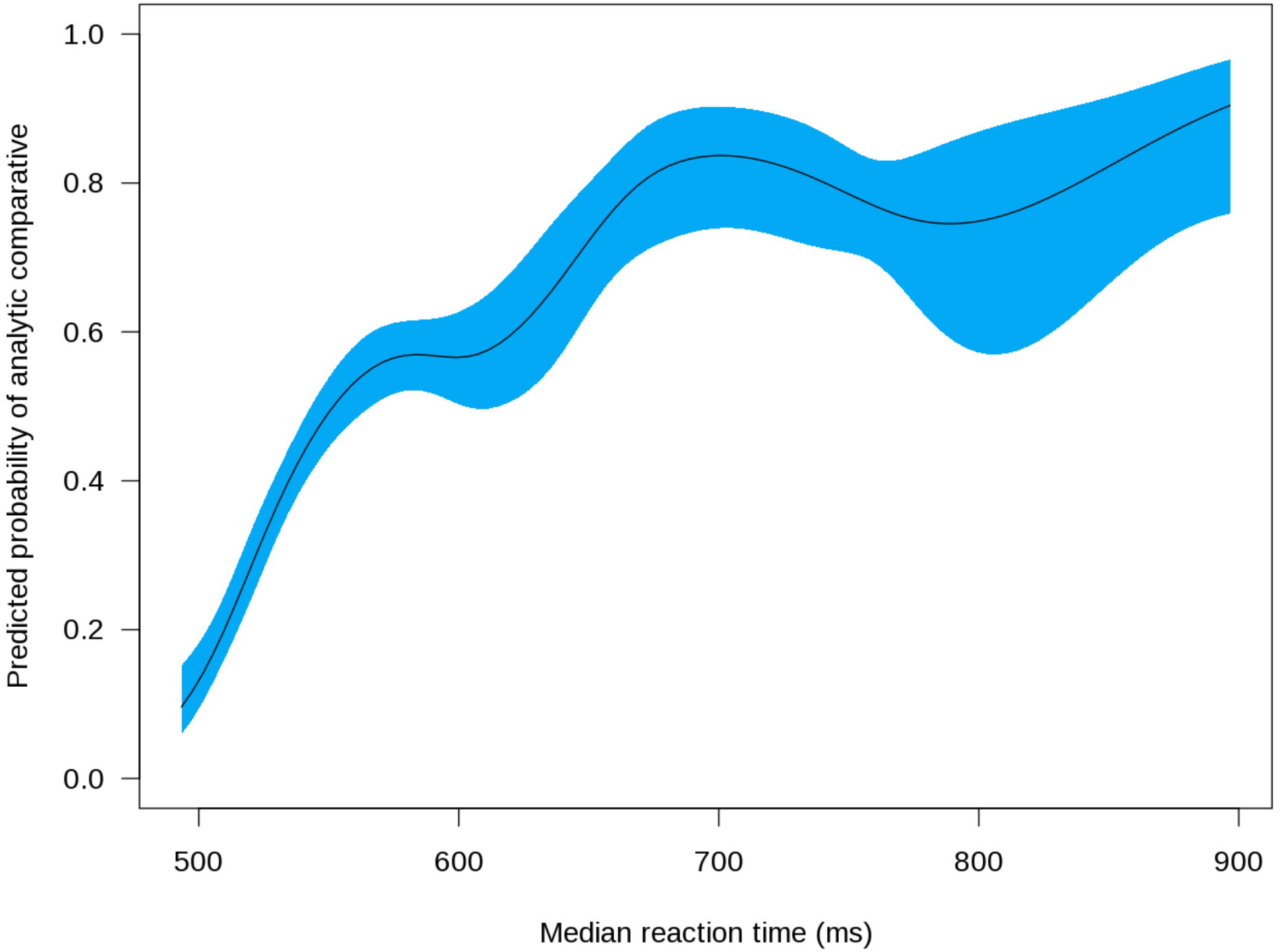
Results



68.9 percent accuracy



74.5 percent accuracy


1.0
0.8
0.6
0.4
0.2
0.0
Predicted probability of analytic comparative
500
600
700
800
900
Median reaction time (ms)

Summary: Production experiment

Reaction times help to predict speaker responses:

- analytic comparatives are preferred if reaction times are high
- synthetic comparatives are preferred if reaction times are low

- Analytic comparatives can be predicted somewhat better
- Individual preferences are an important factor

Speakers are more likely to use analytic comparatives with cognitively complex adjectives.

General summary and discussion

Summary of results

Perception experiment
Listeners do not benefit from *more*-support: Synthetic comparatives are processed faster.

Corpus study
Cognitively complex adjectives occur with a higher proportion of analytic comparatives in a corpus.

Production experiment
Speakers are more likely to use analytic comparatives with cognitively complex adjectives.

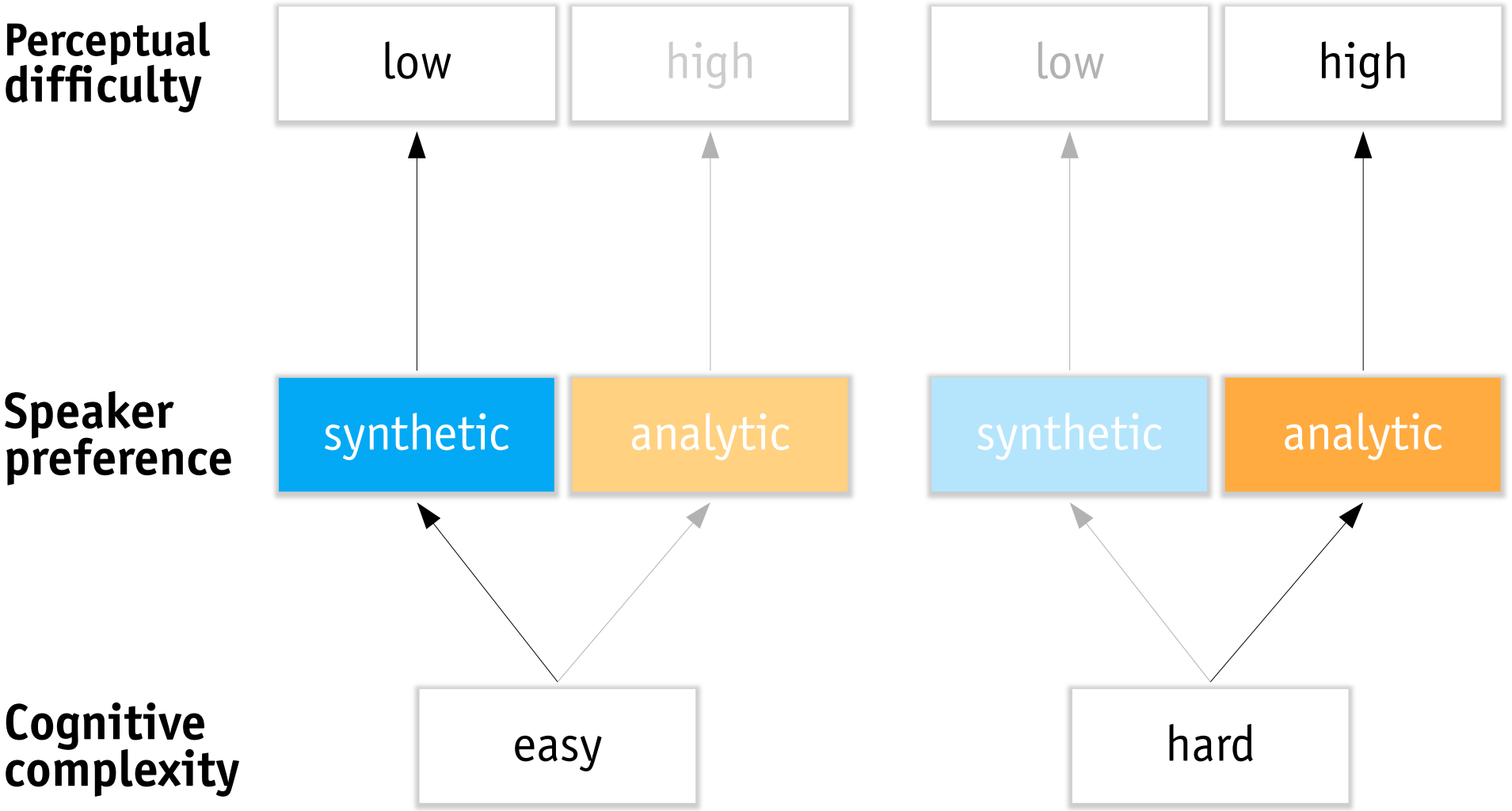
Cui bono?

- Listeners find synthetic comparatives easier to process
- Speakers prefer analytic comparatives for complex adjectives

***More*-support is probably a speaker-oriented compensatory strategy!**

Boyd (2007: 77):
"speakers treated complex items in the same way, regardless of whether a listener was present or not"

Conflicting demands



Conclusion

- Analytic comparatives are more difficult for listeners
- Comparative selection is affected by cognitive complexity
- No support for listener-oriented *more*-support,
  but support for **speaker-oriented *more*-support**
- *More*-support is **not co-operative** (contra “be perspicuous”, Grice 1975: 46)

References

Balota, David A., Melvin J. Yap, Michael J. Cortese, Keith A. Hutchison, Brett Kessler, Bjorn Loftis, James H. Neely, Douglas L. Nelson, Greg B. Simpson & Rebecca Treiman. 2007. The English Lexicon Project. *Behavior Research Methods* 39(3), 445–459.

Boyd, Jeremy. 2007. *Comparatively speaking. A psycholinguistic study of optionality in grammar*. Ph.D. dissertation, University of California, San Diego.

Davies, Mark. 2008-. *The Corpus of Contemporary American English (COCA): 450 million words, 1990-present*. Available online at: http://www.americancorpus.org.

Grice, H. Paul. 1975. *Logic and conversation*. In Peter Cole & Jerry L. Morgan (eds.), *Syntax and semantics*, 41–58. New York: Academic Press.

Hilpert, Martin. 2008. The English comparative. Language structure and language use. *English Language and Linguistics* 12(3), 395–417.

Kuperman Victor, Hans Stadthagen-Gonzalez & Marc Brysbaert. 2012. Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods* 44(4), 978–990.

Mondorf, Britta. 2003. *Support for* more-*support*. In Günter Rohdenburg & Britta Mondorf (eds.), *Determinants of grammatical variation in English*, 251–304. Berlin: Mouton de Gruyter.

Mondorf, Britta. 2009. *More support for* more-*support*. Amsterdam: John Benjamins.

Moscoso del Prado Martín, Fermín, Aleksandar Kostić & R. Harald Baayen. 2004. Putting the bits together. An information theoretical perspective on morphological processing. *Cognition* 94(1), 1–18.