



# Does high frequency lead to automaticity?

## A corpus study

Vsevolod Kapatsinski  
Indiana University



Workshop on Variation, Gradience and Frequency in Phonology  
2007, July 6-8, Stanford, CA

Work supported by NIH Training Grant DC-00012  
and NIH Research Grant DC-00111

### Introduction

#### The theory

- Bybee (2002): articulatory reduction is observed in high-frequency words because high frequency of use makes the production of a word more automatic.
- Automatic behavior is less subject to conscious inhibition.
- Therefore, if high frequency of use automates word production, the production of high-frequency words should be harder to interrupt than the production of low-frequency words.

#### The dependent variable:

Does the speaker interrupt the production of a word s/he intends to replace or repeat?

#### Repetition repair

##### Interrupted:

*It is a fi-, fixed female, by the way.*

##### Not interrupted

*There was another, another amusement park down there.*

#### Replacement repair

##### Interrupted

*It was pathe-, I mean, it was horrible.*

##### Not interrupted

*That's why we were surprised to see 'Toyota' written, I mean, imprinted on the engine.*

#### The hypothesis:

Words that are interrupted have a lower token frequency than words that are not interrupted.

e.g., frequency of *fixed* should be lower than frequency of *another pathetic* / *written*

### Data

Repair tokens from the Switchboard Corpus (Godfrey et al. 1992)  
Checked for transcriber error

#### Repetition repairs:

Included only:

- Single-word repetitions
- more than 1 syllable, 4-8 segments
- more than 1 segment pre-interruption

Excluded:

- Personal names
- Nonce formations
- Words with a higher-frequency homonym
- 'Very', 'really': uninterrupted repetition used for intensification

1018 total

#### Replacement repairs:

Included only:

- Single-word replacements
- Semantically related replacements
- replaced word 3-8 segments in length

1485 total

### Conclusion

- The production of a high-frequency word is less likely to be interrupted than the production of a low-frequency word even if the word is to be replaced
- This is consistent with Logan (1982) who found that typists, asked to stop typing a sentence when presented with a signal, stop less quickly if the signal is presented during the highest-frequency word *the*
- These results support the hypothesis that the production of high-frequency word is more automatic and therefore harder to interrupt than the production of low-frequency words (Bybee 2002)
- The frequencies of the replaced word and the replacement word show a positive correlation: semantically similar confusable words tend to have similar frequencies (Hotopf 1980)
- High-frequency words come to mind before low-frequency words: the replaced word is usually more frequent than the replacement word (not found previously, see Garrett 2001 for a review)

#### References:

Bybee, J. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, 14, 261-90.

Garrett, M. 2001. Now you see it, now you don't: Frequency effects in language production. In E. Dupoux, ed. *Language, brain and cognitive development*, 227-40. MIT Press.

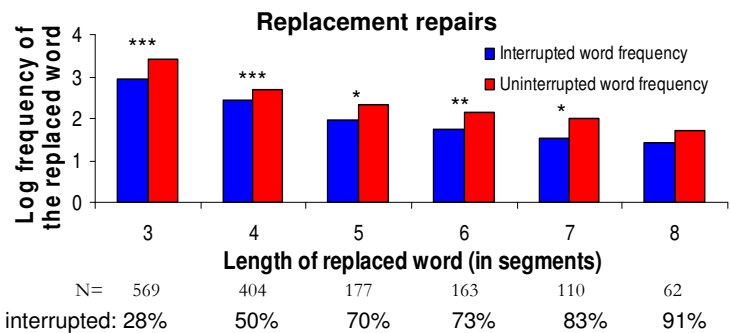
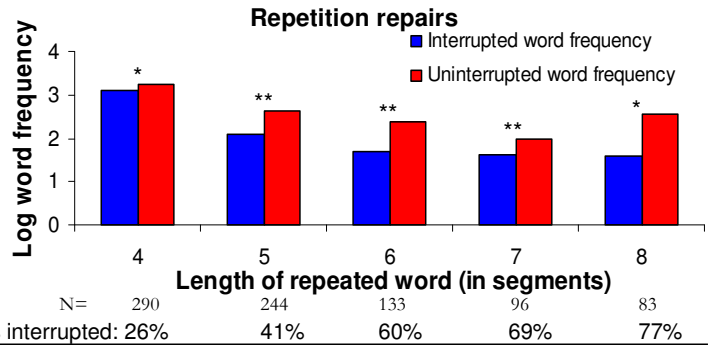
Godfrey, J. J., E. C. Holliman, and J. McDaniel. 1992. SWITCHBOARD: Telephone Speech Corpus for Research and Development. *IEEE ICASSP*, 1:517-20.

Hotopf, W. 1980. Semantic similarity as a factor in whole-word slips of the tongue. In V. Fromkin, ed. *Errors in linguistic performance*, 97-110. Academic Press.

Logan, G. D. 1982. On the ability to inhibit complex movements: A stop-signal study of typewriting. *JEP: Human Perception and Performance*, 8, 778-92.

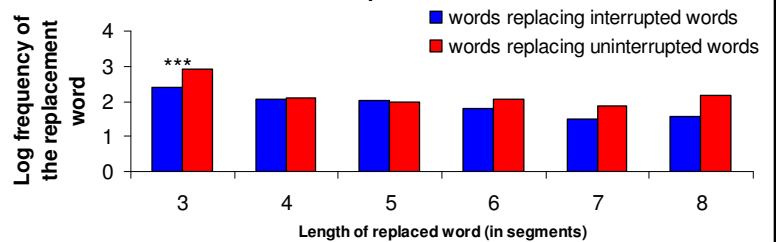
### Results

#### Interrupted words tend to have lower frequency than uninterrupted words



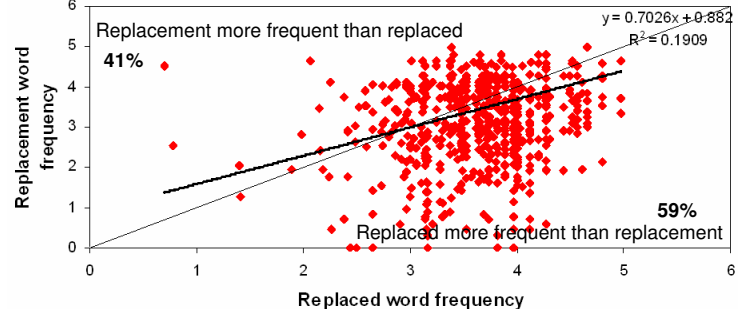
High frequency replacements could become available earlier, leading to interruption

#### But interrupted words actually tend to be replaced by words that are lower in token frequency than the words that replace uninterrupted words



If frequency of the replacement word mattered, we would expect significant differences in the other direction  
**Hence, it is the frequency of the pre-interruption word that matters for whether or not the words are interrupted**

#### The relation between frequency of the replaced word and frequency of the replacement word



- Replacement words tend to be more frequent than replaced words (the more frequent word comes to mind first) ( $p < .001$ )
- The frequency of replaced and the frequency of the replacement are positively correlated ( $p < .001$ )

\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$