# The perception of stop/sibilant clusters in Modern Hebrew

**Kyle S. Jones**
kysjones@email.arizona.edu

## Introduction

In this research, I investigate the causes of categorical stop-sibilant metathesis in *binyan hitpa'el*, the mainly reciprocal and reflexive verbal conjugation in Modern Hebrew. *Binyan hitpa'el* consists of a prefix /hit-/ followed by a verbal stem:

[hit-naʃek] "kiss one another"
[hit-labeʃ] "dress oneself"
[hit-ʁagez] "get excited"
[hit-xamem] "become warm"

However, if the verbal stem begins with a sibilant consonant, the /-t-/ of the prefix categorically metathesizes with a following /s z ʃ t͡s/:

| | |
|---|---|
| [histadeʁ] "get along with" | *[hitsadeʁ] |
| [hiʃtagea] "go crazy" | *[hitʃagea] |
| [hizdaken] "grow old" | *[hidzaken] |
| [hit͡stanen] "catch a cold, cool" | *[hitt͡sanen] |

Which of the two main theories of metathesis adequately accounts for the facts of Modern Hebrew metathesis? The indeterminacy/attestation model of metathesis (Hume 2004) suggests that indeterminacy of the speech signal leads listeners to reinterpret sequences into the more common, attested linear order in their language. On the other hand, Blevins & Garrett (2004) suggest that there are different types of metathesis which have different, distinct phonetic causes. Sibilant metathesis is a case of auditory metathesis, not perceptual metathesis, caused by auditory stream decoupling, the sibilant noise becoming dissociated from the rest of the auditory stream, resulting in listeners having difficulty determining the linear order intended by the speaker.

What do speakers of another language (English) do when faced with ambiguous *hitpa'el* stimuli? Do they reinterpret ambiguous stimuli based on their own native languages, or do they misperceive ambiguous stop/sibilant sequences in a similar direction as the metathesis in Modern Hebrew?

## Methods

Stimuli consisted of 198 *hitpa'el* verbs, ten metathesized and ten un-metathesized verbs in each of ten phoneme categories (except for /l/), drawn from Bolozky (2008). Stimuli were recorded by a native speaker of Modern Hebrew (male, late twenties) two times, both in quiet rooms, using Praat (Boersma & Weenink 2016). Stimuli were normalized for duration and amplitude and combined with a sample of multi-talker babble (Thibodeau n.d.) at a 1:2 SNR using a Praat formula, to replicate noisy conditions in real life and make the task more difficult. The volume for the experiment was set at 20%.

Twenty-one native, monolingual English speakers participated in the experiment, which consisted of a forced choice identification task programmed in DMDX (Forster & Forster 2003). There were four scrambled blocks of stimuli (5 items X 10 phoneme categories = 50 items per block), and participants used shift keys to select one of two choices (metathesized or unmetathesized).

## Results

Three participants ultimately were ultimately excluded. The overall error rate for the remaining 18 participants was 27.1%.

| Participant | Sex | Age | Error Rate |
|---|---|---|---|
| 03 | M | 22 | 20% |
| 05 | F | 20 | 24% |
| 06 | F | 33 | 43% |
| 07 | M | 22 | 34% |
| 09 | F | 22 | 21% |
| 10 | F | 19 | 46% |
| 11 | F | 32 | 12% |
| 12 | F | 24 | 26% |
| 13 | F | 22 | 14% |
| 14 | M | 19 | 23% |
| 15 | F | 24 | 37% |
| 16 | F | 19 | 19% |
| 17 | F | 27 | 29% |
| 19 | M | 21 | 15% |
| 20 | F | 19 | 40% |
| 21 | M | 25 | 18% |
| 22 | M | 31 | 32% |
| 23 | M | 33 | 34% |

Table 1: Participant Results

A logistic regression statistical analysis in Rstudio, with reaction time (Resp) as the dependent variable and metathesis (Meth), phoneme type (Seg), and voicing (Voice) as independent variables, revealed that only phoneme type was significant (each at $p < 0.0001$), with voicing barely significant ($p < 0.05$).

A separate linear mixed effects regression (LMER) revealed that only fricatives, sibilants, and stops were significant ($p < 0.0001$), while sonorants were barely significant ($p < 0.05$). Voicing, instead, was highly significant ($p < 0.0001$), and metathesis was significant as well ($p < 0.01$), but only together with voicing. Statistical results demonstrate that phoneme type is important in determining whether a listener will misinterpret a sound sequence, with voiced sibilants having the greatest tendency to be misinterpreted.

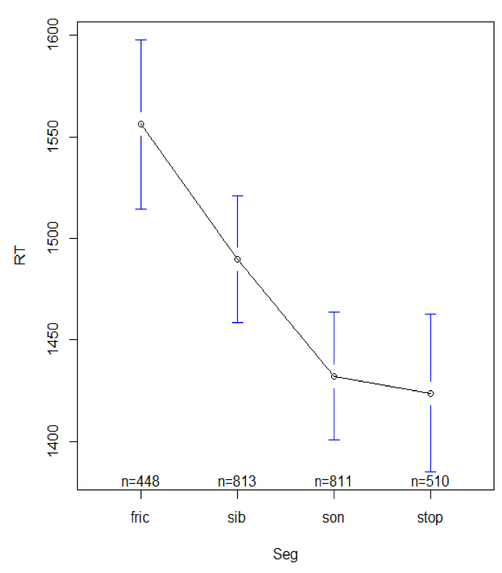


Figure 1: RTs for Correct Response Across Phoneme Types

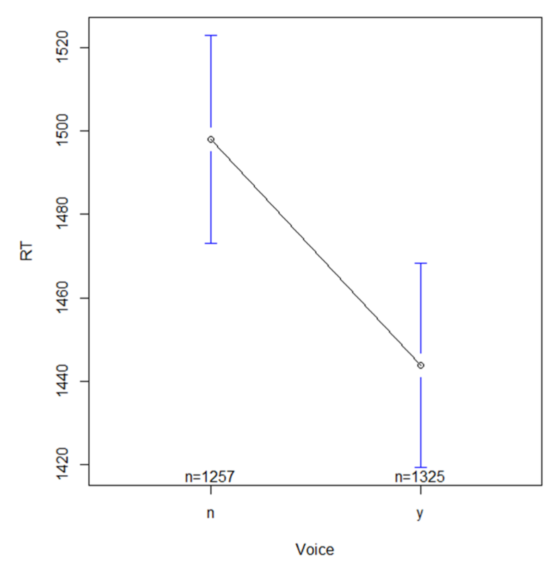


Figure 2: RTs for Correct Response Across Voicing

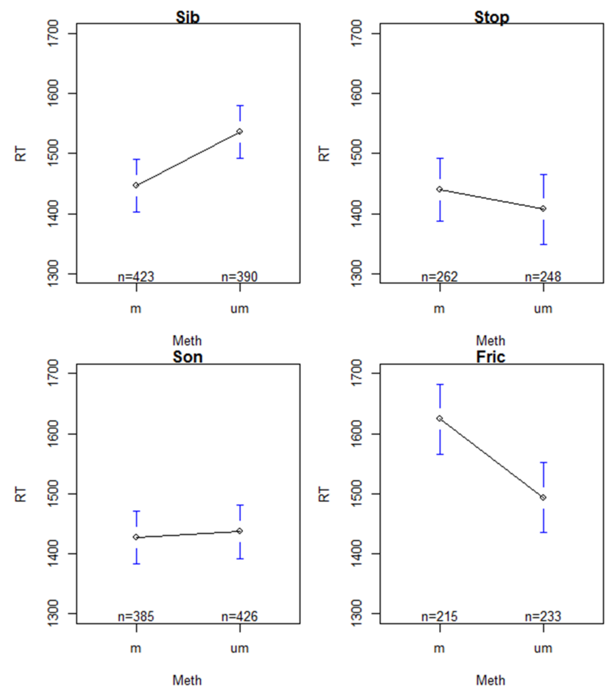


Figure 3: RTs for Correct Response

| Sound Sequence | Phonotactic Frequency | Rate of Confusion |
|---|---|---|
| tp | 5.6% | 37.5% |
| pt | 94.4% | 62.5% |
| tk | 2.9% | 55.7% |
| kt | 97.1% | 44.3% |
| ts | 35.3% | 46.6% |
| st | 64.7% | 53.4% |
| dz | 72.5% | 42.7% |
| zd | 27.5% | 57.3% |
| tt͡s | 0% | 42.9% |
| t͡st | 100% | 57.1% |
| tl | 60.0% | 73.4% |
| lt | 40.0% | 26.6% |
| tm | 93.6% | 52.3% |
| mt | 6.4% | 47.7% |
| tn | 7.8% | 37.9% |
| nt | 92.2% | 62.1% |

Table 2: Phonotactic Frequencies for English (MRC Psycholinguistic Database)

## Conclusions

The indeterminacy/attestation model of metathesis (Hume (2004) can account for the tendency of English speakers to misperceive [ts] as [st] and [tt͡s] as [t͡st], both of which have greater phonotactic frequencies in English. However, the tendency to misperceive [dz] as [zd], even though [dz] has a higher phonotactic frequency in English, is a clear counterexample to Hume (2004).

Evolutionary phonology (Blevins & Garrett 2004) provides a plausible explanation in auditory stream decoupling. Human auditory perception consists of auditory scene analysis, the extremely complex process by which the cacophony of sounds in the environment are organized into meaningful parts. Physical events occur in the world, vibrating specific parts of the cochlea, creating something like a noisy spectrogram. Auditory streams are generated which partition the sounds created by physical events into discrete parts which can be analyzed and interpreted by the brain. In auditory stream decoupling, the sibilant noise is generated on a different auditory stream from the rest of the word. When it comes time for higher order hierarchical organization (in the phonology), the sibilant noise can be mis-/reinterpreted in a linear order, depending on specific characteristics of the language, that differs from the speaker's intended order. This explains why sibilants, in particular, are often involved in metathesis, which Hume (2004) cannot explain.

Metathesis is important for a listener-directed model of language change, like Blevins (2004), in the CHANGE – CHANCE – CHOICE (CCC) model. Sibilant metathesis is an instance of CHANCE, where listeners perceive correctly but reinterpret the speakers' intentions.

How did a listener misinterpretation become a categorical feature of the grammar? According to Hyman (2008), "universal phonetics" → "language-specific phonetics" → "phonology" (238).

## Literature cited

Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge, UK: Cambridge University Press.
Blevins, J., & Garrett, A. (2004). The evolution of metathesis. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology*. Cambridge, UK: Cambridge University Press. 117-156.
Boermsa, P., & Weenink, D. (2016). Praat: doing phonetics by computer [Computer program]. Version 6.0.19, retrieved 13 June 2016 from http://www.praat.org/
Bolozky, S. (2008). *501 Hebrew verbs*. Hauppauge, NY: Barron's Educational Series.
Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35*(1), 116-124.
Hume, E. (2004). The indeterminacy/attestation model of metathesis. *Language, 80*(2), 203-237.
Hyman, L. (2008). Enlarging the scope of phonologization. UC Berkeley Phonology Lab Annual Report.
RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA. Retrieved from http://www.rstudio.com/
Thibodeau, L. (n.d.). Retrieved from http://www.utdallas.edu/~thib/fm_wav.html
Wilson, M. D. (1988). The MRC Psycholinguistic Database: Machine readable dictionary, Version 2. *Behavioural Research Methods, Instruments and Computers, 20*(1), 6-11.

## Acknowledgments

I would like to thank Samira Farwaneh, Adam Ussishkin, Mahmoud Azaz, and Natasha Warner for their help in developing this experiment. I would also like to thank the School of Middle Eastern Studies, the Center for Middle Eastern Studies, and the Graduate and Professional Student Council at the University of Arizona for their financial support.

COLLEGE OF SOCIAL & BEHAVIORAL SCIENCES
School of Middle Eastern & North African Studies