Introduction:

Speech errors have been studied in the lab and recorded in natural speech for large corpora. These studies have led to a wide variety of research on production. Although this research does not aim to support a specific language model (since it is exploratory), it does assume a two stage process of lexical retrieval seen in the ‘Representations and Processing Components Model’ of planning (Jaeger 7). These processes include the retrieval of the lemma (semantic and syntactic information) and the retrieval of the form (phonological, orthographical, and morphological information) (Jaeger 7).

There are a few fundamental features that have been researched which were hypothesized to affect various aspects of speech errors. Harley and MacAndrew did a large analysis of imaginability, frequency, phonological facilitation, semantic relationship, syntactic category, and word-length (2001). Other researchers have expanded this to other factors including neighborhood density, neighborhood frequency, age of acquisition, and phrasal constructions (Vitevitch 1997; Kittredge, Dell, Verkuilen, & Schwartz 2008; Menn & Duffield 2013). Looking at one or a combination of these factors, various types of lexical errors have been explored including paradigmatic lexical substitutions, lexical blends, and syntagmatic lexical errors (Haley & MacAndrew 2001; Vitevitch 1997; Kittredge, Dell, Verkuilen, & Schwartz 2008; Menn & Duffield 2013; Laubstein 1999).

Most of these have studied normal adults or aphasics with only a few developmental studies of speech errors looking at various factors as indicators of the development of processes. Frequency is not one of these longitudinal factors explored in prior literature. In regards to frequency, there has been some debate as to where in the speech production planning process it plays the largest role and the controversy has emerged complex picture of lexical production. Vitevitch found an interaction where malapropisms (when a phonologically but not semantically related word substitutes for another) occurred more with high frequency words in sparse neighborhoods (vice versa for low frequency words) as well as the finding that the targets tended to have lower frequency than the error substitute (1997). Harley and MacAndrew found frequency to be most influential in phonological retrieval in lexical substitutions although the overall picture was complicated (2001). This was also supported in word blends where phonologically related blends were more affected by frequency (Laubstein 1999). However, Kittredge et al. did find some evidence of an effect of word frequency in both the phonological and semantic retrieval (2008). Regardless, as seen in speech errors of various forms, frequency is particularly important in the mental organization and representations of language (Menn & Duffield 2013).

Results:

- One can see in these graphs how the variation in frequencies increases with age in these three general age groups. There are some complications however with a lack in data in some error types. This is an ongoing project however so more data is added for further analysis.

- As for the division between semantics and phonology lexical substitutions, there is a change in frequency in age. The frequencies of the error word are always higher than that of the target word when there is a phonological and a semantic relationship between them although this gap shrinks with age. When young, the target word has a higher word frequency in malapropisms but the reverse is true by middle childhood and remains stable into adulthood. Similarly, the error word is far more frequent when young when there is only a semantic relationship and the reverse occurs and plateaus by middle childhood. Errors arising from words that have no semantic nor phonological similarity follow no trend most likely due to lack of data (as it is the smallest category).

References:


Method:

Data Collection - The data was collected naturally with pen and paper in spontaneous speech and conversation. Ambiguous cases were resolved with the perspective of the error’s speaker and otherwise omitted.

Corpora Included - The young children’s and the middle children’s data were from Dr. Jaeger’s personal database as well as part of the adult database. The rest of the adult data comes from a personal corpus of adult native speakers collected and analyzed by the researcher. The frequency counts used in the calculations came from The Corpus of Contemporary American English (Davies 2008-).

Parameters - Errors that involved proper nouns or non-words (as the target or error) were not included in this analysis because their frequency counts vary greatly among speakers (someone who lives with a “Sandra” will say it more and bias the data). Errors that arose from ‘on the mind’ (aka Freudian slips) or from the environment were excluding since the word source was influenced by non-linguistic factors rather than errors purely from retrieval of forms. As mentioned before, ambiguous cases were also excluded.

Examples:

Lexical Substitution Content Words (Phonologically related only aka malapropism)
AF: ‘That’s not what I decide, what I describe as créamy.’
Lexical Substitution Function Words (Semantically related only)
AF: ‘I’ll let you up, I’ll let you up... Out.’
Lexical Blend of ‘registration’ and ‘registrating’ (Both semantically and phonologically related)
AF: ‘I didn’t see anything about (a thing) sitting in, registering for commencement.’
Lexical Anticipation (Semantically related only)
AF: ‘There are two face- things that I cannot do: faces and places.’
Lexical Perseveration (Neither phonologically nor semantically related)
AF: ‘I don’t want to cite cite thing, one thing...’
Lexical Exchange (Semantically related only)
AF: ‘You would never down a wine of glass... a wine of glass?’