

Asymmetries in Speech Errors and their Implications for Underspecification

Marianne Pouplier and Louis Goldstein
Haskins Laboratories and Yale University

1 Introduction

In speech production research, speech errors have long been appealed to in evaluating the psychological reality of linguistic units. The systematic nature of errors with respect to their occurrence and distribution (cf. for instance Fromkin, 1971; Shattuck-Hufnagel & Klatt, 1979; Dell, 1986; Stemberger, 1991a) allows us to assume that those units that appear to behave independently in speech errors are presumably - at least at some point - units of processing. More specifically, it has been proposed that the asymmetric distribution of single segment errors involving coronals provides evidence for their underspecification (Stemberger, 1991a; Stemberger & Stoel-Gammon, 1991).

However, an important caveat in speech error research is that the method for detecting errors relies on impressionistic transcription. This leaves open the possibility that the inherently segmental nature of transcription carries its own bias into the data. Subphonemic errors or errors resulting in a phonologically ill-formed utterance, for instance, are difficult to transcribe in a segmental system, which may be one of the reasons why this kind of error is so rarely reported (cf. among others Fromkin, 1971; Cutler, 1981; Ferber, 1991; Boucher, 1994). In recent years, studies have begun to investigate speech errors by means of instrumental measurements. Articulatory and acoustic studies (Mowrey & MacKay, 1990; Boucher, 1994; Goldstein et al., in prep.; Frisch & Wright, in press) have shown empirically that errors are not a matter of all or nothing - that is, systematic errors can occur below the level of a segment instead of being confined to a temporal misselection of phonological (abstract) units.

This paper follows a new perspective on speech errors within the framework of Articulatory Phonology, as proposed by Goldstein et al. (in prep.). On the basis of kinematic evidence, their work has demonstrated that speech errors are not restricted to categorical exchanges of position of segmental units, but rather gestures that compose segments can exhibit errors that vary from zero to maximal in magnitude.

Here we report results from two perceptual experiments which use stimuli selected on the basis of their articulatory properties only, covering a range of errorful gestural activations. The outcome of the perceptual experiments suggests that different segments show different degrees of vulnerability to (subsegmental) speech errors: While listeners detected errors reliably for some segments, for other segments the reaction to errorful and non-errorful tokens was not distinct. The data suggest that at least for some error types an asymmetric error distribution arises due to perception, while production itself is not asymmetric. However, for error types involving segments whose gestural compositions stand in a subset relationship to each other (as described below), asymmetries may indeed originate in production due to the overall dominance of a gestural intrusion bias observed in the production data of Goldstein et al. (in prep.).

2 Errors as abstract segment exchanges

The examination of corpora of spontaneous errors as well as results of elicitation experiments have shown that the most commonly occurring errors involve single segments, while single feature errors are rarely reported (cf. for instance Fromkin, 1971; Shattuck-Hufnagel, 1979; 1983; Dell, 1986). Likewise syllables are usually not affected by exchange errors. The notion that phonological segments are primary units of word-form retrieval has thus gained widespread acceptance, as has the view that errors are phonologically well formed: An activated (abstract) segment is categorically shifted to a wrong position within a 'prosodic frame'. In this new position, the segment will be produced 'normally', as if it were the intended segment. Thus allophonic features typically pertain to the new ('wrong') position of a segment (Shattuck-Hufnagel & Klatt, 1979; Shattuck-Hufnagel, 1983).

Errors show further a frequency bias in that more frequent elements are less likely to be affected by error than less frequent elements. Frequency of occurrence is also reflected in a directionality effect: less frequent elements are usually replaced by more frequent elements. For some segments, however, an anti-frequency bias has been reported (in experimental studies as well as in corpora). While /t/ is a more frequent segment in English than /k/, for instance, and thus the expected substitution-directionality is $*/k/ \rightarrow /t/$, the opposite is actually observed: /t/ is more often substituted by /k/ than vice versa. This is also the case for /s/ and /ʃ/, with /s/ as the more frequent element turning more often into /ʃ/ than vice versa. At the same time, Stemberger (1991a) and Stemberger and Treiman (1986) identify an addition bias in cluster environments: In errors, it is more usual for a segment to be added than to be deleted. Stemberger (1991a) reinterprets the anti-frequency effect as surface manifestation of the addition bias by invoking the concept of coronal underspecification: Given that coronals are underspecified for place of articulation, so he argues, the addition bias will lead other segments' place specifications to intrude more easily (since the 'empty space' is a willing host), independent of segment frequency.

3 Gestures as basic units

The framework of Articulatory Phonology (Browman & Goldstein, 1989; 1992; 2001) opens a new perspective on speech errors and the speech production system. Within Browman and Goldstein's framework, dynamically specified gestures are hypothesized to be the basic units of speech production. Errors can thus be interpreted as reflecting the gestural structure of speech, in that they can activate components of gestural structures in varying magnitudes. That is, both individual gestures as well as larger units consisting of tightly cohesive multiple gestures can be involved in erroneous productions. In a magnetometer (EMMA) study by Goldstein et al. (in prep.) support for these assumptions was gained from kinematic speech error data. Errors were observed to be gradient in nature; that is, an individual gesture can take on a continuum of values, varying from zero to maximal. While a segmental approach would be able to accommodate gradient activations at the articulatory output level, there is evidence for gestures as units of higher levels of organization. It was shown that tightly cohesive multigestural constellations can be broken up in an error, and an individual gesture that forms part of a multigestural unit can isolatedly show up at an erroneous temporal location.¹ Further, asymmetries that occur in the articulatory data cannot stem from abstract

¹ Goldstein et al. (in prep.) define two types of errors, 'reduction' and 'intrusion'. A reduction error is defined as an erroneous decrease of the magnitude of the target constriction (e.g., a decreased magnitude of the tongue tip gesture during /t/). An intrusion error is defined as the constriction of a vocal organ that is not controlled in the normal, non-errorful gestural constellation (e.g., addition of a tongue dorsum gesture during /t/).

segment exchanges: The overall observed bias for gestural intrusion as opposed to reduction has the consequence that often two gestures (one appropriate, one intruding) are produced at the same time. For example, during the repetition of the phrase *cop top*, errors are observed in which an intruding /k/-like dorsum (TD) gesture is produced concurrently with the tongue tip (TT) gesture of *top*. Goldstein et al.'s (in prep.) results show that there is no asymmetry in production between /t/ and /k/ - the relevant asymmetry is between reduction and intrusion errors instead. This asymmetry affects /t/ and /k/ equally. For /s/ and /ʃ/, the situation is more complex, since the distinction intrusion - reduction can only be defined in terms of the non-shared vocal tract variable (tongue body; TB).² For TB, the addition bias is confirmed in Goldstein et al.'s data.

These findings lead to the question where the asymmetry that has been reported between coronals and non-coronals might stem from. From the perspective offered in this paper, asymmetries have two potential sources: Asymmetries may originate in *production*: The intrusion bias may lead to one segment prevailing over another. However, production asymmetries between /t/ and /k/ are not evidenced in the production data of Goldstein et al. Asymmetries may originate in *perception*: The intrusion bias might have different perceptual consequences for different segments. In recording errors, perceptually more salient errors may thus come to be overrepresented. That is, gradient errors and their interaction with perceptual biases might account for the asymmetries. In order to put this possibility to test, Goldstein et al.'s kinematic data were used as stimuli in a perceptual experiment with the aim of determining how the articulatory error distribution maps onto perception.

4 Experiment 1: /t/-/k/

4.1 Method

Goldstein et al. (in prep.) used bisyllabic alternating phrases in their articulatory study (e.g. "coptop"; "kiptip"), which were repeated by the subject for about 10 seconds at a time. They distinguish between errorful and non-errorful utterances as follows: Their control condition involved non-alternating phrases (e.g. "copcop", "toptop"), during which no errors occurred. Error definition will be exemplified here for tongue dorsum (TD) magnitude during a /t/ from an alternating trial. The tongue dorsum value during the intended /t/ from the alternating condition is evaluated against the distribution of the non-alternating controls. If the token's gestural magnitude is more than two standard deviations from the TD mean of the /t/-control, it is classified as an 'error'. Further, gradient and categorical errors are distinguished: A gradient error on TD during /t/ is more than two standard deviations from the /t/-control mean, but not as extreme (less than two standard deviations) from the TD values of the /k/ control mean. A categorical error on this particular /t/ token would involve a gestural magnitude which is not only more than two standard deviations from the /t/ control mean but even exhibits a TD value that is within the two standard deviation range of the TD values of the /k/ controls. That is, the TD value of a /t/ defined as categorical error is within the 96% range of the distribution of a non-errorful /k/.

The digitized audio of the original bisyllabic utterances of the production experiment were edited into single syllable utterances (e.g., "cop"). In a simple (not choice) reaction time task, participants were instructed to listen to "short words" presented in random order and

² Browman and Goldstein (2001) hypothesize the gestural specification for /ʃ/ to involve TT and TB, while the gestural specification of /s/ involves TT only. Since an intrusion error is defined by Goldstein et al. (in prep.) as intrusion of a constriction that is not controlled in the non-errorful gestural constellation, this criterion does *per definitionem* not apply to TT during either /s/ or /ʃ/.

decide whether the words begin with a particular consonant sound. Subjects ranged from 18 to 44 years of age and were paid for their participation. 14 subjects were tested. Data from 3 subjects were discarded since their identification rate for the error-free controls was below 50%.

The stimulus list contained 71 single syllable tokens which were selected according to their membership in a particular 'error category' (cf. Table 1).³

	categorical	gradient	no error
Intrusion	<i>t</i> (10)	<i>t</i> (10)	<i>t</i> (10)
	<i>k</i> (12)	<i>k</i> (8)	
Reduction	-	<i>t</i> (5)	<i>k</i> (10)
	-	<i>k</i> (6)	

Table 1. Single syllable stimuli for perceptual task grouped into error categories. Numbers in brackets indicate the number of tokens representing each category.

None of the tokens selected has an error on more than one constriction (i.e. never a reduction as well as intrusion error on the same token). Where possible, a minimum of 10 tokens from each category was selected for both /t/ and /k/. Two major categories of exceptions were categorical and gradient reduction errors: The data collected in Goldstein et al.'s (in prep.) EMMA experiment did not contain any reduction errors of categorical magnitude in TD during /k/ with no accompanying error on TT. Likewise there are no categorical reduction errors on TT during a /t/ with no error simultaneously occurring in TD. Also gradient reduction errors, i.e. errors on the target gesture (TD for /k/ and TT for /t/) with no errorfult intruding gesture (TD for /t/ and TT for /k/) are also underrepresented with only one occurrence for *kip* and *tip*, but 4 during *top* and five during *cop*.⁴

Using Psyscope, the stimuli were presented 12 times overall (6 times each in the 2 different conditions described below), randomized differently each time. Subjects sat in a soundbooth in front of a computer screen and a button box. Stimuli were presented over headphones. Two different monitoring conditions were employed: Subjects were asked to decide whether they heard an initial /t/-sound (condition 1), or an initial /k/-sound (condition 2). If they heard the given sound, they were instructed to press a response button as quickly as possible, otherwise, they were instructed to wait for the next trial. The conditions were blocked in cycles of 3, that is, the program cycled 3 times through the entire /t-k/ stimulus list in 3 different randomizations while subjects monitored for /t/. In the subsequent 3 cycles, subjects were asked to monitor for /k/, then again for /t/ and once more for /k/. Between these blocks of 3 cycles, subjects were given the option to take self-terminated breaks, i.e. subjects end breaks by pressing the response button. During a given block, a letter representing the sound subjects should be monitoring for was displayed on the screen. The time between the onset of two successive stimuli was 2000 ms, partitioned into a 1500 ms response window and 500 ms inter-trial time. During the window of 1500 ms subjects heard the audio stimulus and response time was measured (the response window started with the onset of the audio stimulus). Independent of whether a response was recorded, the next trial came up after 2000 ms.

³ Goldstein et al.'s study included various rate, stress, vowel (*cop top* and *kip tip*), and phrase position conditions. The tokens selected for the present experiment were distributed across all of these conditions.

⁴ The difference between /t/ and /k/ in the gradient-categorical intrusion categories is due to a coding mistake in the experimental setup; these numbers do not reflect genuine differences in frequency of error.

4.2 Results and discussion

The goal of our analysis was to determine how detectability of /t/ and /k/ was affected by error status. Instead of analyzing percent of correct identifications directly, the data were transformed using a non-parametric test for sensitivity (Grier, 1971). This sensitivity measure takes into account subjects' inherent response bias by adjusting the number of 'hits' (i.e. correct identification responses) for the number of 'false alarms' (i.e. incorrect positive responses). The adjustment formula, given in (2), yields a maximal identification value of 1 (only hits, no false alarms) and a minimum value of 0 (only false alarms):

$$(2) I = (1 - P(\text{fa}) + P(\text{hit})) / 2$$

where $P(\text{hit}) = P(r|G)$ and $P(\text{fa}) = P(r|X)$ and r is a response. G is a stimulus with target gesture r , and X is a stimulus not having target gesture r .

Results are shown in Table 4. A 2-way ANOVA with repeated measure on both factors was performed. The two factors are error type (categorical - gradient - no error) and intended target (t-k), whereby intended target refers to the speaker's target in Goldstein et al.'s production study. Each subject contributes one sensitivity value per intended target per error type. Factor error type is significant at $p < 0.01$ ($F(2, 20) = 51.13, p < 0.0001$); factor intended target is not significant ($F(1, 10) < 1$), indicating that the overall sensitivity is the same for /t/ and for /k/. The interaction effect is significant ($F(2, 20) = 21.83; p < 0.0001$). The interaction arises because the effect of error category is stronger for /t/ than for /k/.

error type	intended target <i>t</i>	intended target <i>k</i>
categorical	0.73 (0.11)	0.85 (0.07)
gradient	0.9 (0.04)	0.86 (0.07)
no error	0.97 (0.03)	0.86 (0.08)

Table 2. Means (and standard deviations) for sensitivity results for /t/ and /k/ grouped by error category.

A 1-way ANOVA follow-up with a posthoc (Ryan-Einot-Gabriel-Welsch Multiple Range Test) on the interaction means shows that while for /t/ all three error types are significantly different from each other, none of the means for /k/ differ significantly. These results indicate that errors are asymmetric in that /t/ is perceptually more affected by error than /k/.

As to the apparent anti-frequency bias that has been reported for /t/ and /k/ for error data collected by means of transcription, these results indicate that these asymmetries reflect a property of the perceptual system rather than the production system: if errors are systematically heard more easily on /t/ than on /k/, this perceptual asymmetry will substantially affect the error distribution in corpora.

5 Experiment 2: /s-/ʃ/

5.1 Subjects and experimental setup

The same experimental setup and the same subjects (in a separately scheduled session) were used for /s-/ʃ/; 13 of the earlier subjects were available. Due to coarticulation, the release of the preceding /p/-closure in a *sopshop* phrase is audible during the frication, even after the utterances have been cut up into individual syllables. To ensure that subjects would parse the bilabial release as coda instead of as complex /ps/ or /pʃ/ onset, a syllable /op/ was spliced at the beginning of all tokens.⁵ A silence interval of 100 ms was spliced to the end of the /o/ vowel to make the /p/ closure of constant length. The instructions specified that subjects would hear bisyllabic words with the first syllable always being /op/. It was specifically pointed out to them that there was no consonant at the beginning of the word. Their task was specified as determining whether the second syllable begins with a given consonant sound.

For Experiment 2, data from 2 subjects were discarded since their identification rate for the error-free controls was below 50%. The stimuli distribution is given in Table 9. Selected tokens are distributed across all rate, stress and phrase position conditions.

TB	categorical	gradient	no error
intrusive	s (5)	s (4)	s (5)
reductive	ʃ (5)	ʃ (5)	ʃ (5)

Table 3. Stimulus categories for /s-/ʃ/. Numbers in brackets indicate the number of tokens representing each category.

Since only the activity of one tract variable, i.e. tongue body (TB), can be differentiated in terms of reduction and intrusion as defined by Goldstein et al. (cf. fn 2), fewer tokens were tested than in Experiment 1. In addition, only error data for one vowel condition were available from the EMMA experiment. The stimulus list for the fricatives thus contained 29 tokens.

Recall that for /t-/k/, stimuli were selected such that there are never two co-occurring errors during one token, i.e. no token has an error on both constrictions at the same time. For /s-/ʃ/ this selection criterion had to be modified, since TT and TB receivers are not independent in the way TT and TD are (TT and TB receivers were about 20 mm apart). That is, it is not possible to select tokens with an error on TT only or TB only; for the vast majority of tokens, an error on one tract variable is accompanied with an error on the other tract variable. Neither is it possible to systematically vary error degree (i.e. gradient TT with categorical TB error, categorical TT with categorical TB error, etc.), since not enough representatives of each type are in the EMMA data. Note that non-errorful tokens are truly 'error free'; gestural magnitudes for both constrictions are well within 1 standard deviation of the control mean.

⁵ The syllable was a stressed, fast rate, non-errorful utterance of *shop*: The frication part as well as the first temporal half of the vowel were cut off (resulting in a vowel duration of 55.1 ms).

5.2 Results

Like for /t-/k/, a 2-way repeated measures ANOVA was performed on the nonparametric sensitivity measure with repeated measures on both factors. Factors are error type (categorical-gradient-no error) and intended target (/s-/ʃ/). Factor error type is significant ($F(2,22) = 289.04$; $p < 0.0001$). Factor intended target is not significant ($F(1, 11) = 1.74$; $p = 0.2142$); the interaction effect reaches significance ($F(2,22)=15.12$, $p < 0.001$).

error type	s	ʃ
categorical	0.41 (0.14)	0.36 (0.23)
gradient	0.88 (0.14)	0.6 (0.18)
no error	0.92 (0.1)	0.96 (0.04)

Table 4. Means (and standard deviations) for sensitivity results for /s/ and /ʃ/ grouped by error category.

A 1-way ANOVA on the interaction means with a posthoc test (Ryan-Einot-Gabriel-Welsch Multiple Range Test) shows that the sensitivity values for /ʃ/ are significantly different for categorical, gradient and no error. For /s/, on the other hand, only categorical errors significantly affect sensitivity. There is no statistically significant difference between the no error and gradient error categories. These results were unanticipated: the perception of /ʃ/ was found to be more affected by error than was the perception of /s/.

The results for /s-/ʃ/ show a slight directionality in that the identification of /ʃ/ is more variable under error as the identification of /s/. However, the asymmetry is overall relatively weak, compared to the asymmetry obtained for /t/ and /k/ in Experiment 1. It has to be considered whether the error status of tongue tip affects the outcome of the perceptual results for /s/ and /ʃ/. Defining an error in terms of intrusion and reduction as done in this paper precludes an analysis of TT for /s/ and /ʃ/ analogous to TB. However, errors on TT can be identified for both, /s/ and /ʃ/, since they differ significantly in TT height (/ʃ/ has a higher TT position than /s/). That /s/ and /ʃ/ are equally affected by categorical TB errors might be due to the fact that for these tokens, also the TT gesture is of errorful magnitude. As has been mentioned before, TT and TB do not behave independently in errors due to the close vicinity of the constriction locations. In contrast to /s/, a gestural specification of /ʃ/ further includes an upper lip (UL) protrusion gesture. For the subject whose kinematic data were used for the perceptual experiment, Goldstein et al. (in prep.) could not reliably measure a difference in UL protrusion for /s/ and /ʃ/ in a way that would allow them to statistically determine errorful UL behavior. Nevertheless it cannot be excluded that the presence/absence and magnitude of a UL gesture will interact with perception. The perceptual results for /s/ and /ʃ/ therefore cannot be interpreted in the same way as they can for /t/ and /k/. For the stops, the experiment shows the perceptual consequences of a co-production of two gestures. For the sibilants, the experimental stimuli are less tightly controlled; the experiment shows the effect of the occurrence of at least one errorful constriction. The results nonetheless allow to come to generalizations about asymmetries, since the lack of independence between TT and TB in Goldstein et al.'s (in prep.) experiment suggests that the occurrence of single-constriction errors between these two vocal organs is extremely rare, if not impossible (due to the close vicinity of the constriction locations).

6 General discussion

In the data presented here, asymmetries have been found in production as well as in perception. However, these asymmetries are different in nature from the ones that have been reported by Stemberger (1991a, 1991b). Goldstein et al. (in prep.) show that both /t/ and /k/ exhibit an intrusion bias; there is no production asymmetry between the two stops. An asymmetry does exist in the form of a gestural addition bias, but this phenomenon affects /t/ and /k/ to the same degree. The present experiments demonstrate that it is in perception that the gestural intrusion bias has different consequences. The perception of coronals is more affected by errors than the perception of velars. The anti-frequency effect in substitution errors that has been observed for /t/ and /k/ for error data recorded by means of transcription might thus be due to the transcriber's perceptual biases. Crucially, our data do not require an appeal to coronal underspecification in order to explain the apparent anti-frequency effects in error distributions.

For /s-ʃ/ on the other hand, the perceptual asymmetry cannot explain the directionality effect found in speech error corpora and experiments. On the basis of the perceptual results alone a directionality effect to be expected would be /ʃ/ → /s/ not */s/ → /ʃ/. Since /ʃ/ is more systematically affected by error than /s/, we would expect /ʃ/ to be more often substituted by /s/. This, however, is the opposite of the asymmetry that has been recorded in speech error research. This suggests that the asymmetries cannot be explained by perceptual biases, they must originate in production. Recall that in production, intrusive errors are more frequent than reductive errors. For the sake of clarity, the hypothesized gestural composition of /s/ and /ʃ/ shall be repeated here schematically:

/ʃ/: TT and TB gesture
/s/: TT gesture only

Errors in production are dominated by the intrusion bias. For /ʃ/, there is no intrusion bias in terms of TT, since both TT and TB are actively controlled for in normal production. Thus the most common error should be an intrusive TB error on /s/. This means that /s/ will systematically be more affected by errors compared to /ʃ/. The data obtained in the production experiment confirm this prediction.

It can be concluded that asymmetries can originate in production where the interacting segments are in a subset relationship to each other. Thus intrusive TB errors only affect /s/, not /ʃ/. Note that this supports a more limited notion of underspecification assumed in a gestural framework: gestures have task specific targets for certain tract variables only, and segments typically correspond to sets of gestures that leave several tract variables unspecified (Browman & Goldstein, 1992). Again, there is no need to assume /s/ to be underspecified for place of articulation; it does lack, however, a tongue body constriction gesture. It is to be expected that the 'palatal bias' reported in Stemberger (1991a), in which /t/ turns more often into the affricate /tʃ/ than vice versa, can be explained on the same basis.

The subset relationship between /s/ and /ʃ/ in terms of their gestural composition is an independently motivated assumption within Articulatory Phonology (Browman and Goldstein, 2001) and does not hold of their standard featural differentiation (+/- anterior). The data presented here can be taken as supporting evidence for their asymmetric gestural composition.

7 Conclusions

The experiments presented in this paper provide evidence for systematic asymmetries in the perception of speech production errors. These asymmetries combine with the nature of production errors to explain the patterns of asymmetries reported in speech error corpora. In production, error patterns are generally dominated by an intrusion bias: it is more likely for an errorful gesture to intrude than it is for a target gesture to be reduced. For different segments, this gestural addition bias produces different consequences. For /t/-/k/, intruding TD gestures during /t/ have a systematic perceptual effect, whereas for /k/ intruding TT gestures do not significantly affect identification or reaction times. For /s-ʃ/, perceptual biases are not the source of distributional asymmetries. Rather, the addition bias translates into a 'phoneme bias'. The most likely error to occur is an intruding TB gesture during /s/; the intrusion bias leaves /ʃ/ unaffected by /s/ since /ʃ/ and /s/ are gesturally in a subset relationship to each other. The study shows that the concept of coronal underspecification is not needed to explain asymmetries in speech errors. The data further support the notion that gestures are units of speech production, since the obtained results can be accounted for by the specific assumptions Articulatory Phonology makes about the gestural composition of segments.

Acknowledgments

Thank you to Larissa Chen for comments and much help in practical matters. Thank you also to Alexei Kochetov, Dani Byrd, Donca Steriade, Ken de Jong, and Stefanie Shattuck-Hufnagel for insightful comments and discussions. Work supported by NIH grant HD-01994.

References

- Boucher, Victor J. (1994): Alphabet-related biases in psycholinguistic enquiries: considerations for direct theories of speech production and perception. *Journal of Phonetics*, 22(1), 1-18.
- Browman, Catherine P., & Goldstein, Louis (1989): Articulatory Gestures as Phonological Units. *Phonology*, 6(2), 201-251.
- Browman, Catherine P., & Goldstein, Louis (1992): Articulatory Phonology: An Overview. *Phonetica: International Journal of Speech Science*, 49, 155-80.
- Browman, Catherine P., & Goldstein, Louis (2001): *Articulatory Phonology*. Manuscript.
- Cutler, Anne (1981): The reliability of speech error data. *Linguistics*, 19, 561-582.
- Dell, Gary (1986): A Spreading-Activation Theory of Retrieval in Sentence Production. *Psychological Review*, 93(3), 283-321.
- Ferber, Rose (1991): Slip of the Tongue or Slip of the Ear? On the Perception and Transcription of Naturalistic Slips of the Tongue. *Journal of Psycholinguistic Research*, 20(2), 105-122.
- Frisch, Stefan, & Wright, Richard (in press). The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue.
- Fromkin, Victoria A. (1971): The Non-Anomalous Nature of Anomalous Utterances. *Language*, 47, 27-52.
- Goldstein, Louis, Pouplier, Marianne, Chen, Larissa, & Byrd, Dani (in preparation). Gestural Action Units Slip in Speech Production Errors.
- Grier, J. Brown (1971): Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin*, 75, 424-429.
- Mowrey, Richard A., & MacKay, Ian R. (1990): Phonological Primitives: Electromyographic Speech Error Evidence. *Journal of the Acoustical Society of America*, 88(3), 1299-1312.
- Shattuck-Hufnagel, Stefanie (1979): Speech Errors as Evidence for a Serial-Ordering Mechanism in Sentence Production. In W.E. Cooper & E.C.T. Walker (Eds.), *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett*. Hillsdale, NJ: Lawrence Erlbaum. 295-342.
- Shattuck-Hufnagel, Stefanie (1983): Sublexical Units and Suprasegmental Structure in Speech Production Planning. In P.F. MacNeilage (Ed.), *The Production of Speech*. New York: Springer. 109-136.
- Shattuck-Hufnagel, Stefanie, & Klatt, Dennis (1979): The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior*, 18, 41-55.
- Stemberger, Joseph, & Stoel-Gammon, Carol (1991): The Underspecification of Coronals: Evidence from Language Acquisition and Performance Errors. In C. Paradis & J.-F. Prunet (Eds.), *The Special Status of Coronals. Internal and External Evidence*. San Diego: Academic Press. 181-199.
- Stemberger, Joseph, & Treiman, Rebecca (1986): The Internal Structure of Word-Initial Consonant Clusters. *Journal of Memory and Language*, 25, 163-180.
- Stemberger, Joseph Paul (1991a): Apparent Anti-frequency Effects in Language Production: The Addition Bias and Phonological Underspecification. *Journal of Memory and Language*, 30, 161-185.
- Stemberger, Joseph Paul (1991b): Radical underspecification in language production. *Phonology*, 8, 73-112.