Syllable, Word, and Phoneme Frequency Effects in Spanish Phonological Speech Errors: The David Effect on the Source of the Error*

Julio Santiago
University of Granada, Spain

Elvira Pérez
University of Liverpool, UK

Alfonso Palma
University of Granada, Spain

Joseph Paul Stemberger
University of British Columbia, Canada

The present study examines the frequency of syllable, word, and phoneme units in a corpus of 1477 Spanish phonological speech errors. Phoneme targets were of equivalent frequency to matched controls, whereas phoneme sources and error phonemes were lower in frequency than chance and than target phonemes (the *David effect*). Error, target, and source syllables were of lower frequency than chance, without differences across them. Target and source words were of lower frequency than chance, and error words were lowest in frequency. Contrary to most current theories, which focus on processing of the target, present results suggest that phonological speech errors arise partly from faulty processing of the source phoneme, in the context of weak syllables and words.

1. Introduction

How are syllables represented in the mental lexicon for production? The answer to this question has remained elusive. Whereas there is little debate on whether

---

* The analyses reported herein were carried out while Julio Santiago was a visiting researcher at UBC, hosted by Joe Stemberger. This research was supported in part by grant BS02002-02993 to Alfonso Palma, by NSERC Individual Research Grant #238561-01 to Joseph Paul Stemberger, by NIDCD grant R01DC006948 to Pat O'Seaghdha, and by a travel fellowship from the Universidad de Granada to Julio Santiago. The authors are indebted to Pat O'Seaghdha for many valuable comments and suggestions, and to the members of the Granada research group on Psycholinguistics.

MIT Working Papers in Linguistics 53, 265–303
The State of the Art in Speech Error Research: Proceedings of the LSA Institute Workshop
Carson T. Schütze and Victor S. Ferreira (Eds.)

words and phonemes are represented in the mental lexicon and are actively selected and inserted into structural frames (syntactic or phonological) for production, the status of syllable representations is still far from settled.

Since the initial analyses of speech errors, a seeming *syllable paradox* was noted (Shattuck-Hufnagel 1979; Dell 1986). On the one hand, phonological speech errors show a strong Syllable Position Constraint: most movement errors involving single phonemes and clusters respect syllable position (Nooteboom 1969; Mackay 1970). Onsets interact with onsets, nuclei with nuclei, and codas with codas, and cross-positional interactions are rare. On the other hand, movement errors involving whole syllables are even rarer. For many years, the received interpretation assumed that syllables are structural frames which guide phoneme and segment insertion for production (Shattuck-Hufnagel 1979), but they are not themselves units for insertion (which we will call "content units," following MacKay 1987).

However, there are strong reasons to suppose that syllable content units should be represented in the mental lexicon. The main argument derives from frequency of occurrence of syllables. Schiller, Meyer, Baayen and Levelt (1996) calculated that there are around 12,000 syllable types in Dutch, the 500 most frequent of which accounted for over 85% of the total syllable tokens in their corpus. Linguistic units which get re-used over and over again with such a high frequency would very plausibly be registered in the lexicon as units in some fashion. English is similar to Dutch in this regard, but the argument gets stronger for languages with clearer syllabification and simpler syllable structures, such as Spanish, on which we focus in the present paper. Based on estimates from our own lexical database (to be described below), Spanish has around 1,900 syllable types, and the 500 most frequent account for over 97% of all syllable tokens. On average, each one of those syllable types is 2.27 times more frequent than Dutch types. Nonetheless, Spanish speech errors also show the syllable paradox (del Viso 1992; García-Albea, del Viso and Igoa 1989). In this context, Mandarin Chinese provides an extreme case. With only about 400 syllable types (not counting tone), Mandarin speech errors show frequent involvement of whole syllables, and phoneme errors respect the Syllable Position Constraint (Chen 2000). The case remains open, therefore, at least for the most frequent syllable types in languages like Spanish, English and Dutch.

The two most popular current models of language production (Dell 1986; Levelt, Roelofs and Meyer 1999) include syllable units in their lexical networks, albeit they are treated in a different fashion. For Dell (1986) and other theories of a similar spirit (MacKay 1987; Stemberger 1985), there are both structural and content syllable units which intervene between word and phoneme strata. Content syllables are inserted into word frames, and they lead to the preparation of syllable frames which in turn guide the insertion of phonemes. In contrast, for Levelt et al. (1999), accessing stored syllables is the last stage in word production. They assume a direct connection between words and phonemes, and a set of syllabification rules which build syllables on the fly. As soon as a syllable is built, it contacts the syllabary, a stratum of (at least the most frequent) phonetic syllables, which are conceived of as pre-packaged articulatory gestures.

1.1 Experimental studies of the syllable as content unit

Several decades of research using experimental tasks have not clarified the question on the psychological nature of the syllable. The main lines of research focus on the effects of syllable priming, number of syllables, and syllable frequency. There is also related research examining these variables in speech perception (see Cutler, McQueen, Norris and Somejuan 2001 for a review) and with printed materials in reading tasks (e.g., Jared and Seidenberg 1990; including some Spanish studies, Carreiras, Álvarez and de Vega 1993), but we will focus here only on those studies using pure production tasks.

Following a logic imported from speech perception research (Mehler, Dommergues, Frauenfelder and Segui 1981), syllable priming studies in language production have examined whether primes sharing exactly the first syllable with a following target word would have a stronger priming effect than primes sharing one less or one more phoneme. Ferrand, Segui and Grainger (1996) had French participants perform a picture naming task with very brief masked visual primes, and found a syllable priming effect. This effect was later extended to English by Ferrand, Segui and Humphreys (1997). However, Schiller (2000) carried out an exact replication of the latter study and could not find the syllable priming effect in English (see also Schiller 1999). Finally, Schiller (1998) showed that there is no syllable priming effect in Dutch picture naming either. So far, the only language in which syllable priming effects are uncontested is Mandarin Chinese (Chen, Li and Ferrand 2003), which is consistent with the high rate of whole-syllable speech errors noted above.

A bit more promising approach was pioneered by Meyer (1990, 1991), who presented the prime implicitly, asking Dutch participants to produce several words which shared the initial segments within a block of trials. Meyer (1991) found the expected syllable effect, but this result proved difficult to replicate. Very recently, Cholin, Schiller and Levelt (2004) have been able to find a syllable priming effect using a version of Meyer's original implicit priming task in Dutch, but we will have to wait to see whether this result stands the test of time. Once again, Mandarin Chinese seems to provide a clear case for the syllable using this methodology (Chen, Chen and Dell 2002).

The number-of-syllables effect (longer production latencies for words with more syllables) constitutes a less direct test of the existence of syllable content units in the lexicon, because any such effect might also be due to the use of syllable frames. Adding to the difficulty of interpretation, this line has produced no less frustrating results. The original reports of a number-of-syllables effect in word production by Sternberg, Monsell, Knoll and Wright (1978) and Klapp, Anderson and Berrian (1973) already posed a puzzle that remains unresolved. Sternberg et al. (1978) found the effect in simple reaction time (RT) tasks, in which the response is fully specified before the signal to respond is presented. The effect did not occur in choice RT tasks, where the signal to respond also identifies the response to be made. It is assumed that simple RT tasks allow the programming of the response to be carried out during the preparation period, such that a fully programmed response is kept on hold in some kind of output buffer until the response signal appears. In this context, number of syllables would affect only the motor output buffer, suggesting no role of syllables in the phonological programming of words. In contrast, Klapp et al. (1973) found the effect in choice RT tasks such as word and picture naming, but failed to find it in simple RT tasks like delayed naming. Adding to the complexity of this issue, there are recent reports of replications of the number-of-syllables effect in English picture naming (Santiago, Palma, MacKay and Rho 2000) together with failures to replicate it both in French and English using otherwise comparable procedures and materials (Bachoud-Levi, Dupoux, Cohen and Mehler 1998). Meyer, Roelofs and Levelt (2003) suggested that the effect appears whenever people strategically set a fast criterion to start responding, although prior conflicting results were obtained in conditions that fostered this strategy to an equal extent. Summing up, the field is still a long way from clarifying the factors that govern the effect of the number of syllables on production latency. Interestingly, one of the clearer reports of a number-of-syllables effect in a production task uses a different logic. Roelofs and Meyer (1998) used the implicit priming task and showed that priming from shared initial segments is obtained only when the words also share total number of syllables. Priming could not be found when CV structure but not number of syllables was shared, which supports the use of abstract syllable frames devoid of content information.

The evidence reviewed so far is sparse, mixed, and often difficult to replicate, and only provides a clear case for the syllable as content unit in the case of Mandarin Chinese.

1.2 Syllable frequency

Looking at syllable frequency effects seems a direct way to explore the independent representation of syllables in lexical memory. The finding of syllable frequency effects is taken as a strong test of the existence of syllable content units: either high frequency syllables may be represented whereas low frequency syllables are not, or all syllables are represented but their speed of processing varies. Most current theories of phonological encoding predict effects of syllable frequency: words with high frequency syllables should be produced faster and more accurately than words with low frequency syllables.

Once again, theoretical agreement has not been paired with clear and homogeneous data. The seminal study on syllable frequency in production tasks (Levelt and Wheeldon 1994) showed an effect of the frequency of the last syllable of disyllabic Dutch nouns on latencies in a symbol-naming task. However, their materials confounded syllable and phoneme frequency, and the original effect could not be replicated with more controlled materials (Hendricks and McQueen 1996). Recently, Cholin, Levelt and Schiller (2006) avoided confounds in the materials by resorting to nonwords, and found the expected syllable frequency effect, which this time depended crucially on the frequency of the first syllable. It remains to be seen whether this effect can be extended to words.

A handful of recent studies have examined the influence of syllable frequency on error rates in the speech of special populations. Wilshire and Nespoulous (2003) assessed the reading and repetition of words matched in word frequency and differing in final syllable frequency by two French aphasics with a phonological impairment. Even though the patients produced many errors, syllable frequency made no difference. A possibility suggested by the results of Cholin et al. (2006) is that first-syllable frequency makes a more important contribution than second-syllable frequency. In agreement with this suggestion, Aichert and Ziegler (2004) found an effect of first-syllable frequency in word repetition by German patients with apraxia of speech. Patients tended to produce more errors on low-frequency than on high-frequency syllables, although the effect became significant only when an extreme frequency contrast was used. Stenneken, Hofmann and Jacobs (2005) looked at the neologistic speech of a German jargon aphasic and found it skewed towards high-frequency syllables. In a very detailed study of the speech of Spanish, Italian and French aphasics with phonological impairment, Laganaro (2005) focused on single phoneme substitutions ($N = 304$) and compared the frequency of the target and error syllables, as well as the frequency of the target and error phoneme, in the reading and repetition of a list of words and nonwords matched for syllable length and structure, but varying freely in syllable frequency. In three out of seven subjects, the frequency of the error syllable was higher than the target syllable, and in only one of those was this effect paired with a similar effect of phoneme frequency. The effect of syllable frequency was clearer in nonword than word production. They then looked at the performance of three aphasics in a nonword repetition task where syllable frequency was explicitly manipulated, and found more errors on low-frequency than on high-frequency syllables. Finally, a French-speaking aphasic reading and repeating a list of monosyllabic words also produced error syllables of higher frequency than target syllables, while word frequency showed no effect. In contrast, Nickels and Howard (2004) looked at syllable frequency together with several indices such as number of phonemes, number of syllables, syllable structure, and word frequency, and found no effect of syllable frequency once the effect of the other indices was removed by means of logistic regression.

To sum up, contemporary research on the syllable frequency effect is far from providing clear and complete support for the psychological reality of syllable units. Studies using latency measures have obtained the effect only for nonwords. Studies looking at phonological errors provide clearer results, although the picture is still far from complete. Unfortunately, this research suffers from significant limitations. First of all, the studies available so far only look at the speech of a few patients affected by different speech and language pathologies. It is unclear whether the results can be generalized to normal subjects, or even to other patients with the same conditions, since the observed patterns are quite variable. Secondly, there are hints in the literature that the type of syllable frequency measure that is used may play an important role. Most studies use total token frequency, but position of the syllable in the word seems to be important. Thirdly, most studies only look at the difference in frequency between the target and error syllables, or the target compared to chance level. We will argue below that in order to fully explore the theoretical predictions of current models, it is necessary to include error, target, and (whenever possible) source syllables in the analysis, comparing them to each other as well as to a suitable chance baseline. Finally, syllable frequency (however it is measured) is correlated with both word and phoneme frequency. Effects of both word frequency and phoneme frequency on error rates have been reported in the literature (see below), so syllable frequency effects need to be carefully disentangled from putative effects of these other variables. Other factors, such as syllable length and structural complexity also need to be taken into account. Studies that carefully control all of these factors are very scarce.

1.3 Aims of the present study

The present study is aimed at evaluating the psychological nature of syllable units by looking at syllable frequency effects in a large naturalistic data base: a corpus of over 8000 spontaneous speech errors produced by neurologically healthy Spanish adults which has been collected at the University of Granada, Spain. Looking at naturalistic speech errors will improve the generalizability of our results. The fact that Spanish provides a clearer case for the storage of syllable units increases the chances of finding the expected effects. In this context, we carry out a systematic comparison of error, target, and source syllables and suitable baseline conditions. We also explore a range of different frequency indices, in order to assess which are the most representative measures. Finally, we include measures of word and phoneme frequency, and try to disentangle their effects by means of statistical techniques.

1.4 Theoretical context

It is important to note that any effect of syllable frequency, of any kind, would be important evidence for the existence of syllable units in the mental lexicon. However, far more precise predictions can be drawn from current theories. An example will help to delineate the main characteristics of a phonological speech error and the expectations about frequency from different theoretical points of view. Let's take the following error in which a single phoneme is anticipated:

(1) Me ba**ch**o en he**ch**os. ← Me baso en hechos.
'I rely on facts.'

In this movement error, the phoneme /č/ (spelled *ch*) is anticipated from the source syllable *chos* in the word *hechos* to substitute for the phoneme /s/ (spelled *s*) in the target syllable *so* in the word *baso*. As a result, the error syllable *cho* is produced.

Figure 1 shows schematically the relevant mini-network of Dell's (1986) model. In terms of this model, any factor increasing the level of activation of the intruding phoneme /č/, or decreasing the activation of the target phoneme /s/, at the moment at which it is to be selected for frame insertion will increase the probability of /č/ being anticipated to take the position that /s/ should occupy. This localist view is based on the idea that speech errors arise from interactions limited to the target and source phonemes, an idea shared by most current models of phonological encoding. The causes of the error are the relative weakness of the target in the presence of a stronger competitor. Target processing is thus the key factor in the causation of a speech error.

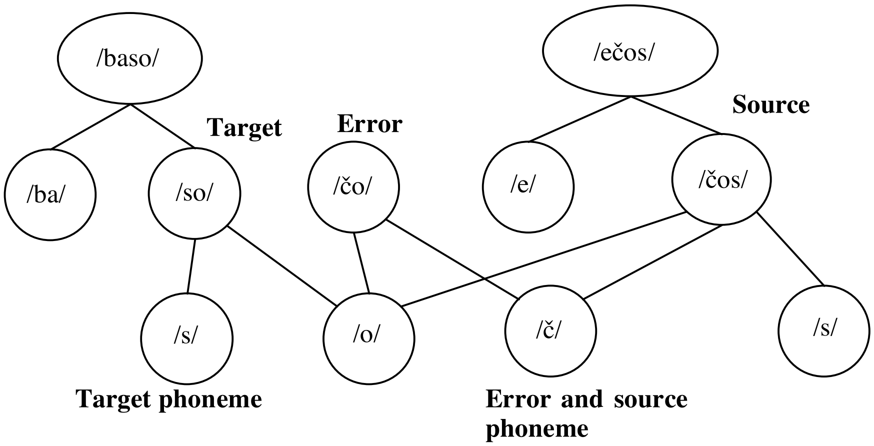

Figure 1
The relevant mini-network from Dell's (1986) model for the anticipation
*Me ba**ch**o en he**ch**os* ← *Me ba**s**o en he**ch**os*

In this context, one cause of relative strength or weakness is the frequency of the syllable node to which a phoneme is connected. Thus, a first prediction from Dell's (1986) theory is that the target syllable should be of lower frequency than chance (that is, a weak syllable), because the concomitant low activation level should predispose all parts of the syllable to error. A second prediction, which has not received any attention so far in the literature, is that the frequency of the source syllable should also be higher than that of the target syllable. Thirdly, the frequency of the error syllable outcome should be higher than the target syllable. It would also be consistent with the model, although not a necessary prediction, for the source syllable to be of higher frequency than chance (a strong syllable) and, finally, for the resulting error syllable to be at least as frequent as expected by chance.

Levelt et al.'s (1999) model shares with Dell's and other connectionist models the localist view on the generation of speech errors. Nevertheless, the placement of the syllable stratum at the very end of the production process, together with the lack of bidirectional connections, has important theoretical consequences. In their view, all speech errors are actually whole-syllable errors. The model always generates the correct sequence of phonemes at the phoneme stratum, which is then syllabified by rule. Resulting syllable addresses are then used to access the syllabary. Errors arise when a wrong phonetic syllable is accessed. Syllabary access is a competitive process, so higher-frequency syllables should be accessed more efficiently than lower-frequency syllables.

Levelt et al.'s view places the origin of speech errors in the competition between the target and the error syllables (contrast this with the connectionist view, in which the error arises as a result of competition between the target and source phonemes). Two predictions thus follow: the target syllable should be weaker (lower frequency) than the error syllable, and the strength of the source syllable is irrelevant (with frequency, on average, equal to chance). Again, an overall weak target syllable (lower frequency than chance) is also expected. It would also be maximally consistent with the model, although it does not necessarily follow from its premises, if the error syllable is of higher frequency than chance.

To widen our theoretical background, let's take a look at two theories which use very different assumptions from connectionist and syllabary models. The first is the PDP sequence-learning model of Dell, Juliano and Govindjee (1993), and the second is the time-varying signal model by Vousden, Brown and Harley (2000). Neither theory uses explicit representations for content syllables. However, both of them succeed to a certain extent in generating speech errors. Dell et al. (1993) simulate word production as a kind of pure sequence learning, where prior phonemic context serves as a cue to produce the next phoneme in the sequence. The model adjusts connection weights by means of a learning algorithm. In this model, a low frequency syllable is a relatively unpracticed sequence. Therefore, production is likely to derail, resulting in a more practiced, higher-frequency sequence (syllable). The model thus predicts that target syllable frequency should be lower than chance, that error syllable frequency should be higher than target syllable frequency, and that error syllable frequency should be higher than source syllable frequency (otherwise, the error syllable would tend to be the same as the source, generating a whole syllable movement error).

Finally, Vousden et al. (2000) use a time-varying signal to implement a syllable frame to which phonemes are associated. Because phoneme sequences are not represented in any way (either as explicit syllable units or as prior phonemic context), this model predicts that the frequency of all error, target and source syllables should be similar and they should not differ from chance.

Summing up, only one theory (Vousden et al. 2000) predicts the absolute absence of any kind of syllable frequency effects. All other theories agree on the prediction that, generally, errors should occur on weak low-frequency syllables, and that target syllables should be of lower frequency than error syllables. Connectionist theories *à la* Dell (1986) add that source syllables should also be of higher frequency than target syllables. Finally, few theories make clear predictions on how error and source syllable frequencies should differ from chance, but the expectation that both error and source syllables should be higher than chance is compatible with most of them.

At this point, it is important to note that the pattern of predictions drawn from each theory for syllable frequency closely matches those for word and phoneme frequency, with the single exception of the Vousden et al. (2000) model, which predicts effects of both word and phoneme frequency but not of syllable frequency.

1.5 Word and phoneme frequency effects on error rates

Although orthogonal to the question of syllable representation, the design of the present study will let us assess the influence of word and phoneme frequency effects on Spanish phonological speech errors.

Word frequency effects on speech errors have been repeatedly reported in the literature, although the pattern is also far from clear. Most reports agree that there are more phonological errors on low-frequency than on high-frequency words, for both healthy adults (Stemberger and MacWhinney 1986; Dell 1990) and neurological patients (Nickels 1995; but see Nickels and Howard 1995). Studies looking at word substitutions also report higher error rates on low frequency items in the case of phonologically related substitutions (malapropisms), although these effects are not observed when the interacting words hold a semantic relation (spoonerisms) (Stemberger 1984; Harley and MacAndrew 2001; del Viso, Igoa and García-Albea 1991; Hotopf 1980). This is compatible with the lexical frequency effect being located at the point when phonological word forms are accessed (Jescheniak and Levelt 1994).

However, the picture is complicated because available data provide mixed support for related theoretical predictions. Harley and MacAndrew (2001) failed to find the expected frequency difference between the target and the intruding (error) word in their collection of malapropisms. Looking at the contrast between error word and chance expectations in phonological errors resulting in real words, Dell and Reich (1981) found a small, non-significant trend toward errors having higher frequency than chance, an effect that was later found more clearly by Gagnon, Schwartz, Martin, Dell and Saffran (1997) with aphasic errors. In contrast, both Garrett (1988) and del Viso et al. (1991) found a trend in the opposite direction, with error words being of lower frequency than chance. A final puzzle is posed by the fact that both semantic and phonological word substitutions show a very clear and strong positive correlation between the frequency of the target and error words, such that words in the same frequency range tend to interact (Garrett 2001). No current model seems to be suited to account for this pattern of observations.

Phoneme frequency effects provide an easier picture to understand in terms of current models, as long as the effects of phonological defaults are taken into account. Following the initial report by Shattuck-Hufnagel and Klatt (1979) of at best very small frequency differences between target and error in single-phoneme speech errors paired with several exceptions in the opposite direction, Stemberger (1991a) showed how these effects are the result of two independent influences. First, there is a Nondefault Bias, a tendency for segments with default features to take on nondefault features from surrounding segments. Segments with default features are generally of higher frequency than similar segments with nondefault features, so the Nondefault Bias is the cause of apparent anti-frequency effects: lower frequency segments replacing higher frequency targets. Second, when the Nondefault Bias is controlled for, equally specified segments show the expected phoneme frequency effect: higher frequency sources replace lower frequency targets. The Nondefault Bias was also observed for Spanish speech errors by del Viso (1992). A preponderance of high frequency phonemes intruding as errors was also observed in the neologisms of jargon aphasics by Robson, Pring, Marshall and Chiat (2003) and by Stenneken et al. (2005). Interestingly, when error phonemes came from the target word, they were often low frequency consonants (Robson et al. 2003), suggesting perhaps a re-cycling of already-selected phonemes in the production of neologisms. Finally, Stemberger (1983, 2002; Stemberger and Middleton 2003) reported phoneme frequency effects in the production of inflected verb forms. When the Nondefault Bias is not relevant, overtensing errors (e.g., *didn't sang* instead of *didn't sing*) are more likely when the vowel of the irregular past-tense form (which is present in the error) is of higher frequency than the vowel of the base form, whereas overregularization errors (e.g. *singed* instead of *sang*) arise more often when the vowel of the base form (which is present in the error) is of higher frequency than the vowel of the past-tense form.

To summarize, there is some suggestive evidence for both word and phoneme frequency effects in the analysis of error rates. Phonological errors and malapropisms tend to occur on low frequency words, and phoneme errors are more frequent than targets as long as they are both specified to the same extent. However, many of the same caveats and limitations discussed for the syllable frequency studies are applicable to this literature. Many effects are difficult to replicate, and some important expected effects (like the higher frequency of word errors versus word targets in malapropisms) are at best very small and often not found at all. Many studies do not control for frequency variables at all three levels (word, syllable and phoneme). Although lexical frequency is simplex (involving counts of word tokens in a corpus), phoneme frequency can be indexed by both type and token measures taking into account position within the syllable (or not), but only absolute token frequency has been used so far. Finally, there are no studies exploring the whole range of theoretically relevant characteristics at word and phoneme levels, which includes the error, the target and (when possible) the source, plus their associated chance estimates.

1.6 Outline of the present investigation

We carry out a broad exploration of frequency effects of word, syllable and phoneme units on Spanish phonological speech errors, including error, target and source, and suitable chance conditions, using a variety of different frequency indices. We will divide this effort into two main steps. Study 1 will assess each relevant pair-wise comparison of error, target, source and their respective control conditions at syllable, phoneme, and word levels over a wide variety of frequency measures, using non-parametric statistical techniques. In Study 2, we will first determine which are the most representative frequency measures, and then submit them to parametric analyses of covariance, in search of main effects of word, syllable and phoneme frequency while discounting cross-influences due to internal correlations.

2. Methods

2.1 The corpus

The source of our data is the corpus of naturalistic Spanish speech errors collected over several years by the students in the Psychology of Language courses at the University of Granada, Spain, under the supervision of one of the authors (Palma). The current size of the Granada corpus is 8031 errors, contributed in varying amounts by a total of 737 students. A detailed description of the training and collection methods and several control procedures can be found in Pérez, Santiago, Palma and O'Seaghdha (2006). The corpus continues growing as new errors are added every year. The present investigation uses all phonological errors collected up to 2001 ($N = 1477$).

The Granada corpus is a multiple-collector effort, drawing on a large number of theoretically naive observers. It might be argued that this method of error collection is more susceptible to perceptual biases than the more common single-collector procedure, in which a small number of highly trained theoretically-informed observers provide most of the errors in the corpus. Pérez et al. (2006) tested this idea by comparing the Granada corpus to the other available Spanish speech error corpus: the single-collector Oviedo corpus of del Viso (1992). A comparison of the two corpora revealed none of the predicted differences if multiple-collector corpora magnify the effects of perceptual bias. We are thus confident of the usefulness of the Granada corpus.

Ninety-three of the total 1477 phonological errors were discarded because they changed the total number of syllables of the target word, a requirement imposed by the way that our syllable frequency measures are calculated (see below). The breakdown into error types of the remaining 1384 errors is shown in Table 1.

Table 1
Phonological error types in the corpus and their frequency

Error type	*Frequency*	
Contextual		
Exchanges (substitutions)		246
Anticipatory side	246	
Perseveratory side	246	
Anticipatory substitutions		414
Unambiguous source	251	
Ambiguous source	163	
Perseveratory substitutions		350
Unambiguous source	145	
Ambiguous source	205	
Shifts (additions)		254
Without loss of source		200
Unambiguous source		143
Anticipatory	82	
Perseveratory	61	
Ambiguous source	57	
With loss of source		54
Target syllable		54
Anticipatory	22	
Perseveratory	32	
Source syllable		54
Anticipatory	22	
Perseveratory	32	
Noncontextual		
Additions	16	
Substitutions	62	
Omissions	42	

Several comments are in order. First, note that both halves of exchange errors are treated independently. Also note that the target and source syllables of shifts with loss of the source element are also entered independently in the analysis, because an error is produced also on the source syllable (due to the omission of the moving unit). Second, some peculiarities of this taxonomy are designed to accommodate two important factors in the analyses to follow. One such factor is whether the source unit can be unambiguously located. Exchanges and shifts with loss of the source unit always have unambiguous sources, as in the following examples:

(2) Que me des el perió**k**i**d**o ← Que me des el perió**d**i**c**o
'Give me the newspaper'

(3) ¿De p**l**ubicidad? ← ¿De pub**l**icidad?
'Of advertising?'

Anticipatory substitutions, perseveratory substitutions and shifts without loss of the source unit usually have an unambiguous source, but the source is ambiguous if there are two or more candidates in the surrounding context. Example (1) above is an unambiguous anticipatory substitution. The following is an example of the ambiguous type, with a possible source at the beginning of the same word and another possible source three words later:

(4) ¿Quién va a ser el pilo**p**o, el tío Pepe o tú?
← ¿Quién va a ser el pilo**t**o, el tío Pepe o tú?
'Who is going to be the pilot, uncle Pepe or you?'

Another important factor is the number of phonemes in the syllable, because of the strong negative correlation between syllable length and syllable frequency. For substitutions (anticipatory, perseveratory, or noncontextual), the error syllable is the same length as the target syllable; in exchange errors, the source syllable is also subject to error, but its associated error syllable is still the same length as the source syllable. For shifts (additions), the error syllable is longer than the target syllable; for shifts without loss of the source there is no error on the source syllable, but for shifts with loss of the source there is an error on the source syllable (omission), and that error syllable is shorter than the source syllable. Both the ambiguity of the source and the length of interacting syllables will be controlled for by selecting only the appropriate cases for a given contrast.

Finally, the same approach will be used with respect to other factors that may be relevant to particular comparisons at the word or phoneme levels. For example, not all errors can be used to look at the lexical frequency of the error word, as phonological errors often result in nonwords. In the same vein, many phonological errors involve more than a single phoneme, therefore precluding the calculation of phoneme frequency. Suitable subsets of the errors will be selected depending on the hypothesis being tested, as will be made clear when relevant.

2.2 The Granada Lexical Database

Both the frequency measures and the procedures for chance estimation rely on the use of a lexicographic tool named the Granada Lexical Database (GRLDB), developed from the Alameda and Cuetos (1995) lexical database and frequency count. Their lexical corpus contains 81,313 word types (orthographic lexemes) from a sample of two million word tokens. The index of lexical frequency is therefore lexeme (not lemma) frequency.

Words in the database were phonologically transcribed and syllabified. The following information was added to each lexical entry: number of syllables, primary stress location, CV structure, number of lexical neighbours, and average frequency of lexical neighbours. Several syllable fields were also added. The phonological syllables of each word are placed into a positional frame of 10 positions, from “1” to “last.” For example, a bisyllabic word has entries in the first and last syllabic fields. Monosyllables fall into the last position. Each syllable also has a field indicating its stress level (0 for unstressed, 1 for stressed) and another with its CV structure.

2.3 Frequency measures and other relevant indexes

Only one index of lexical frequency was used, the orthographic lexeme count provided by Alameda and Cuetos (1995). Twelve different indexes were computed for syllable frequency, resulting from the combination of the following dimensions.

1. *Type versus token frequency*: Type frequency of a syllable counts how many different words possess that syllable. Token frequency results from adding up the lexical frequencies of those words.
2. *Positional frequency in an absolute word frame*: This is a common frame for all words, distinguishing positions first to ninth and last.
3. *Positional frequency in a relative frame*: This measure uses different positional frames for each word length in syllables.
4. *Non-stress versus stress-sensitive*: Stress-sensitive measures treat the different stress values of the same syllable as different syllables.

Table 2 presents an example.

Table 2
Frequency indexes for the syllable /ga/
in the bisyllabic word *gato* (/'.ga.to./) ‘cat’

Frequency	*Category*
776	Tokens (total number of occurrences) as stressed first syllable in bisyllabic words
35	Types (total number of occurrences in different words) in bisyllabic words as stressed first syllable
776	Tokens as stressed first syllable
35	Types as stressed first syllable
956	Tokens as first syllable in bisyllabic words
48	Types as first syllable in bisyllabic words
2134	Tokens as first syllable
290	Types as first syllable
5382	Tokens as stressed syllable
796	Types as stressed syllable
12827	Absolute tokens
2012	Absolute types

In those errors involving only one phoneme, four frequency indexes of phoneme frequency were computed, resulting from crossing two dimensions:

1. *Type versus token frequency*: Phoneme type frequency counts occurrences in different syllables. Token frequency adds up the absolute token frequency of those syllables.
2. *Positional frequency in an absolute syllable frame*: This is a common frame for all syllables with 2 onset positions, 3 vowel nucleus positions, and 2 coda positions.

An example is shown in Table 3.

Table 3
Frequency indexes for the phoneme /g/ in the word *gato*

Frequency	*Category*
76349	Tokens in first onset position
7910	Types (occurrences in different syllables) in first onset position
79356	Absolute tokens
8387	Absolute types

In order to be able to assess the presence of a Nondefault Bias in the set of single phoneme errors, we used the underspecified feature matrices shown in Tables 4 and 5, based on tables for English in Bernhardt and Stemberger (1998) and adapted for the facts of Spanish. Note that “+” and “–” are used for binary features, and "/" denotes the presence of a privative (single-valued) feature or organizing node.

Table 4
Underspecified consonant feature matrix for adult Spanish
(i.e., with no default or redundant features listed)

	Phoneme																	
Feature	p	t	k	b	d	g	č	ǰ	f	θ	s	x	m	n	ŋ	l	r	ɾ
sonorant													+	+	+	+	+	+
consonantal																		
continuant									+	+	+	+					+	+
nasal													+	+	+			
lateral																+		
tense																		+
Laryngeal	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
voiced				+	+	+		+										
spread-glottis																		
Place	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
Labial	/			/					/				/					
round																		
labiodental																		
Coronal							/	/			/				/			
anterior							–	–							–			
grooved											+							
Dorsal			/			/						/						
back																		
high																		
low																		

Table 5
Underspecified vowel feature matrix for adult Spanish
(i.e., with no default or redundant features listed)

	Phoneme				
Feature	i	E	u	o	a
sonorant					
consonantal	–	–	–	–	–
continuant					
nasal					
lateral					
tense					
Laryngeal	/	/	/	/	/
voiced					
spread-glottis					
Place	/	/	/	/	/
Labial					
round					
labiodental					
Coronal					
anterior					
distributed					
grooved					
Dorsal	/		/	/	/
back			+	+	+
high	+		+		
low					+

2.4 Estimation of chance

Our approach to chance estimation differs from prior methods (see Stemberger 1991b for a review). Working from the GRLDB, we were able to randomly select a suitable control unit for each relevant (error, target or source) word, syllable and phoneme. Word controls were matched with the relevant word in number of syllables and stress position. Syllable controls were matched with the relevant syllable in CV structure, stress level, and position in a frame relative to the word's number of syllables. Phoneme controls were matched with the relevant phoneme in syllable position in an absolute frame with 2 onset positions, 3 vowel nucleus positions, and 2 coda positions. Within the set of matched controls for each relevant unit, random extraction took into account a priori probabilities based on the frequency of the candidates in the most specific frequency measures. Table 6 shows a complete record from our data file, corresponding to the vowel anticipation in (5).

(5) Siempre está ab**u**erta la p**u**erta. ← Siempre está ab**i**erta la p**u**erta.
'The door is always open.'

Table 6
Sample record from the data file for the slip *Siempre está ab**u**erta la p**u**erta*

Condition	*Word*	*Syllable*	*Phoneme*
Error	abuerta	/buer/	/u/
Control error	confirma	/gien/	/e/
Target	abierta	/bier/	/i/
Control Target	hacerlo	/θien/	/a/
Source	puerta	/puer/	/u/
Control source	poco	/kuan/	/o/

All indexes were computed for each relevant unit and its matched control, and then submitted to statistical analyses.

3. Study 1: Frequency effects at syllable, phoneme and word levels

The main goal of Study 1 is to provide an exploration of syllable frequency effects across the categories of error, target and source and their control conditions. Other goals are, first, to replicate several well-known speech error phenomena, and second, to explore whether word and phoneme frequency effects can also be found in the present corpus. Overall, we will be using non-parametric statistics (Sign test) to test our hypotheses without worries related to the specific shape of data distributions. Study 2 will build upon this study by choosing the most representative measure of syllable and phoneme frequency and carrying out parametric covariance analyses that will let us disentangle observed effects from the existing cross-correlations across the relevant factors.

3.1 Design

From the corpus described in Table 1, three main datasets were prepared. Firstly, the Error-Target (ET) dataset contained all cases suited for the contrasts between error and target syllables and between error and target controls. This amounts to all errors in which the length of the target syllable remains unchanged: exchanges (both halves, N = 492), anticipatory substitutions (414), perseveratory substitutions (350) and noncontextual substitutions (62). Total N in this dataset is 1318.

Secondly, the Error-Target-Control (ETC) dataset contained all cases suited for the contrast between the error syllable and its control, and between the target syllable and its control. This amounts to the total available data, because control syllables are matched in syllable CV structure and length with key syllables: exchanges (492), anticipatory substitutions (414), perseveratory substitutions (350), noncontextual substitutions (62), shifts of all types (308), additions (16), and omissions (42). The total N in this dataset is 1684.

Finally, the Target-Source-Control (TSC) dataset includes all errors of unambiguous source: unambiguous anticipatory substitutions (251), unambiguous perseveratory substitutions (145), exchanges (492), unambiguous shifts without loss of the source unit (143), and the target syllable of shifts with loss of the source unit (54). Total *N* equals 1085 in this dataset. Contrasts between the target syllable and its control and between the source syllable and its control utilize the whole dataset. For comparisons between target and source syllables, and between target control and source control syllables, only errors with target and source syllables of equal length will be selected.

3.2 Replicating basic findings

Pérez et al. (2006) used the present corpus to replicate several prior findings in the speech error literature. Some of these findings agree with those reported for Germanic languages such as English and Dutch, whereas others reflect crosslinguistic differences which have been reported only for Spanish (del Viso 1992; García-Albea et al. 1989; Berg 1991; see Pérez et al. 2006 for detailed references). To summarize the results at the phonological level, the universal patterns observed in our corpus are a greater incidence of phonological errors between words than within words, and on consonants than on vowels. The Spanish-specific patterns include protection of word-initial position against errors; an equal propensity for errors to occur on stressed and unstressed syllables; and an absence of a lexical bias.

Other well-known effects on phonological speech errors concern structural similarities between target and source words and syllables (MacKay 1970; Stemberger 1990). Using TSC data (all of the errors in which the source could be unambiguously located), there were clearly significant ($p < 0.05$) positive correlations between the number of syllables of target and source words ($r = 0.71$), the position of the target and source syllables in their respective words ($r = 0.17$), and the stress level of target and source syllables ($r = 0.53$). However, Pérez et al. (2006) suggested (with Berg (1991)) that the effect of stress level is an artefact of word position, and that stress does not play a constraining role in Spanish speech errors.

The subset of all single phoneme errors within TSC data was selected to test the well-known effects of phoneme similarity: interacting phonemes in an error are more similar in features than would be expected by chance (MacKay 1970). Two indexes of similarity were used: number of shared features in fully specified matrices and in underspecified matrices. Sign test comparisons let us conclude that the similarity in both indexes between source and target phoneme in the present corpus is greater than the similarity between source phoneme and target control phoneme (fully specified: non-ties = 681, $z = 3.98$, $p < 0.001$; underspecified: non-ties: 403, $z = 2.49$, $p < 0.02$), and that it is also greater than the similarity between the target phoneme and the source control phoneme (fully specified: non-ties = 701, $z = 1.88$, $p = 0.058$; underspecified: non-ties = 382, $z = 3.32$, $p < 0.001$).

Finally, we used this same data set to check the existence of a Nondefault Bias. As an index of the level of feature specification, the number of specified features in the underspecified feature matrices of Tables 4 and 5 was counted for each relevant phoneme (target, control target, source, and control source). Source phonemes showed greater levels of feature specification than target phonemes (non-ties: 558, $z = 2.07$, $p < 0.05$). Moreover, sources were more specified than their matched controls (non-ties: 593, $z = 3.12$, $p < 0.001$), whereas targets were not less specified than their controls (non-ties: 602, $z = 1.18$, $p = 0.23$). Thus, there is a Nondefault Bias in our corpus (also replicating prior findings of del Viso 1992, in the Oviedo corpus), and this bias seems to arise because source phonemes are more highly specified than chance, and not because targets are less specified than chance.

3.3 Syllable frequency effects

The first set of contrasts concerns the target syllable versus its control, and the error syllable versus its control. Over the ETC dataset (i.e., the whole phonological error corpus), targets were of lower frequency than matched control syllables on all twelve measures of syllable frequency (non-ties range = 1552–1613, z range = 3.86–12.64, all $p < 0.001$). Interestingly, error syllables were also of lower frequency than matched controls (non-ties range = 1599–1647, z range = 7.45–19.28, all $p < 0.001$). Therefore, the phonological error does not create a syllable with a frequency level that reaches the level expected by chance.

Using the ET dataset, we now turned to the comparisons between error and target syllables. To our surprise, error syllables were of *lower* frequency than targets (non-ties range = 1274–1310, z range = 3.88–6.12, all $p < 0.001$). As a control, we ran a comparison between the frequency of the control syllables for target and error, which was non-significant on all twelve measures (non-ties range = 1209–1245, z range = 0.05–1.26, all $p > 0.20$). Against most predictions, phonological errors were creating rarer syllables than the syllable targets. They even managed to create 20 syllable "neologisms" (1.51%), completely unattested in the two million word corpus of Alameda and Cuetos (1995). This went up to 89 syllables (6.75%) unattested with the same stress value and in the same position within words of the same number of syllables. It should be kept in mind that in the ET dataset, no error changed the number of phonemes of the target syllable, so these syllabic neologisms are not the result of additions or omissions of phonemes resulting in phonotactically illegal sequences.

We now turn to comparisons involving the source syllable, using the TSC dataset (errors of unambiguous source). Source syllables were systematically of lower frequency than their controls (non-ties range = 985–1029, z range = 5.33–10.88, all $p < 0.001$). Source syllables, therefore, are not merely "normal" syllables and are not stronger than expected; source phonemes come from weak syllables. Selecting the subset of errors with target and source syllables of equal length, targets were clearly of higher frequency than sources on all but one frequency measure. (The only marginally significant probability level corresponded to positional type-frequency on a word-relative frame: non-ties = 581, $z = 1.74$, $p = 0.08$. All other measures: non-ties range = 576–588, z range = 2.18–3.34, all $p < 0.03$.) The comparison between control target and control source syllables was non-significant in all but one frequency measure. (The only significant measure was the positional type-frequency in an absolute frame: non-ties = 591, $z = 2.46$, $p < 0.02$. Type frequency in a word-relative frame was marginally significant: non-ties = 590, $z = 1.85$, $p = 0.06$. All other measures: non-ties range = 582–595, z range = –0.04–1.44, all $p > 0.14$). Finally, we used this dataset to run a direct comparison of error and source syllable frequency. Error syllables were of equal frequency than source syllables in all but two frequency measures, in which they were of lower frequency. (Significant measures were stress-sensitive positional token frequency on a word-relative frame, non-ties = 474, $z = 2.06$, $p < 0.05$; and stress-sensitive token frequency, non-ties = 421, $z = 2.63$, $p < 0.01$. All other measures: non-ties range = 349–474, z range = 0.04–1.51, all $p > 0.12$). Therefore, syllables created by single phoneme errors were not consistently less frequent than source syllables. Instead, on most measures, they were of equal frequency.

Summarizing, we replicated the previously reported result that errors tend to occur on low-frequency target syllables. However, our data uncovered several surprising patterns. First, all error, target, and source syllables are consistently of lower frequency than expected by chance. This will hereafter be called a Weak Units Effect. Second, target syllables are of higher frequency than source syllables. Error syllables created as a result of the error are of lower frequency than targets. By analogy with the biblical story of little David defeating giant Goliath, we will call this the *David Effect*.

Present results pose several interesting puzzles, and open many questions. The first one to pursue is: are these effects limited to syllable frequency? Can we find parallel effects at phoneme and word levels? To answer this question, we carried out similar analyses at those levels.

3.4 Phoneme frequency effects

We looked at phoneme frequency effects by selecting only those errors where a single phoneme is involved. For syllables, the ETC dataset was used to carry out the comparisons between the target phoneme and its control, and the error phoneme and its control. The comparisons of error versus target phoneme, and control error phonemes versus control target phonemes also used this dataset, because this comparison does not require target and error syllables having the same length. To control for the Nondefault Bias, we selected cases where both phonemes in the contrast had the same level of feature specification.

The Weak Units Effect previously reported in the literature was not replicated in the relevant data set in any of four frequency measures ($N = 268$, non-ties range = 268–268, z range = 0.18–0.67, all $p > 0.50$): target and source phonemes were as frequent as their controls. In contrast, errors were of lower frequency than their controls ($N = 248$, non-ties range = 248–248, z range = 3.74–4.00, all $p < 0.001$). Errors were also of lower frequency than targets in all but one of the frequency measures, which approached significance (301 cases; the non-significant measure was the syllable-relative positional token frequency: non-ties = 301, $z = 1.61$, $p = 0.10$; all other measures: non-ties range = 301–301, z range = 2.19–2.76, all $p < 0.05$). Unexpectedly, the check comparison between control error and control target phoneme indicated that the former was more frequent than the latter in the two non-positional measures ($N = 371$; absolute type and token measures: non-ties range: 229–229, z range = 2.24–2.24, both $p < 0.05$; syllable-relative positional measures: non-ties range = 229–229, z range = 1.18–1.45, both $p > 0.14$). We decided to explore this comparison also in the following dataset.

We now turn to comparisons between the target and the source phoneme, using the single-phoneme errors in the TSC dataset (errors with unambiguous source; note that the source and error phoneme are the same). Once again, we controlled for the Nondefault Bias by selecting only cases matched in number of specified features. Source phonemes were of significantly lower frequency than targets in three out of four frequency measures, and marginally in one. ($N = 232$; the marginal measure was syllable-relative positional token frequency: non-ties = 232, $z = 1.77$, $p = 0.07$; all other measures: non-ties range = 232–232, z range = 2.03–2.82, all $p < 0.05$). Therefore, the David Effect is substantiated also at the level of phoneme units. Source phonemes were also significantly less frequent than their controls (non-ties range = 196–196, z range = 2.78–3.21, all $p < 0.01$). Because of the theoretical importance of the contrast between target phoneme and matched controls, we repeated the analysis in this dataset, which is a subset of the previous one. The frequency difference was again non-significant ($N = 188$; non-ties range = 188–188, z range = –0.07–065, all $p > 0.52$). The check comparison between control target and control source phonemes was non-significant in all four measures ($N = 175$, non-ties range = 174–175, z range = 0.15–0.53, all $p > 0.59$). We also checked whether the source phoneme was still less frequent than the corresponding controls when it surfaces as the error phoneme, and this was indeed the case (171 cases, non-ties range = 171–171, z range = 4.12–4.28, all $p < 0.001$).

Finally, the repetition of the check comparison between control error and control target again suggested that control error phonemes have a higher frequency than control target phonemes, significantly in three measures and marginally in one (154 cases, non-ties range = 154–154, z range = 1.85–2.49, all $p < 0.07$). The causes of this result are unclear, as control phonemes are randomly selected from the set of possible phonemes at that syllable position taking into account a priori probabilities. All these conditions are the same for error and target phonemes. This issue will be taken up again in Study 2.

To summarize, contrary to widespread theoretical expectations and some published data, phoneme targets were not less frequent than their controls. Phoneme frequency thus shows a David Effect but not an overall Weak Units Effect: weak sources tend to replace otherwise normal targets, and generate weak-unit errors.

3.5 Word frequency effects

Using the broadest ETC dataset, we replicated the well-established Weak Units Effect at the lexical level: errors tend to occur on words of lower frequency than appropriately matched controls (non-ties = 1670, $z = 20.72$, $p < 0.001$). In this analysis, words that were not present in GRLDB were counted as being of zero frequency. We also carried out the same contrast excluding those cases and their matched controls from the analysis: the Weak Units Effect was still there (non-ties = 1374, $z = 14.86$, $p < 0.001$).

Very interestingly, and again contrary to predictions from most models, errors resulting in words included in GRLDB created words which are less frequent than their controls (non-ties = 348, $z = 12.70$, $p < 0.001$). In this set, the frequency of the error word was even lower than the frequency of the target word (non-ties = 344, $z = 4.36$, $p < 0.001$). Note that this comparison is warranted because all errors changing the number of syllables of the target word were excluded from this analysis, which ensures that word length (in terms of number of syllables, not phonemes) is controlled for.

To assess contrasts with the source word, we turned to the TSC dataset. Source words were also of lower frequency than their controls (non-ties = 1071, $z = 17.41$, $p < 0.001$), but they were comparable in frequency to the target word (non-ties = 554, $z = 0.12$, $p = 0.89$). The results were similar when only words found in GRLDB were included in the analysis (source versus control source: non-ties = 862, $z = 12.29$, $p < 0.001$; source versus target: non-ties = 454, $z = 0.23$, $p = 0.81$). This shows that there is an across-the-board Weak Units Effect in word units, and no David Effect. It also shows that having an error unit of lower frequency than the target is not a necessary consequence of the David Effect. A direct comparison between error and source including only those cases resulting in words found in GRLDB showed that error words are less frequent than source words (non-ties = 199, $z = 2.41$, $p < 0.02$). We will call this effect the *Weak Outcome Bias*.

The David Effect is therefore properly defined as a lower frequency source replacing a higher frequency target. We suggest that the Weak Outcome Bias is the result of the fortuitous rearranging of phonological units in the target (although this is less clear for syllable than word units). At the word level, this produces a noticeable number of nonwords (neologisms) and, because there are more low-frequency than high-frequency words in the lexicon, it also tends by chance to produce a low-frequency word when the result *is* a real word. This suggestion is consistent with the lack of lexical bias in Spanish, and it might point to important cross-linguistic differences in language production processes between Spanish and Germanic languages such as English.

3.6 Conclusions

Study 1 uncovered a complex and in many ways surprising pattern of results. Frequency effects were pervasive and in most cases extremely clear, but their shape poses important challenges to current models of phonological encoding.

Starting from the bottom up, phoneme units show a David Effect: the source phoneme is a weak phoneme, of lower frequency than expected by chance, but which is able nonetheless to replace a stronger target phoneme (which itself has a frequency similar to chance expectations). Once in its new location, the source phoneme (now the error phoneme) is of lower frequency than expected for that location.

At the syllable level, a David Effect was also found: the source syllable was of lower frequency than the target syllable. Syllables created as a result of the phonological error were also of lower frequency than the target. Moreover, there was an across-the-board Weak Units Effect: all three error, target, and source syllables were consistently of lower frequency than their corresponding chance level.

Finally, the word level also showed a kind of Weak Units Effect, with error phonemes coming from low frequency words interacting with phonemes in low frequency target words, and producing error words of even lower frequency. There was no David Effect (frequencies of target and source words were comparable), which suggests that the David Effect is independent of the Weak Outcome Bias.

These results seem to suggest that the current theoretical emphasis on low-quality processing of the target coupled with high-quality processing of the source is misguided, at least for the phonological encoding of Spanish. On the contrary, it appears that low-quality processing of the source unit may bear the primary responsibility for the generation of a speech error. However, before pursuing more fully the theoretical implications of our results, we will attempt to disentangle the effects observed at each level from the influences of cross-level correlations in frequency.

4. Study 2: Disentangling cross-influences between levels

Our first step in that direction is to find out which is the most representative frequency measure out of all twelve syllable and four phoneme measures. Once we reduce our measures to one, we can then turn to parametric statistics (Analysis of Covariance, ANCOVA) to replicate prior results at each level (phoneme, syllable, and word) and to see what remains of them once correlations with variables at other levels are statistically controlled for.

4.1 Factor Analysis of frequency measures at syllable and phoneme levels

We used Factor Analysis by principal components to reduce our data to a single most-representative variable. Factor Analysis finds one or more abstract variables which subsume the combined contributions of all dependent measures. In this sense, the most representative frequency measures would be the factors themselves. However, any future researcher willing to pursue this methodological strategy would be forced to compute the whole range of raw frequency measures over their speech error corpus, use them to calculate the factor(s), and then carry out the rest of their analyses using that/those factor(s) as the single frequency measure.

Another strategy that may be of more use to other researchers is to compute the percentage of total variance over the factor set which is shared by each raw frequency measure. This is called the "commonality" of a variable, and provides a good index of the degree to which a given raw measure is representative of all other measures; the raw measure with the highest commonality is the best available frequency measure. Other researchers may then rely on computing only this kind of frequency and use it in their studies.

For the reduction of the twelve syllable frequency measures, we used the ETC dataset (total available cases) and focused on the target control syllables, because they should constitute a representative sample of the types of syllables involved in Spanish speech errors, while at the same time varying freely across the whole frequency spectrum. Only two factors had eigenvalues greater than 1 and were included in the model (Kaiser 1960): the first factor had an eigenvalue of 7.43 and the second of 2.12. Together, they explained 79.69% of the total variance in the whole set of variables. Table 7 presents the commonalities of the syllable frequency measures.

Table 7
Proportion of common variance shared by each syllable frequency measure (commonality), as indexed by multiple R^2 over the factor set

R^2	*Syllable frequency measure*
0.984	Stress-sensitive token frequency in word-relative positional frame
0.709	Stress-sensitive type frequency in word-relative positional frame
0.948	Stress-sensitive token frequency in word-absolute positional frame
0.969	Stress-sensitive type frequency in word-absolute positional frame
0.983	Token frequency in word-relative positional frame
0.611	Type frequency in word-relative positional frame
0.808	Token frequency in word-absolute positional frame
0.953	Type frequency in word-absolute positional frame
0.881	Stress-sensitive token frequency
0.938	Stress-sensitive type frequency
0.845	Absolute token frequency
0.918	Absolute type frequency

Overall, commonalities of all variables were quite high, showing a noticeable degree of redundancy across the variables. The highest commonality was shown by the measure of token frequency of syllables with a particular stress in a particular position of a frame, relative to the word's number of syllables. We will use this measure in the analyses to follow.

We followed a similar strategy for the reduction of the four phoneme frequency measures. Single phoneme errors were selected from ETC data (1205 cases), and the four frequency measures of the control target phoneme were submitted to factor analysis. Only one factor had an eigenvalue greater than 1.0 (eigenvalue = 3.71, total variance explained = 92.77%). Table 8 shows the commonalities of the variables.

Table 8
Proportion of common variance shared by each phoneme frequency measure (commonality), as indexed by multiple R^2 over the factor set

R^2	*Phoneme frequency measure*
0.9947	Token frequency in syllabic positional frame
0.9943	Type frequency in syllabic positional frame
0.9925	Token frequency
0.9916	Type frequency

Phoneme frequency measures are even more highly correlated than syllable measures, and the differences across them in terms of their representativeness are even smaller. However, the results show once again that the highest commonality is observed for the most specific measure: token frequency in a syllable-position frame. This is the measure that will be used from now on as an index of phoneme frequency.

Overall, the results of the factor analysis of syllable and phoneme frequency measures suggest that the most specific measures outperform more encompassing counts as indexes of frequency, and that syllable-position frames should be relative to the total number of syllables of the word.

4.2 Designs

The overall factorial design included three factors: Unit Type (error, target, and source), Key-control (key unit versus control unit), and Level (word, syllable, and phoneme). Frequency was the dependent measure: lexeme frequency at the word level, stress-sensitive word-relative positional token frequency at the syllable level, and syllable-relative positional token frequency at the phoneme level. The dependent variables were $\log_{10}$ transformed to adjust their distribution to ANCOVA requirements.

The full Overall Design can only be tested in those cases which 1) involve a single phoneme; 2) are of unambiguous source (TSC dataset); 3) have error, target and source syllables of equal length; and 4) have all error, target and source words attested in GRLDB, so that they have frequencies different from zero. This leaves us with only 101 cases, and does not let us test the important set of cases resulting in nonwords.

In order to increase statistical power and include errors producing nonwords in the design, we carried out two analyses of partial designs. Partial Design A contains all three Levels (word, syllable, and phoneme) and the Key-control contrast, but only two Unit Types (target and source). Cases included in this analysis 1) involve single phonemes; 2) have an unambiguous source; 3) have target and source syllables of equal length; and 4) have target and source words attested in GRLDB. Total N in this design is 395.

Partial Design B focuses on two Levels (syllable and phoneme), plus the Key-control contrast and the three Unit Types (error, target, and source). The dataset includes cases which 1) involve single phoneme errors; 2) have an unambiguous source; and 3) have target and source syllables of equal length. Resulting N is 513 cases.

4.3 Overall design: Results

There was a main effect of Unit Type (error versus target versus source), which was marginally significant ($F(2, 200) = 2.93$, $p = 0.055$). There was also a very clear main effect of Key unit versus Control unit ($F(1, 100) = 160.4$, $p < 0.001$). Finally, there were unsurprising differences in total frequency across Levels ($F(2, 200) = 1426.15$, $p < 0.001$).

All interactions were significant. Unit Type interacted with Key-control ($F(2, 200) = 9.97$, $p < 0.001$), indicating differently shaped effects of Unit Type for Key and Control units. Level interacted with Unit Type ($F(4, 400) = 3.51$, $p < 0.01$) and with Key-control ($F(2, 200) = 95.20$, $p < 0.001$), pointing to different effects of both Unit Type and Key-control on word, syllable and phoneme units. Finally, the two-way interaction was also significant ($F(4, 400) = 11.17$, $p < 0.001$). The shape of this complex interaction was explored by means of post-hoc Tukey HSD comparisons. Figure 2 shows cell means. Significant contrasts at the 0.05 level are marked with a black dot or two dots joined by a line. From inspection, it can be concluded that conditions at the word level were primarily responsible for the two-way interaction. Note also that the significant contrast between the control error and control target phoneme that surprised us in section 3.3 above was not replicated in this analysis.

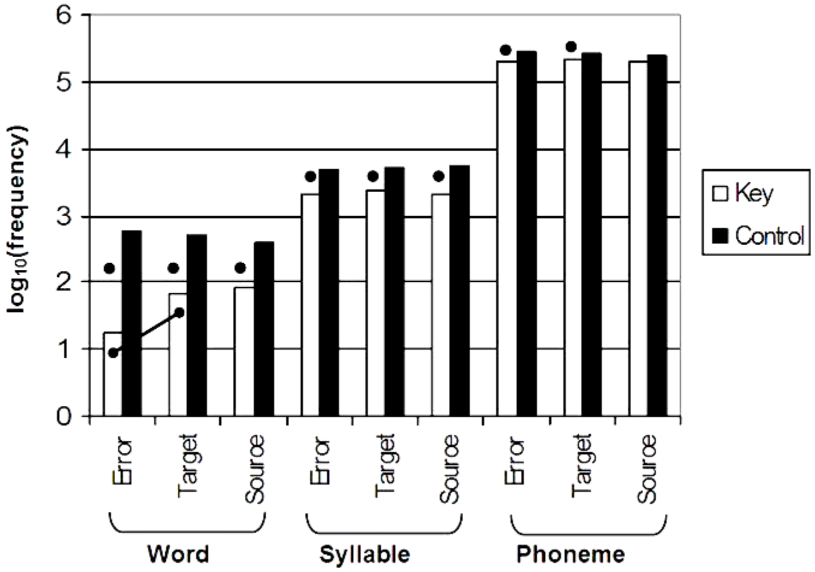

Figure 2
Mean log$_{10}$ frequency for the factors Unit Type, Key-control and Level in Overall Design of Study 2
(Black dots or two dots connected by a line indicate significant contrasts.)

Summing up, the Overall Design showed several effects of frequency at all levels, most of them of the weak-unit type. At the phoneme level, targets and errors were of lower frequency than their controls. At the syllable level, error, target, and source syllables were all of lower frequency than their controls. This was also true at the word level, where there also was a Weak Outcome Bias: error words were of lower frequency than target words. Therefore, this analysis did not replicate all frequency effects found in Study 1, perhaps because of reduced statistical power, or because of its focus on errors creating real words.

What would remain of these frequency effects at each level when measures of frequency at the other levels are introduced as covariates? To answer this question, we carried out independent ANOVAs at word, syllable, and phoneme levels and then repeated the same analyses introducing the other frequency measures as covariates. At the levels of syllable and phoneme frequency, we also introduced the level of featural specification as a covariate.

At the word level, there were clearly significant effects of both Unit Type ($F(2, 200) = 4.30$, $p < 0.02$) and Key-control ($F(1, 100) = 179.76$, $p < 0.001$), and their interaction ($F(2, 200) = 15.80$, $p < 0.001$). When syllable and phoneme frequency were introduced as covariates, the effects remained strong (Unit

Type: $F(2, 196) = 5.62$, $p < 0.01$; Key-control: $F(1, 98) = 83.58$, $p < 0.001$; Interaction: $F(2, 196) = 16.71$, $p < 0.001$).

At the syllable level, the ANOVA revealed only an effect of Key-control, as expected from Figure 2 ($F(1, 100) = 51.76$, $p < 0.001$; other *F*s < 1). When lexical and phoneme frequencies and phoneme featural specification were used as covariates, the effect vanished (all *F*s < 1 except for the interaction: $F(2, 194) = 1.69$, $p = 0.18$).

Finally, at the phoneme level, we also observed only a main effect of Key-control ($F(1, 100) = 15.08$, $p < 0.001$; all other *p* values > 0.15). This effect also disappeared when word and syllable frequencies and feature specification were used as covariates (all *p* values > 0.20).

To conclude, the analysis of the Overall Design by means of ANOVA and ANCOVA revealed strong frequency effects only at the lexical level, where both a Weak Units Effect and a Weak Outcome Bias were observed. All other frequency effects at other linguistic levels vanished after introducing available measures as covariates in the model. The possibility remains, though, that these negative results at syllable and phoneme levels are related to the small statistical power of this design (only 101 cases), so we will now try to corroborate them by using Partial Designs A and B.

4.4 Partial design A: Results

Partial Design A contains only two Unit Types, target and source, factorially crossed with three Levels (word, syllable, and phoneme) and the Key-control contrast, with a total *N* of 395 cases. In a global ANOVA, there was no main effect of Unit Type ($F(1, 394) = 2.39$, $p = 0.12$), while the other factors were significant (Key-control: $F(1, 394) = 309.71$, $p < 0.001$; Level: $F(2, 788) = 5476.07$, $p < 0.001$). The interaction between Unit Type and Key-control missed significance ($F(1, 394) = 2.30$, $p = 0.12$); the interaction between Unit Type and Level was marginally significant ($F(2, 788) = 2.67$, $p = 0.06$); and the interaction between Key-control and Level was clearly significant ($F(2, 788) = 82.30$, $p < 0.001$). Importantly, the two-way interaction of all three factors was significant ($F(2, 788) = 3.50$, $p < 0.05$).

We further explored this interaction by means of post-hoc Tukey HSD comparisons. Figure 3 shows the results: all key-unit types were of lower frequency than their controls with the exception of target phonemes. Moreover, there were clear David Effects at both the syllable and phoneme levels.

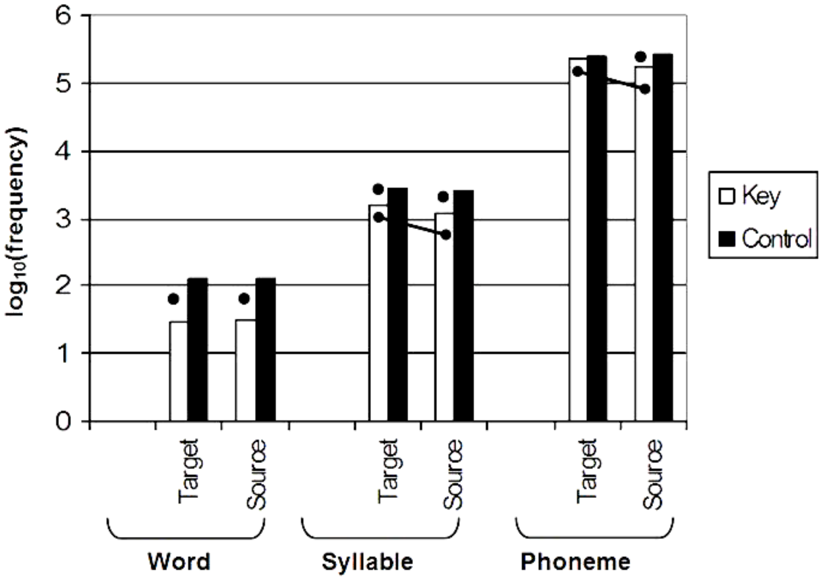

Figure 3
Mean $\log_{10}$ frequency for the factors Unit Type, Key-control and Level
in Partial Design A of Study 2
(Black dots or two dots connected by a line indicate significant contrasts.)

These results were backed by independent ANOVAs at each level, which we later re-ran using the other frequency measures plus degree of feature specification as covariates. At the word level, there was a strong main effect of Key-control only ($F(1, 395) = 197.83$, $p < 0.001$; other Fs < 1). The pattern was the same when both lexical and syllable frequencies were entered as covariates (Key-control: $F(1, 392) = 99.06$, $p < 0.001$; Unit Type: $F(1, 392) = 1.25$, $p = 0.26$; Interaction: $F < 1$).

At the syllable level, the analysis indicated clear effects of both Unit Type ($F(1, 395) = 5.14$, $p < 0.05$) and Key-control ($F(1, 395) = 160.57$, $p < 0.001$), and no interaction ($F(1, 395) = 1.85$, $p = 0.17$). The David Effect that was captured by post-hoc comparisons above was not strong enough to produce a significant interaction in this analysis. When word and syllable frequencies and feature specification were used as covariates, the pattern of results was not lost (Unit Type: $F(1, 391) = 4.21$, $p < 0.05$; Key-control: $F(1, 391) = 48.54$, $p < 0.001$; Interaction: $F < 1$). The significant differences between target and source syllables suggest that a David Effect remains in syllable frequency when correlations with other variables are discounted. However, this effect is also partially captured by the control conditions, producing an interaction which fails to be significant.

Finally, the phoneme level also showed clear main effects (Unit Type: $F(1, 394) = 9.26$, $p < 0.01$; Key-control: $F(1, 394) = 46.25$, $p < 0.001$). Importantly, the factors also interacted significantly ($F(1, 394) = 15.21$, $p < 0.001$). This interaction was unaffected by the introduction of the covariates word and syllable frequency and feature specification (Unit Type: $F(1, 390) = 7.68$, $p < 0.01$; Key-control: $F(1, 390) = 9.25$ $p < 0.01$; Interaction: $F(1, 390)$ 11.49, $p < 0.001$), supporting unambiguously the existence of a real David Effect at the phoneme level.

To summarize, Partial Design A focused on target and source units across the three linguistic levels and the key versus control contrast. Perhaps due to its greater statistical power (with 395 cases, versus 101 in the Overall Design), or because of the inclusion of errors resulting in nonwords, this analysis supported the existence of frequency effects at all three levels which cannot be accounted for by cross-covariations between frequency measures or with feature specification. In short, there was a clear Weak Unit Effect overall (with the exception of phoneme targets, which were as frequent as their controls), and there was also a clear David Effect at the phoneme level and a somewhat less clear David Effect at the syllable level.

4.5 Partial design B: Results

Partial Design B explores only the syllable and phoneme levels, including all three Unit Types (error, target, and source) and the Key-control contrast. Total N is 513 cases. The global ANOVA revealed clear main effects of all three factors (Unit Type: $F(2, 986) = 7.28$, $p < 0.001$; Key-control: $F(1, 493) = 250.19$, $p < 0.001$; Level: $F(1, 493) = 8407.20$, $p < 0.001$). The interaction between Unit Type and Key-control was significant ($F(2, 986) = 11.31$, $p < 0.001$), showing that the differences between key units and their matched controls vary depending on whether the unit is the error, the target, or the source. The interaction between Key-control and Level was also significant ($F(1, 493) = 94.55$, $p < 0.001$), indicating that the differences between key units and controls have varying sizes at the level of syllables and phonemes. Finally, neither the interaction between Unit Type and Level ($F(2, 986) = 1.31$, $p = 0.26$) nor the two-way interaction among all three factors ($F < 1$) reached significance. These results were also further explored by means of post-hoc HSD Tukey comparisons. Cell means and significant comparisons are presented in Figure 4.

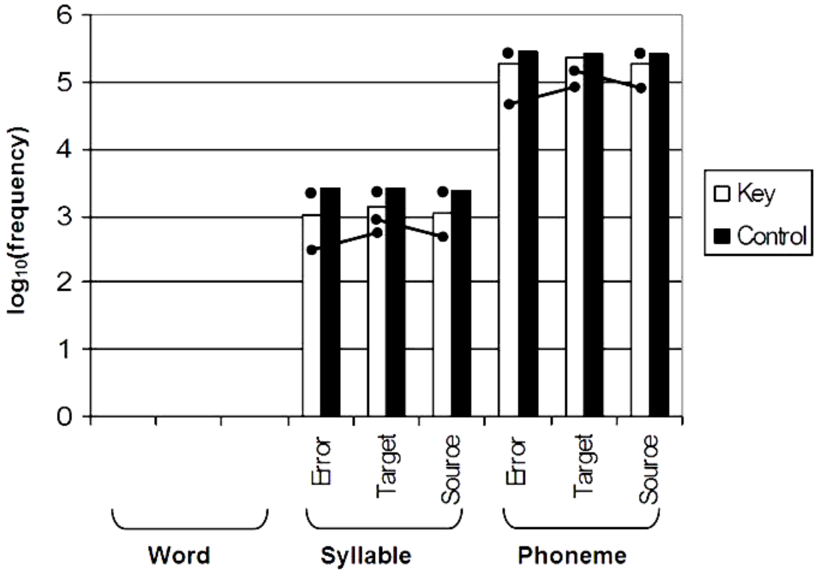

Figure 4
Mean $\log_{10}$ frequency for the factors Unit Type, Key-control and Level
in Partial Design B of Study 2
(Black dots or two dots connected by a line indicate significant contrasts.)

As in Partial Design A, there were across-the-board Weak Unit Effects, with the single exception of phoneme targets, and clear David Effects at both syllable and phoneme levels. Partial Design B also revealed Weak Outcome Biases (error units of lower frequency than target units) at both linguistic levels. As a collateral point, note that once again, this design did not replicate the difference between control error and control target phoneme conditions. Although it is unclear why this significant result arose in the non-parametric analyses of Study 1, it disappears when we focus on the most representative phoneme frequency measure, and we will not pursue this issue any further.

These results were also substantiated by independent ANOVAs for syllables and phonemes, which were then rerun including covariates to separate effects which can be ascribed only to cross-variable correlations from real frequency effects originating at that level.

At the syllable level, the factors showed both main effects and interaction (Unit Type $F(2, 988) = 4.01$, $p < 0.05$; Key-control: $F(1, 494) = 248.23$, $p < 0.001$; Interaction: $F(2, 988) = 4.48$, $p < 0.05$). Interestingly, when phoneme frequency and feature specification level were introduced as covariates, the contrast between key versus control units remained significant ($F(1, 491) = 151.82$, $p < 0.001$), but the effect of Unit Type was much reduced and missed significance ($F(2, 982) = 2.09$, $p = 0.12$).

At the phoneme level, the ANOVA also confirmed the effects of both factors and their interaction (Unit Type: $F(2, 1022)$[1] = 9.01, $p < 0.001$; Key-control: $F(1, 511) = 90.39$, $p < 0.001$; Interaction: $F(2, 1022) = 17.60$, $p < 0.001$). Here the introduction of syllable frequency and feature specification as covariates did not change the pattern of results (Unit Type: $F(2, 982) = 5.04$, $p < 0.01$; Key-control: $F(1, 491) = 10.80$, $p < 0.01$; Interaction: $F(2, 982) = 7.50$, $p < 0.001$).

In conclusion, after discounting cross-correlations across measures, Partial Design B supported prior results from Partial Design A. It showed a clear across-the-board Weak Unit Effect with the single exception of phoneme targets, as well as a clear David Effect at the phoneme level (phoneme sources are of lower frequency than phoneme targets). The David Effect observed at the syllable level seems to originate from the correlation of this measure with phoneme frequency, which suggests that the locus of the David Effect is to be found at the phoneme level. Partial Design B also found a Weak Outcome Bias at the phoneme level, but it is unclear to what extent this effect is independent of the David Effect, because error phonemes are identical to source phonemes and often appear at similar syllable positions. Present data show that the two effects are of a similar size.

5. General discussion

Most current theories of phonological encoding agree on the existence of syllable representations stored as part of the knowledge of the phonological form of words in the mental lexicon. This assumption predicts the existence of frequency effects linked to syllable units, both on vocal production latencies and on error rates. Most models assume that errors arise because of flawed processing of target units, and therefore it is a common expectation that target units should be weaker (lower frequency) than a control base line. Many models also take the view that source units compete directly with targets in the generation of the error, and thus that a direct cause of flaws in target processing originates in strong processing of the source. Source units should therefore be of higher frequency than targets. Finally, if the error involves the incorrect selection of a pre-stored unit instead of the weak target unit, stronger units should have a greater likelihood of being selected as errors than weaker units. This leads to the prediction that errors should be of higher frequency than targets, and perhaps even of higher frequency than a control base line.

These predictions are the same whether we focus on syllable units, our main center of interest, or on word or phoneme units; there is already some support in the literature for some of these predictions, at all three levels. However, as noted in the introduction, the support is limited to small samples, often from pathological populations, and have proven difficult to replicate. This lack of clear evidence is congruent with the very limited support (except in Mandarin

---

[1] The slight variations in degrees of freedom are due to the pairwise deletion of cases where the frequency measure is zero and therefore its $\log_{10}$ is indefinite. This occurs sometimes when syllable measures are part of the analysis (see comments on syllabic “neologisms” in section 3.3).

Chinese) for syllable units found with other methodologies such as vocal production latencies in several languages.

The present investigation is intended to fill this gap by looking at a sizeable corpus of spontaneous speech errors produced by Spanish speakers (a total of 1477 phonological errors). In order to ensure reliable conclusions, several methodological problems had to be resolved. First, it was unclear which measure of syllable and phoneme frequency is the best. Second, syllable frequency effects are highly correlated with frequency effects at both the phoneme and the word levels, and thus it is difficult to rule out whether apparent effects actually result from effects at other levels. Our approach consisted of identifying, for each phonological error, the error and target units, and whenever possible also the source unit at all word, syllable, and phoneme levels. We also measured degree of phoneme feature specification to control for Stemberger's (1991) Nondefault Bias. We computed twelve different measures of syllable frequency, and four measures of phoneme frequency. Finally, we devised a new method for generating a chance base line, consisting of randomly selecting appropriately matched control units, taking into account a priori probabilities.

The first step was then to test existing frequency effects independently at each level (word, syllable, and phoneme) by means of non-parametric statistics. Second, a factor analysis of the syllable and phoneme frequency measures lets us choose the most representative measures, which happened to be the most specific ones: stress-sensitive word-relative positional syllable token frequency, and syllable-relative positional phoneme token frequency. Word, syllable, and phoneme frequency were then $\log_{10}$ transformed, and submitted to parametric analyses of variance and covariance. We replicated all the effects observed with non-parametric techniques, and then disentangled the influences across dependent variables by including them in the design as covariates. Because a full overall design required the narrowing of the analysis to a small subset of errors (those resulting in real words), we also used two partial designs which enjoyed greater statistical power.

Overall, the results were extremely clear and consistent, which heightens the contrast with predictions from current theories. Present results can be summarized as follows.

At the phoneme level, target phonemes are not less frequent than expected by chance. Instead, sources are of lower frequency than chance and than targets, and when they reappear at the error position, they are still of lower frequency than targets and controls. It is the sources of phonological errors which are weak units, not the targets, and the resulting error is also low in frequency. We distinguished two theoretically important frequency effects here: a Weak Unit Effect, wherein relevant units are lower in frequency than their chance base line, and a David Effect, wherein weaker sources replace stronger targets. This David Effect dovetails with the finding that source phonemes have more specified features than expected by chance, whereas targets are specified to the degree expected by chance.

At the syllable level, all three relevant units (error, target, and source) are of lower frequency than chance (Weak Units Effect) to a similar extent (there is no David Effect).

Finally, at the word level, there is also an across-the-board Weak Units Effect (error, target, and source words are of lower-than-chance frequency) with no David Effect (target and source words are of comparable frequency). Interestingly, there is an additional frequency effect: when errors result in real words, these words are of even lower frequency than both target and source words. We called this the Weak Outcome Bias, and suggested that it arises from the fact that most words are of quite low frequency, so phonological errors are most likely to resemble words of low frequency. This is coupled with the lack of a lexical bias in Spanish (see del Viso et al. 1991), also attested in this same data set by Pérez et al. (2006).

Much research still remains to be done to clarify the present pattern of results, and an important line concerns the replication of this study on a corpus of Germanic (English, German, Dutch) speech errors, as would be necessary to assess to what extent the present results are specific to Spanish or generalizable across languages. Our impression at this point is that the Weak Units and David Effects are unlikely to be a Spanish rarity (in spite of available reports of higher-frequency source phonemes replacing lower-frequency targets in English once the Nondefault Bias is controlled for, Stemberger 1991a). The Weak Outcome Bias at the word level might be a specifically Spanish pattern; however, a recent study (Dell, Reed, Adams and Meyer 2000) suggests that a similar effect is observed in English under certain conditions. That study analyzed the phonological errors that participants produced when reciting series of CVC syllables over several sessions. The syllables came from a vocabulary that followed a set of clear phonotactic constraints. One of the predictions tested was that syllables that are not part of the vocabulary should be less likely as error outcomes than syllables that are in the vocabulary and have just received extensive practice. No preference for experimental syllables was observed: errors often created syllables not occurring in the experiment (therefore, of lower frequency than targets), even though they respected the experiment-internal phonotactics to an impressive extent. The authors interpreted this result as indicating the recombination of syllable internal units (CV and VC sequences) without whole-syllable representations. Because only monosyllables were used, it is unclear whether this effect belongs to the syllable level or to the word level, and it may provide an analog of the Weak Outcome Bias reported here.

In agreement with current theories, observed frequency effects support the independent representation of syllables in the lexicon (as well as words and phonemes). However, the shape of these effects suggests important changes to most theories. The David Effect, in which the error tends to be a weaker unit than the target, reveals an interesting bias in the reasoning that has thus far gone into cognitive models of language production. In all probability, it is weakness of the source (which is generally segmentally identical to the error) that underlies the effect. Most models have predicted that the source should be either neutral (and can be either a strong or a weak unit) or else a strong unit. It is a surprise that the source tends to be a weak unit.

Let's pull back for a moment and consider what must be done in order for a word to be processed correctly: 1) The correct set of units must be accessed. 2) Exactly the right number of tokens of each unit must be accessed. 3) Every unit must be placed in the correct position relative to the other units. Consider an addition error such as *blite block* for target *bite block*, in which an /l/ is added to the first word and additionally produced in its proper location in the second word. The first word (*bite*) has accessed all of its segments, with exactly the right number of each, and broadly in the right location relative to the other units (although with the contiguity between the /b/ and the following vowel destroyed). The second word (*block*) has accessed the correct set of units, but with an extra token of /l/, which is located in an incorrect location relative to the other units in the word. There is a sense in which the word *bite* is accurate here, and the word *block* is in error. But the extra segment of *block* has moved imperialistically into the word *bite*, which now has one too many segments. We have heretofore identified *bite* as the "target" word that is in error, and have asked what factors make the target word more susceptible to error. We have identified *block* only as the "source," suggesting either that it is a passive source of elements rather than the driving force behind the error, or that it is the efficient accessing of output units by a strong source that leads to spill-over onto weak targets.

Instead, our results suggest that it is often the source that can be viewed as being in error. The source has failed to keep control of its parts, which wander into nearby susceptible words. Researchers have ignored the necessary function of a word to maintain its integrity and present a set of phonological elements in a particular target order. If we take a new perspective and attach this function to each lexical item during processing, then we expect that strong units will more accurately keep their elements in their proper location, while weak units will do so less accurately. Weak units thus should not only tend to be the recipient of erroneous phonological elements from other words (which models do a good job of explaining) but also should tend to lose track of their phonological elements and contribute erroneous phonological elements to other words. In retrospect, since we expect that strong units will be processed more efficiently and accurately, why did we ever think that strong units would be less capable of keeping control of their phonological parts?

This viewpoint has similarities to other proposals. First, there are models of processing that presuppose that there are limited resources that are used to access all information (Crystal 1987). Anything that stresses this general resource capacity can lead to errors not just on the weak unit, but on other units as well. Thus, a weak unit can facilitate errors elsewhere in the utterance, even if the weak unit itself is processed accurately. Second, there is a parallel in the theory of Optimality Theory (OT) in linguistics (e.g. Prince and Smolensky 2004). The focus of OT is on individual words, but with a recognition that multi-word utterances also fall within the scope of the theory. In order to prevent deletion of segments and also to make morphological infixes a marked type of affix, OT uses the constraints CONTIGUITY (elements that are contiguous in the stored form of the morpheme must be contiguous in the output) and INTEGRITY (the segments in a morpheme hold together, without intrusion of extraneous phonological material of any sort). Both of these constraints are faithfulness constraints that are designed to ensure the faithful output of a morpheme's phonological elements. Bernhardt and Stemberger (1998) proposed that weak units show a lower level of faithfulness than strong units (including low-frequency words but also phonological elements in cognitively weak positions), and CONTIGUITY and INTEGRITY should act like other faithfulness constraints. The /l/ in the first word of ***bl**ite block* violates both INTEGRITY and CONTIGUITY; both words (target and source) are predicted to be weak units, since a higher level of faithfulness in either word is likely to block the error.

The idea that weak units lose control of their elements, leading to errors in nearby words, is conceptually reasonable. The challenge, of course, is integrating it into theories that have not viewed the phonological integrity of a word as something that is related to the accuracy of processing of that word. This remains to be done for models of language production.

References

Aichert, Ingrid and Wolfram Ziegler (2004). Syllable frequency and syllable structure in apraxia of speech. *Brain and Language* 88, 148–159.

Alameda, Ramón and Fernando Cuetos (1995). *Diccionario de frecuencias de las unidades lingüísticas del castellano*. Oviedo: Universidad de Oviedo.

Bachoud-Levi, A. C., E. Dupoux, L. Cohen and J. Mehler (1998). Where is the length effect? A cross-linguistic study of speech production. *Journal of Memory and Language* 39, 331–346.

Berg, Thomas (1991). Phonological processing in a syllable-timed language with pre-final stress: Evidence from Spanish speech error data. *Language and Cognitive Processes* 6, 265–301.

Bernhardt, Barbara H. and Joseph P. Stemberger (1998). *Handbook of phonological development from the perspective of constraint-based nonlinear phonology*. San Diego: Academic Press.

Carreiras, Manuel, Carlos J. Álvarez and Manolo de Vega (1993). Syllable frequency and visual word recognition in Spanish. *Journal of Memory and Language* 32, 766–780.

Chen, Jenn-Yeu (2000). Syllable errors from naturalistic slips of the tongue in Mandarin Chinese. *Psychologia: An International Journal of Psychology in the Orient* Special Issue: Cognitive processing of the Japanese and Chinese languages II, 43, 15–26.

Chen, Jenn-Yeu, Train-Min Chen and Gary S. Dell (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language* 46, 751–781.

Chen, Jenn-Yeu, Wei-Chun Lin and Ludovid Ferrand (2003). Masked priming of the syllable in Mandarin Chinese speech production. *Chinese Journal of Psychology* 45, 107–120.

Cholin, Joana, Willem J. M. Levelt and Niels O. Schiller (2006). Effects of syllable frequency in speech production. *Cognition* 99, 205–235.

Cholin, Joana, Niels O. Schiller and Willem J. M. Levelt (2004). The preparation of syllables in speech production. *Journal of Memory and Language* 50, 47–61.

Crystal, David (1987). Towards a 'bucket' theory of language disability: Taking account of interaction between linguistic levels. *Clinical Linguistics and Phonetics* 1, 7–22.

Cutler, Anne, James M. McQueen, Dennis Norris and A. Somejuan (2001). The roll of the silly ball. In *Language, brain, and cognitive development: Essays in honor of Jacques Mehler,* ed. Emmanuel Dupoux, 181–194. Cambridge: MIT Press.

del Viso, Susana (1992). *Errores espontáneos del habla y producción del lenguaje*. Madrid: Editorial de la Universidad Complutense de Madrid.

del Viso, Susana, José Manuel Igoa and José E. García-Albea (1991). On the autonomy of phonological encoding: Evidence from slips of the tongue in Spanish. *Journal of Psycholinguistic Research* 20, 161–185.

Dell, Gary S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93, 283–321.

Dell, Gary S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes* 5, 313–349.

Dell, Gary S., Cornell Juliano and Anita Govindjee (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science* 17, 149–195.

Dell, Gary S., Kristopher D. Reed, David R. Adams and Antje S. Meyer (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26, 1355–1367.

Dell, Gary S. and Peter A. Reich (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior* 20, 611–629.

Ferrand, Ludovic, Juan Segui and Jonathan Grainger (1996). Masked priming of word and picture naming: The role of syllabic units. *Journal of Memory and Language* 35, 708–723.

Ferrand, Ludovic, Juan Segui and Glynn W. Humphreys (1997). The syllable's role in word naming. *Memory & Cognition* 25, 458–470.

Gagnon, Deborah A., Myrna F. Schwartz, Nadine Martin, Gary S. Dell and Eleanor M. Saffran (1997). The origins of formal paraphasias in aphasics' picture naming. *Brain and Language* 59, 450–472.

García-Albea, José E., Susana del Viso and José Manuel Igoa (1989). Movement errors and levels of processing in sentence production. *Journal of Psycholinguistic Research* 18, 145–161.

Garrett, Merrill F. (1988). Processes in language production. In *Linguistics: The Cambridge survey, Vol. III, Language: Psychological and biological aspects,* ed. Frederic J. Newmeyer, 69–96. Cambridge: Cambridge University Press.

Garrett, Merrill F. (2001). Now you see it, now you don't: Frequency effects in language production. In *Language, brain, and cognitive development: Essays in honor of Jacques Mehler,* ed. Emmanuel Dupoux, 227–240. Cambridge: MIT Press.

Harley, Trevor A. and Sally B. G. MacAndrew (2001). Constraints upon word substitution speech errors. *Journal of Psycholinguistic Research* 30, 395–418.

Hendricks, H. and James McQueen (eds.) (1996). *Max Planck Institute for Psycholinguistics, Annual Report 1995*. Nijmegen: Max Planck Institute for Psycholinguistics.

Hotopf, W. H. N. (1980). Semantic similarity as a factor in whole-word slips of the tongue. In *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand,* ed. Victoria A. Fromkin, 97–109. New York: Academic Press.

Jared, Debra and Mark S. Seidenberg (1990). Naming multisyllabic words. *Journal of Experimental Psychology: Human Perception and Performance* 16, 92–105.

Jescheniak, Jörg D. and Willem J. M. Levelt (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory and Cognition* 20, 824–843.

Kaiser, Henry F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement* 20, 141–151.

Klapp, Stuart T., Wallace G. Anderson and Raymond W. Berrian (1973). Implicit speech in reading, reconsidered. *Journal of Experimental Psychology* 100, 368–374.

Laganaro, Marina (2005). Syllable frequency effect in speech production: Evidence from aphasia. *Journal of Neurolinguistics* 18, 221–235.

Levelt, Willen J. M., Ardi Roelofs and Antje Meyer (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22, 1–38.

Levelt, Willem J. M. and Linda R. Wheeldon (1994). Do speakers have access to a mental syllabary? *Cognition* 50, 239–269.

MacKay, Donald G. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia* 8, 323–350.

MacKay, Donald G. (1987). *The organization of perception and action*. New York: Springer-Verlag.

Mehler, Jacques, Jean-Yves Dommergues, Uli Frauenfelder and Juan Segui (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior* 20, 298–305.

Meyer, Antje (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language* 29, 524–545.

Meyer, Antje (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language* 30, 69–89.

Meyer, Antje S., Ardi Roelofs and Willem J. M. Levelt (2003). Word length effects in object naming: The role of a response criterion. *Journal of Memory and Language* 48, 131–147.

Nickels, Lindsey A. (1995). Getting it right? Using aphasic naming errors to evaluate theoretical models of spoken word production. *Language and Cognitive Processes* 10, 13–45.

Nickels, Lindsey A. and David Howard (1995). Aphasic naming: What matters? *Neuropsychologia* 33, 1281–1303.

Nickels, Lindsey A. and David Howard (2004). Dissociating effects of number of phonemes, number of syllables, and syllabic complexity on word production in aphasia: It's the number of phonemes that counts. *Cognitive Neuropsychology* 21, 57–78.

Nooteboom, Sieb G. (1969). The tongue slips into patterns. In *NOMEN: Leyden studies in linguistics and phonetics,* ed. A. G. Sciarone, 114–132. The Hague: Mouton.

Pérez, Elvira, Julio Santiago, Alfonso Palma and Padraig G. O'Seaghdha (2006). Perceptual bias in speech error data collection: Insights from Spanish speech errors. *Journal of Psycholinguistic Research*, doi:10.1007/s10936-006-9042-7.

Prince, Alan and Paul Smolensky (2004). *Optimality Theory: Constraint interaction in generative grammar*. Oxford: Blackwell.

Robson, Jo, Tim Pring, Jane Marshall and Shula Chiat (2003). Phoneme frequency effects in jargon aphasia: A phonological investigation of nonword errors. *Brain and Language* 85, 109–124.

Roelofs, Ardi and Antje Meyer (1998). Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24, 922–939.

Santiago, Julio, Donald G. MacKay, Alfonso Palma and Christine Rho (2000). Sequential activation processes in producing words and syllables: Evidence from picture naming. *Language and Cognitive Processes* 15, 1–44.

Schiller, Niels O. (1998). The effect of visually masked syllable primes on the naming latencies of words and pictures. *Journal of Memory and Language* 39, 484–507.

Schiller, Niels O. (1999). Masked syllable priming of English nouns. *Brain and Language* 68, 300–305.

Schiller, Niels O. (2000). Single word production in English: The role of subsyllabic units during phonological encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26, 512–528.

Schiller, Niels O., Antje S. Meyer, Harald Baayen and Willem J. M. Levelt (1996). A comparison of lexeme and speech syllables in Dutch. *Journal of Quantitative Linguistics* 3, 8–28.

Shattuck-Hufnagel, Stefanie (1979). Speech error evidence for a serial ordering mechanism in sentence production. In *Sentence processing: Psycholinguistic studies presented to Merrill Garrett,* ed. W. E. Cooper and E. C. T. Walker, 295–342. Mahwah, NJ: Erlbaum.

Shattuck-Hufnagel, Stefanie and Dennis H. Klatt (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Memory and Language* 18, 41–55.

Stemberger, Joseph P. (1983). Inflectional malapropisms: Form-based errors in English morphology. *Linguistics* 21, 573–602.

Stemberger, Joseph P. (1984). Structural errors in normal and agrammatic speech. *Cognitive Neuropsychology* 1, 281–313.

Stemberger, Joseph P. (1985). An interactive activation model of language production. In *Progress in the psychology of language,* vol. 3, ed. Andrew W. Ellis, 143–186. London: Lawrence Erlbaum.

Stemberger, Joseph P. (1990). Wordshape errors in language production. *Cognition* 35, 123–157.

Stemberger, Joseph P. (1991a). Apparent anti-frequency effects in language production: The addition bias and phonological underspecification. *Journal of Memory and Language* 30, 161–185.

Stemberger, Joseph P. (1991b). Radical underspecification in language production. *Phonology* 8, 73–112.

Stemberger, Joseph P. (2002). Overtensing and the effect of regularity. *Cognitive Science* 26, 737–766.

Stemberger, Joseph P. and Brian MacWhinney (1986). Frequency and the lexical storage of regularly inflected forms. *Memory & Cognition* 14, 17–26.

Stemberger, Joseph P. and Christine S. Middleton (2003). Vowel dominance and morphological processing. *Language and Cognitive Processes* 18, 369–404.

Stenneken, P., M. Hofmann and A. M. Jacobs (2005). Patterns of phoneme and syllable frequency in jargon aphasia. *Brain and Language* 95, 221–222.

Sternberg, Saul, Stephen Monsell, R. L. Knoll and C. E. Wright (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In *Information processing in motor control and learning,* ed. George E. Stelmach, 117–152. New York: Academic Press.

Vousden, Janet I., Gordon D. A. Brown and Trevor A. Harley (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology* 41, 101–175.

Wilshire, Caroline and Jean-Luc Nespoulous (2003). Syllables as units in speech production: Data from aphasia. *Brain and Language* 84, 424–447.

(Santiago)
Dept. de Psicología Experimental y Fisiología del Comportamiento
Universidad de Granada
180171-Granada
Spain

santiago@ugr.es
http://www.ugr.es/local/psilcom/jsantiago-i.htm