J Psycholinguist Res (2007) 36:207–235
DOI 10.1007/s10936-006-9042-7

ORIGINAL PAPER

Perceptual Bias in Speech Error Data Collection: Insights from Spanish Speech Errors

Elvira Pérez · Julio Santiago · Alfonso Palma · Padraig G. O'Seaghdha

Published online: 22 December 2006

Abstract This paper studies the reliability and validity of naturalistic speech errors as a tool for language production research. Possible biases when collecting naturalistic speech errors are identified and specific predictions derived. These patterns are then contrasted with published reports from Germanic languages (English, German and Dutch) and one Romance language (Spanish). Unlike findings in the Germanic languages, Spanish speech errors show many patterns which run contrary to those expected from bias: (1) more phonological errors occur between words than within word; (2) word-initial consonants are less likely to participate in errors than word-medial consonants, (3) errors are equally likely in stressed and in unstressed syllables, (4) perseverations are more frequent than anticipations, and (5) there is no trace of a lexical bias. We present a new corpus of Spanish speech errors collected by many theoretically naïve observers (whereas the only corpus available so far was collected by two highly trained theoretically informed observers), give a general overview of it, and use it to replicate previous reports. In spite of the different susceptibility of these methods to bias, results were remarkably similar in both corpora and again contrary to predictions from bias. As a result, collecting speech errors "in the wild" seems to be free of bias to a reasonable extent even when using a multiple-collector method. The observed contrasting patterns between Spanish and Germanic languages arise as true cross-linguistic differences.

Keywords Spanish · Speech production · Naturalistic speech errors · Perceptual bias · Lexical bias

E. Pérez · J. Santiago (✉) · A. Palma
Departamento de Psicología Experimental y Fisiología del Comportamiento,
Universidad de Granada,
Campus de Cartuja, 18179 Granada, Spain
e-mail: santiago@ugr.es

P. G. O'Seaghdha
Department of Psychology, Lehigh University,
Bethlehem, PA, USA

Introduction

Speech errors are little jewels for psycholinguists. At least since the work of Arab linguist Al-Ki-sa'i in the 8th Century AD (Anwar, 1979), it was realized that inadvertent "slips of the tongue" are far from being random. Rudolf Meringer was the first European linguist who devoted attention to them, highlighting that the structured patterns found in collections of speech errors revealed a complex speech production system within the individual mind, and not just a possible source of linguistic change (Fromkin, 1988; Meringer & Mayer, 1895).

Speech errors were instrumental in the process of scientific change that led to the advent of the cognitive revolution in the 50s (Baars, 1986). Based on exchange errors such as *let us always remember that waste makes haste*, Lashley (1951) rejected associative chaining explanations of goal-directed series of actions. When the new scientific context licensed the study of unobservable mental processes, naturalistic speech errors became one of the main empirical domains of a fruitful research tradition on language production mechanisms (see papers compiled in Fromkin, 1980). Fromkin (1971) and Garrett (1975, 1980) soon showed that error data allow drawing a comprehensive and coherent outline of the language production system.

Fromkin (1971) observed that the interacting units in slips of the tongue correspond to a single linguistic definition. Words interact with words, as in the following example:[1]

(1) a laboratory in our own computer (a computer in our own laboratory).

Stems interact with stems, phonemes with phonemes (as in example 2):

(2) a hunk of jeep (a heap of junk)

However, cases of mixed interaction (e.g., words with phonemes) are exceedingly rare.[2] This suggests the existence of representational levels defined by their type of units. These units are processed simultaneously within a given span, leading to occasional errors restricted to them (Garrett, 1975). Examples like 3 also indicate the relative order of processing levels.

(3) a meeting arathon (an eating marathon).

The movement of phoneme /m/ from "marathon" to the beginning of "eating" blocks the application of the morphophonemic rule which adjusts the indefinite article to a following vowel context generating the output "an". Therefore, phoneme movement must have occurred before the realization of morphophonemic adjustments.

Current understanding of language production is strongly rooted in the analysis of the distributional properties of naturalistic speech errors. Arguments such as those just outlined have been complemented with others based on the frequency of occurrence of different types of errors as a function of several variables. For instance, Garrett (1975, 1980) showed that contextual word errors (errors in which the interacting

[1] Hereafter, in exemplifying speech errors we follow the convention of placing first the error utterance followed by the intended utterance in quotes. Sometimes (specially with Spanish examples) we highlight the involved units by writing them in upper case.

[2] This is not to say that all error units have a proper definition in current linguistic theory, as some cases defy known linguistic categories. Such cases are, nonetheless, also quite infrequent.

words come from the surrounding sentential context) usually span longer distances than contextual phonological errors, suggesting greater advance planning at "higher" than "lower" levels. As another example, **?**Shattuck-Hufnagel and Klatt (1980) found that phoneme errors are greatly affected by the similarity in articulatory features between the target and error phoneme, suggesting that features are represented at the phonological level even though they very rarely participate directly in errors (Levelt, 1989; MacKay, 1987).

Speech errors have a compelling ecological validity (Bock, 1996). They do not come from laboratory situations that aim to replicate key aspects of natural communication, but instead from the full-fledged real situation itself. It is always arguable up to what point the replication in controlled settings is successful and whether the observation procedure alters the processes being observed (see Stemberger, 1992, for discussion of experimentally elicited slips, and Bock, 1996, for a wider perspective).

Most current theories are based mainly on data from English, but a growing effort is being devoted to isolate language universal and specific aspects of processing. Cross-linguistic research has a well-established tradition in the realm of speech perception (see Otake & Cutler, 1996, for an overview). In language production, the wide majority of studies encompass only English, Dutch, and German, all of them Germanic languages with many shared structural characteristics. As a matter of fact, few interesting processing differences have been found among them. Only a handful of studies have addressed languages from other families (e.g., Chen, 1999, 2000; Chen, Chen, & Dell, 2002, on Mandarin Chinese; Bachoud-Levy, Dupoux, Cohen, & Mehler, 1998, on French; Costa & Sebastián-Gallés, 1998, on Spanish). To our knowledge, only seven studies have addressed Spanish speech errors (Anton-Mendez, Hartsuiker, Roelstraete & Costa, 2005; Berg, 1991; del Viso, 1992; del Viso, Igoa, & García-Albea, 1991; García-Albea, del Viso, & Igoa, 1989; Hartsuiker, 2002; Igoa, 1996), and all of them worked from the same error collection (del Viso, 1992). In this context, the present study will help drawing a clearer picture of universals and variation in language processing by examining Spanish speech error patterns and comparing them to Germanic error patters.

This paper has several intertwined goals. First, we want to introduce a new collection of slips of the tongue produced by Spanish speakers in everyday conversations, compiled by over 700 theoretically naïve trained observers, and to give a general description of its main error patterns. Second, and more importantly, we want to ascertain whether, and to what extent, possible biases may have affected our procedure of error collection. In order to achieve this goal, we will use the literature on error detection to identify specific error patterns which may be the result of one or more kinds of bias (chiefly perceptual bias). We then undertake a threefold comparison between published results obtained from Germanic corpora (mainly English, Dutch and German), the other Spanish corpus currently available (del Viso, 1992) and our own corpus. Germanic corpora show some patterns which comply with predictions from bias, but they seem to be absent from del Viso's Spanish corpus. This makes its independent replication especially important. As an additional source of constraints, del Viso's corpus was compiled mostly by only two highly theoretically informed trained observers. Consequently, some types of bias may arguably affect her corpus differently from ours in predictable ways. In order to cross-validate both corpora and to replicate prior results on Spanish speech errors, we will first carry out a detailed side-by-side comparison of general error patterns in both corpora and then we will look for some more specific patterns in phonological errors in our corpus. The result of this procedure will allow us to assess the extent to which Spanish corpora are affected by methodological biases, and by extension we will also cross-validate Germanic corpora.

Reliability and validity of naturalistic slips of the tongue

Methodological criticisms of speech error data come mainly in the form of the "bias argument". The usual collection procedure involves one or more observers who collect speech errors as they come across them in their everyday life. More often, corpora are collected by one or a few highly trained and theoretically well-informed observers (which we will call *single-collector methodology*, used by, e.g., del Viso, 1992; Fromkin, 1971; Garrett, 1975; Harley, 1984; Kawachi, 2002; Meringer & Mayer, 1895; Stemberger, 1985). In some cases, a *multiple-collector methodology* has been used (Dell & Reich, 1981; and the present corpus). Here, a higher number of trained but theoretically naïve observers, usually psychology students, are used.[3] Errors are noted in writing, often using the standard Roman alphabet. Tape-recorded corpora are rare (Boomer & Laver, 1968; Garnham, Shillcock, Brown, Mill, & Cutler, 1982; Wijnen, 1992). Consequently, it is often not possible to check the reliability of the observations, opening the door to a variety of possible biases (see Cutler, 1982, 1988, and Bock, 1996, for general discussions). There is also the possibility that single-collector and multiple-collector methodologies are differentially affected by bias.

We will first discuss two potential biases that, even though probably present in the corpora, are unlikely to bias the inferences that are usually drawn from them if appropriate control measures are taken. We will then turn to three biases which can affect the patterns attested in the corpora, point out what shape these influences may take, discuss how they could differentially affect corpora based on single and multiple-collector methods, and identify what would be their symptoms if they do affect a corpus.

First, there is the "distributional bias" (Bock, 1996; Stemberger, 1992): the proportion of errors of a particular type might be incorrectly estimated just because the structural characteristics of the language allow more (or less) opportunities for that error to occur. The problem is that it is often difficult to establish the chance level for the occurrence of a particular type of error. The solution to the problem of distributional bias requires the development of sophisticated methods of chance estimation (see Dell & Reich, 1981; Stemberger, 1991a). Second, there is the "categorization ambiguity bias" (Bock, 1996; Meyer, 1992): many errors are difficult to categorize into a particular error type. If there is any systematic bias in their categorization, it may lead to inflated frequencies of some error categories. The ambiguity bias is counteracted by using standard criteria for inclusion. Standard criteria make different corpora comparable, and also allow comparisons of error frequencies as a function of other variables. Marking ambiguous errors is another standard practice that allows the selection of clear cases when the hypothesis under consideration needs a direct comparison across potentially biased categories (e.g., del Viso, 1992).

Turning to more serious threats to the reliability and validity of slip corpora, we will first consider what may be called the "personal bias". Ferber (1991) compared

[3] Berg (1991), in his reanalysis of del Viso's corpus, asserts that this was compiled using a multiple-collector method. However, between 90% and 95% of all errors in del Viso's corpus were collected by only two observers: Susana del Viso and Jose Manuel Igoa. The rest of the errors were contributed by a small number of other people, mainly Igoa's PhD students (see del Viso, 1992, p. 101 and p. 147).

the errors collected by four trained observers in tape-recorded radio conversations with the errors detected after repeated listening of the same material. She reports surprisingly low levels of error detection (around one-third of total errors) and of between-observer agreement, with not a single slip being recorded by all four observers and most of them being detected by only one observer. A low general level of error detection is not a problem for speech error corpora in itself as long as the detectability does not vary across error categories. Importantly, the four observers did show variations in sensitivity to errors at different levels (phonemic, lexical and grammatical), with two of them failing to detect a single lexical level error and another observer completely missing all phoneme level errors. These results suggest that there might be strong personal biases affecting speech error corpora. Their effects will be stronger the smaller the number of independent observers. Personal biases may sometimes arise out of a theoretical bias towards a certain type of error (see Ellis, 1980, for discussion of Freud's, 1922, possible biases when selecting errors for psychoanalytic interpretation). Ferber's (1991) observers, however, had no particular theoretical biases (only one of them was a linguist), which suggest that personal biases are mostly due to individual differences in attention, memory, perceptual discrimination and other basic cognitive skills. Although expert observers are unlikely to miss entire categories of errors, personal biases may affect corpora compiled using the single-collector method in more subtle ways, but should not affect multiple-collector corpora when the number of observers is high. Comparisons between del Viso's single-collector corpus and our own multiple-collector corpus might therefore reveal the operation of some personal biases in the former.

Linked to the use of a single-collector methodology there is also what we call the "sampling bias". If only one or a few observers compile the corpus, most errors will derive from a very small sample of relatives, acquaintances and the observers themselves. It has been shown that stylistic factors such as speech rate have differential effects on error types (Dell, 1986, 1990; Dell, Burger, & Svec, 1997). People may therefore vary in their characteristic error patterns as they vary in their characteristic speech style. If errors come from a small sample of people, the resulting collection could be biased toward the speakers' "favourite" errors. Again, multiple-collector corpora should reduce sampling bias.

Finally, the most widely discussed potential bias in slip collections is the "perceptual bias", and it will also get the lion's share of attention here. It includes a family of related causes that may lead to some speech errors being more likely to be recorded than others (Bock, 1996; Cutler, 1982; Stemberger, 1992). The main cause is greater perceptual and attentional salience of some errors, but there are also memory factors, as more salient errors are also more likely to be correctly recalled until they are written down. Speech perception studies have shown that perceptual salience varies with a number of factors, leading to very specific expectations about what errors should be more easily detectable.

Perceptual bias predictions are much more detailed at the phonological level than at lexical and supralexical levels. It makes sense that subtle differences in salience may make a stronger impact at the level of more difficult perceptual discriminability. Consequently, it is reasonable to assume that perceptual bias, in any of its forms, would affect more strongly multiple-collector than single-collector corpora. Observers participating in multiple-collector studies are probably less motivated and trained on the average than observers in single-collector studies, making them more prone to miss less noticeable errors (Dell & Reich, 1981).

Testing the perceptual bias hypothesis

Based on the preceding discussion, we propose that a variety of perceptual biases may affect the data contained in speech error corpora by systematically increasing the probability that some errors are detected whereas others escape unnoticed. Self-evident as it might seem, however, that these biases are actually at work in naturalistic slip corpora is still a question open to empirical test. At least one study (Ferber, 1991) found large individual differences in error detection, both at the total rate as well as at the specific error type detection by listener, and no evidence for a higher detectability of some errors over others. These findings are inconsistent with perceptual biases being mandatory, and therefore listeners could be sensitive to training.

An obvious way of validating naturalistic speech error data would be to turn to errors generated under more controlled conditions. There are now available a variety of techniques for generating action slips in the laboratory (see Baars, 1980, 1992a, and contributions in Baars, 1992b). A comparison of slip patterns obtained in the laboratory with those collected during everyday conversations would allow a cross-validation of both methods. Stemberger (1992) undertook just this task. He concluded that most results from naturalistic corpora are replicated in the laboratory, with a few exceptions, which are generally traceable to specific task demands.

Experimental methods, however, suffer from their own problems of ecological validity. In this paper, we follow a complementary strategy: search for the patterns predicted from perceptual bias in naturalistic corpora. As Cutler (1982) concluded, patterns that coincide with the predictions from bias are suspicious, while those that do not coincide, can be taken with considerable confidence. First, we will review available evidence from Germanic corpora. Then we will turn to assess the relevant patterns in del Viso's (1992) Spanish corpus.

In English, Dutch and German corpora, some patterns are found which coincide with the predictions from perceptual bias, while others contradict them. As predicted from perceptual bias, there are more errors:

- on word-initial consonants than medial and final consonants (Boomer & Laver, 1968; MacKay, 1970; Shattuck-Hufnagel, 1987, 1992; Vousden, Brown, & Harley, 2000).
- on syllable-initial than syllable-final consonants (Cohen, 1973; Vousden et al., 2000).
- on low frequency than high frequency words (Dell, 1990; Stemberger, 1984b; Stemberger & MacWhinney, 1986). This is also true of phonologically related, but not of semantically related word substitution errors (Harley & MacAndrew, 2001).
- on stressed than unstressed syllables (Boomer & Laver, 1968; Garrett, 1975; Nooteboom, 1969; Shattuck-Hufnagel, 1986).
- on between-words errors than within-word errors (Garrett, 1975; Stemberger, 1982, 1989; Vousden et al., 2000).
- on anticipations than perseverations (Boomer & Laver, 1968; Cohen, 1973; Dell et al., 1997; Nooteboom, 1969; J.P. Stemberger, unpublished manuscript, 1989; Vousden et al., 2000). Exchanges are the least frequent of movement errors in all of these studies. However, they are the most frequent category in the Toronto corpus of errors (where anticipations also outnumber perseverations, Dell & Reich, 1981). J.P. Stemberger (unpublished manuscript) argued that this was due to a perceptual

bias, and related it to the fact that this is one of the few multiple-collector corpora in the literature.

- on real word versus nonword outcomes of phonological errors (Dell & Reich, 1981; Baars, Motley, & Mackay, 1975, for experimentally induced errors). J.P. Stemberger (unpublished manuscript) argued that this could also be due to a perceptual bias exacerbated in a multiple-collector corpus. Consistently with this idea, he found only a very small lexical bias in his single-collector corpus.

However, many observed patterns in slip collections run contrary to the predictions of perceptual bias. There are more errors:

- on consonants than on vowels (Cohen, 1973; Nooteboom, 1969; Shattuck-Hufnagel, 1986; Stemberger, 1989).
- sharing a greater number of features or, to put it in another way, where target and error phoneme are maximally similar versus maximally different (MacKay, 1970; **?**, 1980; Shattuck-Hufnagel, 1986; Stemberger, 1982, 1989; Vousden, Brown, & Harley, 2000).
- affecting place of articulation than manner and voicing (MacKay, 1970; **?**, 1980; Stemberger, 1989; van den Broecke & Goldstein, 1980).
- where a less frequent more specified phoneme substitutes for a more frequent less specified phoneme (Stemberger, 1991a,b).
- affecting single phonemes, versus bigger units such as clusters and syllables, and more than lexical level errors, although the difference is usually not large (Bock, 1991; Fromkin, 1971; Shattuck-Hufnagel & Klatt, 1979; Stemberger, 1989).

In her analysis of the first corpus of Spanish speech errors, del Viso (1992) found that word-initial consonants were actually *less likely* to participate in errors than word-medial consonants. Errors were equally likely in stressed and in unstressed syllables (see also Berg, 1991), perseverations were somewhat more frequent than anticipations[4], and there was no trace of a lexical bias in phonological errors (del Viso et al., 1991). Apart from these contrasting characteristics, most patterns observed in their corpus agreed with those reported from English, Dutch and German: word-internal syllable-initial consonants were more likely to be involved in errors than syllable-final consonants (Berg, 1991), there were more between-words than within-word errors, there were more errors on consonants than on vowels, there was a featural similarity effect—more single phoneme errors changed place of articulation than manner, and the fewest errors changed sonority—, and although the asymmetries in phoneme substitutions were not tested against the frequency-underspecification hypothesis, the majority of them are in the expected direction (del Viso, 1992).

It would therefore seem that del Viso's corpus is mostly free from perceptual biases, as most diagnostic signatures are not only absent, but actually show the opposite pattern. Only a few aspects remain that could be affected by perceptual biases: more phonological errors occur between words than within word; syllable-initial positions are affected more often; form-based word substitutions tend to affect lower

[4] At the phonological level, perseverations outnumbered anticipations even when assuming that all incompletes are actually anticipations. At the lexical level, this strategy leads to having more anticipations than perseverations. However, if the proportions of clear anticipations and complete exchanges are used to estimate the number of incompletes that belong to each category (as suggested by Stemberger, 1989), the figures are 97 anticipations to 102 perseverations. A safe conclusion is (against Berg, 1991) that there are not more anticipations than perseverations in del Viso's corpus.

frequency words and result in higher frequency error words (del Viso et al., 1991); lexical level errors slightly outnumber phonological errors (1820 vs. 1694 errors. This difference turns much bigger when only unambiguous errors are considered: 1465 vs. 642 errors); and there is a higher proportion of exchanges than in other single-collector corpora (18% of sublexical non-ambiguous movement errors, most of which are probably single-phoneme exchanges, versus 6% of single-consonant errors in Stemberger's, unpublished manuscript, corpus and 9% in Cohen, 1966, cited in J.P. Stemberger, unpublished manuscript).

The syllable-onset effect, however, coupled with a reversed word-onset effect suggests that its probable cause is not perceptual salience. Likewise, the lack of a frequency effect in semantically related word substitutions suggests that its presence in form-based word substitutions is due to reasons other than bias (i.e., a real word frequency effect at the phonological level, see del Viso et al., 1991; Harley & MacAndrew, 2001; Jescheniak & Levelt, 1994). The predominance of phonological over lexical level errors was never that big in other corpora as to preclude an explanation based in chance variation in del Viso's corpus (see above); and finally, the proportion of phonological exchanges is still far from the figures from Dell and Reich's (1981) multiple-collector corpus (54% of syllable-initial single-consonant errors excluding syllables with initial clusters). Only the preponderance of between-words over within-word errors remains.

The cross-linguistic comparison suggests the following conclusion: if we agree that del Viso's single-collector corpus is mostly free from perceptual biases, it logically follows that the patterns predicted by perceptual bias and observed in English, Dutch and German corpora are probably genuine error patterns, and not due to a confounding with perceptual bias. There are no reasons to believe that linguistically trained and theoretically informed Spanish observers should be less affected by perceptual biases than observers from other languages.

Nevertheless, before claiming that the differences between Germanic languages on one side and Spanish on the other are real cross-linguistic differences, there remains the possibility that at least some of the Spanish patterns are due to personal or sampling biases of the two collectors who compiled del Viso's corpus. A comparison with the present multiple-collector corpus allows an independent replication of Spanish findings with a methodology that is less likely to suffer from any personal bias and more likely to be affected by perceptual biases. Contrasting the two corpora will let us identify real cross-linguistic differences, which will then need to be addressed by any general theory of language production.

In what follows, we will first compare general error patterns between del Viso's and our corpus. We will then turn to a more detailed analysis of a subset of the errors in the present corpus in search for more specific patterns predicted by the perceptual bias hypothesis.

Description of the corpus

Error collection

Every year since 1992, psychology students taking part in the Psychology of Language class at the University of Granada undertook the collection of speech errors. At the moment of writing, a total of 737 students have contributed varying amounts of errors to the corpus (from a minimum of 1 to a maximum of 68). The present size of the corpus is 8,031 errors.

Student observers were trained in listening for errors (attending to the phonology of speech, detecting cues that often follow the error like “I mean” or “sorry”), their collection (how to formulate questions to get information about the wrong and intended utterances), and classification. Their task was to record all speech errors that they might come across from neurologically healthy adults of both sexes. If possible, the students interrogated the speaker at the time of error, to verify his or her records and the intended utterance, to exclude intentional humour, and to resolve ambiguities in classification. They outlined the context in which each error occurred, recording what the speaker just said, what he intended to say, what was in the speaker’s mind (if relevant) and any comments or observations that could be useful later on, such as fatigue or distraction of the speaker. Instructions explicitly emphasized that only errors which were clearly heard and remembered should be collected, and that false beginnings and self-corrections should not be collected, since they do not constitute complete errors.

The information contained in each record comprised: error number, error utterance, intended utterance, error type (see Table 1 for definitions), linguistic level of the error unit (see Table 2), context of the error, any comments that could provide useful

Table 1 Error types and definitions

Non-ambiguous or “ traditional ” error types
Contextual errors: the interacting units come from the surrounding sentential context
Anticipation: An upcoming unit appears earlier substituting another item with no omission of the unit in its intended position
Perseveration: An unit is repeated in another position downstream substituting another unit
Exchange: Anticipation plus perserveration
Shift (4 subtypes): (1) An unit is anticipated not substituting any other without losing the original unit. (2) An unit is persevered not substituting any other without losing the original unit. (3) An unit is anticipated not substituting any other unit but losing the original unit. (4) An unit is persevered not substituting any other losing the original unit
Noncontextual errors: the interacting units do not come from the surrounding sentential context
Omission: One or more units are omitted
Blend: Fusion of two units
Addition: An unit substitutes another item
Substitution: A new item is added between two units
Others
Double substitution or bumper car errors: These can be classified as noncontextual if only the first substitution is taken into account, or as contextual if we focus on the second substitution. As the first one is triggering the second one, we felt that these errors should be considered noncontextual
Rule errors: A syntactic rule is violated
Anticipation/Perseveration: The error can be classified as anticipation or perseveration. Two origins are possible
Derivation: An inadequate suffix is added to a stem
Incomplete: The wrong utterance is not concluded
Mixture: There is more than one error in the utterance
Rare: Ambiguous errors that accept more than one classification, or more than three units are involved

Table 2 Linguistic units and levels

Unit type	General level
Phrase	Supralexical
Word	Lexical
Stem	Morphological
Prefix	
Derivational suffix	
Inflectional suffix	
Syllable	Phonological
Simple onset	
Complex onset	
Rime	
Simple vowel nucleus	
Complex vowel nucleus	
Simple coda	
Complex coda	
Coda-onset group	
Others	

information (e.g., mental context of the speaker), identity of the registrar, identity of the person who made the error, his or her educational level, age, date, conscious detection of the error by the speaker, and interval between perception and registration. Each error was later reviewed by the authors, and then introduced into a computerised database. The following fields were also included in the database: general level of the error unit (see Table 2), contextual/noncontextual, doubts (used to mark errors that admit more than one classification and to note down comments), and identity of the person who revised the error. The design of the database takes into account that each error utterance may contain more than one error, and more than one error unit may be involved, allowing a maximum of 27 fields per error utterance.

Control procedures

The control procedures were carried out by the authors, with the general goal of including only "trustable" and precisely recorded errors into the corpus. Errors with poorly registered or inaccurately recorded context, with typographical errors, or implicating more than one language were eliminated. Most errors in which it was not possible to consult the source speaker after the error (e.g., those observed in TV programs) were also eliminated because the intended utterance could not be clearly ascertained.

An important strategy used at this stage is to discriminate between "good" and "bad" collectors (inaccurate listeners, with imprecise or incorrect registration of errors, usually related to a low motivation towards the study). All errors provided by "bad" collectors were eliminated, even though some of them meet the standards. The authors collected and checked the "error recording sheets" in weekly meetings and answered questions from the students. Before the information was introduced in the database, all errors were fully revised again.

Speech error classification

Errors were classified along four dimensions: contextual versus noncontextual, ambiguity of the error, type of error, and error level or linguistic units involved. Ambiguity refers to how easily the error can be included in any of the traditional error types: anticipation, perseveration, exchange, shift, substitution, addition, omission and blend (in our database, the category of shifts is further subdivided into four different types: anticipatory shift with or without losing the source unit, and perseveratory shift with or without losing the source unit). If the error does not fit clearly into any of these categories, it is classified in the category '*Others*'. This category contains some error types that are well-defined but are not part of the traditional categories: derivation (improper morphological derivation or inflection of a morphologically complex word), rule errors (misapplication of syntactic rules), and double substitution or "bumper car errors" (errors in which the first substituted unit shows up later in the sequence substituting for another unit, e.g., "si quieres te dejo una CAMISETA para la TOALLA ('if you want I can lend you a T-SHIRT for the TOWEL') [intended: si quieres te dejo una TOALLA para la PISCINA ('if you want I can lend you a TOWEL for the SWIMMING POOL')]"; see Stemberger, 1985). Other types in this category are clearly ambiguous because they could fit into more than one traditional type (anticipation/perseveration), they affect more than one type of linguistic unit or there is more than one error in a single utterance (mixture errors), or they cannot be clearly classified into any other category (rare errors).

The incomplete category of ambiguous errors deserves some comment: while this category is standardly included in other corpora, our instructions to observers precluded them from collecting incomplete errors. As a result, this category has a frequency close to zero in the present corpus.[5]

Table 3 shows examples of many of our error categories.

General comparison with del Viso's (1992) corpus

This section presents the main error patterns observed in our corpus side by side with those observed in del Viso's corpus. Because the classification system is slightly different between del Viso (1992) and our error database (see Pérez, 2002), we recategorized many errors to comply with del Viso's error taxonomy. Some errors were eliminated because they could not be included in any of del Viso's categories. The resulting classification system is shown in Table 4. Henceforth the present corpus (Pérez, Santiago and Palma) is referred to as PSP and del Viso's is referred to as dV. After adjusting the two corpora, corpus PSP has a total of 7,480 non-ambiguous errors, while corpus dV contains 1,883 errors.

Binomial tests were used to find significant differences between categories within each corpus (a two-tailed 0.05 probability level was always used). If the same categories were significantly different within both corpora, and the direction of the difference was the same, we concluded that they shared the same pattern over the relevant set of categories. When direct cross-corpus comparisons were needed, a chi-squared statistic was used.

[5] All nine errors currently categorised as incompletes are the result of reclassification during the review process of errors included in other categories.

Table 3 Error categories and examples

Error category	Example	Example translated
Contextual, lexical level, word perseveration	a mí me gustan más los GATOS que los GATOS (a mí me gustan más los GATOS que los PERROS)	I like CATS better than CATS (I like CATS better than DOGS)
Contextual, morphological level, stem exchange (stranding error)	voy a estudiar FISIOlogía PSICOlógica (voy a estudiar PSICOlogía FISIOlógica)	I am going to study PSYCHOlogical PHYSIOlogy (I am going to study PHYSIOlogical PSYCHOlogy)
Contextual, phonological level, simple coda exchange	llevamos una paRcaNta al viaje (llevamos una paNcaRta al viaje)	We are taking a banner to the trip
Contextual, phonological level, simple onset anticipation	a mari Caz la vi en el Coche (a mari Paz la vi en el coche)	I saw mari Caz in the Car (I saw mari Paz in the car)
Contextual, phonological level, complex vowel nucleus anticipatory shift without losing the source item	aparte de cogIer unos cIegos (aparte de coger unos cIegos)	Besides getting drunk
Contextual, phonological level, complex vowel nucleus perseveratory shift losing the source item	me ves en la sanUa (me ves en la saUna)	You will see me in the sanUa (you will see me in the saUna)
Noncontextual, supralexical level, sentence blend	a mí las serpientes NO ME DAN GRACIA (a mi las serpientes NO ME HACEN GRACIA/ME DAN MIEDO)	I do not like snakes
Noncontextual, phonological level, syllable omission	dame el cullo (cuCHIllo)	Give me the knife
Others, contextual, phonological level, simple vowel nucleus anticipation/perseveration	sOn judíOs con jamÓn (son judíAs con jamón)	These are beans with ham
Others, contextual, lexical level, word incomplete	son SILLAS… (son CÓMODAS estas sillas)	Are CHAIRS…. (these chairs are COMFORTABLE)
Others, contextual, phonological level, mixture (in this case, more than one error occurs in the utterance)	cual es el archipLiéGago mas grande del mundo? (cual es el archipiéLago mas grande del mundo?)	Which one is the biggest archipLeGago in the world? (which is the biggest archipelago in the world?)
Others, noncontextual, morphological level, derivational suffix derivation error	solo el aprendiMIENTO vicario (solo el aprendiZAJE vicario)	Only vicarious learning
Others, noncontextual, lexical level, double word substitution	si quieres te dejo una CAMISETA para la TOALLA (si quieres te dejo una TOALLA para la PISCINA)	If you want I can lend you a T-SHIRT for the TOWEL (if you want I can lend you a TOWEL for the SWIMMING POOL)
Others, Rare (multiple possible categorizations)	esa nina es la que PIANABA (esa niña es la que TOCABA EL PIANO)	That girl is who was playing the piano

The first column contains information about the error nature (contextual, noncontextual, or other), general level (phonological, morphological, lexical, supralexical), unit involved (coda, word, syllable, etc.), and type of error (anticipation, exchange, substitution, etc.). The second column contains the wrong utterance and in brackets the intended utterance. The third column contains a rough translation of the speech error when possible

Table 4 Classification system used for comparing del Viso's and the present corpora

Contextual	Noncontextual
Anticipation	
Perseveration	Blend
Exchange	Omission
Anticipation/Perseveration	Substitution
Shift	Addition

Table 5 Distribution of errors by general level (data are given in percentages)

	General level		
	Sublexical	Lexical	Supralexical
Present corpus (PSP)	30.01	63.11	6.98
del Viso's corpus (dV)	28	65.32	6.68

Comparison of general distributional patterns

Contextual versus noncontextual errors

Corpus PSP has 3,233 (43.22%) contextual and 4,247 (56.78%) noncontextual errors. Corpus dV contains 864 (43.97%) contextual and 1,101 (56.03%) noncontextual errors. Across corpora, these patterns are nearly identical. The fact that there are fewer contextual than noncontextual errors in dV is completely due to the elimination of incomplete errors (including this category, there are exactly 50% of contextual and 50% noncontextual errors in dV, see del Viso, 1992, tables 4.2 and 4.3). We therefore estimate the rate of incompletes in our corpus the same as in dV, about 6% of total errors.

Comparison by error unit

Both corpora show the same pattern across error levels ($p > .05$), again with very similar proportions (see Table 5). The percentage of errors at the lexical level is highest, followed by the sublexical level, and finally the supralexical level. All differences between categories are significant and in the same direction in both corpora: the lexical level outnumbers by far both the phonological and the supralexical levels.

A higher detectability of lexical level errors is predicted by the perceptual bias hypothesis, and runs contrary to observed patterns in Germanic corpora, where phonological errors usually outnumber lexical errors (see Introduction). However, the fact that the proportion of lexical errors is similar across the Spanish corpora suggests that it is the result of factors other than bias, perhaps a symptom of differences in underlying production mechanisms in Spanish and Germanic languages.

Error types in contextual errors

The main discrepancies in the distribution of error types within contextual errors were found in the amount of complete anticipations and perseverations (see Fig. 1). While PSP has slightly more anticipations (22.67%) than perseverations (20.66%), dV presents the opposite pattern (10.88% anticipations vs. 34.49 % perserverations). This difference seems to be due both to a greater proportion of anticipations in PSP as well

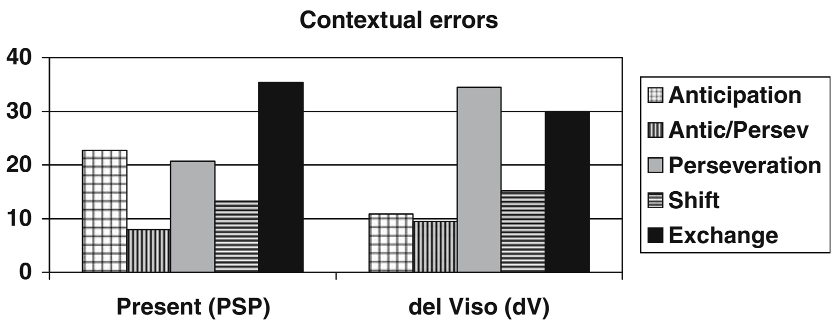

Fig. 1 Distribution of error types in contextual errors (data are in percentages)

as to a smaller proportion of perseverations (anticipations: $\chi^2(1) = 14.37, p < .01$; perseverations: $\chi^2(1) = 8.46, p < .01$). Because the proportions of other categories do not vary significantly across corpora (all $p > .05$), these differences seem to be linked to each other.

Distribution of contextual error types within the sublexical level

The same contrasting pattern of anticipations and perseverations appears when only the sublexical level is considered (see Fig. 2). Anticipations and perseverations are observed in comparable proportions in PSP whereas perseverations are more frequent in dV than anticipations. The difference between PSP and dV regarding anticipations is significant ($\chi^2(1) = 9.41, p < .01$) but the difference between perseverations is not ($p > .05$). However, at this level there are other differences in error proportions which may be taken to indicate that the same pattern of increased perseveratory errors is observed in dV. The main candidate is the anticipation/perseveration category, which is more frequent in dV ($\chi^2(1) = 4.04, p < .04$). dV also shows a lower proportion of shifts than PSP ($\chi^2(1) = 4.92, p < .02$). Either more perseverations in dV fell by chance following contexts where the interacting phoneme is repeated, or some perseveratory shifts in PSP displaced a surrounding phoneme, or both.

A perceptual bias in PSP could be the cause of the higher proportion of anticipations. Shifts in turn might be a more noticeable category of errors than perseverations,

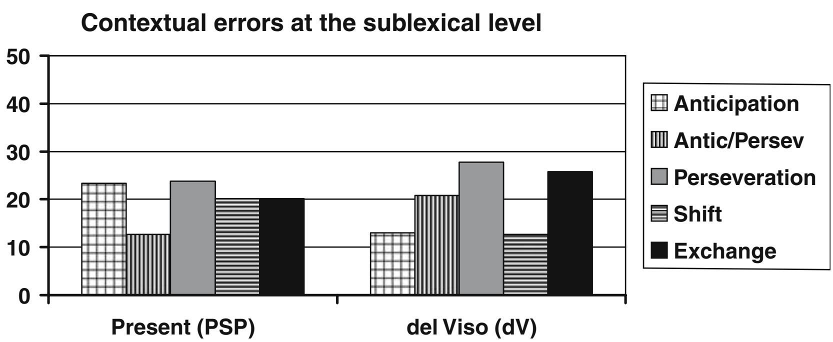

Fig. 2 Distribution of error types in contextual errors at the sublexical level

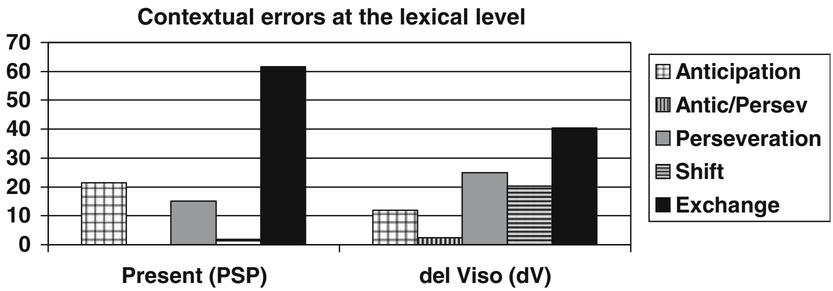

Fig. 3 Distribution of error types in contextual errors at the lexical level

so that the bias could also be responsible for their greater number in PSP. However, a perceptual bias acting in PSP would also predict a concomitant increase in the proportion of exchanges. This proportion is, to the contrary, lower than in dV (although actually not significantly different, $p > .05$, and relatively high in both corpora). We will delay a detailed discussion of this phenomenon until the General Discussion section.

Distribution of contextual error types within the lexical level

As Fig. 3 shows, at the lexical level there is a predominance of exchange errors in both corpora. There are also more anticipations than perseverations in PSP and fewer in dV. The proportion of anticipations is actually greater in PSP than in dV ($\chi^2(1) = 8.27$, $p < .001$) and the proportion of perseverations is smaller ($\chi^2(1) = 5.23$, $p < .02$). Moreover, the proportion of exchanges is clearly higher in PSP than in dV ($\chi^2(1) = 18.92$, $p < .01$). This pattern is consistent with a perceptual bias being at work in both corpora, and to a greater extent in PSP. However, sublexical errors are more difficult to detect than lexical errors. If there were a perceptual bias, it should be clearer at the sublexical level, which is not the case.

In another contrast with sublexical errors, PSP showed a much smaller proportion of shifts ($\chi^2(1) = 20.98$, $p < .01$), suggesting that the increased proportion observed at the sublexical level was not linked to the increase in anticipations nor to a putative perceptual bias.

In general, perseverations outnumber anticipations in del Viso's corpus both at the lexical and sublexical levels. In the present corpus, anticipations are equally or somewhat more frequent than perseverations at both levels. Comparing across corpora, this difference is due to changes in both error categories (i.e., there are more anticipations and less perseverations in PSP than in dV). In the General discussion section we will touch on this issue in more detail.

Distribution of contextual error types at the supralexical level

Patterns in supralexical contextual errors look quite alike in both corpora: the two categories with a non-null frequency in dV differ in the same direction as in PSP (see Table 6). However, the reported percentages come from a very small total number of errors (only 10 non-ambiguous errors in corpus dV), so we did not compare them statistically.

Table 6 Distribution of error types affecting supralexical units (data are given in percentages)

	Contextual error				
	Anticipation	A/P	Perseveration	Shift	Exchange
Present corpus (PSP)	31.25	3.125	28.125	3.125	34.375
del Viso's (dV)	0	0	0	10	90

Error types in noncontextual errors

The observed patterns across noncontextual error types are remarkably similar between the two corpora (Table 7), especially at the sublexical level (Table 8). At the lexical level both corpora show a predominance of word substitutions over all other error types, which is even more pronounced in dV than in PSP, as Fig. 4 shows. This difference is statistically significant ($\chi^2(1) = 4.52$, $p < .03$). As lexical word substitutions are among the most noticeable of errors, this again speaks against a perceptual bias affecting PSP more strongly than dV. The only other difference between the corpora is in the proportion of blends, which is greater in PSP ($\chi^2(1) = 11.22$, $p < .01$). Supralexical patterns are again very similar in both corpora, sharing a predominance of sentence blends over other error types, which appear with very low frequencies (Table 9).

In the following section we focus on more specific analyses addressed to several predictions from perceptual bias on phonological speech errors.

Testing specific predictions from perceptual bias at the phonological level

As part of an on-going research project on syllable frequency effects in phonological speech errors (Santiago, Pérez, Palma, and Stemberger, under review), the set of all phonological errors in a prior stage of our corpus was selected. Total *N* was 1,477 errors (current total number of phonological errors is 2,244 errors, which is the set used in the general analyses above). Because of requirements of that study, errors changing the number of syllables of the affected word (93 cases) or the number

Table 7 Distribution of error types in noncontextual errors (data are given in percentages)

	Noncontextual error type			
	Blend	Omission	Substitution	Addition
Present corpus (PSP)	21.52	10.10	67.36	1.01
del Viso's (dV)	13.80	13.44	70.57	2.17

Table 8 Distribution of noncontextual error types at the sublexical level (data are given in percentages)

	Noncontextual sublexical error type		
	Omission	Substitution	Addition
Present corpus (PSP)	52.03	37.55	10.40
del Viso's (dV)	53.93	43.03	3.03

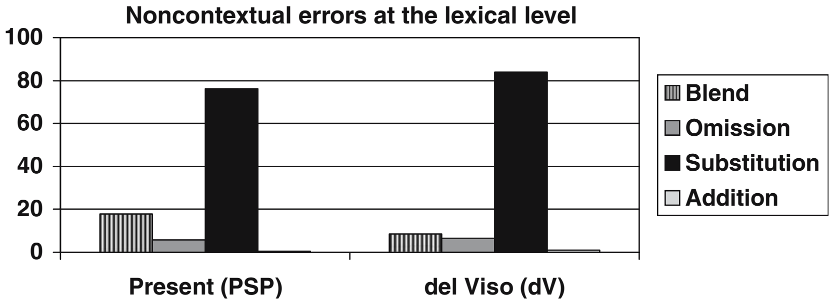

Fig. 4 Distribution of noncontextual errors at the lexical level

Table 9 Distribution of noncontextual error types at the supralexical level

	Noncontextual supralexical error type			
	Blend	Omission	Substitution	Addition
Present corpus (PSP)	20.40	58.77	23.46	17.14
Del Viso's (dV)	27.4	96.42	0	0

Table 10 Error types included in the present analyses and their frequency

Error type	Cases	Cases of unambiguous origin
Anticipations	375	218
Perseverations	327	128
Exchanges	220	220
Noncontextual substitutions	57	–
Total	979	566

of phonemes of the target syllable were discarded (all 254 shifts, 16 additions and 42 omissions). Finally, errors involving more than one phoneme were also discarded (459). Table 10 shows the types and frequency of errors included in the analysis.

We used this dataset to establish whether certain specific phonological error patterns found in del Viso's (1992) corpus which are relevant to the perceptual bias hypothesis are replicated in the present corpus. We will first assess patterns which agree with those reported for Germanic languages: (1) more phonological errors between than within words, consistent with perceptual bias; and (2) higher error rate on consonants than on vowels, which runs contrary to perceptual bias. We will then proceed to test some crucial patterns that have been reported only for Spanish, all of them counter-indicated by the perceptual bias hypothesis: (1) less implication of word-initial consonants in errors; (2) equal propensity to errors on stressed and unstressed syllables; and (3) lack of lexical bias in single-phoneme movement errors.

Throughout, we will employ a lexicographic tool developed by Santiago, Pérez, Palma, and Stemberger (under review). We phonologically transcribed words from a recent Spanish lexical database and frequency count (Alameda & Cuetos, 1995: 81,313 word types from a sample of about 2 million word tokens), divided them into syllables, and located their primary stress. This tool, which we refer to as the Granada Lexical DataBase (GRLDB), allows us to estimate chance probabilities. Berg (1991) followed a standard practice of using the target words in the error corpus to estimate

Table 11 Total number and percentages of errors in each of six word positions, for words ranging in length from 1 to 6 syllables

Total number of syllables	Error syllable						
	1	2	3	4	5	6	Total
1	39 (3.24%)	–	–	–	–	–	39 (3.24%)
2	165 (13.71%)	280 (23.27%)	–	–	–	–	445 (36.99%)
3	100 (8.31%)	162 (13.46%)	119 (9.89%)	–	–	–	381 (31.67%)
4	21 (1.74%)	75 (6.23%)	89 (7.39%)	57 (4.73%)	–	–	242 (20.11%)
5	5 (0.41%)	18 (1.49%)	26 (2.16%)	12 (0.99%)	9(0.74%)	–	70 (5.81%)
6	1 (0.08%)	9 (0.74%)	6 (0.49%)	5 (0.41%)	2(0.16%)	3 (0.24%)	26 (2.16%)
Total	331 (27.51%)	544 (45.22%)	240 (19.99%)	74(6.15%)	10 (0.83%)	3 (0.24%)	1203 (100%)

The anticipatory and perseveratory sides of exchanges have been counted independently

chance. Both approaches have advantages and disadvantages, and we will use chance estimations from both sources.

Between word versus within word errors

Of the 979 single-phoneme errors in the present dataset, the subset of contextual errors for which the source phoneme could be unambiguously located in the environment was selected (see Table 11). A total of 347 of 566 errors spanned two different words (61.3%), while 219 were within-word errors (38.69%). This is a significant difference from a 50–50 distribution ($\chi^2(1) = 5.11$, $p < .05$), showing that phonological errors tend to occur between words more often than within words. However, it is not clear what should be taken as the chance level for this comparison. These percentages are not significantly different from those reported by del Viso (1992): 57% between-word versus 43% within-word errors ($p > 0.05$).

Consonant versus vowel errors

There were 686 errors in the error set involving single consonants (70%) versus 295 (30%) involving single vowels. The question arises whether consonants are involved to a greater extent in errors just because there are more consonants than vowels in Spanish. In GRLDB there were a total of 8,940,031 phoneme tokens, out of which 46.9% were vowels and 53.05% consonants. Consonants are more often implicated in speech errors and vowels less often than would be expected by chance ($\chi^2(1) = 11.50$, $p < .01$).

Word position effects

As shown in Table 12 and Fig. 5, word onsets are protected against speech errors in Spanish. For all word lengths, word-initial syllables were always less error-prone than word-medial syllables. The latter were in turn more error-prone than final syllables. These error distributions were significantly different from a flat distribution at all levels of word length except for trisyllables (two syllables: $\chi^2(1) = 6.68$, $p < .01$; three syllables: $\chi^2(2) = 4.17$, $p > .05$; four syllables: $\chi^2(3) = 17.72$, $p < .01$; five syllables: $\chi^2(4) = 27.55$, $p < .01$; six syllables: $\chi^2(5) = 38.47$, $p < .01$). To calculate the

Table 12 Number of single-phoneme errors as a function of syllable position

Syllable position	Total cases	Cases in bisyllabic words	Corrected cases
Onset	850	312	312
Coda	35	9	19.8

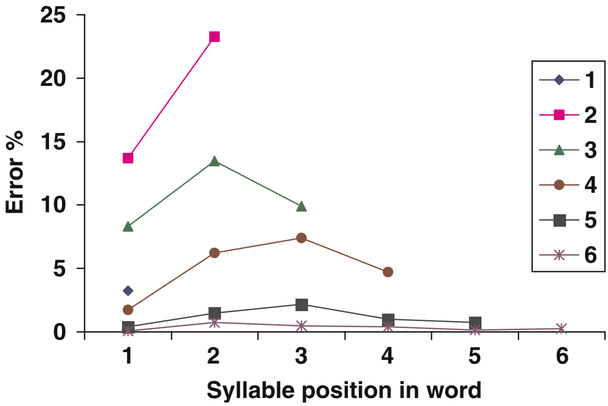

Fig. 5 Error percentages as a function of syllable position in word. A different line is used for words ranging in length from 1 to 6 syllables

probability of an error occurring on an initial syllable just by chance, we need to know the distribution of words of different lengths in Spanish. By looking at target words in del Viso et al. corpus, Berg (1991) estimated that errors should occur on the initial syllable 40% of the time by chance. He observed 27.3% errors in word-initial position, very close to the 27.51% obtained in the present data set. Estimating chance from GRLDB is confounded by the fact that monosyllabic words are included within word-initial syllables in the database. They are often very high frequency function words, which are known to participate in errors rarely (Garrett, 1975, 1980). If all word tokens are included in the calculation, there are a total of 3,936,605 syllable tokens in the database, 49.5% of which are word-initial syllables. The observed proportion of 27.51% word initial errors is significantly different from this chance level ($\chi^2(1) = 19.34, p < .01$). If monosyllables are pulled out, the chance figure goes down to 35.13%, and the difference between this and the observed proportion is nonsignificant ($\chi^2(1) = 2.54, p > .05$). However, the chance level is underestimated in this case, as monosyllabic content words are also discarded. Finally, the observed proportion of word initial errors in the present corpus is also significantly lower than Berg's estimated 40% chance level ($\chi^2(1) = 6.50, p < 0.05$). We conclude that word-initial syllables in Spanish are not more prone to phonological errors than word-medial syllables, and that quite probably they are actually less prone to errors.

Table 10 shows that errors also tended to be syllable-initial in the present error set. The wide majority of single-phoneme errors were concentrated in onset consonants. As del Viso (1992) cautioned, the lower rate of errors on coda consonants could just be a result of the predominance of open syllables in Spanish. Berg (1991) calculated that the ratio of open to closed syllables in Spanish bisyllabic words (after correcting also for onset-less syllables) is 2.2–1. When the same procedure is applied to the GRLDB

Table 13 Number of errors as a function of stress level in target and source syllables

Error syllable	Source syllable	
	Stressed	Unstressed
Stressed	116	50
Unstressed	69	160

word pool, a ratio of 2.15–1 obtains. We tabulated separately errors in bisyllabic words as a function of syllable position and corrected the post-vocalic total using the highest proportion (Berg's). As shown in Table 12, after correcting for the number of coda errors, errors affecting onset consonants still outnumbered coda errors by a ratio of 15.75–1. This is a much greater corrected ratio than the 2.1–1 found by Berg (1991).

Although at present it is unclear why the observed ratios are so different, it can be safely concluded that the patterns first reported by del Viso and co-workers (1992; García-Albea et al., 1989) are actual linguistically specific characteristics of Spanish speech errors: Spanish errors occur preferentially on word-medial syllable-initial positions. This contrasts sharply with patterns reported for Germanic languages (e.g., Shattuck-Hufnagel, 1987), and bears important consequences both for the question of perceptual bias and cross-linguistic differences in phonological encoding mechanisms.

Stress level effects

For the analysis of stress-level effects, we will focus on the subset of unambiguous between-word errors (see Table 13). Spanish has only one stressed syllable per word, therefore within word errors would not allow us to compare if stressed syllables interact with other stressed syllables. Overall, there were 166 (42.02%) errors on stressed syllables (once again, anticipatory and perseveratory sides of exchanges are counted independently), and 229 (57.97%) errors on unstressed syllables, very close to (and not significantly different from) the 40% and 60% figures reported by Berg (1991). Once again, this could be attributed to a higher prevalence of unstressed syllables in Spanish words. Estimating from Berg's (1991) Table 8, Spanish words contain 36.93% stressed and 63.06% unstressed syllables. Nearly identical figures are found when using GRLDB, providing that monosyllables are excluded:[6] 39.34% stressed vs. 60.65% unstressed syllables. The observed error percentages do not differ significantly from chance estimations ($p > .05$), supporting the suggestion than Spanish speech errors are oblivious to stress level (Berg, 1991; del Viso, 1992).

Table 13 cross-tabulates errors as a function of target and source syllable stress level, showing what looks like a compelling association between the two factors. However, it is difficult to deny Berg's (1991) argument that this is no more than an artefact of word-position. When we look only at errors occurring in same versus different positions in target and source words, independently of stress level, we find nearly the same diagonal totals: 253 errors involve syllables in the same positions in their target and source words, versus 142 errors in different positions. On the whole, it seems safe to conclude, with Berg (1991), that stress does not play any significant role in constraining Spanish slips of the tongue.

6 Monosyllabic function words, i.e., the majority of monosyllables in Spanish, are phonologically enclitized to nearby content words, thereby losing their primary stress.

Table 14 Percentages of different types of single-phoneme errors resulting in words versus nonwords (anticipatory and perseveratory sides of exchanges have been counted independently)

	Observed			Word Chance
	Total	Nonword	Word	
Exchange	444 (100%)	356 (80.18%)	88 (19.81%)	38.16% (antic: 31.94%; persev: 44.39%)
Anticipation	375 (100%)	284 (75.73%)	91 (24.26%)	22.84%
Perseveration	327 (100%)	251 (76.75%)	76 (23.24%)	25.50%
Noncontextual substitution	57 (100%)	40 (70.17%)	17 (29.82%)	–
Total	1203 (100%)	931 (77.38%)	272 (22.61%)	28.83%

Lexical bias

A final test of the perceptual bias hypothesis concerns lexical bias in single-phoneme phonological errors in the present corpus. Table 14 shows error percentages resulting in existing words versus nonwords. The criterion to consider an error string as a Spanish word was its presence in GRLDB. In general, our results show even lower proportions of word results than those reported by del Viso, Igoa, & García-Albea (1991): 35% words vs. 65% nonwords. This strongly suggests the absence of any lexical bias effect in our data.

We followed the same procedure as del Viso, Igoa, & García-Albea (1991), del Viso (1992) and also Dell & Reich (1981) to estimate how likely errors are to create words by chance. Only single-phoneme between-words movement errors were used. Errors involving single consonants within clusters as well as words with more than one error were discarded. We decided to make independent counts for syllable-onset and vowel nucleus errors in different word positions, instead of limiting ourselves to word-initial consonant errors, as Dell & Reich (1981) did. This would have left us with too few cases given the low probability of word-initial errors in Spanish. Syllable positions represented by less than three cases were not included. The resulting sample sizes were: 239 anticipations, 221 perseverations, and 35 exchanges.

It should be pointed out that the reason why exchanges seem underrepresented in this sample is that the majority of them occur within a single word, while anticipations and perseverations occur mainly between words. This pattern replicates observations from del Viso and co-workers (García-Albea et al., 1989, Table 4) and contrasts with observations from English (Stemberger, 1982, footnote 8), suggesting that it probably is another real cross-linguistic difference between English and Spanish that needs to be addressed by theories of phonological encoding.

As shown in Table 14, observed percentages of phonological errors resulting in real words were lower than expected for most error types, significantly so for exchanges (exchanges: $\chi^2(1) = 14.26$, $p < .01$; anticipations: $p > .05$; perseverations: $p > .05$). Although the origin of this anti-lexical bias in exchanges is unclear, we can safely conclude that no lexical bias is present in our corpus of Spanish slips of the tongue, again replicating results reported by del Viso and co-workers (del Viso, 1992; del Viso et al., 1991).

In general, these results agree in all respects with prior reports on Spanish speech errors by del Viso and co-workers (1992; García-Albea et al., 1989; del Viso et al.,

1991) and Berg (1991). However, a recent study (Anton-Mendez et al., 2005) claims a lexical bias effect in dV's corpus and also in experimentally elicited Spanish errors by using a different way of constructing the null hypothesis (chance probability = real possible words/total possible substitutions). Nevertheless, these results have not been yet published in a peer-reviewed journal, so further analysis should be taken to clarify any possible effect of methodological dependence in either of the procedures used.

The following General Discussion section addresses the significance of the present results for the validity of speech error corpora and the issue of cross-linguistic differences in language processing.

General discussion

The main overall conclusion that may be drawn from the analysis and comparison of several error patterns between del Viso's (1992) and the present corpus is that similarities are the rule. This conclusion might look surprising given the differences in corpus size (7,480 vs. 1,883 errors after cross-adapting the corpora for comparison) and, more importantly, collection methodology. The present corpus used a multiple-collector methodology, with 737 independent theoretically naïve observers providing errors from a wide variety of everyday situations, speakers, communicative goals and social and economic backgrounds. By contrast, 90–95% of all errors in del Viso's corpus were collected by only two highly trained theoretically informed observers. Perceptual biases, if there is any, should affect much more clearly the present corpus. Personal and sampling biases, if there is any, should affect the later, while they should be absent in the former. In what follows, we will first review the few differences found between corpora and discuss their possible causes. We then compare their common patterns with those found in Germanic languages, and discuss its implications for the question of the reliability and validity of naturalistic speech error collections. Finally, we briefly touch upon the question of cross-linguistic comparisons between Romance (Spanish) and Germanic languages.

As it follows from the literature reviewed in the Introduction, it appears that del Viso's corpus is affected by perceptual bias to a greater extent than the present corpus: the rate of lexical level errors to phonological errors, and the proportions of phonological exchanges and whole-word substitutions are greater in del Viso's, all of which belong to the more detectable error types. Only the higher proportion of word exchanges in the present corpus points in the opposite direction, as exchanges are a very salient type that should be overrepresented as a result of perceptual bias. Differences in error salience, however, should have stronger effects on errors that are more difficult to notice such as phonological errors, which renders a perceptual bias explanation of this pattern less attractive.[7]

The difference in the proportions of anticipations and perseverations between the corpora deserves a more detailed comment. del Viso (1992) reports a high incidence of perseverations and a low incidence of anticipations both at phonological and lexical levels. In our corpus anticipations are equally, or somewhat more frequent, than perseverations at both levels. A first explanation suggests a perceptual bias in the

[7] There are also more shifts at the sublexical level in the present corpus and more at the lexical level in del Viso's (1992) corpus. However, we have no base from speech perception and error detection studies to estimate the relative discriminability of shifts, and therefore, their relevance for the perceptual bias hypothesis.

present corpus, by which anticipations (the more salient type) are detected more often, whereas perseverations are systematically missed. The main problem for this hypothesis is the lack of correlation between this trend and the proportion of exchanges at the phonological and lexical levels (see Figs. 2, 3). If del Viso's (1992) observers were listening more carefully for errors, they should also have caught a smaller proportion of exchanges (another very salient type), and this would have been clearer at the phonological level. Instead, the proportion of phonological exchanges in del Viso's corpus is greater than in our corpus. Another problem is that the preponderance of anticipations over perseverations in the present corpus should be clearer at the phonological level than at the lexical level. On the contrary, anticipations outnumber perseverations only at the lexical level, while they are both equally frequent at the phonological level.

A second possible explanation is the existence of a personal bias, a sampling bias, or both in del Viso's corpus. If one of her two observers was more "tuned" to hearing perseverations, and less "tuned" to anticipations, the observed pattern would result. Alternatively, it is known that any factor that tends to make speech more error-prone will decrease the incidence of anticipations relative to perseverations (Dell et al., 1997; Schwartz, Saffran, Bloch, & Dell, 1994; this phenomena was called "Dell's Law" by Levelt, Roelofs, & Meyer, 1999). If the sample of people from whom errors were collected was small (as it probably was), and some of them had a greater error rate for any reason (e.g., a personal style characterized by a fast speaking rate), a lower anticipatory proportion is expected. However, when all incomplete anticipations in del Viso (1992, Table 4.2) are added to complete anticipations, as Dell, Burger, & Svec (1997) did, the proportion that obtains is .58. This is close to the .60 proportion estimated by Dell, Burger, & Svec (1997) for the adult corpus of Stemberger (1989). In other words, it is within the normal range of values found for English adults. Taking the complementary stance: if we calculate the anticipatory proportion in del Viso's corpus using only complete anticipations, it comes down to .23, which is even lower than the proportion observed for Schwartz, Saffran, Bloch, & Dell (1994) patient FL (.32). But so it is in Stemberger's (1989) adult data, where it turns .32, or in **?**, where it becomes .35. The figures for complete anticipations and perseverations in del Viso (1992) are therefore within the normal range.

The latter reasoning suggests that the key factor underlying the differing proportions of anticipations and perseverations in the two Spanish corpora has to do with the incomplete errors. If we follow Stemberger's (1989) suggestion of using numbers of complete anticipations and exchanges to estimate the proportion of incomplete errors, which are actually anticipations and exchanges, and correct the number of anticipations accordingly, the resulting anticipatory proportion in del Viso's corpus is .37. This is at the lower level of Dell et al.'s (1997) range, comparable to their unpractised tongue twisters. We may then assume for a moment, that the "real" anticipatory proportion for Spanish is somewhere between .37 and .58, including incomplete errors. In the present corpus, incomplete errors were not collected (see Speech Error Classification section). The anticipatory proportion calculated only over complete anticipations and perseverations is .52, much higher than that observed in del Viso (1992) but also in Stemberger (1989) and **?**. This suggests the possibility of a different potential bias: a bias towards categorizing some incomplete errors as complete renderings. This can be considered a form of perceptual or memory bias, as observers probably tend to mentally complete fragmentary utterances, not notice the disfluency and self-corrections associated with them, and/or forget them before notating the error. Those "completed" incomplete errors would then go to increase the numbers of complete anticipations and exchanges, but not of perseverations, raising the anticipatory proportion. This is probably why, overall, the proportion of these two types of errors is greater in our corpus than in del Viso's. It is not a matter of a higher detectability of anticipated or exchanged units in themselves, but a tendency to hear complete utterances even when the speaker stops nearly immediately after the error (or the first component of the error).

Nevertheless, this "completion bias" hypothesis is not free of drawbacks. The main ones are similar to those posed for the perceptual bias explanation above: its effects should vary with error level in very specific ways. Completion bias might increase the number of complete anticipations at phonological and lexical levels to the same extent, after all it is just a matter of disregarding a disfluency (sometimes also a self-correction) on the part of the speaker. However, it should increase the number of phonological exchanges to a greater extent than lexical exchanges. Completing an exchange at the lexical level probably generates a much clearer awareness of active reconstruction than completing a phonological exchange, which should lead to the detection of the error as an incomplete. However, as Figs. 2 and 3 show, the effects of this potential completion bias are seen more clearly at the lexical level in our data than at the phonological level, both on the proportions of anticipations and exchanges. Until this issue receives further investigation, we must leave the question unsettled.

From looking at the differences between the two Spanish corpora, we can conclude that the symptoms of perceptual bias are, if anything, clearer in del Viso's corpus. There are some traces of a possible perceptual (or completion) bias working in our corpus, but the evidence is inconclusive. As perceptual biases of any kind were expected to affect the present corpus to a greater extent than del Viso's corpus, we feel safe to conclude that perceptual bias has, at most, a very limited influence on our corpus, and a still more limited influence in del Viso's corpus. On the other hand, no traces were found of personal or sampling biases operating on the latter corpus.

The reliability of naturalistic speech error corpora against perceptual biases is also supported by our replication of several other patterns attested by del Viso and co-workers, which run contrary to expectations from perceptual biases. We now turn to discuss them in turn.[8]

A higher proportion of errors on consonants than on vowels is consistently found in all published studies of naturalistic speech errors (see Introduction). Its replication using a multiple-collector methodology assures that this is a solid production fact. Together with multiple reports (also from Spanish) that target and error phonemes tend to share a greater number of features, and that target and error phonemes tend to vary in the most confusable feature (place of articulation), we can conclude that error detection does not seem to be affected by the identity of the phonemes involved.

Regarding the position of those phonemes in the speech stream, we also replicated the word position effects reported by del Viso (1992; García-Albea et al., 1989) for Spanish. As discussed in the Introduction, the fact that errors concentrate on syllable-initial positions, coupled with a *lower* proportion of errors in word-initial positions, makes it unlikely that the pattern is being caused by the higher detectability of errors in those positions. We also have shown that the location of interacting phonemes in stressed syllables, which probably makes those errors more noticeable, does not

[8] For the sake of clarity, in the following discussion we will avoid repeating the detailed lists of references given at the Introduction.

influence, to any significant extent, the amount of errors reported, as stress does not have an effect on error rates in either Spanish corpus. The fact that these patterns are replicated with a methodology that is more prone to perceptual biases assures once again that the reported results are quite robust.

Finally, the lexical status of the result of the error does not affect the discriminability of phonological errors in any of the Spanish corpora. Although one can never be certain of negative conclusions, replicating this null result with a methodology that made it more likely to appear (and even finding an anti-lexical bias for exchanges), increases our confidence in the absence of a lexical bias effect in Spanish speech errors. It is quite possible that lexical biases are greatly reduced in naturalistic collections compared to experimentally induced slips also in English (see J.P. Stemberger, unpublished manuscript). It is an open question whether a lexical bias can be found for Spanish in a laboratory situation. What is clear is that perceptual factors are not strong enough to bias a naturalistic corpus in a direction congruent with it.

Two patterns which are expected from perceptual bias remain: a greater proportion of lexical than phonological errors in both Spanish corpora, and a higher number of between word versus within word errors in all the published studies which have looked at this variable. A possible reason for the former is a confusion with ambiguity of the error. del Viso (1992) found nearly the same percentages of phonological (46.91%) and lexical errors (53.08%) in her whole corpus ($N = 3,611$). When ambiguous errors were filtered out ($N = 2,201$), the proportion of lexical errors rose to 66%. In other words, phonological errors were more likely to be ambiguous and therefore rejected from later analysis.[9] All authors practice some kind of filtering out of errors before carrying out their analyses, but the details of these procedures are often left unclear. The fact that there are more lexical errors in the two Spanish corpora might be a result of the filtering procedure. In any case, an explanation based on perceptual bias seems unlikely. Lexical errors are much more easily detectable than phonological errors. Given that perceptual biases are mostly absent from the phonological level (particularly in del Viso's corpus), they are not expected to have stronger effects on the lexical level. A greater proportion of lexical level errors may constitute one more language-specific characteristic of Spanish.

The second "suspicious" pattern is the greater rate of between words versus within word phonological errors, which could be due to perceptual bias because of the greater salience of the former. As Spanish phonological errors have not been shown to be affected by biases linked to phoneme identity, position, or lexical status, it is unlikely that this remaining effect should be a result of bias. However, more substantial arguments can be put forward against the bias hypothesis. If between words and within word errors arise as a result of different production mechanisms, it should be possible to see other proofs of these mechanisms at work. Shattuck-Hufnagel (1983) pointed out that these two types of error seem to behave differently with respect to stress and syllable-position constraints. Stemberger (1982) found that target and intrusion segments are less similar within word than between words. Within word errors also respect the syllable position constraint to a lesser extent. Phoneme shifts create clusters that are similar to their source cluster more often when they span two different words. Finally, there are more anticipations or perseverations of single features within

[9] These proportions cannot be estimated in our corpus, as the filtering procedure is carried out week by week while the corpus is built, and once again over the whole corpus when this has grown enough. Some ambiguous errors are classified into specific categories, but many are disregarded without ever entering the database.

words, but feature exchanges are attested only between words. Our own data from Spanish speech errors also support a processing difference between these two types of errors: most single-phoneme exchanges occur within words, while most anticipations and perseverations occur between words (a finding also reported by García-Albea et al., 1989). This is a language-specific effect not observed in English (Stemberger, 1982). As a conclusion, there are many reasons to believe that the different error rates observed in between versus within word errors reflect deep processing differences and not just a perceptual bias on the side of the collector.

Taken together, present results and prior reports from Spanish suggest that the influence of methodological biases of several kinds (mainly personal, sampling and perceptual biases) is very limited in collections of slips of the tongue compiled using a multiple-collector methodology, and nearly absent in those compiled using a single-collector methodology. As there are no reasons to believe that Spanish error collectors are better observers than those working on Germanic languages, either by training or basic cognitive skills, this means that most "suspicious" patterns found in Germanic corpora are probably actual production phenomena, and not the result of unreliable methods of observation.

We can conclude, therefore, that the following differences between Spanish and Germanic languages such as English, German, and Dutch are substantial cross-linguistic differences. Firstly, Spanish word-initial positions are protected against error (whereas syllable-initial positions share a greater propensity to error with Germanic languages). Secondly, syllable stress does not affect error rates. Finally, there is no lexical bias in phonological errors. Moreover, single-phoneme exchanges are mostly confined to within word errors, and anticipations and perseverations occur most often between words, whereas English does not show clear differences across error types between these two error categories. In addition, but less confidently, we can propose that Spanish might also show a greater proportion of lexical than phonological errors, whereas English shows the opposite; there may be a greater proportion of exchanges in Spanish at both phonological and lexical levels.

The challenge now is to develop a theory of phonological encoding that is able to account for both universal and language-specific aspects of language processing.

Acknowledgments This research served as partial fulfilment of the requirements for a Master's Thesis by Elvira Pérez at the Dept. de Psicología Experimental y Fisiología del Comportamiento, Universidad de Granada, under the supervision of Julio Santiago. We want to dedicate this paper to the students who helped collect the speech errors, and to Daniel Freaker. We also want to gratefully acknowledge the help and enthusiasm of Loli Huertas, Nicolás Gutiérrez, and Elena García-Krafes. Robert Remez and Mark Ball deserve special thanks for their help in bringing the manuscript into its final shape. Portions of this work were presented at the 12th ESCOP Conference. This work was supported by PB 97-0805 and BS02002-02993 grants of the DGESIC, Ministerio de Educación y Cultura (Spanish Ministry of Education and Culture). Elvira Pérez is now at the Dept. of Acoustic Design, Kyushu University, Fukuoka, Japan.

References

Alameda, R., & Cuetos, F. (1995). *Diccionario de frecuencias de las unidades lingüísticas del castellano*. Oviedo: Universidad de Oviedo.

Anton-Mendez, I., Hartsuiker, R., Roelstraete, B., & Costa, A. (2005). *Lexical bias in Spoonish Spanerisms*. Proceedings of the 18th CUNY conference on Human Sentence Processing. Tucson, AZ.

Anwar, M. S. (1979). Remarks on a collection of speech errors. *International Journal of Psycholinguistics, 6*, 59–72.
Baars, B. J., Motley, M. T., & MacKay, D. G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior, 14*, 382–391.
Baars, B. J. (1980). On eliciting predictable speech errors in the laboratory. In V. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand.* New York: Academic Press.
Baars, B. J. (1986). *The cognitive revolution in psychology*. New York: Guilford Press.
Baars, B. J. (1992a). A dozen competing-plans techniques for inducing predictable slips in speech and action. In: B. J. Baars (Ed.), *Experimental slips and human error: Exploring the architecture of volition.* New York: Plenum Press.
Baars, B. J. (Ed.) (1992b). *Experimental slips and human error: Exploring the architecture of volition*. New York: Plenum Press.
Bachoud-Levy, A.-C., Dupoux, E., Cohen, L., & Mehler, J. (1998). Where is the length effect? A cross-linguistic study of speech production. *Journal of Memory and Language, 39*, 331–346.
Berg, T. (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.
Bock, K. (1991). A sketchbook of production problems. *Journal of Psycholinguistic Research, 20*, 141–160.
Bock, K. (1996). Language production: Methods and methodologies. *Psychonomic Bulletin and Review, 3*, 395–421.
Boomer, D. S., & Laver, J. D. M. (1968). Slips of the tongue. *British Journal of Disorders of Communication, 3*, 2–12.
Cohen, A. (1973). Errors of speech and their implication for understanding the strategy of language users. In: V. Fromkin (Ed.), *Speech errors as linguistic evidence*. The Hague: Mouton.
Costa, A., & Sebastián-Gallés, N. (1998). Abstract phonological structure in language production: Evidence from Spanish. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 886–903.
Cutler, A. (1982). The reliability of speech error data. In: Cutler, A. (Ed.), *Slips of the tongue and language production*. Berlin: Mouton.
Cutler, A. (1988). The perfect speech error. In L. M. Hyman & C. N. Li (Eds.), *Language, speech, and mind: Studies in honour of Victoria A. Fromkin*. London: Routledge.
Chen, J.-Y. (1999). The representation and processing of tone in Mandarin Chinese: Evidence from slips of the tongue. *Applied Psycholinguistics, 20*, 289–301.
Chen, J.-Y. (2000). Syllable errors from naturalistic slips of the tongue in Mandarin Chinese. *Psychologia, 43*, 15–26.
Chen, J.-Y., Chen, T.-M., & Dell, G. S. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language, 46*, 751–781.
del Viso, S. (1992). *Errores espontáneos del habla y producción del lenguaje*. Madrid: Editorial de la Universidad Complutense de Madrid.
del Viso, S., Igoa, J. M., & García-Albea, J. E. (1991). On the autonomy of phonological encoding: Evidence from slips of the tongue in Spanish. *Journal of Psycholinguistic Research, 20*, 161–185.
Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*, 283–321.
Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes, 5*, 313–349.
Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review, 104*, 123–147.
Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior, 20*, 611–629.
Ellis, A. W. (1980). On the Freudian theory of speech errors. In V. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand*. New York: Academic Press.
Ferber, R. (1991). Slip of the tongue or slip of the ear? On the perception and transcription of naturalistic slips of the tongue. *Journal of Psycholinguistic Research, 20*, 105–122.
Freud, S. (1922). *Psicopatología de la vida cotidiana*. Spanish translation of the 7th German edition, 1920. Madrid: Biblioteca nueva.
Fromkin, V. (1971). The non-anomalous nature of anomalous utterances. *Language, 47*, 27–52.
Fromkin, V. (Ed.) (1980). *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand.* New York: Academic Press.
Fromkin, V. (1988). Grammatical aspects of speech errors. In F. J. Newmeyer (Ed.), *Linguistics: The Cambridge survey. II.- Linguistic theory: Extensions and implications*. Cambridge: Cambridge University Press.

García-Albea, J. E., del Viso, S., & Igoa, J. M. (1989). Movement errors and levels of processing in sentence production. *Journal of Psycholinguistic Research, 18*, 145–161.
Garnham, A., Shillcock, R. C., Brown, G. D. A., Mill, A. I. D., & Cutler, A. (1982). Slips of the tongue in the London-Lund corpus of spontaneous conversation. In A. Cutler (Ed.), *Slips of the tongue and language production*. Berlin: Mouton.
Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), *The psychology of learning and motivation. Vol. 9*. New York: Academic Press.
Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production. Vol. 1: Speech and talk*. London: Academic Press.
Harley, T. A. (1984). A critique of top-down independent levels models of speech production: Evidence from non-plan-internal speech errors. *Cognitive Science, 8*, 191–219.
Harley, T. A., & MacAndrew, S. B. G. (2001). Constraints upon word substitution speech errors. *Journal of Psycholinguistic Research, 30*, 395–418.
Hartsuiker, R. J. (2002). The addition bias in Dutch and Spanish phonological speech errors: The role of structural context. *Language and Cognitive Processes, 17*, 61–96.
Igoa, J. M. (1996). The relationship between conceptualization and formulation processes in sentence production: Some evidence from Spanish. In M. Carreiras, J. E. García-Albea & N. Sebastián-Gallés (Eds.), *Language processing in Spanish*. Mahwah, NJ: Erlbaum.
Jescheniak, J., & Levelt, W. J. M. (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.
Kawachi, K. (2002). Practice effects on speech production planning: Evidence from slips of the tongue in spontaneous vs. preplanned speech in Japanese. *Journal of Psycholinguistic Research, 31*, 363–390.
Lashley, K. S. (1951). The problem of serial order in behavior. In L. P. Jeffress (Ed.), *Cerebral mechanisms in behavior: The Hixon simposyum*. John Wiley.
Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
Levelt, W. J. M., Roelofs, A., & Meyer, A. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–38.
MacKay, D. G. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia, 8*, 323–350.
MacKay, D. G. (1987). *The organization of perception and action*. New York: Springer-Verlag.
Meringer, R., & Mayer, K. (1895). *Versprechen und verlesen*. Stturgart: Goschensche Verlagsbuchandlung. Reedited in 1978, A. Cutler, & D. Fay (Eds.), Amsterdam: Benjamins.
Meyer, A. (1992). Investigation of phonological encoding through speech error analyses: Achievements, limitations, and alternatives. *Cognition, 42*, 181–211.
Nooteboom, S. G. (1969). The tongue slips into patterns. In A. G. Sciarone (Ed.), *NOMEN: Leyden studies in linguistics and phonetics*. The Hague: Mouton.
Otake, T., & Cutler, A. (1996). *Phonological structure and language processing: Cross-linguistic studies*. Berlin: Mouton.
Pérez, E. (2002). La sílaba y los errores del lenguaje. Unpublished masters thesis. Dept. de Psicología Experimental y Fisiología del Comportamiento, Universidad de Granada, Spain.
Santiago, J., Pérez, E., Palma, A., & Stemberger, J. P. (under review). Syllable, word and phoneme frequency effects in Spanish phonological speech errors: The David effect on the source of the error. In T. S. Carson & V. Ferreira (Eds.). *The state of the art in speech error research: Proceedings of the LSA institute workshop. MIT working papers in linguistics*.
Schwartz, M. F., Saffran, E. M., Bloch, D. E., & Dell, G. S. (1994). Disordered speech production in aphasic and normal speakers. *Brain and Language, 47*, 52–88.
Shattuck-Hufnagel, S. (1986). The representation of phonological information during speech production planning: Evidence from vowel errors in spontaneous speech. *Phonology Yearbook, 3*, 117–149.
Shattuck-Hufnagel, S. (1987). The role of word-onset consonants in speech production planning: New evidence from speech error patterns. In E. Keller & M. Gopnik (Eds.), *Motor and sensory processes in language*. Mahwah, NJ: LEA.
Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering. *Cognition, 42*, 213–259.
Shattuck-Hufnagel, S., & Klatt, D. H. (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.

Shattuck-Hufnagel, S., & Klatt, D. H. (1980). How single phoneme error data rule out two models of error generation. In V. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand*. New York: Academic Press.
Stemberger, J. P. (1982). The nature of segments in the lexicon: Evidence from speech errors. *Lingua, 56*, 235–259.
Stemberger, J. P. (1984b). Structural errors in normal and agrammatic speech. *Cognitive Neuropsychology, 1*, 281–313.
Stemberger, J. P. (1985). An interactive activation model of language production. In A. W. Ellis (Ed.), *Progress in the psychology of language. Vol. 3*. Londres: LEA.
Stemberger, J. P. (1989). Speech errors in early child language production. *Journal of Memory and Language, 28*, 164–188.
Stemberger, J. P. (1991a). Radical underspecification in language production. *Phonology, 8*, 73–112.
Stemberger, J. P. (1991b). Apparent anti-frequency effects in language production: The addition bias and phonological underspecification. *Journal of Memory and Language, 30,* 161–185.
Stemberger, J. P. (1992). The reliability and replicability of naturalistic speech error data: A comparison with experimentally induced errors. In B. J. Baars (Ed.), *Experimental slips and human error: Exploring the architecture of volition*. New York: Plenum Press.
Stemberger, J. P., & MacWhinney, B. (1986). Frequency and the lexical storage of regularly inflected forms. *Memory and Cognition, 14*, 17–26.
van den Broecke, M. P. R., & Goldstein, L. (1980). Consonant features in speech errors. In V. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand*. New York: Academic Press.
Vousden, J. I., Brown, G. D. A., & Harly, T. A. (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology, 41*, 101–175.
Wijnen, F. (1992). Incidental word and sound errors in young speakers. *Journal of Memory and Language, 31*, 734–755.