Spontaneously occurring speech errors in German: BAS corpora analysis

Stefania Marin [1], *Marianne Pouplier*[2]

[1 2]Institute of Phonetics and Speech Processing, LMU Munich, Germany

pouplier|marin@phonetik.uni-muenchen.de

Abstract

This study describes the characteristics of spontaneously occurring speech errors on the basis of an acoustic corpus of German. The results corroborate some of the error patterns observed in transcription corpora (frequency of types of errors, the addition bias of errors), but also disconfirm some others (preference for a word-initial position, phonological well-formedness of errors). This suggests that, especially for cross-linguistic comparisons, available acoustic (semi)-spontaneous corpora are an important tool in speech error research. Depending on the content and quality of the corpus, instrumental analyses similar to those performed on laboratory-elicited errors may however not be possible even on acoustic corpora.

Index Terms: spontaneous speech errors, German, acoustic corpus

1. Introduction

This study aims to evaluate to what extent available acoustic corpora of (semi)-spontaneous speech can be used for speech error analyses, in the context of the debate on the nature of laboratory-elicited vs. naturalistically-occurring speech errors (cf. [1, 2] for a discussion). A first step is to describe the spontaneously occurring errors in such an acoustic corpus and to determine to what extent the observed patterns match those from transcription corpora (i.e. from corpora of "Speech errors collected by listening to everyday speech" [1: 373]) and whether instrumental evidence can be found for similarities between laboratory and spontaneously-occurring errors.

Error analyses based on transcription corpora have identified certain error patterns, such as word and syllable position asymmetries in the distribution of errors, the categorical rather than gradient nature of errors, the addition bias of errors, or the phonological well-formedness of errors [3, 4, 5, 6]. However, it is not entirely clear to what extent these patterns are influenced by attention and perceptual biases specific to these corpora (cf. [1, 7] for a discussion). Given that instrumental studies have found that errors are often not well-formed, and are gradient rather than categorical in nature [2] (e.g. an errorful intended /t/ in *top* in a phrase like *cop top* being produced with a simultaneous tongue tip and tongue dorsum raising), the question that arises in the speech error research is whether transcription studies under-represent such errors due to their inherent limitations or whether laboratory-elicited errors differ in some fundamental way from naturally-occurring ones. In this context, an analysis of errors from a large acoustic corpus of spontaneous speech, where the acoustic signal is available for analysis by multiple experimenters, allows for a more careful perceptual and instrumental description of the errors, although, as we point out in the Discussion section, our corpus also has inherent biases and limitations.

Available language transcription corpora, especially from Germanic languages, have found for example that there are more errors on word-initial than word-medial or final consonants, on syllable-initial than syllable-final consonants, and that there are more anticipations than perseverations (e.g. [8] specifically for German). Such transcription corpus studies have also found that error and target units more often differ in one rather than several features, that there are more errors on consonants than on vowels and that place of articulation is affected more often than manner and voicing (cf. [7] for a review of the relevant literature and specific references). Finally, substitutions have been reported as the most frequent error type [8, 9] (Note that [9], like the current study, is based on an acoustic corpus, but no analyses beyond type of error are provided). Our data consist of spontaneous speech errors from dialogue and monologue German acoustic corpora. Frequency analyses of these errors have been performed, in an attempt to replicate the patterns reported from transcription studies. As we discuss, the errors in these acoustic corpora are unfortunately not suitable for a systematic instrumental analysis.

2. Methods

2.1. Corpus

The data were taken from four BAS (Bayerisches Archiv für Sprachsignale) Corpora, orthographically transcribed by trained transcribers using the Verbmobil Transliteration schema. The corpora were collected for purposes other than speech errors, but they were annotated (among others) for "neologism or mispronunciation", which we exploited for selecting utterances containing potential errors. One corpus, Regional Variants of German [10], consisted of 1-minute monologues from 451 speakers recruited demographically across Germany. 71 utterances, produced by 51 speakers were annotated as "neologism or mispronunciation". The SmartKom corpus [11] consisted of multi-modal human-machine dialogues. From a total of 234 speakers, 43 utterances were annotated as "neologism or mispronunciation". The Verbmobil 1 and 2 corpora [12] consist of face-to-face task-solving dialogues between two native German speakers. A total of 336 "neologism or mispronunciation" utterances came from the two Verbmobil Corpora, produced by 263 speakers. Across corpora, a total of 450 "neologism or mispronunciation" utterances from monolingual monologues and dialogues were available for analysis.

2.2. Classification

The annotated "neologism or mispronunciation" were further classified by the two authors auditorily and by inspecting the acoustic signal. A third transcriber evaluated any cases of disagreement (N=18). Our classification categories with examples are listed in Table 1. Only examples from categories A and B were considered for further error analyses. Note that

what we termed "changes of plan" in category C were not counted as errors since our focus here is on sublexical error patterns. However, full words that were incorrect given the grammatical context (e.g. *Termines* instead of intended *terminlich*) fell under our definition in category B, so our analysis also includes a limited number of word-level errors. For errorful utterances (categories A and B), target and replacer units were also determined (consonant, vowel, consonant/vowel sequence, syllable, morpheme, word), error direction (anticipation, perseveration), as well as type of error: addition or elision of material, substitution, multiple (any combination of addition, elision, substitution), exchange, word blend. Examples are given in Table 2.

Table 1. Classification criteria and examples.

Category	Definition	Example
A. Aborted errorful utterance	Abortion that cannot be completed into a word given the grammatical, semantic context	- nach+--artib--+ Artikel - man +--sula--+ solidariziert sich
B. Errorful utterance	A complete utterance that is a non-word, or not an appropriate word given the grammatical, semantical context.	- zur Mittagspause +--britbringen--+ wollen -bei mir s +--termines--+ terminlich was frei
C. Hesitations, abortions associated with change of context or change of target word	- A fragment utterance that is repeated, and continued without change. - Aborted material is repeated and completed, the context changes. - Aborted material can be completed into a word and changes.	- in der +--mit--+ mittleren - der Hamburger +--Geschäftsr--+ <äh> Hannoverer Geschäftsreise - dann +--bla--+ verbleiben wir
D. Neologisms	Neologisms, colloquial and regional expressions, technical terms, foreign names/words.	- um einen +--Burn-In-Effekt--+ zu handeln - halt +--rucki-zucki--+ das

Table 2. Examples of error type.

Error type	Example	Target
Addition (Ad)	einf[l]ach (C) Konferenzraum[a] (V)	einfach Konferenzraum
Elision (El)	Vorsch[]ag (C) Erled[]ung (VC)	Vorschlag Erledigung
Substitution (Sub)	[v]aiwoche (C) ganzt[a]gig (V) hab[m] (morpheme)	Maiwoche ganztägig habe ich
Multiple (Mult)	[br]itbringen (Ad.+Su.) [ʃöt](El. + Su.)	Mitbringen steht
Exchange	Ter[n]i[m]	Termin
Word blend	Abzubesprechen	abzusprechen + besprechen

There were cases that could afford alternative interpretations, and for such cases the most conservative analysis was selected. For example, /abomnə-/ for intended *Abonnement* could be interpreted as addition of /m/ or as an /m/-/n/ exchange. However, because the word is not complete we favored the former interpretation. Also, an error such as *dass /ks/ dass können wir* could be interpreted as addition of /s/ or as substitution of intended vowel /ö/ in *können* with consonant /s/; in this case we also favored the addition interpretation to a substitution one, which would bias our result towards not observing substitutions across consonants and vowels. Also, some cases could be interpreted as either a feature or segment error: /tembinen/ for *Terminen*, could be interpreted as either addition of segment /b/ or as a nasal feature error (or a timing error on velum raising). Such cases were given a segmental analysis, biasing the results in favor of segmental-level errors, and against sub-segmental/feature-level errors.

3. Results

Of the 450 utterances, two turned out to have been spoken by a non-native speaker and were excluded from any further analysis. One additional case could not be classified, as it could not be determined what the intended utterance was, so it was also excluded. Of the remaining 447 utterances, about half (N= 207, 46%) were classified as errorful by our criteria (73 in category A and 134 in B), and about half as neologism, regionalism etc. (N=197, 44% in our D category) Although hesitations and aborted utterances associated with change of context or of target word had a different coding in the original database annotation ("abortion"), a few cases (N=43, 10%) could also be identified under the "neologism or mispronunciation" coding (our category C).

For further error analyses, the utterances classified in the A and B categories were considered. In terms of corrections, 153 (73%) were corrected overall, with most (97%) A-category cases being corrected, and with 60% of the B-category cases being corrected. In terms of direction of the error(s) being observed, more than half were contextual (anticipation and/or perseveration, exchange), with anticipations occurring most often, as evident from Table 3. In the anticipation + perseveration cases, there were also incorporated two cases of half-exchanges: /relafil/ for *relative* (with /f/ and a coronal exchanging place, but the coronal being /l/ in the error for intended /t/) and /etvender/ for *entweder* (/n/ is elided in the first syllable and added in the second, looking like an /n/-null exchange). The perseveration count included an interesting case of manner exchange: /rausgesygs/ for *rausgesucht*, where the intended velar fricative /x/ is replaced by velar stop /g/ and the intended alveolar stop /t/ by the alveolar fricative /s/.

In terms of error types, the majority of errors were substitutions, followed by additions (Table 4). Additions and elisions together were about half as frequent as the substitutions, with elisions about half as frequent as additions. Exchanges and word blends were overall rare (<4%).

Position of the error is considered next. For word blends, as well as the 3 word-level substitutions that met our definition for category B, this analysis is not relevant hence the totals in Tables 5-7 differ slightly from the totals in Table 4. Position of the error within the word is summarized in Table 5. For both A and B categories, most errors were located medially within a word (for the A category, which was comprised of unfinished utterances, it must be noted that per definition, the error could not be final within the intended utterance). For category B,

initial and final errors were as likely, and were less than half as frequent as the medial ones. Category "multiple" summarizes cases where errors were both initial and medial (3 in total) or both medial and final (4 in total). For example, in *rela[f]ie[l]* for intended *relativ*, two errors occurred, one medial and one final.

Table 3. Frequencies: Error direction.

	A	B	Total A+B
Noncontextual, unclear	26 (35%)	58 (43%)	84 (41%)
Anticipation	26 (35%)	29 (22%)	55 (26%)
Perseveration	11 (14%)	26 (20%)	37 (18%)
Anticipation + perseveration	8 (11%)	18 (13%)	26 (13%)
Exchange	2 (3%)	3 (2%)	5 (2%)
Total	73	134	207

Table 4. Frequencies: Error type.

	A	B	Total A+B
Sub	43 (60%)	76 (56%)	119 (57%)
Ad	15 (20%)	25 (19%)	40 (19%)
El	5 (7%)	13 (10%)	18 (9%)
Mult	7 (9%)	10 (7.5%)	17 (8%)
Blend	1 (1%)	7 (5%)	8 (4%)
Exchange	2 (3%)	3 (2%)	5 (2.5%)
Total	73	134	207

Table 5. Position of error within the word.

	A	B	Total A+B
Initial	10 (14%)	23 (19%)	33 (17%)
Medial	60 (83%)	71 (57%)	131 (67%)
Final		25 (20%)	25 (13%)
Multiple	2 (3%)	5 (4%)	7 (3%)
Total	72	124	196

Position within the word of the syllable containing the error is summarized in Table 6. Overall, 32 errors occurred in intended monosyllabic words, 56 in intended disyllabic words, and 108 in intended polysyllabic words. Per definition, there were more opportunities for an error to occur in initial than medial syllable as both disyllabic and polysyllabic intended words had an initial intended syllable, while only polysyllabic words could have a medial intended syllable. Likewise, there were overall more opportunities for an error to occur in initial than final syllable given that all errors in categories A and B had an initial syllable, but that at least some cases from the A category may have been aborted prior to the intended final syllable. However, the initial syllable was where errors occurred the least. Errors in the medial syllable were the most frequent, although only target polysyllabic words offered the context for a medial syllable error (the two medial errors in intended disyllabic words consisted of medial syllable additions). If we consider polysyllabic words in category B, which regardless of other considerations, offer the opportunity for errors to be made in initial, medial and final syllable, we notice that the errors occurred most frequently in medial syllable, followed by final and followed by initial syllable. Overall, it can be said that polysyllabic words are more likely to be errorful than shorter words (longer words offer more opportunity for error), and it is more likely for the error to occur in later (medial or final) rather than in initial syllables.

We now turn to describing each error type in part in more detail. Given the detail of this analysis, we collapse over the A and B categories. Regarding blends, of the 8 word blends, half were corrected and half not. All exchanges were corrected. One exchange was a cross-word consonant exchange (/fer zil/ for *sehr viel*), and four were within word exchanges of vowels (1 case: /mola-/ for *Mallorca*) or consonants (3 cases: /veim-/ for *Maiwoche*, /ternim/ for *Termin*, /nenm/ for *nehmen*). The consonant exchanges included 2 cross-syllable onset exchanges (/fer zil/, /veim-/) and two within-syllable onset-coda exchanges (/ternim/, /nenm/).

Table 6. Position within word of the syllable containing the error, as a function of number of syllables for the intended utterance.

	A	B	Total A+B
Monosyllabic words	**4**	**28**	**32**
Initial syllable	**17**	**21**	**38**
- Disyllabic words	9	11	20
- Polysyllabic words	8	10	18
Medial syllable	**34**	**30**	**64**
- Disyllabic words	1	1	2
- Polysyllabic words	33	29	62
Final syllable	**12**	**41**	**53**
- Disyllabic words	6	27	33
- Polysyllabic words	6	14	20
Multiple syllables	**5**	**4**	**9**
- Disyllabic words	1	0	1
- Polysyllabic words	4	4	8
Total	72	124	196

Table 7. Correction rate for additions (Ad.), elisions (El.), substitutions (Su.) and multiple (Mu.) errors.

	Ad	El	Sub	Mult	Total
yes	30 (75%)	15 (83%)	94 (79%)	11 (65%)	149 (77%)
no	10 (25%)	3 (17%)	25 (21%)	6 (35%)	44 (23%)
Total	40	18	119	17	194

For additions, elisions, substitutions and multiple errors, details are summarized in Tables 7 – 10. The majority of errors were corrected, and the majority of additions, substitutions and multiple errors were contextual (with anticipation being the most frequent direction). For elision, contextuality cannot be determined, so they are not part of this analysis.

For determining unit of error, a rhotic coda, which is produced vocalized, is counted as a C unit (C and VC units therefore also include cases where C is vocalized R). Diphthongs are counted as a V-unit, and affricates as a C-unit. For additions, the unit reported is the replacer unit (since the target/intended unit is null), for elision it is the target/intended unit (since the replacer unit is null). For substitutions, the target unit is reported, and it must be noted that target and replacer units were the same except for three cases: in two cases, a vowel was replaced by a VC (vowel plus vocalized R) sequence (/ɛrpril/ for *April*, /gemɛr-/ for *gemeinsam*), and in one case a vowel was replaced by an /r/-like non-native sound (/ʃɐn/ for *schön*) (these cases are counted in Table 9 as V units). One must recall that our analysis was biased towards substitutions of the same type of segment, since ambiguous cases such as /ks-/ for intended *können* were analyzed as addition of /s/ rather than substitution of /ö/ with /s/.

Table 8. Direction of error frequencies for additions (Ad.), elisions (El.), substitutions (Su.) and multiple (Mu.) errors.

	Ad	Sub	Mult	Total
Noncontextual	11 (27.5%)	42 (35%)	8 (47%)	61 (35%)
Anticipation	15 (37.5%)	34 (29%)	4 (24%)	53 (30%)
Perseveration	6 (15%)	28 (23%)	2 (11%)	36 (20%)
Anticipation + perseveration	8 (20%)	15 (13%)	3 (18%)	26 (15%)
Total	40	119	17	176

Table 9. Unit of error for additions, elisions, substitutions.

Unit	Ad	El	Sub	Total
C	32 (80%)	8 (44%)	54 (45%)	94 (53%)
V	3 (7.5%)		45 (38%)	48 (27%)
CV	1 (2.5%)	1(6%)	7 (6%)	9 (5%)
VC	1(2.5%)	4 (22%)	2 (<2%)	7 (4%)
VCC			2 (<2%)	2 (1%)
Syllable			2 (<2%)	2 (1%)
Morpheme		1 (6%)	1 (<2%)	2 (1%)
Word			3 (2.5%)	3 (2%)
Other	3 (7.5%)	4 (22%)	3 (2.5%)	10 (6%)
Total	40	18	119	177

Table 10. Error unit (target unit for elisions and substitutions, replace unit for additions).

	A		B		Total A + B	
	C	V	C	V	C	V
Ad	17	3	32	3	49	6
El	12	6	20	6	32	12
Sub	26	36	58	35	84	71
Total	55	45	110	44	165	89

Overall, a single consonant is the error unit in a majority of cases (Table 9). For substitutions, it is almost as likely for a single consonant or vowel to be substituted (45% vs. 38%). Single vowel additions are rare (<8%), and single vowel elisions unattested. Of all additions, 80% of the cases are additions of single consonants. C and VC elisions constitute more than 60% of all elision cases. In some cases, the CV unit was at the same time a syllable: this was the case for the CV elision (/tɛls/ for *Hotels*), and for 6 of the 7 CV substitutions, but it was analyzed at the segmental level as long as the analysis was relatively straightforward (e.g. /informa'ze-/ for *Information*). As to the VC unit, one elision case could also be interpreted as a syllable elision (elision of /er/ in /kundn/ for *erkundigen*), but the other VC errors span two syllables (e.g. elision of /iɡ/ in *erkundigen*; /iɡ/ addition in /forlesiguŋ/ for *Vorlesung*). The remaining syllable-level errors (counted as such in Table 9) involved different syllable types in the target and replacer units, and extremely complicated analyses as segmental- rather than syllable-level substitutions (e.g. /rukʃa-/ for *ruckfliegen*).

As detailed in the Methods section, word substitutions were extremely rare in our analysis, as many such cases would fall under the change-of-plan aborted utterance label. Category "other" included units such as CVC (that did not form one syllable, e.g. elision of /mɪt/ in /naxəks/ for *Nachmittags*), CC, CCC. This category also included a VCCCVCC elision that was not along syllable boundaries: /seix-/ for *Sehenswürdigkeit*. In summary, the most frequent errors were single segment (C or V) errors, and multiple-segment errors were often affiliated to different syllables.

Of the 17 multiple errors, 9 involved only consonants (5 cases of consonant addition + substitution, 2 of consonant elision + addition, 2 of consonant elision + substitution), and 6 a combination of consonants and vowels (4 cases of C addition + V substitution, and one each of CV/VC addition + V substitution). Two cases only involved only vowels.

For the remainder of the analysis, segmental-level additions, elisions and substitutions are considered in more detail. For this level of analysis, multiple errors were split so that consonant and vowel additions, elisions and substitutions overall could be counted. Errors spanning multiple segments within the same utterance were also split for this analysis. Exchanges were included as individual consonant/vowel substitution cases. Overall, 254 additions, elisions and substitutions were available. Substitutions were by far the most frequent error type (N=155), followed by additions (N=55) and elisions (N=44).

A summary of error unit for categories A and B is given in Table 10 (for additions, the unit is the replacer unit, for elisions and substitutions it is the target unit). For both additions and elisions, consonants were targeted more often than vowels. For substitutions, the numbers were more balanced between vowels and consonants, although the predilection for consonant errors was still observed.

Position of the errors by type and classification are summarized in Tables 11-12. Absolute position of the error within word does not differ from the general pattern (most errors are located medially regardless of error type), so is not reported in detail. Position of the syllable containing the error is summarized in Table 12. Beyond the general patterns already discussed from Table 6, the detailed analysis revealed that additions occur almost as likely in syllables that are word-initial as word-medial, while substitutions favor syllables in both medial and final word position (most elisions occur in medial syllables). Finally, position of errorful segment within

a syllable is summarized in Table 12, with most errors in onset position, and the fewest in coda position.

Table 11. Position of syllable containing the error.

	Ad		El		Sub	
	A	B	A	B	A	B
Monosyllables	**1**	**4**		**2**	**3**	**26**
Initial syllable	**9**	**12**	**3**	**8**	**13**	**12**
- Disyllabic words	4	7	2	2	6	5
- Polysyllabic words	5	5	1	6	7	7
Medial syllable	**10**	**10**	**13**	**10**	**32**	**21**
- Disyllabic words	2	1				
- Polysyllabic words	8	9	13	10	32	21
Final syllable		**9**	**2**	**6**	**14**	**34**
- Disyllabic words		4	1	3	8	22
- Polysyllabic words		5	1	3	6	12

Table 12. Position of error within syllable.

	Ad		El		Sub		Total
	A	B	A	B	A	B	A+B
Onset	13	15	8	14	20	33	103
Coda	3	14	4	6	6	25	58
Nucleus	3	4	6	6	36	35	90
Unclear	1	2					3

Table 13. Frequency of cluster creations and simplifications. The numbers in parentheses represent cases, for the initial and final position, where addition/elision happened word-internally (see text for details).

	Position	A	B	A + B
Cluster creation	Initial	6 (5)	8 (6)	14 (11)
	Medial	5	16	21
	Final		7 (5)	7 (5)
Cluster simplification	Initial		3 (2)	3 (2)
	Medial	3	6	9
	Final		2 (1)	2 (1)

We also tabulated in how many cases, the errors resulted in cluster creations and simplifications (Table 13). Overall, there were more cluster creations than simplifications (at a rate of almost 4 to 1), confirming the addition bias for errors reported previously (cf. [6]). Most cases were word-medial (syllabification in this context is not always clear). In terms of clear onset and coda clusters (initial and final positions), there were twice as many onset cluster creations than coda cluster creations, and about the same number of onset and coda simplifications. The onset-coda asymmetry however is likely due to the particularity of aborted errors (category A). When looking only at completed utterances (category B), onset and coda creations are about as frequent in word-initial/word-final position. Also, for initial and final cluster creation/simplification cases, we assessed whether segmental addition/elision happened at word edge or word-internally. Overall, for both cluster creations and simplifications in both initial and final position, in a majority of cases a consonant was added/elided word-internally rather than at the word margin/edge. One cluster creation in initial position included an addition at word edge and one word-internally (/kʃtornʃ-/ for *Schornsteinfegermeister*). Four word cluster creations involved the creation of an illegal cluster (e.g. /kʃtornʃ-/ for *Schornsteinfegermeister;* /md-/ for *Montag;* /tk/*erminkalender* for *Terminkalender*, /kronmgres/ for *Kongres*).

Table 14. Feature description of vowel substitutions

Vowel substitution	A+B
Monophthong > Diphthong	11
Diphthong > Monophthong	5
Height	17
Rounding	6
Place	6
Place + rounding	6
Height + place	5
Height + place + rounding	5
Height + rounding	3
Other/unclear	7
Total	71

Table 15. Feature description of consonant substitutions.

Consonant substitution	A+B
Manner	22
Manner + place	24
Manner + voice	10
Place	22
Place + voice	1
Voicing	2
Other/unclear	3
Total	84

In what follows, substitutions are examined in more detail. For vowels, height substitution was the most common (Table 14): height alone, and height combined with other features accounted for almost 45% of all vowel substitutions. Place alone or combined with other feature accounted for 30% of vowel substitutions. 15% were substitutions of monophthongs by diphthongs, and about half that (7%) were substitutions of diphthongs by monophthongs.

For consonants, manner, place and manner + place substitutions were the most frequent (Table 15), overall accounting for 80% of all substitutions. The feature voice was rarely an error target, but voice plus manner substitutions accounted for 14% of the cases. The manner and manner plus place errors included 10 substitutions of simplex consonants (t, d, z, n, f) with an affricate (ts, tʃ, pf), and 4 substitutions of an affricate (ts) with a simplex consonant (s, z, ʃ, t). Over consonant and vowel errors, it was twice as likely that a simplex segment was substituted with a complex segment (diphthong; affricate) than the other way round. This result

suggests an addition bias also on segmental features (simple vs. complex segments), not only across segments (e.g. cluster creations).

4. Discussion

The current study has confirmed, on the basis of an acoustic corpus some of the patterns revealed by transcription corpora (the high frequency of substitutions, of anticipations, the addition bias of errors), but at the same time it only partly confirmed or even disconfirmed some others. Thus, in our corpus there were more errors in onset than coda overall, but when only non-aborted errorful utterances were considered (which offered equal opportunities for word-initial and word-final errors), this asymmetry was less pregnant (and cluster creations in the B category were as likely in onset as in coda position). Likewise, consonants were targeted more frequently than vowels overall, but for substitutions the numbers were fairly balanced. Finally, in our corpus, errors were most frequent word-medially not word-initially, both manner and place were features frequently targeted by errors and the combination of the two was as frequent. Our corpus also contained errors representing unattested sounds and combinations of sounds (illegal clusters), speaking against a general phonological well-formed-ness of spontaneous errors (with implications for theories of speech production, as addressed elsewhere, e.g. [13]).

Given that our results disconfirmed some of the patterns in a German transcription corpus of errors (e.g. word/syllable initial vs. medial being the locus of most errors, cf. [8]), the differences are likely to be not language-specific but corpus/modality-specific. On the other hand, our corpus replicated in terms of position of errors the results reported from Spanish transcription corpora [7]. While the Spanish pattern has been previously attributed to language specific differences between Germanic and Romance languages [7], our data suggest that the cross-linguistic patterns may in fact be corpus- rather than language-specific. For cross-linguistic comparisons therefore, transcription corpora may not be adequate, as the inherent methodological variability between such corpora and possible biases make it impossible to separate actual cross-linguistic differences from corpus-specific differences.

While in theory, an acoustic corpus also offers the possibility of instrumental analyses that could match to some extent the analyses of laboratory-elicited errors, this was not feasible with the current corpus. A problem of the available acoustic corpus is the fact that errors are so different that each one requires its own analysis; such individual analyses are not trivial to perform given that the available material for a speaker does not allow for good control materials with which to robustly compare the error. Also, the audio quality of the corpora varies considerably which further complicates the issue of what is variability due to for instance a gradient error (e.g. a simultaneous tongue tip and tongue dorsum raising during production of an intended /t/ in *top* in the phrase *top cop*, cf. [2]), and what to inherent noise and overall speech variability. Informally, there were cases that we conservatively interpreted as a full segmental error (due to lack of robust controls), but that could have alternatively be interpreted as a gradient error similar to those elicited in the laboratory (e.g. /tk/*erminkalender* was interpreted as a segment addition of [k], although it could have also been interpreted as a co-production of two articulations, similar to /t/ in *top*, [2]).

Finally, we must point out that auditory and attention biases could not be entirely excluded from our corpus. Although our error descriptions are based on independent analyses from the two authors, in addition to our own perceptual biases, we must also acknowledge the attention and auditory biases of the original annotators of the corpus. We only evaluated the utterances that these annotators labeled as "neologism or mispronunciation", but additional errors (categorical or gradient) may have passed unnoticed in the material annotated as correct productions. Even with such limitations, such acoustic corpora can be very informative about the kinds of errors occurring in natural speech and can serve as a pathfinder for laboratory work.

5. Acknowledgements

Work supported by the ERC under the EU's 7th Framework Programme (FP/2007-2013)/Grant Agreement n. 283349-SCSPL.

6. References

[1] S. Shattuck-Hufnagel. What traditional speech error corpora can tell us (and what they can't): How not to throw the baby out with the bathwater. C.T.Schütze and V. Ferreira (Eds.). *The state of the art in speech error research.* MIT, 373-382, 2007.

[2] M. Pouplier. Articulatory perspectives on errors. C.T.Schütze and V. Ferreira (Eds.). *The state of the art in speech error research.* MIT, 115-132, 2007.

[3] D. S. Boomer and J.D.M. Laver. Slips of the tongue. *British Journal of Disorders of Communication, 3,* 2-12, 1968.

[4] V. Fromkin. The non-anomalous nature of anomalous utterances. *Language, 47,* 27-52, 1971.

[5] D.G. MacKay. Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia, 8,* 323-350, 1970.

[6] S. Shattuck-Hufnagel. Sublexical units and suprasegmental structure in speech production planning. P. MacNeilage (Ed.). *The Production of Speech,* 109-136. Springer, 1983.

[7] E. Pérez, J. Santiago, A. Palma, P. O'Seaghdha. Perceptual bias in speech error data collection: Insights from Spanish speech errors. *J Psych Res 36,* 207-235, 2007.

[8] T. Berg. *Die Abbildung des Sprachproduktionsprozesses in einem Aktivationsflußmodell – Untersuchungen an deutschen und englischen Verpsrechern*. Max Niemeyer Verlag, 1988.

[9] A. Garnham, R. Shillcock, G.D.A. Brown, A. Mill and A. Cutler. Slips of the tongue in the London-Lund corpus of spontaneous conversation. A. Cutler (Ed.) *Slips of the tongue and language production.* 805- 817. Mouton, 1982.

[10] S. Burger and F. Schiel. RVG 1 - A Database for Regional Variants of Contemporary German. *Proc. 1st Int. Conf. Lang. Resources and Evaluation, Granada, Spain,* 1083-1087, 1998.

[11] F. Schiel and U. Türk. Wizard-of-Oz Recordings. W. Wahlster (Ed.) *SmartKom*, 541 – 570, 2006.

[12] S. Burger, K. Weilhammer, F. Schiel and H.G. Tillmann. Verbmobil Data Collection and Annotation. W. Wahlster (Ed.) *Verbmobil: Foundations of Speech-to-Speech Translation,* 537-549, 2000.

[13] M. Pouplier, S. Marin and S. Waltl. Voice onset time in consonant cluster errors: Can phonetic accommodation differentiate cognitive from motor errors? *J. Speech, Language, and Hearing Research,* 57, 1577-1588, 2014.