A corpus study of lexical speech errors in Mandarin

I-Ping Wan, Marc Allassonnière-Tang

Taiwan Journal of Linguistics
Vol. 19.2, 85-118, 2021
DOI: 10.6519/TJL.202107_19(2).0003

A CORPUS STUDY OF LEXICAL SPEECH ERRORS IN MANDARIN*

I-Ping Wan[1] and Marc Allassonnière-Tang[2]
[1]*National Chengchi University*
[2]*University Lyon 2*

ABSTRACT

We investigate a corpus of lexical substitution speech errors in Mandarin conversation data and present how Mandarin speakers produce erroneous lexical items and how these items are related to the intended words. The corpus includes 747 lexical speech errors from 100 participants and applies the part-of-speech definition of the Academia Sinica Corpus. Our results partially match with the observations in Germanic and Romance languages. As an example, the data from Mandarin native speakers shows that erroneously produced words and target words are almost always found in the same parts of speech. Moreover, noun substitutions are the most common type of substitution within the majority of content word pairs. However, the occurrence of verb errors is higher in Mandarin than in other languages, possibly reflecting a word frequency effect.

Keywords: Lexical errors, Speech errors, Mandarin, Nouns, Verbs

*We are grateful for the comments of the reviewers and the editors, which helped to significantly improve the content of this paper. The first author would like to thank Dr. Chain-wu Lee for his continuous cutting-edge high-tech programming support in the construction of all the corpora in the Phonetics and Psycholinguistics Lab at National Chengchi University. The research reported in this paper was supported by funding to the first author under the MOST three-year grant MOST 98-2410-H-004-103-MY2 in Taiwan. The second author is thankful for the support provided under the IDEXLyon fellowship grant 16-IDEX-0005.

1. INTRODUCTION

Speech errors are collected and analyzed as a source of data to investigate how language is processed in the mind. While speech errors occur frequently in natural conversation, they do not occur arbitrarily. Speech errors commonly follow specific patterns and tendencies, which provides insight as to which units and structures are involved in the cognitive representation of language (Harley 2006). For instance, it may not be uncommon to hear a speaker say *bad* instead of *sad*. However, it may be less common to hear a speaker say *pad* instead of *sad. Both bad* and *pad* are phonologically similar to *sad* as all three words share the same nucleus and coda, however, if the error of *bad* is more likely to occur than *pad,* it may be a piece of evidence (but not a proof, since many more factors may also play a role in the process) that lexical entries of the same grammatical category share a stronger connection and are thus probably stored together in the mind. That is to say, *bad* and *sad* are both adjectives, while *pad* is a noun, making it less likely to occur as a speech error of *sad*.[1] This type of error is thus selected to support inferences about how language processing happens since errors as meaningful items can reflect how lexical items are stored in the mind. For instance, if speech errors tend to share the same syntactic category, we can infer that these categories are represented in our inner model of language processing.

In this paper, speech errors at the lexical level are identified as the erroneous selections of lexical items that involve a meaningful morpheme or word[2]. They typically occur when the 'lemmas' of semantically appropriate candidate for lexical items are activated. These *lexical errors* can be broadly categorized in two major types. Erroneous words that do not originate from the surrounding context form one category. For instance, when a speaker says *glass* instead of *cup*, or *more* instead of *less.* On the

[1]The authors are aware that several factors such as semantics and phonology are likely to have a simultaneous effect here. Further details on phonological features are explained in Section 5.2.

[2]Speech errors may also result in meaningless strings of phonemes. By way of illustration, lexical blends occur when two lexical items are activated and inserted into the same syntagmatic slot in a phonologically-blended form, e.g., when a speaker says *perple* instead of *person* or *people*. This paper does not investigate this type of error.

other hand, errors that originate from the surrounding context form another category. For instance, when a speaker says *the glass is in the glass* instead of *the glass is in the fridge.* The latter category is not included in our study since it is context-induced and does not relate directly to the speech processing steps which we are investigating. The term 'lexical errors' thus refers to non-context-induced speech errors in the current paper.

This paper contributes to the existing literature by investigating lexical errors in Mandarin to assess if tendencies attested in Germanic and Romance languages are also found in Mandarin[3]. Current existing models on language processing are mostly based on data from Indo-European languages, which may be subject to Galton's problem, i.e., the tendencies observed in speech errors may be specific to Indo-European languages rather than apply to all languages. Providing additional data from Mandarin may thus shed additional light on the topic and provide stronger support to the existing language production models. Moreover, most of the literature is focused on non-lexical errors (i.e., errors resulting in meaningless strings of phonemes) due to their higher occurring frequency and relevance to phonology (e.g., Cantonese: Alderete et al 2019; Alderete and Davies 2019; Alderete and Tupper 2018; Mandarin: Wan 2016). Studies of speech errors in Mandarin also mostly focused on non-lexical errors (Wan 2007a, 2007b, 2007c, 2016). A few studies have addressed lexical errors in Mandarin (Wan, 2019), but they did not investigate the syntactic relation between targets and errors, which is one of the gaps in knowledge that the current paper aims at filling by using automatized methods based on content in the Academia Sinica Corpus (further details in Section 3).

By focusing on lexical errors, this study provides another type of data of the universal tendencies observed in the semantic level of language processing models. First, Fromkin (1973), Nooteboom (1973), Fay and Cutler (1977), Jaeger and Wilkins (2005), Harley and MacAndrew (2001), and Jaeger (2004) all found that lexical errors shared grammatical categories with their targets since the grammatical patterns of the specific

[3] Context-induced errors are excluded since they are less relevant to the cognitive representation of language, i.e., the cause of context-induced errors is not due to interference in the surface structure of language, but rather to its inner processing.

phrase imposed important restrictions on the selection of words. Moreover, lexical errors related to nouns were the most common. We thus aim at investigating lexical speech errors in Mandarin and seek to verify if the targets and the errors also tend to share the same part of speech, and how the errors are distributed across grammatical categories. Moreover, Jaeger (2004) found that lexical errors were more likely to share a semantic rather than a phonological relationship with the target, thus, a trade-off was suggested between these two types of lexical errors. This tendency has mostly been investigated in Germanic and Romance languages. We also provide a discussion about this matter.

The structure of this paper is listed as follows. In Section 2, we discuss the relevant findings in English and other languages. In Section 3, we explain the methodology for the collection, classification, and analysis of speech errors. Section 4 presents our findings, in which we catalog the possible major factors leading to lexical errors and rank them in terms of importance. Finally, a short comparison of lexical errors between Mandarin and other languages is conducted in Section 5, while Section 6 concludes the paper.

2. PREVIOUS RESEARCH ON SPEECH ERRORS

Traditional language production models generally include three stages: Conceptualization, formulation, and overt execution (Garrett 1975, 1980, 1984, 1993; Levelt 1989; Bock and Levelt 1994) [4] . During conceptualization, a message and a communicative intention are constructed. Processing at this level is thought to involve “pre-linguistic” representations. In the formulation stage, lemmas relevant to the meaning and function of lexical items are activated. Finally, the overt execution refers to the peripheral stage of articulation in the motor encoding, which

[4]Another model attested in the literature is Dell’s spreading activation model, in which the sentence production process is a set of interacting levels of linguistic elements, with the computational activity internal to each independent level (Dell 1984, 1986, 1988). The two models hold a similar view regarding lexical retrieval during language production, although the former deals with a unidirectional top-down process, while the latter deals with interactive bi-directional top-down and bottom-up processes.

converts the phonetic input into motor programs to be sent to the articulators. During this chain of processes, while the most appropriate lexical item is activated by the message representation, lexical items that are semantically related to the target lexical item are also activated; furthermore, when a target lexical item has been selected and its phonological form retrieved, words that are phonologically related to this target word are equally subject to the process of activation which has occurred in the case of the target word. When a lexical error occurs (e.g., *more* instead of *less*), it is because one of the semantically and/or phonologically similar words mistakenly receives higher activation than the intended word and the erroneous word is inserted into the functional/syntactic string (Jaeger 2004).

The majority of speech-error research has been done in relation to the stage of overt execution, i.e., that of the sound structures and of phonological systems, in Germanic languages (Berg 1987; Cutler 1982; Fromkin 1973, 1980; Shattuck-Hufnagel 1979; Stemberger 1983; among others). With regard to the lexical level of formulation, only a few studies have addressed lexical errors due to their scarcity in data (e.g., Hotopf 1980 for English; Arnaud 1999 for French; Jaeger and Wilkins 2005 for English children; Harley and MacAndrew 2001 for English; Jaeger 2004 for English adults and children; Wan and Ting, 2019; Tang and Wan 2019 for Mandarin). The findings of these studies are summarized in the following paragraphs.

Target words and error words almost always share the same grammatical category in the lexical errors found within Germanic and Romance languages (e.g., Fay and Cutler 1977; Fromkin 1973; Jaeger and Wilkins 2005; Harley and MacAndrew 2001; Jaeger 2004; Levelt 1989; Nooteboom 1973; Rapp and Caramazza 1998). A common explanation of this tendency is that the grammatical patterns of syntactic frames are planned and that there is a restriction in place so that frames can only be filled by words of a certain lexical category during the process of speech planning and production. Hotopf (1983) found a consistency of 98% for content words, while Fay and Cutler (1977) reported a consistency of 99.5% for content words. Harley and MacAndrew (2001) observed a consistency of 98% for content words in English, and similarly, Arnaud (1999) discovered a consistency of 97% for content words in French. With

regard to the variation between children and adults, Jaeger and Wilkins (2005) looked at lexical and context-induced speech errors, and claimed that the majority of the children's errors (92.5%) and adults' errors (94%) involved two words from the same lexical category.

Another main convergence within previous studies is the high rate of nouns and the low rate of verb errors. Jaeger (2004) also compared children's data with adult data sets and found that lexical errors involving content words were the most common. In addition, errors involving nouns (common noun, 55%; proper noun: 9.5%) outnumbered errors for other parts of speech within data from adults (c.f., Hotopf 1980). A similar finding was also made in Harley and MacAndrew's (2001) study. With regard to verb errors, verb errors constituted 9% of Fromkin's (1973) content-word substitution errors. Hotopf (1980) only observed 3% and 9% of verb errors in his English and German corpus, respectively. Harley and MacAndrew (2001) showed that nouns outnumbered verbs 4.8 to 1 in their corpus of semantic errors. Based on similar findings, Jaeger (2004) proposed a phonological explanation for why verb errors are not as frequent as other types of errors. In his corpus of errors in English, Jaeger (2004) found that tonic words[5] were frequently present and most often involved common nouns, followed by proper nouns, and then adjectives plus adverbs combined. Tonic words seldom occur as verbs since verbs generally serve as a function of the focus structure; thus, verb errors are rare. These findings relate to our main research question: do the targets and the errors also tend to share the same part of speech, and how are the errors distributed across grammatical categories in Mandarin?

With regard to the semantic and phonological relationships between the targets and errors, past studies slightly diverge in terms of frequency. Fay and Cutler (1977) noticed that purely phonology-induced lexical errors (i.e., errors that only share a phonological relationship with the target) tended to belong to the same grammatical category, have the same number of syllable structure, and the same stress pattern in their corpus. In terms of distribution, 81% of their English corpus of 226 lexical errors was purely phonologically related. However, different ratios are found in other studies. For instance, Hotopf (1983) reported an equal distribution

[5]The prosodic domain of each sentence is composed of a tonic word in English.

of semantically related pairs (52%) and phonologically related pairs (48%) in English data. Arnaud (1999) observed that 67% of errors were semantically related errors and 33% were phonologically related lexical errors in French, while Noteboom (1973) observed that 40% of errors were semantically related errors and 60% were phonologically related lexical errors in Dutch. On the other hand, Harley and MacAndrew (2001) and Stemberger (1989) discovered an opposite ratio as 76% and 83% of the errors in their corpora, respectively, were semantically related compared to only 24% and 17% phonologically related lexical errors. With regard to Mandarin, Wan (2007b, 2007c) found that phonological errors were 1.8 times, and seven times more frequent than semantic errors in corpora of natural errors and aphasic speech. As a theoretical explanation for the divergence of the ratios, Jaeger (2004) suggested that a trade-off occurs between phonological relationships and semantic relationships for lexical errors (for the adult corpus). Word pairs with a very close semantic relationship tend to have a looser phonological relationship, but word pairs with a less close semantic relationship tend to have more phonological properties in common. We will also provide a short discussion on this matter based on data in our corpus.

In summary, the data reported so far shows that lexical errors are more likely to honor the consistency of lexical category cross-linguistically. Moreover, target and error words commonly share a semantic relationship and occasionally have a phonological relationship. No large-scale similar studies known to the authors have been reported yet on natural speech in Mandarin, which is why we attempt to fill this gap.

3. MATERIALS AND METHODS

In this section, we first explain how the corpus of speech errors was built and annotated. Then we summarize how its content was analyzed.

3.1 Data Source and Annotation Process

The speech errors investigated in this study were extracted from a corpus of spontaneous speech recorded from a total of 102 native speakers

of Taiwan Mandarin located in Buffalo and Taipei between 1995 and 2009; thousands of hours have been recorded, and the spoken texts were manually transcribed and annotated by the first author and later her research team. The data for this paper are from a timeline of over ten years and language change may have occurred. However, we consider that the current data are reliable for the purpose of the current analysis since the tendencies in language processing that we are investigating are unlikely to have changed in such a short time frame (see Harley and MacAndrew 2001 for more discussion on this methodology). While other parts of the data set have been used in previous studies (Wan 2007a, 2007b, 2007c, 2016, 2019), a contribution of the current paper is to demonstrate the use of machine-learning techniques to process the spoken data semi-automatically accompanied by the addition of information on the transcription, word segmentation, part-of-speech tags, and phonetic alignment. Figure 1 provides a detailed overview of how the recent speech error corpus is structured.

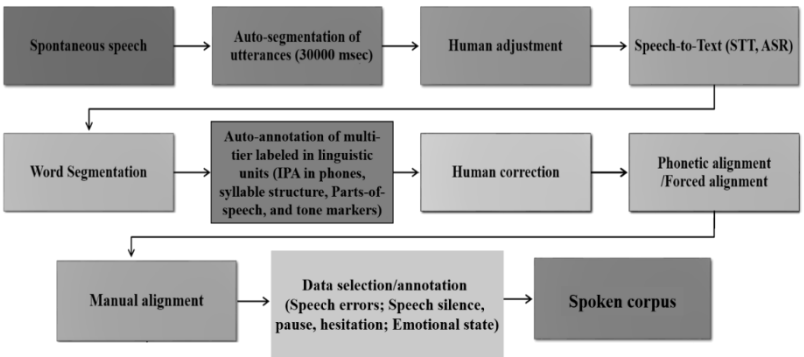

Figure 1. The process of the construction of the corpus.

Each speech session was recorded as an individual audio file. Then, each audio file was pre-processed into auto-segmented frames of 30 seconds at a sampling rate of 16-bit 44.1kHz. To avoid a phone or utterance cut across different files, the beginning and the end points of the frames were manually checked to ensure that each frame contained complete linguistic units in a fully contextual utterance. All the obtained frames were first sent to a Speech-to-Text (STT) system for transcription,

which resulted in an average accuracy of 70%.[6] The output of the STT system was then manually checked. Afterward, the entire transcript was automatically segmented using the CKIP parser (Ma and Chen 2003) and POS tagged by the CKIP tagger from the Chinese Knowledge and Information Processing Group.[7] The parsed and tagged transcription was also checked manually according to the criteria of word segmentation and POS tagging of the Academia Sinica Corpus[8] (Chinese Knowledge an Information Processing [CKIP] 1998), which are commonly applied in corpora such as the Linguistic Data Consortium (Ma and Huang 2006) and the Peking University Corpus (Huang et al 2008:2726). It is important to note that based on the CKIP criteria, the 'adjective' category only includes non-predicative adjectives, while the adjectives in the conventional definition (e.g., *beautiful* in English) are annotated as stative verbs. We are aware of the ongoing debate between linguists as to whether Mandarin has adjectives or not (McCawley 1992; Paul 2010; among others). Our study adopts the second view, since it is the one shared by the CKIP Corpus. This determination may have increased the ratio of verbs in our data compared to the data for other languages, and this matter will be further discussed in the results and discussion. Second, numerals (e.g., one, two, three) are labeled as determiners (numeral determinatives), and we therefore count them as function words in the analysis. Finally, the category of measure words regroups both sortal classifiers and mensural classifiers, even though they are two distinct categories in the literature

[6] The STT Package was further developed from the application pyTranscriber (https://github.com/raryelcostasouza/pyTranscriber) at the Phonetics and Psycholinguistics lab of the second author. The transcription of a 60-min audio file in Chinese characters requires 80 seconds. The accuracy of the output can vary a lot, depending on factors such as voice quality, noise clarity, gender, age, and/or speech speed of speakers. For instance, the accuracy rate varies between 40% and 95% depending on the combination of these factors. In general, an higher degree of accuracy in the result can be obtained in the case of a middle-aged male speaker talking at a rather slow speed in a lecture than when compared with other speakers. This variation in accuracy motivated the need for the manual checking of the output.

[7]The authors are thankful for the open source code from Professor Wei-yun Ma at the Institute of Information Science at Academia Sinica.

[8]The Academia Sinica Corpus contains 11,245,330 words and is the first fully POS-tagged balanced Chinese Corpus (Chen and Huang, 2016).

(Her 2012:1679).

Each audio file was auto-segmented into frames of 30 seconds and the generated transcriptions were then merged in Praat with automatically forced alignment at the phone level with an accuracy of 85 percent[9]. The output of the automatic alignment was adjusted manually when necessary. Finally, a phoneme dictionary was automatically generated to include the following information: Chinese characters, English gloss, POS tag, the utterance coded with vowel, syllable structure and tone, special remarks, the participant, and the emotional state of the participant. A sample of the annotated data is provided in Figure 2.

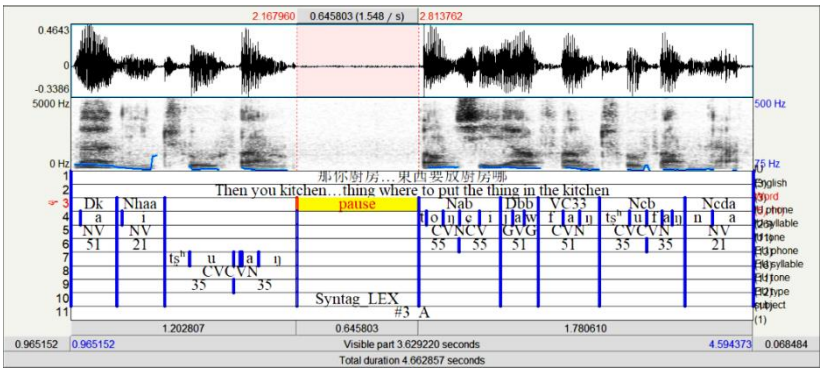

Figure 2. A sample of time-aligned speech errors in the corpus.

In this example, the speaker intended to say *dong1xi1* 'thing' but produced *chu2fang2* 'kitchen', followed by a pause of 645.8 milliseconds. The first tier displays the sentence in Chinese characters. The second tier shows the translation of the sentence. The POS tags are listed in tier three. Tier four to nine include information on phonemes, syllables, and tones.

[9]The current forced aligner is originally based on traditional method HMM with a hidden phoneme dictionary and further refined by the use of DNN models, the purpose of the system of which is to handle spontaneous speech in a challenging noisy environment and to reduce the cost of annotating labeled datasets by trained assistants. The second author appreciates the support of the original source software from Professor Li-hsin Ning and Professor Jiahong Yuan, and the deepest appreciation goes to Dr. Chain-wu Lee for providing training so that the accuracy rate amounted to higher than 85%.

Tier ten lists remarks for the analysis, while the last tier indicates the participant code.

3.2 Detection and Classification of Lexical Speech Errors

Based on a simultaneous analysis of the recordings and the transcriptions, 747 lexical errors (i.e., speech errors resulting in meaningful words) were identified and classified manually according to their phonological structures and semantic relationships (Arnaud 1999; Wan and Ting 2019). The criteria used for detecting these errors were mostly based on auto-correction/self-correction of the speakers, that is to say, when a speaker says an incorrect word and corrects herself/himself immediately afterward (as shown in Figure 2). Other cues such as repetition, speech pauses, phonetic silences, filler words were also considered.

Focusing on auto-correction may attract criticism since it is done manually and can be subject to personal perceptual biases. However, this study is only considering lexical errors, which are not subject to such limitations (see Harley and MacAndrew 2001 for a more detailed discussion on the matter). Another potential pitfall relates to the high rate of false negatives. That is to say, since our identification of speech errors was mostly based on repair by the speakers, we are likely to omit speech errors that occurred without repair. For instance, when a speaker says *bottle* instead of *book* in clauses such as *I want a book*. We preferred to not include such potential errors even if they were identifiable from the context since no formal cues of the existence of an error could be identified in such cases.

The following examples briefly show a sample of the speech errors from our corpus. Lexical errors may occur between two content words (1a), two function words (1b), or across content and function words (1c). By way of illustration in (1a), the noun *ming2ci2* 'noun' is substituted for the noun *dong4ci2* 'verb'. In (1b), the substitution occurs between the two determiners *na4* 'that' and *zhe4* 'this'. Finally, in (1c), the pronoun *wo3men0* is substituted for the noun *wo3fang1*, which represents a substitution of a content word by a function word. These errors are classified as non-context-induced lexical errors since the error is not present in the surrounding context.

(1) Lexical errors related to content words and function words

a. *zhe4xie1 tong1tong1 yao4 jia1 dan1shu4 ming2ci*
these all need plus singular noun
... *dong4ci2*
... verb
'these must be used with singular nouns … singular verbs'

b. *zhe4 ... na4 tian1zai4 bian4lun4 de0 shi2hou4* […]
this … that day at debate DE time
'this … that day during the debate […]'

c. *suo2yi3 wo3-men0 ... wo3fang1 ren4wei2* […]
therefore I-PL ... our side consider
'therefore, we … our side considers that […]'

Context-induced errors such as anticipations, perseveration, and reversals are not included in our analysis since they are not directly relevant to language processing. As an example in (2), the speaker commits a reversal between *duan4zhan4* 'short' and *kuai4le4* 'happiness', which results in a meaningless utterance in (2a), while the intended utterance was as in (2b).

(2) Context-induced error with reversal

a. *zhen1 shi4 kuai4le4 de0 duan3zhan4*
really be happiness DE short
(meaningless)

b. *zhen1* *shi4* *duan3zhan4* *de0* *kuai4le4*
really be short DE happiness
'(This) is indeed a short period of happiness.'

Finally, speech errors resulting in meaningless strings of phonemes are also excluded from this study, since instances of this type of error cannot be identified as words, and their semantics thus cannot be retrieved. As an example in (3a), *mei2you3* is collapsed into the shorter utterance *miu3* with one rhythmic beat (i.e., vowel) omitted; while in (3b), the two lemmas *man4* 'slow' and *ben4* 'stupid' are activated at the same time and phonologically merged into the same utterance *ban4*. In neither case is the error interpretable as a word.

(3) Speech errors resulting in meaningless strings of phonemes
a. *wo3* *mei2you3* *jiang3-dao4* → *wo3 miu3* *jiang3-dao4*
I not.have talk-arrive
'I have not talked about […]' (meaningless)
b. *hen3 man4/ben4* → *hen3 ban4*
very slow/stupid
'very slow/stupid' (meaningless)

Three additional categories of lexical errors are distinguished in the literature: semantic lexical errors, phonological lexical errors, and semantic-phonological lexical errors. Semantic lexical errors refer to errors that are only semantically related to the target, e.g., when a speaker says *I want more* instead of *I want less*. Phonological lexical errors are only phonologically related to the target. They typically result in meaningful words, but ungrammatical or meaningless utterances, e.g., when a speaker says *extra cushion is advised* instead of *extra caution is advised.* Semantic-phonological lexical errors are errors that can be considered both semantically and phonologically related to the target, e.g., when a speaker says *I want you* instead of *I warn you.* We do not distinguish between these three categories in the analysis since their boundaries are fuzzy and hard to identify systematically. Moreover, purely phonological errors are scarce since they occur by coincidence. However, we do provide a short discussion on these distinctions in Section 5.2.

4. RESULTS

Words can broadly be classified as content and function words. Content words include adjectives, adverbs, common nouns, proper nouns, and verbs; while function words contain the other categories such as articles, conjunctions, prepositions, pronouns, among others (Jaeger 2002:225). A comparison with different studies in terms of the content/function word distinction is shown in Table 1. Row one indicates that both the target and the error are content words (1a), the same logic applies for row two with function words (1b). On the other hand, row three involves speech errors in which either the target or the error is a function word and the other a content word (1c).

Table 1. The distribution of errors in content (C) /function (F) words[10]

Language	English	English	German	Mandarin
C	145 (92%)	233 (95%)	339 (90%)	649(87%)
F	12 (8%)	13 (5%)	37 (10%)	91(12%)
C/F	0 (0%)	0 (0%)	0 (0%)	7(1%)
Total	157	246	376	747

In our corpus, 649 lexical errors involved content words, 91 errors involved function words, and seven errors involved content-function words. This figure matches with those of other studies, as the majority of the errors are related to content words (87%). However, the ratio that we found is slightly lower than that in other studies. Our data also included a small percentage (1%) of content/function words speech errors, which were not attested in the corpora of previous studies. Such cases include changes from pronouns to nouns (1c), adverbs to conjunctions (*ke3neng2* 'may' → *ke3shi4* 'but'), adverbs to determinatives (*zen3me0* 'how/why' → *shi2me0* 'what'), prepositions to verbs (*bei4* 'BEI'→ *shou4dao4* 'receive'), and adverbs to postpositions (*si4xia4* 'in private' → *zhi1xia4* 'under'). Based on the 20,000 highest-frequency word count in Mandarin from the CKIP Group of Academia Sinica, the frequency of function

[10]The references for each study from left to right are Jaeger (2004), Hotopf (1983), Meringer (1903), and the current study.

words in Mandarin is 42%, and 58% for content words. This ratio might explain why Mandarin has a slightly higher error rate for content/function words, i.e., the frequency of function words is higher than in other languages, which results in a higher possibility of speech errors across content and function words.

Table 2 shows the number of word pairs with the same vs. different POS tag. We used the main thirteen categories of the CKIP corpus, but extracted the pronouns as a separate category, so that the differentiation between content and function words may be made. The first column refers to the POS of the target word, while each error is labelled in column two or three depending on its shared/distinct POS. The first five rows (Noun, Verb, Adjective, Adverb, Foreign word) indicate sub-categories of content words, whereas the following rows relate to sub-categories of function words. The final column lists the results from a by-category one-way Chi-square test. For instance, the one-way Chi-square test on the ratio of noun errors (390 vs. 10) shows a p-value smaller than .01, which indicates that the observed pattern may occur by chance at the frequency of less than once per hundred.

Table 2. The distribution of errors based on POS tags

Target	shared	distinct	Statistics
Noun	390	10	$\chi^2(1) = 361$, $p < .01$
Verb	213	9	$\chi^2(1) = 187.46$, $p < .01$
Adjective	2	3	N/A
Adverb	16	9	$\chi^2(1) = 1.96$, $p > .05$
Foreign word	0	1	N/A
Conjunction	0	0	N/A
De-construction	0	0	N/A
Pronoun	20	0	$\chi^2(1) = 20$, $p < .01$
Interjection	0	0	N/A
Measure	24	2	$\chi^2(1) = 18.615$, $p < .01$
Determiner	39	3	$\chi^2(1) = 30.857$, $p < .01$
Particle	0	0	N/A
Preposition	2	2	N/A
Postposition	2	0	N/A
Total	708 (95%)	39 (5%)	$\chi^2(1) = 599.14$, $p < .01$

We can thus infer that speech errors with nouns indeed tend to share a POS with the target word. The same tendency is observed in the general distribution or errors, as 708 (95%) errors share the same POS as the target word. Such is the case likewise with most POSs, i.e., verbs, pronouns, measures, and determiners, expect for adverbs. Categories with less than ten observations (e.g., adjectives, foreign words) are not tested statistically and are annotated with N/A in the final column. Similar observations are made when taking into account all the measurable categories (i.e., Noun, Verb, Adverb, Pronoun, Measure and Determiner) as $\chi^2(5) = 63.928$, $p < .01$).[11] Errors related to content words errors generally honor the consistency with regard to the POS (83%, 621/747), with only a minority of the data attested as violations (4%, 32/747). With regard to function words, most cases (12%, 87/747) show that both the target and error share the same POS, and only a few cases (1%, 7/747) have divergent POSs (and the POSs diverge in only a few cases (1%, 7/747)).

The visualization of the Pearson residuals of the Chi-square test on all the major categories allows us to visualize how the distribution of errors differs across POSs. For instance, most errors share the same POS as their target words; but we should also investigate if the errors in a specific POS category have a stronger (or weaker) tendency to share a POS with their target. In Figure 3, the y-axis indicates the value of the residuals, whereas the x-axis refers to the main POS categories. The colors refer to the degree of the tendency of having a shared/distinct POS. A higher value of residuals in a specific category indicates a stronger association with that category in comparison to the general distribution of the data. By way of illustration, the negative value of residuals for 'Distinct.POS' with nouns shows that errors in the noun category are negatively associated with distinct POSs in comparison with the data as a whole. In other words, the POS consistency is stronger with errors having nouns as target words. On the other hand, errors of adverbs are strongly associated with a divergence in the POS, which is what we found in Table 2. We do not investigate this observation further in the current paper due to the small sample size of adverbs (N=25); yet, we speculate that this is due to a by-chance higher

[11]Due to the small sample size of some cells, the Chi-square test may be inaccurate. However, the output of a Fisher test also shows that the observed frequencies are different from the frequency distribution expected under chance ($p < .01$).

ratio of phonologically related errors in adverbs. Further discussion is provided in Section 5.

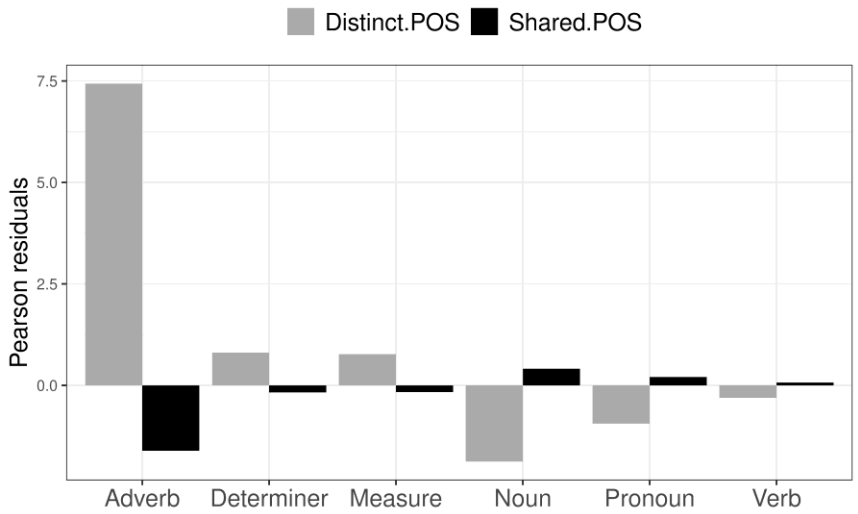

Figure 3. Visualizing the Pearson residuals for the main POS categories

Due to the high ratio of errors involving nouns and verbs, we also display the internal distribution of both categories. Following the definition of the CKIP Corpus, nouns may be classified into common nouns, proper nouns, place nouns, localizers and time nouns. In Table 3, errors belong to the same category as the target if they are annotated with an identical POS, e.g., if both are common nouns such as *ma2yi3* 'ant' and *ren2* 'people'. If the target and the error are from different sub-groups or different main groups, they are annotated as being of a different category, e.g., the target is a common noun (*dian4ti* 'elevator'), but the error is a place noun (*su4she4* 'dormitory'), or the target is a common noun (*qian2* 'money'), but the error is a stative verb (*qiong2* 'poor').

Table 3. The distribution of noun errors based on the POS tags

Target	Shared	Distinct	Statistics
Common	279	20	$\chi^2(1) = 224.35, p < .01$
Proper	32	0	$\chi^2(1) = 32, p < .01$
Place	41	9	$\chi^2(1) = 20.48, p < .01$
Localizer	7	2	N/A
Time	10	0	$\chi^2(1) = 10, p < .01$
Total	369 (92%)	31 (8%)	$\chi^2(1) = 285.61, p < .01$

The general tendency follows our previous observation: 369/400 (92%) of the errors belong to the same category as the target. The visualization of the Pearson residuals (Figure 4) further shows that errors of common nouns, proper nouns, and time nouns tend to be negatively associated with distinct POSs, especially proper nouns. As an example, the targets of common nouns only included 15 errors as place nouns, e.g., *yuan2gong1* 'employee' versus *gong1si1* 'company', and five errors as verbs, e.g., *fa1piao4* 'invoice' versus *fa1pang4* 'get fat'. However, errors of place nouns and localizers are more frequently related to distinct POSs, even though they still follow the general tendency of sharing POSs with the target. Errors with the targets as place nouns involved seven errors as common nouns, e.g., *xin1li3xi4* 'psychology department' versus *xin1li3xue2* 'psychology', with only two errors related to verbs, e.g., *xue2xiao4* 'school' versus *xue2xi2* 'learn'. The two errors in the category of localizers involved substitutions of common nouns, e.g., *di3xia4* 'underneath' versus *wu3tai2* 'stage'. We do not discuss the errors of localizers due to the small sample size (N=9). As for the errors related to place nouns, we do not investigate them in the current paper, but we speculate that their occurrence is related to the fact that common nouns can easily be used in spatial metaphors. The boundary between common nouns and place nouns is thus fuzzier than between other categories such as that between proper nouns and common nouns.

Figure. 4. Visualizing the Pearson residuals for the noun categories

In Table 4, we breakdown the subcategories of verbs into two main groups: active verbs (139 tokens) and stative verbs (83 tokens), the second category being related to the traditional view of adjectives such as 'beautiful' in English.

Table 4. The distribution of verb errors based on the POS tags. The following terms are abbreviated: ps = pseudo, obj = object

Target verb	Shared	Distinct	Statistics
Active intransitive	21	7	$\chi^2(1) = 7$, $p < .01$
Active causative	0	1	N/A
Active ps-transitive	0	3	N/A
Active transitive	44	20	$\chi^2(1) = 9$, $p < .01$
Active + locative obj	0	2	N/A
Ditransitive	3	6	N/A
Active + sentential obj	13	8	$\chi^2(1) = 1.2$, $p > .05$
Active + verbal obj	0	3	N/A
Classificatory	1	2	N/A
you3	1	2	N/A
shi4	0	2	N/A
Stative intransitive	52	10	$\chi^2(1) = 28$, $p < .01$
Stative causative	0	2	N/A
Stative ps-transitive	0	0	N/A
Stative transitive	1	7	N/A
Stative + sentential obj	5	4	N/A
Stative + verbal obj	0	2	N/A
Total	141 (64%)	81 (36%)	$\chi^2(1) = 16$, $p < .01$

Verb errors included 222 speech errors, with the majority (213/222; 96%) respecting consistency with regard to the part of speech, i.e., in most cases the target and the error were both verbs. However, when broken down into sub-groups, the ratio changes: in total, only 64% (141/222) of speech errors belonged to the same sub-group as the target, e.g., when both the target and error were active transitive verbs. We explain this fact by the very detailed classification of the content of the Academia Sinica Corpus utilized in our study: the more categories that there are in place, the higher the chances of interpreting target and errors as members of different sub-groups, even though they are still verbs. Due to the same reason, we do not visualize the Pearson residuals of this table, as most categories have too few tokens. Yet, we still observe that the consistency of the POS is maintained between active and stative verbs, i.e., there is a

high tendency for the error to maintain itself as an active verb (89%, 124/139). A similar effect is observed in the case of stative verbs (77%, 64/83).

5. DISCUSSION

In this section, we first summarize how our Mandarin data matches with other studies in terms of error-type distribution. As an example, we discuss the high occurrence of verb errors in Mandarin with comparison to other languages. Then, we discuss the influence of phonological factors on the speech errors in our Mandarin data.

5.1 A General Comparison with Previous Studies

Our results show that two words mutually involved in a lexical error are usually (95%) of the same POS in Mandarin. Moreover, most of the errors are related to substitutions between two content words (91%). These observations match with the findings of previous studies (Fromkin 1973; Nooteboom 1973; Fay and Cutler 1977; Garrett 1980; Levelt 1989; Rapp and Caramazza, 1998; Harley and MacAndrew 2001; Jaeger 2004). The grammatical patterns of a specific phrase pre-impose important restrictions on the selection of words; the activated words thus tend to conform with the pre-specified lexical category to avoid a violation of the planned grammatical structure (Levelt 1989; Jaeger 2004). Likewise in terms of noun errors, our findings correlate with previous studies by showing a higher frequency of noun errors (54%, 400/747) compared to verb errors (30%, 222/747) and other POSs (Harley and MacAndrew 2001; Jaeger 2004).

Our results based on Mandarin errors diverge with studies on other languages in regard to the higher rate of verb errors (30%, 222/747). Even if we exclude the potentially controversial stative verbs that may be classified as adjectives depending on the theoretical definition, verb errors still account for 19% (139/747) of the data, which outnumbers the rough ratio of 9% of data attested cross-linguistically (e.g., Fromkin 1973; Hotopf, 1980; Harley and MacAndrew 2001; please refer to Section 2 for

more details). This is not entirely surprising based on the high frequency of verbs in Mandarin speech. For instance, verbs account for 27% of the 20,000 most frequent words in the CKIP Corpus and also occur at the highest frequency (23%) before nouns (22%) in the Taiwan Mandarin Conversational Corpus (Tseng 2013). A more frequent usage of verbs would by logic result in a higher possibility of making speech errors[12]. However, not all the error rates are predicted by the frequency of occurrence in corpora. By way of illustration, the ratio of noun errors in our data (54%) exceeds by far the ratio of nouns in the CKIP Corpus (18%, N=20,000). The occurrence of tokens in the entire corpus is therefore not entirely reliable.

We suggest two additional explanations for the high frequency of verb errors in Mandarin. First, Hotopf (1983), Bock and Levelt (1994), and Jaeger and Wilkins (2005) have argued that verbs are the core of a proposition and a clause, and the selection of the right verb may be the key decision in planning an utterance. This is also true in Mandarin; however, a grammatical sentence in Mandarin does not have the same level of restriction on verbs compared to English, in which verbs are the focus structure of sentences. Second, in addition to the centrality of verbs in the proposition, Jaeger (2004) also suspected that the tonic stress in phonological phrases increases the probability of speech errors. Since tonic stress is far more likely to fall on a noun in English than on any other word class, the chances of making errors is thus higher in the case of nouns than in verbs in English. Mandarin does not have a prosodically or pragmatically prominent tonic syllable word at the end of a phrase. Since tone in Mandarin is linked more closely with lexical items**,** while stress is linked more closely with phrasal prosody (Wan 2007a), these elements do not have the same syntagmatic organizing status in phrases. The distribution of verb errors thus differs between languages such as English and Mandarin.

[12] Another possible explanation is the higher frequency of verb errors generated by phonological lexical errors (malapropism). As an example, Fay and Cutler (1977) found a higher ratio of verb errors (32%) in malapropisms. Yet, our data only include 41 phonological errors (among which 31 errors show consistency in the POS and only nine errors are related to verbs); the picture thus does not change much even by excluding those cases.

5.2 The Interaction of Semantic and Phonology

We suggest that the distribution of semantics-induced and phonology-induced errors should not be approached as a binary choice (i.e., is the error related to semantics or phonology**?**). Both semantics and phonology can have an effect on the occurrence of speech errors, which makes it complicated to define which factor plays the major role, which is also why we did not include this distinction in the main analysis. As a preliminary investigation, the lexical errors of our Mandarin corpus were manually tagged by three different annotators in the following way: if both the target and the error tend to have either a semantic [+sem, -phon] or phonological [-sem, +phon] relationship, or both [+sem, +phon][13]. If all the annotators agreed, the label was kept. In the case of partial disagreement, it was determined to follow the consensus of the majority. That is to say, the annotation used by two annotators was kept. In the case of full disagreement, the error was not annotated (this type of case did not actually occur during the annotation process). We are aware that this manual method should ideally be replaced by automatic quantitative classification. However, since we only aim at providing a preliminary overview of the matter, we consider that the method of manual annotation used by previous studies is sufficient for the moment. By way of illustration, (4a) shows a case where the target (*ming2cheng1* 'name') and the error (*shu4zi4* 'number') are considered as semantically related(,) but phonologically unrelated ([+sem, -phon]). In (4b), it is the opposite case, as the target (*lou2ti1* 'stairs') and the error (*OT* 'optimality theory') are phonologically related**(,)** but semantically unrelated ([-sem, +phon]). Finally**(,)** in (4c), the target (*tong3ji4* 'statistics') and the error (*zong3ji4* 'sum') are semantically and phonologically related ([+sem, +phon]).

(4) Different types of semantic and phonological errors in Mandarin

a.	*zhe4*	*ge0*	*shu4zi4*	...	*zhe4*	*ge0*	*ming2cheng1*	*gei3 ni3*
	this	CLF	number		this	CLF	name	give you

'this number … this name is for you'

[13]Errors of the [-sem, -phon] type generally relate to context-induced errors and are not included in the current study.

b. *li3mian4* *you3 OT* ... *lou2ti1* *ma*
inside have OT stairs Q
'are there OT … stairs in there?'

c. *qu4nian2* *hai2* *shi4* *jin1nian2* *de* *zong3ji4* ... *tong3ji4*
last year or be this year DE sum statistics
li3mian4
inside
'in the sum … survey of last year or this year'

In our Mandarin corpus, purely phonological errors of the [-sem, +phon] type account for 5% (41/747) of the data, purely lexical errors of the [+sem, -phon] type account for 47% (348/747), and mixed-type errors [+sem, +phon] account for 48% (358/747) of the data. Purely semantic errors are much more frequent than purely phonological errors, which matches with the findings in other languages that semantic lexical errors are more frequent than phonological lexical errors. However, the definition of the mixed-type has a major effect on the interpretation of the data. On one hand, if we consider that mixed-type errors are semantic errors, the distribution found in Mandarin matches with previous studies in other languages, i.e., the majority of the lexical errors are semantically related (93%), and only 5% of the errors show a pure phonological relation. On the other hand, if we consider that mixed-type errors are phonological errors, the ratio of semantic and phonological errors becomes more balanced (47% vs. 53%). While we do not provide a systematic way to further split the mixed-type errors into semantic and phonological errors, we point out the fact that this high ratio of mixed-type errors differs from what Jaeger (2004) found in English, where the word pairs involved in lexical errors with a very close semantic relationship tended to have a looser phonological relationship, but word pairs with a less close semantic relationship tended to share more phonological properties. This high ratio of mixed-type errors contradicts the hypothesized trade-off in the occurrence of semantic and phonological lexical errors.

While the targets and errors of semantic errors are by default not related phonologically, we can still investigate the purely phonological errors to assess their phonological similarities with the targets. Within the 41 phonological errors, 40 cases shared the same number of syllables. To

be more precise, two examples involved one-syllable word substitution (e.g., *wan51* 'ten thousand' substituted for *wei4* [CLF-person]), 33 cases involved two-syllable word substitutions (e.g., *lou2ti1* 'stairs' substituted for *OT* 'OT [optimality theory]'), and five cases involved three-syllable word substitution (e.g., *kan4bu4qi3* 'look down upon' substituted for *kan4bu4qing1* 'cannot see clearly'). Only one example showed a case in which the target and the error did not share the same number of syllables. In terms of syllable-internal phonological properties, Table 5 displays how many phonological errors shared the same initial consonant, ended in the same rhyme, or carried the same tone. As an overview, 33 word pairs started with the same initial consonants, and seven cases show different initial consonants in the first syllable. There is a statistically significant difference in the phonological errors where the first syllables share the same initial consonants (χ^2 (1)=16.9, p<.01). However, there are no significant differences in the word pairs sharing the same initial consonants in the second syllable (χ^2 (1)=1.6, p>.05).[14] This observation leads to the inference that the initial consonants in the first syllable are more important than the ones in the following syllables.

Table 5. Shared phonological properties within phonological errors

	S_1	D_1	S_2	D_2	S_3	D_3	S_4	D_4
Initial consonants	33	7	15	23	3	2	1	0
Rhymes	31	9	17	21	2	3	1	0
Tones	35	5	29	9	3	2	1	0

In the target-error word substitution involving rhymes, 31 cases show that the word pairs end with the same rhymes, and nine cases show different rhymes in the first syllable, amounting to a significant difference (χ^2 (1)=12.1, p<.01). There are no significant differences when the word pairs involve the same or different rhymes in the second (χ^2 (1)=0.4, p>.05) or in the third syllable (χ^2 (1)=0.2, p>.05). This leads to the inference that the phonological elements in the first syllable are more likely to cause

[14]The statistical test is not applicable in the case of the third syllable due to its small data size.

speakers to produce erroneous word selections. Finally, in the target-error word substitution involving tone, 35 cases show the word pairs have the same tone in the first syllable, and five cases show the involvement of different tones in the same syllable, for a significant difference (χ^2 (1)=22.5, p<.01), and for the second syllable, there is also a significant difference in the word pairs involving the same tone or a different tone (χ^2 (1) =10.5, p<.0). This observation suggests that the tones of the intended words tend to be preserved more than the original rhymes or initial consonants in the case of the second syllable.

As a summary, Jaeger (2004) proposed that when there is no semantic relatedness in target-error word pairs, there is a significant agreement in phonological relatedness. Compared to the results of English studies, Mandarin does not present many phonological errors, and thus the data is not sufficient to directly support such a claim; however, the analysis of the phonological properties in phonological errors does suggest that two lexical items that are not semantically related are likely to be substituted for one another if they meet the following criteria: having the same initial consonant, the same rhyme and/or the same tone in the first syllable. A higher tendency for the preservation of the tone in the second syllable is also found in our data.

6. CONCLUSION

The present study provided data of lexical speech errors in Mandarin and compared the observed patterns with previous studies on other languages. Our findings show that lexical speech errors in Mandarin also tend to preserve the POS of the intended word and that lexical speech errors related to nouns are the most common. Both facts have been equally attested cross-linguistically. However, verb errors are more common in our Mandarin corpus than in languages such as English and German. We hypothesize that this is due to differences in the prosodic systems across languages, which supports the relevance of investigating speech errors in languages of different families. We also investigated the distribution of semantics-induced and phonology-induced lexical speech errors in Mandarin. Half of the errors in our corpus show a potential for a

contribution from either semantics or phonology, which suggests that speech errors should not be considered in a binary manner as either semantics-induced or phonology-induced. This finding does not directly contradict the trade-off between semantics-induced and phonology-induced errors suggested in previous studies. However, it does enhance the motivation to provide more precise criteria to distinguish these two types of errors in future studies.

REFERENCES

AutoTag, C.K.I.P. 1998. *Chinese knowledge information processing group*. Academic sinica: Taiwan.

Alderete, John., and Monica Davies. 2019. Investigating perceptual biases, data reliability, and data discovery in a methodology for collecting speech errors from audio recordings. *Language and speech* 62(2): 281-317.

Alderete, John., Queenie Chan., and H Henny Yeung. 2019. Tone slips in Cantonese: Evidence for early phonological encoding. *Cognition* 191: 103952.

Alderete, John., and Paul Tupper. 2018. Connectionist approaches to generative phonology. In *The Routledge handbook of phonological theory*, eds. Anna Bosch, and Stephen J. Hannahs, pp. 360-390. New York: Routledge.

Arnaud, Pierre J. 1999. Target–error resemblance in French word substitution speech errors and the mental lexicon. *Applied Psycholinguistics* 20(02): 269-287.

Berg, Thomas. 1987. *A Cross-linguistic Comparison of Slips of the Tongue*. Bloomington, IN: Indiana University Press.

Bock, J. Kathryn., and Willem Levelt. 1994. Language production: Grammatical encoding. In *Handbook of Psycholinguistics*, ed. Morton Ann Gernsbacher, pp.945-984. Orlando, FL: Academic Press.

Boersma, Paul., and David Weenink. 2019. Praat: doing phonetics by computer [Computer program] (Version 6.0.53). web. 6 June. 2019. (http://www.praat.org/)

Chen, Keh-Jiann., and Chu-Ren Huang. 2016. Academia sinica balanced corpus. Brill.

CKIP (Chinese Knowledge and Information Processing). 2004. Part-of-speech analysis of Academia Sinica Balanced Corpus of Modern Chinese, *Technical Report* (no.93-05), version 3.

Cutler, Anne. 1982. The reliability of speech error data. In *Slips of the Tongue and Language Production*, ed. Anne Cutler, pp.7-28. Amsterdam: Mouton.

Dell, Gary S. 1984. Representation of serial order in speech: evidence from the repeated phoneme effect in speech errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10(2), 222.

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

Dell, Gary S. 1988. The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of memory and language* 27(2): 124-142.

Fromkin, Victoria A. 1973. *Speech errors as linguistic evidence*. The Hague: Mouton.

Fromkin, Victoria A. 1980. *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand*. NY: Academic Press.

Fay, David., and Anne Cutler. 1977. Malapropisms and the structure of the mental lexicon. *Linguistic Inquiry* 8: 505-520.

Garrett, Merril F. 1975. The analysis of sentence production. In *Psychology of learning and motivation*, ed. Gordon Bower, pp.133-177. NY: Academic Press.

Garrett, Merril F. 1980. Levels of processing in sentence production. In *Language Production*, ed. Brian Butterworth, pp.177-220. London: Academic Press.

Garrett, Merril F. 1984. The Organization of processing structure for language production. In *Biological Perspectives on Language*, eds. David Caplan, Abdré R Lecourse, and Alan Smith, pp.172-193. Cambridge: MIT Press.

Garrett, Merril F. 1993. Errors and their relevance for models of language production. In *Linguistics disorders and pathologies*, eds. Gerthard Blanken, Jürgen Dittmann, Hannelore Grimm, John Marshall and Claus Wallesch, pp.72-92. Berlin: de Gruyter.

Harley, Trevor. 2006. Speech Errors: Psycholinguistic Approach. In *Encyclopedia of Language and Linguistics*, ed. Keith Brown, pp.739-745. Amsterdam: Elsevier Science

Harley, Trevor., and Siobhan MacAndrew. 2001. Constraints upon word substitution speech errors. *Journal of psycholinguistic research* 30: 395-418.

Her, One-Soon. 2012. Distinguishing classifiers and measure words: A mathematical perspective and implications. *Lingua* 122(14): 1668-1691.

Hotopf, William. 1980. Semantic similarity as a factor in whole-word slips of the tongue. In *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand*, ed. Victoria Fromkin, pp.97-109. NY: Academic Press.

Hotopf, William. 1983. Lexical slips of the pen and tongue: What they tell us about language production. *Language production* 2: 147-199.

Huang, C.-T. James., Audery Li, and Yafei Li. 2009. *The syntax of Chinese*. Cambridge: Cambridge University Press.

Jaeger, Jeri. J., and David Wilkins. 2005. Semantic relationships in lexical errors. In *Kid's slips: What young children's slips of the tongue reveal about language development*, ed. Jeri J Jaeger, pp.311-384. Mahwah, NJ: Lawrence Erlbaum Associates.

Jaeger, Jeri J. 2004. *Kids' slips: What young children's slips of the tongue reveal about language development*. Mahwah, NJ: Lawrence Erlbaum Associates.

Levelt, Willem. J. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT press.

Li, Peng-Hsuan., Tsu-Jui Fu., and Wei-Yun Ma. to appear. *Why Attention? Analyze BiLSTM Deficiency and Its Remedies in the Case of NER*. AAAI 2020. https://arxiv.org/abs/1908.11046

Ma, Wei-Yun., and Keh. Jiann Chen. 2003. Introduction to CKIP Chinese word segmentation system for the first international Chinese Word Segmentation Bakeoff. In *proceedings of the second SIGHAN workshop on Chinese language processing (Volume 17)*, pp.168-171, Sapporo, Japan.

Ma, Wei-Yun., and Chu-Ren Huang. 2006. Uniform and Effective Tagging of a Heterogeneous Giga-word Corpus. In *proceedings of the 5th International Conference on Language Resources and Evaluation (LREC-5)*, pp. 2182-2185, Genoa, Italy.

McCawley, James. D. 1992. Justifying part-of-speech assignments in Mandarin Chinese. *Journal of Chinese linguistics* 20(2): 211-246.

Meringer, Rudolf. 1908. *Aus dem Leben der Sprache: Versprechen, Kindersprache, Nachahmungstrieb*. Berlin: Behr.

Nooteboom, Sieb G. 1973. The tongue slips into patterns. In *Speech Errors as Linguistic Evidence*, ed. Victoria. Fromkin, pp.87-95. The Hague: Mouton.

Rapp, Brenda., and Alfonso Caramazza. 1998. Lexical deficits. In *Acquired aphasia*, ed. Martha Sarno, pp.187-227. San Diego, CA: Academic Press.

Shattuck-Hufnagel, Stefanie. 1979. Speech errors as evidence for a serial ordering mechanism in sentence production. In *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*, eds. William. E. Cooper and Edward. C. T. Walker, pp. 295-342. Hillsdale, N. J.: Erlbaum.

Stemberger, Joseph. 1983. Speech Errors and Theoretical Phonology: A review. Unpublished manuscript, Carnegie-Mellon University. Distributed by the Indiana University Linguistics Club, Bloomington, IN.

Stemberger, Jospeh. 1989. Speech errors in early child language production. *Journal of Memory and Language* 28: 164-188.

Tang, Marc., Wan, I-Ping. 2019 Predicting Speech Errors in Mandarin Based on Word Frequency. In *From Minimal Contrast to Meaning Construct*, eds. Qi Su and Weidong Zhan, pp.289-303. Springer Singapore: Singapore.

Tseng, Shu-Chuan. 2013. Lexical coverage in Taiwan Mandarin conversation. *Computational Linguistics and Chinese Language Processing* 18(1): 1-18.

Paul, Waltraud. 2010. Adjectives in Mandarin Chinese: The rehabilitation of a much ostracized category. In *Adjectives. Formal analyses in syntax and semantics*, eds. Patricia Cabredo-Hofherr and Ora Matushansky, pp.115-152. Amsterdam: Benjamins.

Wan, I-Ping. 2007a. On the phonological organization of Mandarin tone. *Lingua* 117: 1715-1738.

Wan, I-Ping. 2007b. Mandarin speech errors into phonological patterns. *Journal of Chinese Linguistics* 35(1): 185-220.

Wan, I-Ping. 2007c. *A phonological investigation in speech errors and aphasic speech in Mandarin*. Taipei: Crane Publishing.

Wan, I-Ping. 2016. Consonant features in Mandarin speech errors. *Concentric* 42(2): 1-39.

Wan, I-Ping and Ting, Jen. 2019. Semantic relationships in Mandarin speech errors. *Taiwan Journal of Linguistics* 17(2): 33-66.

Zwicky, Arnold. 1982. Classical malapropisms and the creation of a mental lexicon. In *Exceptional Language and Linguistics*, eds. Loraine K. Obler and Lise Menn, pp.115–132. New York, NY: Academic Press.

[Received 9 April 2020; revised 12 August 2020; accepted 15 October 2020]

I-Ping Wan
Graduate Institute of Linguistics
National Chengchi University
Taipei, Taiwan, 116
ipwan@nccu.edu.tw

Marc Allassonnière-Tang
Lab Dynamics of Language
University Lyon 2
Lyon, France, 69363
marc.tang@univ-lyon2.fr

APPENDIX

Full list of the POS tagset with examples

POS	Sub-category	Example
Adjective	Non-predicative adjective	*tian1sheng1* 'innate'
Adverb	Adverb	*ke3neng2* 'possibly'
	Pre-verbal adverb of degree	*xiang1dang1* 'quite'
	Post-verbal adverb of degree	*wan4fen1* 'extremely'
	Sentential adverb	*zong3er2yan2zhi1* 'in sum'
	Aspectual adverb	*kan4 le0* 'looked at'
	Quantitative adverb	*jin3* 'only'
Conjunction	Conjunctive conjunction	*gen1* 'and'
	Correlative conjunction	*dan4shi4* 'but'
	Conjunction	*deng3deng3* 'etc'
De-construction		*de0, zhi1, de2, di0*
Determiner	Demonstrative determinative	*zhe4* 'this'
	Quantitative determinative	*quan2bu4* 'all'
	Post-quantitative determinative	*yi3shang4* 'above'
	Specific determinative	*mou3* 'certain'
	Determinative	*liang3* 'two'
Foreign word		*Delete*
Interjection		*o2*
Measure		*yi4 zhi1 bi3* 'a pen'
Noun	Common noun	*shu1* 'book'
	Proper noun	*li3xiao3long2* 'Bruce Lee'
	Place noun	*dong4wu4yuan2*

		'zoo'
	Localizer	*qian2* 'front'
	Time noun	*wan3shang4* 'evening'
	Pronoun	*ta1* 'he/she'
Particle		*ma1*
Postposition		*li3lun4shang4* 'in theory'
Preposition		*cong2* 'from'
Verb	Active intransitive verb	*tiao4wu3* 'dance'
	Active causative verb	*xuan2zhuan3* 'spin'
	Active pseudo-transitive verb	*tui4kuan3* 'refund'
	Active transitive verb	*qu3xiao1* 'cancel'
	Active verb with locative object	*deng1lu4* 'login'
	Distransitive verb	*mai4* 'sell'
	Active verb with sentential object	*wen4* 'ask'
	Active verb with verbal object	*ju4jue2* 'refuse'
	Classificatory verb	*zao4cheng2* 'result'
	Stative intransitive verb	*nu3li4* 'word hard'
	Stative causative verb	*chan3sheng1* 'create'
	Stative pseudo-transitive verb	*wei2zhu3* 'focus'
	Stative transitive verb	*huai2nian4* 'miss'
	Stative verb with sentential object	*you2yu4* 'hesitate'
	Stative verb with a verbal object	*xi2guan4* 'get used to'
	you, shi	*you3, shi4* 'have, be'

台灣華語語意語誤解析

萬依萍[1]、唐威洋[2]
[1]國立政治大學
[2]里昂第二大學

本研究主要利用 747 筆華語語意語誤資料，以中研院詞性分類作為機器訓練之模型基底，並搭配其他具有語意語誤之國際語料庫做一比較，結果發現語言產製中仍出現些許世界通用法則。華語在詞性分類表現與其他外語呈現相同現象，尤其是在實詞中，名詞代換的語意語誤佔絕大多數，然而，華語中的語意語誤中，動詞代換明顯比其他外語高出許多，似乎顯現出詞頻效應。

關鍵字：語意語誤、華語、名詞、動詞