Graded effects of first and second language orthography on pronunciation during second language acquisition

by

Kailen Thomas William Shantz

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

Department of Linguistics
University of Alberta

Abstract

Recent years have seen a rise in the number of studies investigating orthographic influences in second language (L2) populations. This research has found that orthography can impact the acquisition, production and perception of L2 sounds and spoken word-forms. These findings suggest that orthography and its relationship to phonology may play an important role in the instructed learning of L2 pronunciation. The existing research remains limited, however, by its inattention to possible gradient effects of orthography, and a lack of studies exploring if and how orthographic effects interact with L2 exposure.

In order to address these shortcomings, this study aims to assess how frequency-based and probabilistic orthographic measures influence the accuracy of L2 pronunciation in a phonological decoding task for native English speaking learners of German, and a control group with no formal knowledge of German. Data are collected using a reading-aloud task at three time points: 0 months of instruction (control group), 1 month of instruction, and 3 months of instruction. The orthographic predictors of interest are the availability, reliability and validity of English and German one-, two- and three-letter sequences as cues to the German phones expected in the test words. In order, these measures are operationalized as the frequency of letter n-grams, the frequency with which letter n-grams co-occur with the expected phones, and the contingency (delta P) between letter n-grams and the expected phones.

Results indicate that the orthographic predictors of interest influence pronunciation accuracy during phonological decoding, and that this influence is

graded. Findings also reveal interactions with L2 exposure, offering insight into how orthographic effects change during second language acquisition. Finally, one-, two-, and three-letter sequences were found to variably predict pronunciation accuracy, suggesting possible differences in how letter sequences of varying sizes are processed by L2 learners.

Preface

This thesis is an original work by Kailen Shantz. The research project, of which this thesis is a part, received research ethics approval from the University of Alberta Research Ethics Board, Project Name “Reading in a second language”, No. Pro00040755, 28/08/2013.

Acknowledgements

I would like first and foremost to thank my supervisor, Juhani Järvikivi, without whose help and guidance this project would never have been possible. I am also indebted to Ryan Podlubny, for graciously lending me his microphone for an entire semester, and to the various instructors of German 111, in particular Crystal Sawatzky, who permitted me to recruit from their classes. I would also like to thank Terry Nadasdi and the various members of the Centre for Comparative Psycholinguistics and of the Alberta Phonetics Laboratory for their suggestions and feedback on this project.

Thanks are also owed to those who helped to provide encouragement and emotional support, and to remind me that my project was 'cool' when I was feeling disenchanted: Kristina Geeraert, Claudia Heinrich, Caelan Marr, Vince Porretta, Michelle Sims, Tamara Sorenson Duncan, Anne-Michelle Tessier, Yasemin Tulpar and Richard Yanaky. I am also hugely grateful to my family, for continuing to support me as I attempt never to leave university, and to my partner, Phil Oel, for being an unwavering source of support, encouragement and distraction when I need it.

Finally, I am thankful for the generous funding provided to me by the Social Sciences and Humanities Research Council of Canada, by the Faculty of Graduate Studies and Research, and by the Alberta Government, which helped support this project.

Table of Contents

List of Tables

List of Figures and Illustrations

0. Introduction

Decades worth of psycholinguistic research has investigated how various aspects of orthography, such as regularity and consistency, influence linguistic processes, finding effects in visual word recognition (e.g., Forster & Chambers, 1973; Baron & Strawson, 1976; Frederiksen & Kroll, 1976; Stanovich & Bauer, 1978; Ferrand & Grainger, 1992, 1993, 1994, 2003; Grainger, Kiyonaga, & Holcomb, 2006), speech perception (e.g., Seidenberg & Tanenhaus, 1979; Dijkstra, Roelofs, & Fieuws, 1995; Ziegler & Ferrand, 1998; Hallé, Chéreau, & Segui, 2000; Ventura, Morais, Pattamadilok, & Kolinsky, 2004; Ziegler, Ferrand & Montant, 2004; Pattamadilok, Morais, Ventura, & Kolinsky, 2007; Taft, Castles, Davis, Lazendic, & Nguyen-Hoan, 2008), and speech production (e.g., Tanenhaus, Flanigan & Seidenberg, 1980; Lupker, 1982; Damion & Bowers, 2003; Brewer, 2008; Rastle, McCormick, Bayliss, & Davis, 2011). Much of this research has also been done in bilingual populations, finding for example that performance in lexical decision tasks is modulated by the cross-linguistic orthographic similarity of words in a bilingual's languages (e.g., Bijeljac-Babic, Biardeau, & Grainger, 1997; Dijkstra, Van Jaarsveld, & Ten Brinke, 1998; Van Heuven, Dijkstra, & Grainger, 1998; Dijkstra, Grainger, & Van Heuven, 1999; Dijkstra, Timmermans, & Schriefers, 2000; Lemhöfer & Dijkstra, 2004).

While little attention has been paid historically to the effects of orthography during second language acquisition (SLA), recent years have seen a rise in research on such effects, primarily addressing the relationship between first language (L1) and second language (L2) orthography and the ability to accurately produce and perceive L2 sounds and word-forms (e.g., Erdener & Burnham, 2005; Bassetti, 2006, 2007; Hamada & Koda, 2008; Escudero & Wanrooij, 2010; Hayes-Harb, Nicol, & Barker, 2010; Showalter & Hayes-Harb, 2013). This research, however, remains limited by its inattention to possible gradient effects arising from frequency and probabilistic statistics which describe the relationship between orthography and phonology, tending instead to characterize this relationship categorically as consistent vs. inconsistent, or transparent vs. opaque[1]. In addition, these studies tend not to consider how L2 exposure or proficiency might interact with various predictors, as a result of which there is a lack of information about how orthographic effects in L2 processing might emerge and change over time.

Importantly, usage-based accounts of language learning (e.g., MacWhinney, 1987; Lieven, Behrens, Speares, & Tomasello, 2003; Tomasello, 2003; Ellis & Larsen-Freeman, 2009) make predictions about how frequency-based and probabilistic statistics influence language acquisition. These accounts hold that language is an emergent phenomenon shaped by the linguistic input to which we are exposed. That is, the statistical properties of the sounds, words, phrases and sentence structures that a language learner encounters will influence the structure of that learner's developing language through implicit learning. Language acquisition is thus seen as a process of learning to associate co-

[1] But see Ellis & Beaton (1993), who consider letter bigram frequency and orthographic similarity in L2 vocabulary acquisition; Hamada & Koda (2008), who consider mean letter bigram frequency in their stimuli design; and Rafat (2011), who proposes effects of L1 grapheme-phoneme correspondence frequency on L1 transfer.

occurring linguistics items with one another. This is also referred to as form-function mapping (e.g., MacWhinney, 1992), where a linguistic form (e.g., a string of letters, phonemes or morphemes) serves as a cue to one or more linguistic functions (e.g., a particular meaning to be conveyed), and vice versa. In other words, language acquisition involves learning which forms are the cues that predict certain functions, and which functions are the cues that predict certain forms. For example, in order to infer a past tense construal of events in the sentence '*John had walked to the store alone*', a reader must have learned that the orthographic forms '*had'* and '*-ed*' are cues that predict the function 'past tense'. The speed, ease and accuracy with which these form-function mappings are learned in language acquisition is considered to be determined by the strength with which forms and functions are associated in the linguistic input encountered by the language learner. That is, form-function learning is the joint outcome of statistical properties of the input, such as how frequently a particular form occurs, how frequently that form maps onto a particular function, and how reliably that form predicts a particular function (i.e., whether that form maps onto that function 100% of the time, or whether it also maps on to other functions, and whether the function also maps onto other forms). In addition, other psychological factors such as salience, blocking and overshadowing are also assumed to influence form-function learning (see Ellis, 2006b for an overview of these in SLA).

Crucially, the relationship between orthography and phonology is itself a type of form-function mapping, in which one or more graphic symbols serve as a set of cues to a word's pronunciation (i.e., the function, or outcome). If learning how orthography maps onto phonology in an L2 is a usage-based process, then the statistical properties of the relationship between orthography and phonology should predict both the accuracy of a word's pronunciation when reading aloud, and the development of this accuracy over time. In particular, for L2s using alphabetic scripts, pronunciation accuracy during phonological decoding should be predicted by i) the frequency with which letters and letter sequences occur (e.g., how frequently the letter <t> occurs), ii) the frequency with which these letters and letter sequences co-occur in words with one or more specific phones (e.g., the frequency with which <t> occurs in words whose phonological forms contain /t/), and iii) the reliability with which these letters and letter sequences in a word's written form predict one or more specific phones in its spoken form (e.g., how strongly a <t> in a word's written form predicts /t/ in its spoken form). Adopting the terminology of the Competition Model (Bates & MacWhinney, 1982; MacWhinney & Bates, 1989; MacWhinney, 1992), the notion of forms (here, letters and letter sequences) will be referred to as cues, and the statistical measures described above will be referred to respectively as cue availability (i), cue reliability (ii) and cue validity (iii). In addition to these measures predicting pronunciation accuracy, their effect on accuracy should be graded, showing a gradual increase in accuracy as cue availability, reliability and validity increase. Moreover, if learners adjust the association strengths of letters and sounds in their mental lexicons in order to reflect linguistic input, as assumed by usage-based accounts of language learning, then these measures should interact with L2 exposure to predict changes in pronunciation accuracy as exposure increases.

The aim of this study is therefore to investigate 1) whether orthographic input has graded effects on spoken language skills in early SLA and 2) how any such effects interact with L2 exposure. Specifically, this study will use a reading-aloud task to examine whether L2 pronunciation accuracy in phonological decoding is influenced by the availability of L1 and L2 orthographic cues, and the reliability and validity with which these cues correspond to the expected phones. The orthographic cues being used are one-, two- and three-letter sequences, referred to hereafter as unigraphs, bigraphs and trigraphs respectively, or n-graphs collectively. In addition, this study will investigate whether L1 and L2 cue availability, reliability and validity interact with L2 exposure to predict changes in pronunciation accuracy, and whether these measures interact cross-linguistically to jointly predict accuracy and its development over time.

Chapter 1 will outline the relevant research on orthographic effects in language processing and acquisition, giving special attention to the relationship between orthography and phonology. The research to be discussed demonstrates a clear interaction between orthography and phonology in the development of L2 spoken language skills, finding effects of L1 and L2 orthographic depth (Erdener & Burnham, 2005; Hamada & Koda, 2008, 2010), availability of orthographic information (Escudero, Hayes-Harb, & Mitterer, 2008; Showalter & Hayes-Harb, 2013; but see Simon, Chambless, & Alves, 2010; Showalter, 2012), consistency of L2 grapheme-phoneme correspondences (Bassetti, 2006, 2007; Barkley, 2010), and cross-linguistic consistency of grapheme-phoneme correspondences (Hayes-Harb et al., 2010). This research further finds results suggesting that statistical measures describing the relationship between orthography and phonology will predict the development of L2 spoken language skills. Additionally, these studies demonstrate influences of previous L1 and L2 experience on how orthography influences L2 acquisition and processing, indicating that both L2 exposure and L1 measures may interact with L2 statistical measures in predicting the development of L2 spoken language skills.

In Chapter 2, the methodologies both for this current study and for obtaining the orthographic measures under investigation are described. Chapter 3 then presents the results of this study, with separate discussions for unigraphs, bigraphs and trigraphs. Finally, Chapter 4 considers the results of this study in the context of the existing research discussed in Chapter 1, seeking to highlight the contributions made to the current literature and to draw implications for future research.

1. Background

This chapter will provide an overview of research that has investigated orthographic influences on spoken language skills in second language acquisition. It will begin with a brief discussion of writing systems, highlighting the various ways in which L1 and L2 writing systems may vary from one another. The following section will cover research that focuses on whether the availability of orthographic information to L2 learners facilitates or hinders L2 acquisition. In the next section, the focus will be specifically on how SLA is influenced by the relationship between orthography and phonology in the L1 and L2. Finally, the precise research questions to be addressed in this study are outlined and discussed in the context of the research presented below.

1.1 Writing systems

A writing system is "a set of visible or tactile signs used to represent units of language in a systematic way" (Coulmas, 1999: 560). Writing systems can differ from one another with respect to which linguistic units they represent, what script is employed, how transparent the relationship is between graphic symbols and the sounds of a language, and how consistently orthographic forms map onto spoken forms.

The broadest dimension along which writing systems may vary is whether they are meaning-based or sound-based. Meaning-based systems, such as Chinese Hanzi and Japanese Kanji, use graphic symbols to represent morphemes. These are also referred to as morphographic writing systems. Sound-based systems, on the other hand, represent either syllables (e.g., Tibetan) or phonemes (e.g., Greek). These latter systems that represent phonemes, also called phonographic systems, can be further subcategorized as alphabetic or consonantal. Alphabetic systems, like that of English, graphically represent all spoken phonemes, whereas consonantal systems such as Arabic primarily represent consonants and generally omit vowels. Writing systems also vary in which script they employ. For example, English and Russian both use alphabetic writing systems, but English uses the Roman alphabet whereas Russian uses the Cyrillic alphabet.

Another way in which writing systems can vary is in their orthographic depth, or transparency, which refers to the degree to which there is a one-to-one relationship between graphic symbols and sounds. Among alphabetic systems, for example, writing systems like those of Spanish, Italian and Finnish are considered orthographically shallow or transparent, as there is a high degree of one-to-one correspondence between letters and sounds. That is, one letter or letter sequence tends to represent only one sound, and one sound tends to be represented by only one letter or letter sequence. Writing systems like those of French and English, on the other hand, tend to have a one-to-many correspondence between letters and sounds, where a single letter or set of letters can represent multiple sounds or sound sequences, and a single sound or set of sounds can be represented by multiple letters or letter sequences. To illustrate, the letter <a> in English can represent /æ/ (e.g., *bat*), /e/ (e.g., *babe*), /ʌ/ (e.g., *among*) etc. Similarly, sounds like /o/ can be represented variably by <o> (e.g., *tote*), <oa> (e.g., *boat*), <ow> (e.g., *show*) etc. As such, writing systems like this are referred to as being orthographically deep, or opaque. This correspondence between letters and sounds is also often referred to as the grapheme-phoneme correspondence, where graphemes are defined as the written representations of phonemes.

Finally, in the literature investigating how orthography influences SLA, the term consistency has been used to refer to whether a word's orthographic form maps directly onto its spoken form in a one-to-one manner. For example, the word *screen* is considered consistent, as there is a direct one-to-one mapping between its graphemes and its phonemes (<s>-/s/, <c>-k, <r>-/ɹ/, <ee>-/i/, <n>-/n/). The word *computer*, on the other hand, is inconsistent, as the /j/ in the spoken form /kəmpjuɾɚ/ is not represented directly by a grapheme in its written form. The term consistency also sometimes refers to whether the relationship between the L2 orthography and sound system is consistent with the relationship between the L1 orthography and sound system (e.g., whether a letter common to the L1 and L2 corresponds to the same sound in each language as well). For a more detailed explanation of writing systems in the context of second language acquisition, see Cook & Bassetti (2005).

1.2 Orthographic influences in SLA

A long history of L2 research has stressed the important role that input plays in SLA (e.g., Krashen, 1982), which has been further underscored by findings of robust frequency effects (see Ellis, 2002). Bassetti (2008) highlights the fact that written input constitutes a large portion of the L2 input received in instructed SLA. It is perhaps unsurprising, then, that research has found SLA to be influenced by factors pertaining to orthography, such as the availability of orthographic information, orthographic depth, and orthographic consistency; effects of these factors have been reported in speech perception (e.g., Escudero et al., 2008; Escudero & Wanrooij, 2010; Showalter & Hayes-Harb, 2013), speech production (e.g., Young-Scholten, 1998; Young-Scholten, Akita, & Cross, 1999; Young-Scholten, 2002; Erdener & Burnham, 2005; Steele, 2005; Bassetti, 2007; Barkley, 2010; Rafat, 2011), vocabulary learning (e.g., Ellis & Beaton, 1993; Hamada & Koda, 2008; Hayes-Harb et al., 2010), spoken word recognition (e.g., Veivo & Järvikivi, 2013), and reading skills (e.g., Koda, 1990; Hamada & Koda, 2010). In the following sections, studies showing effects in SLA of the factors mentioned above will be elaborated on in order to demonstrate i) that orthography influences the development of L2 spoken language skills, ii) that statistical measures describing the relationship between orthography and phonology may predict the development of L2 spoken language skills, and iii) that previous L1 and L2 experience influences how orthography affects SLA, which suggests that L2 statistical measures describing the relationship between orthography and phonology may interact both with L2 exposure and with similar L1 measures.

1.2.1 Availability of orthographic information during SLA

Research investigating if and how L2 learners are able to use orthographic information to aid the development of L2 spoken language skills has thus far focused on the learning of novel contrasts. It has, moreover, found mixed results, with some studies showing beneficial effects of orthography (Escudero et al., 2008; Showalter & Hayes-Harb, 2013), and others showing no effects of orthography (Simon et al., 2010; Showalter, 2012). Despite these mixed results, the current evidence does suggest that L2 exposure may play a crucial role in if and how L2 learners use orthographic information to facilitate L2 spoken language development.

Showalter & Hayes-Harb (2013), for example, had native English speakers perform a word-learning task with Mandarin pseudowords, in which participants were either given orthographic tone marks, or no orthographic tone marks. In a picture association task designed to assess word-learning, the tone-mark group performed identically to the no-tone-mark group when the auditory form they heard matched the picture they saw, but they outperformed the no-tone-group when there was a mismatch between the picture and the auditory form. In a follow-up experiment where participants matched auditory forms with orthographic forms instead of pictures, the tone-mark group significantly outperformed the no-tone-mark group both when the forms were matched and when they were mismatched. These results suggest that L2 learners are able to use orthographic information to help learn novel acoustic contrasts. As the tone-mark groups in both experiments required more training time to meet criterion than the no-tone-mark groups, it is unclear, however, whether these results truly reflect differences in orthographic input, or whether they stem from the differences in amount of exposure during training.

In a word-learning experiment using eye-tracking, Escudero et al. (2008) had highly proficient Dutch-English bilinguals learn English pseudowords either with or without orthography in order to investigate whether orthography could explain the asymmetrical competition in lexical access found by Weber & Cutler

(2004) and Cutler, Weber, & Otake (2006). These studies found that L2 listeners experience asymmetric confusion when disambiguating between a target word and a competitor whose first syllables are differentiated by a phonemic contrast which is absent in the L1, but which involves a phone present in the L1 and a phone absent in the L1. To illustrate, Dutch, unlike English, does not have a phonemic contrast between /ɛ/ and /æ/. Moreover, it does not use the phone /æ/, but it does have /ɛ/. Accordingly, Dutch learners of English are asymmetrically confused by lexical competitors like *pencil* and *panda*[2] (Weber & Cutler, 2004), the first syllables of which are differentiated by /ɛ/ and /æ/.

To test their hypothesis that orthography influences learners' ability to lexically encode contrasts such as the type described above, Escudero et al. (2008) used the visual word paradigm. In perception experiments using this paradigm, participants are presented with auditory stimuli such as words or sentences while they look at a screen on which words, images or entire scenes are displayed. In Escudero et al.'s (2008) experimental trials, participants saw four images: a target image whose learned name began with a syllable containing either /ɛ/ or /æ/ (e.g., [tɛnzə]), a competitor image whose name began with a syllable differentiated from the target by the /ɛ/-/æ/ contrast (e.g., [tændɔk] if the target is [tɛnzə]), and two distractor images whose initial syllables differed from those of the target and competitor items by more than one phonemic contrast. While the images were displayed on screen, participants heard auditorily presented sentences instructing them to click on the target word. Eye-movements were recorded during each trial

[2] Example taken from Escudero et al. (2008: 347).

in order to examine the proportion of looks to each image during the 300-700ms time window following stimulus onset, which is the same time window used in Weber & Cutler (2004) and Cutler et al. (2006).

Consistent with their hypothesis, Escudero et al. (2008) found that participants who learned English pseudowords without orthography exhibited symmetric confusion for /ɛ/-/æ/ contrasts, as shown by equal looks to target and competitor images in the time window of interest, regardless of whether the auditory stimuli contained /ɛ/ or /æ/. Conversely, those given orthography experienced the same asymmetric confusion reported in previous studies during the 300-700ms time window, looking at target and competitor images when the auditory stimuli contained /æ/, but looking only at images whose name contained /ɛ/ when the auditory stimulus contained /ɛ/. These findings suggest that L2 listeners may use orthography to aid competition resolution during lexical access, i.e., orthography may help L2 listeners to select the appropriate word form when multiple candidate words are activated by ambiguous acoustic cues.

In contrast to the effects reported above, Simon et al. (2010) found no effects of orthography on the ability of L1 English learners of French to learn novel phonemic contrasts. This study differs from the above two in a number of critical ways, however. Rather than phonemic contrasts, Showalter & Hayes-Harb (2013) investigated novel tone contrasts. It is possible, therefore, that there may be differences in how L2 learners perceive, process and acquire tonal contrasts compared to phonemic contrasts. As these studies used different tasks, it furthermore possible that task dependence is responsible for the different results.

Note also that the task used by Simon et al. (2010) was offline, whereas Escudero et al. (2008) used eye-tracking, which provides a high temporal resolution, online measure of information processing as it occurs. Thus, it is possible that although the participants in Simon et al.'s (2010) study did not show orthographic effects offline, they may yet have shown orthographic effects in an online task, and vice versa for the participants in Escudero et al.'s (2008) study. Finally, Simon et al. (2010) used participants with no formal knowledge of the L2 used in the study, whereas Escudero et al. (2008) used highly proficient bilinguals, raising the possibility that the results of these studies may also depend on L2 proficiency and exposure.

Lastly, Showalter (2012) investigated whether learners are able to use entirely novel orthographic symbols to learn a novel contrast. In two experiments, L1 English speakers with no knowledge of Arabic learned the auditory forms of Arabic pseudowords and their "meanings" by associating auditorily presented word forms with pictures. The pseudowords consisted of a number of minimal pairs differentiated by the Arabic contrast between /k/ and /q/. Participants also received orthographic information consisting either of the auditory word form spelled in Arabic script, or of a meaningless Arabic letter sequence in the control condition (<طططط>). Learning was tested in experiment one by having participants indicate whether a picture and an auditory form matched, and in experiment two by having participants indicate whether an orthographic form and auditory form matched. Consistent with Simon et al. (2010) no significant effects of orthographic condition were found in either experiment, adding further evidence to suggest that beginning learners are unable to, or have difficulty using orthographic information to facilitate the learning of novel phonemic contrasts. Alternatively, as suggested by Showalter (2012), these results may reflect difficulty with the /k/-/q/ contrast which is compounded by also learning a novel orthographic script. Longitudinal data may be helpful in determining if and when L2 learners begin using orthography to learn and encode novel phonemic contrasts.

In short, the evidence is currently mixed with regards to whether L2 learners use orthographic information to aid in the development of L2 spoken language skills. While Simon et al. (2010) and Showalter (2012) find that beginning L2 learners do not use orthography to help learn novel phonemic contrasts, the results of Showalter & Hayes-Harb (2013), showing beneficial effects of orthography for beginning L2 learners with novel tonal contrasts, suggest that the ability to use orthographic information may depend on what is being learned. Novel L2 tonal contrasts may simply be easier to learn than novel L2 phonemic contrasts, in which case L2 learners might begin exhibiting positive effects of orthography for phonemic contrasts after greater L2 exposure. Indeed, the effects of orthography for highly proficient bilinguals found by Escudero et al. (2008) indicate that L2 proficiency may play a role in if and how orthographic information is used by L2 learners. As such, a developmental approach using longitudinal or cross-sectional data to investigate L2 use of orthography over time may prove beneficial in determining if, when and how beginning L2 learners use orthography to aid the development of spoken language skills.

1.2.2 Orthography-phonology relationships in SLA

A wealth of L1 research has demonstrated that orthography and its relationship to a language's sound-system influence language processing. Effects of this relationship have been found in speech perception tasks involving both metalinguistic awareness (e.g., Seidenberg & Tanenhaus, 1979; Dijkstra, Roelofs & Fieuws, 1995; Hallé et al., 2000) and online processing (e.g., Ziegler & Ferrand, 1998; Ventura et al., 2004; Ziegler et al., 2004; Pattamadilok et al., 2007; Taft et al., 2008), as well as in speech production (e.g., Tanenhaus et al., 1980; Lupker, 1982; Damion & Bowers, 2003; Brewer, 2008; Rastle et al., 2011) and visual word recognition (e.g., Forster & Chambers, 1973; Baron & Strawson, 1976; Frederiksen & Kroll, 1976; Stanovich & Bauer, 1978; Ferrand & Grainger, 1992, 1993, 1994, 2003; Grainger et al., 2006). The relationship between orthography and phonology also affects the rate and accuracy of reading development for children in their L1: the clearer the relationship is–i.e., the more transparent the orthography–the faster and more accurate children are in learning to read (e.g., Cossu, 1999; Harris & Giannouli, 1999; Durgonoğlu, 2006; Goswami, 2006; Landerl & Thaler, 2006; Lyytinen et al., 2006; Porpodas, 2006; Seymour, 2006). A large body of research has also demonstrated orthographic effects in highly proficient bilinguals (e.g., Bijeljac-Babic et al., 1997; Dijkstra et al., 1998; Van Heuven et al., 1998; Dijkstra et al., 1999, 2000; Lemhöfer & Dijkstra, 2004). Despite the clear effects that the orthography-phonology relationship has in L1 and bilingual processing, considerably less research has investigated the effects of this relationship in early second language acquisition.

Nonetheless, the existing research does find effects of orthography-phonology relationships during early SLA, indicating that statistical measures describing this relationship may prove useful in predicting L2 spoken language skill development.

Erdener & Burnham (2005) had monolingual speakers of Turkish and English perform a pseudoword repetition task using Spanish and Irish pseudowords, varying whether participants also received orthographic information with the auditory stimuli. Crucially, Turkish and Spanish both use shallow orthographies, whereas English and Irish employ deep orthographies, thus enabling the researchers to explore the effects of both L1 and L2 orthographic depth. Results showed that productions were overall more accurate when orthography was given than when it was absent, and that this accuracy is influenced by orthographic depth. When orthography was given, the Turkish participants made significantly more production errors for the Irish stimuli compared to the Spanish stimuli. The L1 English participants, however, performed comparably with both languages. Additionally, Turkish participants outperformed English participants when naming Spanish pseudowords, but were outperformed by the English participant on the Irish stimuli. These results suggest that L1 orthographic depth affects the ease with which L2 orthography-phonology relationships are learned.

The effects of orthographic depth have also been explored using L1s with different types of writing systems (i.e., sound-based or meaning-based). Using a pseudoword naming task, Hamada & Koda (2008) explored the effect of L1 orthographic experience (i.e., type of writing system) on the ability of L2 English learners with Korean and Chinese backgrounds to phonologically decode written regular and irregular English pseudowords to determine their spoken forms. They found that the L1 Korean participants, whose L1 orthography is alphabetic, like that of English, were significantly faster and more accurate in the naming task than the Chinese learners, whose L1 writing system is morphographic and, as such, orthographically deeper than that of Korean. In a separate, word learning experiment, L1 Korean participants also showed better retention of word meanings than L1 Chinese participants, suggesting that the similarity of L1 and L2 orthographic systems may play an important role in L2 word learning. Additionally, participants showed better performance in both experiments for regular words compared to irregular words, indicating that they are sensitive to the statistical patterns of English orthography-phonology relationships. In a later study, Hamada & Koda (2010) partially replicated the results of their 2008 study with L1 Korean, Turkish, Chinese and Japanese learners of English, finding that participants with an alphabetic L1 writing system named English real and pseudowords faster and more accurately than those with a morphographic L1. These results add further evidence to suggest that properties of the orthography-phonology relationship in the L1 (e.g., orthographic depth) will affect how this relationship is learned in an L2.

Some studies have also investigated the effects of orthography on the mental representations of L2 phonology. Bassetti (2006, 2007) for example, reports evidence that orthography influences how L2 learners of Chinese represent Chinese syllables in their mental lexicons. Her results find that syllables with a consistent one-to-one correspondence between letters and phonemes (e.g., /uei/ spelled <wei>) are segmented more accurately than syllables in which phonemes are not all represented in a one-to-one manner in the orthography (e.g., /uei/ spelled <ui>) (2006). Because segmentation was prompted using hanzi script, which does not provide phonological information, rather than pinyin, which does provide phonological information, Bassetti argues that her results reflect how the participants in her study mentally represent the phonological forms of the test words. She also finds that these representations influence pronunciation, with more target-like pronunciations being found for syllables in which all phones are consistently represented in the orthography (2007).

Hayes-Harb et al. (2010) explored whether L1 letter-sound correspondence information influences the learning of phonological forms for novel L2 words. Native English speakers learned auditorily presented pseudowords accompanied by pictures to represent their meanings. Some participants were also given orthographic forms for the pseudowords. Sometimes these forms used letter-sound correspondences consistent with English letter-sound correspondences (e.g., [kɑməd] spelled <kamad>), and other times they used correspondences inconsistent with English letter-sound correspondences, containing either an extra, "silent" letter (e.g., [kɑməd] spelled <kamand>) or an incorrect letter (e.g., [fɑʃə] spelled <faza> rather than <fasha>)[3]. To test learning of the phonological forms, participants were presented with a picture and asked

[3] Examples taken from Hayes-Harb et al. (2010: 372).

whether an auditorily presented form was the correct word for that picture. When the auditory form heard at test did not match the auditory form learned for a picture, participants who were given inconsistent orthographic forms performed worse than those given orthographic forms consistent with English orthography. This suggests that L1 orthographic information may interfere with learning L2 phonological forms when letter-sound correspondences differ in the L1 and L2.

Finally, there has been some investigation specifically of the acquisition of L2 grapheme-phoneme correspondences. Barkley (2010) examined the oral productions of L2 and L3 learners of Portuguese when naming Portuguese real and pseudowords. The analysis was restricted to the correspondences between <z> and /z/, which is a one-to-one mapping, and <s> to /z/, which only occurs intervocalically and before voiced consonants, and is therefore not a one-to-one mapping. Results showed significantly more erroneous productions for the <s>-/z/ correspondence than for the one-to-one <z>-/z/ correspondence, suggesting that the L2 acquisition of orthography-phonology information may be sensitive to how reliably different graphemes predicts particular phonemes. Investigating how L1 transfer is influenced by the cross-linguistic consistency of grapheme-phoneme correspondences, Rafat (2011) found differential rates of L1 transfer for different grapheme-phoneme correspondences in the productions of L1 English learners of Spanish in a word-learning task. These results provide further evidence to suggest that L2 learning is sensitive to the statistical properties describing how orthography corresponds to phonology. Indeed, Rafat proposes that patterns of transfer may be explained by the frequencies of different L1 grapheme-phoneme correspondences.

1.2.3 Summary of orthographic influences in SLA

The research discussed in this section has demonstrated that orthography influences the development of L2 spoken language skills. Orthographic depth has been found to affect the learning of L2 orthography-phonology relationships and spoken word accuracy (Erdener & Burnham, 2005; Hamada & Koda, 2008, 2010). Spoken word accuracy has also been shown to be influenced by the consistency of L2 orthography-phonology relationships (Bassetti, 2007). The evidence discussed further suggests that L2 spoken language skills may be predicted by statistical measures describing the relationship between orthography and phonology. This is indicated by results finding differential rates of L1 transfer and production errors for different grapheme-phoneme correspondence (Barkley, 2010; Rafat 2011), and by results showing sensitivity of L2 learners to consistency (Bassetti, 2006, 2007) and regularity (Hamada & Koda, 2008). It is possible that these differential effects of different correspondences, and of consistency and regularity may be explained by differences in frequency-based and probabilistic measures describing the grapheme-phoneme correspondences in the experimental stimuli. If so, graded effects of such statistics on spoken language skills may emerge in L2 populations.

The research discussed in this section has also shown that previous linguistic experience influences how orthography affects SLA. The ease with which L2 orthography-phonology relationships are learned appears in part to be mediated by L1 orthographic depth (Erdener & Burnham, 2005; Hamada & Koda, 2008, 2010). The learning of L2 orthography-phonology relationships is also affected by how consistent those relationships are with L1 orthography-phonology relationships (Hayes-Harb et al., 2010). Finally, the differing results about whether L2 learners use orthographic information to learn novel contrasts (Escudero et al., 2008; Simon et al., 2010; Showalter, 2012) suggest that L2 proficiency may modulate if and how orthography influences SLA.

In short, the research discussed in this section demonstrates that orthography influences the development of spoken language skills in SLA, that statistical measures describing the relationship between orthography and phonology may predict the development of L2 spoken language skills, and that L2 exposure may affect how such measures influence the development of L2 spoken language skills over time.

1.3 Current study

To address the questions arising from the previous research, the goal of this current study is to investigate whether orthographic input has a graded influence on L2 spoken language skills in early SLA, and whether any effects of orthographic input interact with L2 exposure to predict the development of L2 spoken language skills over time. The research discussed in section 1.2 clearly shows that the relationship between orthography and phonology impacts the development of L2 spoken language skills. However, it has thus far approached this relationship using categorical descriptions, such as consistent/inconsistent or orthographically deep/shallow, rather than using gradient, statistical measures.

Consequently, the possibility of graded orthographic effects in early SLA has been left unexplored. Moreover, the inattention to possible influences of length of exposure has resulted in a lack of knowledge about if and how orthographic effects in L2 populations emerge and develop during SLA as L2 exposure increases. Nonetheless, this research does find results suggesting that statistical measures of the orthography-phonology relationship may predict L2 spoken language skills, and that L2 exposure will play a role in if and how orthographic effects manifest during SLA.

In addition, usage-based approaches to language learning provide an empirically grounded, theoretical reason to expect graded effects of orthography in L2 populations, and an interaction of such effects with L2 exposure. These approaches assume that language acquisition is a process of associative learning, whereby learners gradually adjust their linguistic systems to reflect the frequency and reliability with which linguistic items occur and co-occur in the linguistic input. Accordingly, usage-based models predict that high frequency, high reliability items will tend to be learned earlier in acquisition, and that they will be used more accurately. Indeed, recent research has provided empirical support for these claims in second language acquisition, showing that L2 learners are highly sensitive to the statistical properties of the input (see e.g., Ellis, 2002, 2006a, 2006b; Ellis & Ferreira-Junior, 2009a, 2009b; Ellis & O'Donnell, 2012; *inter alia*).

Given the input structure of the L2 classroom, where written input is often encountered simultaneously with a corresponding spoken form, usage-based theory will predict that instructed L2 learning is sensitive to the statistical properties describing the relationship between the written and spoken input, i.e., how orthography corresponds to phonology. If so, these statistics should predict how accurately L2 learners are able to phonologically decode written forms in order to determine their spoken forms, and how accuracy in this task changes with increased L2 exposure. Orthographic word forms exhibiting high cue availability, reliability and validity in predicting a word's corresponding phonological form should be decoded more accurately, and show higher accuracy earlier in acquisition than orthographic forms with low cue availability, reliability and validity. In other words, the frequency with which a word's graphemes occur, the frequency with which these graphemes co-occur with the word's phones, and the reliability with which these graphemes predict the word's phones should influence L2 phonological decoding and its change over time. It is, moreover, well established that properties of a learner's L1 will influence the learning of an L2, sometimes facilitating this process and sometimes hindering it (see e.g., Kellerman & Sharwood Smith, 1986; Odlin, 1989; Ellis, 2006b). As such, it is furthermore possible that L2 cue availability, reliability and validity will interact with these same measures in a learner's L1 to facilitate or inhibit phonological decoding.

Thus, in order to determine if graded influences of orthographic input are evident in SLA, this study will have L1 English learners of German read German sentences aloud in a phonological decoding task. Each sentence will contain one key test word, selected to reflect a variety of letter and sound combinations.

Importantly, by having participants read aloud, this study is able to implicitly probe the relationship between orthography and phonology by providing participants with the orthographic forms of words and requiring them to determine the phonological forms using this orthographic information. Additionally, in order to control for potentially confounding lexical factors, this study uses a combination of real and pseudowords, and statistically controls for how frequent the real words are, and how familiar the participants are with these real words. If orthographic input has a graded influence on spoken language skills during SLA, we should observe graded effects in the phonological decoding task, emerging as continuous relationships between the availability, reliability and validity of orthographic cues in the test words, and the accuracy with which participants pronounce these words.

The precise orthographic cues under investigation in this study will be one-, two-, and three-letter sequences, henceforth referred to as unigraphs, bigraphs and trigraphs, respectively. In keeping with a usage-based approach, the statistical measures to be used in this study are the following:

1) Cue availability: Measured as relative n-graph frequency, i.e., the frequency per million with which the unigraphs, bigraphs and trigraphs in the experimental stimuli occur in English and German. For the bigraphs in the German word *du*, for example, this is the frequency per million with which '#d', 'du', and 'u#' occur in both English and in German.
2) Cue reliability: Measured as the relative co-occurrence frequency of an n-graph and a phone, i.e., the frequency per million, in English and German,

with which the unigraphs, bigraphs and trigraphs in the experimental stimuli occur in words containing the phones in the prescribed pronunciation of each test word. Using the bigraphs of *du*, again, this is the frequency per million with which '#d', 'du' and 'u#' occur in English and German words containing the phonemes /d/ or /u/. In other words, cue reliability is measured as the relative frequency with which the n-graphs in the test words occur as potential cues to the phones in the expected pronunciations of those words.

3) Cue validity: Measured using delta P, which is a dependency statistic calculating the probability of an outcome occurring in the presence of a cue, minus the probability of that outcome occurring in the absence of that cue (see Allan, 1980; Ellis, 2006a). Delta P is given by the following formula, in which 'P' means probability, 'O' means outcome, and 'C' means cue:

$$\Delta P = P(O|C) - P(O|{-C})$$

As this study investigates phonological decoding, the outcomes here are the phones in a word, and the cues are the n-graphs in that same word. For *du*, the validity of 'u#' as a cue to the outcome /u/ is the probability that the phone /u/ occurs in the pronunciation of a word that contains the bigraph 'u#', minus the probability that a word containing this bigraph does not have the phone /u/ in its pronunciation:

P(/u/|'u#') – P(/u/|–'u#')

> The probability of /u/ occurring given ‘u#’ is the per million frequency of words whose orthographic forms contain ‘u#’ and whose phonological forms contain /u/, minus the total per million frequency with which the bigraph ‘u#’ occurs. The probability of /u/ occurring in the absence of ‘u#’ is the per million frequency of words whose phonological forms contain /u/ but whose orthographic forms do not contain ‘u#’, minus the total per million frequency of all words not containing ‘u#’.

L2 exposure in this study is defined as the length of exposure to L2 orthography and phonology in an instructed setting. It is measured using the total months of instruction received by participants at each time point during data collection. The dependent variable, phonological decoding accuracy, will refer to the accentedness of learners’ productions. Specifically, it is the degree of correspondence between the phonological segments produced by participants for each test word and the phonological segments expected in each word’s prescribed pronunciation (i.e., to what extent participants produce the expected phones in the expected positions in each word). Accuracy is thus measured as the Levenshtein edit distance between the string of phones produced by participants for each test word and the string of phones in each test word’s prescribed pronunciation. The specific research questions to be addressed in this study are:

1) Does orthographic input exhibit graded influences on L2 spoken language skills? In other words, is there a continuous relationship between the statistical measures described above and how accurately the test words are pronounced by participants?

2) Is L2 pronunciation accuracy in reading aloud influenced by the availability of L1 and L2 orthographic cues, and the reliability and validity with which these cues predict the expected phones?
3) Do L1 and L2 cue availability, reliability and validity interact with L2 exposure to predict changes in pronunciation accuracy over time?
4) Do cue availability, reliability and validity interact cross-linguistically to jointly predict accuracy and its development over time? That is, do L1 measures of cue availability, reliability and validity affect how these same measures in an L2 influence pronunciation accuracy and its change over time?

2. Methodology

This study examines the oral productions in German of native English speaking students in a beginner-level German language course, and of a monolingual English control group with no formal knowledge of German. A self-paced reading-aloud task was used to elicit target word productions in carrier sentences. Data were collected from the test group twice during an academic semester: once after approximately one month of instruction, and again after approximately three months of instruction. Linear mixed-effects regression modeling (see e.g., Baayen, 2008) was then used to assess the effects of various orthographic predictors on the accuracy of target word productions, and the change in production accuracy as L2 exposure increased.

2.1 Materials

2.1.1 Test Words

The test words consisted of a total of 50 German real words and 50 German pseudowords. Real words were selected pseudorandomly from a subset of the German Subtitle Lexicon (Brysbaert et al., 2011) to have a minimum per million frequency of five and to represent a variety of letter-to-sound combinations. For selection of the pseudowords, a set of candidates was generated using Wuggy (Keuleers & Brysbaert, 2010). To generate the candidates, the 2000 most frequent words in the German Subtitle Lexicon were loaded into Wuggy, which was set to generate up to ten candidate pseudowords per each real word, matching two-thirds of subsyllabic segments with the input words, and matching these for length of the subsyllabic segments as well. From the list of candidates, pseudowords containing the letter-to-sound correspondences of primary interest in the real words were selected pseudorandomly to avoid any candidates that looked or sounded like real words in German or English. Candidate pseudowords were then loaded back into Wuggy and checked to confirm that neither their orthographic nor phonological forms[4] exist in German.

On subsequent examination of the materials after the data were collected, it was discovered that three of the pseudowords were in fact real words. These items were excluded from analysis.

2.1.2 Sentences

In order to be as authentic as possible, the experimental sentences were created using sentences found in the Mannheim Corpora (Institut für Deutsche Sprache [IDS], 2013). Sentences for the real words were found by searching the corpora using the COSMAS II portal (IDS, 1991-2010) for sentences containing the real words to be tested. Criteria for the sentences were that they be declarative, shorter than 20 words in length, and that the target word occur no earlier than in the third position. From any sentences matching these criteria, the sentence used was selected using the experimenter's judgment.

For pseudoword sentences, words containing the same morphological or orthographic ending–which frequently conveys information about part of speech, number, gender etc. in German–were randomly selected from the German Subtitle Lexicon and then searched for in the Mannheim Corpora. Selection of sentences

[4] Section 2.4 describes how the phonological forms were determined.

followed the same criteria outlined above. Pseudowords were then exchanged with the real words used to find the sentences. Finally, slight modifications were made to sentences where necessary in order to improve sentence flow or to reflect conventional writing style (e.g., contractions were written out in full).

For the test group, one set of counterbalanced lists was created by randomly assigning an order to the test sentences for one list, and reversing this order to yield a second list. For the control group, the 50 sentences containing real words were randomly ordered into a single list. In addition to the German sentences, the control group also read 50 French sentences in a task not reported in this thesis. Whether control participants read the German sentences before or after the French sentences was counterbalanced across participants.

Re-examination of the sentences after data collection revealed spelling mistakes in three of the sentences. As two of these spelling mistakes did not occur in words immediately adjacent to a test word, and resulted in either a real word or a pseudoword, data from the test words in these sentences were not excluded. One spelling mistake, however, occurred in a test word. As such, data from this test word were excluded. All experimental stimuli can be found in Appendix A.

2.2 Participants

Participants were 10 students (six male; *Mean age* = 21) registered in an introductory German course (test group), and 8 female students (*Mean age* = 20) registered in an introductory linguistics course (control group). Both courses took place at the University of Alberta. All participants spoke English as their first and dominant language. Control participants had received no formal education in

German and rated their abilities to speak, understand and read German very low on a ten-point scale (Speaking: $M = 0.625$, $SD = 0.74$; Understanding: $M = 0.875$, $SD = 0.64$; Reading: $M = 0.5$, $SD = 0.53$). One test subject reported having spent two months in a German-speaking country two years prior to data collection. Self-reported abilities to speak, understand and read German given by this participant were all within the range of values given by the other participants. Moreover, this participant did not report having begun learning to read German during his time in a German-speaking country. No other test subjects reported any prior exposure to German beyond that experienced in daily living.

2.3 Procedure

2.3.1 Test Group

The test group participated twice in a single academic semester: once after one month of instruction in German, and again after 3 months of instruction. The procedure for each session was identical, with the exception that participants were debriefed after the second session. Assignment of experimental lists was counterbalanced across participants and sessions, such that each list was seen by half of the participants for the first session, and by the other half of participants for the second session. Each list was thus seen once by each participant. Participants in this group were financially compensated for their participation.

At the beginning of a session, participants were asked to fill out the Language Experience and Proficiency Questionnaire (Marian, Blumenfeld, & Kaushanskaya, 2007). Once the questionnaire was completed, participants were seated in a quiet room and instructed on the procedure for the reading-aloud task. Sentences were presented on a computer screen one at a time in a central position using a Microsoft PowerPoint Presentation (2011). Participants were instructed to read each sentence aloud in a natural manner. Once finished reading a sentence, participants used either the right or down arrow key to proceed to the next sentence. Utterances were recorded on an external microphone at 32 bit with a 44,100 Hz sampling rate using Audacity (2012).

After the reading-aloud task, participants filled out a questionnaire in which they were asked to rate the test words for familiarity and visual English-likeness on a seven-point Likert scale. The scale and description for the familiarity ratings were created following Nusbaum, Pisoni and Davis (1984); only the real words were rated for familiarity. For English-likeness, participants were asked to rate both the real and pseudowords for how much they look like possible English words. The questionnaire can be found in Appendix B.

2.3.2 Control Group

The control group participated just once during a single academic semester. The number of participants seeing each list was counterbalanced, such that four participants saw the German sentences first, and four participants saw them second. Participants in this group did not fill out the LEAP-Q, but rather began their session with the reading-aloud task. Procedure for this task was identical to the procedure used for the test group. Following this task, control participants rated the test words for their English-likeness using the same scale and descriptions as used for the test group. Finally, participants were asked an abbreviated series of questions taken from the LEAP-Q to ensure that they were indeed monolingual native English speakers with no formal knowledge of German. Participants in this group received course credit for their participation.

2.4 Analysis

Sound files were loaded into Praat (Boersma & Weenink, 2013) and test words were phonetically transcribed onto a single tier using the TextGrid editor. Phones were determined largely by ear with the assistance of articulatory cues in the waveform and spectrogram. Phones that could not reasonably be determined were marked with a question mark and not included in subsequent analysis. Partial and whole-word corrections were indicated in the transcription and marked as corrections. Data from the TextGrid tier were then extracted using Lennes' (2009) script for saving conversation tiers as text files.

Because this study is primarily concerned with the implicit knowledge that L2 learners of German possess about how the orthographic form of a word corresponds to its spoken form in German, it was decided to remove corrections made by the test group from subsequent analyses. Where corrections were made, only the portion of the utterance up to where the correction begins is maintained. In order to be as conservative as possible when evaluating learning, however, any whole-word corrections made by the control group were maintained in full, and the portion of the utterance prior to the correction was removed. Thus, for each participant, there remained one production of each test word per session.

Finally, measures of pronunciation accuracy were obtained for each test word production by using the *stringdist* package (Van der Loo, 2013) in R to determine the Levenshtein edit distance of each individual production from its prescribed pronunciation. For the real words, prescribed pronunciations were taken from Duden (2005, 2010). As this was not possible for the pseudowords, Eisenberg's (2006) description of German grapheme-phoneme correspondence rules was used to determine pronunciation[5]. Edit distance scores were then divided by each word's prescribed phonemic length in order to normalize accuracy measures. To facilitate interpretation of the results and figures, the inverse normalized edit distance score is used so that larger scores correspond to higher accuracy.

2.5 Orthographic Predictors

The orthographic predictors of interest in this study are unigraph, bigraph and trigraph frequencies (i.e., cue availability), the frequencies with which these n-graphs occur as potential cues to the phones in a test word (i.e., cue reliability), and the contingency, measured as delta P (see Allan, 1980; Ellis, 2006a), between the target phones and n-graphs in the test words (i.e., cue validity)–that is, how reliably the n-graphs in a word predict the phones in that word's phonological form. While n-graph frequencies were readily available, measures of n-graph-phone frequency and contingency were not. In order to obtain these measures, it was necessary to construct datasets for English and German with which a word's n-graphs could be matched against its phones. The process of creating these

[5] The decision to use grapheme-phoneme correspondence rules here was a pragmatic decision, as native German speakers were not available in sufficient quantity for a norming study. Moreover, the relationship between orthography and phonology in German is relatively unambiguous, making the use of such rules quite reliable.

datasets is described below. To maintain consistency across measures, n-graph frequencies were also obtained from the datasets described below, despite their availability elsewhere. Once measures were obtained, the sums of these measures for each test word were calculated independently for unigraphs, bigraphs and trigraphs. To normalize measures, the sum n-graph frequency of each word was divided by its orthographic length (in letters), and the sum n-graph-phone co-occurrence frequency and sum contingency of each word were divided by its phonemic length (counting diphthongs and affricates as individual phones).

2.5.1 English

In order to obtain the statistical measures described above for English, all words in the US-English Subtitle Lexicon (Brysbaert & New, 2009) with a minimum per million frequency of one were obtained. The pronunciations for these words were sought using the English Lexicon Project (Balota et al., 2007)[6]. Any words not returning a pronunciation from the ELP were removed, resulting primarily in the removal of proper names, contractions, abbreviations and acronyms, and leaving a total of 16,387 words in their orthographic and phonological forms. Modifications were then made to some of the pronunciations in order to better reflect Western Canadian English (e.g., the phones /ɑ/ and /ɔ/ were changed to /ɒ/ where appropriate). Pronunciations were then converted to a modified version of X-SAMPA to make the data readable in R. Finally, any upper case letters in the orthographic forms were converted to lower case.

[6] While the ELP could have been used to generate the words in the target frequency band as well, it was decided instead to use the US-English Subtitle Lexicon in order to make English and German measures as comparable as possible by also using the German Subtitle Lexicon for German measures.

2.5.2 German

For German, all words with a minimum per million frequency of one were taken from the German Subtitle Lexicon (Brysbaert et al., 2011). The CELEX database (Baayen, Piepenbrock, & van Rijn, 1993), accessed via WebCelex (Max Planck Institute for Psycholinguistics [MPI], 2001) was then used to obtain pronunciations for these words, albeit with limited success. Of the 27,364 words matching the frequency criterion, CELEX failed to return results for 11,901, many of which were high frequency words occurring more than 100 times per million. Additionally, numerous systematic errors were noted in the results. To minimize data loss from high frequency words, the remaining words for which no pronunciation had been obtained were run through Wuggy (Keuleers & Brysbaert, 2010) to verify their lexical status. This removed 8923 words, many of which were foreign words and proper names whose frequencies likely reflected the sampling methods of the SUBTLEX corpora rather than their frequencies in daily usage in Germany. After removing contractions, abbreviated titles (e.g., *Mr.*) and misspelled instances of various words, 18,429 words remained.

To obtain pronunciations for the remaining words that CELEX did not return pronunciations for, and to verify the accuracy of the CELEX pronunciations and correct systematic errors, all remaining words were run through Brondsted's Automatic Phonemic Transcriber (2008). To first evaluate the accuracy of this transcriber, its output was checked against the pronunciations given by Duden's *Das Aussprachewörterbuch* (2005) for the 2000 most frequent words. This yielded an accuracy rate of 96.05%, and revealed a small number of systematic errors, which could be corrected for. After correcting for the systematic errors noted in the output of both the transcriber and of CELEX, the final output of each was checked against one another, yielding a disagreement rate of 13.32 %. Each case of disagreement was investigated by hand to correct for idiosyncratic errors such as missing or incorrect phones, yielding a final disagreement rate of 12.34%. The remaining instances of disagreement were largely superficial differences in pronunciation, such as whether a phone was given as a vowel or a glide (e.g., CELEX gives the pronunciation /aktsion/ for *Aktion*, whereas the automatic transcriber gives /aktsjon/). Because CELEX was unable to provide pronunciations for all of the words of interest, and because of the high rate of agreement between the automatic transcriber and Duden (2005), the pronunciations given by the transcriber were used exclusively in the final dataset. Finally, upper case letters in the orthographic forms were converted to lower case in order to simplify comparisons with English measures.

3. Results

3.1 Pronunciation accuracy over time

Mean accuracy scores are given in Table 1 for 0, 1 and 3 months of instruction, and visualized in Figure 1. Pairwise *t*-tests were run to determine if the differences in accuracy scores at each time point are significant. Results show that this is indeed the case. Control participants with no formal instruction in German perform significantly worse than the test group after one month of instruction ($t(719.4) = -13.38$, $p < 0.001$) and after three months of instruction ($t(686.3) = -17.92$, $p < 0.001$). A paired *t*-test also finds that the test group performs significantly better at three months than at one month ($t(959) = 8.02$, $p < 0.001$). Months of instruction is thus taken to be an adequate measure of exposure.

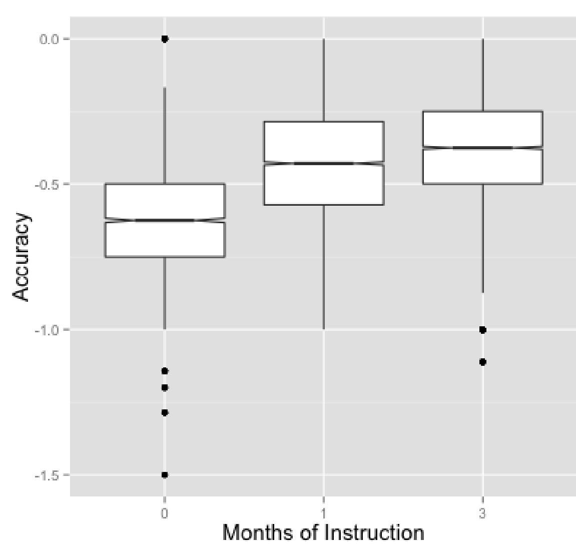

Figure 1: Pronunciation accuracy by months of instruction

Months of Instruction	*Mean Accuracy*
Zero	-0.603
One	-0.438
Three	-0.385

Table 1: Mean accuracy score by months of instruction

3.2 Orthographic predictors

In order to evaluate if and how the orthographic predictors under investigation influence the accuracy of pronunciation during phonological decoding for L2 learners of German, separate linear mixed-effects models were fitted for unigraphs, bigraphs and trigraphs using the *lme4* package (Bates, Maechler, Bolker, & Walker, 2013) in R. Before fitting the models, frequency measures were logarithmically transformed and all numeric predictors then centered. The dependent variable in all models was accuracy, and random effects were subject, word and list. Models were backward fitted by first including all possible predictors and interactions of interest, which are listed in Tables 2 and 3, respectively, and then removing non-significant predictors in a step-wise manner starting with the highest order interactions. Each subsequent model with a predictor removed was compared to the previous model containing that predictor by using the function *anova* in R. This process was continued until no predictor could be removed without making the model worse by at least a marginally significant degree.

To test for collinearity in the models, variance inflation factors were calculated using the *rms* package (Harrell, 2013). Predictors with variance inflation factors greater than 4 were residualized from one or more highly correlated variables as needed until all variance inflation factors were less than 4. In cases where previously significant predictors became non-significant after the reduction of collinearity, removal of predictors continued in the same manner outlined above. In addition to models for unigraphs, bigraphs and trigraphs, a null model was also fitted which contained only the significant non-orthographic predictors. Results of each model are presented in the following sections.

Main Effects	*Abbreviation*
Lexicality (Word or Pseudoword)	Lexicality
Orthographic Length	OrthLength
Phonemic Length of Prescribed Pronunciation	PhonLength
Trial	Trial
Familiarity Rating	Familiarity
Visual English-Likeness Rating	ENLik
Months of Instruction	MoInstr
Word Frequency	WFreq
Sentence Length	SLen
Sentence Position	SPos
Part of Speech	POS
Mean N-Graph Frequency (German)	MGraphFreq_DE
Mean N-Graph Frequency (English)	MGraphFreq_EN
Normalized N-Graph-Phone Co-occurrence Frequency (German)	GPCoFreq_DE
Normalized N-Graph-Phone Co-occurrence Frequency (English)	GPCoFreq_EN
Normalized N-Graph-Phone Contingency (German)	GPCont_DE
Normalized N-Graph-Phone Contingency (English)	GPCont_EN

Table 2: Main fixed effects included in the initial models prior to fitting, with cue availability represented by Mean N-Graph Frequency, cue reliability represented by Normalized N-Graph-Phone Co-occurrence Frequency, and cue validity represented by Normalized N-Graph-Phone Contingency

Interactions
MGraphFreq_DE * MoInstr
MGraphFreq_EN * MoInstr
GPCoFreq_DE * MoInstr
GPCoFreq_EN * MoInstr
GPCont_DE * MoInstr
GPCont_EN * MoInstr
MGraphFreq_DE * MGraphFreq_EN
GPCoFreq_DE * GPCoFreq_EN
GPCont_DE * GPCont_EN
MGraphFreq_DE * MGraphFreq_EN * MoInstr
GPCoFreq_DE * GPCoFreq_EN * MoInstr
GPCont_DE * GPCont_EN * MoInstr

Table 3: Interactions included in the initial models prior to fitting

3.3 Unigraph Model

Results for the unigraph model are given in Table 4. To reduce collinearity in this model, the unigraph-phone co-occurrence frequency (cue reliability) for German was residualized from the German unigraph frequency (cue availability). This value is given in the model as *rGPCoFreq_DE*. A comparison to the null model containing only random effects and control predictors shows that the unigraph model performs significantly better, accounting for 0.6% more variance than the null model (see Table 5).

Results show significant beneficial main effects for the residualized unigraph-phone co-occurrence frequency (cue reliability) in German, and for the English unigraph-phone contingency (cue validity). A calculation of the standardized regression coefficients furthermore indicates that these predictors have the strongest effect sizes on accuracy (Table 6).

Predictor	Coef	SE	t
Intercept	-0.011	(0.022)	-0.5
PhonLength	-0.021	(0.008)	-2.6
Familiarity	0.007	(0.003)	2.8
MoInstr	0.022	(0.004)	5.3
rGPCoFreq_DE	**0.133**	**(0.058)**	**2.3**
GPCont_DE	-0.039	(0.051)	-0.8
GPCont_EN	**0.227**	**(0.055)**	**4.1**
MGraphFreq_DE	-0.068	(0.05)	-1.4
GPCont_DE * MoInstr	0.023	(0.013)	1.7
GPCont_EN * MoInstr	**-0.064**	**(0.016)**	**-4.1**
rGPCoFreq_DE * GPCont_DE	**-0.337**	**(0.146)**	**-2.3**
MGraphFreq_DE * MoInstr	**0.041**	**(0.016)**	**2.8**
rGPCoFreq_DE * GPCont_DE * MoInstr	0.072	(0.038)	1.9

Table 4: Results summary for unigraph model with coefficient estimates *ß*, standard errors SE(*ß*), and associated t-scores for all predictors in analysis. Cue availability is given as MGraphFreq, cue reliability as GPCoFreq, and cue validity as GPCont

Model	Df	AIC	BIC	logLik	Dev	X^2	X^2 Df	Pr(>X^2)	R^2
Null	8	-1608.9	-1563.0	812.45	-1624.9				0.468
Unigraph	17	-1651.2	-1553.6	842.62	-1685.2	60.3	9	< 0.0001	0.474

Table 5: Summary of comparison between unigraph model and null model with degrees of freedom, Aikaike information criterion, Bayesian information criterion, log likelihood, deviance and R^2 for each model, and the X^2 statistic, degrees of freedom and probability estimate for the ANOVA comparison of the models

Predictor	Standardized *(ß)*
rGPCoFreq_DE	0.16863
GPCont_EN	0.25099
GPCont_EN * MoInstr	-0.08409
rGPCoFreq_DE * GPCont_DE	-0.13019
MGraphFreq_DE * MoInstr	0.04647

Table 6: Summary of standardized regression coefficients *(ß)* for all significant predictors of interest. Cue availability is given as MGraphFreq, cue reliability as GPCoFreq, and cue validity as GPCont

As seen by the decreasing slopes in Figure 2, the effect of English unigraph-phone contingency (cue validity) decreases as amount of instruction received increases. This suggests that learners are beginning to rely less on knowledge of English orthography as exposure to their L2 increases. Concurrent with this decrease in the effect of English orthographic knowledge, the effect of German unigraph frequency (cue availability) shows a trend towards a positive slope, depicted in Figure 3. Additionally, the German co-occurrence frequency of unigraphs and phones (cue reliability) appears to modulate the effect of German unigraph-phone contingency (cue validity), such that only low frequency pairings exhibit a clear benefit from increased contingency, whereas mid and high frequency pairings show a nearly flat slope. Figure 4 shows these effects.

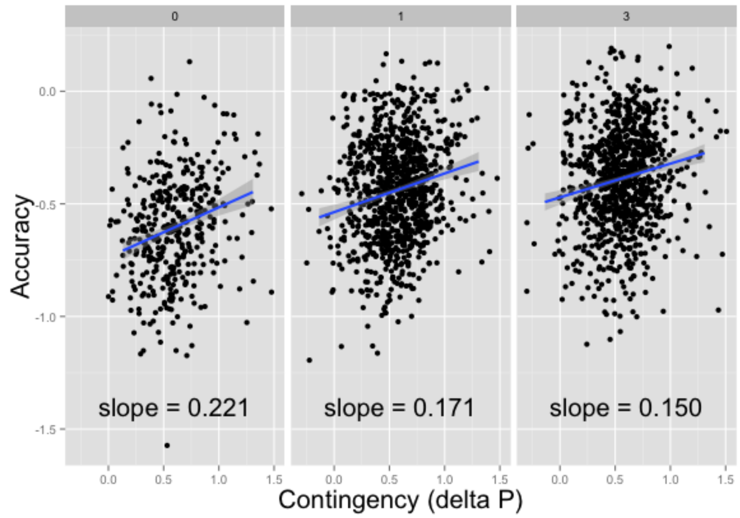

Figure 2: Effect of English unigraph-phone contingency (cue validity) on accuracy for 0, 1 and 3 months of instruction

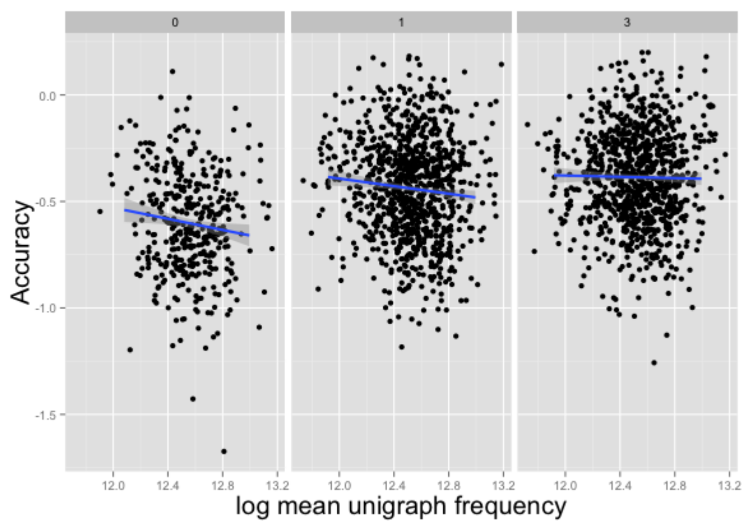

Figure 3: Effect of German unigraph frequency (cue availability) on accuracy for 0, 1 and 3 months of instruction

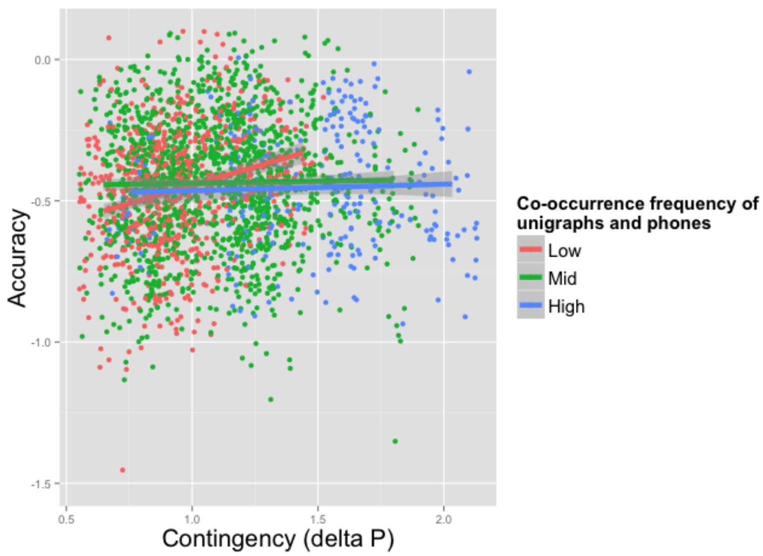

Figure 4: Effects of German unigraph-phone contingency (cue validity) on accuracy for low, mid and high unigraph-phone co-occurrence frequencies (cue reliability)

Finally, although it did not reach significance, the three-way interaction between German co-occurrence frequency (cue reliability), contingency (cue validity) and months of instruction did show a clear trend towards significance and could not be removed from the model without making it worse by a marginally significant degree. As such, the effect of amount of instruction on the interaction between German co-occurrence frequency and contingency are explored in Figure 5. As can be seen in this figure, the modulating effect of frequency on contingency seems to disappear as exposure increases, and all frequency bands begin to show positive effects of contingency, rather than just the low frequency band.

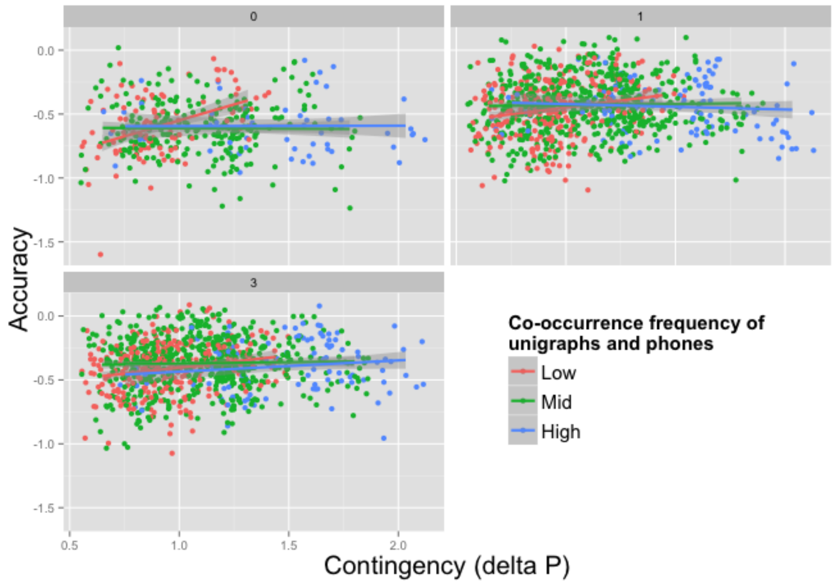

Figure 5: Effect of months of instruction on the interaction between German unigraph-phone contingency (cue validity) and co-occurrence frequencies (cue reliability)

3.3.1 Discussion of unigraph model

Results of the unigraph model indicate that L2 pronunciation accuracy during phonological decoding is indeed influenced by L2 and L1 orthography-phonology statistics, and that these effects are graded. L1 effects include a significant main effect of English cue validity which decreases as amount of instruction increases. This suggests that learners may be aided in the initial stages of learning how L2 orthography corresponds to its sound system by orthographic cues that have high cue validity in the L1, regardless of their validity in the L2. That is, if a set of orthographic cues reliably predicts a set of phones in a learner's L1 regardless of L2 cue validity, the learner will rely on the L1 cue statistics in the initial stages of acquisition when determining pronunciation of a word. As exposure to the L2 increases, however, the learner will begin to rely less on L1 cue-outcome statistics (e.g., cue validity), and rely instead on L2 cue-outcome statistics. In other words, without sufficient L2 exposure for a learner to adequately learn which orthographic cues predict which phones and how reliably, the learner has primarily only L1 orthography-phonology statistics to rely on. It is possible that the positive effects of English cue validity simply reflect L1 transfer, whereby learners inadvertently produce correct phones when strongly predicted by English orthography-phonology statistics. The decrease in slope for this effect as exposure increases does, however, suggests that learners are adjusting their cue-outcome statistics to reflect the L2 input, and that they are beginning to rely less on L1 statistics as they learn which L2 cues reliably predict a word's phonology.

This is further evidenced by the interactions between German cue reliability, cue validity and months of instruction. At zero months of instruction, a beneficial effect of cue validity is only evident for unigraph-phone pairs with low cue reliability, i.e., unigraphs that occur infrequently in the orthographic forms of words whose phonological forms contain the target phones in the prescribed pronunciations of the test words. As amount of instruction increases, however, this difference attenuates and positive effects of cue validity begin to emerge for all co-occurrence frequency (cue reliability) bands. In other words, this suggests that when cue-outcome pairings are sufficiently high frequency (i.e., they have high cue reliability), learners will rely primarily on cue reliability in early L2 acquisition, rather than on cue validity.

A possible explanation for this interaction is that it reflects the input structure of the L2 classroom. During the early stages of instruction, learners are primarily exposed to high frequency words, whose cue-outcome frequencies would also tend to be higher as an artifact of their high lexical frequency–and indeed, amongst the stimuli used for this study, a positive, albeit low correlation exists between German cue reliability and word frequency ($r = 0.14$, $p < 0.0001$). Without evidence to the contrary, a learner will also treat a high reliability cue as having high validity. In other words, if an orthographic cue frequently co-occurs with a particular phone, learners will treat this cue as being a highly reliable predictor of that phone. As such, until cue validity statistics can be adjusted by sufficient exposure to the low reliability counterparts of the high frequency cue-outcome pairs, cue validity will offer little or no predictive power beyond cue reliability, and will thus be uninformative. To illustrate, if a learner of English has only encountered the phoneme /s/ in words whose orthographic forms use <s> (e.g., *six, school*), they will treat the unigraph <s> as a highly valid cue to the phone /s/ due to its high cue reliability. As they begin to encounter words with low reliability cues to /s/ such as *extra* and *century*, the reliability of <s> as a cue to /s/ remains relatively unchanged, but the validity becomes lower to reflect that fact that <s> is not the only cue to /s/.

In short, L1 English learners of German rely primarily on English cue validity during phonological decoding at the earliest stages of acquisition, which likely reflects a lack of sufficient exposure to German for adequate cue-outcome learning to have occurred. As exposure increases, learners begin to rely less on

English cue-outcome statistics and more on those of German. Consistent with the predictions of usage-based accounts of language learning, these results find that orthographic input does indeed have a graded effect on L2 spoken language development. They further show that L2 pronunciation accuracy during phonological decoding is influenced variably by both L1 and L2 measures of cue availability, cue reliability and cue validity. Additionally, results find that the effects orthography-phonology statistics on pronunciation accuracy are influenced by L2 exposure. Results do not, however, find any interactions between L1 measures and L2 measures.

3.4 Bigraph Model

Results of the bigraph model are given in Table 7. The frequency of co-occurrence for German bigraphs and phones (cue reliability) is residualized in this model from the frequency of German bigraphs (cue availability) in order to reduce collinearity. Comparison to the null model indicates that the bigraph model performs significantly better, accounting for 0.4% more variance (Table 8).

As was found for the unigraph model, there is a significant main effect of English unigraph-phone contingency (cue validity), which again exhibits the strongest effect size (Table 9). Though the interaction between English unigraph-phone contingency and months of instruction shows the same directionality as in the unigraph model, it does not quite reach significance. Moreover, whereas there was a significant main effect of the residualized n-graph-phone co-occurrence frequency (cue reliability) for German in the unigraph model, this predictor does not reach significance in the bigraph model.

Predictor	Coef	SE	t
Intercept	-0.006	(0.022)	-0.3
PhonLength	-0.012	(0.008)	-1.6
Familiarity	0.007	(0.002)	2.8
MoInstr	0.022	(0.004)	5.5
MGraphFreq_DE	-0.017	(0.024)	-0.7
GPCont_DE	-0.015	(0.030)	-0.5
GPCont_EN	**0.010**	**(0.029)**	**3.4**
rGPCoFreq_DE	0.012	(0.068)	0.2
GPCoFreq_EN	0.033	(0.031)	1.0
MGraphFreq_DE X MoInstr	**0.016**	**(0.006)**	**2.7**
GPCont_EN X MoInstr	-0.011	(0.006)	-1.8
GPCont_DE X rGPCoFreq_DE	**-0.293**	**(0.088)**	**-3.3**
GPCont_DE X rGPCoFreq_DE X MoInstr	**0.054**	**(0.022)**	**2.4**

Table 7: Results summary for bigraph model with coefficient estimates *ß*, standard errors SE(*ß*), and associated t-scores for all predictors in analysis. Cue availability is given as MGraphFreq, cue reliability as GPCoFreq, and cue validity as GPCont

Model	Df	AIC	BIC	logLik	Dev	X^2	X^2 Df	Pr(>X^2)	R^2
Null	8	-1608.9	-1563.0	812.45	-1624.9				0.468
Bigraph	17	-1640.0	-1542.3	836.98	-1674.0	49.1	9	< 0.0001	0.472

Table 8: Summary of comparison between bigraph model and null model with degrees of freedom, Aikaike information criterion, Bayesian information criterion, log likelihood, deviance and R^2 for each model, and the X^2 statistic, degrees of freedom and probability estimate for the ANOVA comparison of the models

Predictor	Standardized *(ß)*
GPCont_EN	0.21654
MGraphFreq_DE X MoInstr	0.04586
GPCont_DE X rGPCoFreq_DE	-0.15643
GPCont_DE X rGPCoFreq_DE X MoInstr	0.03407

Table 9: Summary of standardized regression coefficients *(ß)* for all significant predictors of interest. Cue availability is given as MGraphFreq, cue reliability as GPCoFreq, and cue validity as GPCont

Additionally, three significant interactions emerged from the bigraph model. As in the unigraph model, the bigraph model reveals significant interactions between German n-graph frequency (cue availability) and months of instruction, and between the co-occurrence frequency (cue reliability) and contingency (cue validity) of the German n-graph-phone pairs. As amount of instruction in German increases, German bigraph frequency starts to show a positive slope in its interaction with accuracy (Figure 6), suggesting a very early sensitivity to bigraph frequency that begins to benefit phonological decoding by three months of instruction. The interaction between German co-occurrence frequency (cue reliability) and contingency (cue validity) also closely resembles that of the unigraph model, with beneficial effects of cue reliability only evident for low reliability pairings of bigraphs and phones (Figure 7). Finally, the three-way interaction between German co-occurrence frequency (cue reliability), contingency (cue validity) and months of instruction does reach significance in the bigraph model, with the modulating effect of cue reliability on cue validity disappearing as amount of instruction increases (Figure 8).

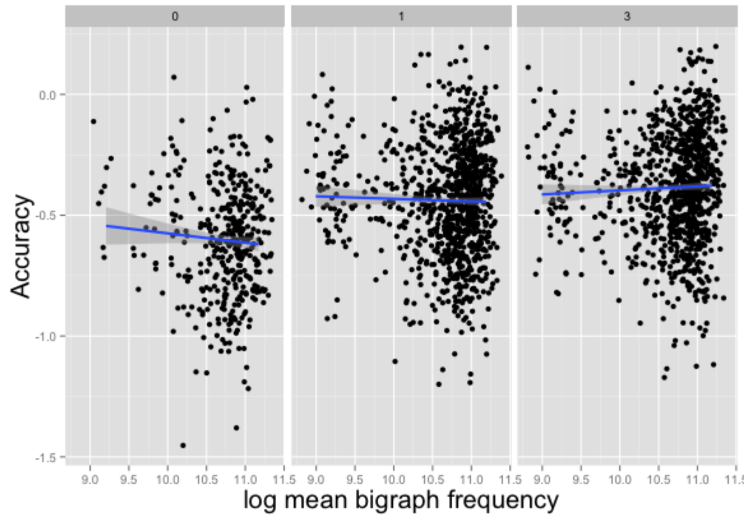

Figure 6: Effect of German bigraph frequency (cue availability) on accuracy for 0, 1 and 3 months of instruction

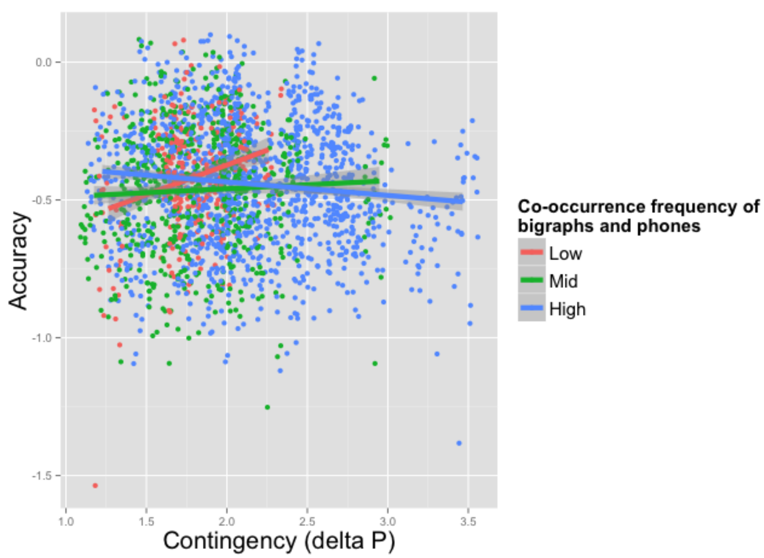

Figure 7: Effects of German bigraph-phone contingency (cue validity) on accuracy for low, mid and high bigraph-phone co-occurrence frequencies (cue reliability)

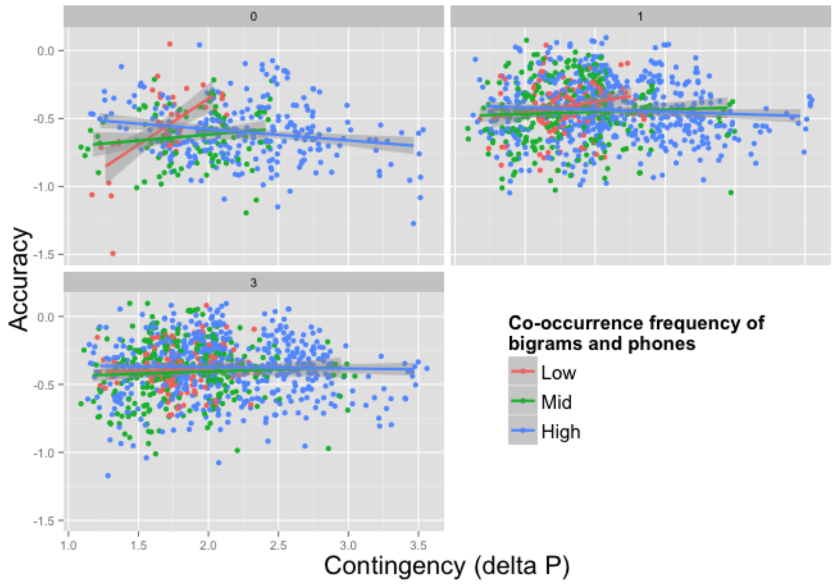

Figure 8: Effect of months of instruction on the interaction between German bigraph-phone contingency (cue validity) and co-occurrence frequencies (cue reliability)

3.4.1 Discussion of bigraph model

Results from the bigraph model corroborate those of the unigraph model, finding sensitivity to the statistical properties of two-letter sequences during L2 phonological decoding. While learners appear again to rely most heavily on English cue validity information, the interaction between this predictor and months of instruction does not quite reach significance, as it does in the unigraph model. In further contrast to the unigraph model, no German orthographic predictors show significant main effects. These findings may reflect the overall lower frequency rates of bigraphs (*mean per million frequency* = 7071) compared to unigraphs (*mean per million frequency* = 130,057). As a result of this low availability, learners will require greater L2 exposure in order to have encountered bigraphs with sufficient frequency to accurately learn the statistical properties describing the availability, reliability and validity of bigraphs as cues to German phones. If this is the case, similar results should be observed for the trigraphs (*mean per million frequency* = 715).

The interaction in this model between German bigraph frequency (cue availability) and months of instruction further indicates that learners require sufficient exposure to German orthography before frequency can exert a positive effect on decoding accuracy, as positive effects of cue availability do not emerge until after 3 months of instruction. The importance of exposure is also indicated by the interaction between German cue reliability, cue validity and months of instruction. In the earliest stages of acquisition, positive effects of cue validity are only evident for lower reliability cue-outcome pairs. With increased exposure, however, the differential effects of cue reliability disappear. As was suggested for the unigraph model, this may reflect the gradual exposure of L2 learners to low reliability cues in the input, which would trigger the adjustment of cue validity statistics to reflect the newly acquired implicit knowledge that high reliability cues do not always have equally high validity. That is, because the accuracy of cue validity depends on both negative evidence (i.e., evidence that a cue and outcome do not always occur together) and positive evidence (i.e., evidence that a cue and outcome occur together), the learning of this measure will necessarily require greater exposure in order for sufficient positive and negative evidence to accumulate for the learners' representations of cue validity to accurately reflect the distribution of these cues and outcomes in the language.

In summary, the results of this model find further evidence that L1 and L2 cue availability, cue reliability and cue validity exhibit graded effects on L2 pronunciation accuracy and its development over time during phonological decoding. Like the unigraph model, this model does not find any interactions between L1 and L2 cue-outcome measures and their influence on pronunciation accuracy. Moreover, not all effects that were significant in the unigraph model reach significance in the bigraph model, indicating that learners are treating one- and two-letter cues differently.

3.5 Trigraph Model

Results for the trigraph model are shown in Table 10. Due to its high variance inflation factor, the co-occurrence frequency (cue reliability) for German trigraph-phone pairs was residualized from German trigraph frequency (cue availability), reducing all variance inflation factors to values less than 4. In an analysis of variance, the trigraph model performs significantly better than the null model, accounting for 0.5% more of the variance (Table 11).

Similar to the unigraph and bigraph models, the trigraph model finds a significant main effect of English n-graph-phone contingency (cue validity) such that words with higher normalized English contingency are produced with greater accuracy. As in the unigraph model, this effect is modulated by amount of instruction, decreasing as months of instruction increases (Figure 9). A main effect of German trigraph-phone co-occurrence frequency (cue reliability) is also present in the model, indicating that learners' pronunciation accuracy is sensitive to the frequency with which trigraphs in German serve as cues to German phones.

Predictor	Coef	SE	t
Intercept	-0.005	(0.022)	-0.2
PhonLength	-0.015	(0.008)	-1.8
Familiarity	0.008	(0.003)	2.9
MoInstr	0.022	(0.004)	5.5
MGraphFreq_DE	0.003	(0.012)	0.2
GPCoFreq_EN	-0.004	(0.015)	-0.2
GPCont_DE	-0.025	(0.017)	-1.4
GPCont_EN	**0.070**	**(0.018)**	**3.8**
rGPCoFreq_DE	**0.129**	**(0.059)**	**2.2**
MGraphFreq_DE X MoInstr	**0.007**	**(0.003)**	**2.6**
GPCont_EN X MoInstr	**-0.012**	**(0.004)**	**-2.8**
GPCont_DE X rGPCoFreq_DE	**-0.119**	**(0.052)**	**-2.3**
GPCoFreq_EN X rGPCoFreq_DE	**-0.138**	**(0.066)**	**-2.1**
GPCoFreq_EN X rGPCoFreq_DE X MoInstr	**0.047**	**(0.015)**	**3.0**

Table 10: Results summary for trigraph model with coefficient estimates *ß*, standard errors SE(*ß*), and associated t-scores for all predictors in analysis. Cue availability is given as MGraphFreq, cue reliability as GPCoFreq, and cue validity as GPCont

Model	Df	AIC	BIC	logLik	Dev	X^2	X^2 Df	Pr(>X^2)	R^2
Null	8	-1608.9	-1563.0	812.45	-1624.9				0.468
Trigraph	17	-1643.9	-1540.5	839.95	-1679.9	55.0	10	< 0.0001	0.473

Table 11: Summary of comparison between trigraph model and null model with degrees of freedom, Aikaike information criterion, Bayesian information criterion, log likelihood, deviance and R^2 for each model, and the X^2 statistic, degrees of freedom and probability estimate for the ANOVA comparison of the models

Predictor	Standardized *(ß)*
GPCont_EN	0.22572
rGPCoFreq_DE	0.15509
MGraphFreq_DE X MoInstr	0.04324
GPCont_EN X MoInstr	-0.04494
GPCont_DE X rGPCoFreq_DE	-0.10824
GPCoFreq_EN X rGPCoFreq_DE	-0.15259
GPCoFreq_EN X rGPCoFreq_DE X MoInstr	0.06182

Table 12: Summary of standardized regression coefficients *(ß)* for all significant predictors of interest. Cue availability is given as MGraphFreq, cue reliability as GPCoFreq, and cue validity as GPCont

Consistent with the results of the unigraph and bigraph models, the effect of German n-graph frequency (cue availability) is modulated by months of instruction, exhibiting a positive-going trend (Figure 10). The interaction between German n-graph-to-phone frequency (cue reliability) and contingency (cue validity) also appears similar for trigraphs as it does for unigraphs and bigraphs, although low frequency trigraph-phone pairings appear to benefit less from increased contingency than do low frequency bigraph-phone and unigraph-phone pairings (Figure 11).

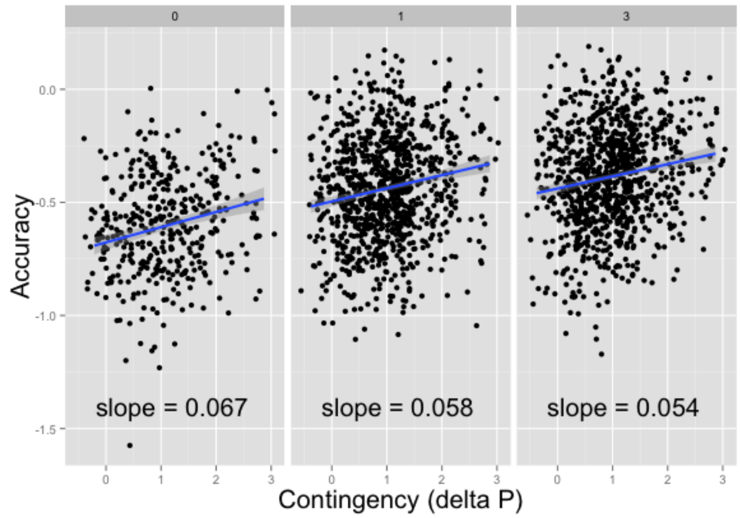

Figure 9: Effect of English trigraph-phone contingency (cue validity) on accuracy for 0, 1 and 3 months of instruction

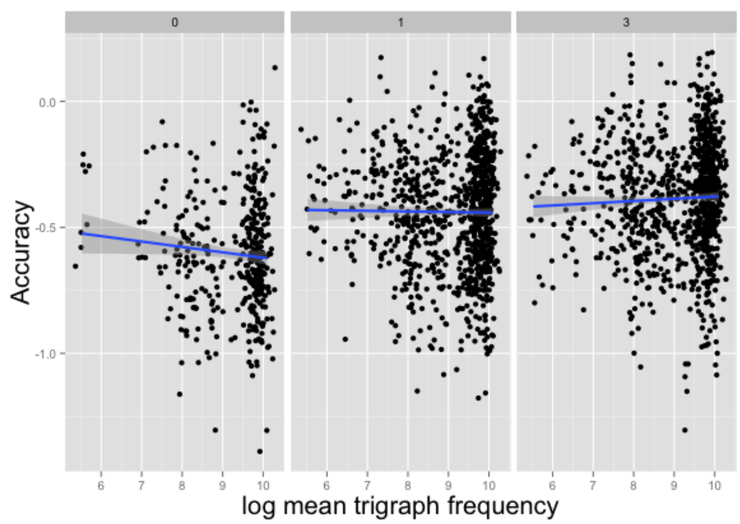

Figure 10: Effect of German trigraph frequency (cue availability) on accuracy for 0, 1 and 3 months of instruction

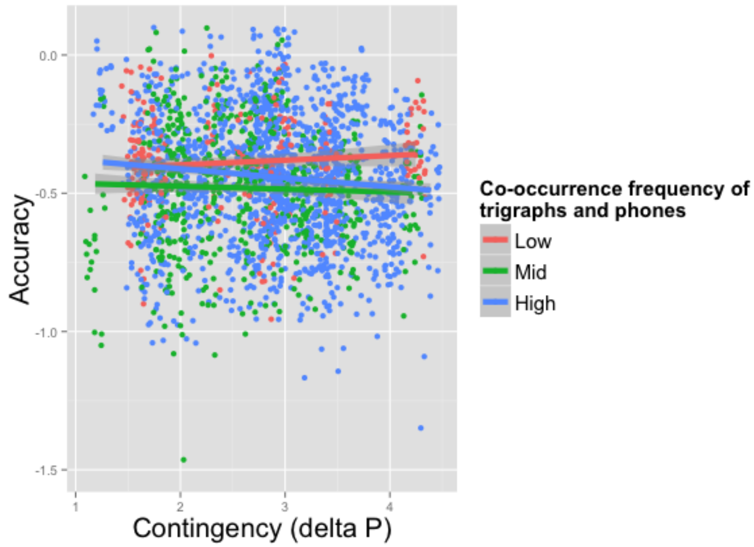

Figure 11: Effects of German trigraph-phone contingency (cue validity) on accuracy for low, mid and high trigraph-phone co-occurrence frequencies (cue reliability)

In addition to the effects discussed thus far, two new interactions emerge as significant in the trigraph model that do not emerge in the unigraph and bigraph models. Figure 12 shows the interaction between German and English trigraph-phone co-occurrence frequencies (cue reliability). It is interesting to note the negative effect that increased English trigraph-phone frequency (cue reliability) has on accuracy for trigraph-phone pairs that have lower frequencies in German. This may indicate that low reliability pairings in German do not benefit from increasing cue reliability in English without sufficient exposure in order to enable learners to learn that these pairings occur in German and that the outcomes are reliably predicted by English cue reliability. Indeed, Figure 13 shows a change in slope for the low frequency German pairings as amount of instruction increases that is consistent with this possibility.

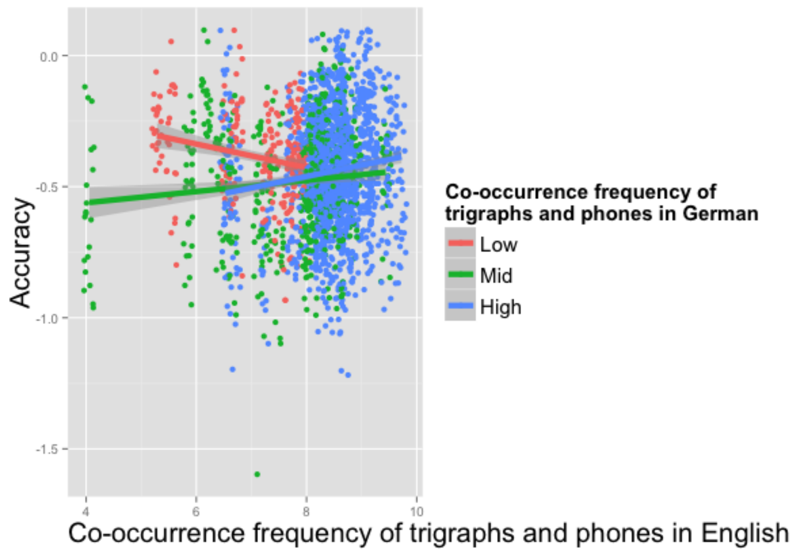

Figure 12: Interaction between German and English trigraph-phone co-occurrence frequencies (cue reliability)

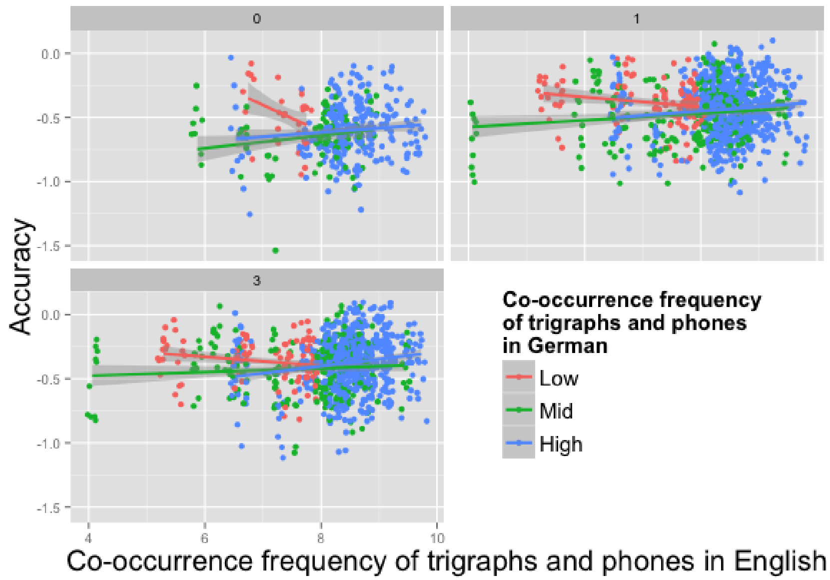

Figure 13: Interaction between German and English trigraph-phone co-occurrence frequencies (cue reliability) at 0, 1 and 3 months of instruction

3.5.1 Discussion of trigraph model

Adding to the results of the unigraph and bigraph models, the trigraph model indicates that L2 learners' sensitivity to the statistical properties of L2 orthography-phonology relationships in phonological decoding also extends to three-letter orthographic cues. Yet again, the predictor showing the strongest effect on pronunciation accuracy is English cue validity. As in the unigraph model, this predictor shows a decrease in its effect on accuracy as amount of instruction in the L2 increases, suggesting that learners begin to rely on other cues with increased exposure. While German cue reliability does not reach significance in the bigraph model, it does show a significant positive main effect in the trigraph model on accuracy, as it does in the unigraph model. Thus, it seems that differential frequency rates of one-, two-, and three-letter sequences are insufficient to explain the non-significance of this predictor in the bigraph model, as was proposed earlier.

It is possible that this result stems from particular statistical properties of German or English, or of the experimental stimuli. Alternatively, this finding may indicate differences in how letter sequences of varying sizes are processed by native-English speaking learners of German. For example, as Ziegler & Goswami's psycholinguistic grain size theory (2005) might predict, these results may reflect differences in how L2 learners use varying grain sizes when learning the relationship between L2 orthography and phonology. Briefly, this theory argues that reading depends, in part, on learning to use appropriate grain sizes during reading to arrive at a word's phonology and meaning. Depending on the consistency with which orthography maps to phonology in a language, the grain sizes may be whole words, syllables, onsets, rimes, graphemes, phonemes, letters, or a combination of these. It may thus be that the variation between models is a result of L2 learners making varying use of different n-graph sizes in order to reliably arrive at a word's pronunciation.

The significant interaction between German trigraph frequency (cue availability) and months of instruction, whereby trigraph frequency begins to exhibit a positive effect on accuracy, indicates that frequency effects for trigraphs arise early in the acquisition of L2 phonological decoding skills. As in the previous models, a modulating effect of German cue reliability is found for cue validity's effect on accuracy. Consistent with these other models, only items with low cue reliability show any positive effect of cue validity on accuracy, which suggests that the learning of accurate cue validity statistics requires greater L2 exposure for high reliability cues, in order for learners to encounter sufficient evidence that these cues also predict other outcomes, and that the outcomes in these pairings are also predicted by other cues. This interaction does not, however, show any significant nor even marginally significant interaction with months of instruction, perhaps due to the lower overall frequency rates of trigraphs, as noted earlier.

Finally, whereas the unigraph and bigraph models do not find any significant interactions between German and English predictors, one such interaction does emerge in the trigraph model between German and English cue reliability. This interaction reveals that the accuracy of words with mid to high cue reliability in German increases as the English cue reliability of these words increases. For items with low cue reliability in German, however, the opposite is observed, with accuracy decreasing as English cue reliability increases. This may reflect a lack of exposure to low-frequency German pairings, which, if greater, would allow learners to recognize that the outcomes are reliably predicted by English cue reliability statistics. Alternatively, this may indicate interference from English for these items with low cue reliability in German, possibly because the spreading activation received by the German phonemes is insufficient to suppress the inaccurate English phonemes also receiving activation. An error analysis would help to determine whether this latter possibility might indeed be the case. This interaction is, however, modulated by amount of instruction, and the negative effect of increased English cue reliability for German items with low cue reliability attenuates as amount of instruction increases.

To summarize, consistent with the unigraph and bigraph models, the trigraph model reveals graded effects of L1 and L2 orthographic measures on L2 spoken language skills during SLA. Cue availability, cue reliability and cue validity are found to variably influence the accuracy of pronunciation during phonological decoding, and the change in this accuracy as L2 exposure increases. Additionally, unlike the other models, this model finds a significant interaction between L1 and L2 measures that furthermore interacts with L2 exposure to influence pronunciation accuracy and its development over time. The following section will now consider the combined results of this study in the context of the existing literature, drawing implications for language learning and for future research.

4. General discussion and conclusion

Consistent with usage-based accounts of language learning, this study finds graded effects of orthography-phonology statistics on pronunciation accuracy for native English speaking learners of German in a reading-aloud task. In answer to research question one, this indicates that graded effects of orthography on L2 spoken language skills do indeed occur during SLA, as shown by the linear relationships found between the accuracy with which L2 learners of German pronounced German words, and the statistical measures used to describe the relationship between orthography and phonology. Results also find positive answers to research questions two and three: the availability, reliability and validity of one-, two- and three-letter orthographic cues are found to influence pronunciation accuracy of L2 learners in a phonological decoding task (research question 2), and the development of accuracy as L2 exposure increases (research question 3). Regarding research question four, results further indicate that L1 and L2 cue-outcome measures do indeed interact, though this is only found for trigraphs. Together, these results indicate that instructed L2 learners do learn the statistical properties describing how L2 orthography corresponds to L2 phonology, and that this learning is a usage-based process driven in part by input frequencies and length of L2 exposure.

Effects of cue availability, reliability and validity were found in each of the unigraph, bigraph and trigraph models. There was, however, variation between models in which predictors and which interactions reached significance. As suggested earlier, this variation may reflect differences in how L2 learners use varying grain sizes while learning how L2 orthography maps to L2 phonology, as Ziegler & Goswami's psycholinguistic grain size theory (2005) might predict. If so, this raises the question of whether the use of grain sizes observed in this study reflects transfer of L1 processing strategies, or whether it reflects the optimal use of grain size given the statistical properties of German orthography-phonology cue-outcome relationships. Alternatively, the variation between models may be a result of differences in frequency, given that larger letter sequences will also tend to occur with a lower mean frequency than smaller letter sequences. If this is the case, we might expect that the models would more closely approximate one another given enough time and exposure to German for participants to reach a saturated state at which new German input no longer significantly changes the cue-outcome statistics of the learners' German systems.

Future research might also examine whether the measures used in this study have predictive power in L2 perception experiments. For example, these measures may help to understand the differential results found by Escudero et al. (2008), Simon et al. (2010), Showalter (2012) and Showalter & Hayes-Harb (2013) with regards to whether L2 learners use orthographic information to aid in the learning of novel L2 contrasts. It may be that the orthographic tone marks used in Showalter & Hayes-Harb's (2013) study were quickly learned as highly valid cues in part because the participants were able to rely on L1 orthography-phonology statistics due to the shared scripts of English and Chinese pinyin. This may have enabled them to more easily learn the phones in the stimuli, easing cognitive load and affording participants a greater opportunity to notice and learn the tones as well. The L1 English participants in Showalter's (2012) study, however, would have been unable to rely on L1 orthography-phonology statistics, as Arabic and English use different scripts. As such these participants may have experienced greater cognitive load than the participants in Showalter & Hayes-Harb's (2013) study. In addition to possible effects of L2 exposure and proficiency, it is furthermore possible that the different results found by Escudero et al. (2008) and Simon et al. (2010), whose studies only used languages employing the Roman alphabet, may reflect differences in L1 and L2 cue availability, reliability and validity for the letter-sound correspondences of interest. Thus, future studies investigating orthographic effects on the learning of novel L2 contrasts and on other aspects of L2 perception might benefit from considering orthography-phonology statistics in their stimuli design and statistical analyses.

Importantly, this study does face limitations due to its use of two closely related languages that have highly similar writing systems. Given that previous research has found different effects of orthography on L2 spoken language skills that appear to depend on how similar L1 and L2 writing systems are (e.g., Koda, 1990; Erdener & Burnham, 2005; Hamada & Koda, 2008, 2010), future research might examine whether different results than those found in this study might emerge if L1s and L2s are used with varying degrees of similarity (e.g., in orthographic depth). Typological distance and L1-L2 script differences might also influence which orthographic measures predict performance, how well, how they interact with one another, and whether L1 effects still emerge. Research might also investigate if and how cue availability, reliability and validity are utilized when learning non-phonographic L2 scripts.

Finally, given that classroom-based SLA exposes learners simultaneously to an L2's orthography and phonology–in many cases learners likely encounter a word's orthographic form before ever hearing its spoken form–it is certainly plausible to expect that cue-outcome statistics describing the relationship between L2 orthography and phonology may more generally predict the acquisition of pronunciation in that L2. Indeed, the results of this study provide preliminary support for this hypothesis, albeit with the crucial limitation that productions were prompted using a phonographic orthography. Previous research has, however, found orthographic effects in spoken language production in the absence of phonographic cues to phonology (Bassetti, 2006, 2007). This raises the question of whether the results of this study will generalize beyond phonological decoding to also predict pronunciation accuracy and its development in spontaneous speech and in tasks such as picture naming, which do not use orthography. If so, this will have implications for models of L2 pronunciation learning (e.g., Flege, 1995), which will need to account for orthographic effects such as those described in this study.

In short, this study finds graded effects of orthography-phonology statistics on pronunciation accuracy in an L2 reading-aloud task. The availability, reliability and validity of English and German unigraphs, bigraphs and trigraphs as cues to German phones were found to variably predict pronunciation accuracy and its development over time as exposure to German increased. These results indicate that L2 learners implicitly learn the statistical properties of the orthography-phonology relationship of their L2, and are able to begin relying on these statistics quite early in order to improve phonological decoding accuracy. The finding of graded effects also raises the possibility that future research might find other such graded effects by complementing the use of traditional categorical variables (e.g., regular/irregular, consistent/inconsistent, deep/shallow) with continuous variables such that those used in this study. This in turn may offer deeper insight into how a second language is learned, and how the statistical properties of the L1 and L2 influence this process.

Works cited

Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society*, *15*(3), 147-149.

Audacity Development Team (2012). Audacity(R): A Free, Cross-Platform Digital Audio Editor (Version 2.0.0) [computer software]. Retrieved from http://audacity.sourceforge.net/.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge, UK: Cambridge University Press.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). The CELEX Lexical Database. Philadelphia: Linguistic Data Consortium.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39*(3), 445-459.

Barkley, S. (2010). The acquisition of orthographic-phonological correspondence rules in L2 and L3 Portuguese: Error resolution, interference and generalizability. Unpublished Ph.D. Dissertation. University of Florida, USA.

Baron, J., & Strawson, C. (1976). Use of orthographic and word-specific knowledge in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(3), 386-393.

Bassetti, B. (2006). Orthographic input and phonological representations in learners of Chinese as a foreign language. *Written Language and Literacy, 9*(1), 95-114.

Bassetti, B. (2007). Effects of hanyu pinyin on pronunciation of learners of Chinese as a Foreign Language. In A. Guder, X. Jian, & Y. Wan (Eds.), *The cognition, learning and teaching of Chinese characters* (pp. 156-179). Beijing: Beijing Language and Culture University Press.

Bassetti, B. (2008). Orthographic input and second language phonology. In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 191-206). Clevedon, UK: Multilingual Matters.

Bates, E., & MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 173-218). New York: Cambridge University Press.

Bates, D., Maechler, M., Bolker, B., Walker, S. (2013). Lme4: Linear mixed-effects models using Eigen and S4. (R package version 1.0.5) [software package]. Retrieved from http://CRAN.R-project.org/package=lme4.

Bijeljac-Babic, R., Biardeau, A., & Grainger, J. (1997). Masked orthographic priming in bilingual word recognition. *Memory & Cognition, 25*(4)*,* 447-457.

Boersma, P., Weenink, D. (2013). Praat: Doing phonetics by computer (Version 5.3.53) [computer software]. Retrieved from http://www.praat.org/.

Brewer, J. (2008). Phonetic reflexes of orthographic characteristics in lexical representations. Unpublished Ph.D. Dissertation. University of Arizona, USA.

Brondsted, T. (2008). Automatic Phonemic Transcriber (Version 1.1). http://tom.brondsted.dk/text2phoneme/ (accessed on 10.02.2014).

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology, 58*(5), 412-424.

Brysbaert, M. & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977-990.

Cook, V. & Bassetti, B. (2005). An introduction to researching second language writing systems. In V. Cook & B. Bassetti (Eds.) *Second language writing systems* (pp. 1-67). Clevedon, UK: Multilingual Matters.

Cossu, G. (1999). The acquisition of Italian orthography. In M. Harris & G. Hatano (Eds.), Learning to read and write: A cross-linguistic perspective (pp. 10-33). New York: Cambridge University Press.

Coulmas, F. (1999). *The Blackwell encyclopedia of writing systems.* Oxford: Blackwell Publishers.

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics, 34*(2), 269–284.

Damian, M. F., Bowers, J. S. (2003). Effects of orthography on speech production in a form preparation paradigm. *Journal of Memory and Language, 49*(1), 119-132.

Dijkstra, T., Grainger, J., & Van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language, 41*(4), 496-518.

Dijkstra, T., Roelofs, A., & Fieuws, S. (1995). Orthographic effects on phoneme monitoring. *Canadian Journal of Experimental Psychology*, *49*(2), 264-

271.

Dijkstra, T., Timmermans, M., & Schriefers, H. (2000). On being blinded by your other language: Effects of task demands on interlingual homograph recognition. *Journal of Memory and Language, 42*(4)*,* 445-464.

Dijkstra, T., Van Jaarsveld, H., & Ten Brinke, S. (1998). Interlingual homograph recognition: Effects of task demands and language intermixing. *Bilingualism: Language and Cognition, 1*(1)*,* 51-66.

Duden (2005). *Das Aussprachewörterbuch*. 6th ed. Mannheim: Dudenverlag.

Duden (2010). *Das Fremdwörterbuch*. 10th ed. Mannheim: Dudenverlag.

Durgonoğlu, A. Y. (2006). How language characteristics influence Turkish literacy development. In R. M. Joshi & P. G. Aaron (Eds.) *Handbook of orthography and literacy* (pp. 219-29). Mahwah, NJ: L. Erlbaum.

Eisenberg, P. (2006). Das Wort: Grundriss der deutschen Grammatik. 3rd ed. Stuttgart: Verlag J.B. Metzler.

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition, 24*(2), 143-188.

Ellis, N. C. (2006a). Language acquisition as rational contingency learning. *Applied Linguistics*, *27*(1), 1-24.

Ellis, N. C. (2006b). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics 27*(2), 164-194..

Ellis, N. C., & Beaton, A. (1993). Psycholinguistic determinants of foreign language vocabulary. *Language Learning, 43*(4)*,* 559-617.

Ellis, Nick C., & Ferreira-Junior, F. (2009a). Construction learning as a function of frequency, frequency distribution, and function. *The Modern*

Language Journal, 93(3), 370-385.

Ellis, Nick C., & Ferreira-Junior, F. (2009b). Constructions and their acquisition: Islands and the distinctiveness of their occupancy. *Annual Review of Cognitive Linguistics*, *7*(1), 187-220.

Ellis, N. C. & Larsen-Freeman, D. (2009). Constructing a second language: Analyses and computational simulations of the emergence of linguistic constructions from usage. *Language Learning, 59*(s1), 90-125.

Ellis, N. C. & O'Donnell, M. B. (2012). Statistical construction learning: Does a Zipfian problem space ensure robust language learning? In P. Rebuschat & J. N. Williams (Eds.) *Statistical learning and language acquisition* (pp. 265-304). Berlin: Mouton de Gruyter.

Erdener, V. D. & Burnham, D. K. (2005). The role of audiovisual speech and orthographic information in nonnative speech production. *Language Learning, 55*(2), 191-228.

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics, 36*(2), 345-360.

Escudero, P. & Wanrooij, K. (2010). The effect of L1 orthography on non-native vowel perception. *Language and Speech, 53*(3), 343-365.

Ferrand, L. & Grainger, J. (1992). Phonology and orthography in visual word recognition: Evidence from masked non-word priming. *Quarterly Journal of Experimental Psychology, 45A*(3), 353-372.

Ferrand, L. & Grainger, J. (1993). The time course of orthographic and phonological code activation in the early phases of visual word recognition. *Bulletin of the Psychonomic Society, 31*(2), 119-122.

Ferrand, L. & Grainger, J. (1994). Effects of orthography are independent of phonology in masked form priming. *The Quarterly Journal of Experimental Psychology, 47A*(2), 365-382.

Ferrand, L. & Grainger, J. (2003). Homophone interference effects in visual word recognition. *The Quarterly Journal of Experimental Psychology, 56A*(3), 403-419.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). Timonium, MD: York Press.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, *12*(6), 627-635.

Frederiksen, J. R., & Kroll, J. F. (1976). Spelling and sound: Approaches to the internal lexicon. *Journal of Experimental Psychology: Human Perception and Performance, 2*(3), 361-379.

Goswami, U. (2006). Orthography, phonology, and reading development: A cross-linguistic perspective. In R. M. Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy* (pp. 463-80). Mahwah, NJ: L. Erlbaum.

Grainger, J., Kiyonaga, K., & Holcomb, P. J. (2006). The time course of orthographic and phonological code activation. *Psychological Science, 17*(12), 1021-1026.

Hallé, P. A., Chéreau, C., & Segui, J. (2000). Where is the /b/ in "absurd" [apsyrd]? It is in French listeners' minds. *Journal of Memory and Language, 43*, 618-639.

Hamada, M. & Koda, K. (2008). Influence of first language orthographic experience on second language decoding and word learning. *Language Learning, 58*(1)*,* 1-31.

Hamada, M. & Koda, K. (2010). The role of phonological decoding in second language word-meaning inference. *Applied Linguistics, 31*(4), 513-531.

Harrell, F. E. (2013). rms: Regression modeling strategies (R package version 3.6-3) [software package] Retrieved from http://CRAN.R-project.org/package=rms/.

Harris, M., & Giannouli, V. (1999). Learning to read and spell in Greek: The importance of letter knowledge and morphological awareness. In M. Harris & G. Hatano (Eds.), *Learning to read and write: A cross-linguistic perspective* (pp. 51-70). New York: Cambridge University Press.

Hayes-Harb, R., Nicol, J., & Barker, J. (2010). Learning the phonological forms of new words: Effects of orthographic and auditory input. *Language and Speech, 53*(3)*, 367-381.*

IDS. (1991-2010). COSMAS I/II corpus search, management and analysis system. Mannheim: Institut für Deutsche Sprache. http://www.ids-mannheim.de/cosmas2/.

IDS. (2013). Das Deutsche Referenzkorpus DeReKo. Mannheim: Institut für Deutsche Sprache. http://www.ids-mannheim.de/kl/projekte/korpora/archive.html.

Kellerman, E. & Sharwood Smith, M. (Eds.) (1986). *Crosslinguistic influence in second language acquisition.* New York: Pergamon Press.

Keuleers, E. & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods, 42*(3), 627-633.

Koda, K. (1990). The use of L1 reading strategies in L2 reading: Effects of L1 orthographic structures on L2 phonological recoding strategies. *Studies in Second Language Acquisition, 12*(4), 393-410.

Krashen, S. (1982). *Principles and Practice in Second Language Acquisition*. Oxford, England: Pergamon Press.

Landerl, K., & Thaler, V. (2006). Reading and spelling acquisition and dyslexia in German. In R. M. Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy* (pp. 121-34). Mahwah, NJ: L Erlbaum.

Lemhöfer, K., & Dijkstra, T. (2004). Recognizing cognates and interlingual homographs: Effects of code similarity in language-specific and generalized lexical decision. *Memory and Cognition, 32*(4)*,* 533-550.

Lennes, M. (2009). Save conversation tiers as text file [source code]. Retrieved from http://www.helsinki.fi/~lennes/praat-scripts/.

Lieven, E., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: A usage-based approach. *Journal of Child Language, 30*(2), 333-370.

Lupker, S. J. (1982). The role of phonetic and orthographic similarity in picture-word interference. *Canadian Journal of Psychology 36*(3), 349-367.

Lyytinen, H., Aro, M., Holopainen, L., Leiwo, M., Lyytinen, P., & Tolvanen, A. (2006). Children’s language development and reading acquisition in a highly transparent orthography. In R. M. Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy* (pp. 47-62). Mahwah, NJ: L. Erlbaum.

MacWhinney, B. (1987). The competition model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 249-308). Hillsdale, NJ: Lawrence Erlbaum.

MacWhinney, B. (1992). Transfer and competition in second language learning. *Advances in psychology: Cognitive processes in bilinguals, 83,* 371-390.

MacWhinney, B. & Bates, E. (1989). *The cross-linguistic study of sentence processing*. Cambridge: Cambridge University Press.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language and Hearing Research, 50*, 940-967.

Max Planck Institute for Psycholinguistics. (2001). *WebCelex.* http://www.mpi.nl/world/celex (accessed on 10.02.2014).

Microsoft. (2011). PowerPoint (Version 14.3.5) [computer software].

Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words. *Research on Speech Perception Progress Report*, *10*, 357-376.

Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning.* Cambridge: Cambridge University Press.

Pattamadilok, C., Morais, J., Ventura, P., & Kolinsky, R. (2007). The locus of the orthographic consistency effect in auditory word recognition: Further evidence from French. *Language and Cognitive Processes, 22*(5), 700-726.

Porpodas, C. (2006). Literacy acquisition in Greek: Research review of the role of phonological and cognitive factors. In R. M. Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy* (pp. 189-200). Mahwah, NJ: L. Erlbaum.

R Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing (Version 3.0.2) [computer software]. Vienna, Austria. Retrieved from http://www.R- project.org/.

Rafat, Yasaman (2011). Orthography-induced transfer in the production of novice adult English-speaking learners of Spanish. Unpublished Ph.D. Dissertation. University of Toronto*,* Canada.

Rastle, K., McCormick, S. F., Bayliss, L., & Davis, C. J. (2011). Orthography influences the perception and production of speech. *Journal of Experimental Psychology: Learning Memory, and Cognition, 37*(6), 1588-1594.

Seidenberg, M. S., & Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory*, *5*(6), 546-554.

Seymour, P. (2006). Theoretical framework for beginning reading in different orthographies. In R. M. Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy* (pp. 441-62). Mahwah, NJ: L. Erlbaum.

Showalter, C. (2012). The influence of novel orthographic information on second language word learning: The case of native English speakers learning Arabic. Unpublished MA Thesis. University of Utah, USA.

Showalter, C. E. & Hayes-Harb, R. (2013). Unfamiliar orthographic information and second language word learning: A novel lexicon study. *Second Language Research, 29*(2)*, 185-200.*

Simon, E., Chambless, D., & Alves, U. K. (2010). Understanding the role of orthography in the acquisition of a non-native vowel contrast. *Language Sciences, 32*(3), 380-394.

Stanovich, K. E., & Bauer, D. W. (1978). Experiments on the spelling-to-sound regularity effect in word recognition. *Memory and Cognition*, *6*(4), 410-415.

Steele, J. (2005). Assessing the role of orthographic versus uniquely auditory input in acquiring new L2 segments. Paper presented at 7ème Journées internationales du réseau français de phonologie (RFP2005), Aix-en-Provence, 2-4 June.

Taft, M., Castles, A., Davis, C., Lazendic, G., & Nguyen-Hoan, M. (2008). Automatic activation of orthography in spoken word recognition: Pseudohomograph priming. *Journal of Memory and Language, 58*(2), 366-379.

Tanenhaus, M. K., Flanigan, H. P., Seidenberg, M. S. (1980). Orthographic and phonological activation in auditory and visual word recognition. *Memory & Cognition, 8*(6), 513-520.

Tomasello, M. (2003). *Constructing a Language: A usage-based theory of language acquisition*. Boston: Harvard University Press.

Van der Loo, M. (2013). stringdist: Approximate string matching and string distance functions (R package version 0.7.0) [software package]. Retrieved from http://CRAN.R-project.org/package=stringdist/

Van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory* *and Language, 39*(3)*,* 458-483.

Veivo, O., & Järvikivi, J. (2013). Proficiency modulates early orthographic and phonological processing in L2 spoken word recognition. *Bilingualism: Language and Cognition*, *16*(4), 864-883.

Ventura, P., Morais, J., Pattamadilok, C., & Kolinsky, R. (2004). The locus of the orthographic consistency effect in auditory word recognition. *Language and Cognitive Processes, 19*(1), 57-95.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*, 1–25.

Wickham, H. & Chang, W. (2013). ggplot2: An implementation of the grammar of graphics (R package version 0.9.3.1) [software package]. Retrieved from http://CRAN.R-project.org/package=ggplot2/.

Young-Scholten, M. (1998). Second language syllable simplification: Deviant development or deviant input? In J. Allan and J. Leather (Eds.), *New sounds '97* (pp. 351-360). Amsterdam: University of Amsterdam.

Young-Scholten, M. (2002). Orthographic input in L2 phonological development. In P. Burmeister, T. Piske and A. Rohde (Eds.), *An integrated view of language development: Papers in honour of Henning Wode* (pp. 263-279). Trier: Wissenschaftlicher Verlag Trier.

Young-Scholten, M., Akita, M. and Cross, N. (1999). Focus on form in phonology: Orthographic exposure as a promoter of epenthesis. In P. Robinson and N. O. Jungheim (Eds.), *Pragmatics and pedagogy. Proceedings of the third PacSLRF* (Vol. 2, pp. 227-233). Tokyo: Aoyama Gakuin University.

Ziegler, J. C. & Ferrand, L. (1998). Orthography shapes the perception of speech:

The consistency effect in auditory word recognition. *Psychonomic Bulletin & Review, 5*(4), 683-689.

Ziegler, J. C., Ferrand, L., & Montant, M. (2004). Visual phonology: The effects of orthographic consistency on different auditory word recognition tasks. *Memory & Cognition, 32*(5), 732-741.

Ziegler, J. C. & Goswami., U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 13*(1), 3–29.

Appendix A: Test words with prescribed pronunciations and carrier sentences

Word	Pronunciation	Carrier Sentence
äffelbar	ɛfəlbaɐ	Doch jetzt sind äffelbar spitze Kostenrechner gefragt.
ähnlich	ɛnlɪç	Unsere Reitweisen sind sehr ähnlich.
überlegt	ybɐlekt	Die Schritte müssen überlegt sein.
übernommen	ybɐnɔmən	Nachher habe ich die Mannschaft wieder übernommen.
ühmen	ymən	Ein Bericht ist in den nächsten Tagen zu ühmen.
angepieft	angəpift	Teils hat er selber Hand angepieft.
angezogen	angətsogən	Es hat mich schon immer mehr angezogen als der Tod.
annehmen	annemən	Ich hoffe, dass sie dieses Geschenk annehmen werden.
aufhören	aufhøʀən	Mehr also die Hälfte davon möchte wieder aufhören.[7]
bänkischen	bɛŋkɪʃən	Damals wurde es aus bänkischen Gründen verschoben.
büther	bytɐ	Schulzeit soll nicht büther werden.
bequem	bəkvem	Die macht bequem ihre fünf und dreizig Sachen.
bestögente	bəʃtøgəntə	Es werden bewusst nicht nur bestögente Traumberufe vorgestellt.
besuchen	bəzuxən	Zwei Anlässe konnte ich bisher besuchen.
bibliothek	bibliotek	Heute wird sie in der Bibliothek mithelfen.
dürblich	dʏɐplɪç	Das vermindert die Probleme dürblich auch.
diejenige	dijenɪgə	Die Konstellation von heute muss nicht diejenige von morgen sein.
draußen	dʀausən	Erst gehen wir nach draußen und dann wieder herein.
durften	dʊɐftən	Die Gefangenen durften sich für die Nacht eingraben.
ebelgannt	ebəlgant	Diese müssten dann wieder durch die Gemeinde ebelgannt werden.
ehemaligen	eəmalɪgən	Fast alle ehemaligen Schüler teilen mit uns diese Meinungen.
eherrau	eɐʀau	Die zweite bestand im Aufbau einer Eherrau.
einfach	ainfax	Wir haben einfach konstant gut gespielt.
einrünnung	ainʀʏnʊŋ	An dieser Einrünnung hat sich nichts geändert.
einverstanden	ainfɛɐʃtandən	Meine Frau war nach kurzer Bedenkzeit einverstanden.
entvorten	ɛntfɔɐtən	Dabei sind auch viele Bilder entvorten.
erstalte	ɛɐʃtaltə	Doch die zweite Hälfte erstalte klar den Gästen aus Basel.
ervältet	ɛɐfɛltət	Das Urteil wird für Freitag ervältet.
ervotten	ɛɐfɔtən	Solche Erfahrungen ervotten in keiner Statistik.
ewigkeit	evɪçkait	Und das ist für die Ewigkeit.
füßen	fysən	Mit den Füßen kam ich auch ganz gut klar.
faßmenden	fasmɛndən	Das Festival wird nächstes Jahr wieder faßmenden.
frohjarm	fʀojaɐm	Eltern sollen die Frohjarm in Ruhe besuchen können.
geboren	gəboʀən	Das Gewitter ist geboren.

[7] This sentence contains the spelling error noted in section 2.1.2 that resulted in a real word. Without the error, the sentence would be ‘*Mehr **als** die Hälfte davon möchte wieder aufhören*’.

geduldig	gədʊldɪç	Auf das Geld warten sie geduldig.
gesehen	gəzeən	Ich habe nun auch gesehen, dass das nicht allgemeiner Konsens ist.
gethogt	gətokt	Den hat nämlich die Strömung auf den Weg gethogt.
gewöhnen	gəvønən	An den werde ich mich nie gewöhnen.
größeren	gʀøsəʀən	Supermärkte könne es nur in größeren Städten geben.
iberkähmen	ibɐkɛmən	Sie hat auch teilweise die Arbeiter iberkähmen.
ihrem	iʀəm	Alle andern sind mit ihrem bisherigen Angebot aufgenommen.
irgendwann	ɪɐgəntvan	Doch auch diese Serie geht irgendwann zu Ende.
jemals	jemals	Er erinnere sich nicht, ihr jemals begegnet zu sein.
jemanden	jemandən	Wir müssten schauen, ob wir jemanden finden.
johren	joʀən	Die Kinder johren die Texte schnell.
kaffee	kafe	Auch Kuchen und Kaffee standen in reicher Fülle bereit.
klöfen	kløfən	Gegen den Wind hatten auch die Läufer zu klöfen.
klappen	klapən	In zwei Wochen wird alles klappen.
kloftig	klɔftɪç	Sie haben uns kloftig unterstützt.
lorsches	lɔɐʃəs	Der Donnerstag entfällt bis auf lorsches.
männer	mɛnɐ	Die jungen Männer rennen davon.
müssen	mʏsən	Zu oft hat sie sie schon hören müssen.
mauchter	mauxtɐ	Wir können neue Themen mauchter aufgreifen.
meeten	metən	Ich kenne sie, und sie meeten mir.
methode	metodə	Die Forscher haben ihre Methode bereits vor Gericht anwenden können.
minuten	minutən	Bis dahin wird der Pilot lange Minuten erleben.
pates	patəs	Wir werden ihm ein pates Andenken bewahren.
pesen	pezən	Ich hoffe, dass wir das Gewerbe pesen können.
quäschen	kvɛʃən	Baden, um gesund zu quäschen.
qualität	kvalitɛt	Sie verbessert die Qualität der Betreuung.
quantam	kvantam	Sie fragen sich, was hier daran so quantam sei.
quantol	kvantol	Wer nichts kaufen will, kann ein quantol plaudern.
quelle	kvɛlə	An der Quelle liege die Wahrheit.
quirtal	kvɪɐtal	Profi zu sein, wäre schon einmal quirtal.
reisen	ʀaizən	Ich bin gewohnt, allein zu reisen.
rihren	ʀiʀən	Zwei junge Männer rihren sie.
schützen	ʃʏtsən	Arme können die Umwelt nicht schützen.
schnaden	ʃnadən	Hinter den Konzerten schnaden Menschen.
schnellen	ʃnɛlən	Das steht quer zum schnellen Blick in der heutigen Zeit.
schwießt	ʃvist	Bei Kaffee und Kuchen schwießt der traditionelle Anlass aus.
schwiller	ʃvɪlɐ	Dieser Aufwand würde schwiller liegen als in der Vergangenheit.
seelen	zelən	Oft brennen auch Kerzen für die armen Seelen.
seltsamer	zɛltzamɐ	Es sei ein sehr seltsamer Zustand gewesen.
sequenz	zekvɛnts	Ein Lied beschreibt nur eine Sequenz aus meinem Leben.

spöhen	ʃpøən	Die Spieler spöhen auf dem Feld ihren wahren Charakter.
spanzen	ʃpantsən	Wer mehr verdienen will, muss dan mehr spanzen.[8]
sprechen	ʃpʀɛçən	Die Damen sprechen über Strickmoden.
staugt	ʃtaukt	Der Charakter staugt sich auf dem Spielfeld.
stießliche	ʃtislɪçə	Das hätte komplizierte stießliche Probleme mit sich gebracht.
türigheit	tyʀɪçhait	Aber wir werden die Mannschaft mit Türigheit verjüngen.
tütig	tytɪç	Sie ist auch tütig bei ähnlichen Vorfällen ein Mittel.
tacht	taxt	Gemeinsam, so wurde immer wieder tacht, sei man stark.
umziehen	ʊmtsiən	Trotzdem wollen die beiden nicht umziehen.
unterschied	ʊntɐʃit	Das ist der Unterschied zwischen ihnen und uns.
verbessern	fɛɐbɛsɐn	Zwei Dinge hätte man vielleicht noch verbessern können.
vermätzeren	fɛɐmɛtsəʀən	Sie dient nur ganz vermätzeren Anlässen.
vermutlich	fɛɐmutlɪç	Die Opfer sind vermutlich Obdachlose.
vielleicht	filaiçt	Ein heisser Sommer macht es vielleicht möglich.
vormee	fɔɐme	Jetzt ist der Zeitpunkt gekommen, die Vormee zu verwirklichen.
vorstellen	foʁʃtɛlən	Darunter kann man sich etwas vorstellen.
wütten	vʏtən	Wir müssen da am Ball wütten.
wertvolle	veɐtfɔlə	Immerhin wurde eine vertvolle Diskussion angeregt.[9]
weseln	vezəln	Wir nehmen ein Thema und weseln es dann weiter.
wichtig	vɪçtɪç	Das Zusammenspiel sei wichtig.
wurigen	vuʀigən	Musik nimmt in seinem Leben einen wurigen Stellenwert ein.
zericheln	tseʀɪçəln	Die haben mich fast zericheln.
zusammen	tsuzamən	Die Welt rückt näher zusammen.

---

[8] This sentence contains the spelling error noted in section 2.1.2 that resulted in a pseudoword. Without the error, the sentence would be '*Wer mehr verdienen will, muss* ***dann*** *mehr spanzen*'.

[9] This sentence contains the spelling error noted in section 2.1.2 that occurred on the test word. '***v****ertvolle'* should be '***w****ertvolle*'.

Appendix B: Post-experiment questionnaire

Instructions

Please rate your familiarity with the German words on this questionnaire. Use the scale below as a guide.

1	2	3	4	5	6	7
I have never heard nor seen this word			I recognize this word, but I don't know its meaning			I am familiar with this word and know its meaning

1. _____ reisen
2. _____ durften
3. _____ sprechen
4. _____ bequem
5. _____ Sequenz
6. _____ zusammen
7. _____ vielleicht
8. _____ jemals
9. _____ gesehen
10. _____ überlegt
11. _____ müssen
12. _____ größeren
13. _____ einfach
14. _____ vermutlich
15. _____ irgendwann
16. _____ Minuten
17. _____ aufhören
18. _____ Kaffee
19. _____ Bibliothek
20. _____ seltsamer
21. _____ geduldig
22. _____ klappen
23. _____ Männer
24. _____ jemanden
25. _____ schnellen
26. _____ geboren
27. _____ verbessern
28. _____ vorstellen
29. _____ Qualität
30. _____ Quelle
31. _____ angezogen
32. _____ einverstanden
33. _____ diejenigen
34. _____ umziehen
35. _____ Füßen
36. _____ schützen
37. _____ draußen
38. _____ besuchen
39. _____ wichtig
40. _____ wertvolle
41. _____ ihrem
42 _____ gewöhnen
43. _____ Seelen
44. _____ Methode
45. _____ übernommen
46. _____ Ewigkeit
47. _____ annehmen
48. _____ ähnlich
49. _____ ehemaligen
50. _____ Unterschied

Please rate the following words for how much they look like possible English words. Use the following scale as a guide

1	2	3	4	5	6	7
This looks nothing like English			This looks like it could be a word in English			This looks exactly like a word in English

1. ______ reisen
2. ______ durften
3. ______ sprechen
4. ______ Bequem
5. ______ Sequenz
6. ______ zusammen
7. ______ vielleicht
8. ______ jemals
9. ______ gesehen
10. ______ überlegt
11. ______ müssen
12. ______ größeren
13. ______ einfach
14. ______ vermutlich
15. ______ irgendwann
16. ______ Minuten
17. ______ aufhören
18. ______ Kaffee
19. ______ Bibliothek
20. ______ seltsamer
21. ______ geduldig
22. ______ klappen
23. ______ Männer
24. ______ jemanden
25. ______ schnellen
26. ______ geboren
27. ______ verbessern
28. ______ vorstellen
29. ______ Qualität
30. ______ Quelle
31. ______ angezogen
32. ______ einverstanden
33. ______ diejenige
34. ______ umziehen
35. ______ Füßen
36. ______ schützen
37. ______ draußen
38. ______ besuchen
39. ______ wichtig
40. ______ wertvolle
41. ______ ihrem
42. ______ gewöhnen
43. ______ Seelen
44. ______ Methode
45. ______ übernommen
46. ______ Ewigkeit
47. ______ annehmen
48. ______ ähnlich
49. ______ ehemaligen
50. ______ Unterschied
51. ______ rossen
52. ______ lorsches
53. ______ staugt
54. ______ quäschen
55. ______ quantol
56. ______ zündern
57. ______ velten
58. ______ johren
59. ______ Eherrau
60. ______ ühmen
61. ______ Einrünnung
62. ______ stießliche
63. ______ tacht
64. ______ dürblich
65. ______ schwießt
66. ______ rihren
67. ______ klöfen
68. ______ vormee
69. ______ gethogt
70. ______ schwiller
71. ______ kloftig
72. ______ pates
73. ______ ervältet
74. ______ pesen

75. _____ schnaden
76. _____ wurigen
77. _____ entvorten
78. _____ erstalte
79. _____ quirtal
80. _____ quantam
81. _____ spanzen
82. _____ ervotten
83. _____ Frohjarm
84. _____ spöhen
85. _____ tütig
86. _____ wütten
87. _____ faßmenden
88. _____ mauchter
89. _____ zericheln
90. _____ weseln
91. _____ iberkähmen
92. _____ bestögente
93. _____ meeten
94. _____ büther
95. _____ vermätzeren
96. _____ Türigheit
97. _____ angepieft
98. _____ äffelbar
99. _____ ebelgannt
100. _____ bänkischen