Spoken and written production of inflectional morphology among L1 Mandarin speakers of English

Spoken and written production of inflectional morphology among L1 Mandarin speakers of English.

Qingyuan Gardner (corresponding author)

Department of Psychology, University of Edinburgh, UK, EH8 9JZ (q.gardner@ed.ac.uk)

Holly P. Branigan

Department of Psychology, University of Edinburgh, UK, EH8 9JZ (holly.branigan@ed.ac.uk)

Vasiliki Chondrogianni

Department of Linguistics and English Language, University of Edinburgh, UK, EH8 9JZ (v.chondrogianni@ed.ac.uk)

Abstract

Second language (L2) speakers frequently make errors when producing L2 inflectional morphology, but the underlying causes of errors remain unclear. We report three experiments investigating how such errors might arise within the language production system, focusing on L2 speakers whose L1 does not use inflectional morphology to indicate temporal properties of events. L1 Mandarin and L1 English participants produced spoken (Experiments 1 and 2) and written (Experiment 3) event descriptions in English involving different temporal contexts. In both spoken and written production, L1 Mandarin participants' production of present (3rd person singular *-s*) and past (*-ed*) inflections was sensitive to L2 temporal cues, but error-prone, with higher omission rates for featurally complex inflections. These results suggest L2 speakers can acquire representations of L1-absent features, but do not consistently activate these features and their corresponding morphological forms during grammatical and morphophonological encoding. Moreover, articulation cannot solely account for all L2 inflectional errors.

Highlights

- L2 speakers can successfully acquire new conceptual distinctions for production
- They can also acquire L1-absent lemma level and morphophonological representations
- These representations may be inconsistently activated during online production
- L2 inflectional accuracy is affected by inflections' feature complexity
- Articulation contributes to but is not the main cause of inflectional errors

Key words

Second language production

Second language acquisition

Morphological processing

Featural complexity

Introduction

Second language (L2) speakers often make errors when producing inflectional markings (henceforth *inflections*) in their L2. For example, L2 English speakers frequently omit the past tense inflection *-ed* when the grammatical context demands it, e.g., **Yesterday the chef shout at the waiter in the restaurant* (where * indicates ungrammaticality). Although there is abundant evidence for erroneous inflectional production by L2 speakers from different native language (L1) backgrounds, there is little agreement over the causes of such inconsistencies (Goad, White, & Steele, 2003; Hawkins & Chan, 1997; Lardiere, 2008; Poulisse, 1999; Prévost & White, 2000), and, in particular, little consideration of how inflectional errors might be accounted for within psycholinguistic models of language production. For instance, do they reflect L2 speakers' failure to acquire conceptual distinctions that are absent in their

L1, inability to acquire and activate L2 syntactic dependencies, inability to represent and/or appropriately activate and retrieve grammatical features, inability to represent and/or consistently activate and retrieve morphological forms, or difficulties in articulating inflections? In this paper, we investigate the locus of erroneous inflections in L2 language production by focusing on the spoken and written production of English tense inflections (i.e., 3rd person singular *-s* [henceforth 3SG *-s*] and past tense *-ed*, as in *shouts* and *shouted*) in L2 learners whose L1 (Mandarin) does not overtly mark for tense morphology.

Many previous studies on inflectional production have found that L1 Mandarin speakers are particularly prone to inflectional errors in L2 English spoken production, especially in comparison with L2 English speakers whose L1s have tense morphology (e.g. Dutch, see Poulisse, 1999). In a series of longitudinal studies, Lardiere (1998a; 1998b; 2000; 2003) found that Patty, an adult from L1 Mandarin-Hokkien background who had been living in the US for more than 10 years, showed only 5.8% accurate regular past tense marking in her spoken production even after prolonged L2 immersion. Similarly, a picture-description study testing L2 English production in upper-intermediate to advanced adult L1 Mandarin speakers of English also showed inflectional production close to chance level (57 % for past *-ed*) or below (28% for 3SG *-s*) after six months of L2 immersion (Goad et al., 2003). Converging evidence from other production tasks confirms the same pattern, and moreover reveals that L1 Mandarin speakers' performance is poorer than L2 English speakers from other L1 backgrounds (Bayley, 1996; Hawkins & Liszka, 2003).

One important factor that might play a role in L1 Mandarin speakers' poor performance on English inflectional production is differences in the two languages' temporal properties: Whereas Mandarin is a non-inflectional language that does not overtly mark for tense on the verb and uses aspectual markers (e.g. *le* for perfective aspect, see 1a) with temporal adverbials to express temporal information (Smith, 1991), English is an inflectional language

that uses obligatory tense and aspectual morphemes on the verb as well as temporal adverbials (see 1b).

1a) *zuo*2 *tian*1 *ta*1 *kan*4 *le*0 *wang*3*qiu*2 *bi*3*sai*4

Yesterday she watch PERFECTIVE ASPECT tennis match

1b) *yesterday she watch -ed (a) tennis match*

yesterday she watch PAST / PERFECTIVE ASPECT tennis match

'Yesterday she watched a tennis match'

Therefore, for L1 Mandarin speakers of English to accurately produce temporal inflections, they must not only be able to conceptualize tense distinctions (e.g., present vs. past) and represent the appropriate inflections that mark these distinctions (e.g., *-s* vs. *-ed*), but also use these distinctions during on-line processing to produce the correct inflections in the appropriate contexts. That is, successful inflectional production implicates factors relating both to representation (competence) and processing (performance).

Linguistic Theories of L2 inflectional production errors

What factors might underlie L1 Mandarin speakers' poor inflectional accuracy in English? Previous research situated within theoretical linguistic frameworks has proposed several possible sources for L2 inflectional errors, implicating both representational and processing deficits. Some theories locate difficulties in inflectional production in specific representational deficits for linguistic features underlying the relevant morphemes (e.g. number, person, tense etc.). For example, L2 speakers might be unable to acquire new L2 featural representations after the critical period if these features are absent in the speaker's L1 (*Failed Functional Feature Hypothesis*; Hawkins & Chan, 1997). Thus L1 Mandarin late learners of English, with no tense feature in their L1, would in principle not be able to

develop the necessary representations for tense features, resulting in absolute omission of inflections (though this is rarely observed in empirical data).

Other representational deficit accounts propose that inflectional production is affected by prosodic constraints, and specifically that L2 inflectional omission (one type of inflectional error) results from L2 speakers' use of L1 prosodic features when producing their L2 (*Prosodic Transfer Hypothesis*; Goad et al., 2003). If the L2 speakers' L1 does not permit certain prosodic structures (e.g. Mandarin does not permit a consonant to be adjoined to a phonological word, attaching [s] *-s* to [ʃaʊt] *shout* to form [ʃaʊts] *shout-s*), L2 speakers would not have the appropriate prosodic representations necessary for inflectional production, and consequently fail to produce the corresponding inflections during L2 production. Accordingly, Goad and White (2006) showed that L1 Mandarin speakers found adjunctions outside the prosodic word (e.g. [hɛlp] *help* becoming [hɛlpt] *helped*) more difficult to produce compared with phonological operations inside the prosodic word (e.g. [drɪŋk] *drink* becoming [drʌŋk] *drunk*). Consistent with this, Lardiere's (2003) Mandarin-Hokkien speaker similarly showed consistent difficulty with word-final consonant clusters on English regular verbs. Moreover, though L1 Mandarin speakers have been found to omit -t/-d phonemes in regular past tense contexts, they also omitted -t/-d phonemes in non-tense contexts (Bayley, 1996; Hawkins & Liszka, 2003), suggesting a key prosodic element to inflectional production. Note that these accounts predict that prosodic constraints should result in inflectional omission primarily in the spoken rather than the written modality, given that phonological representations are most strongly implicated in spoken production (Goad et al., 2003).

Other accounts postulate that L2 inflectional errors are not the result of representational deficits, but rather of inconsistent retrieval (or realization) of L2 morphological forms (*Missing Surface Inflection Hypothesis, MSIH*; Prévost & White, 2000). This account appears to be more in keeping with existing evidence that L2 speakers tend to produce inflections

inconsistently but not randomly (termed 'optional production' in theoretical linguistic accounts) rather than consistently omitting them ('absolute omission'). Cross-linguistically, inconsistent retrieval has been linked to informational complexity (*Featural Complexity Theory*; Hawkins, 2007): Inflections that encode more complex information are more difficult for L2 speakers to produce accurately (e.g., 3SG *-s,* which codes for **PERSON**, **SUBJECT NUMBER** and **TENSE**) are more difficult for L2 speakers to produce accurately than inflections that encode less information (e.g., past *-ed* which codes only for **TENSE**). In support of this, Turkish-English sequential bilingual children (L2 English) show particularly high error rates for 3SG *-s*, compared with past *-ed* (Chondrogianni & Marinis, 2012). Critically, although they produced inflections inconsistently, they were sensitive to the ungrammaticality of inflectional omissions, indicating intact rather than deficient L2 syntactic representations.

These linguistic accounts provide plausible proposals for why L2 speakers might produce inflectional errors, but crucially, they are not embedded within psycholinguistic models of language processing. Therefore, they do not elucidate the specific representational and processing deficits that might lead to inflectional errors in L2 speakers' language production.

Morphological processing in language production

Our concern in this study was to consider how L2 speakers' erroneous inflectional production can be explained within psycholinguistic models of language production. Current modular models standardly assume that L1 production involves stages of constructing a preverbal message (*conceptualization*); activating lexical representations, assigning grammatical functions/syntactic structure, retrieving inflectional morphemes, activating phonological representations, forming phonological words and associated phonetic plans (*grammatical* and *phonological encoding*); and finally executing phonetic articulatory

gestures (*articulation*; Bock, 1982; Bock & Levelt, 1994; Garrett, 1975; Levelt, 1989, 2001; Levelt, Roelofs, & Meyer, 1999; see Dell, 1986, for an alternative computational model assuming interactive activation between levels). Adaptations of this model, assuming the same basic architecture, have been proposed for bilingual language production (de Bot, 1992; 2003; de Bot & Schreuder, 1993). In these adapted models, L1 and L2 share conceptualization and articulatory processes but have separate subsystems for grammatical encoding and lexical access.

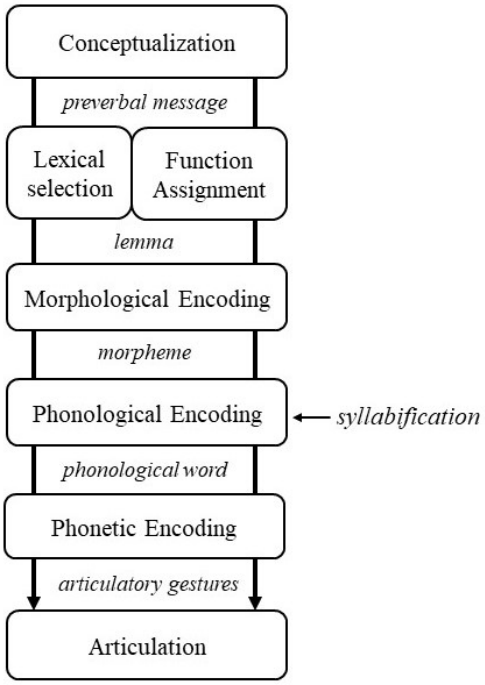

Figure 1. Theoretical outline of lexical access in a spreading activation network (adapted from Bock & Levelt, 1994, and Levelt et al., 1999)

To exemplify the relevant processes, consider an L1 English speaker describing an event in which a chef is shouting at a waiter in a restaurant. During conceptualization, the speaker constructs a preverbal message that contains not only concepts such as *chef, waiter, restaurant* and *shout*, but also semantic relations such as the concepts of *at* and *in,* and crucially, temporal properties of the event. This message is assumed not to be language-

specific, but nevertheless encodes only information strictly relevant to the language of the intended utterance (*microplanning*; de Bot, 1992; Levelt, 1989). Hence the L1 English speaker would code information about the event that included tense and aspect. Note that this is consistent with Slobin's (1987; 1996) 'thinking-for-speaking' principle where the choice of language has been found to shape how a message is conceptualized.

In the following stage, the speaker activates the relevant lexical representations (lemmas; e.g. syntactic component relating to *chef / waiter / shout / restaurant*) with the associated diacritic features such as **NUMBER**, **TENSE** etc. She also determines relevant grammatical functions or syntactic relations, e.g., subject, by consulting the preverbal message. Activation of features at the lemma level underlies subsequent morphological processing of the relevant inflections at the morpheme level (e.g., activation of the **PERFECTIVE ASPECT** and **PRESENT TENSE** features associated with the verb lemma, together with **3RD PERSON** and **SINGULAR** features associated with the subject and verb lemmas – the latter via transmission of these features from the former [Eberhard, Cutting, & Bock, 2005] – underlie subsequent processing of *-s*). The speaker subsequently retrieves relevant phonological representations, including phonemes, syllable and stress information, and carries out syllabification to form the phonological word. The phonological word then undergoes phonetic encoding, where articulatory gestures are planned. Finally, during articulation, the speaker executes the relevant phonetic articulatory gestures to form the sounds for *chef, waiter, restaurant* etc.

Typically, activation flows smoothly from one stage to another, resulting in successful production, but on rare occasions breakdowns in transmitting activation within or between levels can result in a speech error, e.g., **The chef shout at the waiter in the restaurant* (see (Budd, Hanley, & Griffiths, 2011; Dell, 1986; Dell, Chang, & Griffin, 1999; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Foygel & Dell, 2000; Stemberger, 1998). In the model of language production in which we site our work, activation is not binary, but rather graded,

with transmission from one level to the next occurring once a criterial activation level has been reached (see Levelt et al., 1999).

Within this model, we can identify a range of ways in which inflectional errors might in principle arise during L1 Mandarin speakers' production of L2 English. First, errors might arise from representational or processing deficits during conceptualization. If L2 speakers are unable to encode conceptual distinctions that do not exist in their L1, L1 Mandarin speakers (with an L1 that does not grammaticalize tense) could in principle fail to encode event-external information (in our example, how the act of *shouting* as a whole relates to the time of speech) in the preverbal message. As the preverbal message representation drives subsequent linguistic formulation, L1 Mandarin speakers would consequently fail to produce appropriate tense inflections (i.e. omitting *-s* entirely). If L2 speakers are able to represent conceptual distinctions that do not exist in their L1, but experience difficulty in processing such conceptual distinctions, this would result in a tendency to produce tense inflections inconsistently (showing optionality for *-s*; i.e., sometimes correctly but sometimes incorrectly).

Alternatively, errors might arise during formulation. With respect to grammatical encoding, there could be a representational deficit at the lemma level for the relevant diacritic features (consistent with Hawkins and Chan's [1997] account). If L2 speakers (from non-inflectional L1s) can make relevant conceptual distinctions but do not have corresponding diacritic feature representations (as Mandarin does not encode these features), they would not be able to encode the temporal features necessary for subsequent morphological encoding (in our example, the **PRESENT TENSE** feature underlying subsequent processing of *-s*) and so would omit inflections (in English). If L2 speakers represent these diacritic features but experience difficulty in activating them appropriately, this would lead to inconsistent production instead.

There might also be a deficit regarding syntactic dependencies (e.g., subject-verb agreement). If L2 speakers do not have knowledge of these relations (i.e., a representational deficit), they would never activate the appropriate feature representations (e.g., activating the values **3RD** and **SINGULAR** for a verb lemma's **PERSON** and **NUMBER** diacritic features respectively, following a 3rd person singular subject), resulting in absolute omission of inflections. If they instead had a processing deficit that resulted in failure to accurately transmit number information from the subject to the verb (Eberhard et al., 2005), they would produce inflections inconsistently.

At the morphological level, L2 speakers might have representational or processing deficit with regard to (inflectional) morphemes. L2 speakers may successfully acquire and process all relevant morphological feature representations and syntactic dependencies at previous stages yet fail to acquire the appropriate inflectional morphemes that express this relationship. Alternatively, they might have a processing deficit in transmitting activation from morphological feature representations to corresponding (inflectional) morphemes. This would lead L2 speakers to produce inflections inconsistently (consistent with Prévost and White's (2000) Missing Surface Inflection Hypothesis, which claimed difficulties in realizing surface form). In our example, an L2 speaker might have the conceptual distinction of tense, the relevant diacritic feature representations (i.e. **PRESENT TENSE, SINGULAR**) and the appropriate syntactic dependencies (i.e. **3RD PERSON**), but still fail to produce the correct inflectional morphology on some occasions because she could not effectively activate and retrieve the 3SG -*s* inflection.

At the phonological level, L2 speakers might have representational or processing deficits regarding phonemes or prosodic frames necessary for inflectional production. As phonological structures differ across languages, particularly between Mandarin and English (see Roelofs, 2015), L2 speakers might fail to establish the correct phoneme representations

(e.g. [s]) or correctly assemble the relevant phoneme sequences according L2 prosodic frames during phonological encoding to produce an inflected verb (e.g. unaspirated [ts] in [ʃaʊts] as in *shout-s*) . This is partially consistent with Goad et al.'s (2003) account, which claimed that L2 speakers have fundamental difficulties performing phonological adjunctions which are illegal in the L1.

Finally, errors might have an articulatory source: L2 speakers' articulatory gestures may be limited to executing phoneme sequences and associated phonetic plans from their L1, so that they do not acquire additional articulatory gestures for L2 phoneme sequences. This would give rise to consistent omission of specific phoneme sequences in the spoken modality (e.g., failure to articulate 3SG -*s* in [ts] in our example). This would be consistent with Lardiere's (2003) finding of a discrepancy between Patty's written versus oral accuracy in past tense inflection (78% vs. 5.8%).

To summarize, a psycholinguistic model of language production offers several potential loci for L2 inflectional errors in production: Such errors might in principle occur because of representational or processing deficits, and these deficits might be associated with conceptualization, formulation (at the lemma, morphological, and/or phonological level), and/or articulation. Of course, errors might in principle have more than one source, and processing deficits at more than one level of production might act in concert to increase the probability of retrieval failure. This may be exacerbated in spoken production when additional phonetic encoding takes place for overt articulation.

The current study

To investigate whether L2 inflectional errors in production might arise from representational versus processing deficits, and to identify the level(s) at which such deficits

might occur, we now report three experiments that investigated the production of L2 temporal morphology in adult L1 Mandarin speakers of L2 English (and a control group of L1 English speakers). Spoken (Experiments 1 and 2) and written (Experiment 3) responses were elicited using a description paradigm, in which participants produced (under a time limit) descriptions of action scenes, using temporal cues (calendar pictures indicating either Present Habitual or Past temporal contexts), regular verbs (e.g. *shout*) and pictures of people, objects and locations (e.g., *Every day the chef shouts at the waiter in the restaurant*). We analyzed participants' production of inflections (3SG -*s* & past -*ed*) with respect to inflectional accuracy, inflectional type (3SG -*s* and past -*ed*) and inflectional omission.

We investigated two broad possibilities for why L1 Mandarin speakers of L2 English might make errors when producing (temporal) inflectional morphology. Specifically, it is unclear whether the source of error lies with representational deficits or processing breakdowns, and at which stage(s) of language production (see Table 1).

Table 1. *Predictions of representational and processing deficits at multiple stages / interface of inflectional production*

Stage / Interface	Representational	Processing
Stage 1 / Conceptualization	a) absolute omission across temporal contexts.	b) inconsistent 'optional' production across temporal contexts.
Stage 2 / Lemma (diacritic features)	a) absolute omission across temporal contexts.	b) inconsistent production. Accuracy inverse to number of features.
Stage 3 / Lemma (syntactic dependencies)	a) random production across syntactic contexts.	b) inconsistent production; systematic and sensitive to syntactic contexts. Accuracy inverse to number of features.
Stage 4 / Morphological Encoding	a) absolute omission across temporal contexts.	b) inconsistent production; systematic and sensitive to temporal contexts. Accuracy inverse to number of features.
Stage 5 / Phonological / Phonetic Encoding	a) absolute omission for L1-impermissible sequences.	b) inconsistent production; variable depending on phonological context

Stage 6 / Articulation	a) Accuracy: written > spoken; persistent errors across modalities.	b) Accuracy: written > spoken; few or no errors in written modality.

Representational deficits might be implicated at multiple stages of language production (Stage 1a). Firstly, if L2 speakers cannot adaptively conceptualize information relevant to L2 morphological production when the relevant conceptual distinction is not grammaticalized in their L1, and instead use semantic alternatives similar to their L1 (against a strict 'thinking-for-speaking' account), they should not produce the relevant inflections under any circumstances. This account predicts that L1 Mandarin speakers of L2 English should omit 3SG *-s* and past *-ed* inflections entirely (i.e., absolute omission) across temporal contexts (i.e., whether in a Present Habitual context or a Past context).

Secondly, if L2 speakers do not have representations for relevant diacritic features at the lemma level, they should also be unable to produce inflections associated with those features (i.e. absolute omission; Stage 2a). This account predicts that L1 Mandarin speakers should show significantly poorer performance than L1 English speakers across the board (i.e., they would fail to produce both 3SG *-s* and past *-ed* inflections in the appropriate contexts), and that they would do so to the same extent for both inflection types, if at all. Thirdly, if L2 speakers do not have knowledge of (i.e. fail to represent) the relevant syntactic dependencies (i.e., subject-verb agreement), then their inflectional production should be random across syntactic contexts but systematic across temporal contexts (Stage 3a). This account predicts that in the case of English, L1 Mandarin speakers' 3SG *-s* production would be random across singular and plural subjects (syntactic context), but –more accurate in Present Habitual than in Past contexts (temporal context). Lastly, if L2 speakers do not have knowledge of inflectional forms at the morphological level, they should also be unable to produce inflections (i.e. absolute omission; Stage 4a). However, this deficit may be selective for some

inflections but not others. This account predicts that L2 Mandarin speakers should fail to produce 3SG -*s* and past -*ed* in all contexts.

Processing breakdowns might also occur at multiple stages during language production. Firstly, L2 speakers may represent the relevant diacritic features at the lemma level but be unable to activate and integrate them consistently in relation to the verb (e.g., due to inappropriate weights between node connections; Stage 2b). This account predicts inconsistent inflectional production in L1 Mandarin speakers of L2 English which is reflective of the number (or complexity) of diacritic features contained within the inflections. In other words, they should make more errors for inflections requiring activation of both **SUBJECT NUMBER** and **TENSE** features (i.e., 3SG -*s*) than inflections involving only the **TENSE** feature (i.e., past -*ed*). Secondly, L2 speakers may represent the relevant syntactic dependencies, but be unable to activate them consistently under the appropriate syntactic contexts (Eberhard et al., 2005; Stage 3b). If the temporal feature is activated successfully from the previous stages, L1 Mandarin speakers should be more likely to produce both inflections under the correct temporal contexts. At the same time, failure to transmit number will result in poor performance for 3SG -*s* more than past –*ed*, i.e. an effect of feature complexity. If temporal features are not activated for overt marking as proposed previously, L1 Mandarin speakers should perform poorly but should not produce 3SG -*s* less accurately than past -*ed.*

Thirdly, L2 speakers might have appropriate conceptual and lemma level representations and activations as well as knowledge of the relevant syntactic dependencies, but might still experience difficulties activating and retrieving inflectional morphemes (Stage 4b), specifically those which require accurate transmission from more than one feature at the lemma level to the morphological level. This account predicts that L1 Mandarin speakers would be less likely to accurately produce 3SG -*s* than past -*ed* due to number of features, but

should crucially show systematic sensitivity to temporal context for both inflections. That is, they would be more likely to produce 3SG *-s* inflections in Present Habitual contexts than in Past contexts, and more likely to produce past *-ed* inflections in Past contexts than in Present Habitual contexts.

Finally, if articulation difficulties contribute to L2 inflectional errors, L1 Mandarin speakers should be significantly more accurate in written production (which does not involve overt articulation) compared with spoken production (which does involve overt articulation; Stage 6). If articulation is the primary source of such errors (i.e., speakers do not have other representational and processing difficulties during earlier stages of production), L1 Mandarin speakers would produce errors in spoken but not in written production (Stage 6b). If, however, articulatory difficulties only exacerbate other representational and processing sources of error at earlier stages, then L1 Mandarin speakers would produce similar patterns of error in both spoken and written production but the error rate would be higher in spoken than in written production (Stage 6a).

Experiment 1

Methods

Participants

13 native Mandarin (L1 Mandarin) speakers of English aged 19-25 (M=22.46, SD=1.32) and 17 monolingual native English (L1 English) speakers aged 21-33 (M=25.12, SD=3.08) from the University of Edinburgh participated in Experiment 1 and provided valid data; an additional nine participants were also tested but their data were subsequently excluded for several reasons (see below for details). The L1 Mandarin group (i.e., L2 English) consisted of

late learners of English who only had regular exposure to English after the age of eight (exclusion criteria at age five). The monolingual English control group (L1 English) consisted of native English speakers who were not exposed to any other languages before the age of five. The L1 Mandarin participants had achieved an overall score of at least 6.5 on the International English Language Testing System (IELTS, assessing speaking, listening, reading and writing) within the last two years; all were within 24 months of their first arrival in the UK[1].

Materials

For the scene description task, nine transitive experimental verbs with alveolar consonant endings were chosen, eliciting phonologically salient inflectional endings in the past temporal context (see Appendix B). In addition, 36 scenes depicting these transitive actions (four per verb) were created as PNG image files for display on a 1024 x 768 pixels computer screen (see Figure 2 for example). Each scene contained four clip-art items: a calendar image depicting the temporal context of the action (*every day, yesterday*), and three images depicting the entities taking part in the action (an agent, a patient and an instrument or location). The calendar was placed top-center and the three action images were placed below from left to right, congruent with the direction of reading. Nine additional transitive and intransitive filler verbs were chosen, and 36 additional filler scenes were created (Appendix B). 96 entities (people, objects, animals, location etc.) were used multiple times to create 72 action scenes. Singular and plural subjects were counterbalanced across both temporal contexts for each verb. A vocabulary list and a pictorial legend were also provided to familiarize participants with items the scene description task.

[1] See Appendix A for additional information on Mandarin participants' language background.

Figure 2. Experiment 1: Example of trial image from the scene description task, including a temporal cue (calendar image) and entities in the action (chef, waiter, restaurant).

Design

This experiment used a 2 x 2 x 2 mixed design with Subject Number (Singular vs. Plural), Temporal Context (Present Habitual vs. Past) as within-subject variables, and Group (L1 Mandarin vs. L1 English) as a between-subject variable. The experimental design was identical in Experiments 2 and 3.

Procedure

At the beginning of the session, all participants provided demographic details. The Mandarin group also provided information about the history of their L2 acquisition, L2 proficiency, and current L2 usage. Subsequently all participants completed the scene description task.

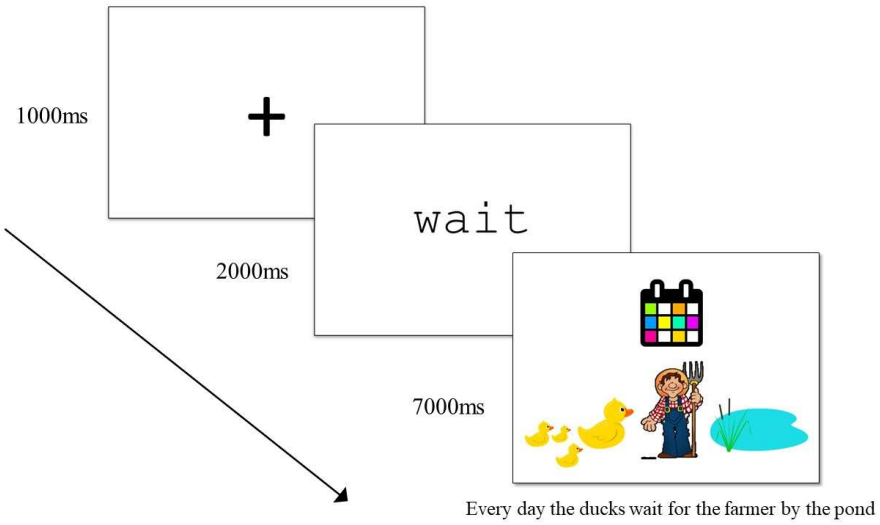

Figure 3. Experiment 1: Three-step trial procedure for the scene description task, including the presentation of fixation (1000ms), target verb (2000ms), and trial image (7000ms).

Before the scene description task, the experimenter explained the interpretation of the calendars to the participants, i.e., that a multi-colored calendar represented that the action in the scene took place habitually (*every day*) and a red-yellow calendar represented that the action in the scene was completed once in the past (*yesterday*). Participants were given further examples of the trial procedure on paper, in which temporal adverbials (*every day* or *yesterday*) appeared at the beginning of each sentence. However, participants were not told explicitly that temporal adverbials were obligatory in their description. Participants then studied the vocabulary list. If they did not understand any concepts, the concepts were explained first in English, and then – if still unclear – in Mandarin.

Participants then completed the scene description task on a computer. Scenes were presented using E-Prime (Version 2.0; Schneider, Eschman & Zuccolotto, 2002). A headset with microphone was prepared to record participants' responses.

On each trial, a fixation point was presented for 1000 milliseconds (ms) followed by the target verb (see Figure 3). The verb was presented on-screen for 2000 ms. This was followed

by the action scene, which was presented for 7000 ms. Participants described the action scene aloud using the given verb and all items on the screen within the given time; responses were recorded via a microphone. Each trial was immediately succeeded by the next trial. Participants had five practice trials before the main experiment began. All participants provided descriptions for all 72 action scenes in two blocks of 36 (18 verbs repeated across singular and plural subjects, in both Present Habitual and Past temporal contexts), separated by a self-paced break. Presentation order was randomized for each participant (participants subsequently repeated this procedure with the same 72 items in a different randomized order, but these data are not discussed further here).

The experiment lasted approximately 30 minutes, and participants were either offered cash (£5) or course credit for their participation.

Coding and Scoring

All trials were recorded as 7000 ms audio files via E-Prime and were transcribed exactly as spoken. Only the first response attempt was coded; any corrections were ignored. 8% of responses with non-target verbs were excluded. The subject of the sentence was coded for number (singular or plural); trials where the subject number in participants' response was incongruent with the image (e.g. the speaker used *duck* instead of *ducks* for an image of more than one duck) were excluded. Target verb inflections were coded into three categories: zero inflection (e.g. *shout*), past tense *-ed* (e.g. *shouted*) and 3rd person singular *–s* (3SG *-s*, e.g. *shouts*); note that we excluded from our analyses responses from participants who, on over 1/2 of the valid trials, failed to carry out the task as instructed (e.g., if they failed to produce target verbs, produced auxiliary forms, e.g., *has been waiting*, or failed to explicitly indicate the temporal context of the sentence as part of their response). Using these criteria we excluded data from participants who had failed to produce target verbs (N=4) or temporal

adverbials (N=3), or produced auxiliary forms (N=2) for more than 18 out of 36 trials. Non-target past *-ed* responses in Present Habitual contexts (past habitual responses) were included as they were necessary for past *-ed* likelihood analyses; note that these responses contained temporal-cue errors (in the context of the current experiments) but were not grammatically incorrect (we return to this point in the Interim and General Discussion).

Verbs with zero inflection, past tense *-ed* and 3SG *–s* responses were then scored for inflectional accuracy (1 or 0) based on the temporal context and subject number (see Table 2). For trials scored as incorrect, error type and subject number were coded as: *omission error* when an obligatory inflection was omitted, e.g. missing 3SG *-s* for singular subject in Present Habitual context or *commission error* when an incorrect inflection was produced, e.g. 3SG *-s* in Past contexts.

Table 2.

Coding and scoring criteria for inflectional production responses in Experiments 1, 2 and 3.

Temporal Context	Subject Number	Verb	Inflection	Example	Accuracy	Error Type
Present Habitual (*Every day*)	Singular (*the chef*)	*shout*	past *-ed*	*shout-ed*	0	commission
			3SG *-s*	*shout-s*	1	--
			zero Infl.	*shout*	0	omission
	Plural (*the ducks*)	*wait*	past *-ed*	*wait-ed*	0	--
			3SG *-s*	*wait-s*	0	commission
			zero infl.	*wait*	1	--
Past (*Yesterday*)	Singular (*the teacher*)	*applaud*	past *-ed*	*applaud-ed*	1	--
			3SG *-s*	*applaud-s*	0	commission
			zero Infl.	*applaud*	0	omission
	Plural (*the children*)	*paint*	past *-ed*	*paint-ed*	1	--
			3SG *-s*	*paint-s*	0	commission
			zero Infl.	*paint*	0	omission

Results

All data files and analysis scripts for experiments reported in this paper can be found online (via the Open Science Framework at https://doi.org/10.17605/OSF.IO/N6E9G).

Outcome variables (response accuracy, responses of different inflectional types, and of inflectional omissions) from Experiment 1 were analyzed using logistic mixed effects regression models (LMEs). We used a forward model building strategy with a maximal random effects structure (Barr, Levy, Scheepers & Tily, 2013). Predictor variables (Group, Temporal Context and Subject Number) were contrast-coded before being included as fixed effects predictors. We included Participant as a random intercept, and Item and Temporal Context as random slopes if they significantly improved model fit. Log-likelihood ratio tests (chi-squared test) were used to compare alternative logit regression models in order to decide whether the new model with additional fixed and random effect variables significantly improved the goodness-of-fit.

We carried out three sets of analyses on the data from the scene description task. The first and second set focused on the overall accuracy of inflection depending on temporal context and number regardless of morpheme (with accurate responses coded as **1**, and inaccurate responses coded as **0**) and the likelihood of production for each type of morpheme (3SG *-s* and past *-ed*). For both sets of analyses, a logistic mixed effects regression model was built with Group, Temporal Context and Subject Number as fixed effect predictors. Participant was included as a random intercept, and Item was included as a random intercept or slope if the log-likelihood chi-squared model comparison showed it significantly improved the fit of the model. Separate subgroup analyses were also conducted for L1 Mandarin (L2 English) and L1 English groups, deducting Group as a fixed effects predictor but keeping all other variables the same.

The third set of analyses focused on responses involving inflectional omissions (the most common type of error, as opposed to errors of commission), which provided additional insight into the likelihood of omitting 3SG *-s* or past *-ed* inflections across conditions for both L1 Mandarin and L1 English groups and thus informs the nature of such errors. For these

analyses, Bayesian logistic mixed effects models (BLMEs) were built to address the problem of partial separations (Rainey, 2016; Zorn, 2005). This was caused by the missing response category of Plural Subject omission in the Present Habitual temporal context, i.e. participants cannot make omission errors if the condition requires zero inflection, leading to consistent scores of 0 in this error category. This in turn led to the maximum likelihood estimate (Wald's Test) of a non-Bayesian logistic regression model tending towards infinity for the outcome variable (see Hauck & Donner, 1977). Consequently, a BLME model was used to impose a fixed prior to the fixed effect parameters, improving parameter estimates for inflectional omissions. Note that although Subject Number was not included as a predictor for omission error analyses (for reasons stated above), few number of errors across conditions and groups still qualified the use of a Bayesian model in addition to the missing response category. As error patterns are expected to differ substantially across groups, subgroup analyses were conducted for L1 Mandarin and L1 English groups. The subgroup BLME models consisted of Temporal Context as the only fixed effects predictor, and participant and item as random intercepts. Other instances of model non-convergence were dealt with using the '*bobyqa*' algorithm for constrained optimization by increasing the number of iterations to 10000.

Our presentation focuses on key main effects and interactions; see Tables 3-6 for complete inferential statistics for each model.

Overall Inflectional Accuracy

Response accuracy in each temporal context and subject condition was first analyzed (i.e., 3SG *-s* responses in the Present Habitual Singular Subject condition; zero-inflection responses in the Present Habitual Plural Subject conditions; and past *-ed* responses in the Past Singular / Plural Subject conditions; Figure 4).

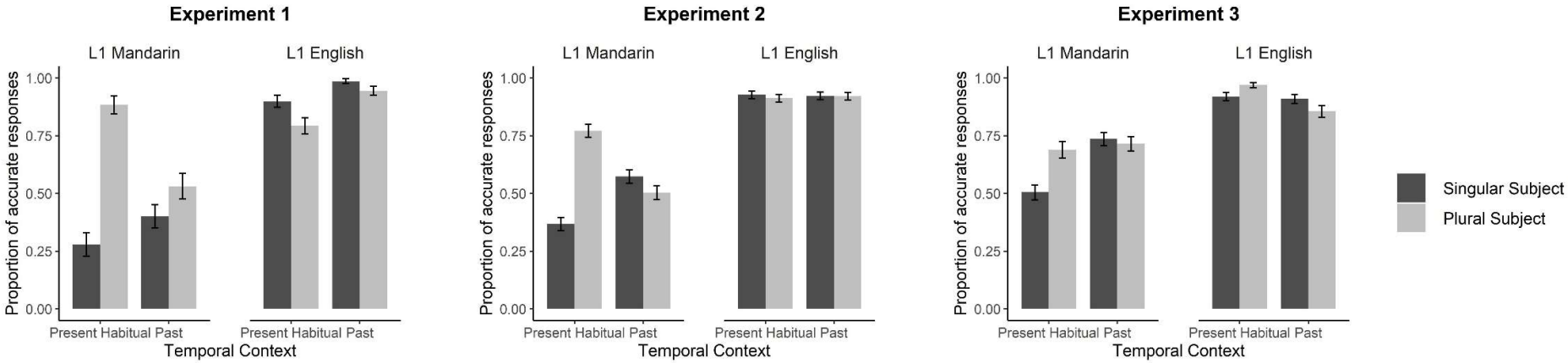

Figure 4. Experiments 1, 2 and 3: Average proportion of accurate inflectional responses in Present Habitual and Past temporal contexts in Scene description task for L1 Mandarin and L1 English groups (N=13;17; N=37;34; N=45;42). Error bars denote +/- 1 SE.

There was a significant main effect of Group. Inflectional accuracy was more variable across conditions in the L1 Mandarin group than in the L1 English group, with the L1 Mandarin group producing most accurate responses in the Present Habitual Plural Subject condition (which did not require any inflection) and fewest accurate responses in the Present Habitual Singular Subject condition (which required the 3SG *-s* inflection; M=0.88 vs. M=0.28; L1 English: M=0.79 vs. M=0.90). There was a two-way interaction between Group and Temporal Context. Participants produced more accurate responses in the Present Habitual contexts than in the Past contexts in the L1 Mandarin group (M=0.56 vs. M=0.46), but the inverse was true for the L1 English group (M=0.85 vs. M=0.96). There was also a significant three-way interaction between Group, Temporal Context and Subject Number (Table 3). Subgroup analyses revealed that in the L1 Mandarin group, Temporal Context interacted with Subject Number; in the L1 English group, there was no such interaction.

Inflectional Type

3rd Person Singular -s (3SG -s) responses

Further analyses were conducted 3SG *-s* responses of each condition (Figure 5). Note that a 3SG *-s* response was a grammatically correct response in the Present Habitual Singular Subject condition, but an error (i.e., production of an incorrect inflection [error of commission]) in all other conditions.

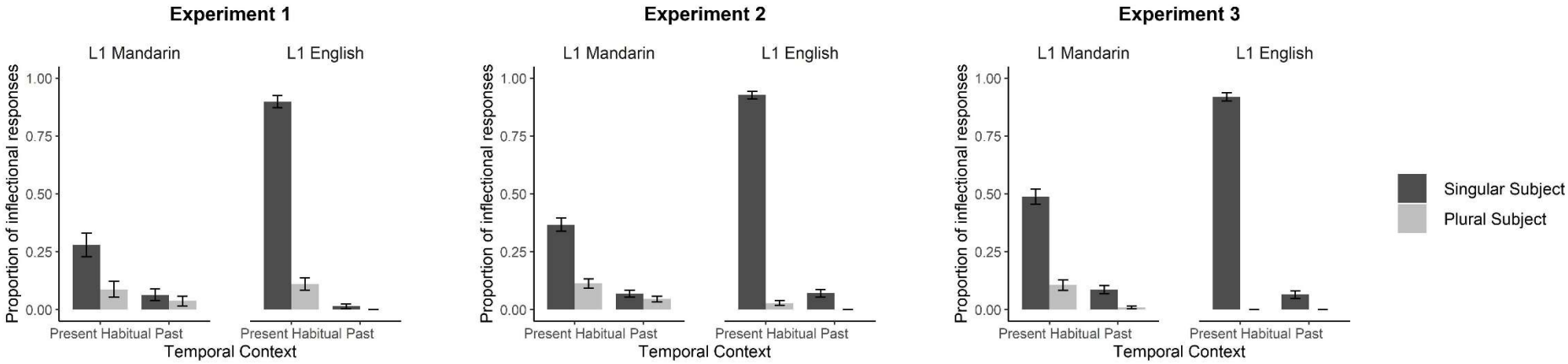

Figure 5. Experiments 1, 2 and 3: Average proportion of 3SG -s inflectional production across Present Habitual and Past temporal contexts for L1 Mandarin and L1 English groups (N=13;17; N=37;34; N=45;42). Error bars denote +/- 1 SE.

There was a significant main effect of Temporal Context (Table 4): Although the L1 Mandarin group produced more 3SG *-s* inflections in the Present Habitual Singular Subject context than in other contexts, they did so to a lesser extent than the L1 English group (M=0.28 vs. M=0.90).

Subgroup analyses confirmed the effect of Temporal Context in each group. They also revealed that in the L1 Mandarin group, there was a significant effect of Subject Number, with participants significantly more likely to produce 3SG *-s* inflections following a singular subject than a plural subject (M=0.16 vs. M=0.06). But critically, there was no a significant interaction between Subject Number and Temporal Context: Participants did not produce significantly more 3SG *-s* inflections in the Present Habitual Singular Subject condition than in other conditions.

In the L1 English group, there was a significant effect of Subject Number, with participants more likely to produce 3SG *-s* inflections following a singular subject than a plural subject. There was also a significant interaction between Subject Number and Temporal Context: Participants produced more 3SG *-s* inflections in the Present Habitual Singular Subject condition than in other conditions.

Past -ed responses

Further analyses examined past *-ed* responses in each condition (Figure 6). Note that a past *-ed* response constituted a grammatically correct response in the Past contexts, but an error in the Present Habitual contexts. Further analyses examined past *-ed* responses in each condition.

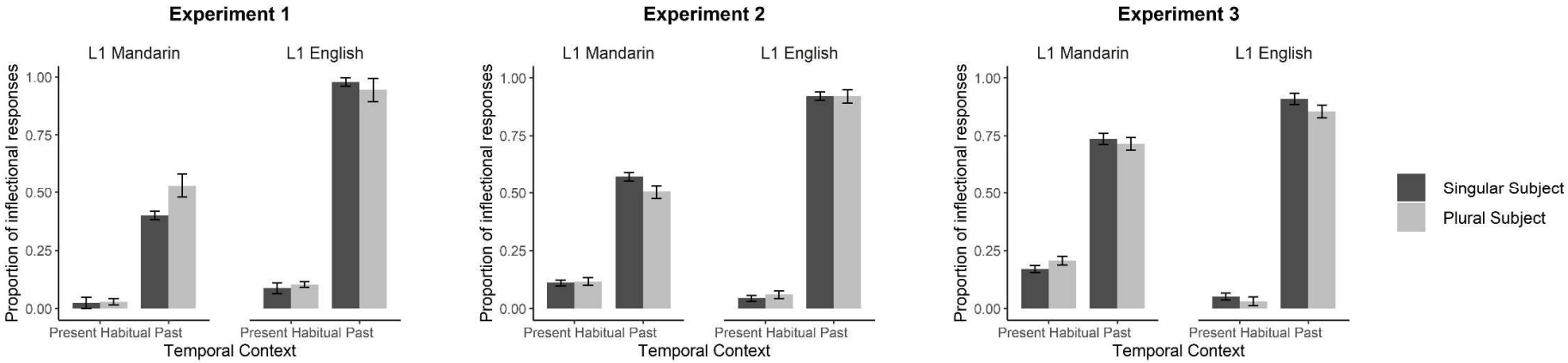

Figure 6. Experiments 1, 2 and 3: Average proportion of past -ed inflectional production across Present Habitual and Past temporal contexts for L1 Mandarin and L1 English groups (N=13;17; N=37;34; N=45;42). Error bars denote +/- 1 SE.

There was a significant main effect of Group and of Temporal Context, with a significant two-way interaction between Group and Temporal Context (Table 5): Although the L1 Mandarin group produced more past *-ed* inflections in the Past contexts than in the Present Habitual contexts (M=0.46 vs. M=0.03), they did so to a lesser extent than the L1 English group (M=0.96 vs. M=0.10). Subgroup analyses revealed that in the L1 Mandarin group, there was a significant effect of Temporal Context but no other significant effects, and that in the L1 English group, there was similarly a significant effect of Temporal Context but no other significant effects.

Inflectional errors

Participants made two types of inflectional error: omission errors (where the appropriate inflection was omitted), and errors of commission (where an incorrect inflection was produced).

Inflectional omission responses varied across groups and conditions. Inflectional omission responses were analyzed across the Present Habitual Singular Subject and Past Singular / Plural conditions (Table 6). Note that participants could not make omission errors in the Present Habitual Plural Subject condition, where a zero inflection would be grammatical (e.g., *Every day the chefs shout*). A BLME model was therefore used to analyze the likelihood of inflectional omissions out of all inflectional errors using Group and Temporal Context as predictors; Subject Number was not included as a predictor due to the missing response category for the Present Habitual Plural Subject condition.

There was a significant main effect of Group (Table 6): The L1 Mandarin group was significantly more likely to produce inflection omission responses than the L1 English group (M=0.44 vs. M=0.02 [calculated over all responses]). There was also a main effect of Temporal Context: On average, there were more inflection omission responses in the Past

contexts than in the Present Habitual contexts (M=0.21 vs. M=0.14; given no possibility of omission in Present Habitual Plural Subject condition). There was also a significant interaction between Group and Temporal Context, indicating that the effect of Temporal Context differed significantly between the L1 Mandarin and L1 English groups. Subgroup analyses revealed that there was no significant effect of Temporal Context in the L1 Mandarin group but there was in the L1 English group.

Participants made relatively few inflectional errors that were errors of commission, thus no additional inferential statistical analyses were performed. Overall, the L1 Mandarin group made fewer errors of commission than the L1 English group (M=0.06 vs. M=0.08). We note that both L1 Mandarin and L1 English groups produced past habitual forms in the present habitual contexts that were grammatical but erroneous under the intended (present habitual) interpretation of the temporal cue (e.g., '*every day the ducks waited for the farmer…';* M=0.01 vs. M=0.05).

Table 3.

Experiments 1, 2 and 3: Logistic mixed-effects statistics for inflectional accuracy for L1 Mandarin and L1 English groups.

		Experiment 1 (N=13;17)		Experiment 2 (N=37;34)		Experiment 3 (N=45;42)	
		B (SE)	***p***	**B (SE)**	***p***	**B (SE)**	***p***
Main Model							
	Intercept	2.14 (0.29)	**<.001**	1.59 (0.15)	**<.001**	1.81 (0.14)	**<.001**
	Group (***L1 Mandarin*** *vs. L1 English*)	3.04 (0.56)	**<.001**	2.69 (0.29)	**<.001**	2.10 (0.28)	**<.001**
	Temporal Context (***Present Habitual*** *vs. Past*)	0.89 (0.31)	**.004**	-0.08 (0.18)	.667	-0.12 (0.16)	.442
	Subject Number (***Singular*** *vs. Plural*)	-0.16 (0.31)	.601	0.38 (0.18)	**.031**	0.29 (0.16)	.070
	Group × Temporal Context	3.15 (0.56)	**<.001**	0.21 (0.27)	.441	-1.62 (0.32)	**<.001**
	Group × Subject Number	-3.45 (0.56)	**<.001**	-0.95 (0.27)	**<.001**	-0.17 (0.32)	.598
	Temporal Context × Subject Number	-1.56 (0.56)	**.010**	-1.09 (0.35)	**.002**	-1.35 (0.33)	**<.001**
	Group × Temporal Context × Subject Number	2.63 (1.09)	**.016**	2.62 (0.54)	**<.001**	-0.72 (0.65)	.267
L1 Mandarin							
	Intercept	0.23 (0.39)	.559	0.24 (0.17)	.159	0.77 (0.16)	**<.001**
	Temporal Context	-1.11 (0.32)	**<.001**	-0.18 (0.19)	.340	0.70 (0.20)	**<.001**
	Subject Number	2.02 (0.31)	**<.001**	0.84 (0.19)	**<.001**	0.38 (0.20)	.055
	Temporal Context × Subject Number	-3.23 (0.63)	**<.001**	2.38 (0.38)	**<.001**	-0.94 (0.40)	**.012**
L1 English							
	Intercept	3.58 (0.50)	**<.001**	3.02 (0.29)	**<.001**	3.07 (0.31)	**<.001**
	Temporal Context	2.22 (0.58)	**<.001**	0.01 (0.24)	.968	-0.96 (0.28)	**<.001**
	Subject Number	-1.54 (0.57)	**.003**	-0.09 (0.24)	.706	0.20 (0.28)	.477
	Temporal Context × Subject Number	-0.69 (1.09)	.524	0.22 (0.48)	.633	-1.77 (0.57)	**.002**

Table 4. *Experiments 1, 2 and 3: Logistic mixed-effects statistics for 3SG -s responses for L1 Mandarin and L1 English groups.*

		Experiment 1 (N=13;17)		Experiment 2 (N=37;34)		Experiment 3 (N=45;42)	
		B (SE)	***p***	**B (SE)**	***p***	**B (SE)**	***p***
Main Model							
	Intercept	-3.77 (0.61)	**<.001**	-2.76 (0.36)	**<.001**	-2.91 (0.50)	**<.001**
	Group (***L1 Mandarin*** *vs. L1 English*)	0.32 (0.55)	.558	0.93 (0.34)	**.006**	0.72 (0.30)	**.015**
	Temporal Context (***Present Habitual*** *vs. Past*)	-4.85 (1.07)	**<.001**	-2.69 (0.64)	**<.001**	-3.20 (0.96)	**.001**
	Subject Number (***Singular*** *vs. Plural*)	-	-	-	-	-	-
	Group × Temporal Context	-4.68 (0.89)	**<.001**	-2.43 (0.39)	**<.001**	-1.99 (0.44)	**<.001**
	Group × Subject Number	-	-	-	-	-	-
	Temporal Context × Subject Number	-	-	-	-	-	-
	Group × Temporal Context × Subject Number	-	-	-	-	-	-
L1 Mandarin							
	Intercept	-2.61 (0.37)	**<.001**	-2.61 (0.28)	**<.001**	-2.54 (0.26)	**<.001**
	Temporal Context	-1.55 (0.45)	**.001**	-1.80 (0.26)	**<.001**	-2.67 (0.41)	**<.001**
	Subject Number	-1.03 (0.45)	**.022**	-1.16 (0.26)	**<.001**	-2.41 (0.44)	**<.001**
	Temporal Context × Subject Number	0.91 (0.88)	.301	1.43 (0.51)	**.005**	0.10 (0.86)	.903
L1 English							
	Intercept	-4.48 (0.97)	**<.001**	-3.58 (0.82)	**<.001**	-5.08 (1.61)	**.002**
	Temporal Context	-7.63 (1.70)	**<.001**	-4.12 (1.24)	**.001**	-4.58 (2.31)	**.048**
	Subject Number	-	-	-	-	-	-
	Temporal Context × Subject Number	-	-	-	-	-	-

Table 5. *Experiments 1, 2 and 3: Logistic mixed-effects statistics for past -ed responses for L1 Mandarin and L1 English groups.*

		Experiment 1 (N=13;17)		Experiment 2 (N=37;34)		Experiment 3 (N=45;42)	
		B (SE)	***p***	**B (SE)**	***p***	**B (SE)**	***p***
Main Model							
	Intercept	-0.47 (0.31)	.134	-0.76 (0.24)	**.001**	-0.47 (0.19)	**.013**
	Group (***L1 Mandarin*** *vs. L1 English*)	2.97 (0.63)	**<.001**	1.34 (0.47)	**.005**	-0.21 (0.38)	.583
	Temporal Context (***Present Habitual*** *vs. Past*)	5.70 (0.40)	**<.001**	5.23 (0.25)	**<.001**	5.13 (0.26)	**<.001**
	Subject Number (***Singular*** *vs. Plural*)	-0.13 (0.33)	.699	0.00 (0.17)	.993	-0.31 (0.18)	.090
	Group × Temporal Context	2.71 (0.77)	**<.001**	4.60 (0.49)	**<.001**	3.47 (0.47)	**<.001**
	Group × Subject Number	-0.69 (0.69)	.312	0.36 (0.34)	.295	-0.68 (0.37)	.063
	Temporal Context × Subject Number	-0.65 (0.67)	.334	-0.43 (0.34)	.217	-0.32 (0.37)	.383
	Group × Temporal Context × Subject Number	-1.66 (1.39)	.232	0.01 (0.69)	.992	0.17 (0.73)	.813
L1 Mandarin							
	Intercept	-2.45 (0.59)	**<.001**	-1.50 (0.35)	**<.001**	-0.37 (0.25)	.136
	Temporal Context	4.09 (0.61)	**<.001**	3.08 (0.27)	**<.001**	3.37 (0.23)	**<.001**
	Subject Number	0.31 (0.54)	.563	-0.20 (0.26)	.436	0.03 (0.19)	.869
	Temporal Context × Subject Number	0.42 (1.12)	.709	-0.43 (0.51)	.405	-0.41 (0.38)	.289
L1 English							
	Intercept	0.63 (0.34)	**.066**	-0.09 (0.36)	.805	-0.58 (0.29)	**.046**
	Temporal Context	6.50 (0.52)	**<.001**	7.58 (0.50)	**<.001**	6.91 (0.50)	**<.001**
	Subject Number	-0.40 (0.42)	.349	0.18 (0.29)	.535	-0.65 (0.31)	**.037**
	Temporal Context × Subject Number	-1.25 (0.83)	.135	-0.42 (0.59)	.472	-0.24 (0.63)	.699

Table 6.

Experiments 1, 2 and 3: Bayesian logistic mixed-effects statistics for inflectional omission for L1 Mandarin and L1 English groups.

		Experiment 1 (N=13;17)		Experiment 2 (N=37;34)		Experiment 3 (N=45;42)	
		B (SE)	***p***	**B (SE)**	***p***	**B (SE)**	***p***
Main Model							
	Intercept	2.26 (0.70)	**.001**	1.08 (0.21)	**<.001**	0.40 (0.26)	.123
	Group (***L1 Mandarin*** *vs. L1 English*)	-2.85 (1.03)	**.006**	-2.00 (0.46)	**<.001**	-0.52 (0.55)	.384
	Temporal Context (***Present Habitual*** *vs. Past*)	3.32 (1.17)	**.005**	1.57 (0.25)	**<.001**	1.76 (0.34)	**<.001**
	Subject Number (***Singular*** *vs. Plural*)	-	-	-	-	-	-
	Group × Temporal Context	4.35 (1.53)	**.004**	0.89 (0.71)	.213	0.96 (0.84)	.250
	Group × Subject Number	-	-	-	-	-	-
	Temporal Context × Subject Number	-	-	-	-	-	-
	Group × Temporal Context × Subject Number	-	-	-	-	-	-
L1 Mandarin							
	Intercept	4.02 (1.51)	**.008**	1.13 (0.44)	**.010**	-0.02 (0.76)	.980
	Temporal Context	2.23 (1.40)	.111	2.98 (0.75)	**<.001**	3.41 (1.18)	**.004**
	Subject Number	-	-	-	-	-	-
	Temporal Context × Subject Number	-	-	-	-	-	-
L1 English							
	Intercept	-0.19 (0.78)	.804	-0.63 (0.33)	.055	-0.06 (1.10)	.955
	Temporal Context	5.75 (1.41)	**<.001**	2.00 (0.66)	**.002**	3.29 (1.66)	**.047**
	Subject Number	-	-	-	-	-	-
	Temporal Context × Subject Number	-	-	-	-	-	-

Discussion

Although L1 Mandarin speakers made errors when producing English temporal inflections, with an overall accuracy rate of only 52% across conditions, they nevertheless showed sensitivity to temporal context. That is, they were more likely to produce 3SG -*s* and past -*ed* inflections in a temporal context that was appropriate for those inflections than in a temporal context that was inappropriate for those inflections. However, their performance was not uniform across inflections. Notably, they showed particularly low accuracy in the Present Habitual Singular Subject condition (requiring 3SG -*s*), relative to L1 English speakers' performance in the same condition, and relative to their own performance in the Past Tense conditions (requiring past -*ed*). These results suggest that L1 Mandarin speakers were able to conceptualize and linguistically encode relevant tense distinctions, but that they were not able to produce them consistently, with one inflection type being more susceptible to error than another.

In contrast, our L1 English speakers made few errors, and these 'errors' did not tend to be ungrammatical: They tended to be inflections that were temporally erroneous for the experimental condition of the task but grammatically correct (e.g., '*every day the ducks waited for the farmer…*'), rather than the agreement violations that L1 Mandarin speakers produced (e.g., '*every day the ducks waits for the farmer…*'). One likely explanation for these responses is that they sometimes did not interpret the temporal cue as intended in the context of the experiment (i.e., sometimes erroneously interpreted *'every day'* as indicating past habitual rather than present habitual).

In Experiment 2, we sought to replicate these findings with a larger sample and a more robust experimental paradigm. In Experiment 1, which used speeded presentation, participants misrecalled the verb on 8% of trials. Consequently, Experiment 2 used self-paced verb presentation to increase the proportion of valid responses.

Experiment 2

Methods

Participants

37 L1 Mandarin speakers of English aged 20-29 (M=23.38; SD=1.83) and 34 L1 English speakers aged 19-46 (M=24.21; SD=5.65) participated in Experiment 2 and provided valid data; an additional 15 participants were also tested but their data were subsequently excluded for several reasons (see below for details). In addition to the participant recruitment criteria for the L1 Mandarin group (L2 English) in Experiment 1, Mandarin participants were also required to have at least a score of 5.5 on the spoken component of the IELTS exam. Recruitment criteria for the L1 English group was identical to those in Experiment 1.

Materials

We used the images, verbs, vocabulary list and pictorial legend aid from Experiment 1, with minor adjustments to remove ambiguity in some items. Each combination of experimental verb and scene was presented only once (hence, the total number of trials was halved to 72 with 36 experimental and 36 filler trials). Verb presentation in each trial was self-paced to allow participants more time to use the verb.

Design

The experimental design was identical to Experiment 1.

Procedure

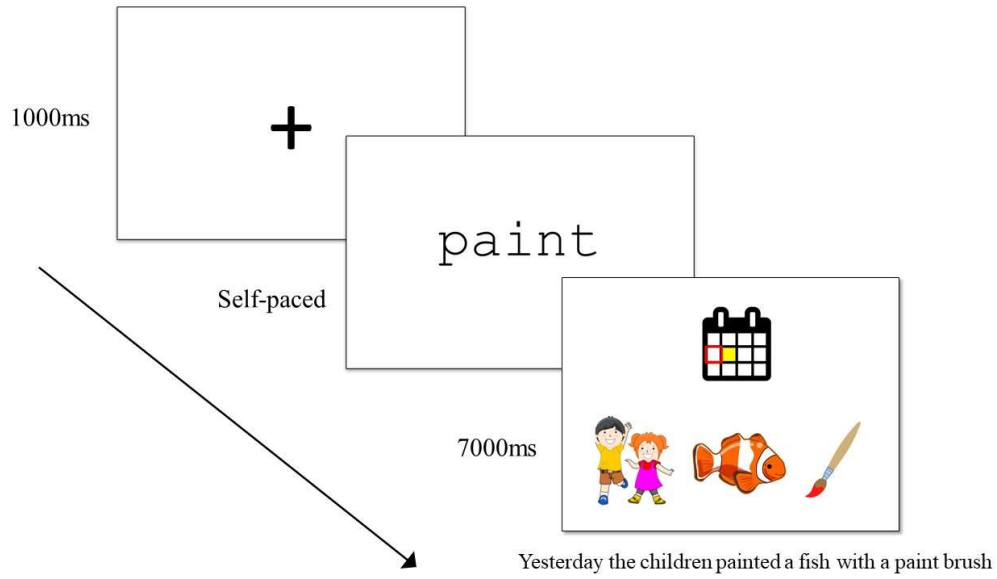

Figure 7. Experiment 2: Three-step self-paced trial procedure for the scene description task, including presentations of fixation (1000ms), target verb (self-paced) and target image (7000ms).

The experimenter followed the same protocol as Experiment 1 but emphasized the self-paced element of the scene description task (Figure 7). Before the start of the experiment, participants were reminded that they should remember the verb before viewing the action scene, use the objects in the scene from top to bottom and left to right, and avoid using auxiliary verbs in descriptions in each trial (see Gardner, Branigan & Chondrogianni, 2018).

The trial procedure was identical to Experiment 1, except that participants controlled the progress of each trial by pressing the [**SPACE BAR**] after reading the target verb. Participants had five practice trials before commencing the formal experiment. Participants were paid £5 in cash or given course credit for their time.

Coding and Scoring

The coding and scoring procedures for the scene description task were identical to Experiment 1. As in Experiment 1, we excluded data from participants who failed to produce target verbs (N=5) or temporal adverbials (N=6) as instructed, or produced auxiliary forms (N=1) for over 1/2 of the experimental trials (more than 18 out of 36 trials). Three Mandarin participants were excluded as they did not meet the English proficiency requirements or had been living in the UK for longer than 24 months.

Results

Analyses were conducted as in Experiment 1, except where otherwise stated.

Overall Inflectional Accuracy

There was a significant main effect of Group and of Temporal Context, with a significant three-way interaction between Group, Temporal Context and Subject Number (Table 3). Inflectional accuracy was again more variable across conditions in the L1 Mandarin group than in the L1 English group, with the L1 Mandarin group producing most accurate responses in the Present Habitual Plural Subject condition and fewest accurate responses in the Present Habitual Singular Subject condition (L1 Mandarin: M=0.77 vs. M=0.37; L1 English: M=0.93 vs. M=0.91; Figure 4).

Subgroup analyses revealed that in the L1 Mandarin group, there was a significant interaction between Temporal Context and Subject Number; in the L1 English group, there was no such interaction.

Inflectional Type

3rd person singular -s (3SG -s) responses

Group and Temporal Context were used as fixed effects predictors for the 3SG *-s* analysis BLME model. Subject Number was dropped due to the missing category problem in the L1 English group (i.e., no 3SG *-s* response for Past Plural Subject condition; Figure 5)

There was a significant main effect of Group and of Temporal Context, with a significant interaction between Group and Temporal Context (Table 4): Although the L1 Mandarin group produced more 3SG *-s* inflections in the Present Habitual contexts than in the Past contexts (M=0.25 vs. M=0.06), they did so to a lesser extent than the L1 English group (M=0.47 vs. M=0.03; Figure 5).

Subgroup analyses revealed that in the L1 Mandarin group, Subject Number was a significant predictor, with participants being more likely to produce 3SG *-s* inflections following a singular subject than a plural subject.

Past –ed responses

There were significant main effects of Group and of Temporal Context, with a significant two-way interaction between Group and Temporal Context (Table 5): Although the L1 Mandarin group produced more past *-ed* inflections in the Past contexts than in the Present Habitual contexts (M=0.54 vs. M=0.11), they did so to a lesser extent than the L1 English group (M=0.92 vs. M=0.05; Figure 6). Subgroup analyses revealed that in the L1 Mandarin group, there was a significant effect of Temporal Context but no other significant effects or interactions; likewise, in the L1 English group, there was a significant effect of Temporal Context but no other significant effects or interactions.

Inflectional errors

Similar to Experiment 1, participants across L1 Mandarin and L1 English groups again made omission errors and errors of commission.

Inflectional omission responses varied across groups and conditions. There was a significant main effect of Group (Table 6): The L1 Mandarin group was significantly more likely to produce inflection omission responses than the L1 English group (M=0.35 vs. M=0.03). There was also a main effect of Temporal Context: Participants across L1 Mandarin and L1 English groups were more likely to make omission errors in the Past contexts than in the Present Habitual contexts (M=0.23 vs. M=0.15; given no possibility of omission in Present Habitual Plural Subject condition). There was no significant interaction between Group and Temporal Context, indicating that the effect of Temporal Context did not differ significantly between the L1 Mandarin and L1 English groups. Subgroup analyses confirmed the significant effect of Temporal Context across both L1 Mandarin and L1 English groups.

There were again relatively few inflectional errors that were errors of commission, thus no additional inferential statistical analyses were performed. Overall, the L1 Mandarin group made more errors of commission than the L1 English group (M=0.11 vs. M=0.05), including past habitual forms in the present habitual contexts (M=0.05 vs. M=0.03).

Discussion

Experiment 2 replicated the finding that although L1 Mandarin participants made inflectional errors (omissions and incorrect production), they nevertheless showed sensitivity to temporal context. As in Experiment 1, they showed particularly low accuracy in the Present Habitual Singular Subject condition (requiring 3SG -*s*), relative to L1 English speakers' performance in the same condition, and relative to their own performance in the

Past Tense conditions (requiring -*ed*). These results provide further evidence that L1 Mandarin speakers are able to conceptualize and linguistically encode relevant tense distinctions (i.e., that they have the necessary conceptual and linguistic representations) but are not able to process them consistently during production. They also replicate the pattern whereby the (featurally complex) 3SG -*s* inflection with subject number and tense features is more susceptible to error than the (featurally less complex) past -*ed* inflection with a tense feature only. However, Experiment 2 does not distinguish whether this difference was due to inconsistent retrieval of morphological forms or errors in oral articulation.

In Experiment 3, we therefore examined whether these patterns of inflectional error would remain when participants did not orally articulate their responses. To do this, we used the same task as Experiment 2, but we asked participants to produce typed responses. If L1 Mandarin participants' inflectional errors in Experiments 1 and 2 resulted from oral articulatory failures alone, we should see significantly higher inflectional accuracy in Experiment 3 compared to Experiment 2. On the other hand, if retrieval failures accounted for the inflectional errors previously observed, similar patterns of inflectional production should remain even with the articulatory component of the task removed.

Experiment 3

Methods

Participants

45 L1 Mandarin speakers of English aged 18-31 (M=22.93; SD=2.29) and 42 monolingual L1 English speakers aged 17-20 (M=18.21; SD=0.74) participated in Experiment 3 and provided valid data; an additional four participants were also tested but

their data were subsequently excluded for different reasons (see below for details). Participants were recruited based on identical criteria as Experiment 2 with no additional requirements.

Materials

24 experimental items and 24 filler items were used from Experiment 2 (retaining 12 of the 18 verbs: six experimental and six filler verbs) in order to shorten the experimental session (as participants were considerably slower to produce typed than spoken responses). The trial structure, pictorial legend and illustrative trial examples were identical to Experiment 2. Individual images of animal, place, food, activity etc. were labelled with their corresponding names (e.g. dog, park, etc.) were created as JPGs as part of a vocabulary training session.

Procedure

Prior to the formal experimental session, participants passively viewed images with their corresponding names as part of a vocabulary training session in order to familiarize themselves with the trial images. The instructions for the scene description task were identical to Experiment 2 except that participants were told that during the presentation of each action scene, they must type out their descriptions on a computer keyboard, with no option to edit their responses (i.e., written scene description task). The duration of each action scene was increased to 15000 ms to reflect response modality. Afterwards, participants were paid £5 for their time.

Coding and Scoring

The coding and scoring procedures for the written scene description task were identical to the spoken scene description task in Experiment 1 and 2. As in the previous two

experiments, we excluded data from participants who had failed to produce target verbs (N=2) or produced auxiliary forms (N=2) for over 1/2 of experimental trials (more than 12 out of 24 trials). One additional participant was excluded as the participant could not provide valid English proficiency scores for Experiment 3.

Results

Overall Inflectional Accuracy

There was a significant main effect of Group, with a significant interaction between Group and Temporal Context (Table 3). Inflectional accuracy was more variable across temporal contexts in the L1 Mandarin group than in the L1 English group. The L1 Mandarin group showed greater disparity between performance in the Past and Present Habitual contexts than the L1 English group (L1 Mandarin M=0.73 vs. M=0.58; L1 English: M=0.88 vs. M=0.94; Figure 4). Notably, however, there was no significant three-way interaction between Group, Temporal Context, and Subject Number.

Subgroup analyses revealed that in the L1 Mandarin group, there was a significant effect of Temporal Context and a marginal effect of Subject number, with a significant interaction between Temporal Context and Subject Number. Similarly, in the L1 English group, there was a significant effect of Temporal Context and a significant interaction between Temporal Context and Subject Number; however, there was no significant effect of Subject Number for the L1 English group.

Inflectional Type

3rd person singular -s responses

There was a significant main effect of Group and of Temporal Context, with a significant interaction between Group and Temporal Context (Table 4): Although the L1 Mandarin

group produced more 3SG *-s* inflections in the Present Habitual contexts than in the Past contexts (M=0.33 vs. M=0.05), they did so to a lesser extent than the L1 English group (M=0.47 vs. M=0.03; Figure 5).

Subgroup analyses revealed that in the L1 Mandarin group, there was a significant effect of Subject Number, with participants more likely to produce 3SG *-s* inflections following a singular subject than a plural subject; there was also a significant interaction between Temporal Context and Subject Number, with participants most likely to produce 3SG *-s* inflections in the Present Habitual Singular Subject condition. In the L1 English group, there was a significant effect of Temporal Context.

Past –ed responses

There was a significant main effect of Temporal Context, with a significant two-way interaction between Group and Temporal Context (Table 5): Although the L1 Mandarin group produced more past *-ed* inflections in the Past contexts than in the Present Habitual contexts (M=0.73 vs. M=0.19), they did so to a lesser extent than the L1 English group (M=0.88 vs. M=0.04; Figure 6). Subgroup analyses revealed a significant effect of Temporal Context for the L1 Mandarin group but no other significant effects; likewise, there was a significant effect of Temporal Context but no other significant effects for the L1 English group.

Inflectional errors

Inflectional omission responses varied across groups and conditions. There were no significant main effects of Group (M=0.21 vs. M=0.04) but there was a marginal main effect of Temporal Context in the main analysis (M=0.15 vs. M=0.10; Table 6). There was a marginal effect of Temporal Context in the L1 Mandarin group but not the L1 English group in the subgroup analysis.

There were again relatively few inflectional errors of commission. Overall, the L1 Mandarin group made more errors of commission than the L1 English group (M=0.15 vs. M=0.04), including past habitual forms in the Present Habitual contexts (M=0.09 vs. M=0.02).

Between Experiment Comparisons

Three sets of analyses were conducted to test the effect of production modality (spoken vs. written) on inflectional accuracy, inflectional type (3SG -*s* and past -*ed*), and inflectional omission (including existing predictors Group, Temporal Context and Subject Number). Data from Experiment 2 and Experiment 3 were combined, including only verbs which were common across the two experiments (verbs which were used in Experiment 2 but not Experiment 3 were excluded). For full statistics, see Tables 7-10.

There was no significant main effect of Modality on inflectional accuracy overall. Participants overall were not more likely to produce an accurate response in the written modality compared with the spoken modality (M=0.79 vs. M=0.74). Interestingly, there was a two-way interaction between Group and Subject Number irrespective of Modality. There were greater differences between singular and plural subjects in the L1 Mandarin group than the L1 English group (L1 Mandarin: M=0.57 vs. M=0.69; L1 English: M=0.92 vs. M=0.92). There was a three-way interaction between Group, Temporal Context and Modality and a four-way interaction also including Subject Number (Table 7).

Subgroup analyses revealed a similar picture. Different from the individual experiment analyses reported above, there were significant effects of Temporal Context for both L1 Mandarin and L1 English groups. However, the two groups showed different effects of Temporal Context. Whilst L1 Mandarin participants were less likely to produce accurate inflections in the Present Habitual than the Past Context (M=0.57 vs. M=0.65), L1 English

participants were more likely to produce accurate responses in the Present Habitual Context than the Past Context (M=0.94 vs. M=0.90) irrespective of production modality. Critically, there was a marginal main effect of Modality in the L1 Mandarin, but not in the L1 English group. Additionally, whilst there was a two-way interaction between Temporal Context and Modality, and a three-way interaction between Temporal Context, Subject Number and Modality for the L1 Mandarin group, there were no such interactions for the L1 English group.

For 3SG *-s* production, there was no significant main effect of Modality overall. Participants were not more likely to produce 3SG *-s* in the written modality compared with the spoken modality (M=0.22 vs. M=0.21). Modality did not interact with any other predictors (Group, Temporal Context and Subject Number). Subgroup analyses did not reveal any significant effects of Modality, nor any interactions (Table 8).

For past *-ed* production, there was no significant main effect of Modality. Similar to 3SG *-s*, participants were not more likely to produce past *-ed* in the written modality compared with the spoken modality (M=0.46 vs. M=0.41). The interaction between Group and Modality was close to significance (Table 9). Subgroup analyses revealed a significant effect of Modality for the L1 Mandarin group for past *-ed* production, but not for the L1 English group. No other interactions were found involving Modality in either group.

For inflectional omissions, there was a marginal main effect of Modality overall (Table 10): Numerically, participants omitted fewer inflections in the written modality than the spoken modality (M=0.13 vs. M=0.18). There was also a marginal interaction between Group and Modality. Subgroup analyses revealed a significant main effect of Modality on inflectional omission for the L1 Mandarin group, but not the L1 English Group: L1 Mandarin participants were more likely to make omission errors in the spoken modality than in the

written modality, but the L1 English participants did not show modality effects as they made very few errors overall and were close to ceiling levels in terms of accuracy in both modalities. Interestingly, L1 Mandarin speakers also produced more past habitual forms in the written modality compared with the spoken modality (Exp. 3 - 0.09 vs. Exp. 2 – 0.05), whilst L1 English speakers produced the similar proportions of past habitual forms (Exp. 3 – 0.02 vs. Exp. 2 – 0.03).

Table 7.

Experiments 2 and 3: Between-experiment analyses on the effect of production modality on inflectional accuracy.

		B (SE)	*p*
Main Model			
	Intercept	1.76 (0.11)	**<.001**
	Group ***(L1 Mandarin** vs. L1 English)*	2.36 (0.21)	**<.001**
	Temporal Context ***(Present Habitual** vs Past)*	-0.16 (0.14)	**.274**
	Subject Number ***(Singular** vs Plural)*	0.30 (0.14)	**.036**
	Modality ***(Spoken** vs Written)*	0.18 (0.21)	.372
	Group × Temporal Context	-1.00 (0.24)	**<.001**
	Group × Subject Number	-0.46 (0.24)	**.051**
	Group × Modality	-0.51 (0.41)	.212
	Temporal Context × Subject Number	-1.19 (0.29)	**<.001**
	Temporal Context × Modality	0.07 (0.24)	.764
	Subject Number × Modality	-0.00 (0.24)	.992
	Group × Temporal Context × Subject Number	0.61 (0.48)	.199
	Group × Temporal Context × Modality	-1.42 (0.48)	**.003**
	Group × Subject Number × Modality	0.66 (0.48)	.169
	Temporal Context × Modality × Subject Number	-0.33 (0.48)	.495
	Group × Temporal Context × Subject Number × Modality	-2.95 (0.96)	**.002**
L1 Mandarin			
	Intercept	0.57 (0.13)	**<.001**

	Temporal Context	0.34 (0.16)	**.029**
	Subject Number	0.53 (0.16)	**.001**
	Modality	0.44 (0.23)	.054
	Temporal Context × Subject Number	-1.47 (0.32)	**<.001**
	Temporal Context × Modality	0.79 (0.23)	**.001**
	Subject Number × Modality	-0.32 (0.24)	.173
	Temporal Context × Subject Number × Modality	1.15 (0.47)	**.015**
L1 English			
	Intercept	3.09 (0.22)	**<.001**
	Temporal Context	-0.68 (0.21)	**.001**
	Subject Number	0.07 (0.21)	.751
	Modality	-0.07 (0.39)	.866
	Temporal Context × Subject Number	0.93 (0.42)	**.026**
	Temporal Context × Modality	-0.66 (0.42)	.112
	Subject Number × Modality	0.30 (0.42)	.477
	Temporal Context × Subject Number × Modality	-1.84 (0.84)	**.028**

Table 8. *Experiments 2 and 3: Between-experiment analyses on the effect of production modality on 3SG -s production.*

		B (SE)	*p*
Main Model			
	Intercept	-2.68 (0.41)	**<.001**
	Group ***(L1 Mandarin** vs. L1 English)*	0.77 (0.22)	**.001**
	Temporal Context ***(Present Habitual** vs. Past)*	-2.77 (0.79)	**<.001**
	Modality ***(Spoken** vs. Written)*	-0.01 (0.22)	.970
	Group × Temporal Context	-2.17 (0.30)	**<.001**
	Group × Modality	-0.14 (0.44)	.744
	Temporal Context × Modality	-0.13 (0.29)	.656
	Group × Temporal Context × Modality	0.63 (0.59)	.285
L1 Mandarin			
	Intercept	-2.46 (0.26)	**<.001**
	Temporal Context	-2.27 (0.43)	**<.001**
	Modality	0.08 (0.31)	.795
	Temporal Context × Modality	-0.36 (0.36)	.325
L1 English			
	Intercept	-4.77 (1.32)	**<.001**
	Temporal Context	-5.49 (2.24)	**<.014**
	Modality	-0.27 (0.44)	.539
	Temporal Context × Modality	0.73 (0.67)	.280

Table 9. *Experiments 2 and 3: Between-experiment analyses on the effect of production modality on past -ed roduction.*

		B (SE)	*p*
Main Model			
	Intercept	-0.63 (0.16)	**<.001**
	Group ***(L1 Mandarin** vs. L1 English)*	0.40 (0.31)	.197
	Temporal Context ***(Present Habitual** vs Past)*	5.41 (0.24)	**<.001**
	Subject Number ***(Singular** vs Plural)*	-0.14 (0.18)	.436
	Modality ***(Spoken** vs Written)*	0.33 (0.31)	.289
	Group × Temporal Context	3.97 (0.38)	**<.001**
	Group × Subject Number	-0.26 (0.28)	.351
	Group × Modality	-1.35 (0.62)	**.029**
	Temporal Context × Subject Number	-0.46 (0.37)	.207
	Temporal Context × Modality	-0.26 (0.37)	.474
	Subject Number × Modality	-0.33 (0.29)	.257
	Group × Temporal Context × Subject Number	0.11 (0.57)	.846
	Group × Temporal Context × Modality	-0.77 (0.73)	.294
	Group × Subject Number × Modality	-0.94 (0.57)	.102
	Temporal Context × Subject Number × Modality	0.31 (0.58)	.590
	Group × Temporal Context × Subject Number × Modality	0.17 (1.15)	.884
L1 Mandarin			
	Intercept	-0.83 (0.22)	**<.001**

	Temporal Context	3.44 (0.24)	**<.001**
	Subject Number	0.00 (0.22)	.992
	Modality	1.01 (0.41)	**.013**
	Temporal Context × Subject Number	-0.53 (0.43)	.223
	Temporal Context × Modality	0.13 (0.34)	.706
	Subject Number × Modality	0.14 (0.30)	.641
	Temporal Context × Subject Number × Modality	0.26 (0.61)	.673
L1 English			
	Intercept	-0.43 (0.23)	.066
	Temporal Context	7.35 (0.41)	**<.001**
	Subject Number	-0.30 (0.24)	.212
	Modality	-0.35 (0.47)	.459
	Temporal Context × Subject Number	-0.43 (0.48)	.377
	Temporal Context × Modality	-0.66 (0.65)	.313
	Subject Number × Modality	-0.80 (0.49)	.098
	Temporal Context × Subject Number × Modality	0.37 (0.97)	.701

Table 10.

Experiments 2 and 3: Between-experiment Bayesian analyses on the effect of production modality on inflectional omissions.

		B (SE)	*p*
Main Model			
	Intercept	0.82 (0.17)	**<.001**
	Group ***(L1 Mandarin** vs. L1 English)*	-1.53 (0.37)	**<.001**
	Temporal Context ***(Present Habitual** vs. Past)*	1.68 (0.20)	**<.001**
	Modality ***(Spoken** vs. Written)*	-0.58 (0.33)	.078
	Group × Temporal Context	1.03 (0.58)	.077
	Group × Modality	1.60 (0.72)	**.026**
	Temporal Context × Modality	0.06 (0.42)	.882
	Group × Temporal Context × Modality	-0.14 (1.13)	.903
L1 Mandarin			
	Intercept	1.09 (0.21)	**<.001**
	Temporal Context	1.51 (0.22)	**<.001**
	Modality	-0.88 (0.41)	**.031**
	Temporal Context × Modality	0.08 (0.46)	.853
L1 English			
	Intercept	-0.42 (0.24)	.080
	Temporal Context	2.13 (0.51)	**<.001**
	Modality	0.65 (0.48)	.173
	Temporal Context × Modality	0.00 (0.88)	.999

Discussion

Experiment 3 replicated the key findings of Experiments 1 and 2 in the written modality: L1 Mandarin speakers showed sensitivity to temporal context, but exhibited particularly low accuracy in the Present Habitual Singular Subject condition (requiring 3SG -*s*), relative to L1 English speakers' performance in the same condition, and relative to their own performance in the Past temporal contexts (requiring past -*ed*). Critically, L1 Mandarin speakers were (marginally) less likely to be accurate in the spoken modality (Exp. 2) than in the written modality (Exp. 3), and were more likely to make omission errors in the spoken modality than in the written modality. L1 English speakers did not show such effects. These findings are consistent with a processing account of inflectional errors, based on inconsistent retrieval of morphological forms; they are not compatible with an account that attributes inflectional errors purely to articulatory difficulties.

General Discussion

Previous research has established that L2 speakers frequently produce inflectional errors but has not reached consensus over the loci of such errors. Specifically, it is unclear whether erroneous inflectional production arises from deficits in representation versus processing, and where within the production system any such deficits might be located.

In three experiments, L1 Mandarin and L1 English participants produced descriptions (Experiments 1 and 2: spoken; Experiment 3: written) of events involving different temporal contexts. As expected, in all three experiments L1 Mandarin participants produced fewer accurate inflections overall than L1 English participants in both Present Habitual and Past temporal contexts. However, like L1 English participants, L1 Mandarin participants' production of inflections was sensitive to temporal context, so that they were more likely to

produce inflections that were appropriate for the temporal context than inappropriate for the temporal context (i.e., more likely to produce 3SG -*s* in the Present Habitual context than in the Past context, and more likely to produce past -*ed* in the Past context than in the Present Habitual context), thus discriminating the locus (loci) of inflectional errors (see next section for discussion). This pattern held for both 3SG -*s* and past -*ed*, with a higher proportion of past -*ed* produced appropriately across all three experiments than 3SG -*s,* and held across both spoken and written production.

The Locus of L2 Inflectional Errors

These results are informative about the nature of L2 speakers' erroneous inflectional production within a processing model of language production and cast light on previous accounts that have been proposed from a theoretical linguistic perspective. At the start of this paper, we outlined possible representational and processing sources of errors in L2 inflectional production within current models of language production (Levelt, 1989; Levelt et al., 1999), together with linguistic accounts of L2 inflectional errors (Hawkins & Chan,1997; Prévost & White, 2000; Goad et al., 2003). These include conceptualization failures, missing or inconsistent activation of lemma level diacritic features, missing or inconsistent activation of syntactic dependencies, missing or inconsistent retrieval of inflectional morphemes and articulation failures. Our findings are consistent with some but not all of these accounts. Taken as a whole, they provide evidence that errors in L2 inflectional production more likely reflect processing breakdowns rather than representational deficits.

Broadly speaking, our findings refute representational deficit accounts of inflectional errors at the conceptualization and lemma level (Table 1 – Stages 1, 2a and 3a). Firstly, refuting a conceptualization deficit account (which predicted absolute omission across the board regardless of modality), our L1 Mandarin speakers were sensitive to L2 temporal

distinctions at the level of conceptualization and showed above chance inflectional accuracy in the spoken and written modalities (Exp. 1 - 51%; Exp. 2 - 54% and Exp. 3 - 66%). They were also sensitive to temporal distinctions appropriate for individual inflections (3SG *-s* and *-ed*). Secondly, refuting representational deficits in the form of missing diacritic features (e.g. **TENSE**) at the lemma level (Stage 2a), our L1 Mandarin speakers systematically (but variably) produced both 3SG *-s* and past *-ed* consistent with temporal context, indicating that they had acquired both temporal and syntactic distinctions for subsequent retrieval of morphological forms. This is in keeping with previous account of 'optional' inflectional production in second language acquisition research (i.e. sometimes producing and sometimes omitting the appropriate inflections). Thirdly, refuting representational deficits in the form of missing syntactic dependencies (e.g., activating the values **3RD** and **SINGULAR** for a verb lemma's **PERSON** and **NUMBER** diacritic features respectively, following a 3rd person singular subject), our L1 Mandarin speakers were able to appropriately use 3SG *-s* in the correct syntactic contexts, thus demonstrating that they had established appropriate syntactic dependencies for subject number (even though 3SG *-s* production was poorer overall).

Hence our data clearly demonstrate that erroneous inflectional production was not the result of failure to acquire relevant diacritic representations or syntactic dependencies. As such, they argue against Hawkins and Chan's (1997) *Failed Functional Feature Hypothesis*, which claimed that it was not possible for L2 speakers to acquire grammatical features that do not exist in the speaker's L1. Our findings demonstrate that L2 speakers of English whose L1 does not use inflectional morphology were capable of acquiring L2 temporal distinctions and inflectional features after the critical period (all our participants acquired L2 English after age five). Instead, our results are in line with accounts that attribute inflectional error to processing breakdowns. Within a psycholinguistic model of production, inflectional errors could be caused by a processing breakdown which failed to activate the relevant diacritic

features at the lemma level and the appropriate syntactic dependencies, which in turn would lead to failure to activate the corresponding morphological forms (in the same way as speech errors can arise in L1 production; Dell et al., 1997; Stages 2b, 3b and 4b). This account would predict that speakers' production of specific morphological forms would be sensitive to temporal context, but that it would be susceptible to error especially under processing load (e.g., time constraints), and would show an effect of featural complexity. Particularly, considering the transmission of activation between lemma level representations and inflectional morphemes, inflectional morphemes involving more complex features (e.g., distinctions based on both subject number and tense) would rely on accurate transmission of activation from multiple feature nodes, making successful retrieval less likely. This stands in contrast with inflections involving singular or less complex features (e.g. tense only), which only require transmission of activation from one feature node, making successful retrieval more likely. To put it another way, the probability of successful activation (retrieval) is reduced with each additional feature (e.g. if the probability of activating a single feature is *p*, the probability of activating multiple features would reduce to *p***p*).

Our findings are compatible with this account. 3SG -*s*, requiring both subject number and tense activation, was indeed less likely to be successfully retrieved (and therefore more frequently omitted) than past -*ed*, which requires only tense activation. These findings therefore support Hawkins' (2007) account of featural complexity, and are consistent with data from Chondrogianni and Marinis (2012), where 3SG -*s* was found to be more difficult to produce accurately than past -*ed*.

This finding can be considered in conjunction with theories that seek to account for L1 speech errors, for example Dell (1997). In Dell's theory, patterns of L1 aphasic speech errors are accounted for in part by inappropriate weights between connections that impact on accurate transmission of activation. In the same way, patterns of L2 errors may have a partial

source in inappropriately weighted connections between feature nodes for activating the correct syntactic dependencies or morphological forms where context requires them. For example, the transfer of number information from the noun phrase to the verb requires appropriate activation of syntactic nodes for subject-verb agreement (Eberhard et al., 2005). In the case of L2 speakers, such inappropriate weights between node connections might be the result of L1 transfer or a lack of learning (see Poulisse, 1999 for a review). On a broader level, inappropriately weighted connections at multiple stages of production (lexical, semantic, phonological) may reduce the probability of successful retrieval by exacerbating activation failures at previous stages (see also Budd et al., 2011).

Our findings are also compatible with Prévost and White's (2000) Missing Surface Inflection Hypothesis. Under this account, inaccurate or optional production of inflections was attributed to difficulties in morphological encoding and activating the relevant morphological forms rather than representational deficits. By implication, this indicates fundamentally a processing difficulty where inflections involving more complex features (i.e., 3SG -*s*) should be more susceptible to this difficulty. In our study, L1 Mandarin participants still processed L2 temporal and subject number information during L2 inflectional production (more 3SG -*s* responses in Present Habitual Singular Subject context than any other context). This suggested difficulties in activation and integration during retrieval of morphological forms rather than the lack of representations. Note that the current data cannot tease apart activation of features and retrieval of morphological forms.

Turning to an account locating L2 speakers' inflectional errors in articulation failures, production data across spoken (Exp. 2) and written (Exp. 3) modalities were especially revealing about the role of articulation in inflectional production (Stage 6). Despite the overall increase in inflectional accuracy by L1 Mandarin speakers of English in written compared with spoken production, inflectional error patterns from spoken production

persisted in written production even when no overt articulation was involved. This clearly indicated that articulation difficulties alone were not the primary cause of inflectional errors and cannot solely account for erroneous inflectional production. The source of inflectional error must primarily occur earlier on in the production process. However, this does not preclude potential phonological processing difficulties in the L2 that might contribute to higher incidence of errors in spoken production (Stage 5), especially when inflections also create additional syllables in our stimulus set. In other words, additional phonological / phonetic processing for articulation in spoken production may have reduced the probability of successful production, whereas this is not the case for written production. Additionally, the higher inflectional accuracy in the written modality (i.e. *'writing'* on a keyboard) might be linked to the increased response time permitted in Experiment 3 (Experiments 1 and 2: 7000ms; Experiment 3: 15000ms), which gave L1 Mandarin speakers more time to activate morphological representations and retrieve corresponding morphological forms.

Limitations and Future Directions

Experiments 1, 2 and 3 provide convincing evidence that patterns of inflectional errors are likely to be attributable to processing breakdowns during formulation rather than fundamental representational deficits, but do not distinguish whether these breakdowns are associated with activation of diacritic features, activation of syntactic dependencies and retrieval of morphological forms). Moreover, current evidence cannot tease apart possible difficulties in phonological processing from other forms of processing errors during L2 inflectional production. Assuming that phonological mediation occurs in both spoken and written production (Friederici, Schoenle, & Goodglass, 1981; Zhang & Damian, 2010), it remains plausible that differences between L1 and L2 phonological properties may have constrained the speaker's ability to encode (and syllabify) the relevant phoneme sequences correctly according to L2 prosodic frames.

We note that L1 Mandarin and L1 English groups produced *-ed* inflections in present habitual contexts (past habitual forms) across spoken and written modalities. Aside from possible incorrect interpretation of the temporal cue in our study (for both groups), one other explanation for the production of past habitual forms across groups is their high incidence of usage in spoken and written English. *-ed* being a way of grammaticalizing past events, regardless of aspectual properties, requires no subject number agreement and is more frequently used in speech and written language (including in participles in passive structures). Exposure to *-ed* usage in different contexts may have indirectly increased the frequency of production as well as led to better accuracy in our study. However, our current study do not allow us to tease apart frequency effects in inflectional production.

Our results are also informative regarding debates around the storage and processing of inflected forms (see Stemberger & Macwhinney, 1986; Ullman, 2001). As demonstrated by 'base-form' errors in our data, where inflectional errors most frequently take the form of omission (instead of errors of commission), there is good ground to argue that fully inflected forms are not stored in the lexicon (at least for our stimulus set) and are 'grammatically computed' on a rule-based principle during real-time production.

Our results also raise interesting questions about the role of explicit or metalinguistic knowledge in real-time production. Our L1 Mandarin speakers of L2 English all acquired English after the first critical period (AoA > 5 years), suggesting that lemma level representations (diacritic features), syntactic dependencies and their associated morphological forms may be established through explicit instruction. However, this does not necessarily mean L2 learners can activate these representations consistently during real-time production. In other words, explicit knowledge enables 'competence' at a representational level, but does not necessarily enable proficient 'performance' at a processing level.

A related question is how L2 learners respond to or learn from their own inflectional errors (e.g., omitting an obligatory inflection). Inflectional errors reoccur even when speakers intentionally attempt to avoid these errors in real-time L2 production. Why might L2 learners fail to learn from their own mistakes? One possibility could be lack of self-monitoring. Particularly, insufficient attentional resources (e.g., limited working memory capacity whilst carrying out the task) may diminish the L2 speaker's ability to monitor overt speech via the speech comprehension system (Levelt, 1989; Levelt et al., 1999; see Broos, Duyck, & Hartsuiker, 2016 for a review), directly contributing to the lack of conscious learning (Kormos, 1999). Another possibility could be down to the non-salient nature of inflectional morphemes as semantic or temporal cues. For example, temporal adverbials being a more salient cue to temporal information may overshadow the efficacy of inflectional morphemes in the same context, reducing the importance of inflectional errors in L2 speech (see Wulff & Ellis, 2018). Consequently, L2 speakers may not receive relevant feedback from inflectional errors, as communication is not impeded in these contexts.

A further question concerns how late L2 learners might acquire the relevant L2 knowledge that underpins their language production. Our results are consistent with the possibility that L2 speakers carry over functionally equivalent features from their L1, as proposed in Lardiere's (2008) Feature Reassembly Hypothesis. Under this account, L2 speakers initially use L1 features for L2 acquisition, before creating new features for the L2. In the context of our study, we may speculate that in the absence of tense inflections in the L1, L1 Mandarin participants may have viewed English tense inflections as functionally equivalent to Mandarin aspectual markers, and mapped past *-ed* onto the Mandarin aspectual marker *le* (which does not require subject number agreement).

Finally, we suggest that a future line of research might profitably consider the acquisition of other, morphologically richer L2s that may provide more data to support and inform a

processing account of inflectional errors. English is morphologically impoverished compared with other Indo-European languages (e.g., German, Dutch, Russian). Investigating such languages might therefore allow us to observe different types of L2 inflectional errors in L1 Mandarin speakers beyond the omission or incorrect production of a single morpheme. For example, for L2 words with multiple inflections, inflectional omission might in principle occur for some but not all of the obligatory inflectional morphemes. Moreover, we might see more selective commission errors whereby a more frequent inflectional morpheme is used in the place of a less frequent one.

Taking our results as a whole, they provide compelling evidence from both spoken and written L2 production that erroneous L2 inflectional production has a source in processing breakdowns, and specifically lemma and/or morpho-phonological processing, rather than a representational deficit. More importantly, we have shown how these findings can be interpreted within an approach that integrates both psycholinguistic models of language production and linguistic theories of L2 inflectional errors. We suggest that this reconciliation of theoretical perspectives allows deeper insight into the nature of L2 language production.

Acknowledgements

We would like to thank Professor Leah Roberts, Patrick Sturt and our three anonymous reviewers for their valuable comments and suggestions.

References

Bardovi-Harlig, K. (2000). Tense and Aspect in Second Language Acquisition: Form, Meaning, and Use. *Language Learning: A Journal of Research in Language Studies, 50*, 1.

Bayley, R. (1996). *Competing Constraints on Variation in the Speech of Adult Chinese Learners of English*. 97120. https://doi.org/10.1075/sibil.10.05bay

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review*, *89*(1), 1. https://doi.org/10.1037/0033-295X.89.1.1

Bock, K. (1995). *Sentence production: From mind to mouth*. In J. L. Miller & P. D. Eimas (Eds.), *Handbook of Perception and Cognition Vol. 11. Speech, language, and communication* (p. 181–216). Academic Press. https://doi.org/10.1016/b978-012497770-9/50008-x

Bock, K., & Levelt, W. (1994). *Language production: Grammatical encoding*. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (p. 945–984). Academic Press.

Broos, W. P. J., Duyck, W., & Hartsuiker, R. J. (2016). Verbal Self-Monitoring in the Second Language. *Language Learning*, *66*(September), 132–154. https://doi.org/10.1111/lang.12189

Budd, M. J., Hanley, J. R., & Griffiths, Y. (2011). Simulating children's retrieval errors in picture-naming: A test of Foygel and Dell's (2000) semantic/phonological model of speech production. *Journal of Memory and Language*, *64*(1), 74–87. https://doi.org/10.1016/j.jml.2010.08.005

Chondrogianni, V., & Marinis, T. (2012). Production and processing asymmetries in the acquisition of tense morphology by sequential bilingual children. *Bilingualism: Language and Cognition, 15*(1), 5-21. https://doi.org/10.1017/S1366728911000368

de Bot, K. (1992). A bilingual production model: Levelt's 'speaking' model adapted. *Applied linguistics*, *13*(1), 1-24.

de Bot, K. (2003). Bilingual speech: from concepts to articulation: Mündliche Produktion in der Fremdsprache. *Fremdsprachen Lehren und Lernen*, *32*, 92-103.

de Bot, K., & Schreuder, R. (1993). Word production and the bilingual lexicon. *The Bilingual Lexicon*, *191*, 214.

Dell, G. S. (1986). A Spreading-Activation Theory of Retrieval in Sentence Production. *Psychological Review*, *93*(3), 283. https://doi.org/10.1037/0033-295X.93.3.283

Dell, G. S., Chang, F., & Griffin, Z. M. (1999). Connectionist models of language production: Lexical access and grammatical encoding. *Cognitive Science*, *23*(4), 517-542. https://doi.org/10.1207/s15516709cog2304_6

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical Access in Aphasic and Nonaphasic Speakers. *Psychological Review,* 104, No. 4 (1997): 801. https://doi.org/10.1037/0033-295X.104.4.801

Eberhard, K. M., Cutting, J. C., & Bock, K. (2005). Making syntax of sense: Number agreement in sentence production. *Psychological Review*, *112*(3), 531–559. https://doi.org/10.1037/0033-295X.112.3.531

Foygel, D., & Dell, G. S. (2000). Models of Impaired Lexical Access in Speech Production. *Journal of Memory and Language*, *43*(2), 182-216. https://doi.org/10.1006/jmla.2000.2716

Friederici, A. D., Schoenle, P. W., & Goodglass, H. (1981). Mechanisms underlying writing and speech in aphasia. *Brain and Language*, *13*(2), 212–222. https://doi.org/10.1016/0093-934X(81)90091-2

Gardner, Q., Branigan, H., & Chondrogianni, V. (2018). Investigating the Production of L2 Inflectional Morphology Using a Picture Description Paradigm. SAGE Publications Ltd.

https://dx.doi.org/10.4135/9781526446657

Garrett, M. F. (1975). The analysis of sentence production. *Psychology of Learning and Motivation - Advances in Research and Theory*, (Vol. 9, pp. 133-177). Academic Press. https://doi.org/10.1016/S0079-7421(08)60270-4

Goad, H., White, L., & Steele, J. (2003). Defective syntax or L1-constrained prosodic representations? *Canadian Journal of Linguistics*, *48*(2), 243-263.

Hauck, W. W., & Donner, A. (1977). Wald's test as applied to hypotheses in logit analysis. *Journal of the American Statistical Association, 72*(360a), 851-853. https://doi.org/10.1080/01621459.1977.10479969

Hawkins, R. (2007). Emergent and innate sources of knowledge in the early L2 acquisition of English verbal morphology. In *Martin Atkinson – The minimalist muse* (pp. 135–156).

Hawkins, R., & Chan, C. Y. H. (1997). The partial availability of Universal Grammar in second language acquisition: The 'failed functional features hypothesis.' *Second Language Research*, *13*(3), 187-226. https://doi.org/10.1191/026765897671476153

Hawkins, R., & Liszka, S. A. (2003). *Locating the source of defective past tense marking in advanced L2 English speakers*. *Language Acquisition and Language Disorders*, *30*, 21-44. https://doi.org/10.1075/lald.30.03haw

Kormos, J. (1999). Monitoring and self-repair in L2. *Language Learning*, *49*(2), 303–342. https://doi.org/10.1111/0023-8333.00090

Lardiere, D. (1998). Case and Tense in the 'fossilized' steady state. *Second Language Research*, *14*(1), 1-26. https://doi.org/10.1191/026765898674105303

Lardiere, D. (2000). Mapping features to forms in second language acquisition. In *Second language acquisition and linguistic theory* (pp. 102–129).

Lardiere, D. (2003). Second Language Knowledge of [±Past] vs. [±Finite]. *6th Generative Approaches to Second Language Acquisition Conference (GASLA 2002)*, 176–189.

Lardiere, D. (2008). Feature Assembly in second language acquisition. In *The Role of Formal Features in Second Language Acquisition, 106*, 140. https://doi.org/10.4324/9781315085340-5

Levelt, W. J.M. (1989). *Speaking: From intention to articulation*. MIT Press.

Levelt, W. J.M. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(23), 13464-13471 https://doi.org/10.1073/pnas.231459498

Levelt, W. J.M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1-38. https://doi.org/10.1017/S0140525X99001776

Poulisse, N. (1999). *Slips of the Tongue in L2 Production 1*. 201–224. John Benjamins.

Prévost, P., & White, L. (2000). Missing Surface Inflection or Impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research*, *16*(2), 103–133. https://doi.org/10.1191/026765800677556046

Rainey, C. (2016). Dealing with separation in logistic regression models. *Political Analysis, 24*(3), 339-355. https://doi.org/10.1093/pan/mpw014

Roelofs, A. (2015). Modeling of phonological encoding in spoken word production: From Germanic languages to Mandarin Chinese and Japanese. *Japanese Psychological Research*, *57*(1), 22–37. https://doi.org/10.1111/jpr.12050

Slobin, D. I. (1987). Thinking for speaking. In *Annual Meeting of the Berkeley Linguistics Society* (Vol. 13, pp. 435-445).

Slobin, D. I. (1996). From “thought and language” to “thinking for speaking”. In J. J. Gumperz & S. C. Levinson (Eds.), *Studies in the social and cultural foundations of language, No. 17. Rethinking linguistic relativity* (p. 70–96). Cambridge University Press.

Smith, C. S. (1991). The Aspectual System of Mandarin Chinese. In *The Parameter of Aspect* (pp. 343–390).

Stemberger, J. P. (1998). Morphology in Language Production with Special Reference to Connectionism. *Hanndbook of Morphology*, 428-452.

Stemberger, J. P., & MacWhinney, B. (1986). Frequency and the lexical storage of regularly inflected forms. *Memory & Cognition*, *14*(1), 17-26.

Ullman, M. T. (2001). The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research*, *30*(1), 37–69. https://doi.org/10.1023/A:1005204207369

Wulff, S., & Ellis, N. C. (2018). Usage-based approaches to second language acquisition. In *Bilingual Cognition and Language* (pp. 37-56). John Benjamins. https://doi.org/10.1075/sibil.54.03wul

Zhang, Q., & Damian, M. F. (2010). Impact of phonology on the generation of handwritten responses: Evidence from picture-word interference tasks. *Memory and Cognition*, *38*(4), 519-528. https://doi.org/10.3758/MC.38.4.519

Zorn, C. (2005). A Solution to separation in binary response models. *Political Analysis*, *13*(2), 157-170. https://doi.org/10.1093/pan/mpi009

Appendices

Appendix A.

Experiments 1, 2 and 3: L1 Mandarin (L2 English) participant language background information.

	Experiment 1 (N=13)		Experiment 2 (N=37)		Experiment 3 (N=45)	
	M	SD	M	SD	M	SD
IELTS Overall*	7.19	0.38	6.93	0.43	7.03	0.42
IELTS Spoken* (Exp. 2)	-	-	6.28	0.55	-	-
Written* (Exp. 3)	-	-	-	-	6.92	3.14
AoA for L2 English (years)	10.69	2.43	9.51	2.39	8.82	2.37
Length of Stay (months)	4.15	2.51	8.51	1.35	6.08	3.17
L2 Contact (hours)	4.23	3.47	3.20	2.57	4.50	2.46

* IELTS scores are assessed on a scale of 0 – 9 in 0.5 increments for individual components of language ability and for the average overall score. Upper intermediate L2 English proficiency corresponds approximately to scores of 5.5 - 6.5 on the scale, and advanced L2 English proficiency corresponds approximately to scores of 7 – 8.5 on the scale.

Appendix B.

Experiments 1, 2 and 3: Experimental and filler verbs used in the scene description task.

Experimental	Filler
Shout	*Watch*
Wait	*Cook*
*Load**	*Write**
Start	*Listen*
End	*Drink*
*Applaud**	*Go**
Need	*Run*
*Print**	*Speak**
Paint	*Sit*

*Items not included in Experiment 3.