Heritage Speakers and Morphosyntactic Processing in Bilinguals
Shi Zhang

Recent research on heritage language (HL) acquisition has tremendously improved our understanding about how ethnolinguistically minority languages are acquired and how early bilingualism could affect the outcomes of language acquisition, but relatively few studies have been concerned with the theoretical aspects of HL morphosyntactic processing. This paper argues that investigation into issues such as whether HL processing mechanisms would become non-monolingual-like because of HL attrition and/or extensive second language (L2) exposure could help us construct a model capable of predicting and explaining HL morphoysntactic processing. Furthermore, it argues that testing the shallow structure hypothesis and its competing theories among HSs would enable us to determine the nature of the potential discrepancies between HL and monolingual L1 processing.

1. Introduction

In the last two decades, there has been an increasing interest in using psycholinguistic approaches to investigate language processing in bilinguals, with a focus on adult second language (L2) speakers. One central question in this line of research is whether there is any qualitative processing difference between monolinguals and bilinguals[1], as well as how to account for the possible differences.

In their influential work on adult L2 processing, Clahsen and Felser (2006a) proposed the shallow structure hypothesis (SSH), which claims that, firstly, the L1 processing mechanisms for morphology and syntax are essentially the same for children and adults, and secondly, L2 speakers do not exploit structural information at a monolingual-like level when processing complex L2 grammatical structures (e.g. anaphora resolution), and thus employ *shallow processing* in L2 processing (Felser & Cunnings, 2012; Felser & Roberts, 2007; Marinis, et al., 2005). The second claim has caused a big controversy since its publication, as some recent studies reported that advanced and near-native L2 speakers were able to employ monolingual-like processing for complex grammatical structures (e.g. E. Kim, et al., 2014). Many researchers argue that L2 speakers do not necessarily adopt shallow processing, and the observed L1/L2 differences could be explained by individual

---

[1] There is a range of views over what precisely constitutes a bilingual, a monolingual or a native speaker, and how far they are comparable (e.g. Cenoz 2013). This debate is beyond the scope of this article, so here, for ease of reference, I define monolingual as using one language predominantly from birth, and taken to represent a widely-used norm particularly in language processing research. Definitions of bilinguals and heritage speakers are detailed further below.

differences in cognitive abilities or memory retrieval operations (Cunnings, 2017; Hopp, 2015; also cf. Roberts, 2012). However, most of these studies on bilingual language processing only concerned adult L2 speakers who started L2 acquisition after puberty, and largely ignored early bilinguals, such as heritage speakers (HSs).

Recently, more and more researchers have started looking into the acquisition of heritage languages (HLs) by HSs. HLs usually refer to the non-societal and non-majority languages in a given social context, such as Spanish in the US (Valdés, 2005); HSs are often defined as bilinguals who have acquired the HL and the societal language naturalistically in early childhood (Pascual y Cabo & Rothman, 2012). Existing research has shown that HSs are highly heterogeneous in terms of the outcomes of HL acquisition: some adult HSs are almost indistinguishable from their monolingual peers in most linguistic domains, while some adult HSs can merely understand the HL (cf. Oh, et al., 2003; Rothman, 2007). These studies also suggest that factors such as onset age of bilingualism, language attrition and quality of input can affect the outcome of HL acquisition (Montrul, 2008; Polinsky, 2011; Rothman, 2007; 2009). From a psycholinguistic perspective, by studying HSs, we could tease out if early bilingualism would make HL processing different from monolingual L1 processing and further understand how factors like onset age of bilingualism and L2 influence would shape language processing in bilinguals.

Until now, there has been a number of studies on HL lexical processing, and an increasing number of studies on morphosyntactic processing in HSs (such as Knospe & Felser, 2016; Montrul, et al., 2014; Villegas, 2016); this paper will concentrate on morphosyntactic processing in HSs. Although empirical studies on the HL processing of morphosyntax are emerging, only limited studies have discussed the relevant theoretical aspects of HL processing, such as how to explain and predict the potential differences between HL and monolingual L1 processing. Bolger and Zapata (2011) attempted to do this by extending Ullman's (2001) Declarative/Procedural model, but as it will be pointed out later, their attempt is in fact problematic, because Ullman's model does not adequately serve the purpose of predicting whether and how the same grammatical structures would be processed differently by HSs and monolinguals.

This paper argues that, rather than extending an existing model, we could develop a working HL processing model by examining some issues deal with in other L1 and L2 processing models, such as Clahsen & Felser (2006a; 2006b), Cunnings (2017) and Dussias (2003; 2004). In doing so, it is not only possible to develop an HL processing model, but also possible to address some important issues in bilingual language processing, such as whether there is an interaction between the development of L1/L2 knowledge and L1 processing mechanisms in bilinguals.

The rest of this paper is organized as follows: Section 2 describes the complex routes and outcomes of HL acquisition, reviews a selection of existing studies on HL morphosyntactic processing, and provides a critical review of Bolger and Zapata's (2011) attempt to extend Ullman's (2001) model. Section 3 discusses how studying the issues concerned by several other L1/L2 processing models could help us develop an HL processing model. Finally, Section 4 provides a conclusion, and points out what other issues should be considered when constructing an HL processing model.

2. Heritage Language Acquisition and Processing

2.1 The Routes and Outcomes of Heritage Language Acquisition

As Carreira (2004) points out, precisely defining HSs has always been a problem, as HSs can be defined as anyone who have connections with a minority language on the one end, and someone who grew up learning the heritage language and has some proficiency in it on the other end (Benmamoun, et al., 2013). As this paper mainly concerns language processing, and language processing requires at least some proficiency in a language, the term HSs will refer to the second type of HSs here. I will define HSs based on Montrul's (2008) classification of bilinguals and Pascual y Cabo & Rothman's (2012) definition of HSs.

Based on the onset age and the sequence of language acquisition, Montrul (2008) identified two types of bilinguals: simultaneous bilinguals and sequential bilinguals. Simultaneous bilinguals acquire two languages simultaneously before the age of three, and sequential bilinguals start to acquire an L2 after having acquired some aspects of an L1. Sequential bilinguals are further categorized into early sequential bilinguals, who start to acquire an L2 between ages four to twelve, and late sequential bilinguals, who start to acquire an L2 after the age of twelve. As this classification enables us to easily label different bilinguals, it is adopted here, yet it is worth noting that extensive future research will be essential to validate the linguistic differences between each group of bilinguals.

Pascual y Cabo and Rothman (2012) define HSs as bilinguals who have naturalistically acquired an HL and a majority/societal language in early childhood. HSs always acquire the HL as an L1 through early exposure to input from their parents or other caregivers, and either acquire the majority language as one of their L1s, making them simultaneous bilinguals, or as an L2, making them early sequential bilinguals. As HSs always acquire the HL as an L1 naturalistically in early childhood, they should be deemed as L1 speakers of the HL. Despite being L1 speakers of the HL, depending on the route of HL acquisition, HSs can deviate from their monolingual peers in terms of linguistic knowledge and language use. For instance, some adult HSs only differ from monolinguals in comprehending and using some specific grammatical structures (Rothman 2007), but some adult HSs only possess limited HL proficiency (Oh, et al., 2003). These facts mean that HSs and monolingual L1 speakers can differ from each other in terms of the outcomes of L1 acquisition.

Existing research suggests that at least three factors have an impact on the outcomes of HL acquisition: onset age of bilingualism, quality of input and language attrition. Regarding the effects of onset age of bilingualism, Montrul (2008) argues that, if an HS begins acquiring the majority language before the age of eight, they are likely to receive reduced quantity and types of input, which will further result in what Montrul terms 'incomplete acquisition' of some aspects in the HL. For example, Montrul (2002) demonstrated that only simultaneous and early sequential HSs of Spanish were significantly different from monolinguals in interpreting and producing the perfective/imperfective tense-aspect morphology in Spanish, while late sequential bilinguals with L1 Spanish and L2 English were not. Similarly, J.-H. Kim, et al. (2009) showed that, in comparison to monolingual Korean speakers and late sequential bilinguals with L1 Korean and L2 English, simultaneous HSs of Korean in the US demonstrated less monolingual-like knowledge of Korean reflexives. Conversely, Kupisch, et al., (2014)

found no difference in understanding French gender marking, articles and prepositions between simultaneous HSs of French who grew up predominantly in Germany and simultaneous bilingual speakers of French and German who grew up predominantly in France, although the HSs had a less monolingual-like accent. This finding suggests that an early onset age of bilingualism may not always cause differences in linguistic knowledge between adult HSs and L1 speakers, and more finely-grained research is needed to reveal how other variables may also constrain HL acquisition.

Quality of input can also affect the outcomes of HL acquisition, leading to differences between HSs and monolinguals. For instance, an HS who is highly proficient in the HL may still differ from a monolingual in understanding or using some specific grammatical structure; this fact does not necessarily mean the HS has failed to fully acquired the HL, as the HL input may not even contain the target structures. Rothman (2007) examined advanced Brazilian Portuguese (BP) HSs' knowledge of inflected infinitives, of which the input is available in formal BP but not colloquial BP. It was found that, despite the fact that all the HSs heard and used BP daily, those HSs who had not received formal input in BP did not show monolingual-like knowledge of inflected infinitives, while the HSs who had received such input did. This finding was consolidated by Pires and Rothman (2009), who observed that HSs of European Portuguese who had not received formal education in European Portuguese had robust knowledge of inflected infinitives, which were always available in European Portuguese. Therefore, when HSs do not show monolingual-like knowledge for some specific grammatical properties, this HS/monolingual discrepancy may be caused by the quality of HL input.

Language attrition, or the loss of acquired linguistic knowledge, can also play a role in shaping the outcomes of HL acquisition. Polinsky (2011) found that while the child HSs of Russian demonstrated monolingual-like comprehension of subject and object relative clauses, the adult HSs of Russian did not show comparable comprehension for these relative clauses. This finding indicates that the erosion of once-acquired grammatical knowledge can take place among HSs, and affect the outcomes of HL acquisition. Unfortunately, it seems that few other studies have studied attrition among HSs in such a rigorous way, though there is a rich body of research on L1 attrition among late bilinguals (see Schmid 2016).

The reviewed studies suggest that, despite both being L1 speakers of a language, HSs and monolinguals may experience different routes and outcomes of HL acquisition. While there remains the possibility that HSs and monolinguals may also differ in language processing, only limited studies have yet examined this possibility. The following section will review a selection of these studies, focusing on research concerning HL morphosyntactic processing.

2.2 Studies on Heritage Language Processing

At present, studies on HL processing are emerging but still limited in number. Most of the existing studies focused on the processing of lexical items (e.g. Montrul, et al., 2014; Moreno & Kutas, 2005), but there are also a few studies on morphosyntactic processing, which are reviewed here.

Montrul (2006) investigated whether simultaneous HSs of Spanish in the US had monolingual-like knowledge and processing patterns for unergative and unaccusative verbs in Spanish and English. Based on Sorace's (2000) unaccusative hierarchy, Montrul

classified a selection of unergative and unaccusative Spanish and English verbs as core, less core and peripheral verbs. According to her predictions, during sentence processing, core unaccusative verbs should be processed faster than core unergative verbs, as unaccusative verbs do not need any movement of the subject while unergative verbs do (compare *John walked* with *John arrived __*). Meanwhile, as the less core and peripheral verbs are more subject to grammatical variability, speakers should be less determinate and slower in processing sentences containing these two types of verbs. The participants performed a grammaticality judgement task (GJT) and an on-line visual probe recognition task in Spanish and English respectively. The GJT required the participants to determine the acceptability of a series of sentences containing the target verbs, and assessed their knowledge of unaccusativity. The on-line visual probe recognition task required the participants to read sentences clause by clause, then determine if a specific word had appeared in the sentence they had just read, and finally answer questions based on the meaning of the sentences. The author observed that, in general, the HSs demonstrated monolingual-like knowledge of unaccusativity in both English and Spanish, and showed monolingual-like patterns for processing Spanish and English unergative and unaccusative verbs. However, in comparison to the Spanish and English monolinguals, the HSs had slower RTs in both languages.

Foote (2010) used a speeded production task to investigate if both simultaneous and early sequential HSs and late L2 learners of Spanish would show a distributivity effect for subject-verb agreement as monolingual Spanish speakers do (see Vigliocco, et al., 1996). If such an effect existed for all Spanish speakers, all the participants should produce more errors when the grammatical number and the conceptual number were mismatched (as in *the label on the bottles is yellow*, where *the label* is grammatically singular but conceptually plural) than when the grammatical number and conceptual number are matched (as in *the road to the mountains is long*). The participants were asked to watch a picture and listen to a sentence fragment (e.g. *the label on the bottles ...*), and complete the sentences by producing fragments like ... *is/are yellow* as quickly and accurately as possible. The results suggest that both groups demonstrated a distributivity effect, but the magnitude of this effect seemed to be modulated by their Spanish proficiency. However, this study did not include a monolingual Spanish control group or any RT measures, so it was not informative enough to suggest what on-line processing differences might lie between HSs and other Spanish speakers.

Keating, et al. (2016) studied Spanish HSs' on-line resolution of null and overt pronouns using a self-paced reading task. Two off-line tasks were also employed to assess the participants' knowledge of null and overt pronouns in Spanish. For the self-paced reading task, the researchers used sentences like *Después de que el sospechoso habló con el policía,* pro/*él admitió su culpabilidad* (After the suspect spoke with the policeman, *pro*/he admitted his guilt), and manipulated the verbs in the main clauses to create contexts biased to the subject or the object antecedent in the subordinate clauses. They examined if Spanish monolinguals and HSs would prefer to interpret the null pronoun as linked to the subject antecedent (*el sospechoso*) and interpret the overt pronoun as linked to the object antecedent (*el policía*) – if this was the case, the participants should process the sentences with null pronouns faster when the context was biased to the subject antecedent, and process the sentences with overt pronouns faster when the context was biased to the object antecedent. The data showed that the monolinguals and the HSs were similar in processing

the sentences with null pronouns, but only the monolinguals processed the sentences with overt pronouns faster when the context was biased to the object. As the off-line tasks suggested these HSs had monolingual-like mastery of null and overt pronouns in Spanish, this processing difference could not be attributed to differences in linguistic knowledge. The authors thus argued that this difference should be a quantitative one, which might be caused by the HSs' difficulties in integrating various types of linguistic information when processing pronouns.

These studies suggest that, when HSs and monolinguals show comparable knowledge for certain grammatical structures, their respective processings of these structures does not always resemble each other. While HSs show a monolingual-like processing pattern for early acquired morphosyntactic structures (e.g. unaccusativity in Spanish), they also show varied processing patterns for later acquired morphosyntactic structures (e.g. interpreting overt pronouns in Spanish). Although two of these studies suggested that HL proficiency and individual differences might correlate with HL and monolingual L1 processing differences, none of them have actually looked into the effects of these factors.

As HL morphosyntactic processing is a relatively new research area, it is not surprising that the above studies have only investigated a limited range of grammatical structures and languages. However, comparing to these empirical studies on HL morphosyntactic processing, even fewer studies have attempted to address the problem of how to predict and explain behaviours in HL morphosyntactic processing. In the following section, I will critically review Bolger and Zapata (2011), who attempt to provide a theoretical model of HL processing.

3. Towards a Heritage Language Processing Model

 3.1 Bolger and Zapata (2011)'s View on HL Processing

Bolger and Zapata (2011) suggested that Ullman's (2001) Procedural/Declarative model could be extended to explain the differences between HL and monolingual L1 processing. The Procedural/Declarative model suggests that, the representation and processing of L1 lexicon depend on declarative memory, which is subserved by temporal lobe structures, while the representation and processing of L1 grammar rely on procedural memory and are subserved by left frontal/basal-ganglia structures. In contrast, L2 grammatical representation and processing rely more heavily on declarative memory (and the underlying neural substrates) rather than procedural memory, and the degree of reliance on declarative memory is expected to increase with older onset age of L2 acquisition and less use of the L2.

Bolger and Zapata (2011) argued that, if an HS started acquiring an L2 early, this HS's HL grammar could be partly or wholly taken over by the L2 grammar, leading to a non-monolingual-like processing pattern for the HL. However, this extension of the Procedural/Declarative model is very problematic for the following reasons.

Firstly, the authors' prediction about the development of HL grammar is vague and difficult to falsify. Bolger and Zapata did not explain what "one grammar replacing another" means, or why and how an HL grammar could be replaced by an L2 grammar simply due to early onset of bilingualism. Based on their argument, it is difficult, if not impossible, to predict how an HS would process an HL if his/her HL grammar is "partly

or wholly taken over" by the L2 grammar. Moreover, it is doubtful whether an HL grammar could be wholly replaced by an L2 grammar, since few studies have observed the complete loss of an HL in HSs. For instance, Au, et al. (2008) reported that, in a series of tasks measuring productive and receptive grammatical knowledge, even HSs with limited HL proficiency and input were more monolingual-like than late sequential L2 learners with comparable proficiency, suggesting that it is unlikely for HSs to completely lose HL grammatical knowledge.

Secondly, the authors did not extend Ullman's model in the right direction. As Ullman (2001) is mainly concerned with how the neural substrates subserving procedural and declarative memory also subserve language processing, his model should be adapted to predict which neural substrates will be involved in HL and L1 processing respectively, rather than to predict any behavioural differences between HL and L1 processing, such as whether HSs and monolinguals would show the same level of processing speed and preference for interpreting null pronouns in Spanish.

Therefore, Bolger and Zapata's attempt to extend the Procedural/Declarative model is in fact problematic, as it lacks the power to predict and account for the potential differences between HL and monolingual L1 morphosyntactic processing. In the next section, I will argue that the theoretical issues concerned in a few other L1/L2 processing models, such as Clahsen & Felser (2006a; 2006b), Dussias (2003; 2004), Hopp (2015) and Cunnings (2017), could be examined in HL morphosyntactic processing research, and doing so will enable us to develop an HL processing model in the future.

 3.2 Developing a Heritage Language Processing Model

An appropriate model of HL processing should at least be able to predict whether and how HL morphosyntactic processing will be different from monolingual L1 processing, as well as to explain what factors could explain the potential discrepancies between HL and monolingual L1 processing.

In respect to predicting whether HL and monolingual L1 morphosyntactic processing would be different, we could examine if there is a *continuity* of HL processing, and if HL processing is subject to L2 influence. If any HL and monolingual L1 processing differences were observed, we could examine if the nature of such differences could be explained by the SSH and its competing hypotheses (Clahsen & Felser, 2006a; 2006b; Cunnings, 2017; Hopp, 2015), by investigating whether HSs, monolinguals and adult L2 speakers have similar abilities in utilizing structural information, integrating different types of information and retrieving cues from memory during morphosyntactic processing.

In the following section I will first look at how *the continuity of parsing hypothesis* and theories of L2 influence on L1 processing could be used to predict and explain whether there will be any differences between HL and monolingual L1 morphosyntactic processing. Then we will discuss how the SSH and its competing hypotheses could be used to explain the nature of the potential discrepancies between HL and monolingual L1 processing.

4. Issues in Bilingual Language Processing

4.1 Continuity of HL/L1 Processing

As mentioned in the introduction, the SSH consists of two hypotheses. The first hypothesis, i.e. *the continuity of parsing hypothesis* (Clahsen & Felser, 2006b: 107), assumes that, as long as children have acquired the relevant L1 knowledge for certain grammatical structures, they would resemble their adult counterparts in processing these grammatical structures. Moreover, the L1 processing mechanisms do not need to develop and do not change over time. Several studies suggested that children largely resembled adults in processing sentence ambiguity and syntactic dependencies in their L1s (Booth, et al., 2000; Felser, et al., 2003; Roberts, et al., 2006; Traxler, 2002). Clahsen and Felser (2006a) argued that the differences observed in children's and adults' L1 processing are largely cognitive, e.g. when children relied less on lexical-semantic information than structural information when resolving sentence ambiguity, this can be explained by their differences in working memory capacities. This hypothesis is in line with Fodor's (1998; 1999) proposal that the ability to parse an L1 does not require development.

However, as Clahsen and Felser themselves have pointed out, their evidence mainly comes from children older than six, who should in principle have acquired most knowledge in their L1 (2006b: 111). Therefore, there remains the possibility that children may experience a change of language processing mechanisms during early L1 acquisition. During sentence processing, very young children may have to mainly rely on lexical-semantic information as they have not acquired much syntax, but older children will rely mainly on structural information after having acquired the relevant syntax. However, this speculation has not yet been systematically examined.

Although *the continuity of parsing hypothesis* mainly concerns monolingual L1 processing, this hypothesis could also be tested among HSs, and it would allow us to examine whether there will be a continuity of HL processing. It is worth noting that, if an interaction between the development of linguistic knowledge and the development of processing mechanisms does exist, its effect should also be observable in HSs who experienced HL attrition. On the one hand, if an HS who experienced HL attrition only demonstrated changes in the grammatical knowledge, but not the processing patterns for the HL grammatical structures under attrition, it is likely that HL/L1 processing mechanisms indeed do not change over time, even when the once acquired L1 knowledge has been changed or lost. On the other hand, if changes in both the knowledge and the processing patterns for some specific HL grammatical structures were observed, the observation will implicate that L1 processing depends on L1 linguistic knowledge and needs development; then, at least for early bilinguals, Clahsen and Felser's claim that L1 processing do not change over time should be falsified.

Based on evidence gained from examining *the continuity of parsing hypothesis* among HSs, we could construct a model capable of predicting whether HL morphosyntactic processing will resemble monolingual L1 processing in terms of *development paths*. The next section will argue that, by examining whether L2 experience influence HL/L1 processing, we will be able to predict whether HL processing will resemble or differ from monolingual L1 processing because of *different language experience*.

4.2 L2 Influence on HL/L1 Processing

Although less commonly framed as a theoretical model, some researchers have hypothesized that a bilingual's L1 processing might become less monolingual-like if this bilingual has extensive L2 experience. Dussias (2003; 2004) studied whether exposure to L2 English might affect the resolution of relative clause attachment in L1 Spanish using self-paced reading and eye-tracking. She observed that, when encountering sentences such as *Peter fell in love with the daughter of the psychologist who studied in California*, Spanish monolinguals and Spanish-English bilinguals with limited exposure to English (average time of exposure = 3.7 years) preferred to interpret the relative clause as being attached to NP1 (*the daughter*), but English monolinguals and Spanish-English bilinguals with more exposure to English (average time of exposure = 11 years) preferred to interpret the relative clause as being attached to NP2. These findings suggest that extensive exposure to an L2 may affect L1 processing strategies, but it is not clear whether early bilinguals' L1 on-line processing would also be affected by extensive exposure to an L2. Furthermore, it is not clear whether this L2 influence on L1 processing could be found in both early bilinguals and late bilinguals.

As HSs are bilinguals exposed to an L2 since early childhood, they could provide us the ground for examining how early extensive L2 experience might affect HL/L1 processing. If input frequency plays a crucial role in shaping language processing, then HSs who are dominant in an L2 should be seriously influenced by the L2 when processing the HL. For instance, an English-dominant Spanish HS should demonstrate NP2 attachment preference for Spanish relative clause attachment. In contrast to this hypothesis, Jegerski, et al. (2014) reported that Spanish-English HSs demonstrated an NP1 attachment preference, suggesting that an L2 may not always affect L1 processing in early bilinguals. It is possible that early bilinguals have developed two separate processing strategies for HL/L1 and L2 from early on, and they are able to adopt monolingual-like processing for HL morphosyntactic structures.

If future studies concerning L2 influence on HL processing found more supportive evidence that HL morphosyntactic processing in early bilinguals is immune to L2 influence, we would be able to predict that HL morphosyntactic processing will resemble monolingual L1 processing, regardless of HSs' and monolinguals' different language experience; otherwise, the opposite prediction could be made. In either case, we could construct a model with falsifiable predictions.

4.3 Language Processing Differences in Bilinguals

Enquiries into the above two issues will allow us to predict whether the continuity of HL processing and L2 influence would lead to differences between HL and monolingual L1 processing because of differences in *development paths* and *language experience*. In this section, I argue that the theoretical approaches of Clahsen and Felser (2006a), Hopp (2015) and Cunnings (2017) can be used to predict if the potential differences between HL and monolingual L1 processing are different in nature, i.e. qualitative or quantitative.

Existing studies comparing L1 and L2 processing have observed that L1 speakers and advanced/near-native late L2 speakers differed in processing complex morphosyntactic structures (e.g. Felser & Cunnings, 2012; Felser & Roberts, 2007; Marinis, et al., 2005). For instance, Felser and Cunnings (2012) employed eye-tracking to study the processing of English reflexives (*herself/himself*) by L1 speakers and German L2 learners of English.

They found that, when encountering sentences like *James has worked at the army hospital for years. He noticed that the soldier had wounded himself while on duty in the Far East* (Felser & Cunnings, 2012: 579), the L2 learners considered the grammatically inaccessible but discourse-prominent matrix antecedent (i.e. *James*) as a possible resolution during the early processing while L1 speakers did not. According to the second hypothesis of the SSH (a brief description of this hypothesis is provided in the introduction), such findings suggest L2 speakers underuse structural information and employ *shallow processing* for an L2, and indicate a qualitative difference between L1 and L2 processing.

However, recent studies have found that late L2 speakers were able to perform like L1 speakers in processing various structures. For instance, Hopp (2015) observed monolingual-like performance in using morphosyntactic information to resolve syntactic ambiguity among German L2 speakers of English who were good at exploiting syntactic information encoded in lexical items, and argued that L1/L2 processing differences might not be qualitative, but quantitative if individual differences in cognitive abilities were taken into consideration.

Furthermore, an eye-tracking study by E. Kim, et al. (2014) reported that Korean L2 speakers of English demonstrated monolingual-like processing for English reflexives when the reflexives were discourse-prominent; meanwhile, they differed from L1 speakers in processing English pronouns by considering the grammatically inaccessible but discourse-prominent antecedent as a possible resolution. Cunnings (2017) suggests that L1 and L2 speakers may differ in memory retrieval operations during language processing. In the case of anaphoric resolution, L1 speakers predominantly prefer to retrieve an antecedent which matches the syntactic constraint cue required by the reflexives, whereas L2 speakers prefer to retrieve an antecedent which matches the semantic/discourse cue required by the reflexives. Such difference implicates differences in L1/L2 processing preferences, but not inevitable L2 shallow processing.

When HSs differed from monolingual L1 speakers in processing certain HL morphosyntactic structures, it is also worth investigating whether the difference was due to that HSs lack the ability of utilizing structural information at a monolingual-like level (and thus qualitative), or due to individual differences in cognitive abilities and/or memory retrieval strategies (and thus quantitative). In doing so, we will be able to predict whether HL morphosyntactic processing would become qualitatively, or simply quantitatively, different from monolingual L1 processing, and determine whether HL processing and monolingual L1 processing are different in nature.

5. Conclusion

This paper has argued that studying a series of issues in L1 and L2 processing among HSs could help us to develop a working model for predicting and explaining HL morphosyntactic processing. HSs are bilinguals with diverse onset ages of bilingualism and heterogeneous routes of HL/L1 acquisition, and studies suggested that HSs might show non-monolingual-like processing for some HL morphosyntactic structures.

By looking into the continuity of HL processing and L2 influence on HL processing, we could develop a model capable of predicting whether HL morphosyntactic processing would differ or resemble monolingual L1 processing based on the HSs' and the

monolinguals' difference in development paths and language experience. Meanwhile, by investigating HSs' ability of utilizing structural information and individual differences in cognitive abilities and memory retrieval operations, it is possible to develop a model that could explain and predict the nature of the potential discrepancies between HL and monolingual L1 morphosyntactic processing.

This paper has only discussed limited issues worth investigating when developing an HL processing model. Other issues, such as whether the properties of different linguistic structures, HL and L2 proficiency would affect HL processing, are also worth considering when constructing an HL processing model. The effect of such issues could be explored and discussed in the future.

References

Au, T. K., Oh, J. S., Knightly, L. M., Jun, S.-A., & Romo, L. F. (2008). Salvaging a Childhood Language, *Journal of Memory and Language*, **58**(4): 998-1011.

Benmamoun, E., Montrul, S., & Polinsky, M. (2013). Heritage Languages and their Speakers: Opportunities and Challenges for linguistics. *Theoretical Linguistics*, **39**(3-4): 129-81.

Bolger, P. A., & Zapata, G. C. (2011). Psycholinguistic Approaches to Language Processing in Heritage Speakers. *Heritage Language Journal*, **8**(1): 1-29.

Booth, J. R., Mac Whinney, B., & Harasaki, Y. (2000). Developmental Differences in Visual and Auditory Processing of Complex Sentences. *Child Development*, **71**(4): 981-1003.

Carreira, M. (2004). Seeking Explanatory Adequacy: A Dual Approach to Understanding the Term "Heritage Language Learner". *Heritage Language Journal*, **2**(1): 1-25.

Cenoz, J. (2013). Defining Multilingualism. *Annual Review of Applied Linguistics*, **33**: 3-18.

Clahsen, H., & Felser, C. (2006a). Grammatical processing in language learners. *Applied Psycholinguistics*, **27**(1): 3-42.

Clahsen, H., & Felser, C. (2006b). Continuity and shallow structures in language processing. *Applied Psycholinguistics*, **27**(1): 107-26.

Cunnings, I. (2017). Interference in Native and Non-Native Sentence Processing', *Bilingualism: Language and Cognition*, 1-10. Available at: < https://www.cambridge.org/core/journals/bilingualism-language-and-cognition/article/interference-in-native-and-nonnative-sentence-processing/75C3F27EBEE7309BB163B0C393C3CB54> [accessed on: 28.05.2017].

Dussias, P. E. (2003). Syntactic ambiguity resolution in L2 learners: Some effects of bilinguality on L1 and L2 processing strategies, *Studies in Second Language Acquisition*, **25**(4): 529-557.

Dussias, P. E. (2004). Parsing a first language like a second: The erosion of L1 parsing strategies in Spanish-English Bilinguals. *International Journal of Bilingualism*, **8**(3): 355-71.

Felser, C., & Cunnings, I. (2012). Processing reflexives in a second language: The timing of structural and discourse-level constraints. *Applied Psycholinguistics*, **33**(3): 571-603.

Felser, C., Marinis, T., & Clahsen, H. (2003). Children's Processing of Ambiguous Sentences: A Study of Relative Clause Attachment. *Language Acquisition*, **11**(3): 127-63.

Felser, C., & Roberts, L. (2007). Processing wh-dependencies in a second language: a cross-modal priming study. *Second Language Research*, **23**(1): 9-36.

Fodor, J. D. (1998). Learning to Parse?. *Journal of Psycholinguistic Research*, **27**(2): 285-319

Fodor, J. D. (1999). Learnability Theory: Triggers for Parsing with. In: Klein E. C. & Martohardjono G. (eds). *Language Acquisition and Language Disorders 18*. John Benjamins: Amsterdam, pp. 363-406.

Foote, R. (2010). Age of acquisition and proficiency as factors in language production: Agreement in bilinguals. *Bilingualism: Language and Cognition*, **13**(2): 99-118.

Hopp, H. (2015). Individual difference in the second language processing of object-subject ambiguities. *Applied Psycholinguistics*, **36** (2): 129-73.

Jegerski, J., Keating, G. D., & VanPatten, B. (2014). On-line relative clause attachment strategy in heritage speakers of Spanish. *International Journal of Bilingualism*, **20**(3): 254-268.

Keating, G. D., Jegerski, J., & VanPatten, B. (2016). Online processing of subject pronouns in monolingual and heritage bilingual speakers of Mexican Spanish. *Bilingualism: Language and Cognition*, **19**(1): 36-49.

Kim, E., Montrul, S., & Yoon, J. (2014). The on-line processing of binding principles in second language acquisition: Evidence from eye tracking. *Applied Psycholinguistics*, **36**(6): 1317-1374.

Kim, J.-H., Montrul, S., & Yoon, J. (2009). Binding Interpretations of Anaphors by Korean Heritage Speakers. *Language Acquisition*, **16**(1): 3-35.

Knospe, G.-M., & Felser, C. (2016). The role of proficiency in Turkish heritage speakers' processing of pronouns. Paper presented at the Heritage Language Acquisition: Breaking New Ground in Methodology and Domains of Inquiry, Tromsø. Available at: < http://site.uit.no/lava/files/2016/07/Knospe-Felser-HLA-workshop-2016.pdf> [accessed on: 28/05/2017].

Kupisch, T., Lein, T., Barton, D., Schröder, D. J., Stangen, I., & Stoehr, A. (2014). Acquisition outcomes across domains in adult simultaneous bilinguals with French as weaker and stronger language. *Journal of French Language Studies*, **24**(3): 347-76.

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in Second Language Sentence Processing. *Studies in Second Language Acquisition*, **27**(1): 53-78.

Montrul, S. (2002). Incomplete acquisition and attrition of Spanish tense/aspect distinctions in adult bilinguals. *Bilingualism: Language and Cognition*, **5**(1): 39-68.

Montrul, S. (2006). On the bilingual competence of Spanish heritage speakers: syntax, lexical-semantics and processing. *International Journal of Bilingualism*, **10**(1): 37-69.

Montrul, S. (2008). *Incomplete Acquisition in Bilingualism: Re-Examining the Age Factor*. Amsterdam: John Benjamins.

Montrul, S., Davidson, J., de la Fuente, I., & Foote, R. (2014). Early language experience facilitates the processing of gender agreement in Spanish heritage speakers. *Bilingualism: Language and Cognition*, **17**(1): 118-38.

Moreno, E. M., & Kutas, M. (2005). Processing semantic anomalies in two languages: an electrophysiological exploration in both languages of Spanish–English bilinguals. *Cognitive Brain Research*, **22**(2): 205-20.

Oh, J. S., Jun, S.-A., Knightly, L. M., & Au, T. K. (2003). Holding on to childhood language memory. *Cognition*, **86**(3): B53-64.

Pascual y Cabo, D., & Rothman, J. (2012). The (Il) Logical Problem of Heritage Speaker Bilingualism and Incomplete Acquisition. *Applied Linguistics*, **33**(4): 450-455.

Pires, A., & Rothman, J. (2009). Disentangling sources of incomplete acquisition: An explanation for competence divergence across heritage grammars. *International Journal of Bilingualism*, **13**(2): 211-38.

Polinsky, M. (2011). Re-Analysis in Adult Heritage Language. *Studies in Second Language Acquisition*, **33**(2): 305-28.

Roberts, L. (2012). Individual Differences in Second Language Sentence Processing. *Language Learning*, **62**: 172-88.

Roberts, L., Marinis, T., Felser, C., & Clahsen, H. (2006). Antecedent Priming at Trace Positions in Children's Sentence Processing. *Journal of Psycholinguistic Research*, **36**(2): 175-88.

Rothman, J. (2007). Heritage speaker competence differences, language change, and input type: Inflected infinitives in Heritage Brazilian Portuguese. *International Journal of Bilingualism*, **11**(4): 359-89.

Rothman, J. (2009). Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International Journal of Bilingualism*, **13**(2): 155-63.

Schmid, M. S. (2016). First Language Attrition. *Language Teaching*, **49**(2): 186-212.

Sorace, A. (2000). Gradients in Auxiliary Selection with Intransitive Verbs, *Language*, **76**(4): 859-90.

Traxler, M. J. (2002). Plausibility and subcategorization preference in children's processing of temporarily ambiguous sentences: Evidence from self-paced reading. *The Quarterly Journal of Experimental Psychology Section A*, **55**(1): 75-96.

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: the declarative/procedural model. *Bilingualism: Language and Cognition*, **4**(2): 105-22.

Valdés, G. (2005). Bilingualism, Heritage Language Learners, and SLA Research: Opportunities Lost or Seized?. *The Modern Language Journal*, **89**(3): 410-26.

Vigliocco, G., Butterworth, B., & Garrett, M. F. (1996). Subject-verb agreement in Spanish and English: Differences in the role of conceptual constraints, *Cognition*, **61**(3): 261-98.

Villegas, Á. (2016). What can an eye-tracker tell us about the linguistic knowledge of heritage speakers?. Presented at the Heritage Language Acquisition: Breaking New Ground in Methodology and Domains of Inquiry, Tromsø. Available at: < http://site.uit.no/lava/files/2016/07/Villegas.pdf> [accessed on: 28.05.2017].

Bio

Shi Zhang is a PhD student at the University of Reading, and his current work uses eye tracking to examine the on-line processing of perfective/durative aspect marking and reflexive binding among Mandarin-English bilingual speakers. His research interests include language processing, language attrition and language acquisition.