

BUCLD 42 Proceedings
To be published in 2018 by Cascadilla Press
Rights forms signed by all authors

L2 Listeners Rely on the Semantics of Classifiers to Predict

Theres Grüter, Elaine Lau, and Wenyi Ling*

Prediction, or the use of currently available information to create expectations about what will happen next, is an integral part of human cognition and behavior, including language processing (see Kuperberg & Jaeger, 2016, for review). There is ample evidence that native (L1) speakers make incremental use of linguistic information of all kinds – phonological, morphosyntactic, semantic – to create probabilistic expectations about how words, sentences and discourses will continue. More recently, the question has been raised as to the extent to which these findings generalize to non-native language (L2) processing. In other words, to what extent do L2 listeners (or readers) use information in the incoming signal to incrementally update their expectations about what they will hear (or read) next? Early studies on predictive processing with L2 learners, which adapted experimental designs previously used with L1 speakers, reported no or reduced effects in L2 compared to L1 groups (Kaan et al., 2010; Lew-Williams & Fernald, 2010; Martin et al., 2013). Findings such as these prompted Grüter et al. (2014, 2017) to propose that L2 speakers may have *Reduced Ability to Generate Expectations (RAGE)* during language processing. The goal of the RAGE hypothesis was to instigate further research aimed at investigating how and under what circumstances predictive processing in an L2 may differ from what had been observed in L1 processing. More recent studies have shown clear effects of prediction in some groups of L2 learners under certain circumstances (see Hopp & Lemmerth, 2016, for review), thus making clear that the question cannot be whether L2 learners do or do not predict, but what the modulating factors in L2 prediction may be (see also Kaan, 2014).

Here we explore one potential source of differences, namely the relative weight that L1 and L2 listeners allocate to different types of linguistic cues to predict what will come next. More specifically, we investigate how L1 and L2 speakers of Mandarin Chinese use information encoded by prenominal sortal classifiers (e.g., 一条绳子, ‘one CL_{LONG-THING} rope’) to create expectations about upcoming nouns. Based on previous research discussed in more detail

* Theres Grüter, University of Hawai‘i at Mānoa, theres@hawaii.edu; Elaine Lau, Chinese University of Hong Kong, elaine@cuhk.edu.hk; Wenyi Ling, University of Hawai‘i at Mānoa, wenyi9@hawaii.edu. We are grateful to Craig Chambers for sharing the materials from Tsang & Chambers (2011) with us, and to Victoria Lee for assistance with data collection. This study was supported by a grant from *Language Learning’s* Small Grant Research Program to TG.

below, we hypothesized that L1 speakers would rely primarily on information about form class – learned through extensive experience with co-occurrence relations in the input – while L2 speakers would rely more heavily on the semantics associated with classifiers themselves. To test this hypothesis, we adapted the materials from a previous Visual-World eye-tracking study with L1 speakers of Cantonese (Tsang & Chambers, 2011; Experiment 3) for L1 and L2 speakers of Mandarin. Here we report the results from 24 L1 speakers as well as an initial cohort of 19 L2 speakers of Mandarin.

1. Classifiers in Chinese

Classifiers – also known as measure words – are free morphemes marking the class of the noun they co-occur with. Their presence is obligatory when the noun phrase includes a numeral or demonstrative, as illustrated in (1).

- (1) 一 条 绳子
yī * (tiáo) shéngzi
one CL_{LONG-THING} rope
‘a rope’

Sortal classifiers, the type of interest here, categorize nouns into classes broadly based on inherent properties of the object they denote, such as shape, natural kind, and function (Gao & Malt, 2009). While several hundreds of classifiers have been documented in Mandarin Chinese (Chen et al., 1988), according to a recent estimate, approximately 50 to 70 sortal classifiers are used frequently in modern Chinese (Li, 2013). The number of nouns associated with a particular classifier varies greatly, and although classifiers are generally associated with certain semantic properties, nouns that belong to the same classifier class do not necessarily form a homogeneous semantic set. The classifier *tiáo*, for instance, is generally associated with nouns that conform to the description “slender, long-shape thing, often flexible” (Gao & Malt, 2009, p. 1171); yet not all nouns that co-occur with *tiáo* fit this description to the same degree. For example, both *shéngzi* ‘rope’ and *gǒu* ‘dog’ are members of the *tiáo* class, yet only the former straightforwardly fits the semantic criteria associated with the class. Put in terms of incremental processing, the *semantics* associated with *tiáo* is predictive of *shéngzi* ‘rope’ to a greater degree than it is of *gǒu* ‘dog’. However, if predictions are based purely on class membership as determined by co-occurrence statistics in the language, then *tiáo* should be equally predictive of both.

Classifiers are both similar to and different from gender-marked determiners found in many Indo-European languages, which have been the focus of several previous studies on predictive processing in L1 and L2 (see Section 2). Like gender-marking on a determiner, pronominal classifiers allow the listener to narrow down the set of potential nouns that might follow. Unlike gender classes, however, where semantic criteria play no role in class assignment for

inanimate nouns, there is a significant, though not exhaustive, semantic component to the noun class system in classifier languages like Chinese, as illustrated above. In this study, we take advantage of the fact that classifiers can present informative cues about upcoming nouns at two related but not isomorphic levels: (i) semantic, and (ii) purely form-class, or co-occurrence related. For prototypical members of a class, such as *shéngzi* ‘rope’, both cues will be effective. In the case of non-prototypical members, such as *gǒu* ‘dog’, however, only the latter will be effective; i.e., if only the semantics of *tiáo* is used to activate potentially upcoming nouns, ‘dog’ is unlikely to be among them. Moreover, reliance on the semantics of the classifier could lead to the activation of nouns that possess relevant semantic features, but do not happen to be part of that classifier class. The experimental design of our study – borrowed from Tsang and Chambers (2011) – takes advantage of the potential for such dissociations to investigate listeners’ relative reliance on the two levels of informativity of classifiers, dissociations that do not arise in the context of noun classes marked by grammatical gender.

2. Use of prenominal cues to upcoming nouns in L1 and L2 processing

Gender marking on prenominal determiners has been shown to facilitate L1 speakers’ processing of nouns in a number of mostly Indo-European languages (e.g., Grosjean et al., 1994; Dahan et al., 2000; Lew-Williams & Fernald, 2007). L1 speakers of classifier languages appear to make similar use of prenominal classifiers. In a visual world study, Huettig et al. (2010) showed that L1 Mandarin speakers were faster to look at a target object when it was named preceded by an informative classifier, i.e., a classifier compatible only with the target noun in the visual scene, than when no classifier was present in the linguistic signal (see also Klein et al., 2012). In two visual world experiments with L1 speakers of Cantonese, Tsang and Chambers (2011) further probed to what extent the facilitatory effect of classifiers on referential processing was due to *semantic* information pertaining to shape encoded by sortal classifiers, versus to what they termed “grammatical constraints”, i.e., knowledge of class-membership based on accumulated knowledge of co-occurrence relations, irrespective of meaning. Their findings indicated that semantic information did not play a strong role when the target object was a prototypical member of the class (i.e., displaying all its defining semantic features, e.g., long, narrow and flexible, for the classifier *tiu4*). In this case, upon hearing the classifier *tiu4*, there was no increase in listeners’ looks to objects that possessed relevant semantic features but did not belong to the *tiu4* class (e.g., *kei4* ‘flag’, which can be described as long and flexible, but cannot co-occur with *tiu4*). Only competitors from the same classifier class attracted looks; among those, slightly stronger effects were observed for prototypical than for non-prototypical members of the class, indicating that semantic information contributed to listeners’ expectations, but only within the set of class-consistent nouns. In a second experiment, Tsang and Chambers (2011) sought to increase the potential

for competition by including targets that themselves were not prototypical members of their class. In this case, participants attended somewhat more to competitors that displayed prototypical semantic features, including competitors not from the same classifier class; yet competition from classifier-consistent competitors remained overall stronger. The authors concluded that shape classifiers influence L1 speakers' predictive processing "primarily through their grammatical constraints" (p. 1065), with classifier semantics acting as secondary cue that becomes apparent only in certain circumstances, such as when non-prototypical members of a class are involved. In the study reported here, we followed the design of Tsang and Chambers' (2011) second visual world experiment (their Experiment 3) to create an analogous experiment in Mandarin, with the goal of testing our hypothesis that L2 learners rely more strongly on semantic information encoded by classifiers than L1 speakers.

Previous work probing potentially facilitatory effects of prenominal cues in L2 processing has produced mixed results. Several studies reported no facilitatory effects of gender-marking on determiners on L2 learners' processing of inanimate nouns (where semantic information is irrelevant), including studies with highly proficient L2 learners (e.g., Guillelmon & Grosjean, 2001; Grüter et al., 2012). Others have found facilitation under certain circumstances, notably for learners whose L1 also marks gender (Dussias et al., 2013; Hopp & Lemmerth, 2016), for learners who independently demonstrated fast lexical access speed in the L2 (Hopp, 2013), or who had acquired lexical gender assignment with little remaining variability in the system (Hopp, 2016). Of note, in an experiment otherwise identical to the visual world studies that showed no facilitatory effects with intermediate and near-native L2 learners of Spanish (Lew-Williams & Fernald, 2010; Grüter et al., 2012), Lew-Williams and Fernald (2009, Experiment 3) included animate, gendered nouns, such as *la niña* ('the_{FEM} girl'), and observed a clear facilitatory effect for both L1 and intermediate-proficiency L2 speakers of Spanish. Thus it appears that even intermediate-proficiency L2 learners were able to make use of the *semantic* information encoded by the gender-marked article, but they were not able to exploit the more abstract cue signaling noun-class membership in the case of inanimate nouns during online processing.

In a visual-world study with L2 learners of Mandarin Chinese, using a design closely following that used by Lew-Williams and colleagues, Lau and Grüter (2015) reported evidence suggesting that intermediate-proficiency L2 learners of Chinese were able to take advantage of information encoded on the classifier to facilitate processing of the subsequent noun. Based on this finding, and referring to Lew-Williams and Fernald's (2009) findings from semantic gender marking, Lau and Grüter (2015) hypothesized that L2 learners of Chinese benefit from the *semantic* informativity of classifiers, a cue that is not available with gender-marked determiners, except in the rare case of nouns denoting animate, gendered entities.

3. Hypothesis and predictions

The present study was directly motivated by Lau & Grüter's (2015) findings, and their suggestion that L2 learners' of Chinese derive processing facilitation from classifiers due to their semantic informativity. The experimental paradigm used by Tsang and Chambers (2011, henceforth T&C), described above, offers an ideal scenario for testing this hypothesis. In particular, we predict that if L2 learners' of Chinese derive processing benefits from classifiers because they use their semantics to generate expectations about upcoming nouns, then upon hearing a classifier, they will direct their attention to *all* competitors that match the classifier's semantics, including those that are not from the same classifier class. T&C found no (Experiment 2) or only very small (Experiment 3) competition effects from class-inconsistent competitors with matching semantic features among L1 speakers of Cantonese. We predict that in a Mandarin adaptation of T&C's Experiment 3, these effects will be enhanced among L2 learners, and greater compared to those in L1 speakers of Mandarin.

4. Methods

4.1 Participants

Twenty-four L1 speakers of Mandarin Chinese ($M = 26$ years, range: 19-39) and 19 English-dominant L2 speakers of Chinese ($M = 28$ years, 21-57) participated in this study. All except two L2 participants were taking or had recently taken 3rd- or 4th-year Chinese classes at the University of Hawai'i. Their mean self-rated proficiency in Mandarin was 5.7/10 ($SD = 1.9$); their mean age of first exposure to Mandarin was 17 years ($SD = 12$). Four L2 participants reported some exposure to Chinese during childhood from one or both parents and are best characterized as heritage speakers. Yet all four indicated English as their dominant language, and none considered themselves native speakers of Chinese; all were taking Chinese language classes and had been placed in the same 3rd or 4th year Chinese classes as the other L2 participants. Since they had been exposed to similar classroom instruction, and this instruction was a significant and recent contributor to their acquisition of Mandarin, we decided to retain these heritage speakers in the present L2 sample.

4.2 Materials and procedure

4.2.1 Visual World experiment

The visual world experiment was designed following T&C's Experiment 3. Linguistic stimuli consisted of questions as in (2). In experimental items, they contained one of three sortal classifiers (条, *tiáo*, '~long, flexible'; 支, *zhī*, '~stick-like, long'; 张, *zhāng*, '~flat, spread open') paired with a noun that is not a prototypical member of the classifier class. For example, *gǒu* 'dog' was a target noun after the classifier *tiáo*, generally associated with objects that can be described as long, slender and flexible, descriptors that do not straightforwardly

apply to ‘dog’; nevertheless, *gǒu* ‘dog’ commonly co-occurs with *tiáo*. Adopting T&C’s terminology, target items can thus be characterized as ‘G+’, reflecting a grammatical match between classifier and noun, and ‘S-’, indicating a mismatch between the semantic features of the noun and those associated with the classifier (class). We followed T&C’s rationale for using G+S- targets in order to increase the potential for competition from other nouns.

- (2) 哪 一 条 是 狗?
Nǎ yī tiáo shì gǒu?
Which one CL is dog?
‘Which one is a/the dog?’

We created 12 item sets, with each set consisting of a target, three competitors, and a distractor. Each set included (i) a competitor from the same class matching the semantic features of the class (G+S+; e.g., *shéngzi* ‘rope’ for *tiáo*), (ii) a competitor from a different class that nevertheless possesses some of the target class’ semantic properties (G-S+; e.g., *shǒubiǎo* ‘wristwatch’), and (iii) a semantically unrelated competitor from a different class (G-S-; e.g., *píngguǒ* ‘apple’). Distractors were similar to (iii), i.e., grammatically and semantically unrelated to the target noun. Target and competitor nouns were selected by first translating T&C’s Cantonese stimuli into Mandarin. If the translation belonged to the same class (e.g., *tiú4/tiáo*), was compatible with only one of the classifiers used in this study, and was likely to be familiar to 3rd-year learners of Chinese (based on inspection of textbook vocabulary), the item was retained (29/48). Otherwise items were replaced with Mandarin nouns that fit the requirements of the design. All nouns appear >50 times in the SUBTLEX-CH corpus (Cai & Brysbaert, 2010), and were judged to be familiar to 3rd-year learners of Chinese.

Auditory stimuli were recorded by a female native speaker of Mandarin in a sound-proof booth at 44.1 kHz, and edited using Praat (Boersma & Weenink, 2017). Experimental stimuli were constructed by concatenating extracted tokens of *nǎ yī* (‘which one’), the classifier, *shì* (‘is’), and the target noun. The duration of the first three parts was held constant across all items. Silence was added after the classifier and after *shì* such that the duration from classifier onset to noun onset was exactly 1,150 ms in each experimental item. This 1,150-ms time period constitutes the critical region for analysis. Two native Mandarin speakers and two L2 learners checked and confirmed the naturalness of all stimuli.

Visual scenes consisted of 3 objects: the target, one of the 3 competitors, and a distractor (Figure 1). Items in the 3 conditions (G+S+, G-S+, G-S-) were rotated across 3 lists, such that each participant saw a given target object only once, for a total of 12 experimental trials, with 4 items in each of 3 conditions. An additional 24 filler trials were included. The experiment was conducted on an SMI RED250 eye-tracker sampling at 250 Hz. Each trial began with a 2000 ms display of the visual scene, followed by the question. Participants were instructed to click on the correct image in answer to the question. Gaze and mouse-click responses were recorded through SMI *Experiment Suite* software.



Figure 1. Examples of visual scenes for target [gōu, dog] for each condition.

4.2.2 Vocabulary test

A vocabulary test independent of the visual world experiment was created to assess whether participants have knowledge of the target classifier-noun pairings. Participants were asked to complete a 4-alternative forced-choice cloze test, which consisted of a total of 50 items, including the 12 target nouns from the Visual World experiment. In these 12 critical items, participants had to complete a numeral phrase (NUM CL N) with the most suitable classifier. An example is shown in (3). English translation of the phrase was given to ensure the intended meaning of the phrase was clear.

- (3) 一 _____ 魚
 a/one fish
 條 張 只 台

The vocabulary test was implemented as a web-based survey, which the participants were asked to complete at least 4 days before the test session in order to minimize any priming effects from the target items in the vocabulary test on performance in the visual world experiment. While not all participants adhered to this schedule, all completed the vocabulary test at least 1 day before the visual world task, and most completed it substantially earlier ($M = 8.7$ days, $SD = 6.7$ days, range: 1 – 30 days).

5. Results

We begin by reporting the results from the vocabulary test, as they inform our analysis of the eye-gaze data. Mean accuracy across all 50 items was 95% (range: 88-100) in the L1 and 73% (44-90) in the L2 group. For the 12 target classifier-noun items only, accuracy was substantially lower, especially in the L2 group (L1: $M = 90\%$, range: 75-100; L2: $M = 46\%$, 17-75). That some L1 speakers performed below 100% on these items likely reflects that some nouns can occur with more than one classifier. Thus an unexpected response on this task does not necessarily indicate lack of knowledge of the target CL-N pairing, but could reflect an individual preference for a different pairing.

Taking into consideration the findings from the vocabulary test, we first present eyegaze data from all trials, followed by a second presentation including

only those trials for which the participant had selected the expected classifier for the target noun on the vocabulary test. The first analysis preserves balanced numbers of items per participant, at the risk of including items for which participants may have lacked knowledge of the target CL-N pairing. The second analysis affords greater confidence that participants have this knowledge, yet we lose statistical power due to reduced and unbalanced data.

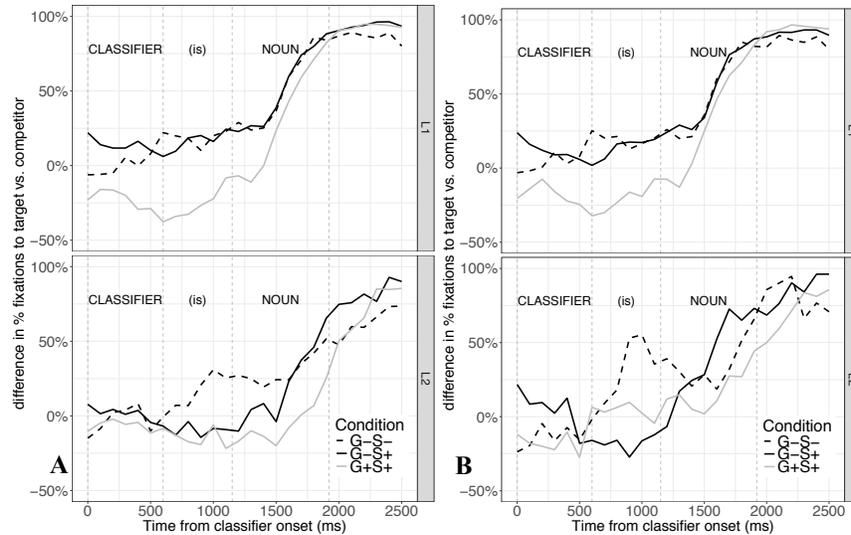


Figure 2. Difference in proportion fixations to target vs. competitor, by Condition and Group. A: all trials; B: Trials with incorrect classifier selection for target noun on Vocabulary Test excluded.

Trials with incorrect mouseclick responses were excluded from all analyses. This affected 1.4% of the eye-gaze data in the L1, and 10.1% in the L2 group. Figure 2 illustrates L1 and L2 participants' looking behavior in the three experimental conditions (panel A: all trials; panel B: excluding trials with incorrect classifier selection for the target on the vocabulary test). For each participant group and condition, we plotted the difference in the proportion of looks to the target minus the proportion of looks to the competitor. Thus a positive value indicates more looks to the target than to the competitor, a negative value indicates more looks to the competitor, while zero indicates an equal proportion of looks to the target and the competitor. Visual inspection of looking patterns in the L1 group indicates a steady increase in looks to the target after the onset of the classifier in both the G-S- and G-S+ conditions, with very similar patterns in the two conditions. In the G+S+ condition, by contrast, looks to the target decrease following the onset of the classifier, indicating that native speakers were more likely to look at the competitor – the prototypical member of the classifier class – than at the non-prototypical target. In the L2 group, we

observe an increase in looks to the target following the classifier in the G-S- condition only. Note that the increase starts well before the onset of the noun, indicating that L2 participants appear to be using information on the classifier predictively. In both the G+S+ and G-S+ condition, we see a trend in the L2 group towards more looks to the competitor, with looks to the target increasing only after the onset of the noun. Notably, and unlike in the L1 group, we observe a very similar pattern in the G+S+ and G-S+ conditions in the L2 group.

For statistical analysis, we calculated weighted empirical logits over looks to target vs. competitor, aggregated over subjects and over items (Barr, 2008), during the critical region starting from the acoustic onset of the classifier and extending to the onset of the noun (1,150 ms). Fixations initiated before 200 ms after classifier onset were excluded (following T&C). We used linear mixed-effect regression (lmerTest; Kuznetsova et al., 2016) to model the data. We constructed analogous models for the two sets of data (by-participant and by-item) with Condition and Group as fixed effects, and random intercepts for Participants and Items, respectively. Group was contrast-coded (L1 = -.5, L2 = .5) and centered. Condition was treatment-coded with G-S+ as the baseline, thus allowing for the two theoretically motivated contrasts between G-S+ and G+S+ and between G-S+ and G-S-. Note that the effect of Group in this model indicates a simple effect, i.e., the effect of Group in the baseline G-S+ condition. Table 1 presents the relevant model output for the dataset corresponding to Figure 2A, i.e., not excluding trials based on performance on the independent vocabulary test. Both the by-participant and by-item models indicate a significant difference between the G-S+ and the G+S+ conditions, with overall fewer looks to the target (and more to the competitor) when the competitor matches the classifier in both form class (G+) and semantics (S+). Importantly, this effect is qualified by a significant interaction with Group. Separate follow-up models on the data from each group show a significant difference between the G-S+ and G+S+ conditions in the L1 ($b_1 = -.93$, $p = .001$; $b_2 = -1.14$, $p < .001$) but not in the L2 group ($b_1 = -.03$, $p = .9$; $b_2 = -.09$, $p = .8$), indicating that native speakers, but not L2 learners, experience more competition from class-consistent (G+) than from class-inconsistent (G-) competitors that match semantic features of the classifier. In other words, L1 speakers appear to privilege G+/- information, and are affected by semantic cues only when the comparison is between competitors from within the same classifier class (prototypical versus non-prototypical). L2 learners, on the other hand, do not distinguish between form class-consistent and -inconsistent competitors as long as they instantiate semantic properties associated with the classifier. It thus appears that L2 speakers privilege S+/- information over G+/- . This observation is further supported by the simple effect of Group (significant only by items), with the negative beta value indicating that in the (baseline) G-S+ condition, L2 learners were less likely to look at the target (and more likely to look at the competitor) than native speakers. In other words, L2 learners appear to experience somewhat more competition than L1 speakers from competitors that match the classifier in semantics but not in form class. Notably, when the same

models are run with G+S+ as the baseline condition, the effect of Group is not significant ($b_1 = .45, p = .13; b_2 = .41, p = .21$), indicating that the difference in the G-S+ condition is unlikely to be due to L1 speakers generally being less affected by competition than L2 speakers.

Table 1. Model output for analysis of all trials (with correct click responses).

	by participants		by items	
	b_1	p	b_2	p
(Intercept)	.11	.5	.11	.5
ConditionG+S+	-.52	.01 *	-.62	.007 **
ConditionG-S-	.04	.8	-.18	.4
Group	-.48	.10	-.65	.04 *
ConditionG+S+ × Group	.93	.03 *	1.06	.02 *
ConditionG-S- × Group	.56	.2	.58	.2

* $p < .05$ ** $p < .01$

We applied the same models to the data after excluding trials for which the participant selected an unexpected classifier on the vocabulary test (Figure 2B). Recall that these exclusions affected approximately 10% of the L1, and as much as 50% of the L2 data. The dataset for this second analysis is thus considerably smaller and unbalanced. In these models, the overall difference between the G-S+ and the G+S+ conditions observed in the first analysis is no longer significant ($b_1 = -.31, p = .2; b_2 = -.33, p = .2$), yet importantly, the interaction with Group remains fully significant ($b_1 = 1.23, p = .02; b_2 = 1.53, p = .005$). Separate follow-up models for each group continue to show a significant difference between the G-S+ and G+S+ conditions in the L1 ($b_1 = -.83, p = .008; b_2 = -1.05, p < .001$) but not in the L2 group ($b_1 = .40, p = .3; b_2 = .57, p = .2$). As in the first analysis, the simple effect of Group is significant by items only ($b_1 = -.51, p = .16; b_2 = -.79, p = .02$). Thus even when confining the analysis to trials where we can be reasonably confident that participants knew the association between the classifier and the target noun, we find that L1 listeners experience more competition from class-consistent (G+S+) than class-inconsistent (G-S+) competitors, while L2 listeners appear to consider them equally.

5. Discussion

The goal of this study was to test the hypothesis that L2 learners of Chinese rely more strongly on the *semantic* informativity of sortal classifiers than L1 speakers during real-time comprehension. In support of this hypothesis, we found that while L1 speakers experienced less competition from class-inconsistent (G-) than from class-consistent (G+) competitors that match semantic features of the target noun, L2 speakers seem to experience equal competition from the two types of competitors. The pattern observed in the L1

group is consistent with T&C's conclusion that L1 speakers rely on form-class information as a primary predictive cue, while the semantics associated with the classifier comes into play only when competitors from within the same classifier class (prototypical versus non-prototypical) are pitted against each other. By contrast, the looking behavior of our L2 group suggests that L2 listeners are using classifier-semantics as a primary cue to create expectations about upcoming nouns, with form-class information playing a more minor role.

We note, however, that the data presented here are at least suggestive of L2 listeners being able to use form-class information predictively as well: When no object in the visual scene matched the semantic features associated with the classifier, as was the case in the G-S- condition, L2 listeners showed a clear increase in looks to the target (vs the competitor) well before the onset of the noun (Figure 2). Since targets in this experiment were always non-prototypical members of their classifier class (G+S-), this facilitation cannot be attributed to the semantic informativity of the classifier. We must caution, however, that the statistical models reported here were not designed to assess the robustness of this particular effect, which here we describe only based on visual inspection of Figure 2. Additional analyses will be needed to examine this further.

In sum, it appears that the differences observed in this study between the L1 and L2 groups are best captured in terms of relative weight allocated to the different levels at which classifiers can be informative for predictive processing. While L1 listeners treat form-class as primary, and semantic information as secondary cues, the opposite appears to be the case for L2 listeners. This is consistent with prior proposals from various theoretical backgrounds which have argued for greater reliance on lexico-semantic, and lesser reliance on grammatical, information in L2 processing (e.g., Felser et al., 2003; Van Patten, 2004). The findings reported here support an extension of this claim to the realm of predictive processing.

Our account of L1/L2 differences in terms of differences in relative weighting of cues is compatible in particular with perspectives that view L1/L2 differences as gradual rather than categorical (see also Clahsen & Felser, 2017). Yet the question that remains to be addressed is *why* L2 learners would rely more strongly on semantically informative cues than L1 speakers. As we pointed out above, it does not seem to be the case that L2 learners are *unable* to use form-class information on classifiers, at least when there is no semantic match available in a given context. Instead, we believe that lesser reliance on classifier form-class information in L2 than in L1 processing may stem from attention to different unit sizes in L2 vs. L1 learning. As demonstrated in experimental studies on so-called blocking effects in language learning (Ellis, 2006), learners first exposed to unsegmented multiword units are more successful at learning semantically uninformative co-occurrence relations between determiners and nouns than learners who are first exposed to individual nouns associated with their meanings (Arnon & Ramscar, 2012; Siegelman & Arnon, 2015; see Paul & Grüter, 2016, for replication with classifier-noun sequences in Chinese). As Arnon and colleagues have argued, the former is reflective of L1 acquisition,

where speech segmentation is a crucial initial task for the language learning infant, whereas the latter is more similar to L2 learning, where learners come to the task with sophisticated knowledge of language from their L1, which allows them to home in more efficiently on the meaning of individual words. Yet L2 learners' relative efficiency at identifying the meaning of words means that they have to rely relatively less (compared to L1 learners) on information contained in co-occurrence relations between nouns and words they often co-occur with, such as determiners and classifiers. As a consequence, it appears likely that information on co-occurrence statistics will be a less prominent part of L2 than of L1 representations. If these assumptions are on the right track, then the stronger reliance on semantic vs form-class cues in L2 vs. L1 learners of Mandarin can be seen as reflecting listeners' adaptive use of prediction during language processing under different circumstances (Kuperberg & Jaeger, 2016). In other words, listeners favor the cue that has the best predictive value within their relevant system of representation; in an L1 system, this is likely to be the cue derived from co-occurrence statistics, within an L2 system, it is more likely to be one related to semantic informativity.

6. Limitations and Conclusions

The findings presented here constitute a preliminary report from an ongoing study. As pointed out above, the current L2 sample includes some speakers with early/heritage exposure to Chinese, as well as speakers whose performance on the vocabulary test indicated limited knowledge of classifiers, which led to substantial data loss, and consequent loss of statistical power, in some of our planned analyses. We are currently collecting data from additional sequential L2 learners with more extensive exposure to Mandarin in China, with the goal of conducting future analyses on equally sized samples of L1 and advanced sequential L2 speakers of Mandarin.

This study has provided at least preliminary evidence from the processing of classifier-noun sequences by L1 and L2 speakers of Mandarin suggesting that L2 listeners rely more strongly than L1 listeners on the semantic informativity of classifiers to create expectations about upcoming nouns. This aligns well with previous findings that showed intermediate-proficiency L2 learners of Spanish were able to use semantically informative, but not semantically uninformative, gender marking on determiners. These findings are consistent with accounts of L2 processing that propose generally greater reliance on lexico-semantic information in L2 than in L1 processing, and extend the observation to the realm of predictive processing. We argue that the observed differences between L1 and L2 listeners do not stem from L2 listeners' inability to use grammatical or form-class cues, but that due to differences in attention to different unit sizes during L2 vs. L1 learning, semantic cues have relatively greater utility than form-class cues for prediction in an L2 compared to an L1.

References

- Arnon, Inbal, & Ramscar, Michael (2012). Granularity and the acquisition of grammatical gender. *Cognition*, 122(3), 292-305.
- Barr, Dale J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59, 457-474.
- Boersma, Paul & Weenink, David (2017). Praat: doing phonetics by computer [Computer program]. Version 6.0.36, retrieved 11 November 2017 from <http://www.praat.org/>
- Cai, Qing, & Brysbaert, Marc (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *Plos ONE*, 5(6), e10729.
- Chen, Baocun, Che, Guicheng, Chen, Hao, & Zhang, Zaizhan. (1988). *Hanyu liangci cidian* [A dictionary of Chinese numeral classifiers]. Fuzhou, China: Fujian Renmin Chubanshe [Fujian People's Publishing House].
- Clahsen, Harald, & Felser, Claudia (2017/FirstView). Some notes on the Shallow Structure Hypothesis. *Studies in Second Language Acquisition*.
- Dahan, Daphne, Swingle, Daniel, Tanenhaus, Michael K., & Magnuson, James S. (2000). Linguistic gender and spoken-word recognition in French. *Journal of Memory and Language*, 42, 465-480.
- Dussias, Paola E., Valdes Kroff, Jorge R., Tamargo, R. E. G., & Gerfen, C. (2013). When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition*, 35(2), 353-387.
- Ellis, Nick C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, saliency, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics*, 27(2), 164-194.
- Felser, Claudia, Roberts, Leah, Marinis, Theodore, & Gross, Rebecca (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics*, 24(3), 453-489.
- Gao, Ming Y., & Malt, Barbara C. (2009). Mental representation and cognitive consequences of Chinese individual classifiers. *Language and Cognitive Processes*, 24(7/8), 1124-1179.
- Grosjean, François, Dommergues, Jean-Yves, Cornu, Etienne, Guillelmon, Delphine, & Besson, Carole (1994). The gender marking effect in spoken word recognition. *Perception & Psychophysics*, 56(5), 590-598.
- Grüter, Theres, Lew-Williams, Casey, & Fernald, Anne. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research*, 28(2), 191-215.
- Grüter, Theres, Rohde, Hannah, & Schafer, Amy J. (2014). The role of discourse-level expectations in non-native speakers' referential choices. In Will Orman & Matthew J. Valteau (Eds.), *Proceedings of the 38th Annual Boston University Conference on Language Development* (pp. 179-191). Somerville, MA: Cascadilla Press.
- Grüter, Theres, Rohde, Hannah, & Schafer, Amy J. (2017). Coreference and discourse coherence in L2: The roles of grammatical aspect and referential form. *Linguistic Approaches to Bilingualism*, 7(2), 199-229.
- Guillelmon, Delphine, & Grosjean, François (2001). The gender marking effect in spoken word recognition: The case of bilinguals. *Memory & Cognition*, 29(3), 503-511.
- Hopp, Holger. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, 29(1), 33-56.
- Hopp, Holger (2016). Learning (not) to predict: Grammatical gender agreement in non-native processing. *Second Language Research*, 32(2), 277-307.
- Hopp, Holger, & Lemmerth, Natalia (2016/FirstView). Lexical and syntactic congruency in L2 predictive gender processing. *Studies in Second Language Acquisition*.

- Huettig, Falk, Chen, Jidong, Bowerman, Melissa, & Majid, Asifa. (2010). Do language-specific categories shape conceptual processing? Mandarin classifier distinctions influence eye gaze behavior, but only during linguistic processing. *Journal of Cognition and Culture*, 10, 39-58.
- Kaan, Edith (2014). Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism*, 4(2), 257-282.
- Kaan, Edith, Dallas, Andrea, & Wijnen, Frank (2010). Syntactic predictions in second-language sentence processing. In Jan-Wouter Zwart & Mark de Vries (Eds.), *Structure preserved. Festschrift in the honor of Jan Koster* (pp. 207–213). Amsterdam: John Benjamins.
- Klein, Natalie M., Carlson Greg N., Li, Renjie, Jaeger T. Florian, & Tanenhaus, Michael K. (2012). Classifying and massifying incrementally in Chinese language comprehension. In Diane Massam (Ed.), *Count and mass across languages* (pp. 261-282). Oxford, UK: Oxford University Press.
- Kuperberg, Gina R., & Jaeger, T. Florian (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, 31(1), 32-59.
- Kuznetsova, Alexandra, Brun Brockhoff, Per, & Christensen, Rune H. B. (2016). lmerTest. Retrieved from <http://CRAN.R-project.org/package=lmerTest>
- Lau, Elaine, & Grüter, Theres (2015). Real-time processing of classifier information by L2 speakers of Chinese. In Elizabeth Grillo & Kyle Jepson (Eds.), *Proceedings of the 39th Annual Boston University Conference on Language Development* (pp. 311-323). Somerville, MA: Cascadilla Press.
- Lew-Williams, Casey, & Fernald, Anne. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, 18, 193-198.
- Lew-Williams, Casey, & Fernald, Anne. (2009). Fluency in using morphosyntactic cues to establish reference: How do native and non-native speakers differ? In Jane Chandlee, Michelle Franchini, Sandy Lord & Gudrun-Marion Rheiner (Eds.), *Proceedings of the 33rd Annual Boston University Conference on Language Development* (pp. 290-301). Somerville, MA: Cascadilla Press.
- Lew-Williams, Casey, & Fernald, Anne. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, 63, 447-464.
- Li, XuPing (2013). *Numeral Classifiers in Chinese: The Syntax-Semantics Interface*. Berlin: De Gruyter.
- Martin, Clara D., Thierry, Guillaume, Kuipers, Jan-Rouke, Boutonnet, Bastien, Foucart, Alice, & Costa, Albert (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, 69, 574-588.
- Paul, Jing Z., & Grüter, Theres (2016). Blocking effects in the learning of Chinese classifiers. *Language Learning*, 66(4), 972-999.
- Siegelman, Noam, & Arnon, Inbal (2015). The advantage of starting big: Learning from unsegmented input facilitates mastery of grammatical gender in an artificial language. *Journal of Memory and Language*, 85, 60-75
- Tsang, Cara, & Chambers, Craig G. (2011). Appearances aren't everything: Shape classifiers and referential processing in Cantonese. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(5), 1065-1080.
- VanPatten, Bill (2004). Input processing in second language acquisition. In Bill VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 5-31). Mahwah, NJ: Erlbaum.