

Effects of morphosyntactic complexity on Pronoun Interpretation: Classifiers in Mandarin

Pronoun interpretation depends on many factors such as grammatical role ([1],[2],[3]), discourse prominence and referential form (null pronoun vs. overt pronoun) ([4],[5]). In *pro-drop* languages like Mandarin, overt pronouns show a bias towards a subject interpretation, and null pronouns show an even stronger subject bias ([4],[6]). However, the reasons behind this “subject” bias are unclear. One explanation is the *Topichood hypothesis* ([2],[7]): subjects are generally topical referents and are therefore more likely to be re-mentioned in subsequent dialogue, in short/reduced (e.g., as a pronoun); [1]. Topicality influences both production and interpretation of pronouns in Mandarin ([4],[8]): Importantly, fronting a subject to the left of the a sentence-initial word like “yinwei” (“because”) increases the likelihood that both null and overt pronouns are interpreted as referent with that subject, but null pronouns more so than overt ones ([4],[9],[10],[11]). Another (related) potential factor is complexity: Because pronouns are short elements, they may favor a corresponding, or “parallel” ([12]) short-form antecedent, and topics themselves are usually the most morphologically or prosodically reduced NPs in a sentence, which would make choosing them as an antecedent give rise to COMPLEXITY PARALLELISM. We test our COMPLEXITY PARALLELISM hypothesis through two experiments on Mandarin, by manipulating morphosyntactic complexity of the object (Exp.1) and subject (Exp.2) antecedent NPs via presence vs. absence of nominal classifiers (small words denoting NP’s shape and function), and complexity of the pronoun itself (null or overt).

We ran two 2x2x2 forced-choice sentence comprehension experiments (Topical/Canonical word order, Classifier Presence/Absence, Null/Overt pronoun, following [4], [8] sentence comprehension experiments (within subs, within items)). Exp1 manipulates presence of a classifier on the object NP; Exp2 manipulates presence of a classifier on the subject NP. The task is to read sentences like in Table 1 and then answer a comprehension question asking the interpretation of the null/overt pronoun in the second conjunct. The question accompanying the conditions in Table 1 is “Who hit the rock?” with three response options of subject, object, and neither: “horse”, “donkey”, “someone else”. For each experiment, we have 24 itemsets and 16 fillers. Responses are coded as 1 when the “subject” answer is chosen, 0 otherwise.

Expt 1 (Object NP modified by classifier). Data from 64 L1 Mandarin speakers, recruited from Prolific, was analyzed. Results are shown in Fig 1 (left). A 2x2x2 logistic MEM reveals a main effect of Topicality ($z=6.09$, $p<.0001$), meaning that topicality leads to more subject reference choice. Pronoun type was also observed ($z=-3.44$, $p=.0006$), meaning that null pronouns significantly increase subject/topic referent choice. Importantly, we found a main effect of Classifier ($z=-3.2$, $p=.001$), indicating that an object modified by a classifier increases the likelihood of choosing the subject as an antecedent; there were no interactions ($ps > .2$).

Expt 2 (Subject NP modified by classifier). Data from a different 63 L1 Mandarin speakers, recruited from Prolific, was analyzed. Results are shown in Fig 1 (right): We found a main effect of Topicality ($z=9.69$, $p<.0001$), meaning that meaning that the topical order increases the likelihood of choosing the subject as the referent, compared with canonical order. This supports the idea that discourse prominence (topicality) influences pronoun resolution. There was also a main effect of Pronoun type ($z=-3.52$, $p=.0004$), meaning that overt pronouns significantly reduce subject/topic referent choice. This indicates that null pronouns are more likely to refer to the most topical element (consistent with *pro-drop* languages like Mandarin), as in [4], [6]. There were no interactions ($ps > .5$), suggesting that topicality effects override complexity of the subject NP.

To compare Experiments 1 and 2, we analyzed the two datasets as a whole, by including “Experiment” as an additional fixed effect in the logistic MEM. The analysis revealed a main effect of Topicality ($z=11.11$, $p<.0001$), and Pronoun type ($z=-4.87$, $p<.0001$). Critically, there are two interactions with Experiment: one was a significant positive Classifier \times Experiment interaction ($z=3.06$, $p=.002$), meaning that classifier absence/presence has a stronger effect when it is manipulated on the *object* NP than the subject NP. The other effect was a Topicality \times Experiment interaction ($z=2.72$, $p=.007$): A stronger subject bias for the canonical word order in Exp 1; when

the object was marked with a classifier. Overall, we find support for the view that effects of topicality on pronoun interpretation are independent of morphological reduction [8]. However, we also find effects of morphological complexity on subject antecedent selection, but only when the complexity of the clause-mate object is manipulated: We hypothesize that this is because the predictability that the subject will be mentioned next is *relative* to that of the object: when the object is morphologically complex (i.e., modified by a classifier), then the subject is a comparatively *more* likely next-mention candidate than the subject having the *more* reduced form. Conversely, complexity manipulations of the subject do not alter the next-mention probability of the subject; the subject is *already* the more topical referent.

Table 1. Sample item: Topical/Canonical (a/c vs. b/d) × Classifier present/absent (a/b vs c/d) × Null/overt pronoun (italics)

a. Topical – Subject (Exp2) / Object (Exp1) Classifier present – Overt/Null Pronoun									
<i>yi</i>	<i>pi</i>	ma	yinwei	zhuigan	<i>yi</i>	<i>pi</i>	lv,	suoyi	<i>ta/ ∅ ...</i>
<i>one</i>	<i>CL</i>	horse	because	chase	<i>one</i>	<i>CL</i>	donkey,	so	<i>it/ ∅ ...</i>
b. Canonical – Subject (Exp2) / Object (Exp1) Classifier present – Overt/Null Pronoun									
yinwei	<i>yi</i>	<i>pi</i>	ma	zhuigan	<i>yi</i>	<i>pi</i>	lv,	suoyi	<i>ta/ ∅ ...</i>
because	<i>one</i>	<i>CL</i>	horse	chase	<i>one</i>	<i>CL</i>	donkey,	so	<i>it/ ∅ ...</i>
c. Topical – Classifier absent – Overt/Null Pronoun									
		ma	yinwei	zhuigan	lv,	suoyi	<i>ta/ ∅</i>	...	
		horse	because	chase	donkey,	so	<i>it/ ∅</i>	...	
d. Canonical – Classifier absent – Overt/Null Pronoun									
		Yinwei	ma	zhuigan	lv,	suoyi	<i>ta/ ∅</i>	...	
		because	horse	chase	donkey,	so	<i>it/ ∅</i>	...	
<i>“The/(one) horse chased (one) donkey, so (it)”</i>									

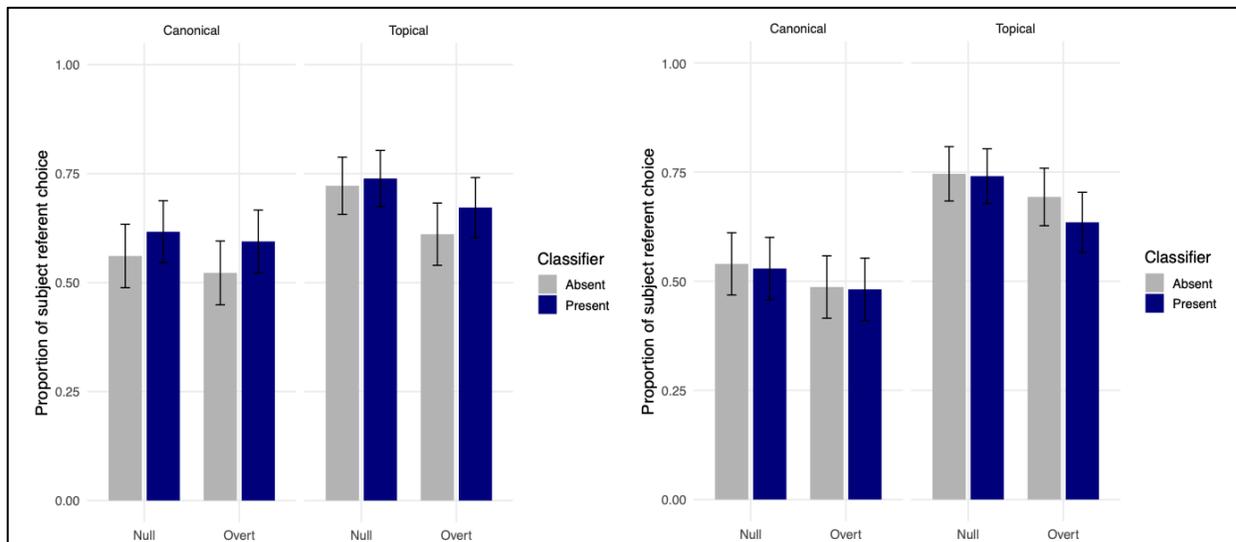


Fig 1. The proportion of subject referent choice by condition: object NP modified by classifier (left) and subject NP modified by classifier (right). Error bars show 95% CI.

References: [1] Arnold, 2001, Discourse Processes. [2] Gordon et al., 1993, Cognitive Science. [3] Rohde & Kehler, 2014, LCN. [4] Lam & Hwang, 2022, Cognitive Science. [5] Ariel, 1990, Anaphoric antecedents. [6] Zhang & Kwon, 2022, J of Linguistics. [7] Givon, 1983, Typological Studies in Language. [8] Lam & Hwang, 2025, LCN. [9] Ng, Hwang, & Cui, 2025, Discourse Processes. [10] Arnold, 1998. PhD Thesis, Stanford. [11] Liao et al., 2024, Glossa. [12] Chambers & Smyth, 1998, JML.