Experimental evidence for the discourse potential of bare nouns in Mandarin

Jess H.-K. Law & Syrett Kristen
Rutgers University

1. Introduction

Many studies on bare nouns have come to the agreement that these nominals should have a special status that distinguishes them from regular indefinites, either in terms of syntactic incorporation (Geenhoven 1998) or semantic (pseudo-)incorporation (Farkas & de Swart 2003, Chung & Ladusaw 2004, Dayal 2011, Krifka 2016, a.o.). However, there is far less consensus on whether and how the special status has a reflex in the discourse domain. Chung & Ladusaw (2004) observe that incorporated nominals in Chamorro are transparent in discourse, i.e., they support anaphora just as regular indefinites do, but Farkas & de Swart (2003), Dayal (2011) and Krifka (2016) argue that bare nouns in Hungarian, Hindi and Persian have reduced discourse transparency as compared with indefinites. Within the second camp, Farkas & de Swart (2003), based on the investigation of bare nouns in Hungarian, maintain that reduced transparency leads to unacceptable anaphora, while Dayal (2011), Modarresi (2014), Krifka (2016) allow bare nouns to support anaphora with added pragmatic steps.

This study addresses the controversy over the discourse transparency of bare nouns by offering an experimental study of bare nouns based on Mandarin. Specifically, by comparing pronominal anaphora supported by bare nouns, as exemplified in (1), and that supported by regular indefinites, as exemplified by (2), we ask to what extent bare nouns share the hallmark property of regular indefinites in supporting pronominal anaphora.

---

*This work has benefited from discussions with Simon Charlow, Veneeta Dayal, Ken Safir. Funding was generously provided by a startup grant to K. Syrett. Special thanks go to the research assistants in Laboratory for Developmental Language Studies, who helped conduct the norming study.

1A note of clarification. Reduced discourse transparency has been attributed to number-neutral bare nouns in Hungarian and Persian. It is not clear what role number morphology on bare nouns plays in Hindi. This paper only concerns the discourse property of number-neutral bare nouns.

2Bare nouns are number-neutral in Mandarin and have been suggested to be compatible with both singular and plural pronouns (Rullmann & You 2006). Pronouns do not contrast in terms of gender in spoken Mandarin but have gender-specific orthographic forms.
(1) Pronominal anaphora supported by a bare noun
Wo kanjian-le **xuesheng**, Tā/tāmén gen wo wenhao.
I see-Asp student he/they with me greet
‘I saw a student/students. He/she/they greeted me.’

(2) Pronominal anaphora supported by an indefinite
Wo kanjian-le **yi-ge** xuesheng, Tā gen wo wenhao.
I see-Asp one-Cl student he with me greet
‘I saw a student. He/she greeted me.’

In more detail, we conducted an experiment with an offline, acceptability judgment task and an online, self-paced reading task to find out how native speakers of Mandarin perceive and process pronominal anaphora with bare nouns and regular indefinites. It is found that the two types of pronominal anaphora were judged as having comparable acceptabilities, but pronominal anaphora involving bare nouns have longer processing time. The results suggest that pronominal anaphora with bare nouns incur extra processing effort not found in pronominal anaphora with regular indefinites, providing support for studies that argue for extra steps for bare nouns to support discourse anaphora (e.g., Dayal 2011, Krifka 2016).

The rest of the paper is organized as follows: Section 2 outlines the existing theories and their implications for the experiment. Section 3 details the experiment design and Section 4 discusses the experimental results. Section 5 concludes.

2. Three hypotheses

2.1 What they are

The literature on the discourse transparency of bare nouns offers three hypotheses: the opacity hypothesis, the transparency hypothesis, and the translucency hypothesis. In this section, I elaborate on these hypotheses, focusing on how the current study serves to test the predictions derivable from each of them.

The opacity hypothesis is first discussed by Dayal (1999) and later defended at great length by Farkas & de Swart (2003). According to Farkas & de Swart (2003), bare nouns in Hungarian cannot support pronominal anaphora, unlike regular indefinites, as evidenced by the contrast in (3) and (4) (examples cited from Farkas & de Swart 2003: (38–39)).

(3) Hungarian pronominal anaphora with a bare noun
a. János, **beteg** vizsgált.Past a rendelőben.
   Janos patient.Acc examine.Past the office.in
   ‘Janos patient-examined in the office.’

b. ??pro, Túl sulyosnak találta **öt** és beutáltatta pro a korházba.
   too severe.Dat find.Past he.Acc and intern.Cause.Past pro the hospital.in
   ‘He found him too sick and sent him to hospital.’
Experimental evidence for the discourse potential of bare nouns in Mandarin

(4) **Hungarian pronominal anaphora with an indefinite**


b. pro, Túl súlyosnak találta ótj és beutaltatta pro j a kórházba. pro too severe.Dat find.Past he.Acc and intern.Cause.Past pro the hospital.in ‘He found him too sick and sent him to hospital.’

The degraded status of (3) is attributed to the interaction between the bare noun and the pronoun: the pronoun is an index carrying a presupposition that a discourse referent with the same index is already present in the context. Since the bare noun fails to introduce such a discourse referent, the presupposition cannot be met. As a result, presupposition failure kicks in and severely degrades the pronominal anaphora with a bare noun.

And the opposite end of the hypothesis spectrum is the **transparency hypothesis**, which takes bare nouns to be transparent in discourse, just like indefinites. While this hypothesis has not drawn as much attention compared to the opacity hypothesis, it has been defended by a few studies. An example is Chung & Ladusaw (2004), who found that incorporated nominals in Chamorro are transparent in discourse, as cited in (5) (Chung & Ladusaw 2004: 122). Since they argue that bare nouns are also incorporated, they can be taken to advocate the transparency hypothesis.

(5) **Chamorro incorporated nominal**

Káda unu ni gai-[haga], siempri ha-po’lu na bunita gui’, each one Comp WH[nom].have-daughter surely Agr-assume Comp pretty she ‘Everyone who has a daughter, thinks that she is beautiful.’

Lying somewhere between the two polarizing hypotheses is the **translucency hypothesis**. This hypothesis takes pronominal anaphora with bare nouns to appear transparent. However, the transparency is argued to differ from the transparency of pronominal anaphora with regular indefinites. Concretely, bare nouns are argued to not introduce discourse referents the way that regular indefinites do. Nonetheless, pronominal anaphora with bare nouns is achieved via adjustments in pragmatics, which are not required for pronominal anaphora with regular indefinites. Dayal (2011) and Krifka (2016) are representative studies along these lines.3

2.2 How they were tested

The experiment consisted of an off-line, acceptability judgment task and an on-line, non-cumulative self-paced reading task. The acceptability judgment task was a standard method in testing linguistic grammaticality. If pronominal anaphora involving bare nouns shows

---

3For Dayal (2011), the pragmatic adjustment can be implemented as a kind of ‘bridging’. For Krifka (2016), the discourse referents introduced by bare nouns are closed by subordinated existential closure. To retrieve them, additional steps in the forms of abstraction and summation are necessary.
comparable acceptability with the baseline, i.e., pronominal anaphora involving indefinites, then the transparency and translucency hypothesis is favored over the opacity hypothesis. Otherwise, the opacity hypothesis is favored. The non-cumulative self-paced reading task, also known as the moving window paradigm, is well known for its ability to reflect online processing difficulty as a positive function of the reading time (Just et al. 1982). It was intended to probe the processing effort incurred in pronominal anaphora supported by bare nouns. If such a sentence required longer processing time than a similar sentence with an indefinite baseline, then it lends support to the opacity hypothesis and the translucency hypothesis. The following table summarizes the relevant potential results:

<table>
<thead>
<tr>
<th>Potential result</th>
<th>Lends support to</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pronominal anaphora with BNs acceptable and fast</td>
<td>Transparency hypothesis</td>
</tr>
<tr>
<td>Pronominal anaphora with BNs unacceptable and slow</td>
<td>Opacity hypothesis</td>
</tr>
<tr>
<td>Pronominal anaphora with BNs acceptable but slow</td>
<td>Translucency hypothesis</td>
</tr>
</tbody>
</table>

3. **Experiment**

3.1 **Stimuli**

The stimuli consisted of 27 targets, 18 controls, and 65 fillers. They were grouped into 3 lists using a Latin-Square design with pseudo-randomization, and were presented in Simplified Chinese characters (font size 28). Each list had 9 targets, 6 controls, and 31 to 33 fillers. The same targets and controls showed up in the self-paced reading task and the acceptability judgment task. A subset of the fillers used in the self-pacing reading task were also used in the acceptability judgment task.

Each target and control stimulus consisted of two sentences, as exemplified in (7). The first sentence (S1) introduced an antecedent noun phrase in a minimal context. The second sentence (S2) began with a (possibly covert) pronoun that was unambiguously anaphoric to the noun phrase introduced in S1.

(7) Sample stimulus

S1: Women zai bianlidian kandao-le xiaotou. we at convenience.store saw-Asp THIEF
    ‘We saw THIEF in a convenience store.’

S2: {Ta/tamen/pro} tou-wan dongxi zhihou jiu likai-le xianchang.
    he/they/pro steal-Asp thing after then leave-Asp scene
    ‘He/they/pro left after stealing things.’

All the S2s shared the same schema, as shown in (7). Each S2 could be divided into 6 fragments, including a pronoun region and the spill-over region, which was the three words following the pronoun (pronoun+1 ... pronoun+3). Previous studies have demonstrated that
Experimental evidence for the discourse potential of bare nouns in Mandarin

a spill-over region of roughly this size was appropriate for self-paced reading studies on pronouns (e.g., Chow et al. 2014). Each target fragment was controlled for length, log word frequency, and visual complexity. Parts of speech (PoS) were controlled for the pronoun+1 region only. A linear-mixed effects model was fit to the data and it revealed that PoS did not have a main effect on the reading speed. Therefore, letting PoS vary in the spill-over region is not of concern. In the self-paced reading experiment, all sentences were displayed fragment by fragment.

(8) Sample stimulus format

<table>
<thead>
<tr>
<th></th>
<th>S2</th>
<th>Ta/tamen/pro</th>
<th>tou-wan</th>
<th>dongxi</th>
<th>zhihou</th>
<th>jiu</th>
<th>likai-le.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region</td>
<td>pronoun</td>
<td>pronoun+1</td>
<td>pronoun+2</td>
<td>pronoun+3</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Length</td>
<td>1-2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>3-5</td>
<td></td>
</tr>
<tr>
<td>PoS</td>
<td>pronoun</td>
<td>verb</td>
<td>varies</td>
<td>varies</td>
<td>varies</td>
<td>varies</td>
<td></td>
</tr>
</tbody>
</table>

The target stimuli involved two factors, namely, bare noun type and pronoun type, each having three levels. Crossing three levels of bare noun type and three levels of pronoun type resulted in nine types of target stimuli. To achieve a Latin-square design, three sentences with different bare nouns used in different contexts were created for each type of target stimuli, giving rise to 27 target stimuli in total.

The three levels in bare noun type are: (i) bare nouns with a neutral number bias (BN-Neutral), (ii) bare nouns with a singular number bias (BN-Singular), and (iii) bare nouns with a plural number bias (BN-Plural). Bare nouns with different number biases were included because previous studies had taken number bias to be a factor relevant for pronominal anaphora with bare nouns (Modarresi 2014). The number bias of the bare nouns used in this study was determined by a norming study. The three levels in pronoun type are: (i) overt 3rd-person singular pronoun (SG), (ii) overt 3rd-person plural pronoun (PL), and (iii) covert pronoun.

There were two types of control stimuli. The first type had numeral phrases as the referential antecedent and the second type had coordinated bare nouns as the referential antecedent. Numeral phrases, qua regular indefinites in Mandarin, were included to tease apart the transparency, opacity and translucency hypotheses. By comparing the behaviors of pronominal anaphora with bare nouns and that with regular indefinites, we can determine which of these hypotheses best accords with the experimental results. Coordinated bare nouns were included to test a few different variants of the translucency hypothesis. However, for reasons of space, we do not try to tease these variants apart in this paper and hence do not discuss the results from from coordinated bare nouns.

4Stimuli involving the covert pronoun turned out to have both low acceptability and long processing time, regardless of whether the antecedent is a bare noun or an indefinite. As it is likely that there are other complexities involved with processing the covert pronoun, they were excluded from the analysis.
3.2 Participants

30 native Mandarin speakers were recruited to participate in the experiment, which took place in a laboratory. They were compensated with a seven-dollar bookstore gift card. All participants participated in the acceptability judgement task and the self-paced reading task.

3.3 Procedure

In the self-paced reading task was conducted before the acceptability judgment task. The experiment was conducted using a 21” computer screen with SuperLab 5. Instructions and practice trials were presented before each task.

In the self-paced reading task, participants read each item fragment by fragment, pressing the space bar to advance through the sentence pair. The reaction time between key presses were recorded. Comprehension questions appeared after some number of the stimuli, to ensure that participants paid attention to the task and correctly interpret the coreference relation intended by the pronoun in each item. Data points with incorrectly answered comprehension questions were excluded in data analysis.

After a self-timed break, participants proceeded to the acceptability judgment task. For each trial, participants had to rate, based on a scale of 1 to 5, how acceptable the sentences presented to them were. Participants were instructed to use the number keys on the keyboard to indicate their choice. The Chinese equivalents of the following descriptive labels were used with the scale: 5: Perfectly acceptable; 4: Acceptable; 3: Marginally acceptable; 2: Hardly acceptable; and 1: Completely unacceptable.

4. Results

All statistical analyses reported in this paper were conducted by linear mixed-effects modeling with lme4 package for the statistical language R (R Core Team 2016). The merit of mixed-effects models is that they can take independent variables as fixed effects and incorporate both random effects of subjects and items within a single analysis (Baayen et al. 2008). All models used to generate the statistical analyses in this study are mixed-effects models with random effects of subject and item, unless specified otherwise. Models are fit using a restricted maximal likelihood technique. Probabilities were estimated by means of the function summary in the package lmerTest.

4.1 Acceptability judgment

A total of 450 ratings were collected from the target stimuli in the acceptability judgment experiment. No outliers were identified. The results are presented in (9), with the left fig-

---

5A potential concern with using linear regression for analyzing Likert scales is that Likert scales are ordinal in nature and do not satisfy the linearity and normality assumptions. However, Norman (2010) has dismissed this worry as unnecessary as Likert scales can be appropriately analyzed with linear-mixed effects modeling. Gibson et al. (2011) have also recommended using linear mixed-effects modeling to analyze Likert scales.
Experimental evidence for the discourse potential of bare nouns in Mandarin

As shown in the left figure in (9), the control items had expected acceptability. In particular, participants refuted the pronominal anaphora when a pronoun mismatched an indefinite in terms of number, showing that they made use of the lower end of the Likert scale to indicate unacceptability. In addition, they rated pronominal anaphora with indefinites as highly acceptable (mean = 4.85), indicating that they also made use of the higher end of the scale to indicate acceptability and that the stimuli with indefinites were well-chosen to serve their role as the baseline.

The results from the target stimuli are presented in the right figure. There are a few points to note about this figure. First, in contrast to the sensitivity to different pronouns shown by indefinites, bare nouns are more tolerant to pronouns with different numbers. Relatedly, neutral bare nouns without a number bias can support a singular pronoun and a plural pronoun equally well. These results are expected as bare nouns are number-neutral in Mandarin. That being said, number bias does show its effect-bare nouns with a particular number bias prefers a pronoun with the same number.

Secondly, focusing on bare nouns supporting pronouns that are compatible with their number bias, we found that they were also judged highly acceptable (mean = 4.57), with neutral bare nouns reaching the rating 4.59, bare nouns with a singular bias reaching the rating 4.43, and bare nouns with a plural number bias reaching the rating 4.73. Although the target stimuli had a slightly lower rating than the control stimuli when it comes to supporting the right pronoun, the difference is not statistically significant ($\beta = 0.25$, SE = 0.19, $t = 1.31$, $p = 0.192$). What the high acceptability suggests is that bare nouns are either transparent or translucent in discourse, but not opaque.

6In addition to excluding data points from stimuli involving coordinated bare nouns and those involving the covert pronoun, a set of data involving indefinites in the form of a vague numeral phrase ji-ge NP "several NP" were also excluded. The main reason is that participants gave unexpected tolerance to sentences with singular pronouns referring to these expressions. A potential reason might be that it was mistaken to be an interrogative word, which has the same form and roughly means ‘how many’ in Mandarin.
Overall the comprehension questions received a high accuracy rate (mean = 95.38%), indicating that the participants paid attention in the reading task and processed the reading material. In addition, there is a negative correlation between reaction times and acceptability ratings in the data collected ($r = -0.33$, $n = 45$, $df = 43$, $p < 0.05$), confirming the assumption that participants generally spend more time reading less acceptable sentences.

A total of 14292 response time data points were collected, 2550 of which were from the targets and controls. Response times larger than 1092ms and smaller than 200ms were taken to be outliers and removed from analysis. The upper cutoff point was computed as the upper quartile (498ms) plus three times the interquartile range ($3 \times 189$ms), as done in Sturt et al. (2002). The lower cutoff point is set at 200ms because any time shorter than this is believed to be likely due to unintended button presses rather than responses after meaningful semantic processing (Jegerski 2014, Baayen & Milin 2010).

In addition, all response time data from items whose comprehension questions had been incorrectly answered were removed. This series of data treatment affected 9.29% of the target and control data points, leaving 2313 target data points to be analyzed. The response time data were then log-transformed to make the distribution less eschewed and better fit the statistical analyses. The result (pre-log-transformation) is presented in (10) below.

(10) Results from the self-paced reading task

The dash-and-dot line decorated with circles at the bottom of the graph represents the response time profile of the stimuli involving indefinites. The solid line decorated with triangles represents the results from the stimuli involving bare nouns. The bare noun line is further decomposed into three lines, the BN-Neutral line, the BN-Plural line, and the BN-Singular line, to show the internal composition.
Let us pay attention to the indefinite line and the bare noun line first. In the pronoun region, they hardly differ (394ms vs. 391ms). However, they start to diverge in the pronoun+1 region (313ms vs. 372ms), and continue to do so in the pronoun+2 region (316ms vs. 389ms) and the pronoun+3 region (329ms vs. 369ms). The overall difference is statistically significant ($\beta = 0.12$, SE = 0.04, $t = -2.99$, $p < 0.05$). In addition, the three lines that collectively constitute the bare noun line are similar in shape. More importantly, they all show a processing profile different from the indefinite line’s, indicating that whatever that trigger a slower processing profile was shared by all the bare nouns stimuli, regardless of the number bias of the bare nouns.

Recall that due to the lag time in processing, a slow down is often not seen in the ‘trigger’, but reflected on the spill-over region of that trigger. In this case, the pronoun, which requires an antecedent, is the trigger, and the region following the pronoun is the spill-over region. Since a pronoun does trigger a slow down when its antecedent is a bare noun, we can then conclude that there is more processing effort associated with pronominal anaphora supported by bare nouns. This result, when interpreted independently from the result of the acceptability judgment task, is compatible with a view that bare nouns have reduced discourse transparency.

4.3 Discussion

We started out with three hypotheses regarding the discourse transparency of bare nouns in Mandarin: the opacity hypothesis, the transparency hypothesis, and the translucency hypothesis. The acceptability judgment task showed that bare nouns support pronominal anaphora just as well as indefinites, refuting the opacity hypothesis. The self-paced reading task revealed that there was more processing effort associated with pronominal anaphora supported by bare nouns, as compared to pronominal anaphora supported by indefinites, refuting the transparency hypothesis. What is left is the translucency hypothesis, which predicts that bare nouns have reduced discourse transparency, but anaphora can still be successful, as reflected by the high acceptability, given extra pragmatic steps, as reflected by the longer processing time.

A more general message from this study is that the discourse domain is a valid place to look for evidence in support of the special syntactic and/or semantic status of bare nouns. However, one has to use the right combination of measures to detect the effect.

5. Conclusion

The findings of this study contribute to the literature on the discourse property of bare nouns at a few levels. At the theoretical level, it provides novel, empirical support for the translucency hypothesis, which is otherwise difficult to tease apart from the transparency hypothesis. At the methodological level, it is the first study to investigate the discourse property of bare nouns in an experimental setting, with tasks sensitive to both acceptability and processing. It is this combination of experimental tasks that helped us single out the translucency hypothesis.
References


Jess H.-K. Law, Kristen Syrett
jesshk.law@rutgers.edu, kristen.syrett@rutgers.edu