1. Introduction

Sluicing, originally discussed by Ross (1969), is exemplified by the pair of English sentences in (1).

(1) a. Kim$_1$ saw Sue somewhere, but I don’t know where she$_1$ saw Sue.
    b. Kim saw Sue somewhere, but I don’t know where Δ.

Descriptively, it is possible to omit the clausal material expected to follow the *wh*-phrase, as shown in (1b). As a supposed species of ellipsis, a significant amount of work has gone into determining what, if any, linguistic material is present in the sluicing site Δ.

A portion of the literature has converged on the idea that, at some level of representation, there is full clausal syntax in the omission site. Even so, the exact identity of the clausal material has been up for debate. An influential analysis proposes that (1b) has the representation in (2), involving ellipsis of an IP from which the remnant *wh*-phrase has been extracted (e.g., Ross 1969, Merchant 2001). It has also been proposed—especially for *wh*-in-situ languages—that sluices have representations like (3), resulting from ellipsis of either a copular or cleft construction (e.g., Merchant 1998, Potsdam 2007, Adams and Tomioka 2012, Gribanova 2013, and others).

(2) Kim saw Sue somewhere, but I don’t know where$_1$ [ Kim saw Sue$_1$ ].
(3) Kim saw Sue somewhere, but I don’t know where$_1$ [ it was (that Kim saw Sue$_1$) ].

With this in mind, our goal in this paper is to investigate the syntax of sluicing-like constructions (SLCs) in Mandarin Chinese, a *wh*-in-situ language. A relevant pair of examples is provided in (4).

---

1 Alternative analyses of sluicing in this vein propose that the ellipsis site is supplied its content post-syntactically (e.g., Chung et al. 1995), contains a proform that is pragmatically resolved (e.g., Lobeck 1995, Messick et al. 2019), or lacks linguistics content all together (e.g., Culicover and Jackendoff 2001).
(4) a. Zhangsan, zai moudi kanjian-le Lisi,
    Zhangsan at somewhere saw-ASP Lisi
dan wo bu zhidao ta, zainali kanjian-le Lisi
    but I NEG know he where saw-ASP Lisi
    ‘Zhangsan saw Lisi somewhere, but I don’t know where he saw Lisi.’
b. Zhangsan, zai moudi kanjian-le Lisi, dan wo bu zhidao (shi) zainali
    Zhangsan at somewhere saw-ASP Lisi, but I NEG know COP where
    ‘Zhangsan saw Lisi somewhere, but I don’t know where.’

The example in (4b) demonstrates the ability to omit the clausal material expected to follow the wh-phrase shui ‘who.’ We argue that Mandarin employs at least two different strategies to generate SLCs: an extraction strategy, like (2), and a cleft reduction strategy, like (3). Which strategy is employed is reliably betrayed by the presence or absence of the copula shi.

The motivation for this analysis comes from the account it provides for a well-known puzzle regarding the distribution of the element shi in SLCs, which we illustrate in section 2. Our treatment of SLCs presented in section 3 finds additional support from previously observed connectivity effects and a novel corpus study, which are presented in section 4. Finally, we show in section 5 how this analysis provides a way to understand a set of otherwise puzzling facts regarding the distribution of shi with the remnant zenmeyang ‘how.’ We conclude the paper in section 6.

2. The Puzzling Asymmetric Distribution of shi in Mandarin SLCs

A well-known property of Mandarin SLCs is the variable presence of the element shi, which is normally identified as a copula or focus marker. As was shown by the example in (4b), shi may be optional in an SLC. The non-optionality of shi in examples like (5) have led to the generalization that shi is optional only with adjunct wh-phrases, see (4b), but obligatory with argument wh-phrases (e.g., Adams 2004, Wang and Wu 2006).

(5) Zhangsan, kanjian-le mouren, dan wo bu zhidao *(shi) shui
    Zhangsan saw-ASP someone, but I NEG know COP who
    ‘Zhangsan saw someone, but I don’t know who.’

Wang and Wu (2006: 383) argue that the optionality of shi reflects is role as a Case assigner to the remnant. As an argument wh-phrase, shui ‘who’ in (5) requires Case, which is asserted to be assigned at PF under adjacency. The presence of shi fulfills this need of Case assigning, while the adjunct zainali in (4b), does not need Case or does not have to rely on shi for Case. Thus, shi is optional in these instances.

There is a reason to think that this empirical generalization does not accurately characterize the data and, therefore, that this analysis does not sufficiently account for the distribution of shi. Adams and Tomioka (2012: 223, (9d)) provide the example in (6), in which an argument wh-phrase appears in an SLC with shi optionally present.

(6) Lisi bu xihuan yishou ge, danshi wo bu zhidao (shi) na yi shou ge
    Lisi NEG like one-CL song but I NEG known COP which one CL song
    ‘Lisi doesn't like a song, but I don’t know which one/song.’
The remnant in (6) would be expected to require Case assignment at PF under adjacency, just like the remnant in (5). The optionality of shi, then, is incorrectly predicted.

In response to this and other data, Adams and Tomioka (2012: 222) state the necessity of shi as a function of the identity of the wh-phrase. We state the generalization as in (7).

(7) A Mandarin SLC has shi
   i. obligatorily before simplex argument wh-phrases (shui ‘who’, shenme ‘what’) and
   ii. optionally before other wh-phrases (zainali ‘where’, han shui ‘with whom’, etc.)

This generalization has been used to motivate or bolster analyses that tie the presence or absence of shi instead to the type of predication relationship that can be established inside the SLC (e.g., Adams 2004, Wei 2004, Park & Li 2013). Space precludes doing proper justice to these analyses. However, we would recommend that the reader see the discussion in Li & Wei 2017 for considerations against these approaches. It is for these reasons that we present the alternative account of (7) that is found below.

3. Multiple Paths to Sluices

3.1. Sluicing and Pseudo-sluicing

As stated above, we propose that Mandarin makes use of two different strategies for generating SLCs. Moreover, which strategy is employed is reliably betrayed by the presence or absence of the copula shi.

In an SLC without the copula shi, we propose that the wh-phrase is a remnant of IP-ellipsis. This is intended to resemble genuine sluicing in a language like English, but to be more closely analogized to the focus-driven sluicing argued for by Wang and Wu (2006) and Song and Yoshida (2017). The example in (8) illustrates the syntax of the example in (4b) with what we will refer to as IP-Sluicing.

(8) IP-Sluicing
    … dan wo bu zhidao [cp zainali₁ [zhangsan₁, kanjian-le Lisi]]
    but I NEG know where Zhangsan see-ASP Lisi
    ‘… but I don’t know where Zhangsan saw Lisi.’

For concreteness, we assume with the research cited just above that the wh-phrase undergoes focus-driven movement to a clause peripheral position. In this position, it escapes ellipsis of an embedded IP constituent.

SLCs that contain the copula shi, we suggest, have a different underlying syntax. The presence of shi in (4b) signals the Pseudo-sluicing strategy provided in (9).

(9) CP-Pseudo-Sluicing
    … dan wo bu zhidao [cp pro expo, shi zainali₁ [zhangsan₁, kanjian-le Lisi de]]
    but I NEG know COP where Zhangsan see-ASP Lisi DE
    ‘… but I don’t know where it was that Zhangsan saw Lisi.’
Like Adams and Tomioka (2012), we suppose that the presence of shi signals an embedded copular construction. However, we more closely follow Wang and Wu (2006) and Song (2016) in supposing that the remnant wh-phrase is formally related to a position in relative-like clause. In other words, we propose that shi signals what we will descriptively refer to as a reduced shi-de-cleft strategy, whereby an embedded CP is elided.2

The claim, therefore, is that there are two different clause-reduction strategies for generating SLCs in Mandarin. In the following sections, we will see how this provides a handle on the generalization in (7). However, we can observe here that treating both strategies as elliptical configurations leads us to expect the binding connectivity effects observed by Song and Yoshida (2017), independent of the presence of absence of shi. A relevant example is provided in (10).

(10)  
Ta1 gen san-ge Lisi de pengyou qu kan-le dianying, dan wo bu zhidao ...  
He with three-CL Lisi GEN friend go watch-ASP movie but I NEG know

a.* gen ji-ge Zhangsan1 de pengyou2 [IP te-x2 qu kan-le dianying]  
with how.many-CL Zhangsan GEN friend he go watch-ASP movie

b. shi gen ji-ge Zhangsan1 de pengyou2 [CP ta-x2 qu kan-le dianying de]  
SHI with how.many-CL Zhangsan GEN friend he go watch-ASP movie DE

*‘He, went to the movies with three of Lisi’s friends, but I don’t know with how many of Zhangsan,’ friends he went to the movies.’

The disjoint reference effects associated with Condition C arise in the absence of shi, as in (10a), and in the presence of shi, in (10b). As Song and Yoshida (2017) point out, this is expected from a clause reduction analysis of Mandarin SLCs. This follows from our proposal, wherein the two available strategies, betrayed the presence or absence of shi, are both generated by a clause-reduction mechanism.

3.2. Wh-fronting in simple sentences predicts shi in SLCs

We proposed in (8) that Mandarin SLCs without the copula shi involve focus-driven movement of the wh-remnant out of an elliptical IP. The requirement for shi with simplex argument wh-phrases, therefore, can be understood as a reflection of the general inability of these elements to appear in IP-Sluicing configurations. Consider the contrast below:

(11)  
*Shui1 [IP Zhangsan kanjian-le x1]?  (12)  
Zainali [IP Zhangsan x1 kanjian-le Lisi ]?  
who Zhangsan see-ASP where Zhangsan see-ASP Lisi

‘Who did Zhangsan see?’ ‘Where did Zhangsan see Lisi?’

These examples demonstrate that simplex argument wh-phrases like shui ‘who’ in (11) cannot independently appear clause-initially. On the other hand, other wh-phrases, including

2 We will remain intentionally vague with regard to the internal syntax of the cleft construction in (9). The remnant wh-phrase may be extracted from the embedded CP or may be mediated with a CP-internal position via an operator chain (see Park & Li 2013). Choosing between these two options must be left for future research. As we will see, the important property of this construction is its status descriptively as a type of shi-de construction. Thus, we are also not committed to the idea that CP-Pseudo-Sluicing can be fed by just any type of shi-de construction (see Cheng 2008 and Paul & Whitman 2008).
zainali ‘where’ in (12), can independently appear clause-initially. This can be taken to reveal that the underlying syntax that we take to be responsible for SLCs without shi (IP-Sluicing), which is shown in (13), is not available to simplex argument wh-phrases.

\[(13) \quad * \ldots \text{dan wo bu zhidaowu}_{[ip} \text{Zhangsan kanjian-le}_x,\text{]}
\begin{align*}
& \text{but I NEG know who } \text{Zhangsan saw-ASP} \\
& \text{‘Zhangsan saw someone, but I don’t know who.’}
\end{align*}
\]

Given this, the obligatoriness of shi with simplex argument wh-phrases follows from the fact that they can only be generated by CP-Pseudo-Sluicing. That is, the example from (6) necessarily has the representation in (14).

\[(14) \quad \ldots \text{dan wo bu zhidaowu}_{[CP \text{pro-exp]} \text{shi}} \text{ shui}_{1} \quad \text{[CP Zhangsan kanjian-le}_x, \text{de ]}
\begin{align*}
& \text{but I NEG know COP who Zhangsan saw-ASP DE} \\
& \text{‘Zhangsan saw someone, but I don’t know who.’}
\end{align*}
\]

The optionality of shi with other wh-phrases can be understood, not strictly as the optionality of the copula shi, but as an option between two different mechanisms for sluicing. The independent ability to front the non-simplex wh-phrases, like zainali ‘where’ in (12) reveals the availability of IP-Sluicing as a mechanism for generating SLCs like (4b), in which case there will be no shi.

\[(15) \quad \ldots \text{dan wo bu zhidaowu}_{[ip} \text{Zhangsan-x, kanjian-le Lisi}]
\begin{align*}
& \text{but I NEG know where Zhangsan saw-ASP Lisi} \\
& \text{‘Zhangsan saw Lisi somewhere, but I don’t know where.’}
\end{align*}
\]

On the other hand, when these wh-phrases appear with shi, we propose the SLC has been generated by CP-Pseudo-Sluicing. Thus, (4b) may have the representation below in (16).

\[(16) \quad \ldots \text{dan wo bu zhidaowu}_{[CP \text{pro-exp]} \text{shi}} \text{ zainali}_{[CP \text{Zhangsan-x, kanjian-le de } ]}
\begin{align*}
& \text{but I NEG know COP where Zhangsan saw-ASP DE} \\
& \text{‘Zhangsan saw Lisi somewhere, but I don’t know where.’}
\end{align*}
\]

While the analysis that has been presented here provides an account of (7), it remains to be demonstrated that the presence of shi in an SLC necessarily indicates an elliptical shi-de cleft. To this end, we provide the non-elliptical versions of (4b) and (5) in (17) and (18), respectively.

\[(17) \quad \text{Zhangsan zai moudi kanjian-le Lisi,}
\begin{align*}
& \text{Zhangsan at somewhere saw-ASP Lisi,} \\
& \text{dan wo bu zhidaowu}_{[CP \text{pro-exp]} \text{shi}} \text{ zainali}_{[CP \text{Zhangsan-x, kanjian-le *(de) } ]}
\begin{align*}
& \text{but I NEG know COP where Zhangsan saw-ASP DE} \\
& \text{‘Zhangsan saw Lisi, but I don’t know where Zhangsan saw Lisi.’}
\end{align*}
\]

---

3 See Song (2016) for a similar observation about other focus constructions.
The sentences in (17) and (18) are unacceptable without the element *de at the end of the sentences. These non-elliptical examples demonstrate, first, that simplex argument *wh-phrases and other *wh-phrases are both independently able to appear in what we have referred to as *shi-de-clefts. This supports the claim that, in SLCs with *shi, the elliptical material contains an elided CP of a *shi-de-cleft. Second, these examples illustrate that in non-elliptical versions of the sentences of interest, the sentence final particle *de is obligatory. This provides further support for the claim that, in the presence of *shi, a Mandarin SLC has elided an embedded relative-like CP of the shape provided in these examples.

4. Treebank Corpus Search Evidence

To further support our claim of multiple sluicing strategies in Mandarin Chinese, we conducted a corpus search in the Chinese Penn Treebank 7.0 (CTB7). Our search was conducted with two questions.

4.1. Evidence for previous observations on simplex *wh-phrases

First, we asked about the validity of the previous observations in the literature regarding the obligatory *ness of *shi with simplex *wh-phrases; see (7i). We searched for instances of SLC with a simplex *wh-phrase and the occurrences of the copula *shi preceding the *wh-phrase. We labeled the data with [SHI+simplex-*wh] for the *wh-phrase appearing with *shi and [-SHI+simplex-*wh] for the ones without *shi. If (7i) is correct, we will expect to only find [SHI+simplex-*wh] examples, as Table 1 shows, which are representatives for this kind of construction in the corpus. On the other hand, we expect to observe a zero (or near zero) co-occurrences of simplex argument *wh-phrases without *shi.

<table>
<thead>
<tr>
<th>Corpus Examples (elliptical)</th>
<th>Translations</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ...kao gu xue jia bu zhidaosh** shi** shenme</td>
<td>“… the archaeologists don’t know what.”</td>
</tr>
<tr>
<td>b. ...houlai you you jiquan da zai wo lian shang, bu zhidaosh** shi** shui</td>
<td>“… more punches came into my face, (I) don’t know who.”</td>
</tr>
</tbody>
</table>

Table 1: Examples of elliptical version of [SHI + simplex-*wh] in CTB7

The result of corpus counting is shown in Table 2 on the following page. The comparison of the two elliptical versions in Table 2 reflects the predictions of (7i). In this 780,000-word corpus, simplex argument *wh-phrases appear without *shi in only 2.1% of the relevant examples, while the ones with *shi appear much more frequently in 46.9% of the examples.

4 The remaining constructions in the total count were non-embedded instances of *shi-de constructions and, thus, not immediately relevant to the questions at hand.
Elliptical version | Counts | Percentage  
---|---|---  
[ SHI + simplex-wh ] | 45/96 | 46.9%  
[ -SHI + simplex-wh ] | 2/96 | 2.1%  

Table 2: The counts of the presence versus the absence of shi with simplex wh-phrases in the elliptical version in CTB7

In addition to confirming the previous observations in (7i), the search results in Table 2 also support what we propose in this paper: that the absence of shi indicates an application of IP-sluicing (8), whereas the presence of shi indicates an application of the CP-Pseudo-sluicing (9). The high versus low percentages of these two kinds of constructions suggest that IP-sluicing is not accessible to simplex wh-phrases, but CP-Pseudo-sluicing is.

4.2. Evidence for shi...de construction as an underlying form

Second, we asked whether non-elliptical instances of embedded CPs provided evidence for the claim surrounding (17) and (18) that the presence of shi in SLCs implicates the presence of a shi-de constructions underlingly, before applying CP-Pseudo-Sluicing.

To answer this question, we searched for non-elliptical embedded wh-clauses. Among these selected sentences, we compared the rate of occurrence of embedded wh-clauses with shi-de surrounding the wh-phrase to the occurrence rate of those constructions with only shi preceding a wh-phrase and no subsequent de. The two kinds of constructions are labelled as [+SHI+DE] and [+SHI-DE] in the data, respectively. If what we proposed in section 3 is correct, we should expect to observe high occurrences of the [+SHI+DE] construction, as shown in the example sentences in Table 3. On the other hand, we should expect to see very low occurrences of the [+SHI-DE] constructions.

<table>
<thead>
<tr>
<th>Corpus Examples (non-elliptical)</th>
<th>Translations</th>
</tr>
</thead>
<tbody>
<tr>
<td>c. ... na ge shaonian shi zai nali dedao yizhi qiang de</td>
<td>“… (asked) where the young man got the gun.”</td>
</tr>
<tr>
<td>d. ...kankan duoshu ren shi zenme xiang de</td>
<td>“… see how the majority thinks of this.”</td>
</tr>
</tbody>
</table>

Table 3: Examples of non-elliptical version of [+SHI +DE] in CTB7

The search results of these two kinds of constructions are shown in Table 4. They bear out the prediction of the claim made above. The sentences with the [+SHI +DE] construction occur more frequently than the [+SHI -DE] construction in the corpus with a ratio of 65:13.

<table>
<thead>
<tr>
<th>Non-elliptical version</th>
<th>Counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ +SHI + DE ]</td>
<td>65</td>
</tr>
<tr>
<td>[ +SHI - DE ]</td>
<td>13</td>
</tr>
</tbody>
</table>

Table 4: The counts of the presence versus the absence of de with wh-phrases in the non-elliptical version in CTB7
The ratio shows that the instances of embedded \textit{wh}-clauses containing \textit{shi-de} constructions outnumber those with only \textit{shi}. For the purpose of this particular search, we excluded many kinds of \textit{shi-de} constructions, such as relative clauses and main clause questions (\textit{e.g., ni shi nali de}, ‘where are you from’). If we would have included these less relevant examples, the resulting ratio could be even greater due to the higher number of occurrences of \textit{shi-de}.

To sum up this section, the two corpus search results provide supporting evidence for the previous observation of the obligatoriness of \textit{shi} with simplex argument \textit{wh}-phrases, as well as our proposal for the distinct underlying representation of SLCs containing \textit{shi}.

5. The Zenmeyang-Puzzle

The generalization in (7) is a fairly well-accepted description of the distribution of \textit{shi} in Mandarin. However, there is a well-known, but little understood, complication to these facts. We extend our analysis here to account for this complication.

5.1 Zenmeyang ‘how’ behaves like simplex argument \textit{wh}-phrases

The \textit{wh}-adverbial zenmeyang ‘how’ appears in SLCs but, contrary to expectations, necessarily co-occurs with \textit{shi}.\textsuperscript{5} In other words, zenmeyang patterns with simplex argument \textit{wh}-phrases despite being neither (obviously) simplex nor an argument \textit{wh}-phrase. The example in (19) is provided to illustrate this observation.

\begin{verbatim}
(19) Zhangsan xiuru-le Lisi, dan wo bu zhidao *(shi) zenmeyang.
    Zhangsan humiliate-ASP Lisi, but I NEG know COP how
    ‘Zhangsan humiliated Lisi, but I don’t know how.’
\end{verbatim}

The analysis that was presented in the previous section would assert that this pattern reflects the fact that zenmeyang can not be generated as a \textit{wh}-remnant by way of the IP-Sluicing strategy introduced in (8). Instead, the analysis would assert, only the CP-Pseudo-Sluicing strategy from (9) is available for generating SLCs with zenmeyang ‘how.’

This makes specific predictions about how zenmeyang ‘how’ will behave in non-elliptical constructions. Specifically, we expect that zenmeyang ‘how’ will continue to pattern with simplex argument \textit{wh}-phrases in non-elliptical constructions. The following examples are provided to show that this prediction is borne out.

\begin{verbatim}
(20) *Zenmeyang, [\textit{ip} Zhangsan x, xiuru-le Lisi ]?
         how                   Zhangsan humiliate-ASP Lisi
    ‘How did Zhangsan humiliate Lisi?’
(21) Zhangsan xiuru-le Lisi,
    Zhangsan humiliate-ASP Lisi,
         dan wo bu zhidao [\textit{cp} pro exp\ shi zenmeyang [\textit{cp} Zhangsan x, xiuru-le *(de)]]
        but I NEG know COP how                     Zhangsan humiliate-ASP DE
    ‘Zhangsan humiliated Lisi, but I don’t know how Zhangsan humiliated Lisi.’
\end{verbatim}

\textsuperscript{5} There are some alternatives of \textit{zenmeyang: zenme, zenyang, and ruhe} to refer to ‘how’ in Mandarin. We believe that the claims here can be extended to these additional \textit{wh}-phrases as well.
The example in (20) shows that zenmeyang ‘how’ is not independently able to be fronted to a clause-initial position from its underlying position as a modifier of the predicate. In the same way as above, this supports the claim that IP-Sluicing is unavailable for making zenmeyang ‘how’ a wh-remnant. The effect is that CP-Pseudo-Sluicing will be the required strategy, which has the effect of requiring that shi be present (19). The example in (21) demonstrates that zenmeyang ‘how’ is independently able to appear in a shi-de-cleft and, moreover, that the presence of shi requires the presence of de.

5.2 Corpus evidence on Zenmeyang

With the goal of testing whether zenmeyang ‘how’ behaves like simplex argument wh-phrases, and whether CP-Pseudo-sluicing underlies examples like (19), we repeated the second study from section 4 on only those constructions containing zenmeyang ‘how.’ We searched for non-elliptical embedded wh-clauses with zenmeyang ‘how.’ We then compared the rate of occurrence of shi-de surrounding the wh-phrase to the occurrence rate of those constructions with only shi was present.

The result is shown below in Table 5.

<table>
<thead>
<tr>
<th>Non-elliptical version of zenmeyang</th>
<th>Counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>-SHI -DE</td>
<td>5</td>
</tr>
<tr>
<td>+SHI -DE</td>
<td>0</td>
</tr>
<tr>
<td>+SHI +DE</td>
<td>60</td>
</tr>
</tbody>
</table>

Table 5: The counts of the presence versus the absence of shi-de with zenmeyang ‘how’ in the non-elliptical version in CTB7

Looking at the left column in Table 5, [-SHI -DE] indicates examples that have zenmeyang ‘how’ occurring without either shi or de; [+SHI -DE] indicates zenmeyang without de; and [+SHI +DE] indicates zenmeyang surrounded by shi...de. That evidence again supports our claims that, first zenmeyang syntactically behaves like simplex argument wh-phrases. There is an overwhelming preference for zenmeyang to occur with shi. Secondly, when shi is present, it must co-occur with de. This suggests in the same way as above that CP-Pseudo-Sluicing is the underlying representation for SLCs containing zenmeyang.

6. Conclusion

In this paper we have argued that Mandarin employs two different strategies for generating sluicing configurations. These include a IP-ellipsis strategy familiar from languages like English as well as a reduced cleft construction that employs CP-ellipsis. This treatment allows for an account of the apparent optionality of shi that also predicts a well-known asymmetry between simplex argument wh-phrases and other wh-phrases. We also showed how this multiple-strategy analysis can be extended to provide an understanding of the otherwise puzzling observation that the wh-remnant zenmeyang ‘how’ patterns with simplex argument wh-phrases in requiring shi.
Selected references


