Discourse Factors Involved in Distribution of Null-Subject Phenomenon in Mandarin Chinese

Siqi Shi

Nanyang Junior College, Singapore
Email: mmjudy@126.com

Abstract

This paper aims to examine the null-subject phenomenon in Mandarin Chinese, a discourse-oriented radical pro-drop language, and the factors that contribute to the presence/absence of overt subject pronouns in the language. A questionnaire survey was conducted among 11 Chinese high school students to collect data regarding the usage of null subjects by native Mandarin Chinese speakers. Three discourse factors that affect the distribution of null subjects, namely: 1) register, 2) speakers' agency, and 3) nature of information conveyed by subject pronouns, are identified and analysed based on the collected data. Therefore, we should pay more attention to pragmatic factors when refining syntactic theories regarding null-subjecthood in future studies.

Keywords

1. The Null-Subject Phenomenon

Null subject or pro-drop is the phenomenon where the subject position in a sentence can be left empty. There are mainly two types of pro-drop languages. Canonical pro-drop languages like Spanish permit pro-drop in almost all situations. In (1) below, the pronominal subject 'we' can be left unexpressed. Null-subjecthood in Spanish is commonly attributed to its relatively rich verbal inflection, in which the verb already provides information about the number and person of the subject pronoun, rendering the explicit expression of the pronominal subject redundant.

1) (*nosotros*) *hemostrabajadotodo el día* [Spanish]
We have-1SG.PRES worked all the day
"We worked all day." (D'Alessandro 2014:6) [1]

However, in radical pro-drop languages like Mandarin Chinese, due to the absence of agreement inflection, there is no morphological indication that provides reference to the null subject, and this information can only be extracted from the context. As shown in (2), the speaker answering the question may leave the subject *ta* (he) omitted.

2) Speaker A: *ZhangsankanjianLisi le ma*? [Mandarin Chinese]

Zhangsan see Lisi LE Q

"Did Zhangsan see Lisi?"

Speaker B:

a) *ta kanjian ta le*

He see he LE

"He saw him."

b) *e kanjian ta le*

see him LE

"[He] saw him." (Huang 1984: 533) [2]

This article presents new data that suggests a potential new factor (*i.e.* speakers' agency) in the licensing of null subjects in radical pro-drop languages. Section 2 of this paper is a review of previous theories accounting for the distribution of null-subject languages. Section 3 consists of results from a questionnaire survey conducted among native Mandarin Chinese speakers. An analysis of the data is presented in Section 4, which identifies three independent factors contributing to the occurrence of pro-drop from the survey results. Section 5 presents some further implications of the data and analysis. Section 6 is the conclusion of this paper.

2. Factors Affecting Null-Subjecthood

2.1. Verbal Agreement

The occurrence of null-subject phenomenon in various languages is widely ascribed to the setting of the null-subject parameter under Chomsky's Principles and Parameters (P & P) framework. Hence, a research question that has received wide attention is the correlation between the occurrence of null subjects in a language and other specific characteristics of the language (D'Alessandro 2014) [1]. Perlmutter (1971) [3] is the first to point out the relationship between null-subjecthood and the presence/absence of verbal inflections determined by person morphology. Taraldsen (1980) [4] further ascribes NSLs' ability to extract the pronominal subject across an overt complementiser to their rich inflection. Rizzi's (1982) [5] formulation of the classical NSP comprehensively summarises all proposals at that time related to NSLs and null-subject phenomenon. These observations have been a crucial part of our understanding of the presence/absence of null-subject phenomenon in different languages.

An example of NSL is Italian, where verbs inflect for six different person-number combinations, as shown in **Table 1**.

However, one exception is the subjunctive mood in Italian: it presents a

Table 1. Agreement inflection of Italian verbs.

Inflection forms of *Amare* "love"	Translation
(*io*) am-o	"I love"
(*tu*) am-i	"You love"
(*lui/lei*) am-a	"He/she loves"
(*noi*) am-iamo	"We love"
(*voi*) am-ate	"You love"
(*loro*) am-ano	"They love"

syncretic inflection for the first, second, and third person singular pronouns, yet only allows pro-drop in the first and third persons singular (Peverini 2004) [6].

4) *Credonochepro₁sia₁colpevole* (*pro* = io/*tu/lui) [Italian]

Believe.3pl that be.subj. guilty

"They believe me/*you/*him to be guilty". (Peverini 2004:109) [6]

As shown in (4), the fact that pro-drop is allowed in first and third persons singular presents what is contrary to our expectation according to Perlmutter's (1971) postulation[3].

According to Chomsky (1981, 1982) [7] [8], a main and necessary feature of pro-drop languages is recoverability (*i.e.*, that the referent of null subjects can be identified), which is widely thought to be achieved through agreement. However, in Standard Italian, because of the syncretic inflection in the subjunctive mood, the omission of first and third-person pronouns cannot be recovered from verbal agreement, but recoverability can be achieved through the context. Huang (1984, 1989) [2] [9] hence proposes a generalised control theory and argues that "identification hypothesis is essentially correct, but that it must be more broadly interpreted than is assumed in the agreement-based theory".

2.2. Pragmatic Factors

Despite the common impression that the null subject phenomenon has received an adequate account under the syntactic parameter theory, the P & P-based (or a purely formal syntactic) explanation is not really complete. These gaps in the account for the occurrence of null subject phenomena suggests greater emphasis be given to pragmatic factors.

In a similar attempt to investigate the effect of such pragmatic factors, Soares, Miller & Hemforth (2019) [10] reexamined present-day Brazilian Portuguese (BP) corpus data, and reach the conclusion that ambiguous (syncretic) agreement markings have little influence on the overall number of overt pronouns: only about 5% more subjects are overt in ambiguous verbal tense types than in exclusive verbal tense types in BP. However, when the first-person singular is used in a tense type that does not exclusively reveal the discourse person of the subject, the overt form is preferred to the null form. Soares, Miller & Hemforth (2019) [10] suggest that the impoverishment of verbal agreement results in a bias in favour of maximal informativeness in BP.

3. Data

3.1. Participants

A total of eleven high school students aged 18 or 19 participated in the survey. They were all born in Mainland China and consider Mandarin as their first language. A total of eight of them have received formal education with English as the main language of instruction in another country, while three have always lived in China.

3.2. Methodology

The questionnaire was designed with Q1 - Q4 including subject pronouns that convey only old information (*i.e.* information that has already been mentioned in the preceding context) and do not involve specific registers other than casual exchange of information. Subject pronouns in Q5 - Q7 also convey old information, but the respondents are free to interpret them as containing an element of emphasis or formality. Q8 - Q11 involve unambiguously more formal situations where the respondents are compelled to use a more serious or formal register in their chosen or proposed manners of speaking. Lastly, Q12 - Q15 involve subject pronouns that convey new information.

3.3. Results

3.3.1. Group I: Subject Pronouns Convey Old Information

Table 2 below shows responses to Q1 - Q4, which enquire information regarding a specific person whose identity is reflected as a pronoun (e.g., he) in the provided options, which creates the condition for the pronoun to be omitted. All of them are yes/no questions. Q1 - Q3 are multiple choice questions, where participants are asked to choose the answer that they find the most comfortable or natural, while Q4 is a rating question where participants are required to evaluate the four options on a scale of 1 - 5.

The respondents have shown a strong preference for the omission of both subject and object pronouns when the pronouns convey old information and their referents are already present in the preceding discourse. Both Q1 and Q2 have a total of 9 out of 11 responses preferring omitted subjects and objects, while in Q3, all respondents perceived the option without overt subject and object

Table 2. Participants' responses to Q1 - Q4.

Features	Q1	Q2	Q3	Q4	
	No. of responses			Ave	Std Dev
Overt subject only	0	0	0	3.9	0.79
Overt object only	0	1	0	2.5	0.38
Both overt	1	1	0	3.6	1.37
Both omitted	9	9	11	4.8	0.39
No difference	1	0	0	-	-

pronouns to be the most natural. In Q4, the option with the same characteristics is also assigned the highest average mark of 4.8, indicating it is the most comfortable way of expression, and its second lowest standard deviation of 0.39 shows that such perception among all 11 participants is rather consistent.

3.3.2. Group II: Respondents Can Freely Employ Different Registers

Questions in group II also involve pronouns whose referents are already specified in the preceding discourse, but the situations here may be interpreted as more serious than those in Group I (e.g. Q5 and Q6 are conversations with teachers), or just as casual exchange of information similar to those in Group I. Q5 and Q6 are multiple choice questions, and Q7 is a rating question (**Table 3**).

This involvement of speakers' agency in retaining or dropping pronouns is most salient in Q5, where the two participants choosing the option with overt subject and object pronouns expressed that, since the context is to speak to a teacher, they would adopt a more serious attitude to show respect. Two participants who used overt object also felt that this option conveys their seriousness to the task given by the teacher. In contrast, respondents who view the questions as casual communication has consistently chosen options without overt pronouns.

However, in Q7 where a friend seeks to clarify a piece of information, a more complete response with both pronouns overt is slightly more preferred than that with only subject pronoun. One participant explained that overt pronouns convey information in a clearer, less equivocal manner, it would be clearer to include both pronouns to ensure the message is accurately conveyed.

3.3.3. Group III: A More Formal or Serious Register Is Required

For Group III questions, Q8 - Q11 involve contexts that are more formal than those of the preceding questions, e.g. talking to a police (Q10), being interrogated by a classmate (Q8), or speaking to a teacher under pressure (Q9 and Q11). These contexts entail a more formal register through which the respondents are expected to reinforce the idea they wish to convey, such as highlighting their innocence or the appropriateness of their requests. Q8 and Q11 are open questions, while Q9 and Q10 are multiple choice questions that allow respondents to write down their own phrasing should they find all options uncomfortable (**Table 4**).

Table 3. Participants' responses to Q5 - Q7.

Features	Q5	Q6	Q7	
	No. of responses	No. of responses	Ave	Std Dev
Overt subject only	0	3	-	-
Overt object only	4	0	3.4	0.26
Both overt	2	0	4.0	0.36
Both omitted	3	8	-	-
No difference	2	0	-	-

Table 4. Participants' responses to Q8 - Q11.

Features	No. of responses			
	Q8	Q9	Q10	Q11
Overt subject only	2	8	0	4
Overt object only	0	0	1	0
Both overt	3	1	10	6
Both omitted	6	2	0	0
Sentence-final particle	6	-	-	7
Interjections to express certainty	6	-	2	0

In Group III questions where emphasis of information through a more serious register is required, respondents resort to various ways to incorporate these elements in their responses. 6 out of 11 and 7 out of 11 responses to Q8 and Q11 respectively include sentence-final particles to further buttress their claims, either pointing out the absurdity of the speaker's accusation (Q8) or explaining the motives of their requests to a teacher (Q11). Q8 has also seen short, abrupt, and terse responses that seek to construct a greater force of self-defense, as shown by 6 responses with omitted subject and object pronouns. Moreover, responses to Q9 and Q10 also show a consistent preference for longer answers, with 8 out of 10 adopting sentences with overt subject in Q9 and 10 out of 11 responses with both pronouns overt in Q10.

3.3.4. Group IV: Subject Pronouns Convey New Information Only

In Group IV questions, the respondents are prompted with wh-questions that are broad in nature (e.g. "What happened?"). As such, subject and object pronouns both convey new information in the provided options. Q12 - Q14 are all multiple choice questions, and Q15 is a rating question (**Table 5**).

Responses to Q12 - Q15 show a consistent inclination to employ overt subject and object pronouns. The majority of participants chose options where both pronouns at subject and object position are explicitly expressed, with 10 for Q12 and 9 for Q13 and Q14. Respondents have also assigned the same option in Q15 the highest average score of 4.5 upon 5.0 with a low standard deviation of 0.2.

It can be observed from the data above that participants' knowledge for omission of pronouns seem to be consistent. This implies that, although certain grammatical factors license null subjects in various cases, the occurrence of null subjects is also determined by the specific context where the discourse takes place.

4. Analysis

As discussed in Sections 1 and 2, previous theories on the null-subject phenomenon (Jaeggli 1982; Jaeggli and Safir 1989; Huang 1989; Chomsky 1981, 1982) [7] [8] [9] [11] [12] have mainly attempted to give overall summaries (agreement features and retrievability from context) regarding the distribution of

Table 5. Participants' responses to Q12 - Q15.

Features	Q12	Q13	Q14	Q15	
	No. of responses			Ave	Std Dev
Overt subject only	0	0	0	0	-
Overt object only	1	2	2	3.3	0.3
Both overt	10	9	9	4.5	0.2
Both omitted	0	0	0	0	-

pro-drop in all contexts. However, we need to recognise the nature of languages as a medium for communications, and hence the necessary presence of contexts in which the communications happen. A pragmatic-oriented approach that takes into account the situations where individuals produce certain utterances with overt or omitted pronouns is required to further refine these theories.

4.1. Register

Responses to Group I - III questions show a distinct relationship between the use of null subject and register. The more casual the register, the greater the tendency for speakers to drop subject or object pronouns. Evidence for this includes question Q1 in Group I, where the protagonist's friend wishes to clarify whether the incident of Xiaoming hitting Zhangsan has truly happened. Most of the respondents interpreted this situation as a casual one, in that the protagonist and his friend most likely share a close relationship; moreover, considering that the friend's intention is merely to clarify whether the incident in question has really taken place, a short answer would be succinct and straight to the point. As such, 9 out of 11 respondents have chosen the option with no overt pronouns while giving similar explanation.

In contrast, questions from Group III are of more serious situations. In Q11, where Zhangsan was questioned by a policeman investigating a burglary at his neighbour's house, the context involves a more serious social occasion than that in Q1, and correspondingly participants have quite consistently used more complete sentences.

4.2. Speakers' Agency

Having established that null-subjecthood is also conditioned by the registers employed by speakers, we can see other problems arise regarding how and when speakers decide to employ certain registers and drop/retain their pronouns. The judgment of seriousness or formality of different situations, in turn, varies across speakers based on their own interpretation of the context. Their omitting/retaining the pronouns is therefore also dependent on their own agencies in understanding the situation and reacting differently.

4.2.1. Speakers' Agency with Respect to Seriousness

In everyday contexts, there sometimes may not be a clear distinction between formal and informal situations, as social hierarchical relationships between indi-

viduals are often complex, varying, and to some extent subjective. Because of this ambiguity associated with the social relationship between speakers and listeners, speakers would have an agency to see the situation as more/less serious, and react accordingly.

For example, Q5 is about answering a teacher's question (regarding whether the protagonist has informed his classmate of an announcement). Some participants (3 out of 11) see it as a similar exchange of information as questions in Group I in that the teacher's question, as they have identified, is essentially seeking simple affirmation or negation, and hence preferred short answers with no overt subjects. However, one of them expressed that, as teachers should be of a superior position, the use of overt subject is a sign of respect. Furthermore, two respondents also took into account the issue at question (*i.e.*, that the teacher has asked the student to inform Lisi of the lesson's timing), stating that "the explicit mentioning of *ta* (him) implies the student has treated the task seriously".

The respondents' differing responses and various ways of approaching the question suggest that, even when asked the same question in the same context, native speaker who have the same proficiency for a particular language may provide different answers, and these in turn are dependent on their differing backgrounds, personal experiences and knowledge, characters, etc. The use of null and overt subjects in different situations is therefore partly determined by speakers' own choices, *i.e.* their personal agency.

4.2.2. Speakers' Agency with Respect to Emphasis

The factor of speakers' agency also contributes to the dropping of pronouns through varying interpretations of emphasis. An example is Q7, where Zhangsan's intention is merely to ask his friend for a name mentioned by his teacher, which he did not hear clearly. Two of the four respondents who rated A or B (two options with overt subject pronouns) higher than C (the option without overt pronouns) explained that "an answer that only includes the name Zhangsan is asking for would be the most succinct, and it also answers his question directly without incurring more irrelevant information or simple repetition". This view varies from the five respondents who ranked C as the most comfortable. One of them gave an explanation as stated in Section 3.3.2—a complete answer with overt subject pronoun yields greater clarity—an attribute more crucial than brevity in this case. A further discussion of this factor is presented in Section 5.

4.3. Whether the Pronoun(s) Convey Old/New Information (Principle of Least Effort)

In developing an ecological account of human behaviour, Zipf (1949: vii) [13] emphasises the role played by a Principle of Least Effort as "the primary principle that governs our entire individual and collective behaviour". In the realm of linguistic behaviour, Zipf (1949) [13] distinguishes the opposing pressures exerted by the speakers' economies of effort: minimal linguistic articulation and maximal explicitness [14] (Carston 2005: 271).

This conflict can be seen from Group I and III questions, where the pronouns provide old and new information respectively. As presented in Sections 3.3.1 and 3.3.3, the prevailing trend of utilising omitted subject and object pronouns in Group I questions and making both pronouns explicit in Group III questions is observed.

Previous theories on recoverability (Jaeggli 1982; Huang 1989; Chomsky 1981, 1982) [7] [8] [9] [11] have already speculated such pattern. However, we still need to notice that the uses of explicit pronouns are still clearer and more informative. The Principle of Least Effort takes a step further to provide an account for the motivation to omit pronouns at the expense of clarity of expression—In common situations where speakers do not see the need to express their ideas with extra clarity, they would naturally express in a more succinct manner, giving rise to the above discussed phenomenon of pronoun omission.

5. Further Implications

The analysis in this paper is similar to some previously proposed theories. Chomsky (2005) [15] identifies three separate factors that jointly determine the (I-) language attained: genetic endowment, experience, and the "third factor" (*i.e.*, principles not specific to the faculty of language). Johansson (2013) [16] suggests that the third factor can be "everything that is part of the explanation of language but is not language-specific" and therefore turns out to be "a heterogeneous collection of component factors" which are not limited to general theoretical principles, but also include biological and human-specific factors.

The Principle of Least Effort is drawn in 4.3 to explain the inclination of participants to convey required information with maximum clarity while using the shortest possible utterances. The two constraints are largely of social (maximal explicitness) and biological (minimal articulation) nature. It is therefore reasonable to suggest that the factor of Principle of Least Effort constitutes a 'third factor' in the sense of Chomsky (2005) [15], in that it affects the language used by an individual in a way that is not language-specific.

6. Conclusion

Despite the less significant positions of pragmatic factors in previous theories under the P & P framework, this paper has endeavoured to establish that pragmatic factors in fact play a significant role in the distribution of null-subject phenomenon in Mandarin Chinese. In view of the complexity of the human language and its essential aim of transmitting information, this paper would also like to propose greater emphasis be given to pragmatic factors when future studies attempt to provide a more refined theory for the occurrence of null subjects.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] D'Alessandro, R. (2014) The Null Subject Parameter: Where Are We and Where Are We Headed. Leiden University Centre for Linguistics.

[2] Huang, J.C.T. (1984) On the Distribution and Reference of Empty Pronouns. *Linguistics Inquiry*, 531-574.

[3] Perlmutter, D. (1971) Deep and Surface Constraints in Syntax. Holt, Rinehart and Winston, New York.

[4] Taraldsen, K.T. (1980) On the Nominative Island Condition, Vacuous Application and the That-Trace Filter. Indiana University Linguistics Club, Bloomington, IN.

[5] Rizzi, L. (1982) Issues in Italian Syntax. Foris, Dordrecht.
https://doi.org/10.1515/9783110883718

[6] Peverini, C. (2004) Pro-Drop and the Italian Subjunctive: An Areal Investigation. *The Italianist*, **24**, 107-114. https://doi.org/10.1179/ita.2004.24.1.107

[7] Chomsky, N. (1981) Lectures on Government and Binding. Foris, Dordrecht.

[8] Chomsky, N. (1982) Some Concepts and Consequences of the Theory of Government and Binding. MIT Press, Chicago.

[9] Huang, J.C.T. (1989) Pro-Drop in Chinese: A Generalised Control Theory. In: *The Null Subject Parameter*, Springer, Dordrecht, 185-124.
https://doi.org/10.1007/978-94-009-2540-3_6

[10] Soares, E.C., Miller, P.H. and Hemforth, B. (2019) The Effect of Verbal Agreement Marking on the Use of Null and Overt Subjects: A Quantitative Study of First Person Singular in Brazilian Portuguese. *Forum Linguistico*, **16**, 3579-3600.
https://doi.org/10.5007/1984-8412.2019v16n1p3579

[11] Jaeggli, O. (1982) Topics in Romance Syntax. Foris, Dordrecht.

[12] Jaeggli, O. and Kenneth, J.S. (1989) The Null Subject Parameter and Parametric Theory. In: *The Null Subject Parameter*, Springer, Dordrecht, 1-44.
https://doi.org/10.1007/978-94-009-2540-3_1

[13] Zipf, G.K. (1949) Human Behaviour and the Principle of Least Effort.

[14] Carston, R. (2005) A Note on Pragmatic Principles of Least Effort. *UCL Working Papers in Linguistics*, **17**, 271-278.

[15] Chomsky, N. (2005) Three Factors in Language Design. *Linguistic Inquiry,* **36**, 1-22.
https://doi.org/10.1162/0024389052993655

[16] Johansson, S. (2013) Biolinguistics or Physicolinguistics? Is the Third Factor Helpful or Harmful in Explaining Language? *Biolinguistics*, **7**, 249-275.