

Haplology of Cantonese BEI: Evidence in Production

Chun Hau Ngai

Indiana University, Bloomington

Abstract

This study examines the validity of empty categories in the context of Cantonese dative constructions. Specifically, it explores the emergence of a second 'give' verb, *bei2*, conditioned by phonological distance. This research builds upon Tang's (1998), and Xu and Peyraube's (1997) proposal that the Cantonese serial verb construction originates from the canonical double object construction. Here, the second *bei2* is replaced by a null morpheme to avoid identical syllables. However, the second *bei2* reappears when the phonological distance between the verbs increases. To examine this effect, 30 Cantonese speakers participated in a picture description task in which logographs primed the use of *bei* as the main verb in either a canonical double object construction or a serial verb construction. Although the fixed effect of phonological distance did not reach significance, a significant intercept was found, suggesting a preference for canonical double-object constructions when the theme is short. The results provide initial evidence for a syllable-based effect on the use of the double-*bei* construction. The findings are discussed in relation to psycholinguistic concepts, such as weight, and implications for the acquisition of Cantonese grammar. This study contributes to the understanding of haplology in Cantonese and sheds light on language production and acquisition.

Keywords: Haplology, Cantonese, Psycholinguistics

Haplology of Cantonese *BEI*: Evidence in Production

A typical lexical item, such as ‘*dog*’ or ‘*eat*’, is comprised of various components, including phonological, semantic, and syntactic information. However, according to Jackendoff (1997, 2002), not all lexical entries are composed of these facets. Expressions like ‘*ouch*’ and ‘*uh*’ lack syntactic properties, yet they still constitute part of the mental lexicon, despite these deficiencies. In some theoretical linguistic traditions, entries consisting of only syntactic properties are sometimes proposed to capture certain syntactic patterns (e.g., Chomsky's (1981) Government and Binding Theory). These entries are commonly referred to as ‘empty categories.’ The following sentences from Oshita (2004, p. 2) are examples of such patterns due to the presence of a trace, one type of deficient lexical entry:

(1a) Which prize_{*i*} do you want to/wanna win *t_i*?

(1b) Which contestant_{*i*} do you want *t_i* to/*wanna win the prize?

To many English speakers, the ‘wanna contraction’ is deemed grammatical only in sentence (1a), and not in sentence (1b). This divergence can be explained by the presence of a trace that is left after *want* when the constituent *which contestant* is displaced. This example demonstrates that empty categories can impose constraints on the set of grammatical sentences.

From a psycholinguistic perspective, the presence of empty categories presents a theoretical challenge that requires justification. Specifically, one might wonder how speakers generate and process these lexical entries. According to Fodor (1989), comprehenders face a challenge in processing empty categories since these categories are not directly embedded in the acoustic signal. Thus, the presence of such entries requires an inference from the linguistic context and associations with previous antecedents (e.g., filler-gap dependencies as described in Fodor (1989)). On the other hand, regarding production, the question is how empty categories

are accessed during sentence formation. For most lexical entries, words are stored as separate lemmas and lexemes, which hold syntactic and phonological information, respectively Kempen and Huijbers (1983). The next question is how such items are retrieved and whether the retrieval of empty categories involves a similar process or whether retrieval is completely by-passed given the absence of phonological content.

In light of these theoretical questions regarding the processing and production of empty categories, the psychological reality of such phenomena requires justification. This article explores the occurrence of such empty categories in production by examining null morphemes in one Cantonese dative construction. First, the various constructions used to convey object transfer in Cantonese are discussed, including the four possible constructions that may occur with *bei2* (俾; 'give'), including examples and constraints. The article then elaborates on Tang (1998) and Xu and Peyraube (1997) null morpheme hypothesis, which suggests that a second *bei2* is omitted in the double-object construction to avoid redundancy. This study examines this hypothesis by using a picture description task where the number of syllables of the theme is manipulated. If Tang (1998) analysis is correct, speakers are predicted to use the double-*bei* construction more as the length of the theme increases.

Dative Constructions in Cantonese

The act of giving is a fundamental aspect of human interaction, and as such, it is not surprising to find that a verb 'give' is present in many languages. Generally, the concept of giving entails the transfer of an object from one person to another (Newman, 1996). Conceptually, giving involves three arguments: an agent (the initiator of the event), a theme (the object being transferred), and a recipient (the individual receiving the object).

In Cantonese, there are multiple constructions used to convey the transfer of an object from one person to another, making it one of only a few languages that permit such variation (Matthews & Yip, 2013). Here, I discuss the four Cantonese dative constructions described in Matthews and Yip (2013). Despite having a rigid word order, Cantonese provides multiple constructions, allowing speakers to express similar ideas. For instance, the Cantonese verb *bei2* can be used in both the canonical [(V)erb (T)heme (R)ecipient] and non-canonical [V R T] constructions¹. The four possible constructions in which *bei2* occurs are as follows:

- Canonical double object construction: [*bei2* T R]
- Non-canonical double object construction: [*bei2* R T]
- Serial-verb construction: [V T *bei2* R] (including the special case of double-*bei*)
- *Zoeng1 (ba)* construction [*zoeng1* T *bei2* R]

Due to the scarcity of Cantonese corpora, there are currently no studies known to the author that exclusively examine the use of dative constructions by adults. Therefore, I only discuss studies examining the use of the dative constructions by both adults and children. By examining longitudinal data in the Hong Kong Cantonese Child Language Corpus (Lee & Wong, 1998) and the Hong Kong Bilingual Child Language Corpus (Yip et al., 2001), Chan (2003) found that children do not apply the same constraints to the dative constructions as adults. Unlike adults, the use of the non-canonical double object construction and double-*bei* construction are not affected by register or the length of the theme. See below for the specific conditions for the use of these constructions.

¹ The use of the non-canonical [V R T] construction is subject to specific constraints. Details of the constraints on the non-canonical [V R T] construction can be found in the section on non-canonical double object constructions.

The Cantonese examples discussed in this paper are transcribed using the JyutPing Romanization system. Tones are indicated by the numbers in the transcriptions. Abbreviations used are: CL (classifier), SG (singular), MOD (modifier), PFV (perfective marker), EXP (experiential), and SFP (sentence final particle).

The Canonical Double Object Construction

The canonical word order for denoting transfer of possession (henceforth TOP) in Cantonese is [V T R]. In this construction, both the theme and the recipient are located in post-verbal position without additional morphology (2-3) (Chan, 2003; Yip & Matthews, 2007).

- (2) 我 俾 本 書 你
Ngo5 bei2 bun2 syul nei5
 1.SG give CL book 2.SG
 'I give the book to you.'
- (3) 我 俾 錢 你
Ngo5 bei2 cin2 lei5
 1.SG give money 2.SG
 'I give you money.'

The Non-Canonical Double Object Construction

Another construction is the non-canonical double object construction, where the word order is [V R T]. Corpus data from bilingual Cantonese-speaking children between three and four years old shows less frequent use of this construction than the canonical variant (Chow, 2005). In (4), the recipient immediately follows *bei2*, and the theme follows the recipient. Additionally, verbs in this construction are not restricted to *bei2* (5).

- (4) 我 俾 你 本 書
Ngo5 bei2 nei5 bun2 syul
 1.SG give 2.SG CL book
 'I give you a book.'

- (5) 我 教 你 中文
Ngo5 gaau3 nei5 zung1 man4
 1.SG teach 2.SG Chinese
 ‘I teach you Chinese.’

The Serial-Verb Construction

The use of serial-verb constructions is also prevalent in colloquial Cantonese (Tang, 1998).

As illustrated in 6-7; (Yip & Matthews, 2007), the theme precedes the recipient, and the two are separated by a co-verb.

- (6) 我 遞 本 書 俾 你
Ngo5 dai6 bun2 syu1 bei2 nei5
 1.SG hand CL book give 2.SG
 ‘I hand you a book.’
- (7) 我 寄 (封) 信 俾 你
Ngo5 gei3 fung1 seon3 bei2 nei5
 1.SG mail CL letter give 2.SG
 ‘I mail you a letter.’

The *Zoeng1 (Ba)* Construction

The *zoeng1* construction takes the form [*zoeng1* T *bei2* R]². This construction is used to express displacement. In (8-9), *zoeng1* precedes the co-verb *bei2*, and the theme intervenes between them.

- (8) 我 將 本 書 俾 你
Ngo5 zoeng1 bun2 syu1 bei2 nei5
 1.SG put CL book give 2.SG
 ‘I remove the book and give it to you.’
- (9) 我 將 啲 資料 俾 你
ngo5 zoeng1 di1 zi1-liu2 bei2 nei5
 1.SG zoeng1 CL information give 2.SG
 ‘I remove the information and give it to you.’

² This construction is also known as the *ba* (把)-construction in Mandarin (Bauer & Matthews, 2003; Cheung, 2006; Matthews & Yip, 2013).

The *zoeng1* construction is rarely used in spontaneous speech due to its restricted nature. As discussed in Cheung (1992) and Matthews and Yip (2013), this construction is highly restrictive, allowing only definite or generic direct objects and bounded events. Therefore, it is unsurprising that the use of *zoeng1* constructions is very infrequent in child speech. In fact, Mai et al. (2018) investigated child language corpora and discovered that the use of this construction is exceedingly rare. Out of the seventeen children studied, only one monolingual and two bilingual children produced *zoeng1* constructions.

The Movement Hypothesis and the Null Morpheme Hypothesis

One key observation is the variation among the dative constructions concerning the verbs they permit. Table 1 summarizes the ability of various verbs to occur in the different constructions. ‘Give’ is allowed in most constructions discussed, but verbs like ‘pluck’ and ‘send’ are restricted to the serial-verb construction (Chan, 2003).

Table 1. The occurrence of various verbs as the main verb of the dative constructions.

	‘give’	‘send’ and ‘pluck’	‘teach’	<i>zoeng1</i>
Canonical double-object construction [V T R]	☑	*	*	*
Non-canonical double-object construction [V R T]	☑	*	☑	*
Serial-verb construction [V T V R]	☑	☑	*	*
Zoeng1 construction [zoeng1 T bei2 R]	*	*	*	☑

The Movement Hypothesis

Movement has been offered as a possible explanation for the overlap of verbs in these constructions. Cheng (1988) and Huang and Zhan (1983) postulated that the canonical [V T R] construction is derived from the non-canonical [V R T] construction through movement of the

object. However, this account is challenged by the 'teach'-type verbs, which are only permitted in the non-canonical double object construction. Furthermore, the movement account has difficulty explaining the distribution of semantically ambiguous verbs like *ze3*, which can mean either 'lend' or 'borrow'. The ambiguity is retained in the non-canonical double-object construction (10) but is ungrammatical in the canonical double-object construction (11). Therefore, unless semantic conditions are imposed on transformational rules, it is unlikely that a derivational relationship exists between canonical and non-canonical double-object constructions.

- (10) 我 借 佢 一百蚊
ngo5 ze3 keoi5 jat1-baak3-man1
 1.SG borrow/lend 3.SG one hundred dollars
 'I borrowed one hundred dollars from him.'
 'I lent him one hundred dollars.'
- (11) *我 借 一百蚊 佢
ngo5 ze3 jat1-baak3-man1 keoi5
 1.SG lend one hundred dollars 3.SG
 'I lent one hundred dollars to him.'

The Null Morpheme Hypothesis

Alternatively, Tang (1998) Xu and Peyraube (1997) proposed that the canonical double-object [V T R] construction is derived from the serial-verb [V T V R] construction³. Under this hypothesis, the *bei2* occupying the co-verb position in the serial-verb construction is removed in the surface phonetic form, resulting in the canonical double-object construction (12).

- (12) *bei2* T *bei2* R → *bei2* T \emptyset R

Phonetic data reveals a distinct pause or syllable lengthening between the theme and recipient in the canonical double-object construction (Qiao, 1966). This pause, indicated by '//',

³ See Chin (2011) and Huang (2012) for a similar explanation for Hakka.

is only permitted in the canonical double-object construction (13), not the non-canonical construction (14). According to Selkirk's (1986) Silent Demibeat Addition Hypothesis, pauses indicate syntactic boundaries, and the null *bei2* requires two demibeats (an extended pause or syllable lengthening) between the theme and recipient. As illustrated in the partial structures of (13-14), because the null morpheme (denoted as ' \emptyset ') constitutes a separate verb phrase with the recipient (15), an additional pause is needed to reflect the structural difference between canonical and non-canonical double-object constructions (compare 16).

- (13) 我 俾 咗 個 本 中文書// 小明
ngo5 bei2 zo2 go3 bun2 zung1 man4 syu1 siu2 ming4
 1.SG give PFV that CL Chinese book Siu-Ming
 'I gave that Chinese book to Siu-Ming.'
- (14) ?我 俾 咗 小明// 本 中文書
ngo5 bei2 zo2 siu2 ming4 bun2 zung1 man4 syu1
 1.SG give PFV Siu-Ming CL Chinese book
 'I gave Siu-Ming a Chinese book.'
- (15) Chinese book]_N]NP [[[\emptyset]_V [Siu-Ming]_N]NP]VP
- (16) Siu-Ming]_N]NP [[Chinese book]_N]NP

The existence of a null morpheme is further supported by the dispreference for the double-*bei* construction, where two identical *bei2* verbs occupy the main verb and co-verb positions. (17-19) are arranged in descending order by the number of syllables separating the two instances of *bei2*. In (17), the verbs are separated by an aspect marker and a modified noun phrase, resulting in a distance of ten syllables. In contrast, (18) has only two syllables between the two verbs, resulting in a degraded sentence. To generate a sentence with adjacent *bei2*'s, the theme is topicalized⁴ in

⁴ Although Cantonese possesses a rather rigid word order, topicalization is permitted (Matthews & Yip, 2013).

(19). (19) is ill-formed due to the adjacency of homophonous *bei2* tokens. As the phonological distance decreases, acceptability diminishes.

- (17) 我 俾 咗 本 用 中文 寫 嘅 語法
Ngo5 bei2 zo2 bun2 jung6 zung1man4 se2 ge3 jyu5faat3
 1.SG give PFV CL use Chinese write MOD grammar
 書 俾 佢
syu1 bei2 keoi5
 book give 3.SG
 ‘I gave a grammar book written in Chinese to him/her.’
- (18) ??我 俾 本 書 俾 佢
Ngo5 bei2 bun2 syu1 bei2 keoi5
 1.SG give CL book give 3.SG
 ‘I give a book to him/her.’
- (19) *嗰 本書 我 會 俾 俾 佢
Go2 bun2 syu1 ngo5 wui5 bei2 bei2 keoi5
 that CL book 1.SG will give give 3.SG
 ‘As for that book, I will give it to him.’

To satisfy the constraint on repeating identical elements, native speakers tend to replace the second *bei2* with a null morpheme. As the phonological distance between the identical syllables widens, the second occurrence of *bei2* is less likely to be deleted, resulting in the double-*bei* construction. This hypothesis also accounts for the types of verbs permitted in the canonical double-object construction: *bei2* is the only verb that leads to repetition, making it the only verb permitted in the canonical double-object construction. Other verbs used in the serial-verb construction do not constitute a repetition, thus they do not trigger the same deletion.

In summary, the Movement Hypothesis fails to explain the distribution of permitted verbs, and the Null Morpheme Hypothesis offers a more viable explanation by positing a derivational relation between the canonical double-object construction and the serial-verb construction: the deletion of the co-verb.

Current Study

The goal of this study is to examine the validity of the Null Morpheme Hypothesis by determining whether the use of the double-*bei* construction is sensitive to the phonological distance between the two *bei2* verbs. Since Tang (1998) and Xu and Peyraube's (1997) hypotheses are grounded in syntax, this hypothesis does not specify the conditions for the phonological mechanism and does not provide concrete means for determining phonological distance. This deficiency underscores the need for research aimed at illuminating the conditions underlying the use of *bei2* in Cantonese. The Null Morpheme Hypothesis predicts an increase in the use of double-*bei* and a gradual decrease in the use of the canonical-double object construction as the phonological distance between the main verb and co-verb increases.

To evaluate the association between phonological distance and construction use, a working operationalization of phonological distance is necessary. Here, phonological distance is defined as the number of syllables between the first syllable after the main verb *bei2* and the first syllable of the recipient. The calculation of phonological distance is exemplified in (20), which has a total of six syllables, including the perfective aspect marker (*zo2* 咗) theme (*go3 saang1-jat6-daan6-gou1* 個生日蛋糕).

- (20) 個 男仔 俾咗 個 生日蛋糕 個女仔
Go3 naam4 zai2 bei2 zo2 go3 saang1-jat6-daan6-gou1 go3 nei5 zai2
 CL boy BEI PFV CL birthday cake CL girl.
 'The boy gave a birthday cake to the girl.'

Method

Participants

Thirty-one Cantonese speakers residing in Hong Kong were recruited for the experiment through online postings and posters in various public areas. One participant was excluded from

the dataset due to self-identified limited Cantonese proficiency. The remaining participants were between the ages of 21 and 26 and identified themselves as Cantonese native speakers (n=27) or fully proficient in Cantonese (n=3). The subject pool consisted of 22 males and 8 females. None reported speech or hearing impairments.

Materials

Participants completed a picture description task. Prior to the experiment, a pilot study was conducted to ensure that the experimental drawings could naturally elicit events of 'giving.' All pictures displayed events involving two individuals of different genders, introduced as 'the girl' (個女仔) or 'the boy' (個男仔). Of the 32 drawings, 16 were the target items depicting a transfer of possession event with various objects. One of the challenges of the current study was to elicit themes with the desired syllable count. To this end, the drawings displayed themes with varying numbers of expected syllables. There were two items per expected syllable count, ranging from two syllables (e.g., 'the cup' 個杯 *go3 bui1*) to five syllables (e.g., 'the birthday cake' 個生日蛋糕 *go3 saang1 jat6 daan2 gou1*)⁵. This range was chosen because the themes used in free conversation from the PolyU Corpus of Spoken Chinese in which *bei2* was used as the main verb only ranged from one to five syllables. For each expected syllable count, there were two items, totaling eight target items (Appendix A). All 32 line drawings from the pilot study were used in the final experiment.

⁵ Although a classifier is mandatory in noun phrases, numerals are optional prior to the classifier (Matthews & Yip, 2013). Therefore, the number of syllables could also range between two and six.

To mask the purpose of the experiment, a superimposed character was added to each presentation of a drawing. For the 16 target drawings, three versions of each drawing were generated, resulting in a total 48 drawings used. The three versions differed only in the superimposed logograph (*bei2* 俾, *zoeng1* 將, or *sung3* 送) used to prime the use of that verb. All logographs were well-suited to describe the TOP drawings. *Bei2* could be used as the main verb or co-verb in the canonical double-object construction [*bei2* T R], the non-canonical double object construction [*bei2* R T], and the serial-verb construction [V T *bei2* R]. *Sung3*, which roughly means 'deliver', could be used in the serial-verb construction [V T *bei2* R]. Finally, *zoeng1* could be used in the *zoeng1* construction [*zoeng1* T *bei2* R]. The remaining 36 drawings were fillers, resulting in 96 total items. All fillers depicted intransitive actions with logographs suited to describe that drawing superimposed. For example, a drawing depicting the male character standing had the character for *kei5* 'stand' (企) overlaid. The target drawings without the superimposed logographs can be found in Appendix B.

All stimuli were equally distributed into three blocks, with each block consisting of 16 filler and 16 target drawings. Within each block, there were three pictures each corresponding to the *zoeng1*, *bei2*, and *sung3* logographs. Drawings that only differed in logograph were not in the same block. The blocks were pseudo-randomized based on the following criteria:

- a. Target drawings were not allowed to occur in three successive trials.
- b. The location of the characters (left-side vs. right-side) could not be identical in three successive trials.
- c. Identical events with different logographic characters were prohibited from occurring in successive trials (excluding TOP verbs)
- d. Identical logographic characters were prohibited from occurring in three successive trials.

At the end of the experiment, a survey was issued to the participants to ensure their native or near-native proficiency in Cantonese. The survey included questions about their language background, such as native language, parental language ability, and Cantonese proficiency, as well as biographical information, such as family background and education level. See Appendix C for the complete questionnaire. Participants were also asked to posit the goal of the experiment.

Procedure

All experiments were conducted and recorded online via Zoom meeting software with informed consent. Pictures were displayed using Microsoft PowerPoint. The pictures were superimposed with the primed verb presented in its written form. The verbs were written in red and located in the middle of the screen to ensure that participants were aware of the characters. The display of the written logographs and the drawing was synchronous. Participants were instructed to describe the picture and were given the option to use the target verb or not. The instructions were only displayed prior to the first block. The message remained on the screen until the participant initiated the experiment. Prior to each block participants completed three practice items. Since this study is only interested in the choice of construction, a fixation cross with no time limit was displayed to introduce separation between trials. Upon completing the description of a drawing, participants pressed the space bar on their keyboard to proceed to the next trial. There was no time limit, and the experiment lasted approximately 20 minutes. After the picture description task, participants completed the background survey.

Scoring and analysis

A total of 219 responses with a main verb *bei2* in the canonical double object construction or serial verb construction were recorded and coded by the author. Responses that

were not one of these constructions or used a different verb were not analyzed (e.g., non-canonical double-object constructions, *zoeng1* constructions, and serial-verb constructions with *sung3* as the main verb). Responses were coded for single- or double-*bei* and phonological distance. The following examples illustrate how responses were coded:

(21) Double-*bei*; phonological distance = 3

個 男仔 俾 個 鬧鐘 俾 個 女仔
 go3 naam4 zai2 bei2 go3 naau6 zung1 bei2 go3 nei5 zai2
 CL boy BEI CL alarm clock BEI CL girl.
 ‘The boy gives an alarm clock to the girl.’

(22) Single-*bei*; phonological distance = 6

個 女仔 俾 咗 個 生日蛋糕 個 男仔
 go3 nei5 zai2 bei2 zo2 go3 saang1 jat6 daan2 gou1 go3 naam4 zai2
 CL girl BEI PFV CL birthday cake CL boy
 ‘The girl gave a birthday cake to the boy.’

Results

To examine Tang's (1998) and Xu and Peyraube (1997) Null Morpheme Hypothesis, the frequency of tokens across phonological distances was analyzed. As illustrated in Table 2, the number of tokens (collapsing the two analyzed constructions) sharply decreases after three syllables. There appears to be a general tendency to shift from the double-object construction to the double-*bei* construction as phonological distance increases.

Table 2. Number of tokens by construction and phonological distance.

	Phonological distance					
	2	3	4	5	6	7
Canonical double object	72	73	8	2	1	0
Double-<i>bei</i>	12	34	3	5	3	1
Percentage of double-<i>bei</i>	14.29%	31.78%	27.27%	71.43%	75%	100%
Total	84	107	11	7	4	1

Since the assumptions of independence of residuals were not met, a logistical generalized linear mixed model (GLMM) was fitted to examine the relationship between construction type and distance while accounting for random effects of participant. For the purpose of this analysis, the use of single-*bei* was coded as 0 and double-*bei* as 1.

All statistical analyses were conducted using R Studio (RStudio Team, 2020). The model assumptions of linearity and homoscedasticity were met⁶. As seen in Table 3, results of the GLMM revealed a significant intercept ($\beta = -8.2435$, $SE = 2.3079$, $z = -4.012$, $p < 0.001$) and a non-significant effect of distance on the choice of construction ($\beta = 0.2623$, $SE = 0.3051$, $z = 0.598$, $p = 0.389929$). The random intercept variance for subjects was estimated as 53.69 ($SD = 7.327$). The test suggests no interaction between phonological distance and choice of construction.

Table 3. Output of the GLMM.

	Estimate	SE	z	p
Intercept	-8.2435	2.3079	-4.012	0.001
Phonological distance	0.2623	0.3051	0.598	0.389929

Discussion

Despite the non-significant effect of phonological distance on construction choice, the present study provides initial evidence supporting the Null Morpheme Hypothesis. A visual examination of the distribution of tokens provides some evidence that an effect exists. Upon examining the frequency of the double-*bei* construction across phonological distance, it was found that the percentage of double-*bei* tokens increases as the phonological distance between

⁶ For the complete statistical output, see the OSF repository: <https://osf.io/hcsvd/>

two identical elements grows. This finding aligns with observations made by Tang (1998) and Xu and Peyraube (1997). According to Tang (1998) and Xu and Peyraube (1997), the canonical double object construction originates from the double-*bei* construction, and the acceptability of such sentences depends on the presence of the homophonous second *bei2*. As the phonological distance between the two verbs increases, the co-verb *bei2* reappears overtly. Notably, the responses here indicate that the number of syllables required for the double-*bei* construction to be produced is less than ten syllables, contrary to the example in Tang (1998). Additionally, the current experiment found an increase in the use of the double-*bei* construction above the four-syllable threshold, suggesting a gradient phonological rule rather than a dichotomous one.

While the findings of this study provide interesting insights into the relationship between phonological distance and the choice of construction in Cantonese, it is important to acknowledge the limitations of the current study and ways to improve on it. The lack of a significant effect of distance in this study could be attributed to two factors. First, given the unrestrictive nature of the task, the majority of collected responses involved constructions that were not the focus of this study, such as non-canonical double object constructions, serial-verb constructions, and *zoeng1* constructions. Second, most of the tokens were found to cluster around lower phonological distance. Out of the 219 tokens using the construction-of-interest, only a minority of participants (N = 9), in only 23 tokens, produced a phonological distance greater than four syllables. For future studies, it would be recommended to incorporate a larger sample size (n > 30) and implement stricter control over output. By limiting possible productions and increasing the number of participants, future studies could increase the number of relevant responses, which in turn increases the likelihood of detecting a significant effect of phonological distance on construction choice.

As noted by reviewers, the emergence of the co-verb *bei* in Cantonese parallels the use of resumptive pronouns in more complex structures in English, both aimed at alleviating processing challenges. In Cantonese, where morphological marking is absent, the parser heavily relies on word order to identify the theta roles of arguments. With Cantonese possessing both [V T R] and [V R T] word order, word order alone is insufficient in distinguishing the theme from the recipient. The introduction of the co-verb *bei* can be seen as a strategy to resolve ambiguity by introducing an explicit marker between the two nouns, similar to the use of resumptive pronouns in English. This parallel is particularly evident when considering resumptive pronouns in complex relative clauses. While resumptive pronouns are generally disallowed in English, they are used in sentences containing long filler-gap dependency to compensate for difficulties in accessing certain constituents (Ferreira & Swets, 2005; Hawkins, 2004; Kroch, 1981; Shlonsky, 1992). Thus, the use of the co-verb *bei* in Cantonese and the use of resumptive pronouns in English appear to share a common motivation rooted in facilitating the processing of syntactic structures.

Nevertheless, the results indicate that the use of the double-*bei* construction in Cantonese is a nuanced and subtle phenomenon sensitive to various factors, such as the number of syllables and the weight of the theme. These findings empirically support Tang's (1998) and Xu and Peyraube's (1997) accounts with the deletion of the second *bei*₂ in Cantonese, suggesting that production-based factors may influence the choice of syntactic constructions. The role of constituent weight in determining word order has been extensively studied in linguistics (Behaghel, 1909; Benor & Levy, 2006; Quirk, 2010; Wasow, 2002). For instance, there are two equivalent constructions available to convey transfer of possession in English, namely the double object dative construction (e.g., John gave Mary the book) and the prepositional dative

construction (e.g., John gave the book to Mary). Speakers prefer placing longer and more complex constituents at the end of the sentence (i.e., the ‘short before long’ preference) (Arnold et al., 2000; Stallings et al., 1998). This preference may also influence the acceptability of the double-*bei* construction in Cantonese. Although the order of constituents remains the same, the choice of construction is sensitive to the weight of the theme, with the second *bei2* phonologically realized when there is sufficient distance between two identical word forms.

Although the findings are not conclusive, the current study raises some questions about the acquisition of Cantonese syntax. Following Yip and Matthews (2007), the double-*bei* construction, a type of serial-verb construction, is thought to be important for acquiring [V T R] word order in Cantonese. However, if children struggle to acquire the haplology of the co-verb *bei2*, it would suggest that they have not fully acquired the [V T R] word order. As argued in Chomsky (1986), grammatical judgements reflect linguistic competence, so children who have yet to acquire the ability to use the double-*bei* construction would have a more limited repertoire of grammatical constructions. Future research should explore this idea by using graded judgement task. In this task, children are asked to participate in a game where a fictional character describes a photo. The children are asked to determine whether the statement corresponds to the picture (Van Hout et al., 2010; Veenstra et al., 2017). Such studies would provide additional insights into the acquisition of Cantonese syntax and shed light on the mechanisms underlying language development.

In conclusion, this study provides some valuable insights into the relationship between phonological distance and the choice of dative construction in Cantonese. However, the findings should be interpreted with caution, and future studies should consider the limitations of the study and explore alternative means for testing the Null Maker Hypothesis. The nuanced and subtle

nature of the double-*bei* construction suggests a need for a larger sample size ($n > 30$) in future studies to obtain accurate and reliable results. Furthermore, the implications of these findings extend to the acquisition of Cantonese syntax, where the double-*bei* construction is a prerequisite for the [V T R] word order. Examining Chomsky (1986) proposition that grammaticality judgement reflects linguistic competence through graded judgment tasks with Cantonese-speaking children could provide further insights into language acquisition and the construction's underlying mechanisms. By examining the complex interplay between phonological distance and other linguistic features, future studies can provide a more comprehensive understanding of the factors that influence the choice of construction in Cantonese.

References

- Arnold, J. E., Losongco, A., Wasow, T., & Ginstrom, R. (2000). Heaviness vs. Newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, 76(1), 28–55. <https://doi.org/10.1353/lan.2000.0045>
- Bauer, R. S., & Matthews, S. (2003). *Cantonese*. Routledge.
- Behaghel, O. (1909). Beziehungen zwischen umfang und reihenfolge von satzgliedern. *Indogermanische Forschungen*, 25, 110.
- Benor, S. B., & Levy, R. (2006). The chicken or the egg? A probabilistic analysis of English binomials. *Language*, 233–278.
- Chan, A. (2003). *The Development of Bei2 Dative Constructions in Early Child Cantonese* (Issue November).
- Cheng, L. S. (1988). *Dative constructions in Mandarin and Cantonese*. MIT.
- Cheung, K. S. (2006). *Processing factors in language comprehension and production: The case of Cantonese dative constructions*. Hong Kong University.
- Cheung, S. H. N. (1992). The pretransitive in Cantonese. *Zhongguo Jingnei Yuyan Ji Yuyanxue*, 1, 241–303.
- Chin, A. C. (2011). Grammaticalization of the Cantonese double object verb [pei 35] in typological and areal perspectives. *Language and Linguistics*, 12(3), 529–563.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht.
- Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. Greenwood Publishing Group.

- Chow, C. (2005). *The effects of parental input on the early development of bei2 dative constructions in Cantonese children*. The University of Hong Kong (Pokfulam, Hong Kong).
- Ferreira, F., & Swets, B. (2005). The production and comprehension of resumptive pronouns in relative clause “island” contexts. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (pp. 263–278). Routledge.
- Fodor, J. D. (1989). Empty Categories in Sentence Processing. *Language and Cognitive Processes*, 4(3–4), S1155–S1209. <https://doi.org/10.1080/01690968908406367>
- Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. OUP Oxford.
- Huang, H. (2012). Dative constructions in Hakka: A constructional perspective. *Hakka Studies*, 5(1), 39–72.
- Huang, J. J., & Zhan, B. (1983). Guangzhou fangyan zhong de teshu yuxu xianxiang [Exceptional cases in word order of Guangzhou dialect]. *Yuyan Yanjiu [Studies in Languages]*, 5, 121–126.
- Jackendoff, R. (1997). *The architecture of the language faculty*. MIT press.
- Jackendoff, R. (2002). English particle constructions, the lexicon, and the autonomy of syntax. In N. Dehé, R. Jackendoff, A. McIntyre, & S. Urban (Eds.), *Verb-particle explorations* (pp. 67-94). Mouton de Gruyter.
- Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition*, 14(2), 185-209.
- Kroch, A. S. (1981). On the role of resumptive pronouns in amnestying island constraint violations. *Chicago Linguistic Society*, 17, 125–135.

- Lee, T., & Wong, C. (1998). CANCORP: The Hong Kong Cantonese Child Language Corpus. *Cahiers de Linguistique Asie Orientale*, 27(2), 211–228.
- Mai, Z., Kwan, C., & Yip, V. (2018). Expressing displacement in heritage Cantonese: Cross-linguistic influence and structural vulnerability. *International Journal of Bilingualism*, 22(6), 603–618.
- Matthews, S., & Yip, V. (2013). *Cantonese: A comprehensive grammar*. Routledge.
- Newman, J. (1996). *Give: A cognitive linguistics study*. Walter de Gruyter.
<https://doi.org/10.1515/9783110811933.607>
- Oshita, H. (2004). Is there anything there when there is not there? Null expletives and second language data. *Second Language Research*, 20(2), 95–130.
- Qiao, Y. (1966). *Guangzhouhua Kouyuci de Yanjiu [A study of Cantonese colloquial expressions]*. Overseas Chinese Languages Publishing Company.
- Quirk, R. (2010). *A comprehensive grammar of the English language*. Pearson Education India.
- RStudio Team. (2020). *RStudio: Integrated Development Environment for R*.
- Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology*, 3, 371–405.
- Shlonsky, U. (1992). Resumptive pronouns as a last resort. *Linguistic Inquiry*, 23(3), 443–468.
- Stallings, L. M., MacDonald, M. C., & O’Seaghdha, P. G. (1998). Phrasal ordering constraints in sentence production: Phrase length and verb disposition in heavy-NP shift. *Journal of Memory and Language*, 39(3), 392–417. <https://doi.org/10.1006/jmla.1998.2586>
- Tang, S.-W. (1998). On the ‘inverted’ double object construction. *Studies in Cantonese Linguistics*, 35–52.
- Van Hout, A., Harrigan, K., & de Villiers, J. (2010). Asymmetries in the acquisition of definite and indefinite NPs. *Lingua*, 120(8), 1973–1990.

- Veenstra, A., Hollebrandse, B., & Katsos, N. (2017). Why some children accept under-informative utterances: Lack of competence or pragmatic tolerance? *Pragmatics & Cognition*, 24(2), 297–313.
- Wasow, T. (2002). *Postverbal behavior*. CSLI Stanford.
- Xu, L., & Peyraube, A. (1997). On the double object construction and the oblique construction in Cantonese. *Studies in Language. International Journal Sponsored by the Foundation "Foundations of Language,"* 21(1), 105–127.
- Yip, V., & Matthews, S. (2007). Dative construction with bei2 “give” in bilingual children’s Cantonese. In *The bilingual child: Early development and language contact*. (pp. 90–113). Cambridge University Press. <https://doi.org/10.1002/9780470756997.ch4>
- Yip, V., Matthews, S., & Huang, Y.-Y. (2001). *The Hong Kong Bilingual Child Language Corpus*.

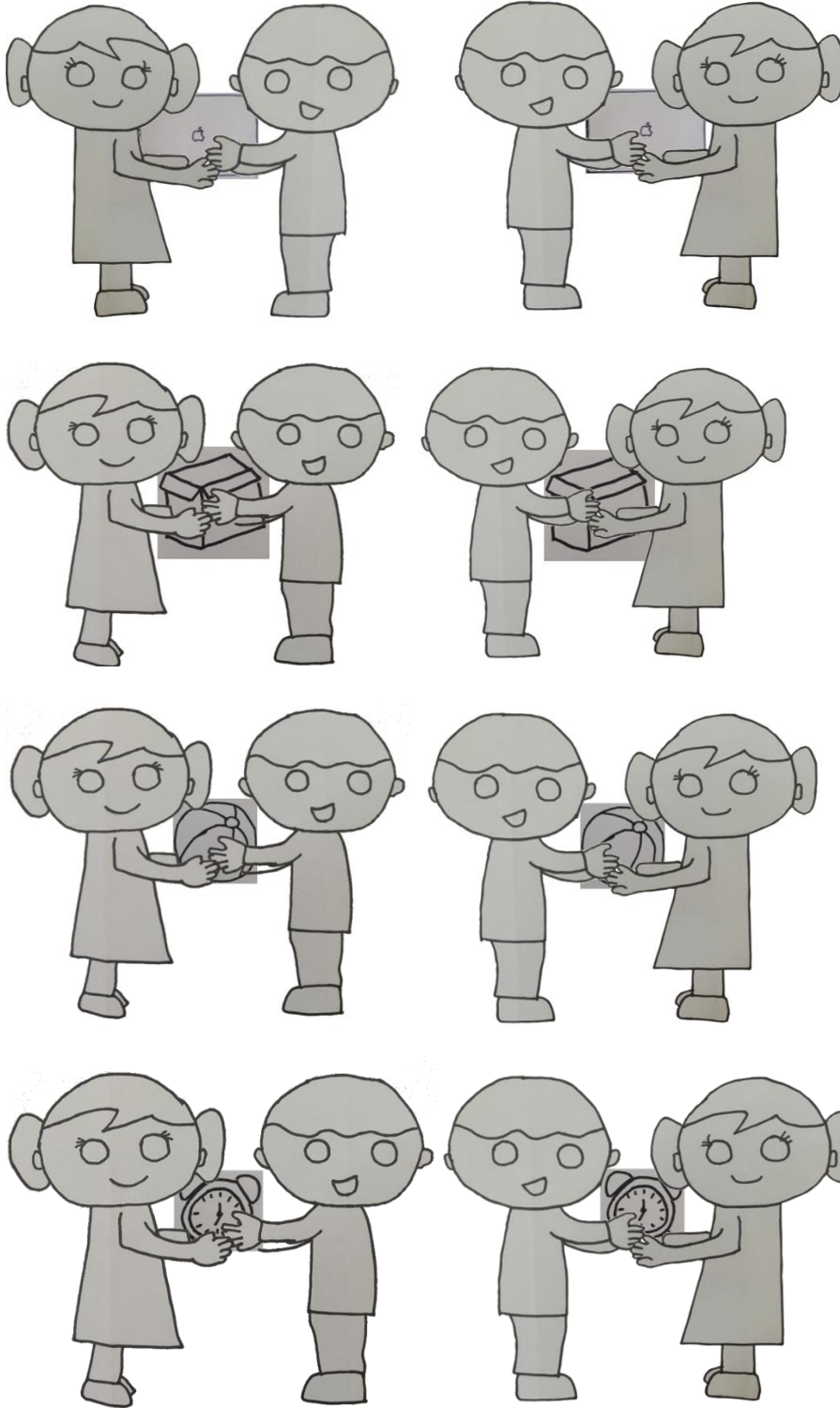
Appendix A

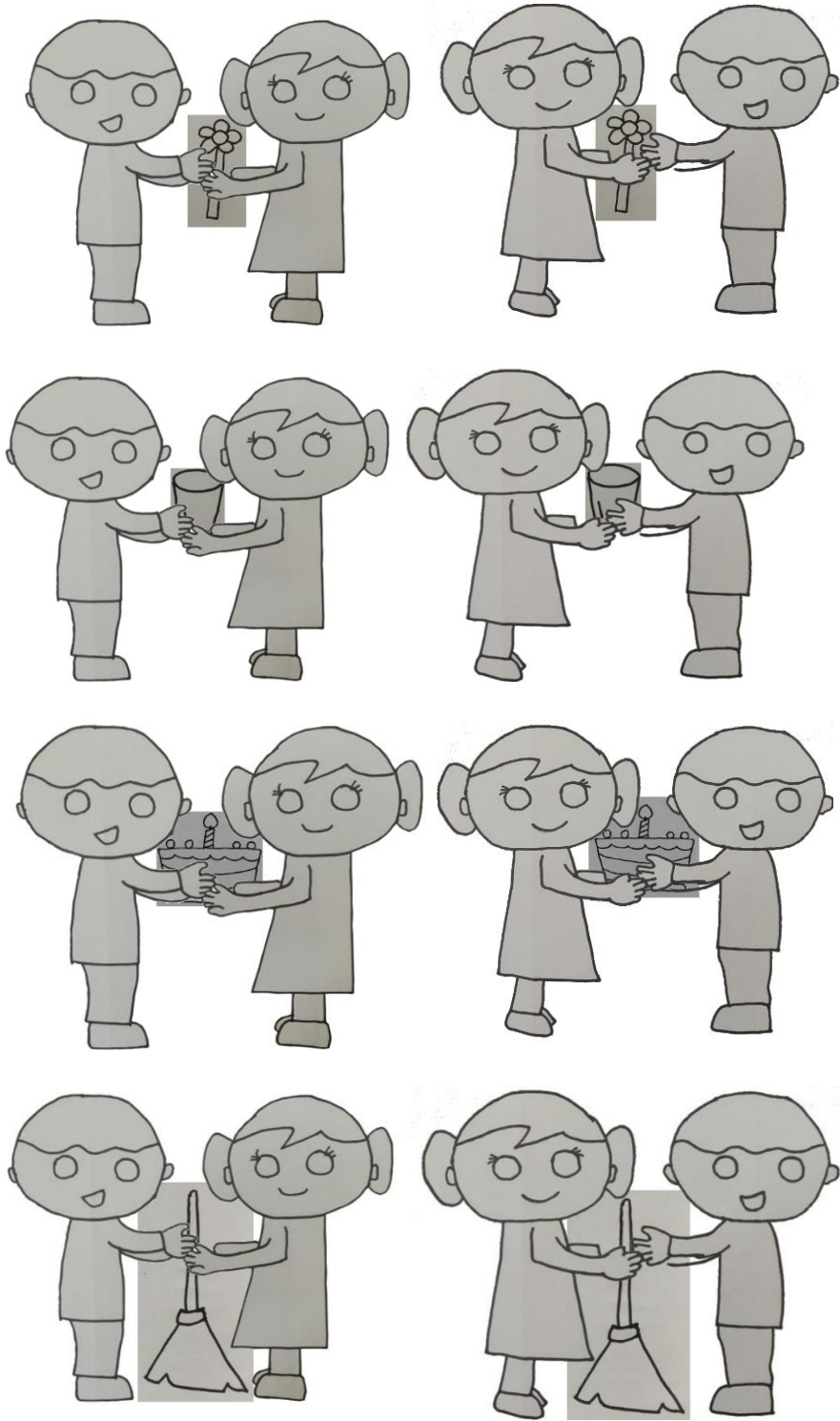
Expected themes

Syllable Count	2	3	4	5
	個波 go3 bo1 'The ball'	個鬧鐘 go3 naau6 zung1 'The alarm clock'	個紙皮箱 go3 zi2 pei4 soeng1 'The cardboard box'	個生日蛋糕 go3 saang1 jat6 daan2 gou1 'The birthday cake'
	朵花 do2 faa1 'The flower'	把掃把 baa2 sou3 baa2 'The broom'	個玻璃杯 go3 bo1 lei4 bui1 'The glass'	部手提電腦 bou6 sau2 tai4 din6 nou5 'The laptop'

Appendix B

Sample pictures





Appendix C

Background Questionnaire

Question No.	Questions in Cantonese	Questions in English
Q001	你的實驗編號	Participant ID:
Q002	實驗進行日期	Date of Experiment(day/month/year)
Q003	年齡	Age
Q004	性別	Gender
Q005	教育程度	Education Level
Q006	你的廣東話水平	Cantonese Proficiency
Q007	你的英語水平	English Proficiency
Q008	你於什麼年齡起開始接觸英語	At what age were you exposed to the language of English?
Q009	你有就你的英語水平參與公開考試嗎?	Have you taken any <i>open exam</i> on your proficiency on English?
Q010	承上題, 你參與的公開考試成績為幾分? (如沒有的話, 請填 " 沒有 ")	As mentioned above, what is your respective grade?
Q011	你認為你的母語為	What is your <i>native language</i> ?
Q012	你還會什麼其他語言	What other languages do you know?
Q013	你於家中最常用作溝通的語言為	What is the language of communication at home?
Q014	你父親的母語為	What is your <i>parental (father's)</i> native language?
Q015	你父親最高的教育水平	What is your <i>parental (father's)</i> highest education level?
Q016	你母親的母語為	What is your <i>maternal (mother's)</i> native language?
Q017	你母親最高的教育水平	What is your <i>maternal (mother's)</i> highest education level?
Q018	你認為實驗的目的是什麼?(如不清楚, 可填寫不知道)	What do you think is the purpose of the experiment? (If you are uncertain, please fill in 'uncertain')
Q019	你在實驗進行期間有使用什麼策略嗎? (如沒有, 可填寫沒有)	What strategy did you use throughout the experiment? (If there