

Information Structure in Cantonese: Scenes and Topic

Wilson Lui

The University of Oxford

Proceedings of the LFG'22 Conference

Miriam Butt, Jamie Y. Findlay and Ida Toivonen

(Editors)

2022

CSLI Publications

<http://csli-publications.stanford.edu/>

Abstract

This study presents an LFG information structure proposal that accounts for different types of Cantonese constructions involving one or multiple ‘topics’, a concept loosely defined in the literature. The proposal resolves the puzzle in understanding and categorising ‘topics’ by utilising the LFG framework which dissociates information structure roles from rigid syntactic positions. It also has implications for other topic-prominent languages which encode information structure through constituents with relatively flexible structural positions and word orders. This study proposes that Cantonese ‘topic’ constructions involve zero, one, or multiple SCENES that provide the setting of a proposition, and one and only one TOPIC that indicates what the sentence is about. SCENES and TOPIC are both generated at the IP level.

1 Introduction

Language involves the exchange of information, and information structure refers to ‘the level of sentence organisation that represent how sentences are structured in a particular context in order to facilitate information exchange’ (Dalrymple et al. 2019:366; see also Dalrymple and Nikolaeva 2011:45). As such, differences in how information is desired to be exchanged and packaged by the interlocutors will affect the structures of the sentences.

Cantonese presents a puzzle in information structure. As a topic-prominent language (Li and Thompson 1976:459–461), Cantonese encodes information structure through constituents with relatively flexible structural positions and word orders. For instance, the sentence ‘topic’¹ can be indicated through the left-dislocation of a word or phrase to the beginning of the sentence (Matthews and Yip 2011:84). Moreover, two or more ‘topics’ can appear consecutively at the beginning of the sentence (Matthews and Yip 2011:86–87; for Mandarin Chinese, see, e.g., Badan and del Gobbo 2010:72–81). There are ‘topics’ at the left periphery which seem to bear no grammatical relation to the predicate (see, e.g., Fung 2007; Matthews and Yip 2011:86). These are also known as ‘dangling topics’ in Mandarin Chinese (see, e.g., Shi 2000; Huang and Ting 2006; Pan and Hu 2008). The generation of these ‘topics’ ‘may be optional from a strictly syntactic point of view, but in fact are motivated by discourse factors such as topic, focus and background information’ (Butt and King 1996:14). When Li and Thompson (1976) distinguished between subject-prominent languages (those which can be ‘more insightfully described’ as taking the notion of subject as basic) and topic-prominent languages

[†]I thank Dr Louise Mycock for supervising an earlier draft of this paper. I also thank the two anonymous reviewers for helpful and constructive comments. The Faculty of Linguistics, Phonology and Phonetics at the University of Oxford and the International Lexical Functional Grammar Association provided financial support for conference attendance.

¹The label of ‘topic’ is used loosely here. In this study, I distinguished two discourse functions SCENE and TOPIC within the notion of ‘topic’ as understood in the previous literature on Cantonese. TOPIC is defined in a narrower sense in this study: see Section 3 below.

(those taking the concept of topic as basic), it was also implicitly considering the differences between languages which determine word order based on particular grammatical functions and information structure categories, respectively (see also Snijders 2015:104–105).

One notable approach under the Chomskyan tradition that is frequently deployed to analyse ‘topic’ constructions is the cartographic approach (see, e.g., Rizzi 1997; Benincà and Poletto 2004; Badan and del Gobbo 2010). Under this approach, information structure roles are assigned to particular phrase structure positions which are projected at certain specified levels, moving and appearing as the specifier of the appropriate Topic Phrase (TopP) (see, e.g., Cheung 2015:119). However, the cartographic approach comes with various deficiencies (Dalrymple and Nikolaeva 2011:59–61). One major criticism is the proliferation of multiple semantically similar projections at different points of the hierarchy. Paul (2015) further showed that the cartographic approach does not fit well with the Mandarin Chinese data, as the flexible permutations of multiple topics and their semantic differences cannot be captured even by using multiple sub-projections.

This study considers the extent to which constructions involving ‘topics’ and multiple ‘topics’ in Cantonese can be accounted for under the Lexical-Functional Grammar (LFG) framework, and in particular, its general information structure architecture. Section 2 provides an overview of the development and current analysis of two notions of information structure—topic and scene—through a survey of the previous literature. It also surveys and evaluates the modelling of information structure under the LFG approach. Section 3 then suggests an analysis of information structure for Cantonese with a new set of phrase structure rules which incorporates the necessary discourse functions (DFs) of the information structure. Section 4 further illustrates the proposal and its operation with examples of ‘topic’ and ‘multiple topic’ constructions in Cantonese. Section 5 concludes the study and suggests further research directions.

2 Information structure and Lexical-Functional Grammar

2.1 What is a topic?

As discussed in Section 1, Cantonese is a topic-prominent language. This means that the ‘grammatical meaning’ of subject and predicate in a sentence is topic and comment (rather than actor and action): the subject is literally the ‘topic’ to talk about, and the predicate is the ‘comment’ that the speaker makes on the ‘topic’ (Chao 1968:40). In other words, topics in Cantonese are not necessarily grammatical subjects, and grammatical subjects are not necessarily topics either.

Having said these, the notions of subject and topic are still closely related. For example, in many languages, the subject is the default topic (see, e.g., Lambrecht 1994:136; Dalrymple et al. 2019:386). Lambrecht (1994) observed that there is a ‘strong correlation between subject and topic’: across languages (including Cantonese and English), ‘the subject of a sentence will be interpreted as its

topic and the predicate as a comment about this topic unless the sentence contains morphosyntactic, prosodic or semantic clues to the contrary'. In his words, the subject is the 'unmarked topic expression' and the topic-comment structure is the 'unmarked presuppositional structure' of a sentence.

From an information structure point of view, topic is usually considered under two views: (1) the aboutness view and (2) the frame view. The aboutness view defines topic as 'the thing which the proposition expressed by the sentence is ABOUT' (Lambrecht 1994:118), 'expressing information which is relevant to and which increases the addressee's knowledge of this referent' (Lambrecht 1994:131). In other words, the speaker announces a topic as the 'centre of attention' or the 'theme of discourse', and then says something about the topic (Li and Thompson 1976:462). In example (1), the sentence is talking about a person, SiuMing, and expresses the information that he knows French. SiuMing is the centre of attention and the theme of the sentence, and a comment is then added to this entity.

- (1) *siu2ming4 sik1 faat3man2.*
 SiuMing know French
 'SiuMing knows French.'

This view remains influential and is accepted in many other works (see, e.g., Li and Thompson 1981:85; Reinhart 1982:80, adopting the 'file card' metaphor; Dalrymple and Nikolaeva 2011:48–49). However, one difficulty of the aboutness view is its inability to categorise topics which are not referential expressions, such as adverbial expressions. For example, temporal expressions at the left periphery do not represent what the sentence is about; they only give the temporal setting of the subsequent proposition. In example (2), the sentence is about SiuMing, and expresses the information that he has seen the boss. The temporal expression *gam1jat6* 'today' at the sentence-initial position gives the time when SiuMing 'sees' the boss, but does not 'announce' what the sentence is about.

- (2) *gam1jat6 siu2ming4 gin3-gwo3 lou5baan2.*
 today SiuMing see-EXP boss
 'Today, SiuMing has seen the boss.'

On the other hand, the frame view proposes that topic is the 'frame' that 'sets a spatial, temporal, or individual framework within which the main predication holds ... limiting the applicability of the main predication to a certain restricted domain' (Chafe 1976:50). It is broader than the aboutness view because this view does not presume an explicit relationship between the topic and the subsequent predicate. The frame view is motivated by the desire to account not only for topics which are arguments, but also those which are only loosely associated with a proposition (Lambrecht 1994:118). More examples will follow.

Let us now consider some Cantonese constructions that have one or multiple 'topics' (as loosely defined in the literature). Example (3) is derived from example

(1) and shows a left-dislocated ‘topic’ where the object *faat3man2* ‘French’ is displaced to the beginning of the sentence. This example can be considered to convey the information of either ‘talking about French, SiuMing knows it’ (the aboutness view), or ‘within the “frame” (or context) of French, SiuMing is someone who knows it’ (the frame view). Here, these two views do not seem to differ much.

- (3) *faat3man2 siu2ming4 sik1.*
 French SiuMing know
 ‘French, SiuMing knows.’

Example (4) is an example where the two views differ. Here, there is a ‘topic’ *ji4gaal ge3 tin1hei3* ‘the weather now’, which, similar to the temporal expression in example (2) above, seems to have no grammatical relation with the predicate. It does not indicate what the sentence is about: in Cantonese, example (4) is about catching a cold. This interpretation fits well with the analysis of topics as frames, indicating the circumstantial context in which it is very easy to catch a cold.

- (4) (Matthews and Yip 2011:86)
ji4gaal ge3 tin1hei3 zeoi3 ji6 soeng1fung1.
 now POSS weather most easy cold
 ‘It is very easy to catch a cold in this weather.’

Example (5) seems to consist of two ‘topics’, appearing consecutively at the left periphery of the sentence.² In these sentences, it is difficult to identify precisely what the sentence is ‘about’, and the aboutness view will therefore face considerable obstacles. However, if we take the frame view, then the first of these ‘topics’ can be taken as the frame of the sentence, providing the framework within which the subsequent, main predication holds. The second ‘topic’, *ngo5* ‘I’, is the subject of the predicate. Thus, example (5) can be understood under the frame view to mean ‘With regard to the Chiuchow dialect, I do not know a single word of it.’

- (5) (Matthews and Yip 2011: 87)
ciu4zau1-waa6 ngo5 jat1 geoi3 dou1 m4-sik1 gong2.
 Chiuchow-language 1SG one phrase also NEG-know speak
 ‘The Chiuchow dialect, I do not know a single word of it.’

Mandarin Chinese is also considered to be productive in terms of these ‘multiple topic’ constructions (see, e.g., Badan and del Gobbo 2010:81). Example (6) shows a sentence with two ‘topics’, with the first one showing no particular grammatical relation to the main predicate. In this example, the aboutness view will face a similar difficulty in example (5) above. However, the frame view will be able to take the first ‘topic’ *shengwulunlixue* ‘bioethics’ as the frame which anchors the subsequent predicate (‘I’ being an outsider) to a specific point of reference, that is, the academic area of bioethics. Similarly, the second ‘topic’ *wo* ‘I’ is the subject of the predicate.

²Matthews and Yip (2011:87) described these constructions as having two subjects.

- (6) Mandarin Chinese (Huang 1994:163)
shengwulunlixue wo shi menwaihan.
 bioethics 1SG be outsider
 ‘(With regard to) bioethics, I am an outsider/layman.’

The data above shows that there seems to be two kinds of topics in Cantonese (and Mandarin Chinese). The first kind of topic has no apparent grammatical relation with the predicate, and they are much better accounted by the frame view which provides the context for the subsequent predicate. The other kind of topic relates to what the sentence is about; for example, it gives the subject of the predicate directly. This study argues that these two kinds of topics should be treated differently. It proposes that in LFG, the former kind of topic should be defined as SCENES, while the latter kind of topic should be defined as TOPICS.

2.2 What is a scene?

Unlike ‘topic’, ‘scene’ is not a commonly recognised notion in information structure. Despite the unfamiliar label, we saw in examples (4), (5) and (6) above that there are ‘topics’ as ‘frames’ at the left periphery which seem to bear no grammatical relation to the predicate. This study argues that, instead of categorising them as ‘topics’ or frames, the discourse function of SCENE in LFG will be able to account for these constructions.

The proposal to use SCENE as a discourse function in LFG was first seen in Andréasson (2007) when analysing Swedish examples. That work defined SCENE as ‘constituents that relate the proposition to a temporal, spatial, or circumstantial context’ (Andréasson 2007:34). In doing so, it referred to the definition of topics as ‘frames’ in Chafe (1976:50), cited in Section 2.1 above. It is not immediately clear what is the difference between the two ideas. I venture to suggest that they are essentially the same in substance; the difference (if there is one) is merely that of context between information structure in general and in terms of information structure and packaging in LFG specifically.

I define SCENES as constituents that provides the setting of a proposition, which is slightly broader than Andréasson’s (2007) definition as it does not limit the context that SCENE has to relate to. It can be a PP or NP relating to the spatial or temporal elements of the setting, or the event or circumstances in which the proposition is established. At first sight, this definition for SCENE seems to be similar to that for TOPIC. In the following paragraphs, I identify two key differences between SCENES and TOPICS.

First, SCENE is optional, and when it is present, it may take one or more than one value. However, TOPIC is mandatory, and there must be one and only one value of TOPIC. The assumption for this proposition is that there is one and only one TOPIC in any given sentence in Cantonese. If one takes Reinhart’s (1982) understanding of ‘topic’ with the metaphor of file cards, then one sentence can only have one ‘centre’, ‘theme’ or ‘subject matter’: it is the file card to be extracted

and filled out. As for the so-called thetic (all-focus) sentences, which apparently lack any topic, two arguments can be made. First, as Cantonese is a ‘topic-drop’ or ‘discourse pro-drop’ language, topics may be omitted based on pragmatic presupposition (Nikolaeva 2001:5). Second, implicit stage topics can be assumed which provides ‘the spatio-temporal parameters of the sentence (the here-and-now of the discourse)’ to assess the truth value of the proposition (for more details, see Erteschik-Shir 2019:224). Stage topics are different from SCENES because SCENES (if they exist) must be explicitly stated in the sentence.

Second, based on the definition above, SCENES are not necessarily related to the clause; they can merely ‘provide the setting’. However, the TOPIC must relate to the clause, either by way of being a dislocated topic through the DIS function, or by being the grammatical subject of the sentence, under the strong correlation referred to at the beginning of Section 2.1 above.

After looking at how TOPIC and SCENE are to be defined, we now turn to the approach in LFG in which these notions are used to account for these Cantonese ‘topic’ and ‘multiple topic’ constructions through a new perspective.

2.3 Lexical-Functional Grammar

Little has to be said about the fundamental assumptions and breakthroughs of the Lexical-Functional Grammar (LFG) framework (see, e.g., Bresnan 1978; Kaplan and Bresnan 1982; Bresnan 2001; Bresnan et al. 2016; Dalrymple et al. 2019, Dalrymple 2023) as a generative, non-transformational and constraint-based theory, which readers will be familiar with.

In relation to information structure in LFG, Zaenen (2023) provides an excellent overview on the *status quo*. This study continues to speak and contribute to the literature in this area, which is yet to be fully understood and appreciated.

The traditional (now rendered obsolete) approach towards topic and focus in LFG is to treat them as ‘grammaticalised discourse functions’ – namely, TOP and FOC – which are discourse functions that must be linked to an argument function through the Extended Coherence Condition; in other words, they must be ‘linked to the semantic predicate argument structure of the sentence in which they occur, either by functionally or by anaphorically binding an argument’ (Bresnan and Mchombo 1987:746). As they have syntactic roles to perform, they can therefore be represented and directly encoded in the f-structure (Bresnan and Mchombo 1987; Bresnan 2001; Bresnan et al. 2016:196–199).

This approach used the labels of ‘topic’ and ‘focus’ in a ‘deliberately ambiguous way’ to ‘capture a syntactic relation and an information structure relation simultaneously’ (Dalrymple et al. 2019:376). It was quickly realised that there were often mismatches between grammatical functions and discourse functions in many languages. For example, the topic or focus may only match a part of an f-structure constituent, and so FOC will not be able to link with any f-structure that consists only of the focused material (King 1997:8–9). Moreover, in addition to syntax, other aspects of the linguistic structure, such as morphology (Dalrymple

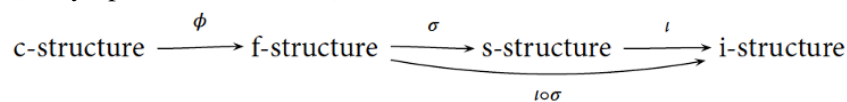
and Nikolaeva 2011) and prosody (Mycock 2013) can encode certain information structure DFs such as ‘focus’. As such, the grammaticalised discourse functions FOC and TOP would not be able to fully model the pragmatics or information structure of discourse functions. A purely syntactic representation of information structure DFs will be insufficient, and the failure to distinguish between the syntactic relations and the information structure relations is therefore undesirable.

To resolve the problems above, King (1997) first made an explicit formal proposal in LFG for a separate, independent level of information structure (i-structure). An i-structure is represented, like an f-structure, by an attribute value matrix. TOPIC and FOCUS, for instance, are primitive features of i-structure and assumed since its inception (King 1997:9). This approach has been followed in most subsequent works of LFG (see, e.g., Mycock 2006; Dalrymple and Nikolaeva 2011; Dalrymple et al. 2019:366–394).

Earlier proposals modelled the i-structure as a projection from the c-structure through the $\iota(x)$ function, and there will be a direct connection between the two (King 1997). The i-structure is independent of the f-structure, which is another projection from the c-structure through the $\phi(x)$ function. In this study, I adopt the more recent proposal in Dalrymple and Nikolaeva (2011), refined in Dalrymple et al. (2019:381–393), which suggests that i-structure is projected directly from s-structure through the $\iota(x)$ function; in other words, the information structure also contains semantic information. This can resolve what is known as ‘the granularity problem’ by ensuring that only the meaning constructors contributed by the head(s) of the f-structure are associated with the respective i-structure roles (for more details, see Dalrymple et al. 2019:379–381, 393).

Figure (7) below shows the LFG architecture assumed in this study, with the levels of representation and how they are associated by the mapping functions.

(7) (Dalrymple et al. 2019: 384)



In sum, LFG assigns an autonomous role to information structure. This solves the difficulties outlined above by allowing the dissociation of discourse functions from rigid syntactic positions. The features and phrase structure rules that will be adopted for the information structure in LFG for Cantonese are detailed in Section 3 below.

3 The proposed architecture of information structure in LFG for Cantonese

Section 2 above presents an overview of the existing literature on information structure, on LFG, and on the Cantonese data. To summarise, the aim of this study is to account for the (potentially multiple) ‘scene(s)’ as well as the ‘topic’ at the left

periphery. The search is for a consistent and simple solution, and the LFG framework provides such a solution. This section details the new proposal for information structure in LFG for Cantonese. Section 3.1 states the features of information structure in Cantonese and their definitions. Section 3.2 provides the phrase structure rules to generate an appropriate analysis.

3.1 Features

Incorporating the discussion in Sections 2.1 and 2.2 above, this proposal identifies a total of four discourse functions (DFs) for the i-structure for Cantonese: SCENE, TOPIC, FOCUS and BACKGROUND. It is expected that this inventory of DFs will also be applicable to some other languages, although further investigations of its applicability will have to be left for a future study.

1. SCENE is a constituent that provides the setting of a proposition (see Section 2.2 above; modified from Andréasson 2007:34). It must be an adjunct, and is not necessarily related to a position inside the clause. This also includes the category of ‘dangling topics’ as described in the literature (eg Xu and Langendoen 1985; Shi 2000).
2. TOPIC is the constituent that indicates what the sentence is about (see Section 2.1 above; Lambrecht 1994:118). There can be one and only one TOPIC. It may be a displaced phrase from the predicate, or the subject of the predicate.
3. FOCUS is ‘the information in a statement that is intended to increase the listener’s knowledge’ (Andréasson 2007:29).³
4. BACKGROUND is ‘informationally old knowledge that may be necessary for syntactic reasons or to make it clear how new information fits with what is already known’ (Dalrymple and Nikolaeva 2011:65, citing Butt and King 1996).

3.2 Phrase structure rules

The annotated phrase structure rules for Cantonese required to license the information structure DFs of SCENE and TOPIC are in (8) and (9). How they can capture all the features described in Section 3.1 above is explained in full below. The other phrase structure rules in Cantonese are summarised at (10).⁴

Most of these phrase structure rules are similar to those in English. In particular, I assume IP adjunctions for SCENE and TOPIC as ‘topicalised phrases’ are also adjoined as specifiers to IP in English (see, e.g., Bresnan et al. 2016:16–17).

³Lambrecht (1994:207) defined FOCUS as the ‘portion of a proposition which cannot be taken for granted at the time of speech’. As this study is primarily interested in how (multiple) ‘topics’ can be accounted, the different formulations of FOCUS are not discussed in detail.

⁴These phrase structure rules (particularly the rules for the VP) are simplified to some extent (see, e.g., Lam 2008), but this will not affect the analysis of information structure, which (in this study) concerns exclusively the left periphery.

- (8) $IP_1 \longrightarrow \{NP|PP\}^* \left(\{NP|PP\} \right) \{NP|PP\}^* IP_2$
 $\downarrow \in (\uparrow ADJ) \quad \downarrow \in (\uparrow DIS) \quad \downarrow \in (\uparrow ADJ) \quad \uparrow = \downarrow$
 $\uparrow_{\sigma_i} = \downarrow_{\sigma_i} \quad (\uparrow COMP^* GF) = \downarrow \quad \uparrow_{\sigma_i} = \downarrow_{\sigma_i}$
 $(\downarrow_{\sigma} DF) = SCENE \quad \uparrow_{\sigma_i} = \downarrow_{\sigma_i} \quad (\downarrow_{\sigma} DF) = SCENE$
 $(\downarrow_{\sigma} DF) = TOPIC$
- (9) $IP_2 \longrightarrow \left(NP \right) I'$
 $(\uparrow SUBJ) = \downarrow \quad \uparrow = \downarrow$
 $\uparrow_{\sigma_i} = \downarrow_{\sigma_i}$
 $((\downarrow_{\sigma} DF) = TOPIC)$
- (10) $I' \longrightarrow \left(I \right) VP$
 $\uparrow = \downarrow \quad \uparrow = \downarrow$
 $VP \longrightarrow V'$
 $\uparrow = \downarrow$
 $V' \longrightarrow \left(AdvP \right) V'$
 $\downarrow \in (\uparrow ADJ) \quad \uparrow = \downarrow$
 $V' \longrightarrow V \left(NP \right) \left(NP \right)$
 $\uparrow = \downarrow \quad (\uparrow OBJ_{\theta}) = \downarrow \quad (\uparrow OBJ) = \downarrow$
 $NP \longrightarrow N'$
 $\uparrow = \downarrow$
 $N' \longrightarrow \left(AdjP \right) N'$
 $\downarrow \in (\uparrow ADJ) \quad \uparrow = \downarrow$
 $N' \longrightarrow N$
 $\uparrow = \downarrow$
 $PP \longrightarrow P'$
 $\uparrow = \downarrow$
 $P' \longrightarrow P \quad NP$
 $\uparrow = \downarrow \quad (\uparrow OBJ) = \downarrow$
 $AdvP \longrightarrow Adv'$
 $\uparrow = \downarrow$
 $Adv' \longrightarrow Adv$
 $\uparrow = \downarrow$

Phrase structure rule (8) generates all SCENES and the displaced TOPIC (if any). SCENES must necessarily be adjuncts, and they can appear before or after the TOPIC. The Kleene stars are to indicate zero, one, or multiple SCENES at either position. The TOPIC in IP_1 is optional, and if present, must be a member of the set value of the DIS(placed) attribute of IP_1 (see also Dalrymple et al. 2019:659). The $(\uparrow COMP^* GF) = \downarrow$ equation, where $GF \equiv \{SUBJ|OBJ|OBJ_{\theta}|COMP|XCOMP|OBL_{\theta}|ADJ \in |XADJ \in\}$, means that the topic can be displaced from an embedded clause (Dalrymple et al. 2019:205–208). The set membership symbol is used to allow reference to some member of a set. The $\uparrow_{\sigma_i} = \downarrow_{\sigma_i}$ equations are to ensure that the i -structures for the mother node and the daughter node are the same.

Phrase structure rule (9) generates the SUBJ, and the optional $((\downarrow_{\sigma} DF) = TOPIC)$ equation means that the SUBJ will also receive the DF as TOPIC if there is no displaced TOPIC (Dalrymple and Nikolaeva 2011:81), bearing in mind the strong as-

sociation between subject and topic (Lambrecht 1994:136). This NP is also optional, to account for pro-drop in Cantonese (Luke et al. 2001:2). As discussed in Section 2.2 above, these sentences have an implicit TOPIC based on pragmatic presupposition (Nikolaeva 2001:5). Again, there are also the $\uparrow_{\sigma_i}=\downarrow_{\sigma_i}$ equations.

As for the DFS of FOCUS and BACKGROUND, they are not generated through any phrase structure rules but are inferred whenever appropriate from the context. How these DFS are inferred is not the focus of this essay (for more details, see, e.g., Dalrymple and Nikolaeva 2011). In Section 4 below, I assume all verbs in the examples have a DF of BACKGROUND, although they do not necessarily have to. It is merely to demonstrate the full inventory of features available for the i-structure.

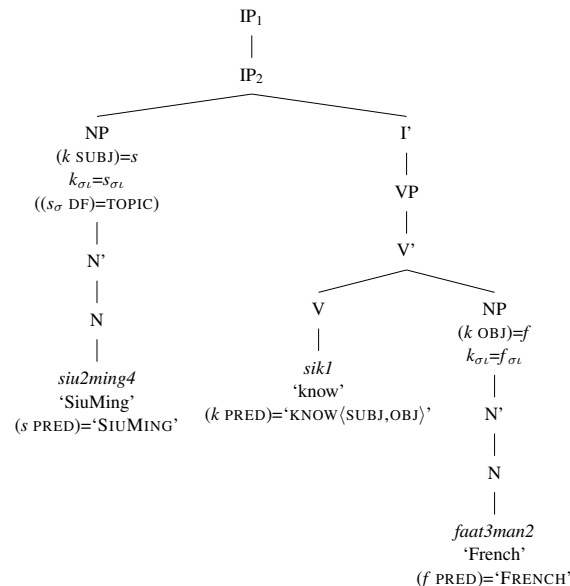
4 Applying the LFG analysis to Cantonese

In this section, the different combinations of TOPIC and SCENES are presented with their respective LFG analyses. As proposed in Section 3 above, in Cantonese, there can be zero, one or multiple SCENES, zero or one displaced TOPIC, and the grammatical subject as the optional TOPIC. There must be one and only one TOPIC.

4.1 One TOPIC, zero SCENE

Let us first consider a simple Cantonese sentence with a canonical SVO order. We use example (1) again. The c-structure, f-structure, s-structure and i-structure of example (1) are shown at (11).

(11) a. **c-structure:**⁵



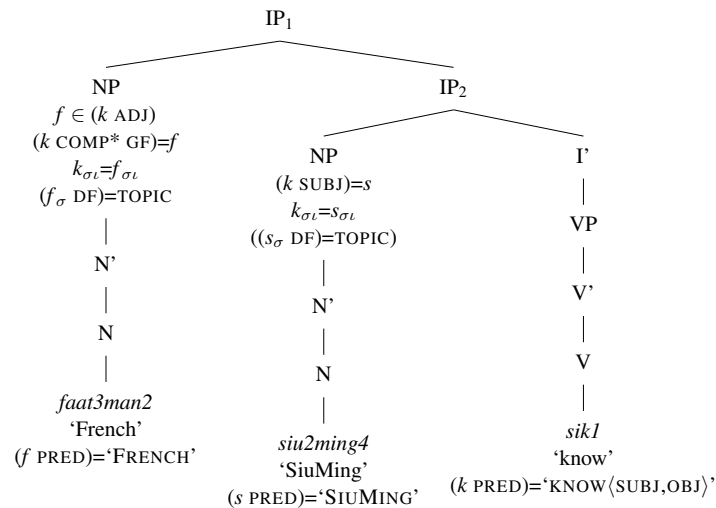
⁵IP₁ will have been excluded by Economy. It is only included here to demonstrate the effects of the phrase structure rules proposed in Section 3.2 above.

- b. **f-structure:**
$$\left[\begin{array}{l} \text{PRED 'KNOW(SUBJ,OBJ)'} \\ \text{SUBJ } \left[\begin{array}{l} \text{PRED 'SIUMING'} \\ \text{OBJ } \left[\begin{array}{l} \text{PRED 'FRENCH'} \end{array} \right] \end{array} \right] \end{array} \right]$$
- c. **s-structure:**
 SiuMing : s_σ [DF TOPIC]
 French : f_σ [DF FOCUS]
 $\lambda y. \lambda x. \text{know}(x,y) : s_\sigma \multimap (f_\sigma \multimap k_\sigma)$ [DF BACKGROUND]
- d. **i-structure:**
$$\left[\begin{array}{l} \text{TOPIC} \quad \text{SiuMing : } s_\sigma \\ \text{FOCUS} \quad \{ \text{French : } f_\sigma \} \\ \text{BACKGROUND} \quad \{ \lambda y. \lambda x. \text{know}(x,y) : s_\sigma \multimap (f_\sigma \multimap k_\sigma) \} \end{array} \right]$$

Example (1) is a standard Cantonese sentence following the canonical SVO order. Following the default relationship between subject and topic, in example (1), *siu2ming4* ‘SiuMing’ is both the subject and the topic. In (11), the c-structure NP node *siu2ming4* is mapped to SUBJ at the f-structure through the ϕ function. As there is no other topic, the optional (\downarrow_σ DF)=TOPIC equation in phrase structure rule (9) applies, and gives *siu2ming4* a DF value of TOPIC in the s-structure when mapped from the f-structure through the σ function. This is then further mapped on the i-structure as a TOPIC through the ι function. This captures formally how *siu2ming4* is the TOPIC of the sentence at the i-structure.

Now consider example (3), where the object *faat3man2* ‘French’ is displaced and is now at the left periphery (the beginning) of the sentence. The c-structure, f-structure, s-structure and i-structure of example (3) are shown at (12).

- (12) a. **c-structure:**



- b. **f-structure:**
$$\left[\begin{array}{l} \text{PRED 'KNOW(SUBJ,OBJ)'} \\ \text{DIS } \left\{ \left[\text{PRED 'FRENCH'} \right] \right\} \\ \text{SUBJ } \left[\text{PRED 'SIUMING'} \right] \\ \text{OBJ } \left[\text{PRED 'FRENCH'} \right] \end{array} \right]$$

c. **s-structure:**

French : f_σ [DF TOPIC]

SiuMing : s_σ [DF FOCUS]

$\lambda y. \lambda x. \text{know}(x,y) : s_\sigma \multimap (f_\sigma \multimap k_\sigma)$ [DF BACKGROUND]

d. **i-structure:** $\left[\begin{array}{l} \text{TOPIC} \quad \text{French} : f_\sigma \\ \text{FOCUS} \quad \quad \{ \text{SiuMing} : s_\sigma \} \\ \text{BACKGROUND} \quad \left\{ \lambda y. \lambda x. \text{know}(x,y) : s_\sigma \multimap (f_\sigma \multimap k_\sigma) \right\} \end{array} \right]$

In example (3), the constituent *faat3man2* ‘French’ is displaced. It meets the proposed definition of TOPIC under phrase structure rule (8). Moreover, it is clear that example (3) can be an answer to the following question (3A), which proves the TOPIC status of *faat3man2* as what the sentence is about:

(3A) *faat3man2 bin1go3 sik1?*

French who know

‘Who knows French?’ (lit. ‘French, who knows?’)

This shows that the context also plays an important role in retrieving the information structure DF of a constituent (Dalrymple and Nikolaeva 2011). If *siu2ming4* ‘SiuMing’ is the answer to the question (3A), then it achieves FOCUS status. The fronted phrase *faat3man2* ‘French’ in example (3) has the DF of TOPIC and shows what the sentence is about. The FOCUS and BACKGROUND of the sentence provides further information about the TOPIC. This is reflected in (12). First, as the c-structure NP node *faat3man2* is displaced, it is now licensed by IP_1 and receives the TOPIC by applying the $(\downarrow_\sigma \text{ DF}) = \text{TOPIC}$ equation in phrase structure rule (8). Through the same projections from c-structure to f-structure (ϕ function), from the f-structure to s-structure (σ function), and from the s-structure to i-structure (ι function), *faat3man2* ‘French’ is formally captured as the TOPIC of the sentence. As there is already one TOPIC, the optional $((\downarrow_\sigma \text{ DF}) = \text{TOPIC})$ equation in phrase structure rule (9) does not operate, and the subject NP *siu2ming4* will not receive a DF of TOPIC in the s-structure. Recall that there can never be more than one TOPIC in a Cantonese sentence. The optionality of the equation enforces that assumption.

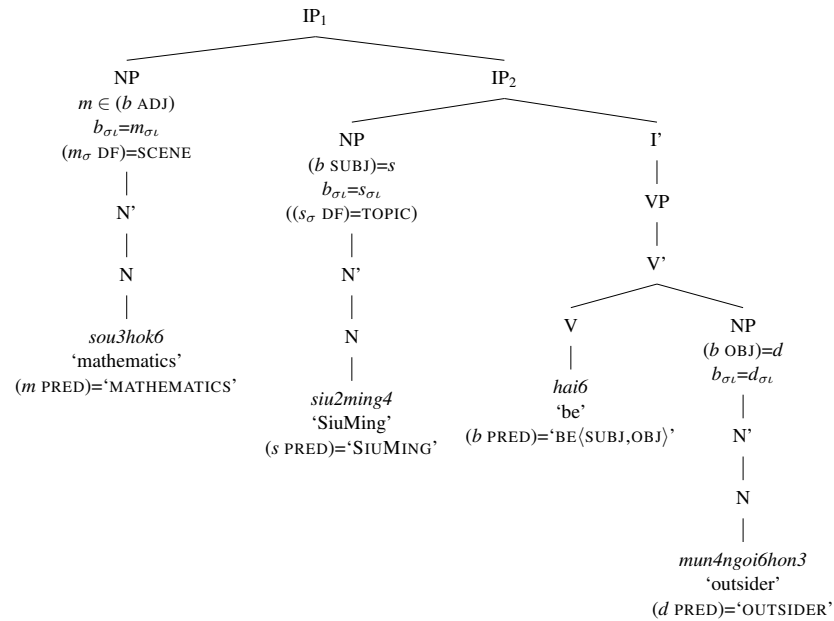
4.2 One TOPIC, one SCENE

This situation is where there is a ‘topic’ that is only loosely related to the predicate, also known as ‘Chinese-style’ topics in the literature (see, e.g., Chafe 1976; Shi 2000). Examples include example (5) above and a new example (13) below, which is similar to the Mandarin Chinese example (6). The c-structure, f-structure, s-structure and i-structure of (13) are in (14).⁶

⁶For simplicity, I treat *hai6* as a verb here, but note the discussion of PREDLINK for copula constructions (see, e.g., Dalrymple et al. 2019:32–33; Laczko 2021:196–198).

- (13) *sou3hok6 siu2ming4 hai6 mun4ngoi6hon3.*
 mathematics SiuMing be outsider
 ‘(With regard to) mathematics, SiuMing is an outsider/layman.’

(14) a. **c-structure:**



b. **f-structure:**

$$\left[\begin{array}{l} \text{PRED 'BE(SUBJ,OBJ)'} \\ \text{SUBJ } \left[\begin{array}{l} \text{PRED 'SIUMING'} \end{array} \right] \\ \text{OBJ } \left[\begin{array}{l} \text{PRED 'OUTSIDER'} \end{array} \right] \\ \text{ADJ } \left\{ \left[\begin{array}{l} \text{PRED 'MATHEMATICS'} \end{array} \right] \right\} \end{array} \right]$$

c. **s-structure:**

mathematics : m_σ [DF SCENE]
 SiuMing : s_σ [DF TOPIC]
 outsider : d_σ [DF FOCUS]
 $\lambda y. \lambda x. \text{be}(x,y) : s_\sigma \multimap (d_\sigma \multimap b_\sigma)$ [DF BACKGROUND]

d. **i-structure:**

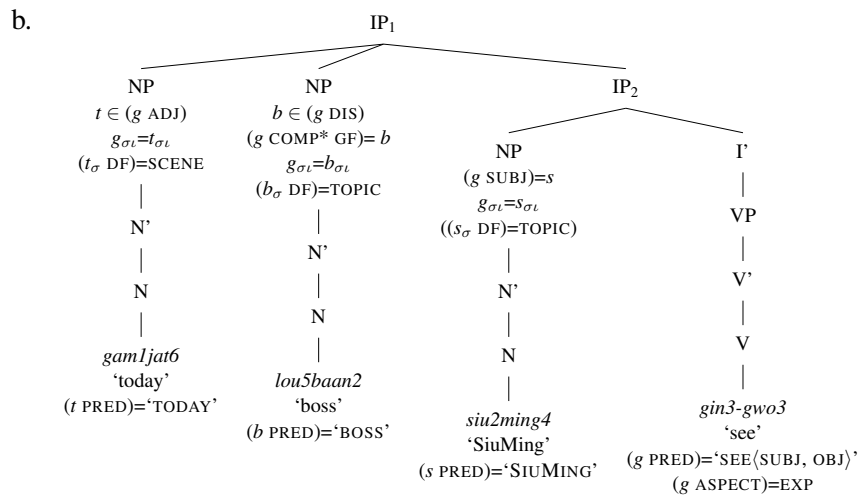
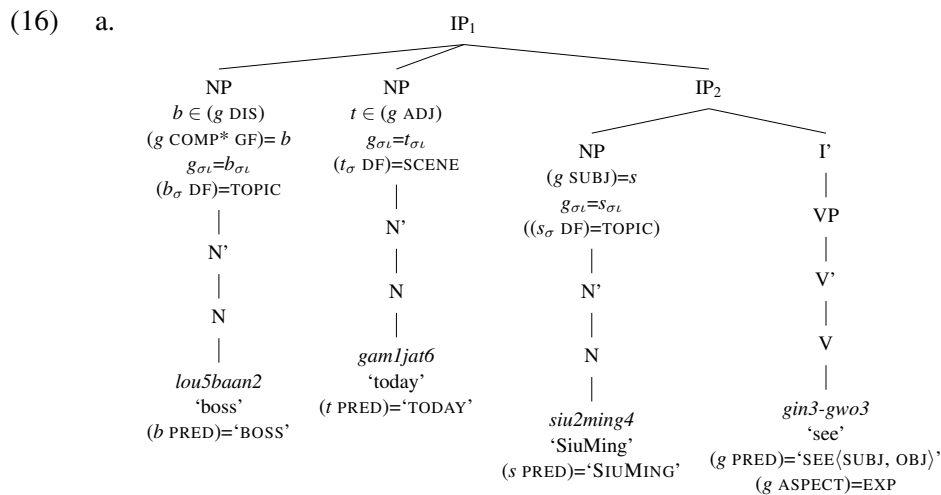
$$\left[\begin{array}{l} \text{SCENE } \left\{ \text{mathematics : } m_\sigma \right\} \\ \text{TOPIC } \text{SiuMing : } s_\sigma \\ \text{FOCUS } \left\{ \text{outsider : } d_\sigma \right\} \\ \text{BACKGROUND } \left\{ \lambda y. \lambda x. \text{be}(x,y) : s_\sigma \multimap (d_\sigma \multimap b_\sigma) \right\} \end{array} \right]$$

In (13), *sou3hok6* ‘mathematics’ carries the information structure role of SCENE. The reasoning process consists of several steps. First, *sou3hok6* ‘mathematics’ is not required by the predicate, and is therefore an adjunct. As it is not a displaced constituent of the predicate, it therefore cannot be a TOPIC that is licensed by IP₁ under phrase structure rule (8). As there is no displaced TOPIC, the TOPIC in this sentence is licensed by IP₂ under phrase structure rule (9) through the default option, that is, *siu2ming4* ‘SiuMing’ as the grammatical subject of the predicate. On

the premise that there can be one and only one TOPIC, the constituent *sou3hok6* is therefore a SCENE, which, in this sentence, appears before the TOPIC.

I now turn to the different permutations of constituents between the following examples (15a) and (15b). Paul (2015) suggested that there seems to be slight semantic differences between examples like these in Mandarin Chinese. Applying his observations, example (15a) expresses the meaning of ‘Regarding the boss, today is the time when SiuMing has seen him.’, while example (15b) expresses the meaning of ‘During today, the boss is the person who SiuMing has seen.’. The c-structure of examples (15a) and (15b) are in (16a) and (16b) respectively.

- (15) a. *lou5baan2 gam1jat6 siu2ming4 gin3-gwo3*
 boss today SiuMing see-EXP
 ‘The boss, today, SiuMing has seen.’
- b. *gam1jat6 lou5baan2 siu2ming4 gin3-gwo3*
 today boss SiuMing see-EXP
 ‘Today, the boss, SiuMing has seen.’



Although there is a difference in the position of the constituents, there is no difference in terms of the analysis in respect of information structure. In both (15a) and (15b), *lou5baan2* ‘boss’ is the object of the predicate ‘see’ and is therefore a displaced constituent. Under phrase structure rule (8) for IP₁, it has the DF of TOPIC. The other fronted element *gam1jat6* ‘today’ is a temporal expression which is an adjunct. Removing it does not render the sentence ungrammatical. Based on a deduction exercise like example (13) above, we can conclude that *gam1jat6* is a SCENE. It can appear before or after the TOPIC *lou5baan2* ‘boss’, and this is captured by the two positions of SCENES in phrase structure rule (8). As such, regardless of the position of *lou5baan2* with respect to *gam1jat6* (before or after), the TOPIC is *lou5baan2* ‘boss’ and the SCENE is *gam1jat6* ‘today’. Both examples (15a) and (15b) therefore have the same f-structures, s-structures and i-structures, which are shown in (17).

(17)	a. f-structure:	<table style="border-collapse: collapse; width: 100%;"> <tr> <td style="padding: 2px;">PRED</td> <td style="padding: 2px;">‘SEE(SUBJ,OBJ)’</td> </tr> <tr> <td style="padding: 2px;">ASPECT</td> <td style="padding: 2px;">EXP</td> </tr> <tr> <td style="padding: 2px;">DIS</td> <td style="padding: 2px;">{ [PRED ‘BOSS’] }</td> </tr> <tr> <td style="padding: 2px;">SUBJ</td> <td style="padding: 2px;">[PRED ‘SIUMING’]</td> </tr> <tr> <td style="padding: 2px;">OBJ</td> <td style="padding: 2px;">[PRED ‘SIUMING’]</td> </tr> <tr> <td style="padding: 2px;">ADJ</td> <td style="padding: 2px;">{ [PRED ‘TODAY’] }</td> </tr> </table>	PRED	‘SEE(SUBJ,OBJ)’	ASPECT	EXP	DIS	{ [PRED ‘BOSS’] }	SUBJ	[PRED ‘SIUMING’]	OBJ	[PRED ‘SIUMING’]	ADJ	{ [PRED ‘TODAY’] }
PRED	‘SEE(SUBJ,OBJ)’													
ASPECT	EXP													
DIS	{ [PRED ‘BOSS’] }													
SUBJ	[PRED ‘SIUMING’]													
OBJ	[PRED ‘SIUMING’]													
ADJ	{ [PRED ‘TODAY’] }													
	b. s-structure:													
		today : t_σ [DF SCENE]												
		boss : b_σ [DF TOPIC]												
		SiuMing : s_σ [DF FOCUS]												
		$\lambda y.\lambda x.\text{see}(x,y) : s_\sigma \multimap (b_\sigma \multimap g_\sigma)$ [DF BACKGROUND]												
	c. i-structure:	<table style="border-collapse: collapse; width: 100%;"> <tr> <td style="padding: 2px;">SCENE</td> <td style="padding: 2px;">{ today : t_σ }</td> </tr> <tr> <td style="padding: 2px;">TOPIC</td> <td style="padding: 2px;">boss : b_σ</td> </tr> <tr> <td style="padding: 2px;">FOCUS</td> <td style="padding: 2px;">{ SiuMing : s_σ }</td> </tr> <tr> <td style="padding: 2px;">BACKGROUND</td> <td style="padding: 2px;">{ $\lambda y.\lambda x.\text{see}(x,y) : s_\sigma \multimap (b_\sigma \multimap g_\sigma)$ }</td> </tr> </table>	SCENE	{ today : t_σ }	TOPIC	boss : b_σ	FOCUS	{ SiuMing : s_σ }	BACKGROUND	{ $\lambda y.\lambda x.\text{see}(x,y) : s_\sigma \multimap (b_\sigma \multimap g_\sigma)$ }				
SCENE	{ today : t_σ }													
TOPIC	boss : b_σ													
FOCUS	{ SiuMing : s_σ }													
BACKGROUND	{ $\lambda y.\lambda x.\text{see}(x,y) : s_\sigma \multimap (b_\sigma \multimap g_\sigma)$ }													

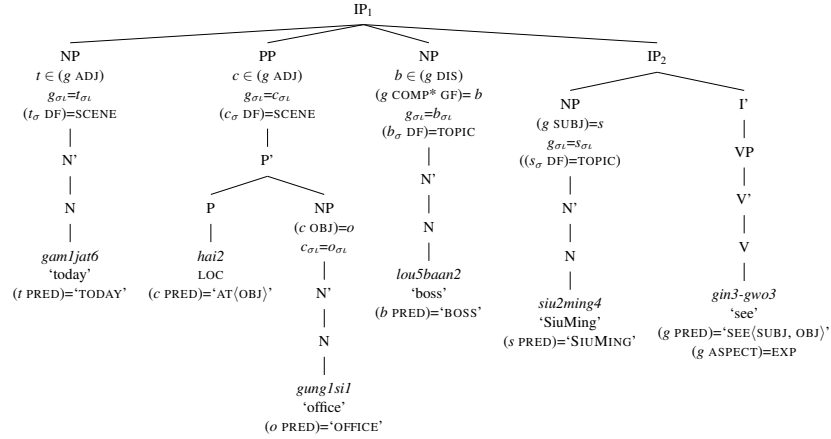
Therefore, the semantic differences alluded to in Paul (2015) is not resolvable based on the information structure proposed in this study. It may be due to other factors, such as prosody. This can be a possible future research direction.

4.3 One TOPIC, two or more SCENES

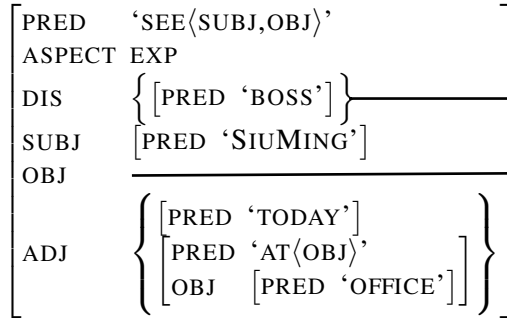
Two SCENES are possible in Cantonese, and this is usually done by combining a spatial SCENE and a temporal SCENE as in example (18). The c-structure, f-structure, s-structure and i-structure of example (18) are in (19).

- (18) *gam1jat6 hai2 gung1si1 lou5baan2 siu2ming4 gin3-gwo3*
today LOC office boss SiuMing see-EXP
‘Today, at office, the boss, SiuMing has seen.’

(19) a. **c-structure:**



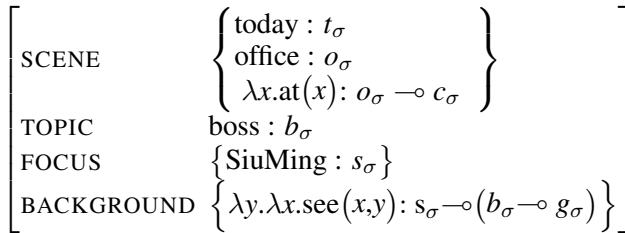
b. **f-structure:**



c. **s-structure:**

office : o_σ [DF SCENE]
 $\lambda x.at(x) : o_\sigma \multimap c_\sigma$ [DF SCENE]
today : t_σ [DF SCENE]
boss : b_σ [DF TOPIC]
SiuMing : s_σ [DF FOCUS]
 $\lambda y.\lambda x.see(x,y) : s_\sigma \multimap (b_\sigma \multimap g_\sigma$ [DF BACKGROUND])

d. **i-structure:**

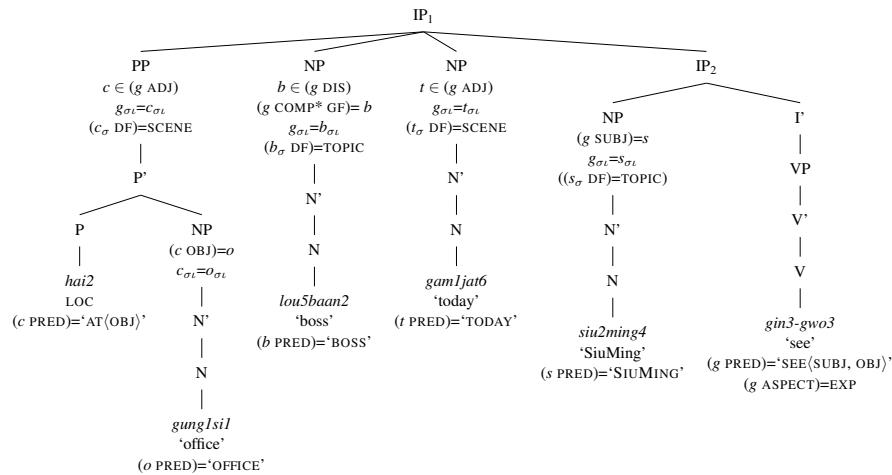


Following a similar reasoning process like (14), *lou5baan2* is the object of the predicate 'see' and is therefore the displaced TOPIC in IP₁ under phrase structure rule (8). The other two fronted elements are adjuncts and are therefore SCENES. Here, the two SCENES are both specifiers to the IP₁ as enforced in phrase structure rule (8). In the i-structure, they are interchangeable elements within the set of values of SCENE. This also correctly predicts that there is no difference in meaning between examples (18) and (18A). Example (18A) shows two SCENES, one before

and one after the TOPIC. The c-structure of (18A) is shown in (20). The f-structure, s-structure and i-structure of (18A) are the same as (18) and are already shown in (19) above.

(18A) *hai2 gung1si1 lou5baan2 gam1jat6 siu2ming4 gin3-gwo3*
 LOC office boss today SiuMing see-EXP
 ‘At office, the boss, today, SiuMing has seen.’

(20) **c-structure:**



Theoretically speaking, there can be an infinite number of SCENES. However, one must bear in mind what Xu and Langendoen (1985:17) insightfully observed: ‘Needless to say, there are practical limits to the number of topics and perhaps also to their relative lengths. It is difficult to accept a sentence with three topics, and harder still to accept one with more than three. . . . Any limitation on the number of topics in a topic structure may therefore be considered a matter of performance.’. This quotation used the term ‘topics’; in the context of this study, it means the total number of all SCENES plus the TOPIC.

5 Conclusion

This study has presented an LFG information structure proposal that accounts for different types of ‘topic’ constructions in Cantonese. I argue that these constructions involve zero, one, or multiple SCENES that provide the setting of a proposition, and one and only one TOPIC that indicates what the sentence is about. SCENES and TOPIC are both generated at the IP level, through the two proposed phrase structure rules (8) and (9).

The proposal of this study takes an important step forward for the literature on Cantonese grammar by resolving the previous puzzle in understanding and categorising the notion of ‘topic’ in Cantonese. It also furthers our understanding

on how the LFG framework can be utilised to account for similar ‘topic’ constructions in other topic-prominent languages, by dissociating information structure roles from rigid syntactic positions.

Further research can look at the cross-linguistic application of this proposal. In particular, it will be interesting to consider and test the proposal against other topic-prominent languages. Mandarin Chinese is a clear candidate. Other languages may include Japanese and Korean.⁷ Another research direction is to refine the proposal by considering how FOCUS and BACKGROUND should be contextually inferred, and (when they are inferred) their interactions with SCENES and TOPIC.⁸

References

- Andréasson, Maia. 2007. The architecture of i-structure. In *Proceedings of the LFG '07 Conference*, pages 26–43, Stanford: CSLI Publications.
- Badan, Linda and del Gobbo, Francesca. 2010. On the syntax of topic and focus in Chinese. In Paola Benincà and Nicola Munaro (eds.), *Mapping the Left Periphery: The Cartography of Syntactic Structures*, volume 5, pages 63–90, Oxford: Oxford University Press.
- Benincà, Paola and Poletto, Cecilia. 2004. Topic, focus and V2: Defining the CP sublayers. In Paola Benincà and Nicola Munaro (eds.), *The Structure of CP and IP: The Cartography of Syntactic Structures*, volume 2, pages 52–75, Oxford: Oxford University Press.
- Bresnan, Joan. 1978. A realistic transformational grammar. In Joan Bresnan Morris Halle and George A. Miller (eds.), *Linguistic Theory and Psychological Reality*, page 1–59, The MIT Press.
- Bresnan, Joan. 2001. *Lexical-functional Syntax*. Stanford: CSLI Publications.
- Bresnan, Joan, Asudeh, Ash, Toivonen, Ida and Wechsler, Stephen. 2016. *Lexical-functional Syntax*. Malden: Wiley-Blackwell.
- Bresnan, Joan and Mchombo, Sam A. 1987. Topic, pronoun, and agreement in Chicheŵa. *Language* 63, 741–782.
- Butt, Miriam and King, Tracy Holloway. 1996. Structural topic and focus without movement. In *Proceedings of the LFG '96 Conference*, Stanford: CSLI Publications.
- Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, subjects, topics

⁷A conference participant pointed out that examples similar to (13) are possible in Japanese.

⁸A reviewer questioned that in some carefully designed question-and-answer pairs, a predicted SCENE actually bears the FOCUS role. In these situations, it seems that context should take precedence. The reviewer then went on to question why should SCENES and TOPIC not be contextually inferred as well. An answer might be this: unlike FOCUS and BACKGROUND, SCENES and TOPIC appear invariably at the left periphery (in Cantonese), which motivates this proposal to capture these through formal tools. If one only relies on context to assign these information structure roles, then one will fall (again) into the profusion and confusion of terminologies and concepts in the literature of information structure.

- and point of view. In Charles N. Li (ed.), *Subject and Topic*, page 26–55, New York: Academic Press.
- Chao, Yuan-Ren. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.
- Cheung, Candice Chi Hang. 2015. On the fine structure of the left periphery. In Wei-Tien Dylan Tsai (ed.), *The Cartography of Chinese syntax*, page 75–130, Oxford: Oxford University Press.
- Dalrymple, Mary. 2023. *The Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.
- Dalrymple, Mary, Lowe, John J. and Mycock, Louise. 2019. *The Oxford Reference Guide to Lexical Functional Grammar*. Oxford: Oxford University Press.
- Dalrymple, Mary and Nikolaeva, Irina. 2011. *Objects and Information Structure*. Cambridge: Cambridge University Press.
- Erteschik-Shir, Nomi. 2019. Stage topics and their architecture. In Verner Egerland Valéria Molnár and Susanne Winkler (eds.), *Architecture of Topic*, page 223–248, Boston: Walter de Gruyter.
- Fung, Suet-Man. 2007. *Topic and focus in Cantonese: An OT-LFG account*. Masters Thesis, The University of Hong Kong.
- Huang, Ray Rui-Heng and Ting, Jen. 2006. Are there dangling topics in Mandarin Chinese? *Concentric: Studies in Linguistics* 32, 119–146.
- Huang, Yan. 1994. *The Syntax and Pragmatics of Anaphora*. Cambridge: Cambridge University Press.
- Kaplan, Ronald M. and Bresnan, Joan. 1982. Lexical-Functional Grammar: A formal system for grammatical representation. In Joan Bresnan (ed.), *The Mental Representation of Grammatical Relations*, page 173–281, Cambridge: The MIT Press.
- King, Tracy Holloway. 1997. Focus domains and information-structure. In *Proceedings of the LFG '97 Conference*, Stanford: CSLI Publications.
- Laczkó, Tibor. 2021. On the inventory of grammatical functions in LFG from a Hungarian perspective. In *Proceedings of the LFG '21 Conference*, page 181–201, Stanford: CSLI Publications.
- Lam, Olivia Shi-Ching. 2008. *Object Functions and the Syntax of Double Object Constructions in Lexical Functional Grammar*. Ph. D.thesis, University of Oxford.
- Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press.
- Li, Charles N. and Thompson, Sandra A. 1976. Subject and topic: A new typology of language. In Charles N. Li (ed.), *Subject and Topic*, page 457–489, New York: Academic Press.
- Li, Charles N. and Thompson, Sandra A. 1981. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley: University of California Press.
- Luke, K. K., Bodomo, A. and Nancarrow, O. 2001. The subject condition in Can-

- tonese. In *Proceedings of the LFG '01 Conference*, Stanford: CSLI Publications.
- Matthews, Stephen and Yip, Virginia. 2011. *Cantonese: A Comprehensive Grammar*. Oxon: Routledge, second edition.
- Mycock, Louise. 2006. *The Typology of Constituent Questions: A Lexical Functional Grammar Analysis of 'Wh'-Questions*. Ph. D.thesis, University of Manchester.
- Mycock, Louise. 2013. Discourse function of question words. In *Proceedings of the LFG '13 Conference*, page 419–439, Stanford: CSLI Publications.
- Nikolaeva, Irina. 2001. Secondary topic as a relation in information structure. *Linguistics* 39, 1–49.
- Pan, Haihua and Hu, Jianhua. 2008. A semantic–pragmatic interface account of (dangling) topics in Mandarin Chinese. *Journal of Pragmatics* 40, 1966–1981.
- Paul, Waltraud. 2015. *New Perspectives on Chinese Syntax*. Berlin: De Gruyter Mouton.
- Reinhart, Tanya. 1982. *Pragmatics and Linguistics: An Analysis of Sentence Topic*. Bloomington: Indiana University Linguistics Club.
- Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman (ed.), *Elements of Grammar*, page 281–337, Dordrecht: Kluwer.
- Shi, Dingxu. 2000. Topic and topic-comment constructions in Mandarin Chinese. *Language* 76, 383–408.
- Snijders, Liselotte. 2015. *The Nature of Configurationality in LFG*. Ph. D.thesis, University of Oxford.
- Xu, Liejiong and Langendoen, D. Terence. 1985. Topic structures in Chinese. *Language* 61, 1–27.
- Zaenen, Annie. 2023. Information structure. In Mary Dalrymple (ed.), *The Handbook of Lexical Functional Grammar*, Berlin: Language Science Press.