OBJECT Topicalization in Cantonese*

Cauvis Suet Man Fung

The University of Hong Kong

TOPIC and FOCUS topicalization constructions are commonly found in Cantonese, a Chinese language spoken by 71 million people. This paper accounts for the phenomenon of topicalization in Cantonese within the framework of Optimality-Theoretic Lexical-Functional Grammar (OT-LFG), focusing on constructions where the TOPIC or FOCUS fills the OBJ function.

While the [+New] feature is what licenses FOCUS topicalization in Cantonese, the NEW-L constraint cannot be directly applied to the language since a Cantonese FOCUS does not necessarily contain purely [+New] information and the [+New] information is not obliged to occur in the leftmost position of the FOCUS. The constraint [+New]-FOCUS ∧ FOCUS-L, along with other new constraints, are proposed in this study to account for the phenomenon. One hierarchy and two subhierarchies are established for TOPIC and FOCUS topicalization respectively.

1 INTRODUCTION

This paper investigates the phenomenon of OBJ topicalization in Cantonese within the OT-LFG framework, adopting Rosén’s (1998: 184) definition of topicalization which defines it as ‘a construction in which a leftmost constituent is understood as filling a missing constituent in the sentence’. The relevant constraints are identified and new constraints are introduced in order to establish constraint hierarchies for topicalization in Cantonese.

This paper is organized as follows. Section 2 discusses the methodology. Section 3 presents some observations on topicalization in Cantonese based on the results of a questionnaire survey. Section 4 identifies the relevant constraints and introduces new constraints. In section 5, constraint hierarchies for TOPIC and FOCUS topicalization are established. The paper is concluded in section 6.

2 METHODOLOGY

Based on the results of a questionnaire survey on topicalization in Cantonese completed by 40 native Cantonese speakers (see Fung, in preparation), some observations on Cantonese topicalization are outlined and the relevant constraints are identified. New constraints are introduced whenever necessary. Constraint hierarchies are established for topicalization.

3 OBSERVATIONS ON TOPICALIZATION IN CANTONESE

In this section, some observations on topicalization constructions in Cantonese are presented, based on the results of the questionnaire survey conducted.

When the phrase bearing the OBJ function is at the same time the discourse topic of the sentence, it can either be extracted to the initial position to become a topicalized discourse

* I would like to thank my supervisor, Dr. Adams Bodomo, for his valuable comments on this paper and the participants of the Fifth Cambridge Postgraduate Conference in Language Research for their useful feedback on my presentation. I am also grateful to Olivia Lam for her insightful comments on an earlier version of this paper.

© 2007 by Cauvis Suet Man Fung
CamLing 2007: 80-87
OBJECT Topicalization in Cantonese

81

topic (hereafter TOPIC) or stay in its canonical post-verbal position. The case is similar for a topicalized discourse focus (hereafter FOCUS). The OBJ carrying [+New] information can either be topicalized to become a FOCUS or stay in its canonical position. A FOCUS may contain [-New] information together with the [+New] element and it is not obligatory that all [+New] information appear inside the FOCUS (Fung 2006).

An OBJ can be partially topicalized. With an OBJ consisting of a Q0, a CL0 and an NP, the NP can be topicalized to become the TOPIC or the FOCUS, being separated from its CL0. It is however impossible to extract the CLP (consisting of a CL0 and an NP), leaving the Q0 in situ. Whenever the CL0 is topicalized, it has to be topicalized together with its NP complement. Constructions where the NP complement is left in situ are ill-formed.

4 THE RELEVANT CONSTRAINTS

The following are the constraints relevant to topicalization in Cantonese, according to the above observations:

4.1 DisTopic-L

One of the main goals of this paper is to find out when and how a discourse topic is topicalized to become a TOPIC. Though a discourse topic usually stretches across a number of utterances, what is being dealt with here is the phrase in a sentence which designates the discourse topic. The DisTopic-L constraint is introduced, which requires that the left edge of the phrase designating the discourse topic align with the left edge of the immediate sentence containing that phrase. Every element (to be defined below) intervening between the left edge of the discourse topic and the left edge of the sentence incurs one violation to the constraint.

4.2 Abut-NP(CL-HD) and Abut-CLP(Q-HD)

The Abut-NP(CL-HD) constraint, which says that an NP has to abut with its CL0, is introduced based on the observation that an NP can be topicalized leaving its CL0 in situ. Extracting a CLP separating it from its Q0 is however impossible. This illegitimacy has to be accounted for by the Abut-CLP(Q-HD) constraint, which requires that a CLP abut with its Q0. One violation is incurred to these constraints by each element occurring between the specified edges. The following phrase illustrates how these constraints work:

(2)  [OP saaml[CLP bun2[NP zidin2]]]
     three CL dictionary
     ‘three dictionaries’
This phrase satisfies Abut-CLP(Q-HD) since the left edge of the CLP, *saam1 ngo5 bun2 zi6din2* (three-1.SG-CL-dictionary), where the CLP, bun2 zi6din2, and the Q0, saam1, is separated by the D0, ngo5. The phrase in (2) also satisfies Abut-NP(CL-HD), with the left edge of the NP zi6din2 abutting with the right edge of the CL0, bun2. *Bun2 saam1 zi6din2* (CL-three-dictionary), for instance, incurs one violation to this constraint by separating the NP zi6din2 and the CL0 bun2 by the Q0, saam1.

In the above discussion, it is stated that the number of violations incurred is the number of *elements* intervening between the edges. The word *element* is interpreted as a categorial unit in this study. The example below illustrates what an element refers to:

(3) saam1 bun2 zi6din2 aa3 ngo5 maa5zo2
    three CL dictionary PART 1.SG buy.PERF
    ‘It is three dictionaries that I have bought.’

In (3), the sentence-initial QP, which consists of three elements including a Q0 (saam1), a CL0 (bun2) and an NP (zi6din2), and the particle aa3 standing in front of the SUBJ ngo5 incur in total four violations to the SUBJ-L constraint.

### 4.3 [+New]-FOCUS ∧ FOCUS-L

The [+New]-FOCUS ∧ FOCUS-L constraint is derived particularly for FOCUS topicalization. The two occurrences of 'FOCUS' in the constraint refer to a topicalized discourse focus. This constraint requires that there be a FOCUS in the sentence whenever there exists some [+New] information and that the FOCUS should occur in the leftmost position. One violation is incurred either when there is some [+New] information but no FOCUS in the sentence or when the FOCUS does not align left.

Choi (1999) argues that topicalization is employed in English to encode information prominence instead of newness. PROM-L is then a relevant constraint for characterizing English topicalization. As discussed in Fung (2006), FOCUS in Cantonese can encode discourse focus bearing either the [+Prom] or [-Prom] feature and it is the [+New] feature, rather than the [+Prom] feature, which licenses FOCUS topicalization. Choi’s (2001) NEW-L, however, cannot be directly applied to Cantonese. Although the [+New] feature licenses FOCUS topicalization, a FOCUS may also at the same time contain some [-New] information, and it is not obligatory for the [+New] element to appear in the leftmost position of the sentence. Neither is it necessary that all [+New] information appear inside the FOCUS. Based on these observations, the conjunctive constraint [+New]-FOCUS ∧ FOCUS-L is introduced, which states that there must be a FOCUS when there exists some [+New] information in the sentence and that the FOCUS should align left. This constraint conjunction does not require that all [+New] information should appear inside the FOCUS or that the FOCUS should contain only [+New] information. Neither does it demand that the [+New] information align left.

### 4.4 *Topicalize-FC

*TOPICALIZE-FC is introduced based on the observation that sentence (4) is well-formed while sentence (5) is ill-formed (assume that zi6din2 and saam1 bun2 bear the [+New] feature in (4) and (5) respectively, so that they are topicalized to satisfy [+New]-FOCUS ∧ FOCUS-L):

(4) zi6din2 aa3 ngo5 maa5zo2 saam1 bun2
    dictionary PART 1.SG buy.PERF three CL
    ‘I have bought three dictionaries.’
While the topicalized phrase in (4) is an NP, the topicalized elements Q₀ and CL₀ in (5) do not form a maximal projection. This, however, does not necessarily lead to ungrammaticality. Consider example (6).

(6) *maai5zo2 laa1 ngo5 go2 bun2 syu1
    buy.PERF PART 1.SG DEF CL book
    ‘I have already bought that book.’

In this example, the focus maai5zo2 is a V, which is not a maximal projection either. This construction shows that a topicalized element is not necessarily a maximal projection.

Another difference between constructions (4) and (5) is that the topicalized phrase in (4) is an NP which takes no complement. In (5), it is a Q₀ and a CL₀ which are topicalized, with the NP complement of the CL₀, zi6din2, being left in situ. Leaving the complement in situ, again, cannot explain its unacceptability. In (6), the V maai5zo2 is topicalized to become the focus, also leaving its CLP complement go2 bun2 syu1 in its canonical OBJ position.

Though the complements of the topicalized CL₀ and V₀ in sentences (5) and (6) respectively are both left in situ, the two constructions differ in the syntactic category of the topicalized element. In (6), the stranded CLP is the complement of the topicalized V₀, which is a lexical category, while the stranded NP in (5) is the complement of the topicalized CL₀, which is a functional category. Based on this observation, the *TOPICALIZE-FC constraint is proposed, which disallows topicalizing only a functional category with its complement being left in situ. One violation is incurred to this constraint for every topicalized functional category whose complement stays in its canonical position.

It should, however, be noted that, since Cantonese is a pro-drop language, the complement can be dropped, provided that it can be recovered from the context. If the NP zi6din2 in (5) is dropped, the sentence becomes grammatical:

(7) *saam1 bun2 aa3 ngo5 maai5zo2
    three CL PART 1.SG buy.PERF
    ‘It is three (dictionaries) that I have bought.’

To conclude, what *TOPICALIZE-FC prohibits is that the functional category is topicalized with the complement being left in situ. The functional category is free to be topicalized when the complement can be recovered from the context and is dropped.

5 THE CONSTRAINT HIERARCHIES

The following two hierarchies are arrived at based on the observations presented in section 3:

(8) (a) Constraint hierarchy for TOPIC topicalization:
    DistOPIC-L >> SUBJ-L, Abut-OBJ(V-HD), Abut-NP(CL-HD)

(b) Constraint hierarchy for FOCUS topicalization:
    Abut-CLP(Q-HD), *TOPICALIZE-FC >> [+New]-FOCUS ∧ FOCUS-L >> SUBJ-L, Abut-OBJ(V-HD), Abut-NP(CL-HD)

Since both the canonical structure and the topicalization construction are acceptable in some cases, the notion of free ranking (Kager 1999) is involved. This paper, however, concentrates on the phenomenon of topicalization, and therefore only considers the hierarchies for selecting the topicalization constructions as the optimal outputs.
In this paper, it is assumed that the OT input consists of an underspecified f-structure and an i-structure (King 1997). King (1997) mentions that every constituent is assigned a discourse label in the i-structure. It is therefore normal to find discourse topic, [+New] and [-New] information coexisting in an i-structure. Though it is possible for some constructions to have TOPIC as well as FOCUS in an attempt to satisfy both DISTOPIC-L and [+New]-FOCUS ∧ FOCUS-L, this kind of construction is not commonly found in the language. Since a language is assumed to have one single hierarchy (subhierarchies are possible where options exist), it is reasonable and desirable to combine the above-established hierarchies for TOPIC and FOCUS constructions into one hierarchy for Cantonese. This issue will however be left for future research. The inputs given in the following will therefore be independently evaluated against the hierarchy for TOPIC topicalization and the one for FOCUS topicalization, and constructions with multiple topicalized phrases will not be included in the candidate set. The constraint ranking for TOPIC constructions will be examined first. Consider the following input:

(9) (a) 

\[
\begin{align*}
\text{PRED} & \quad \text{‘deu6} <x, y>’ \\
\text{ASPECT} & \quad \text{PERF} \\
\text{GF}_1 & \quad [\text{PRED} \quad \text{‘Pro’} \\
& \quad \text{NUM} \quad \text{SG} \\
& \quad \text{PERS} \quad 1 \quad x] \\
\text{GF}_2 & \quad [\text{PRED} \quad \text{‘zi6din2’} \\
& \quad \text{CL} \quad \text{bun2} \\
& \quad \text{DEF} \quad + \quad y] \\
\end{align*}
\]

Diagram 1

Input I – f-structure

(b) 

\[
\begin{align*}
\text{DISCOURSE TOPIC} & \quad \{\text{zi6din2}\} \\
+\text{New} & \quad \text{ngo5 deu6zo2} \\
\end{align*}
\]

Diagram 2

Input I – i-structure

Table 1 shows how the optimal output is selected with hierarchy (8a).

(10) 

<table>
<thead>
<tr>
<th>a. ngo5 deu6zo2</th>
<th>bun2 zi6din2</th>
<th>DISTOPIC-L</th>
<th>SUB-L</th>
<th>Abut-OB(V-HD)</th>
<th>Abut-NP(CL-HD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSG throw away.PERF</td>
<td>CL dictionary</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. bun2 zi6din2</td>
<td>ngo5 deu6zo2</td>
<td>***</td>
<td>**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CL dictionary PART LSG throw away.PERF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1

Evaluation of Input I with hierarchy (8a)
Candidate (10a), with the canonical structure is ruled out by violating the highest-ranked Distopic-L. With Distopic-L outranking Subj-L, Candidate (10b), the topicalization construction, becomes the optimal output.

Now consider another input:

\[
\begin{array}{c}
(11) \\
\text{(a)}
\end{array}
\]

\[
\begin{array}{c}
\text{Diagram 3} \\
\text{Input II – f-structure}
\end{array}
\]

\[
\begin{array}{c}
\text{(b)}
\end{array}
\]

\[
\begin{array}{c}
\text{Diagram 4} \\
\text{Input II – i-structure}
\end{array}
\]

Since the discourse topic is at the same time the Subj ngo5, it is obvious that the canonical SVO structure, which satisfies both Distopic-L and Subj-L, is the optimal output with respect to the hierarchy for Topic topicalization. The following discussion will then concentrate on the evaluation against the hierarchy for Focus topicalization in (8b). The tableau below shows the evaluation:

\[
\begin{array}{c}
(12)
\end{array}
\]

\[
\begin{array}{c}
\text{Table 2} \\
\text{Evaluation of Input II with hierarchy (8b)}
\end{array}
\]
The canonical construction (12a) violates [+New]-FOCUS ∧ FOCUS-L by having no FOCUS. Candidate (12c) violates Abut-CLP(Q-HD) by topicalizing the CLP bun2 zi6din2, separating it from the Q0 saam1. Since SUBJ-L, Abut-OBJ(V-HD) and Abut-NP(CL-HD) are on the same stratum of the hierarchy, candidate (12d) wins over candidate (12b) by incurring fewer cumulative violations to the three constraints.

According to the results of the questionnaire survey, both candidates (12b) and (12d) are acceptable given that the [+New] information falls on the NP zi6din2. The hierarchy has to be revised since it only selects (12d) as the optimal output. The number of violations to SUBJ-L and Abut-NP(CL-HD) differ for the two candidates. With SUBJ-L outranking Abut-NP(CL-HD), candidate (12b) becomes optimal. With Abut-NP(CL-HD) outranking SUBJ-L, candidate (12d) is optimal. The tableaux below show how the two optimal outputs are selected with different subhierarchies:

Table 3
Evaluation of Input II with SUBJ-L outranking Abut-NP(CL-HD)

<table>
<thead>
<tr>
<th></th>
<th>Abut-CLP(Q-HD)</th>
<th>+New]-FOCUS ∧ FOCUS-L</th>
<th>Abut-OBJ(V-HD)</th>
<th>SUBJ-L</th>
<th>Abut-NP(CL-HD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>ngo5 maai5zo2 saam1 bun2 zi6din2 1.SG buy.PERF three CL dictionary</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>zi6din2 aa3 ngo5 maai5zo2 saam1 bun2 dictionary PART 1.SG buy.PERF three CL</td>
<td>**</td>
<td>**</td>
<td>*****</td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>bun2 zi6din2 aa3 ngo5 maai5zo2 saam1 CL dictionary PART 1.SG buy.PERF three</td>
<td>*!</td>
<td>**</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td>saam1 bun2 zi6din2 aa3 ngo5 maai5zo2 three CL dictionary PART 1.SG buy.PERF</td>
<td>**</td>
<td>****</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4
Evaluation of Input II with Abut-NP(CL-HD) outranking SUBJ-L

<table>
<thead>
<tr>
<th></th>
<th>Abut-CLP(Q-HD)</th>
<th>+New]-FOCUS ∧ FOCUS-L</th>
<th>Abut-OBJ(V-HD)</th>
<th>Abut-NP(CL-HD)</th>
<th>SUBJ-L</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>ngo5 maai5zo2 saam1 bun2 zi6din2 1.SG buy.PERF three CL dictionary</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>zi6din2 aa3 ngo5 maai5zo2 saam1 bun2 dictionary PART 1.SG buy.PERF three CL</td>
<td>**</td>
<td>**</td>
<td>*****</td>
<td>**</td>
</tr>
<tr>
<td>c.</td>
<td>bun2 zi6din2 aa3 ngo5 maai5zo2 saam1 CL dictionary PART 1.SG buy.PERF three</td>
<td>*!</td>
<td>**</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td>saam1 bun2 zi6din2 aa3 ngo5 maai5zo2 three CL dictionary PART 1.SG buy.PERF</td>
<td>**</td>
<td>****</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
It has been assumed that SUBJ-L, Abut-OBJ(V-HD) and Abut-NP(CL-HD) are on the same stratum. According to Tesar and Smolensky (1998), all constraints are on the highest level of the hierarchy at the initial stage for constraint demotion. Therefore, when Abut-NP(CL-HD) in Tableau 3 or SUBJ-L in Tableau 4 is demoted, Abut-OBJ(V-HD) stays with the higher-ranked constraint for further demotion if necessary.

6 CONCLUSIONS

In this paper, an OT-LFG account has been proposed for the phenomenon of topicalization in Cantonese. Some observations on this phenomenon have been presented, based on which the relevant constraints have been arrived at. Five new constraints have been introduced, which are DISTOPIC-L, Abut-NP(CL-HD), Abut-CLP(Q-HD), [+New]-FOCUS ∧ FOCUS-L and *TOPICALIZE-FC. The constraints have been ranked against each other and one hierarchy and two subhierarchies have been established for TOPIC topicalization and FOCUS topicalization respectively.

It is hoped that this OT-LFG account of OBJ topicalization in Cantonese will contribute to the field of Cantonese syntax by demonstrating how the framework can be applied to the language. Topicalization constructions involving different grammatical functions can be investigated in the future in order to provide a comprehensive account of topicalization.

REFERENCES


Cauvis Suet Man Fung

Linguistics Office,
School of Humanities,
The University of Hong Kong,
Pokfulam Rd,
HONG KONG

cauvis@graduate.hku.hk