

# A Constraint-based Analysis of A-NOT-A Questions in Mandarin Chinese

Wenjie Wang

Nanyang Technological University

Sanghoun Song

Incheon National University

Francis Bond

Nanyang Technological University

Proceedings of the 22nd International Conference on  
Head-Driven Phrase Structure Grammar

Nanyang Technological University (NTU), Singapore

Stefan Müller (Editor)

2015

CSLI Publications

pages 195–214

<http://csli-publications.stanford.edu/HPSG/2015>

Wang, Wenjie, Song, Sanghoun, & Bond, Francis. (2015). A Constraint-based Analysis of A-NOT-A Questions in Mandarin Chinese. In Stefan Müller (Ed.): *Proceedings of the 22nd International Conference on Head-Driven Phrase Structure Grammar, Nanyang Technological University (NTU), Singapore* (pp. 195–214). Stanford, CA: CSLI Publications.

## Abstract

The *A*-NOT-*A* structure is one way to express polar questions in Mandarin Chinese. The present study provides a constraint-based analysis of *A*-NOT-*A* questions in Mandarin Chinese within the framework of HPSG (Pollard & Sag, 1994) and MRS (Copestake et al., 2005). We propose two possible approaches to analysing the *A*-NOT-*A* structure — a morphological/lexical approach as well as a syntactic approach — and illustrate their implementation, as well as their respective strengths and weaknesses.

# 1 Introduction

## 1.1 Basic Properties

The *A*-NOT-*A* structure is one way to express polar questions in Mandarin Chinese. The structure is so termed because it consists of an element (*A*) that is followed immediately by the same element but of negative polarity (NOT-*A*). For ease of reference, we shall refer to these elements as  $A_1$  and  $A_2$  respectively.

The *A*-NOT-*A* structure exists in various forms, which are exemplified below:

- (1) a. Basic: *A*-NOT-*A*  
张三 喜欢<sub>A1</sub> 不 喜欢<sub>A2</sub> 狗 ?  
Zhāngsān xǐhuān bù xǐhuān gǒu ?  
Zhangsan like NOT like dog PU
- b. Contracted: *A'*-NOT-*A*  
张三 喜 不 喜欢 狗 ?  
Zhāngsān xǐ bù xǐhuān gǒu ?  
Zhangsan like NOT like dog PU
- c. Phrasal: *AO*-NOT-*AO*  
张三 喜欢 狗 不 喜欢 狗 ?  
Zhāngsān xǐhuān gǒu bù xǐhuān gǒu ?  
Zhangsan like dog NOT like dog PU
- d. Phrasal: *AB*-NOT-*A*  
张三 喜欢 狗 不 喜欢 ?  
Zhāngsān xǐhuān gǒu bù xǐhuān ?  
Zhangsan like dog NOT like PU

All variations presented in (1) convey almost the same meaning: “Does Zhangsan like dogs?”

### 1.1.1 Reduplication

As shown in (1),  $A_1$  and  $A_2$  are reduplicates of each other. Reduplication for *A*-NOT-*A* can be performed *partially*: (1b) shows that  $A_1$  can be reduplicated with just its first character/syllable, while (1d) shows that the verb can be reduplicated

without its complement. Note that in both cases,  $A_2$  must itself be *fully* reduplicated. As such, the following are ungrammatical:

- (2) a. \* 张三 喜欢 不 喜 狗 ?  
 \* Zhāngsān xǐhuān bù xǐ gǒu ?  
 Zhangsan like NOT like dog PU
- b. \* 张三 喜欢 狗 不 喜 ?  
 \* Zhāngsān xǐhuān gǒu bù xǐ ?  
 Zhangsan like dog NOT like PU

### 1.1.2 What can be $A$ ?

All lexical types capable of behaving as a syntactic head of predicates in Mandarin Chinese, such as verbs, adjectives, and prepositions, can participate in the  $A$ -NOT- $A$  structure as  $A$  elements (Tseng, 2009). In the examples below, adjectives and prepositions (co-verbs) are shown playing the role of  $A$  elements:

- (3) a. 张三 高 不 高 ?  
 Zhāngsān gāo bù gāo ?  
 Zhangsan tall NOT tall PU  
 ‘Is Zhangsan tall (or not tall)?’
- b. 张三 在 不 在 家 ?  
 Zhāngsān zài bù zài jiā ?  
 Zhangsan at NOT at home PU  
 ‘Is Zhangsan at home (or not at home)?’

Adverbs are not allowed to be  $A$  elements, with the exception of frequency adverbs such as 常 *cháng* “often”:

- (4) a. \* 张三 很 不 很 高 ?  
 \* Zhāngsān hěn bù hěn gāo ?  
 Zhangsan very NOT very tall PU  
 (Intended: ‘Is Zhangsan very tall?’)
- b. 张三 常 不 常 迟到 ?  
 Zhāngsān cháng bù cháng chí-dào ?  
 Zhangsan often NOT often late-arrive PU  
 ‘Is Zhangsan often late?’

$A$  elements cannot be reduplicated elements themselves. As such, although the frequency adverb 常 *cháng* can be an  $A$  element, its reduplicated form 常常 *cháng cháng* cannot.

- (5) a. \* 张三 常常 不 常常 迟到 ?  
 \* Zhāngsān cháng-chang bù cháng-chang chí-dào ?  
 Zhangsan often NOT often late-arrive PU  
 ‘Is Zhangsan often late?’

### 1.1.3 What can be NOT?

Mandarin Chinese employs two negative operators (不 *bù* and 没 *méi*), the choice of which hinges on the aspectual property of the verbal item that they are attached to: 不 *bù* for statives and imperfectives, and 没 *méi* for bound events and perfectives. This is exemplified in (6).

Both of them can participate in the *A*-NOT-*A* structure as NOT, and likewise the aspect of the *A* element determines which is used. They also have slightly different co-occurrence constraints.

- (6) a. 去 不 去 ?  
 qù bù qù ?  
 go NOT go PU  
 ‘Are you going?’
- b. 去 没 去 ?  
 qù méi qù ?  
 go NOT go PU  
 ‘Have you gone (somewhere)?’

## 1.2 Basic Constraints

### 1.2.1 Modifiability of *A* elements

The *A* elements in *A*-NOT-*A* cannot take modifiers, such as degree adverbs, or aspectual markers:

- (7) a. \* 张三 很高 不 很高 ?  
 \* Zhāngsān hěn gāo bù hěn gāo ?  
 Zhangsan very tall NOT very tall PU  
 ‘Is Zhangsan very tall?’
- b. \* 张三 去了 不 去了 ?  
 \* Zhāngsān qù le bù qù le ?  
 Zhangsan go LE NOT go LE PU  
 (Intended: ‘Zhangsan went?’)

The exception is the A-MEI-A sub-pattern, which can be post-modified by the experiential aspectual marker 过 *guò*.

- (8) 张三 去过 没 去过 ?  
 Zhāngsān qù guò méi qù guò ?  
 Zhangsan go GUO NOT go GUO PU  
 'Has Zhangsan been there before?'

### 1.3 Co-occurrence Constraints

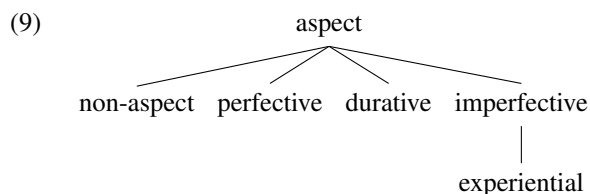
#### 1.3.1 Sentence-final particles

A-NOT-A questions are not permitted to occur with certain sentence-final particles. In the cases of 了 *le*, 吗 *ma*, 吧 *ba*, 哦 *o* and 耶 *ye*, it is because only propositions can be used with these sentence-final particles, whereas A-NOT-A is a question.

Other sentence-final particles like the emphatic markers 嘛 *ma*, 呀 *ya* and 呢 *ne* do not, however, restrict themselves to only propositions and are therefore permitted to be used with A-NOT-A.

#### 1.3.2 Aspectual markers

Chinese is an aspect-based language, in which aspect is linguistically and necessarily expressed, and plays an important role in syntax. The aspect hierarchy of Chinese (as implemented in ZHONG [ | ]) is roughly sketched out in (9):



Grammatical aspect in Chinese is largely expressed by verbal markers. There are three aspectual markers in Mandarin Chinese: 了 *le*, 着 *zhe*, and 过 *guò*, which indicate the perfective, durative, and experiential aspects respectively. Since each verb lexically selects these markers, not all these three items can be necessarily attached to all verbs. For example, 去 *qù* 'go' does not canonically co-occur with *zhe*. These markers are collectively known as LE-ZHE-GUO or LZG, and they are hierarchically constrained as described in (10) in Type Definition Language.

- (10) +vjp :+ [ LZG lzg ].  
 lzg := avm.  
 le := lzg.  
 zhe := lzg.  
 guo := lzg.

no-lzg := lzg.  
 le+zhe := le & zhe.  
 le+guo := le & guo.  
 zhe+guo := zhe & guo.  
 le+zhe+guo := le & zhe & guo.

The LE-ZHE-GUO markers are also restricted in their co-occurrence with *A-NOT-A*, either with the entire *A-NOT-A* phrase, or with the individual *A* elements (See Section 1.2.1). The markers *lè* and *zhè* are not allowed to co-occur with *A-NOT-A* at all, while *guò* can only occur with *A-NOT-A* if the NOT element is 没 *méi*.

#### 1.4 Versus MA-questions

The MA-question is another type of polar question, in which a sentence-final particle 吗 *mā* is used.

For example, (11) has a similar meaning to (1).

- (11) 张三 喜欢 狗 吗 ?  
 Zhāngsān xǐhuān gǒu ma ?  
 Zhangsan like dog MA PU

‘Does Zhangsan like dogs?’

On the surface, both (1) and (11) are translated as “Does Zhangsan like dogs?”, and thus appear allo-structural and the semantic representation should be almost the same in order for one form to be paraphrased into the other form. However, there are at least three reasons for believing that they are not equivalent:

Firstly, they are pragmatically different. MA-questions are seen as being bi-ased towards the overtly indicated proposition (*p*), whereas *A-NOT-A* questions are neutral as both propositions (*p* and  $\neg p$ ) are indicated (Liing, 2014), barring the differences arising due to sequential order.

Secondly, they differ in terms of information structure. MA-questions can have focus on any of its constituents. For instance, in (11), either the subject 张三 *Zhāngsān*, the object 狗 *gǒu*, or the verb 喜欢 *xǐhuān* can be evaluated as containing focus. Should focus be required, the asker employs a specific prosodic clue and/or the focus marker 是 *shì*. (12) presents that different constituents in MA-questions can be freely clefted.

- (12) a. 是 张三 喜欢 李四 吗 ?  
 shì Zhāngsān xǐhuān Lǐsì ma ?  
 SHI Zhangsan like Lisi MA PU

‘Is it Zhangsan (and not anyone else) who likes Lisi?’

- b. 张三 是 喜欢 李四 吗 ?  
 Zhāngsān shì xǐhuān Lǐsì ma ?  
 Zhangsan SHI like Lisi MA PU  
 ‘Is it that Zhangsan likes Lisi?’
- c. 张三 喜欢 的 是 李四 吗 ?  
 Zhāngsān xǐhuān de shì Lǐsì ma ?  
 Zhangsan like DE SHI Lisi MA PU  
 ‘Is it Lisi whom Zhangsan likes?’

This is because the scope of *mā* is not explicitly observable from the sentence itself. By contrast, *A-NOT-A* does not signal focus to any other elements but the structure itself (i.e., no ambiguity). The subject and the object in *A-NOT-A* questions cannot pass the cleft test exemplified in (12). In other words, *A-NOT-A* always bears focus (i.e., predicate focus).

Thirdly, they differ semantically. When a universal quantifier 都 *dōu* is used, a scope ambiguity happens with MA-questions but not with *A-NOT-A* questions, as shown in (13). (McCawley, 1994)

- (13) a. 他们 都 喜欢 不 喜欢 开车 ?  
 tāmen dōu xǐhuān bù xǐhuān kāichē ?  
 they all like NOT like drive PU  
 ‘Do they all like to drive?’
- b. 他们 都 喜欢 开车 吗 ?  
 tāmen dōu xǐhuān kāichē ma ?  
 they all like drive MA PU  
 ‘Do they all like to drive?’ or  
 ‘Do all of them like to drive?’

## 2 HPSG Account

This section proposes two possible approaches to handling the *A-NOT-A* structure: 1) the morphological/lexical approach and 2) the syntactic approach.

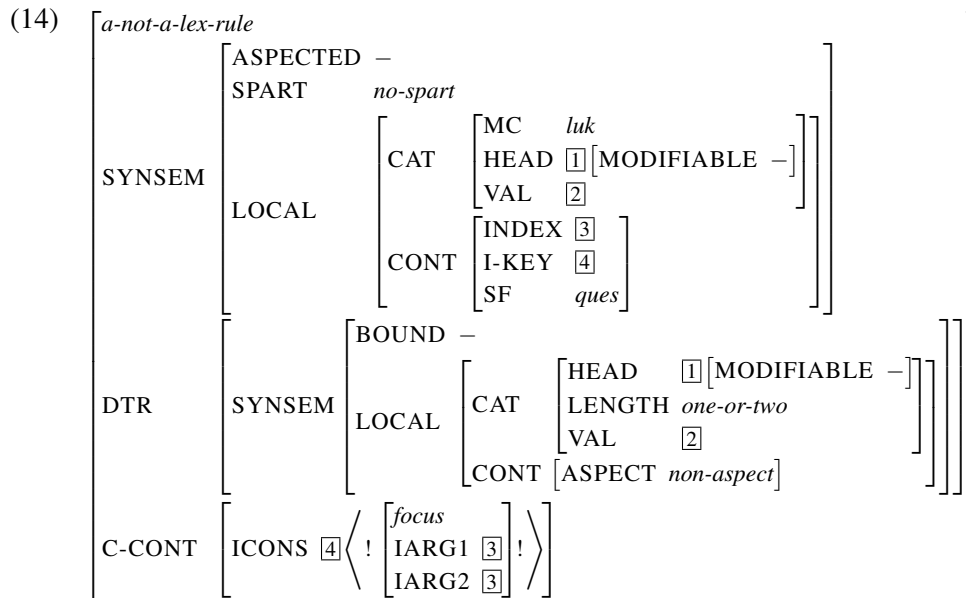
### 2.1 Approach 1: Morphological/Lexical Approach

In this approach, the *A-NOT-A* structure is handled from the lexicon and thus its morphology. The *A-NOT-A* structure is dealt with as a single morphological word, and this allows us to treat the *A-NOT-A* element as a single predicate in the semantics. This approach aligns with the implementation chosen for reduplicated

adjectives in ZHONG [ ] (Fan et al., 2015). The treatment of the *A-NOT-A* structure as a “monolithic” morphological word also means that the modification of the *A* element is naturally prevented from happening. Constraints on the modification of the entire *A-NOT-A* structure are also much more easily implemented.

### 2.1.1 Parent/Super Lexical Rule

A super-type lexical rule (*a-not-a-lex-rule*) provides the general constraints and definition of the structure. This rule is responsible for the conversion of any lexical entry that can participate as *A* elements into *A-NOT-A* and thereafter provides the relevant information for the structure. Key sections of the *a-not-a-lex-rule* are illustrated below.



The lexical rule indicates that the *A-NOT-A* structure bears the sentence force (SF) of *ques*. A feature type MODIFIABLE is used to state that the *A-NOT-A* structure cannot be modified. Focus is represented via ICONS (Individual CONstraints) (Song, 2014). The I-KEY feature points to ICONS, indicating that *A-NOT-A* is the focus of the sentence. IARG1 and IARG2 both point to the INDEX of the *A-NOT-A* structure itself.

Two lexical rules will inherit from this parent lexical rule. These two child lexical rules are for the *A-不-A* and *A-没-A* sub-patterns discussed earlier in Section 1.1.3.

### 2.1.2 Lexical Rule for *A-不-A* sub-pattern

This child lexical rule handles the *A-不-A* sub-pattern, and inherits from *a-not-a-lex-rule*. The additional constraints are indicated below:



$$(15) \left[ \begin{array}{l} a\text{-not-a-bu-lex-rule} \\ \text{SYNSEM} \mid \text{CONT} \left[ \begin{array}{l} \text{ASPECT } non\text{-aspect} \\ \text{LZG } no\text{-lzg} \end{array} \right] \end{array} \right]$$

The ASPECT has been indicated as *non-aspect*, which prevents sentence-final particles from modifying the structure. The LZG feature is given a value of *no-lzg*, which prevents the aspectual markers from modifying the structure.

### 2.1.3 Lexical Rule for A-没-A sub-pattern

This child lexical rule handles the A-没-A sub-pattern, and inherits from *a-not-a-lex-rule*. The additional constraints are indicated below:

$$(16) \left[ \begin{array}{l} a\text{-not-a-mei-lex-rule} \\ \text{SYNSEM} \mid \text{CONT} \left[ \begin{array}{l} \text{ASPECT } imperfective \\ \text{LZG } guo \end{array} \right] \end{array} \right]$$

The ASPECT feature is given the value of *imperfective*. As this sub-pattern allows the experiential aspectual marker 过 *guò* to modify the structure, we provide the LZG feature with a value of *guo*.

### 2.1.4 Handling A'-NOT-A

From (1b), duplicated here as (17), we see an example of the A'-NOT-A sub-pattern.

$$(17) \text{ Contracted: A'-NOT-A} \\ \text{张三 喜 不 喜欢 狗 ?} \\ \text{Zhāngsān xǐ bù xǐhuān gǒu ?} \\ \text{Zhangsan like NOT like dog PU}$$

To recap, this pattern exhibits partial reduplication, where only the first syllable/character of  $A_1$  is reduplicated. Nevertheless, apart from surface form, it is identical to its fully reduplicated counterpart. As such, this pattern is first transformed into the fully reduplicated pattern before being handled by the lexical rules. The mechanism for this is described in Section 2.1.6.

### 2.1.5 Sample Derivation

Using the sentence 张三喜欢不喜欢狗? ‘Does Zhangsan like dogs?’, we derive the MRS in (18):

$$(18) \left[ \begin{array}{l} \text{INDEX } \boxed{2} \left[ \begin{array}{l} \text{SF } \textit{ques} \\ \text{ASPECT } \textit{non-aspect} \end{array} \right] \\ \\ \text{RELS } \left\langle \begin{array}{l} \left[ \begin{array}{l} \textit{named\_rel} \\ \text{LBL } \boxed{4} \\ \text{CARG } \textit{‘张三’} \\ \text{ARG0 } \boxed{3} \end{array} \right], \left[ \begin{array}{l} \textit{proper\_q\_rel} \\ \text{LBL } \boxed{6} \\ \text{ARG0 } \boxed{3} \\ \text{RSTR } \boxed{7} \\ \text{BODY } \boxed{8} \end{array} \right], \\ \left[ \begin{array}{l} \textit{喜欢\_v\_I\_rel} \\ \text{LBL } \boxed{1} \\ \text{ARG0 } \boxed{2} \\ \text{ARG1 } \boxed{3} \\ \text{ARG2 } \boxed{9} \end{array} \right], \left[ \begin{array}{l} \textit{狗\_n\_I\_rel} \\ \text{LBL } \boxed{10} \\ \text{ARG0 } \boxed{9} \end{array} \right], \left[ \begin{array}{l} \textit{exist\_q\_rel} \\ \text{LBL } \boxed{11} \\ \text{ARG0 } \boxed{9} \\ \text{RSTR } \boxed{12} \\ \text{BODY } \boxed{13} \end{array} \right] \end{array} \right\rangle \\ \\ \text{HCONS } \left\langle \left[ \begin{array}{l} \textit{qeq} \\ \text{HARG } \boxed{0} \\ \text{LARG } \boxed{1} \end{array} \right], \left[ \begin{array}{l} \textit{qeq} \\ \text{HARG } \boxed{7} \\ \text{LARG } \boxed{4} \end{array} \right], \left[ \begin{array}{l} \textit{qeq} \\ \text{HARG } \boxed{12} \\ \text{LARG } \boxed{10} \end{array} \right] \right\rangle \\ \\ \text{ICONS } \left\langle \left[ \begin{array}{l} \textit{focus} \\ \text{IARG1 } \boxed{2} \\ \text{IARG2 } \boxed{2} \end{array} \right] \right\rangle \end{array} \right]$$

The semantic head  $\boxed{2}$  has [SF *ques*], which indicates that the sentence is interrogative. Within the semantics, the *A-NOT-A* structure has only a single predicate *喜欢\_v\_I\_rel*. The element in ICONS is specified as *focus*, and both IARG1 and IARG2 are coindexed with the INDEX of the verb. This means that the *A-NOT-A* structure is the focus within the clause.

In most areas, the MRS for the *A-NOT-A* structure is close to that of MA-questions, with a number of key differences explained in an earlier section pertaining to the areas such as the focus and the aspect. Barring these, MA-questions can technically be generated from our implementation of the *A-NOT-A* structure, and can likewise be provided alongside *A-NOT-A* questions as suitable candidates during machine translation.

### 2.1.6 Implementation in Zhong [1]

In a nutshell, the input is first cleaned up by a regular expression preprocessor (REPP) and readied for parsing. The cleaning up includes removal of spaces left over from segmentation (Eg: 高不高 → 高不高), and replacing the reduplicated parts with the character 々<sup>1</sup> (Eg: 高不高 → 高不々). The segment 不々 is treated by the parser as a suffix, which it can then remove and reduce the structure to just the *A* element, and subsequently match with its appropriate lexical entry (Eg: 高不々 → 高). This lexical item will then be passed through the *a-not-a-lex-rule* and the relevant child lexical rule, and will then be given the features and semantics of the *A-NOT-A* structure.

<sup>1</sup>This character is adopted from Japanese, which uses it to indicate the reduplication of the character that precedes it.

As explained earlier in Section 2.1.4, the A'-NOT-A sub-pattern (the contracted pattern) is identical — apart from surface form — to its fully reduplicated counterpart. As such, the REPP will pick up these contracted patterns in the input and transform them into their fully-reduplicated forms, removing any spaces along the way (Eg: 喜 不 喜欢 → 喜欢不喜欢), and also replace the reduplicated element with the character 々 (Eg: 喜欢不喜欢 → 喜欢不々). As with the above, the structure is then reduced to the *A* element, and then be passed through the *a-not-a-lex-rule*.

It should be noted that the implementation is done based on the functions and limitations of the system, and it does not reflect any assumptions on the actual parsing of the structure by a speaker.

### 2.1.7 Limitations

This method does not allow us to constrain the objects to be identical in the AO-NOT-AO pattern, as the *O* elements can be diverse and be too vast to feasibly implement. The *O* elements can also, potentially, be of an arbitrarily long length as long as it is a grammatically correct verb phrase. However, such long sentences are not necessarily accepted by speakers due to the cumbersome nature of it, even if they do not violate any grammatical rules.

Also, because of the treatment of the *A*-NOT-*A* structure as a single morphological word, the formation of the structure remains opaque to the grammatical system, which only sees the structure as a single lexical entry.

This approach does not allow us to cover A-MEI-A patterns where the *A* element is modified by *guò*, as illustrated in (19):

- (19) a. (Unmodified)  
       吃 没 吃  
       chī méi chī  
       eat MEI eat
- b. (With Experiential GUO)  
       吃 过 没 吃 过

Using this approach, it will require that this pattern be also generated in the morphology, or automatically detected when parsing, and then given additional rules that take into account the aspectual marker.

An initial issue that we had believed might arise from this approach was that the lexicon could become very large if each *A* element were to have a separate lexical entry for its respective *A*-NOT-*A* form(s). However, with the *A*-NOT-*A* structure now being automatically detected and pre-processed (such that separate lexical entries are no longer needed), this disadvantage is largely removed.

## 2.2 Approach 2: Syntactic Approach

The syntactic approach builds the *A*-NOT-*A* structure as three components:  $A_1$ , NOT and  $A_2$ .

### 2.2.1 Characters

The *A* elements in *A*-NOT-*A* are full or partial reduplicates of each other. One such form is that only the first character of  $A_1$  is reduplicated. With this in mind, we introduce new feature types to the lexicon entries, as underlined in (20):

(20)

+vjp												
STEM		☐										
BOUND		<i>luk</i>										
SPART		<i>spart</i>										
HEAD		CHAR	<table style="border-collapse: collapse; margin: 0 auto;"> <tr> <td style="padding: 2px 5px;"><i>char</i></td> <td></td> </tr> <tr> <td style="border-bottom: 1px solid black; padding: 2px 5px;">FCHAR</td> <td style="padding: 2px 5px;"><i>string</i></td> </tr> <tr> <td style="border-bottom: 1px solid black; padding: 2px 5px;">WCHAR</td> <td style="padding: 2px 5px;">☐</td> </tr> <tr> <td style="border-bottom: 1px solid black; padding: 2px 5px;">LENGTH</td> <td style="padding: 2px 5px;"><i>length</i></td> </tr> </table>	<i>char</i>		FCHAR	<i>string</i>	WCHAR	☐	LENGTH	<i>length</i>	
<i>char</i>												
FCHAR	<i>string</i>											
WCHAR	☐											
LENGTH	<i>length</i>											
PRED		P-KEY	☐									
			☐									

The feature types WCHAR and FCHAR specify all characters and the first character of a lexical entry, respectively. The feature WCHAR is identical to the STEM of the lexical entry. Next, the LENGTH specifies that an entry has *one* or *more-than-one* character. Finally, the *luk* feature BOUND specifies if an entry is a bound or non-bound form.<sup>2</sup> This is to ensure that one-character  $A_1$  forms of a multi-character word are not used outside of *A*-NOT-*A*, as they are not independent morphemes. The P-KEY feature is identical to the PRED feature so as to block homographs from co-occurring as the *A* elements. An example of such a homograph is 撒 *sā* / *sǎ*, which can mean ‘let go’ and ‘scatter’, respectively. These two will have different PRED values: 撒\_v\_1\_rel and 撒\_v\_2\_rel. Finally, the SPART feature indicates the type of sentence-final particle that can co-occur with the structure.

To provide a clearer idea, the entries in (21) illustrate the bound and non-bound forms of 喜欢, respectively. As they are identical to each other apart from length, they take the same PRED value.

<sup>2</sup>The *luk* constraint consists of three components, such as +, −, and *na* (not-applicable).

(21) a. 
$$\left[ \begin{array}{l} \text{喜} \\ \text{STEM} \quad \square \langle \text{'喜'} \rangle \\ \text{BOUND} \quad + \\ \text{CHAR} \quad \left[ \begin{array}{l} \text{FCHAR} \quad \text{'喜'} \\ \text{WCHAR} \quad \square \\ \text{LENGTH} \quad \text{one} \end{array} \right] \\ \text{PRED} \quad \text{\_喜欢\_v\_rel} \end{array} \right]$$

b. 
$$\left[ \begin{array}{l} \text{喜欢} \\ \text{STEM} \quad \square \langle \text{'喜欢'} \rangle \\ \text{CHAR} \quad \left[ \begin{array}{l} \text{FCHAR} \quad \text{'喜'} \\ \text{WCHAR} \quad \square \\ \text{LENGTH} \quad \text{more-than-one} \end{array} \right] \\ \text{PRED} \quad \text{\_喜欢\_v\_rel} \end{array} \right]$$

The use of the features FCHAR and WCHAR to access the characters of a word is due to a limitation in the present system. It is expected that future iterations will store the characters as a list, and that the characters will be accessed via their indices.

### 2.2.2 Supertype

The present analysis uses the NOT element as the “origin” of the *A*-NOT-*A* structure, which will then select the *A* elements. A generic *A*-NOT-*A* lexical type A-NOT-A-ADV-LEX is defined for this NOT element. As shown in (22), the element of MOD goes for  $A_1$ , the element of COMPS goes for  $A_2$ , and both take +*vjp* (verb, adjective or preposition) as their head type. Both *A* elements are semantically identical, so they take the same SUBJ and COMPS, and share the same ASPECT and P-KEY values.  $A_1$ , being the head of the structure, bears the sentential force (SF) of *ques*.

(22)

$a\text{-not-}a\text{-adv-lex}$ POSTHEAD +	$\left[ \begin{array}{l} +vjp \\ SF \quad \textit{ques} \\ I\text{-KEY} \quad \boxed{1} \\ INDEX \quad \boxed{2} \\ P\text{-KEY} \quad \boxed{3} \\ ASPECT \quad \boxed{4} \\ SUBJ \quad \boxed{5} \\ COMPS \quad \boxed{6} \\ SPART \quad \textit{no-spert} \end{array} \right]$
MOD	$\left\langle \left[ \begin{array}{l} +vjp \\ P\text{-KEY} \quad \boxed{3} \\ ASPECT \quad \boxed{4} \\ SUBJ \quad \boxed{5} \\ COMPS \quad \boxed{6} \\ SPART \quad \textit{no-spert} \\ BOUND \quad - \end{array} \right] \right\rangle$
COMPS	$\left\langle ! \boxed{1} \left[ \begin{array}{l} \textit{focus} \\ IARG2 \quad \boxed{2} \end{array} \right] ! \right\rangle$
ICONS	

The focus meaning is represented via Individual CONStraint (Song, 2014). Its I-KEY feature points to the ICONS element, which indicates that the *A-NOT-A* structure is the focus of the sentence. Thus, IARG2 in ICONS is identical to INDEX of  $A_1$ . In addition,  $A_2$  has the constraint [BOUND –], as bound forms cannot participate as  $A_2$ .

(23) a.

$\text{不}_{\text{polar\_basic}}$ STEM $\langle \text{‘不’} \rangle$ COMPS $\left\langle \left[ \begin{array}{l} ASPECT \quad \textit{non-aspect} \\ LZG \quad \textit{no-lzg} \end{array} \right] \right\rangle$
---

b.

$\text{没}_{\text{polar\_basic}}$ STEM $\langle \text{‘没’} \rangle$ COMPS $\left\langle \left[ \begin{array}{l} ASPECT \quad \textit{imperfective} \\ LZG \quad \textit{guo} \end{array} \right] \right\rangle$
--

As mentioned before, the NOT element can be either 不 *bù* or 没 *méi*, depending on the *A* elements’ aspectual property. As we see in (23), the aspectual properties

of their *A* elements are indicated in their respective COMPS' ASPECT constraints. When the NOT element is *bù*, the *A* elements cannot co-occur with any of the LE-ZHE-GUO markers (*no-lzg*), whereas when the NOT element is *méi*, it can co-occur with *guò*.

As we have seen in Section 1.1, there are a few patterns for the *A*-NOT-*A* structure. With the generic *A*-NOT-*A* lexical type we defined in (22), we create two sub-types for *A*-NOT-*A* and *A'*-NOT-*A*, as shown in Section 2.2.3 and Section 2.2.4.

### 2.2.3 Subtype: A-NOT-A

The sub-type for the basic form is as follows:

$$(24) \left[ \begin{array}{l} a\text{-not-a-basic-adv-lex} \\ \text{MOD} \left\langle \left[ \begin{array}{l} \text{LIGHT} \quad + \\ \text{WCHAR} \quad \boxed{\text{L}} \\ \text{BOUND} \quad - \end{array} \right] \right\rangle \\ \text{COMPS} \left\langle \left[ \begin{array}{l} \text{LIGHT} \quad + \\ \text{WCHAR} \quad \boxed{\text{L}} \end{array} \right] \right\rangle \end{array} \right]$$

The basic form of *A*-NOT-*A* contains two identical *A* elements, as shown in (25):

$$(25) \begin{array}{l} \text{张三} \quad \text{喜欢} \quad \text{不} \quad \text{喜欢} \quad \text{狗} \quad ? \\ \text{Zhāngsān} \quad \text{xǐhuān} \quad \text{bù} \quad \text{xǐhuān} \quad \text{gǒu} \quad ? \\ \text{Zhangsan} \quad \text{like} \quad \text{NOT} \quad \text{like} \quad \text{dog} \quad \text{PU} \end{array}$$

As such, both MOD ( $A_1$ ) and COMPS ( $A_2$ ) have identical WCHAR values. The MOD is constrained to [BOUND -] to block it from parsing the contracted form. Lastly, both MOD and COMPS are constrained with [LIGHT +] such that the *A*-NOT-*A* structure will be treated as a single lexical item instead of as a phrase (cf. Abeillé & Godard (2001)). The constraints presented so far will account for the following ungrammatical sentences:

$$(26) \begin{array}{l} \text{a. *张三} \quad \text{讨厌} \quad \text{不} \quad \text{喜欢} \quad \text{狗} \quad ? \\ \quad \text{Zhāngsān} \quad \text{tǎoyàn} \quad \text{bù} \quad \text{xǐhuān} \quad \text{gǒu} \quad ? \\ \quad \text{Zhangsan} \quad \text{hate} \quad \text{NOT} \quad \text{like} \quad \text{dog} \quad \text{PU} \\ \text{b. *张三} \quad \text{喜} \quad \text{不} \quad \text{喜} \quad \text{狗} \quad ? \\ \quad \text{Zhāngsān} \quad \text{xǐ} \quad \text{bù} \quad \text{xǐ} \quad \text{gǒu} \quad ? \\ \quad \text{Zhangsan} \quad \text{like} \quad \text{NOT} \quad \text{like} \quad \text{dog} \quad \text{PU} \end{array}$$

### 2.2.4 Subtype: A'-NOT-A

The sub-type for the contracted form is as follows:

$$(27) \left[ \begin{array}{l} a\text{-not-a-contracted-adv-lex} \\ \text{MOD} \left\langle \left[ \begin{array}{ll} \text{LIGHT} & + \\ \text{WCHAR} & \boxed{1} \\ \text{BOUND} & + \\ \text{LENGTH} & one \end{array} \right] \right\rangle \\ \text{COMPS} \left\langle \left[ \begin{array}{ll} \text{LIGHT} & + \\ \text{FCHAR} & \boxed{1} \\ \text{LENGTH} & more-than-one \end{array} \right] \right\rangle \end{array} \right]$$

In the A'-NOT-A variant, only the first character of  $A_1$  is reduplicated, as shown in (28):

$$(28) \text{张三 喜 不 喜欢 狗 ?} \\ \text{Zhāngsān xǐ bù xǐhuān gǒu ?} \\ \text{Zhangsan like NOT like dog PU}$$

As such, the LENGTH value of MOD ( $A_1$ ) is constrained to *one*, while its WCHAR — being a single-character word — is identical to the FCHAR of COMPS ( $A_2$ ). As it is a bound form, we constrained it to [BOUND +]. In order to block it from parsing A-NOT-A sentences where the A elements are both single-character words, COMPS is given an additional constraint of *more-than-one* to its LENGTH feature. Finally, as with the basic form, both MOD and COMPS are indicated as [LIGHT +] to treat it as a single lexical item instead of a phrase.

### 2.2.5 AO-NOT-AO

The constraints for this type is shown in (29):

$$(29) \left[ \begin{array}{l} ao\text{-not-ao-adv-lex} \\ \text{MOD} \left\langle \left[ \begin{array}{ll} \text{verb} & \\ \text{LIGHT} & - \\ \text{WCHAR} & \boxed{1} \end{array} \right] \right\rangle \\ \text{COMPS} \left\langle \left[ \begin{array}{ll} \text{verb} & \\ \text{LIGHT} & - \\ \text{WCHAR} & \boxed{1} \end{array} \right] \right\rangle \end{array} \right]$$

The AO-NOT-AO form's A elements are restricted to being verbs, and they are phrases instead of words. As with the basic form, the WCHAR value of the two





$$(31) \left[ \begin{array}{l} \text{INDEX } \boxed{2} \left[ \begin{array}{l} \text{SF } \textit{ques} \\ \text{ASPECT } \textit{non-aspect} \end{array} \right] \\ \\ \left[ \begin{array}{l} \textit{named\_rel} \\ \text{LBL } \boxed{4} \\ \text{CARG } \textit{“张三”} \\ \text{ARG0 } \boxed{6} \end{array} \right], \left[ \begin{array}{l} \textit{proper\_q\_rel} \\ \text{LBL } \boxed{7} \\ \text{ARG0 } \boxed{6} \\ \text{RSTR } \boxed{8} \\ \text{BODY } \boxed{9} \end{array} \right], \\ \\ \text{RELS } \left\langle \left[ \begin{array}{l} \textit{\_喜欢\_v\_I\_rel} \\ \text{LBL } \boxed{10} \\ \text{ARG0 } \boxed{2} \\ \text{ARG1 } \boxed{6} \\ \text{ARG2 } \boxed{11} \end{array} \right], \left[ \begin{array}{l} \textit{\_不\_r\_rel} \\ \text{LBL } \boxed{1} \\ \text{ARG0 } \boxed{12} \\ \text{ARG1 } \boxed{13} \end{array} \right], \left[ \begin{array}{l} \textit{\_喜欢\_v\_I\_rel} \\ \text{LBL } \boxed{14} \\ \text{ARG0 } \boxed{2} \\ \text{ARG1 } \boxed{6} \\ \text{ARG2 } \boxed{11} \end{array} \right] \right\rangle \\ \\ \left[ \begin{array}{l} \textit{\_狗\_n\_I\_rel} \\ \text{LBL } \boxed{15} \\ \text{ARG0 } \boxed{11} \end{array} \right], \left[ \begin{array}{l} \textit{exist\_q\_rel} \\ \text{LBL } \boxed{16} \\ \text{ARG0 } \boxed{11} \\ \text{RSTR } \boxed{17} \\ \text{BODY } \boxed{18} \end{array} \right] \\ \\ \text{HCONS } \left\langle \left[ \begin{array}{l} \textit{qeq} \\ \text{HARG } \boxed{8} \\ \text{LARG } \boxed{4} \end{array} \right], \left[ \begin{array}{l} \textit{qeq} \\ \text{HARG } \boxed{13} \\ \text{LARG } \boxed{10} \end{array} \right], \left[ \begin{array}{l} \textit{qeq} \\ \text{HARG } \boxed{17} \\ \text{LARG } \boxed{15} \end{array} \right] \right\rangle \\ \\ \text{ICONS } \left\langle \left[ \begin{array}{l} \textit{focus} \\ \text{IARG1 } \boxed{2} \\ \text{IARG2 } \boxed{2} \end{array} \right] \right\rangle \end{array} \right]$$

The semantic relations are indicated in the MRS.  $A_1$  and  $A_2$  are given the same indexes: ARG0 for the verb itself, ARG1 for the subject *Zhāngsān*, and ARG2 for the object *gǒu*. This means that they share the same argument structure. The second element in the HCONS list is responsible for the scope of negative operator *bì*: HARG is co-indexed with the ARG1 of the scopal modifier (i.e.  $\boxed{13}$ ), and LARG is co-indexed with the label of  $A_1$  (i.e.  $\boxed{10}$ ). The element in the ICONS list is specified as *focus*, and the values of IARG1 and IARG2 are both co-indexed with the verb’s INDEX. This means that the *A-NOT-A* structure is associated with focus within the clause. Finally, the semantic head  $\boxed{2}$  has [SF *ques*], which indicates that the sentence is interrogative.

### 2.2.7 Limitations

This approach demonstrates the underlying syntactic rules of the *A-NOT-A* structure. However, in doing so, it offers a semantic analysis that contains two predicates — each of the *A* elements are treated as a separate predicate, with constraints to make them identical. This was found to be an unsatisfactory semantic analysis of the structure, which should instead have only a single predicate since it is not a true disjunctive. This analysis, however, does not permit the presence of only one

predicate.

Secondly, the present approach does not block the modification of the *A*-NOT-*A* structure itself, and so cause the over-generation of sentences such as 张三 [ 很 [喜欢狗 不喜欢] ] 狗? , where the *A*-NOT-*A* structure is modified by the degree adverb 很 ‘very’.

Thirdly, like the first approach, this approach does not have the ability to restrict the *O* (object) elements in the *AO*-NOT-*AO* structure to be identical due to limitations of the system, and would thus cause over-generation of structures like 张三喜欢狗<sub>O1</sub> 不喜欢猫<sub>O2</sub>? .

### 3 Conclusion

In this paper, we provided two HPSG accounts for the *A*-NOT-*A* structure in Mandarin Chinese, approaching it syntactically as well as morphologically, and looked at their respective strengths and weaknesses. Overall, the morphological approach is preferred, as it provides more accurate semantics, and we are better able to restrict the modification of the entire structure, even if the formation of the *A*-NOT-*A* structure is not as transparently illustrated with this approach.

Both methods are unable to reliably account for the *AO*-NOT-*AO* structure as it remains non-trivial to constrain *O* to be identical, or to account for the arbitrariness in length that *O* can be. It is hoped that future work will be able to cover this particular pattern.

#### Acknowledgements

We would like to thank participants of the Linguistic Analysis Design (LAD) sessions (DELPH-IN ’15 summit), attendees of the HPSG 2015 conference as well as all others who had given their useful input.

This research was supported in part by the MOE Tier 2 grant *That’s what you meant: a Rich Representation for Manipulation of Meaning* (MOE ARC41/13).

#### References

- Abeillé, Anne & Daniele Godard. 2001. A Class of “Lite” Adverbs in French. In Joaquim Camps & Caroline R. Wiltshire (eds.), *Romance syntax, semantics and l2 acquisition: Selected papers from the 30th linguistic symposium on romance languages, gainesville, florida, february 2000*, 9–26. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Copestake, Ann, Dan Flickinger, Carl Pollard & Ivan A. Sag. 2005. Minimal Recursion Semantics: An Introduction. *Research on Language & Computation* 3(4). 281–332.

- Fan, Zhenzhen, Sanghoun Song & Francis Bond. 2015. Building Zhong [], a Chinese HPSG meta-grammar. In *Proceedings of the 22nd International Conference on Head-Driven Phrase Structure Grammar (HPSG 2015)*, 97–110.
- Liing, Woan-Jen. 2014. *How to Ask Questions in Mandarin Chinese*: City University of New York dissertation.
- McCawley, James D. 1994. Remarks on the Syntax of Mandarin Yes-No Questions. *Journal of East Asian Linguistics* 3(2). 179–194.
- Pollard, Carl & Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago, IL: The University of Chicago Press.
- Song, Sanghoun. 2014. *A Grammar Library for Information Structure*: University of Washington dissertation.
- Tseng, Wen-Hsin Karen. 2009. A Post-syntactic Approach to the A-not-A Questions. *UST Working Papers in Linguistics, Graduate Institute of Linguistics* 5(National Tsing Hua University).