

A-not-A Questions in Mandarin Chinese: An HPSG Account

Wenjie WANG

April 20, 2015

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 3 |
| 1.1 | Brief Look at A-not-A | 3 |
| 1.1.1 | Reduplication in A-NOT-A | 4 |
| 1.1.2 | “A” of A-NOT-A | 5 |
| 2 | Existing Accounts | 8 |
| 2.1 | Restrictions on A-not-A | 11 |
| 2.1.1 | On the Modifiability of A-NOT-A | 12 |
| 2.1.2 | Co-occurrence with LE-ZHE-GUO | 12 |
| 2.1.3 | Sentence-Final Particles | 13 |
| 2.1.4 | Quantified Subjects | 13 |
| 2.2 | Other types of questions in Mandarin Chinese | 14 |
| 2.2.1 | 还是 <i>hàishì</i> Disjunctive | 14 |
| 2.2.2 | MA-questions | 16 |
| 2.3 | Existing HPSG-based accounts | 20 |
| 3 | Purpose and Scope | 21 |

| | | |
|----------|--|-----------|
| 4 | HPSG Account | 23 |
| 4.1 | Brief Introduction to HPSG and MRS | 23 |
| 4.2 | Groundwork | 23 |
| 4.2.1 | A-not-A as coordinate structure? | 23 |
| 4.2.2 | Character/Syllable List | 24 |
| 4.2.3 | The Headedness of A-not-A | 25 |
| 4.3 | Basic A-not-A / Basic Form | 26 |
| 4.4 | A-not-A contracted / Contracted Form | 30 |
| 4.4.1 | Single-character entries for bound forms | 31 |
| 4.5 | VP-not-VP | 33 |
| 4.5.1 | AB-not-AB | 33 |
| 4.5.2 | More complex VP-NOT-VP structures | 35 |
| 4.6 | Other Variants | 36 |
| 4.6.1 | VV-compounds | 36 |
| 4.6.2 | VP-neg | 36 |
| 4.7 | The “monolithic” approach | 38 |
| 5 | Implementation | 40 |
| 5.1 | The ZHONG [[]] grammar | 40 |
| 5.2 | Preliminary Implementation | 40 |
| 5.3 | Test-suite | 42 |
| 5.3.1 | Sentences | 42 |
| 5.3.2 | Coverage | 42 |
| 5.4 | Generation | 43 |
| 6 | Limitations and Ongoing Investigation | 44 |
| 7 | Conclusion | 44 |

Abstract

In this paper, I look at A-NOT-A questions in Mandarin Chinese, and create an account based on the framework of Head-driven Phrase Structure Grammar (HPSG) and Minimal Recursion Semantics (MRS). While the A-NOT-A structure has seen extensive research in the past decades, they have largely been movement-based. Thus, this paper attempts to provide a non-movement but constraint-based HPSG/MRS account, which has thus far not been performed. Secondly, I have also begun initial implementation of said account into the HPSG-based ZHONG [1] computational grammar for Chinese. While the basic forms of the A-NOT-A structure were accounted for and implemented, it was found that limitations in the system and formalisation prevented VP-NOT-VP questions from being successfully implemented.

1 Introduction

In this paper, I will be looking at the A-NOT-A question structure of Mandarin Chinese and provide an account of this phenomenon based on the frameworks of Head-driven Phrase Structure Grammar (HPSG) and Minimal Recursive Semantics (MRS) (Pollard and Sag, 1994; Copestake et al., 2005). Simultaneously, this account will be implemented into the HPSG/MRS-based ZHONG [1] computational Chinese Grammar developed by Nanyang Technological University (NTU).

This paper is divided into six (6) main sections. In the current section (§1), I will run through the basic properties of A-NOT-A questions. Next, in §2, I look more into the existing accounts and their handling of the A-NOT-A structure, as well as two other question types in Mandarin Chinese and how they compare to A-NOT-A questions. §3 gives a general idea of the scope of this paper’s analysis and the intended coverage. §4 will provide the HPSG account and §5 will touch briefly on the implementation of the HPSG account into the ZHONG [1] grammar. Finally, §6 will mention the limitations of the current implementation, as well as the work that remains to be done.

1.1 Brief Look at A-not-A

The A-NOT-A structure is one of the methods available in Mandarin Chinese for posing alternative or yes/no questions. It is thus named because the structure is composed of an element A being immediately followed by a copy of itself that is negated (Li and Thompson, 1989). For ease of reference, the two copies of A are labelled as A_1 and A_2 , according to their linear (left-to-right) order. In Mandarin Chinese, NOT can be either of the negators 不 *bù* or 没 *méi*, the choice of which depends mainly on the aspect of the A element (stative or imperfective, and bound events or perfective, respectively), similar to when they are used as actual negators.¹

¹In this paper, I shall use NOT to refer to either of these where specificity is not required.

1.1.1 Reduplication in A-NOT-A

In its most basic form, A is reduplicated in full; however, it can also be partially reduplicated. The following examples illustrate some of these possibilities:

- (1) a. 张三 喜欢 狗 不 喜欢 狗?
 zhāngsān xǐhuān gǒu bù xǐhuān gǒu ?
 Zhangsan like dog NOT like dog?
 ‘Does Zhangsan like dogs?’
- b. 张三 喜欢 不 喜欢 狗?
 zhāngsān xǐhuān bù xǐhuān gǒu ?
 Zhangsan like NOT like dog?
 ‘Does Zhangsan like dogs?’
- c. 张三 喜 不 喜欢 狗?
 zhāngsān xǐ bù xǐhuān gǒu ?
 Zhangsan xi- NOT like dog?
 ‘Does Zhangsan like dogs?’
- d. * 张三 喜欢 不 喜 狗?
 zhāngsān xǐ bù xǐ gǒu ?
 Zhangsan like NOT xi- dog?
 ‘Does Zhangsan like dogs??’
- e. * 张三 喜 不 喜 狗?
 zhāngsān xǐ bù xǐ gǒu ?
 Zhangsan xi- NOT xi- dog?
 ‘Does Zhangsan like dogs?’
- f. 张三 喜欢 狗 不 喜欢?
 zhāngsān xǐhuān gǒu bù xǐhuān ?
 Zhangsan like dog NOT like?
 ‘Does Zhangsan like dogs?’
- g. * 张三 喜欢 狗 不 喜?
 zhāngsān xǐhuān gǒu bù xǐ ?
 Zhangsan like dog NOT xi-?
 ‘Does Zhangsan like dogs?’
- h. * 张三 喜 狗 不 喜欢?
 zhāngsān xǐ gǒu bù xǐhuān ?
 Zhangsan xi- dog NOT like?
 ‘Does Zhangsan like dogs?’

As shown in the examples above, partial reduplication can result in either the verb phrase being reproduced without its complement, or the verb phrase being reproduced without its complement *and* with only its first character/syllable. The latter form of reduplication can be seen as a violation of lexical integrity, as the single-character forms are bound forms and cannot exist independently (see §4.2.2 and §4.4.1 for more).

Partial reduplication is not equally applicable to both A_1 and A_2 , as seen in (1g). For A_2 , only one type of partial reduplication — deletion of complement — is permitted.² Moreover, (1h) shows that A_1 cannot go through single-character reduplication if its complement is also copied, indicating that leaving out constituents from the reduplication process can only occur linearly up to the left-most element.³

1.1.2 “A” of A-NOT-A

Apart from being verbs or verb phrases as shown previously, the A elements can also be adjectives, prepositions (or coverbs) and modals, as long as they can function as syntactic heads of a predicate (Tseng, 2009). The same possibilities for full or partial reduplication is extended to these other types as well.

(2) a. **Adjective**

张三 高 不 高?
 zhāngsān gāo bù gāo?
 Zhangsan tall NOT tall
 ‘Is Zhangsan tall?’

b. **Adjective (optional partial reduplication)**

张三 健(康) 不 健康?
 zhāngsān jiàn(kāng) bù jiànkāng?
 Zhangsan healthy NOT healthy
 ‘Is Zhangsan healthy?’

c. **Preposition/Co-verb**

张三 在 不 在家?
 zhāngsān zài bù zài jiā?
 Zhangsan at NOT at home
 ‘Is Zhangsan at home?’

²A type of reduplication unique to A_2 is one where A_2 is completely absent, resembling a form of ellipsis. This particular variant is known as the VP-NEG or A-NOT pattern. I look at this type briefly in §4.6.2.

³An alternate procedure could be that there is deletion/elision occurring after a *full* reduplication to produce the partially-reduplicated forms, with deletion being required to occur linearly towards the left-most element.

d. **Modal Verbs**

张三 要 不 要 吃 苹果?
 zhāngsān yào bù yào chī píngguǒ?
 Zhangsan want NOT want eat apple
 Does Zhangsan want to eat apples?

e. **Frequency Adverbs**

张三 常 不 常 迟到?
 zhāngsān cháng bù cháng chídào?
 Zhangsan often NOT often late-arrive
 ‘Does Zhangsan often come late?’

f. **Degree Adverbs [Tseng (2009), Eg 1]**

* 张三 很 不 很 喜欢 李四?
 *zhāngsān hěn bù hěn xǐhuān Lisi?
 Zhangsan very NOT very like Lisi
 ‘Does Zhangsan like Lisi very much?’

Not all adverbs can participate as the A element. For instance, as seen in (2f), degree adverbs such as 很 *hěn* ‘very’ cannot be the A element. Tseng (2009) provided a reason, suggesting that the A-NOT-A operator applies to the most immediate morpho-syntactic word (MWd) under its command, and this MWd needs to be a syntactic head. Since *hěn* is not considered such a head, the A-NOT-A operator is prevented from attaching to and operating on it.

Another constraint on adverbs is that their reduplicated forms are not permitted as A elements. The two sentences in (3) show the contrast:

- (3) a. 张三 常 不 常 迟到?
 zhāngsān cháng bù cháng chídào?
 Zhangsan often NOT often late-come
 ‘Does Zhangsan often come late?’
- b. * 张三 常常 不 常常 迟到?
 zhāngsān chángchang bù chángchang chídào?
 Zhangsan often NOT often late-come
 ‘Does Zhangsan often come late?’

Liu (2010) suggests that this restriction abides by Dai (1990)’s Disyllabification Rhythm Rule, which states that a phonological rule based on metrical requirements determines the reduplication process of the A-NOT-A structure.

A special type of A-NOT-A question involves the use of the copula 是 *shì* as the A element:

(4) 张三 是 不 是 喜欢 狗?

zhāngsān shì bù shì xǐhuān gǒu

Zhangsan BE not BE like dog

‘Is it (true) that Zhangsan likes dogs?’

At first glance, it resembles the result of a movement of the question tag 是不是? *shìbùshì* “isn’t it?”.

(5) 张三 喜欢 狗, 是 不 是?

zhāngsān xǐhuān gǒu , shì bù shì

Zhangsan like dog, BE not BE

‘Zhangsan likes dogs, doesn’t he?’

Due to *shì* frequently being used as a translation of English “to be”, such a variant has been termed B-NOT-B questions Liu (2010). While interesting, I will not deal with this variant in this paper.

2 Existing Accounts

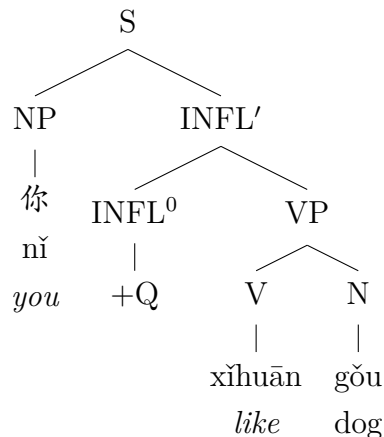
The A-NOT-A structure has seen extensive research in the past decades. Among the early accounts of this structure is Huang (1991), who proposes modular accounts for the various patterns and forms of A-NOT-A questions. He suggests that A-NOT-A questions are realised as a constituent that contains an A-NOT-A question feature [+Q]. Using example (6), this feature is realised by re-duplicating the constituent that immediately follows it (the VP), and then inserting the appropriate negator between the copies:

(6) [Adapted from Huang (1991), Example 44]

Example Declarative Sentence:

你 喜欢 狗
 nǐ xǐhuān gǒu
 you like dog

‘You like dogs’



As we have previously seen in §1, this targeted constituent can be partially or fully reduplicated. Therefore, [+Q] can reduplicate the entire VP, or it can reduplicate just the V or even its first character. The addition of this [+Q] feature, he argues, is similar to what is used to form the WH-questions, and makes A-NOT-AB syntactically on-par with WH-questions.

Huang’s account, however, was not entirely complete. In particular, McCawley (1994) remarks on the A-NOT-A structure, and pin-pointed Huang’s vague account of the NOT element, stating Huang’s failure to state the presence of there being two types — 不 *bù* and 没 *méi* — each of which are used under different circumstances, the former being used for statives and imperfectives, while the latter is used to negate bound events or perfectives.

Huang et al. (2009) modifies to Huang (1991)'s initial appraisal and analysis, by stating that the NOT-A part of the structure is generated based on morphologically-motivated reasons. The NOT element is no longer inserted as an independent element, but is instead borne from the appropriate negation of the second A element (A_2), and that the aspectual properties of A will determine the form of the negator. As such, a compound verb such as 看懂 *kàndǒng* "read.understand" — which is composed of an action ("read") plus a result ("understand") — must take *méi* as the negator.

Liing (2014) makes similar criticisms of the analyses in Huang (1991) as well as Huang et al. (2009). She said that although the latter paper had extended its coverage of the NOT element, it was still inadequate. Firstly, it failed to take into account that some words which are ungrammatical when negated can nevertheless be negated when used in the A-NOT-A structure:

(7) [Adapted from Liing (2014), examples 12a and 12a']

- a. *他 不 无情。
 tā bù wúqíng
 he NOT non.empathic
 'He is not without empathy'
- b. 他 无情 不 无情?
 tā wúqíng bù wúqíng
 he non.empathic NOT non.empathic
 'Does he lack empathy?'

Also, Liing argues that morphological rules will not have accounted for situations where the same character/word is "morphologically unvarnished", and can be negated by either negator:

- (8) a. 去 不 去
 qù bù qù
 go NOT go
 'Are you going?'
- b. 去 没 去
 qù méi qù
 go NOT go
 'Have you gone/been to (somewhere)'

Ernst (1994) looked at the co-occurrence of adjuncts with A-NOT-A questions, and highlighted that many adjuncts such as 一定 *yídìng* "definitely" and 乱 *luàn* "chaotically" cannot be used with or modify the A-NOT-A structure (not in fact other question types), but certain adjuncts such as 今天 *jīntiān* "today" can be used. It was proposed that the latter types of adjuncts behave in an

argument-like manner, making them different in scope on the logical form and thus able to modify questions.

A particularly distinct account is Gasde (2001), who approached the structure with the premise of Chinese being originally SOV. He argued that the underlying deep structure (DS) of an A-NOT-A structure — or any existing SVO sentence, for that matter — is in fact SOV, and is realised as thus:

(9) Adapted from Gasde (2001), Examples 4.11, 4.12

- a. [你 [电影 看不看]]
 [ni [diànying kàn-bù-kàn]]
 you movie watch-not-watch
- b. [你 [电影 看不]]
 [ni [diànying kàn-bù]]
 you movie watch-not

The constituent *kàn-bù-kàn* is made up two components: the stem *kàn* and what he termed the “semi-suffix”, which is *bù-kàn* in (9a) and *bù* in (9b). These semi-suffixes, he suggests, “can be ‘taken along’ or ‘left behind’” when generating the final structure. As such, the A-NOT-A structure can be realised by the movement (or not) of the components, with the corresponding traces still in place. This, as he claims, properly accounts for sub-patterns of A-NOT-A, such as A-NOT-AB, AB-NOT-A and even A(B)-NOT.

(10) From Gasde (2001), Examples 4.1’, 4.2’, 4.9’

- a. [你 看不看_i [电影 t₁]]
 [ni kàn-bù-kàn_i [diànying t_i]]
 you watch-not-watch movie
- b. [你 看₁ [电影 t_i -不看]]
 [ni kàn_i [diànying t_i -bù-kàn]]
 you watch movie -not-watch
- c. [你 看₁ [电影 t_i -不]]
 [ni kàn_i [diànying t_i -bù]]
 you watch movie -not

Huang (2008) explores a possible variant of A-NOT-A questions known as VP-neg questions. In such a variant, the right A is completely elided, with NOT marking the end of the sentence, thus making a type of A-NOT structure:

- (11) 张三 喜欢 狗 不?
 zhāngsān xǐhuān gǒu bu?
 Zhangsan like dogs NOT
 ‘Does Zhangsan like dogs or not?’
- (12) 张三 睡觉 了 没(有)?
 zhāngsān shuìjiào -le mei(you)?
 Zhangsan sleep-PERF NOT
 ‘Has Zhangsan slept or not?’

This structure has also been called the VP-NEG QUESTION structure or NEGATIVE PARTICLE QUESTION structure, and Huang has been deemed it “controversial” as there has been no consensus among linguists on whether the VP-NEG form is a reduced/derived form of A-NOT-A . (See §4.6.2 for more on the VP-NEG structure)

In proposing the non-context-free nature of Mandarin Chinese, with A-NOT-A questions as a point of investigation, Radzinski (1990) raised the question of whether the reduplication (or copying) process was a semantic or syntactic process. In the former, it has been suggested that the NOT-A part is derived by producing a semantically-negated form of the A element. This semantic approach appears to be attractive because it allows a perfectly duplicated copy of A_2 , since A_2 is simply a semantically negative form of A_1 and is otherwise identical. Radzinski countered this by saying that were A-NOT-A truly the juxtaposition of two semantically-opposite forms, then the juxtaposition of two antonyms would likewise be grammatical and valid in the A-NOT-A forms. This, however, is not possible and Radzinski therefore concluded that the copying process is a syntactic process instead.

Overall, the existing accounts had relied heavily on the mechanisms of movement and transformations in order to account for the A-NOT-A structure. These mechanisms, however, are not subscribed to in the HPSG formalism, which instead works on a constraint-based approach. As such, while the main observations of the A-NOT-A structure in the existing accounts can be used as reference (including observations such as occurrences, grammaticality, restrictions, etc), the actual mechanisms and analyses will have to be re-thought for my account.

2.1 Restrictions on A-not-A

Apart from the basic features investigated in the previous sections, there are also addition constraints and restrictions on the occurrence of the A-NOT-A structure. In this section I shall look briefly at a few of these restrictions.

2.1.1 On the Modifiability of A-NOT-A

To recall, Ernst (1994) and Law (2006) had noted the difference between different types of adjuncts/adverbs in terms of their ability or inability to modify A-NOT-A structures, such that adverbs like *yídìng* “definitely” are forbidden while those like *jīntiān* “today” are permitted.

The restrictions discussed in those studies, however, occur on the A-NOT-A *phrase* level. It should be noted that modification of the A elements themselves is also not allowed. For example, McCawley (1994) as well as Liing (2014) noted the restrictions on negation for A elements which already has a negative element as its first character/syllable. As stated, this was in part, this has been attributed to the already-present negative element in the A-NOT-A structure.

In fact, it could be extended that any modification to the A elements themselves does not appear to be permitted. For example, McCawley (1994) noted that the predicate phrases modified by elements such as the degree adverb 很 *hén* “very” and adverbs such as 只 *zhǐ* “only” and 也 *ye* “also” are not permitted, as illustrated by the following examples:

- (13) a. *你 很 喜欢 狗 不 很 喜欢 狗?
 nǐ hén xǐhuān gǒu bù hén xǐhuān gǒu
 you very like dog NOT very like dog
- b. *你 只 不 只 喜欢 狗?
 nǐ zhǐ bù zhǐ xǐhuān gǒu
 you only NOT only like dog
- c. *你 也 不 也 喜欢 狗?
 nǐ yě bù yě xǐhuān gǒu
 you also NOT also like dog

2.1.2 Co-occurrence with LE-ZHE-GUO

This restriction on modification extends to the post-modifying elements 着 *zhè*, 了 *le* and 过 *guò*. These elements, which I’ll refer to collectively as the LE-ZHE-GUO markers, serve as aspectual markers. *Zhè* is the durative marker, *le* indicates the perfective and *guò* indicates the experiential aspect.

As a whole, aspectual markers are not necessarily permitted with the *bù*-form of A-NOT-A. The marker *guò*, however, can co-occur if used with the *méi*-form A-MÉI-A structure. This restriction is identical to that of an ordinary negative sentence involving *bù* or *mèi*.

As with the ordinary negation, *bù* modifies states and imperfectives, while *mei* modifies bound events and perfectives.

2.1.3 Sentence-Final Particles

Not all sentence-final particles (SFPs) can be used with A-NOT-A questions. For instance, the SFP variant of 了 *le* cannot be used with A-NOT-A questions because it can only take a declarative or proposition under its scope, and cannot have scope over a question. Likewise, the question SFP 吗 *mā* cannot take scope over A-NOT-A because it is already a question.

2.1.4 Quantified Subjects

A particular property of A-NOT-A questions noticed by Wu (1997) is the restriction on quantified noun phrases (NP) in subject position, as the following examples show:

(14) **Adapted from Wu (1997), example 3**

- a. ? 每 个 人 都 跑 不 跑?
 měi gè rén dōu pǎo bù pǎo?
 every CL person DOU run NOT run
 ‘Does everyone run or not run?’
- b. * 有 人 跑 不 跑?
 yǒu-rén pǎo bù pǎo??
 someone run NOT run?
 ‘Does someone run or not run?’
- c. * 没 有 人 跑 不 跑?
 méi-yǒu-rén pǎo bù pǎo?
 nobody run NOT run
 ‘Does nobody run or not run?’

The first instance is highly “deviant” (which Wu reasoned is due to the ambiguity that arises from *meigeren* being either a quantifier or a group-denoting NP), while the remaining two are completely unacceptable.

This, as a result, makes it different from typical yes/no questions. Wu establishes that an A-NOT-A question is made up of two “cells” which are mutually exclusive but jointly exhaustive, which means they cover all sets of possibilities — there is no third option. While this gels with referential NPs as the subject, it does not with quantified NPs because such NPs consist of many members, each of whose truth value pertaining to the A-NOT-A question might differ. Thus, their choices will become non-mutually-exclusive and non-jointly-exhaustive, which should not be possible with A-NOT-A questions. Therefore, such questions are semantically anomalous.

2.2 Other types of questions in Mandarin Chinese

In addition to the A-NOT-A structure, there are other ways to ask questions in Mandarin Chinese. This section takes a brief look at two such question types — the 还是 *háishì* disjunctive and the 吗 *ma*-question — and compares them to the A-NOT-A structure.

2.2.1 还是 *háishì* Disjunctive

The 还是 *háishì* “or” disjunctive typically involves distinct choices, such that the general structure can be A HÁISHÌ B ([A or B]). If more than two choices are involved, then *háishì* “or” is typically used only to conjoin the last two conjuncts, with the rest separated conventionally by commas (or the “list-commas” used in Chinese writing), similar to how *or* or *and* is used in English. The disjunctive can be exclusive and non-exclusive, and the responses can be either one of the choices, *both* or even *neither*.

A variant of this is the A HÁISHÌ NOT-A ([A] or [not A]) form, where the second choice is simply the negated form of the first. In this form, the disjunctive is strictly exclusive and a choice has to be made between the two. This is because both disjunctive propositions — p or $\neg p$ — cover the entire set of possibilities, with no “third choice”. Among the various question types, this particular variant A HÁISHÌ NOT-A disjunctive question appears to be the closest to A-NOT-A questions.

A key difference between the two question structures is their permitted order of A or not-a. As McCawley (1994) pointed out, while the *haishi* disjunctive has a preference for the positive choice to come first, it still permits the negative choice to come before, even if there is a shift in bias. In the A-NOT-A structure, however, such a reversal is not allowed:

- (15) a. 你 不 去 还是 去?
 nǐ bù qù háishì qù?
 you NOT go HAISHI go?
- b. *你 不 去 去?
 nǐ bù qù qù?
 you NOT go go?
- c. [From McCawley (1994), Example 13a]
 你 不 喜欢 他 还是 喜欢 他?
 nǐ bù xǐhuān tā háishì xǐhuān tā?
 you NOT like him HAISHI like him?
- d. [From McCawley (1994), Example 13a']
 *你 不 喜欢 他 喜欢 他?
 nǐ bù xǐhuān tā xǐhuān tā?
 you NOT like him like him?

Although the order of the disjunctive elements in the HÁISHÌ disjunctive can be reversed, it should be noted again that the positive disjunct being first is still the preferred, conventional order. As such, Liing (2014) believes that the A-NOT-A construction could in fact be a grammaticalisation of such a preference, which explains why the reverse (NOT-A A) is not permitted.

A second difference is the acceptability of modification of the A elements in the two question types. For example, the HÁISHÌ disjunctive can have its A elements modified by a degree adverb like 很 *hén* “very”, but this cannot be done for A-NOT-A questions, as we have seen in §2.1.1 (McCawley, 1994):

- (16) a. 你 很 喜欢 狗 还是 不 很 喜欢 狗?
 nǐ hén xǐhuān gǒu hái shì bù hén xǐhuān gǒu
 you very like dog HAISHI NOT very like dog
 ‘Do you like dogs very much or not like dogs very much?’
- b. *你 很 喜欢 狗 不 很 喜欢 狗?
 nǐ hén xǐhuān gǒu bù hén xǐhuān gǒu
 you very like dog NOT very like dog

Finally, the A-NOT-A structure exhibits the effects of island constraints, whereas the HÁISHÌ disjunctive does not. The HÁISHÌ disjunctive can also be used within a relative clause, while A-NOT-A cannot.

- (17) a. 张三 去 还是 不 去 比较 好?
 zhāngsān qù hái shì bù qù bǐjiào hǎo?
 zhangsan go HAISHI not go comparatively better?
 ‘Is it better that Zhangsan go (there) or not go (there)?’
- b. *张三 去 不 去 比较 好?
 zhāngsān qù bù qù bǐjiào hǎo?
 zhangsan go not go comparatively better?
 ‘Is it better that Zhangsan go (there) or not go (there)?’

(18) (From Huang (1991), Examples 34 and 35)

- a. 你 喜欢 [认识 你 还是 不 认识 你] 的 人?
 nǐ xǐhuān [rènshì nǐ hái shì bù rènshì nǐ] de rén?
 you like know you HAISHI not know you DE person
 ‘Do you like people who know you or people who do not know you?’
- b. *你 喜欢 [认识 你 不 认识 你] 的 人?
 nǐ xǐhuān [rènshì nǐ bù rènshì nǐ] de rén?
 you like know you NOT know you DE person
 ‘Do you like people who know you or people who do not know you?’

2.2.2 MA-questions

These questions are marked by the sentence-final particle (SFP) 吗 *mā*. Hereafter I shall refer to these as MA-questions.

It has been suggested by some (such as Ernst (1994) and Law (2001)) that MA-questions and A-NOT-A questions are semantically equivalent and thus interchangeable, such that a MA-question can be paraphrased into an A-NOT-A question (and vice versa):

- (19) a. 张三 喜欢 不 喜欢 李四?
 zhāngsān xǐhuān bù xǐhuān Lisi
 Zhangsan like NOT like Lisi
 ‘Does Zhangsan like Lisi?’
- b. 张三 喜欢 李四 吗?
 zhāngsān xǐhuān Lisi ma?
 Zhangsan like Lisi MA
 ‘Does Zhangsan like Lisi?’

However, there are some differences between the two structures in at least three aspects which suggest they are not, in fact, allostructures of each other.

Firstly, in terms of semantics, scope ambiguity can arise when the universal quantifier 都 *dōu* is used with MA-questions, but not with A-NOT-A questions, and the meaning of their responses can differ, particularly when a negative response is given (McCawley, 1994). Taking the two questions in examples (20a) and (20b) below, I tabulate their possible responses (and corresponding meaning) in (21), based on McCawley (1994).

(20) [Adapted from McCawley (1994), Examples 21 & 23]

a. **A-NOT-A question:**

他们 都 喜欢 不 喜欢 开车?
 Tamen dou xǐhuan bu xihuan kaiche?
 they DOU like NOT like drive
 ‘Do they all like to drive?’

b. **MA-question**

他们 都 喜欢 开车 吗?
 Tamen dou xihuan kaiche ma?
 they DOU like drive MA
 ‘Do they all like to drive? / Do all of them like to drive?’

(21) [Adapted from McCawley (1994), Examples 21 & 23]

| Response / Question | A-not-A | MA-question |
|---------------------------------|-----------------------------|-------------------------------|
| 喜欢 <i>xǐhuān</i> [like] | They do (all like to drive) | |
| 不喜欢 <i>bù xǐhuān</i> [not like] | None of them likes to drive | |
| 不 <i>bù</i> [“No”] | None of them likes to drive | Not all of them like to drive |

As can be seen from the above, the negative responses have the same meaning (that none of the people likes to drive) for the A-NOT-A question. This is because the negator *bù* is in the scope of *dōu* in both answers ([*dōu* [*bù*]] and [*dōu* [*bù xǐhuān*]]), and there is no ambiguity.

On the other hand, the two negative responses can differ in meaning for the MA-question, because of the ambiguous scope of *dōu*: with a response of “*bù xǐhuān*”, the negator *bù* is under the scope of *dōu* ([*dōu* [*bù xǐhuān*]]), just like it would for A-NOT-A . However, with a response of “*bù*”, it is *bù* that takes scope over everything else: [*bù* [*dōu xǐhuān*]], which could be roughly translated as “not all (of them) like” (versus “all (of them do) not like”).

Secondly, while not mentioned in any previous accounts that I’m aware of, I suggest a difference in information structure between the two types of questions, and that focus or topic of a MA-question is not immediately known on the syntactic level if the question is a plain, unmarked sentence such as (22):

(22) 张三 喜欢 李四 吗?
 zhāngsān xǐhuān Lisi ma?
 Zhangsan like Lisi MA
 ‘Does Zhangsan like Lisi?’

Using the sentence in (22), we can have at least three different interpretations, if based solely on the syntax:

(23) a. [张三]_F 喜欢 李四 吗?
 [zhāngsān]_F xǐhuān Lisi ma?
 Zhangsan like Lisi MA
 ‘Is it Zhangsan (and not anyone else) who likes Lisi?’

b. 张三 [喜欢]_F 李四 吗?
 zhāngsān [xǐhuān]_F Lisi ma?
 Zhangsan like Lisi MA
 ‘Is it that Zhangsan likes Lisi?’

- c. 张三 喜欢 [李四]_F 吗?
 zhāngsān xǐhuān [Lìsì]_F ma?
 Zhangsan like Lisi MA
 ‘Is it Lisi whom Zhangsan likes?’

It could be argued that the verb-focused interpretation — that is, focus being on *xǐhuān* — could be the default one, likely arising from it being the head of the sentence, but it does not exclude the possibility of the focus being different without any change in syntax. This is likely because the scope of *mā* is the entire sentence and not specific constituents. Should focus be required, the asker will need to either rely on prosodic clues such as intonation or stress, or employ syntactic tools such as topic markers like 是 *shì*, in order to bring the elements into focus:

- (24) a. 是 张三 喜欢 李四 吗?
Shì zhāngsān xǐhuān Lìsì ma?
 SHI Zhangsan like Lisi MA
 ‘Is it Zhangsan (and not anyone else) who likes Lisi?’
- b. 张三 是 喜欢 李四 吗?
 zhāngsān **shì** xǐhuān Lìsì ma?
 Zhangsan SHI like Lisi MA
 ‘Is it that Zhangsan likes Lisi?’
- c. 张三 喜欢 的 是 李四 吗?
 zhāngsān xǐhuān **dè shì** Lìsì ma?
 Zhangsan like DE SHI Lisi MA
 ‘Is it Lisi whom Zhangsan likes?’

In A-NOT-A questions, however, there is no ambiguity in focus — the A-NOT-A structure is always the constituent in focus — and as such the answer would therefore be directly in response to the A-NOT-A question.

In a multi-verb or modal+verb construction, the A-NOT-A structure cannot be applied to the second (or non-first) verb, since the A-NOT-A structure, when used as an interrogative, is generally not permitted to be part of an embedded clause. As such, focus cannot be shifted to other verbs by means of applying A-NOT-A mechanisms to it.

- (25) a. 张三 喜欢 不 喜欢 看书?
 zhāngsān xǐhuān bù xǐhuān kànshū?
 Zhangsan like NOT like read?
 ‘Does Zhangsan like to read?’

- b. * 张三 喜欢 看(书) 不 看书
 zhāngsān xǐhuān kàn(shū) bù kànshū?
 Zhangsan like read NOT read
 ‘Does Zhangsan like to read?’

Thirdly, they are pragmatically different. Liing (2014) sees MA-questions as a type of confirmation question, where the asker’s already-expressed stance is either confirmed or denied, while A-NOT-A questions are just yes/no questions and neutral. She adds that MA-questions and A-NOT-A questions are generally viewed to be the same only because when in use, they typically receive the same answer. This, however, does not mean the two question types are interchangeable.

Similarly, Yuan and Hara (2013) proposed that MA-questions are simple questions, while A-NOT-A questions have the additional assertion of ignorance on the asker’s part. They suggested that MA-questions can be neutral or biased (towards assertion p , but never to $\neg p$), whereas A-NOT-A questions can only be neutral. This was based on Gunlogson (2004)’s “biased context” concept: when assertion p is explicitly/publicly indicated whereas $\neg p$ is not, then the context is biased towards p . Therefore, since MA-questions make only one visible (public) proposal (p), they can be biased towards p , while A-NOT-A questions, with both p and $\neg p$ being asserted, remains neutral.

However, I believe there are counter-examples for MA-questions where the non-asserted $\neg p$ could be the intended/biased proposal. Imagine a scenario where our protagonist Zhangsan was given a certain task, in which his competence is doubted. In such a case, the doubter might ask the following to another person:

- (26) 张三 会 跳舞 吗?
 zhāngsān huì tiàowǔ mā?
 Zhangsan know/can dance MA?
 ‘Does Zhangsan (really) know how to dance?’

In this scenario, the person could be posing the question in mild disbelief or even mockery, and thus the question could instead be biased to assertion $\neg p$, that the subject *Zhangsan* does not actually know how to dance. In other words, while MA-questions do not overtly indicate $\neg p$, they can still be biased towards that, given the right context.

In such a use, this variant bears close resemblance to the Cantonese question type that is marked with the SFP *me1*, which is biased towards $\neg p$ (Hara, 2014). The examples in (27) demonstrates this, with the asker’s assumption in brackets.

- (27) a. 张三 识 做 咩?
 Zeong1saan1 sik1 zou6 me1?
 Zhangsan know do ME?
 ‘Does Zhangsan know how to do it? (It seems to me he does not know how)’
- b. 张三 唔 识 做 咩?
 Zeong1saan1 m4 sik1 zou6 me1?
 Zhangsan not know do ME?
 ‘Does Zhangsan not know how to do it? (It seems to me he does know how)’

2.3 Existing HPSG-based accounts

Apart from ZHONG [1], there are at least two other HPSG-based grammars of Mandarin Chinese, namely the Mandarin Grammar Online (ManGO) (Yang, 2007) and the Mandarin Chinese Grammar (MCG) (Zhang et al., 2011), each with varying coverage of Mandarin Chinese. As of writing, none of these grammars — including ZHONG [1] — have coverage of the A-NOT-A structure. There are also no HPSG-based account of A-NOT-A questions outside of these grammars. As such, the present paper seeks to fill this area both in the HPSG literature as well as in the computational grammars.

3 Purpose and Scope

My purpose for this project is two-fold: Primarily, it is to derive a formal HPSG and MRS account of the A-NOT-A structure. Secondly, it is to begin an implementation of said account into the ZHONG [1] grammar developed by the Nanyang Technological University (NTU), Singapore. Within the timeframe of this project, the implementation will only be in the preliminary stages, and the state of the implementation will be noted in a later section. During the course of the implementation, limitations with the current grammatical system were also encountered. **These will be noted, and proposed changes will be stated wherever possible.**

The literature reviewed thus far has indicated several rules and constraints of the A-NOT-A structure, all of which will ideally need to be accounted for in order to have an extensive coverage of the phenomenon. However, in order to tame the scope for this paper, not all of these will be looked into and accounted for.

Firstly, only the A-NOT-A forms with 不 *bù* as NOT will be covered, swo A-NOT-A forms with 没 *méi* as NOT will not be dealt with. Secondly, this paper will limit itself to the following most basic and common instances of A-NOT-A :

- **Basic A-not-A**
 - The basic pattern, with A fully reduplicated.
- **Basic A-not-A (contracted)**
 - The partially reduplicated pattern, where only the first character is reduplicated.
- **VP-not-VP / AB-not-AB**
 - Where B represents the object of a verb being used as A.
- **AB-not-A**
- **A-not-AB**

I will also describe in brief these other patterns of A-NOT-A :

- **VV-compound as A element**
- **VP-neg**

In addition, the present analysis will also include the restrictions imposed on the A-NOT-A structure in terms of where and when it can and cannot occur, as well as the elements which can be used in conjunction with it.

To further reduce the scope for the present paper, the analysis as well as the implementation for the ZHONG [] grammar will attempt to cover only simple sentences and questions that contain only a single clause. Therefore, sentences with more complex structures such as embedded clauses and relative clauses — as well as the restrictions they impose — will not be implemented at the current stage.

Finally, although it has been shown in the previous sections that the A-NOT-A structure can be used in a non-interrogative sentence, similar to how WH-words can be embedded in a declarative sentence, such a usage arguably behaves in a slightly different way, with its own sets of constraints — particularly in the semantics. As such, for the sake of scope, this paper will only focus on the interrogative and non-embedded usage of A-NOT-A .

4 HPSG Account

4.1 Brief Introduction to HPSG and MRS

This paper makes use of the Head-driven Phrase Structure Grammar (HPSG) and Minimal Recursion Semantics (MRS) frameworks (Pollard and Sag, 1994; Copestake et al., 2005) to describe the A-NOT-A question structure and implement that as part of the ZHONG [1] Chinese grammar.

In a nutshell, HPSG is a formalism that consists of structured representations of grammatical categories (realised as typed feature structures), and is motivated by a constraint-based descriptive approach. Its structured, modular and uniform approach makes it an ideal framework for computational grammars. It is highly lexicalised, and its constraints comprise several parts, including a lexicon made up of basic words, lexical entries that build up derived words, as well as other grammatical rules and principles. In these aspects, it is similar to the Government and Binding (GB) theory, wherein the use of detailed lexical entries and grammatical principles determines structure. However, the notion of transformation is absent in the HPSG formalism, and it relies on structure sharing instead of movement to relate attributes between structures, therefore making it non-derivational. (Pollard and Sag, 1994; Levine and Meurers, 2006)

The MRS framework provides the semantic formalism used in conjunction with HPSG (or other typed feature structure formalisms), and was designed as a computational semantic framework built for parsing and generation. Like HPSG, the semantic representations in the MRS framework can be expressed easily using typed feature structures. It makes use of a flat semantic structure that eases representation of transfer rules (between languages), while still maintaining scoping information where required, hence making it an attractive tool for fields such as Machine Translation. (Copestake et al., 2005)

4.2 Groundwork

4.2.1 A-not-A as coordinate structure?

In an earlier section, I covered the analysis by Huang (1991) and McCawley (1994) which dispute the direct derivation of A-NOT-A from the coordinate disjunctive A *haishi* NOT-A. Instead, they claim that it is a distinct disjunctive question type. The differences between the two question types were later brought by Liing (2014) to mean that A-NOT-A questions are not, in reality, disjunctive questions to begin with, but are yes-no questions making a single proposition.

Could it then be that the A-NOT-A structure is not even a coordinate structure at all? Huang (1991)'s mention of the apparent violation of the Directionality Constraint by A-NOT-A suggests

that it differs somewhat from standard coordinate structures. McCawley (1994)’s examples about the non-interchangeability of the supposed coordinates (the A elements) in A-NOT-A lend further support to the idea that the structure might not be an actual coordinate structure, since in most coordinated structures, the coordinated elements are interchangeable, at least on the syntactic level. Liing (2014)’s suggestion that A-NOT-A questions are yes-no questions further strengthens this, as it highlights that underlying the structure is only one proposition.

4.2.2 Character/Syllable List

The implementation of the A-NOT-A structure necessitates the comparison of the A elements to ensure they are identical. To do so in our grammar will require that the lexical rules be able to “see” the characters or syllables of each lexical entry. As such, in addition to the existing features of a lexical entry, this analysis introduces two new features, WCHAR and FCHAR. These are the whole word (all characters) and the first character, respectively. It also introduces a LENGTH feature, which has two possible values: *one* and *more-than-one*, for a word which has one character or multiple characters, respectively.

Both of these features are implemented as part of the HEAD, and WCHAR is identical to the STEM (which is the orthographical representation of the lexical item), while FCHAR includes only the first character. These are illustrated with the following example lexical entries for the verbs 叫 *jiào* “to call” and 喜欢 *xǐhuān* “to like”:

| | | |
|------|-------------|------------|
| (28) | <i>word</i> | |
| | STEM | ① ⟨‘叫’⟩ |
| | FCHAR | ‘叫’ |
| | WCHAR | ① |
| | LENGTH | <i>one</i> |

| | | |
|------|-------------|----------------------|
| (29) | <i>word</i> | |
| | STEM | ① ⟨‘喜欢’⟩ |
| | FCHAR | ‘喜’ |
| | WCHAR | ① |
| | LENGTH | <i>more-than-one</i> |

In (28), we have a single-character word *jiào* “to call”. As described, its WCHAR value is linked to the STEM value, as they are identical. This being a single-character word, its FCHAR does not differ from WCHAR, and its LENGTH value is *one*. In (29), we have a multi-character word *xǐhuān* “to like”. As with the previous word, the WCHAR and STEM are identical. This time, FCHAR is not the same as WCHAR and its LENGTH value is *more-than-one*.

The roles of each of these features will be detailed in the §4.2 on the basic form.

4.2.3 The Headedness of A-not-A

In the existing accounts covered, the head of the A-NOT-A structure is one of the A elements, and the generation of the A-NOT-A structure begins from this head, with reduplication, elision and negation applied to it or its copy. However, whether that be A_1 or A_2 differs from analysis to analysis, depending on where the analysis determines this “starting point” to be. This, of course, poses a problem.

The presence of two semantically identical elements (even if only partially reduplicated) makes it difficult to convincingly determine which of the two should be the head.

A possibility is there actually being two heads belonging to two separate predicates that is joined by a non-overt/unseen conjunction, as is typical of a coordinate structure. However, as discussed in an earlier section, there is some reason to believe the A-NOT-A is not a coordinating structure, or one which is sufficiently different from the typical coordinating structure.

On the other extreme is the possibility that the A-NOT-A structure could be non-headed, making it similar to proposed analyses for serial-verb constructions (see Mueller and Lipenkova (2009) for more). But unlike serial-verb constructions where the two or more verbs involved are different, the A elements in A-NOT-A constructions are identical or near-identical, one essentially a copy of the other. This suggests the presence of a single head, with the other being “vestigial” to the other. The idea of a non-headed structure is, of course, also slightly uncomfortable, taking into account the “head-driven” nature of HPSG.

Based on what we know about the A-NOT-A structure, A_1 can be elided to have only its first character remaining. Such a form is usually unable to stand on its own outside of A-NOT-A or other reduplicative sequences, making it a bound or even “parasitic” form. On the other hand, the A_2 element must preserve its integrity and cannot be separated or decomposed further, making it a better candidate to be the head. However, despite this, the bound form is apparent only on the morphological level; on the syntactic and semantic level both A elements are still the same, so this admittedly is not a strong case. Of course, for purely arbitrary and aesthetic reasons, the completeness of A_2 would still be an attractive candidate as the head.

The seemingly “monolithic” nature of A-NOT-A — that is, nothing can apparently come between the structure — makes it difficult to apply headedness tests that requires mechanisms such as movement or modification of the individual components. This monolithic structure could also lend support to the idea that the A-NOT-A structure could be a single, “morphological word” that is a single predicate. In this case, whether A_1 or A_2 is the head might actually be unimportant, because

the head is simply the entire A-NOT-A morpho-word itself.

In a previous section, I brought up Liing (2014)’s claim that A-NOT-A questions are not disjunctive questions but are instead yes/no questions. In other words, there is only one proposition (p), whose truth value is determined by the responses “yes” or “no”. In this way, Liing’s account also adds credence to the notion of a single predicate upon which the truth value is solely determined.

The lack of a definitive answer (at the moment) for this suggests the possibility of two different ways to approach the A-NOT-A structure, particularly in the area of parsing. I have elected to mainly focus on the non-monolithic approach, in order to have a better account of the mechanics behind the A-NOT-A structure. I will, however, still briefly mention the “monolithic” approach in a later section.

Nonetheless, as the non-monolithic approach requires one of the A elements to be the head, it was decided to choose one of these — A_1 — to be the head.

4.3 Basic A-not-A / Basic Form

The BASIC FORM is one where the A elements are reduplicated in full. As described in the introduction section, the A element can be a verb, adjective or preposition. In this present account, only *words* are treated as A elements of the basic form. Their phrasal equivalents — namely verb phrases — are treated as a separate type (See the section on VP-NOT-VP for more details)

For computational efficiency, the non-changing element NOT of all A-NOT-A was decided to be used as a single “point of origin” which selects the appropriate A elements. This is opposed to a much more intensive (and by extension tedious) approach of having a A-NOT-A -specific lexical entry for every word that then points to a lexical rule. Additionally, it does not appear to be possible under the present system to indicate a specific word (such as 不 $bù$) to participate in a phrase, unless a feature (akin to BU=true) is assigned specifically for that word which we can then use as a constraint.

With that, a general lexical rule for this variant of $bù$ is created, as illustrated in (30). As with the normal negator, this NOT element is derived from the BASIC-SCOPAL-ADVERB-LEX.

$$\begin{array}{c}
 (30) \quad \left[\begin{array}{l}
 \text{a-not-a-adv-lex} \\
 \text{CAT} \quad \left[\begin{array}{l}
 \text{HEAD} \quad \left[\text{MODIFIABLE} \quad - \right] \\
 \text{POSTHEAD} \quad + \\
 \text{COMPS} \quad \left\langle \begin{array}{l}
 +vjr p \\
 \text{MODIFIABLE} \quad \boxed{2} \quad - \\
 \text{COMPS} \quad \boxed{3} \\
 \text{SUBJ} \quad \boxed{4} \\
 \text{ASPECT} \quad \boxed{5} \quad \text{non-aspect}
 \end{array} \right\rangle \\
 \text{MOD} \quad \left\langle \begin{array}{l}
 +vjr p \\
 \text{MODIFIABLE} \quad \boxed{2} \\
 \text{COMPS} \quad \boxed{3} \\
 \text{SUBJ} \quad \boxed{4} \\
 \text{ASPECT} \quad \boxed{5}
 \end{array} \right\rangle
 \end{array} \right] \\
 \text{CONT} \quad \left[\begin{array}{l}
 \text{SF} \quad \text{ques} \\
 \text{ASPECT} \quad \text{non-aspect}
 \end{array} \right]
 \end{array} \right]
 \end{array}$$

In the above feature structure, MOD is the A_1 element while COMPS is the A_2 element. Both the head types are indicated as $+vjr p$, which restricts the A elements to be verbs (v), adjectives (j), adverbs (r) or prepositions (p). A constraint of [MODIFIABLE -] prevents modifiers from acting on it. As both A elements share the same subject and complement(s), their SUBJ and COMPS and ASPECT values are indicated to be identical. Additionally, they are also both not modifiable.

With the above, and using a simple sentence as an example, it creates the following Minimal Recursion Semantics representation in (31):⁴

(31) Example Sentence:

张三 喜欢 不 喜欢 狗?
zhāngsān xǐhuān bù xǐhuān gǒu ?
Zhangsan like NOT like dog?

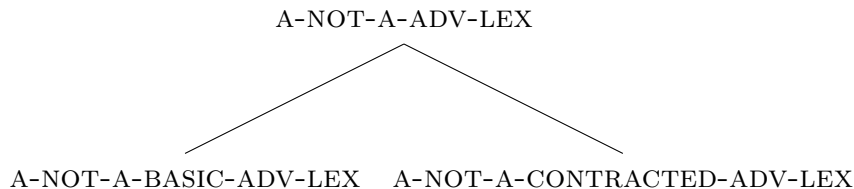
‘Does Zhangsan like dogs?’

⁴In this paper, I use labels like “LIKER” and “LIKED” for ease of understanding. The computational grammar uses generic labels like ARG0, ARG1, etc, to represent the same things.

$$\left[\begin{array}{l} \text{INDEX } \boxed{2} \left[\begin{array}{l} \text{SF } \textit{ques} \\ \text{ASPECT } \textit{imperfective} \end{array} \right] \\ \\ \text{RELS } \left[\begin{array}{l} \text{RELN } \textit{name} \\ \text{NAME } \textit{‘张三’} \\ \text{INDEX } \boxed{1} \end{array} \right], \left[\begin{array}{l} \text{RELN } \textit{喜欢} \\ \text{INDEX } \boxed{2} \\ \text{LIKER } \boxed{1} \\ \text{LIKED } \boxed{5} \end{array} \right], \left[\begin{array}{l} \text{RELN } \textit{不} \\ \text{INDEX } \boxed{3} \\ \text{NEGATED } \boxed{4} \end{array} \right], \left[\begin{array}{l} \text{RELN } \textit{喜欢} \\ \text{INDEX } \boxed{2} \\ \text{LIKER } \boxed{1} \\ \text{LIKED } \boxed{5} \end{array} \right], \left[\begin{array}{l} \text{RELN } \textit{狗} \\ \text{INDEX } \boxed{5} \end{array} \right] \end{array} \right]$$

The MRS representation in (31) illustrates the semantic relations between the elements in the simple sentence. Both the A elements are indicated to have the same subject (the LIKER), as well as the same object (the item being LIKED), since they are semantically identical. The INDEX of the entire sentence is identical to the INDEX of the A₁ element, as it was decided for it to be the head of the sentence (see §4.2.3 for a discussion on this). The SF (sentential force) feature has a value of *ques*, which indicates that it is an interrogative. This feature also serves as a constraint that prevents sentence-final particles such as *lè* and *mā* from co-occurring with it (see §2.1.3 for a brief account of SFPs and A-NOT-A questions).

This general lexical type will serve as the parent from which two other types — for the basic form and the contracted form — will be derived. As such, we shall have the following type hierarchy:



With that in place, let us proceed to the sub-type for the basic form. As the A-NOT-A-BASIC-ADV-LEX is a sub-type of the general lexical entry, it inherits all its properties. For the basic form of the A-NOT-A structure, a few additional features will be required. For simplicity, (32) indicates only these additional features:

$$(32) \left[\begin{array}{l} \textit{a-not-a-basic-adv-lex} \\ \\ \text{SYNSEM } \left[\begin{array}{l} \text{COMPS } \left\langle \left[\begin{array}{l} \text{WCHAR } \boxed{1} \\ \text{BOUND } - \end{array} \right] \right\rangle \\ \\ \text{MOD } \left\langle \left[\begin{array}{l} \text{WCHAR } \boxed{1} \\ \text{BOUND } - \end{array} \right] \right\rangle \end{array} \right] \end{array} \right]$$

As described in earlier sections, the BASIC FORM of the A-NOT-A structure requires that both A elements (A₁ and A₂) be identical. As such, in (32), the WCHAR of both the MOD (A₁) and COMPS

(A₂) are constraint to be identical. A further constraint restricts the A element to be non-bound forms. These forms are single characters of a multi-character word, and they cannot exist outside of structures like A-NOT-A (§4.4.1 covers more of this).

With the above constraints added, it is able to predict the following:

- (33) a. 张三 喜欢 不 喜欢 狗?
 zhāngsān xǐhuān bù xǐhuān gǒu ?
 Zhangsan like NOT like dog?
 ‘Does Zhangsan like or not like dogs?’

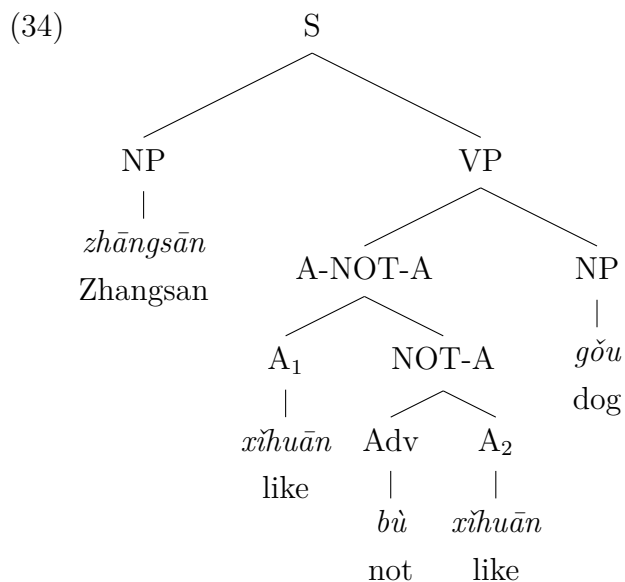
b. [Non-identical A elements]

- * 张三 讨厌 不 喜欢 狗?
 zhāngsān tǎoyàn bù xǐhuān gǒu ?
 Zhangsan hate NOT like dog?
 ‘Does Zhangsan hate or not like dogs?’

c. [Bound forms being used]

- * 张三 喜 不 喜 狗?
 zhāngsān xǐ bù xǐ gǒu ?
 Zhangsan XI NOT XI dog?
 ‘Does Zhangsan like or not like dogs?’

The above analysis will also give rise to a tree structure similar to (34), taking (33a) as the example sentence:



4.4 A-not-A contracted / Contracted Form

The A-NOT-A CONTRACTED, or the CONTRACTED FORM, is a variant of the BASIC FORM, and can be applied when the A element is a multi-character word. This form sees the contraction of the first A element A_1 such that only the first character (FCHAR) is retained, with the rest deleted. This was illustrated earlier in (1c) and (2b), duplicated here as (35a) and (35b):

- (35) a. 张三 喜 不 喜欢 狗?
 zhāngsān xǐ bù xǐhuān gǒu ?
 Zhangsan like NOT like dogs
 ‘Does Zhangsan like dogs or not like dogs?’
- b. 张三 健 不 健康?
 zhāngsān jiàn bù jiànkāng
 Zhangsan healthy not healthy
 ‘Is Zhangsan healthy or not healthy?’

Our previous implementation for the BASIC FORM will be inadequate for such cases, as the whole strings (WCHAR) for A_1 and A_2 are not identical in the CONTRACTED FORM. A different lexical rule — also a sub-type of and inheriting from the parent — will need to be created. This lexical rule is implemented as A-NOT-A-CONTRACTED-ADV-LEX:

$$(36) \left[\begin{array}{l} a\text{-not-a-contracted-adv-lex} \\ \text{SYNSEM} \left[\begin{array}{l} \text{COMPS} \left\langle \left[\begin{array}{l} \text{FCHAR} \\ \text{LENGTH} \end{array} \right] \begin{array}{l} \boxed{1} \\ \text{more-than-one} \end{array} \right\rangle \\ \text{MOD} \left\langle \left[\begin{array}{l} \text{WCHAR} \\ \text{LENGTH} \end{array} \right] \begin{array}{l} \boxed{1} \\ \text{one} \end{array} \right\rangle \end{array} \right. \end{array} \right]$$

As illustrated above, the FCHAR of the COMPS (A_2) is constraint to be identical to the WCHAR of the single-character MOD (A_1). The additional constraint of the LENGTH on the complement (A_2) element prevents this rule from also being used to parse BASIC A-NOT-A structures, such that it will parse sentence (37a) but not (37b), which is supposed to be handled by the lexical rule for the basic form.

- (37) a. 张三 健 不 健康?
 zhāngsān jiàn bù jiànkāng
 Zhangsan healthy not healthy
 ‘Is Zhangsan healthy or not healthy?’
- b. 张三 高 不 高?
 zhāngsān gāo bù gāo?
 Zhangsan tall NOT tall
 ‘Is Zhangsan tall?’

4.4.1 Single-character entries for bound forms

In addition to the new lexical rule, these single-character forms will also require their own entries in the lexicon (if they do not already exist), to allow the parser to identify them. These forms, however, always exist as part of the multi-character word, and cannot occur separated or independently outside of certain structures (such as the reduplicative adverbs formed from adjectives, as well with A-NOT-A). These are therefore considered *bound* forms, and are consequently given the constraint of [BOUND +]. As an example, (38a) shows the full (or normal) unbound form of the verb *xǐhuān* and (38b) shows its corresponding bound form.

(38) a. [Unbound/Normal Lexical Entry]:

| | | | |
|-------|---------------|----------------------|-----|
| 喜欢 _v | | | |
| STEM | [1] | ⟨ ‘喜欢’ ⟩ | |
| CHAR | [| FCHAR | ‘喜’ |
| | WCHAR | [1] | |
| | LENGTH | <i>more-than-one</i> | |
| PRED | ‘_ 喜欢 _v_rel’ | | |

b. [Bound Form Lexical Entry]:

| | | | |
|-------|---------------|------------|-----|
| 喜 _v | | | |
| STEM | [1] | ⟨ ‘喜’ ⟩ | |
| BOUND | + | | |
| CHAR | [| FCHAR | ‘喜’ |
| | WCHAR | [1] | |
| | LENGTH | <i>one</i> | |
| PRED | ‘_ 喜欢 _v_rel’ | | |

The relation between the bound forms and the unbound (full) forms are indicated in PRED, whose value is equivalent to that of the full form's. This is essential as there are other multi-character Chinese words with the same first character, and constraining the PRED will prevent these from being used with the wrong unbound forms.

In (39a) and (39b), I use 说话 *shuōhuà* “to speak” and 说笑 *shuōxiào* “to joke” to illustrate this. Both of these multi-character words have the same first character, 说 *shuō*, but each of them will still have their own single-character entries for the bound form. Their bound forms have almost the same features, except PRED, which semantically links them to their full forms.

- (39) a.
$$\left[\begin{array}{l} \text{说话_v} \\ \text{STEM } \boxed{1} \langle \text{‘说话’} \rangle \\ \text{CHAR } \left[\begin{array}{l} \text{FCHAR } \text{‘说’} \\ \text{WCHAR } \boxed{1} \\ \text{LENGTH } \textit{more-than-one} \end{array} \right] \\ \text{PRED } \text{‘_说话_v_rel’} \end{array} \right] \left[\begin{array}{l} \text{说_v} \\ \text{STEM } \boxed{1} \langle \text{‘说’} \rangle \\ \text{BOUND } + \\ \text{CHAR } \left[\begin{array}{l} \text{FCHAR } \text{‘说’} \\ \text{WCHAR } \boxed{1} \\ \text{LENGTH } \textit{one} \end{array} \right] \\ \text{PRED } \text{‘_说话_v_rel’} \end{array} \right]$$
- b.
$$\left[\begin{array}{l} \text{说笑_v} \\ \text{STEM } \boxed{1} \langle \text{‘说笑’} \rangle \\ \text{CHAR } \left[\begin{array}{l} \text{FCHAR } \text{‘说’} \\ \text{WCHAR } \boxed{1} \\ \text{LENGTH } \textit{more-than-one} \end{array} \right] \\ \text{PRED } \text{‘_说笑_v_rel’} \end{array} \right] \left[\begin{array}{l} \text{说_v} \\ \text{STEM } \boxed{1} \langle \text{‘说’} \rangle \\ \text{BOUND } + \\ \text{CHAR } \left[\begin{array}{l} \text{FCHAR } \text{‘说’} \\ \text{WCHAR } \boxed{1} \\ \text{LENGTH } \textit{one} \end{array} \right] \\ \text{PRED } \text{‘_说笑_v_rel’} \end{array} \right]$$

4.5 VP-not-VP

As the name of the structure suggests, verbs or verb phrases are the A elements.

4.5.1 AB-not-AB

A basic sub-type of this structure is the AB-NOT-AB or VO-NOT-VO structure, where the element B is the object (direct, indirect or both) of the verb A. This was illustrated in Example (1a), and has been duplicated here as (40):

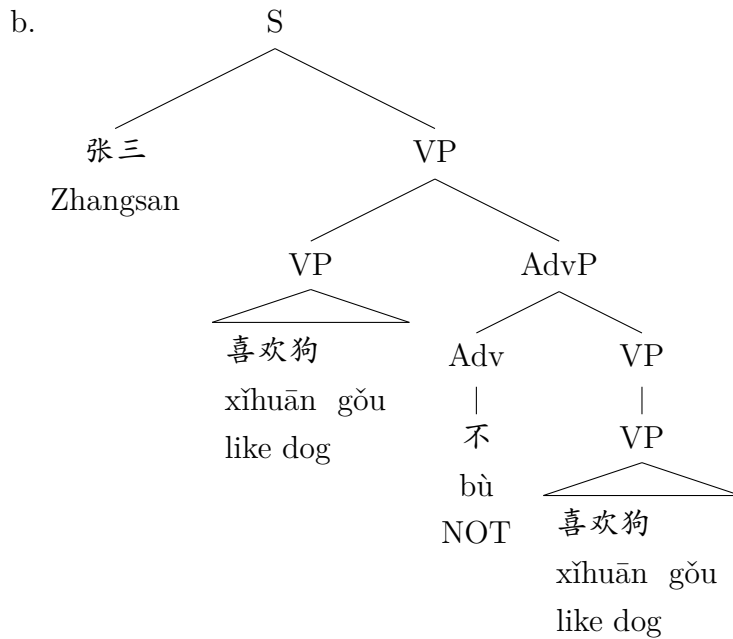
- (40) 张三 喜欢 狗 不 喜欢 狗?
 zhāngsān xǐhuān gōu bu xǐhuān gōu?
 Zhangsan like dogs NOT like dogs
 ‘Does Zhangsan like dogs or not like dogs?’

As opposed to the previous two sections for the BASIC FORM and CONTRACTED FORM, the objects of either A₁ or A₂ form a verb-phrase before they are selected by NOT, instead of only the verb being selected. This difference is illustrated in the example below. (41a) represents the basic form, while (41b) represents the tree for the AB-NOT-AB form.

- (41) a.
-
- ```

graph TD
 S[S] --- Zhangsan[张三
Zhangsan]
 S --- VP1[VP]
 VP1 --- VP2[VP]
 VP1 --- NP[NP]
 VP2 --- VP3[喜欢 不 喜欢
xǐhuān bù xǐhuān
like NOT like]
 NP --- Dog[狗
gǒu
dog]

```



The lexical rule for the *ab-not-ab-adv-lex* is largely similar to the general A-NOT-A rule introduced in §4.2, except that the A elements are specified to be phrases instead of words, and that the head of these A elements can only be verbs. Also, the subject must not be filled, to block phrases where the subject is included as part of A<sub>1</sub>. The below illustrates only the parts different from the general form (however, this lexical rule is *not* a sub-type, unlike the rules for the basic and contracted form):

$$(42) \left[ \begin{array}{l} \textit{ab-not-ab-adv-lex} \\ \text{SYMSEM} \left[ \begin{array}{l} \text{COMPS} \left\langle \begin{array}{l} \textit{phrase} \\ \text{HEAD } \textit{verb} \\ \text{SUBJ } \boxed{1} \end{array} \right\rangle \\ \text{MOD} \left\langle \begin{array}{l} \textit{phrase} \\ \text{HEAD } \textit{verb} \\ \text{SUBJ } \boxed{1} \langle [ ] \rangle \end{array} \right\rangle \end{array} \right] \end{array} \right]$$

Notice that while the above rule states that the verb must be the same (based on the WCHAR), there is no similar constraint on the object (the B element). This, unfortunately, is a limitation of the current system, and there is currently no mechanism that can ensure the object be the same, as these information, being non-head, are not automatically passed-up to the mother node. This is essential for the AB-NOT-AB implementation to be parsed successfully. However, the passing up of such elements remains unrecommended or disallowed in the present formalism. In other words, the present account remains incomplete until a change in the formalism/framework is done.

### 4.5.2 More complex VP-NOT-VP structures

As verb phrases are the A elements in this structure, significantly more complex structures can be used on either side of NOT, so long as they are permitted as verb phrases in Mandarin Chinese. The complexity of the VP usually arises from the length of the NP within, which can be an arbitrarily long, lacking an actual upper bound, as it can contain a large number of smaller conjoined elements.

We see an example of this VP-NOT-VP structure below:

- (43) 张三 喜欢 吃 便宜 的 面 不 喜欢 吃 便宜 的 面?  
 zhāngsān xǐhuān chī piányí de miàn bù xǐhuān chī piányí de miàn  
 Zhangsan like eat cheap DE noodles NOT like eat cheap DE noodles  
 ‘Does Zhangsan like to eat cheap noodles?’

Here, however, is an area where actual acceptability versus grammaticality can be at odds. While in theory the VP can be very complex, such a structure would have been deemed troublesome by native speakers, even the much simpler VO-NOT-VO structure<sup>5</sup>, if only because there is a limit to how much information can be manipulated at once by the human mind. As such, in cases of the above, such sentences will almost-always be simplified to the either the BASIC FORM — where the object is elided — or the CONTRACTED FORM — where the object and any non-first character of the verb is elided — as described in §4.4.

- (44) [Reduced to Basic Form]

张三 喜欢 不 喜欢 吃 便宜 的 面?  
 zhāngsān xǐhuān bù xǐhuān chī piányí de miàn?  
 Zhangsan like NOT like eat cheap DE noodles

‘Does Zhangsan like to eat cheap noodles?’

- (45) [Reduced to Contracted Form]

张三 喜 不 喜欢 吃 便宜 的 面?  
 zhāngsān xǐ bù xǐhuān chī piányí de miàn?  
 Zhangsan like NOT like eat cheap DE noodles

‘Does Zhangsan like to eat cheap noodles?’

<sup>5</sup>Interestingly, Liu (2010) appears to dispute the validity of the VO-NOT-VO structure in his analysis, as his analysis only moves the verb or part of it, and forbids the movement of the internal argument/object. I will, however, remain with the more general acceptance of VO-NOT-VO in this paper.

## 4.6 Other Variants

### 4.6.1 VV-compounds

VV compounds, or RESULTATIVE COMPOUNDS, are a type compound verb in Chinese where the event represented by the first verb ( $v_1$ ) in the compound and the event or state represented by the second verb ( $v_2$ ) in the compound have a causal relation. (See Li (1990) for more details.)

VV compounds, like normal verbs, can also be A elements in A-NOT-A questions. As with normal two-character words, the non-first character can optionally be deleted:

- (46) 张三 气(死) 不 气死 你?  
 zhāngsān qì-(sǐ) bù qì-sǐ nǐ?  
 zhangsan anger(-die) NOT anger-die you?  
 ‘Does Zhangsan anger you to death?’

While A-NOT-A was believed to be neutral, it is possible to have a certain bias, as the above two sentences can also be translated to “Doesn’t Zhangsan anger you?”, a likely rhetorical question which demonstrates a pre-disposition of the asker towards agreement/a YES answer. While this is *might* be apparent in ordinary V-NOT-V, the possibility of bias-ness appears to be stronger in VV-NOT-VV. This could be because of the presence of the resultative  $v_2$ , which can suggest a pre-determined conclusion that strengthens the possibility of bias.<sup>6</sup>

### 4.6.2 VP-neg

The VP-neg structure, also known as the negative particle questions, is a sub-type of A-NOT-A questions. It takes the following forms:

- (47) a. 张三 买书 不?  
 zhāngsān mǎi shū bù?  
 Zhangsan buy books NOT?  
 ‘Is Zhangsan buying books?’
- b. 张三 吃饭了没(有)?  
 zhāngsān chīfàn le méi(you)?  
 Zhangsan eat LE NOT?  
 ‘Has Zhangsan eaten?’

<sup>6</sup>This also raises a question on whether different VV compounds used in A-NOT-A can cause different degrees of bias (or not at all). It is, however, out of the scope of this paper.

This particular structure is thus named because of the negator ends the sentence, and is used similarly to a sentence-ending particle. With A-NOT-A as a basis, it appears that such a structure elides  $A_2$ , resembling what could be termed the A-NOT structure. However, is it really a derivation of the A-NOT-A structure?

The VP-NEG structure has caused some debate among linguists. In an analysis of this structure, Huang (2008) provides an overview of the literature pertaining to it, and noted the divided opinions among linguists. Some, as he noted, see it as simply an ellipsed form of the A-NOT-A structure, while others see the NOT element as a type of question particle, not dis-similar to 吗 *mā*. Huang himself was in favour of the view that it is a derivation of the A-NOT-A structure and proposed a

Cheng et al. (1996), as well as Gasde (2001) claim that historically, the VP-NEG structure could not have been a derivation of the A-NOT-A structure, because it is an older structure than the latter, having existed since the Classical Chinese era while A-NOT-A had only entered use during the Sui and Tang Dynasty. Therefore, as Gasde explained, it was more likely an independent, distinct structure. While not disputing this, Huang (2008) believes that in Modern Chinese, this alternation has become much simpler and “fairly free”, and that this structure in Modern Chinese might in reality be different from that of Classical Chinese’s. The examples below are extracted from Huang (2008), which he used to demonstrate the difference. The (a) parts are the questions, while (b) are the replies:

(48) **Modern Chinese:**

- a. 你 去 不?  
 nǐ qù bù?  
 you go NOT?  
 ‘Will you go or not?’
- b. 不 去。  
 bù qù  
 NOT go  
 ‘(I will) not go.’

(49) **Classical Chinese**

- a. 知 可 否?  
 zhī ké fǒu?  
 know possible NOT?  
 ‘Do(es) (someone/you) know whether it is possible or not?’

- b. \* 否 可。  
 fǒu ké  
 NOT possible  
 ‘(It is) not possible.’

In the above, Huang (2008) showed that while the negator in VP-NEG of Modern Chinese can negate a following predicate, the negator of the Classical Chinese equivalent cannot, and this supports the claim that the Classical Chinese and Modern Chinese question types are different from each other.

Despite the difference that Huang (2008) brought up, it could have been due more to the difference between the words 否 *fǒu* and 不 *bù* themselves (like the environment in which they can or cannot occur), rather than their syntactic functions in the above two examples. The difference illustrated in (48) and (49) works under the assumption that *fǒu* is a negator, rather than just a word that expresses a negative meaning.

Even if it is a negator, it should not be assumed to behave similarly to *bù*. For example, a similar negator-like word like 非 *fēi* can also not be used to negate everything (Contrast 非寻常 *fēi xúncháng*/不寻常 *bù xúncháng* “not normal” [both] with 不知道 *bù zhīdào* “not know” / \*非知道 *fēi zhīdào*). Following this, it could also be reasoned that the sentence-final *bù* could be different from the negator *bù*, appearing only identical on the surface. Still, such comparisons and debates are out of the scope of this paper.

From the computational grammar point of view, treating the negator in this structure as a sentence-final question particle or as a question marker does indeed simplify the analysis, since it can make use of existing analyses and accounts for such particles. Regardless, this structure requires further investigation in future accounts and, once again, it is not in the scope of this paper.

## 4.7 The “monolithic” approach

The abovementioned approach has its set of pros and cons. As it is based more on the syntactic structure of the A-NOT-A structure and how it is built-up, it provides a clearer account of the mechanics behind the formation of the structure. However, it might not be as accurate in terms of the semantics, as the single-predicate analysis cannot be satisfactorily accounted for.

An alternate approach to the analysis and parsing of the A-NOT-A structure is by treating the structure as a single morphological word, and therefore approach it from the lexicon. In such a case, most words that are permitted to be part of A-NOT-A will also have A-NOT-A versions as lexical entries. As such, each word will have three copies: 1) the normal word, 2) the A-NOT-A

basic form and 3) the A-NOT-A contracted, which can easily be generated programmatically.

Such a treatment has its own advantages and disadvantages. As a whole, it can be more semantically accurate as it will allow the representation of a single predicate. Also, it increases the computational efficiency of the parsing, as it reduces the number of combinations (or “edges”) the parser has to form out of the possible elements and constituents. For example, in “A B C D E F”, there are six atomic elements which the parser will attempt to combine to form AB, BC, CD, DE, EF, and these will be further combined based on the lexical and phrasal rules defined. The lexicon approach treats A-NOT-A structures as a single, atomic element, and thus reduces the number of elements that needs to be combined, forming something like “A B CDE F”, which only needs to create “AB, BCDE, CDEF” in the initial combination. Furthermore, this removes the ambiguity that can arise regarding the variant of negator that is involved, if the parser initially sees the NOT-A constituent as a negated phrase with a standard negator (instead of being a part of the A-NOT-A structure), thus reducing the number of “false positivies” in the parsing.

However, this implementation will inevitably lead to a very large lexicon, since, as mentioned earlier, we will need A-NOT-A versions of the words.<sup>7</sup> Secondly, it does not provide much information on the syntactic aspects of the A-NOT-A structure, such as its formation, in the grammar itself, since the A-NOT-A phrases will essentially be added manually or with a script, both of which are performed outside the grammar’s environment. Thirdly, such an analysis will not be able to account for AB-NOT-AB without producing an even larger lexicon that contains all the possible AB combinations.

Neither system at the moment provides reliable parsing of the AB-NOT-AB (or VP-NOT-VP) structure. To reiterate, the first method currently has no mechanism that permits the “checking” of the B elements — the object or the complement of the head verb — to ensure they are identical. Therefore, the account provided for it is only an “idealised” one which at present cannot be performed in the grammatical system.

In the second method, as explained earlier, the lexicon would have to be uncomfortably large to cater for the various possible complements/objects that can exist for the AB-NOT-AB or VP-NOT-VP structure, thus making it unfeasible.

---

<sup>7</sup>In a way, this might not be as big a problem as imagined, as the lexicon can be generated programmatically, assuming there is no issue with filesizes.

## 5 Implementation

This section deals with further details of the implementation of the A-NOT-A phenomenon into the ZHONG [1] Chinese computational grammar.

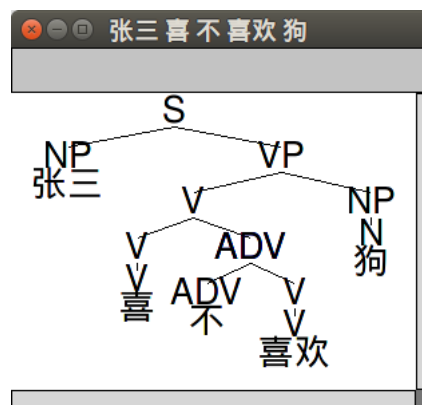
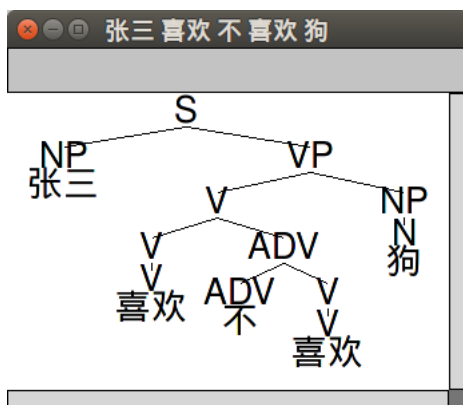
### 5.1 The ZHONG [1] grammar

ZHONG [1] is a HPSG/MRS-based (Pollard and Sag, 1994; Copestake et al., 2005) computational grammar developed by Nanyang Technological University (NTU), Singapore. It is based off the Mandarin Grammar Online (ManGO) computational grammar (Yang, 2007), although its coverage also includes other Chinese languages such as Cantonese and Min.

### 5.2 Preliminary Implementation

As laid down in the §3 on the purpose and scope of this paper, the implementation of the grammar into the ZHONG [1] has begun on the preliminary level, and it can at present already parse the basic and contracted forms of the A-NOT-A structure, albeit currently on a rudimentary level with simple sentences. To illustrate, here are some screenshots of sample parses, based on the following simple sentences in (50):

- (50) a. 张三 喜欢 不 喜欢 狗?  
 zhāngsān xǐhuān bù xǐhuān gǒu?  
 Zhangsan like NOT like dog?  
 ‘Does Zhangsan like dogs?’
- b. 张三 喜 不 喜欢 狗?  
 zhāngsān xǐ bù xǐhuān gǒu?  
 Zhangsan xi- NOT like dog?  
 ‘Does Zhangsan like dogs?’

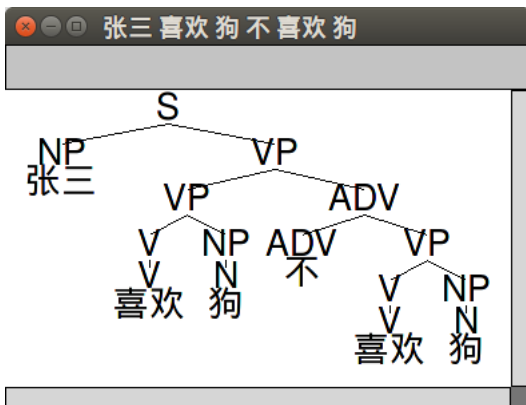




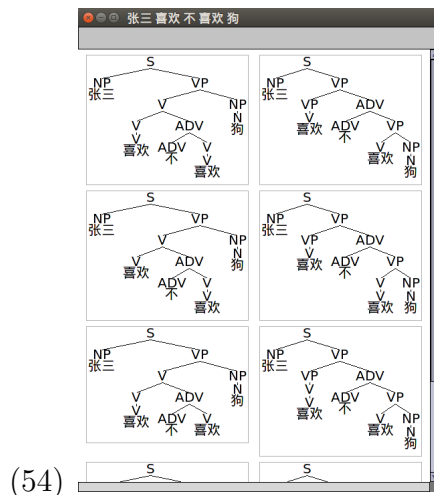
In addition to being able to parse basic A-NOT-A sentences, it also correctly rejects ungrammatical in A-NOT-A sentences, such as when  $A_1$  and  $A_2$  are different from each other.

Although the limitations of the current system means we cannot at present check that the B elements in the AB-NOT-AB structure are identical, it is nonetheless able to parse an AB-NOT-AB sentence and provide a satisfactory tree structure, albeit still parsing it if the B elements are different:

- (52) 张三 喜欢 狗 不 喜欢 狗?  
 zhāngsān xǐhuān gǒu bù xǐhuān gǒu ?  
 Zhangsan like dog NOT like dog?  
 ‘Does Zhangsan like dogs?’



The above screenshots are just one of many possible parses that the current grammar generates for the sentences. For example, for the sentence in (50a), there are at least eight parse trees generated by the computational grammar:



Not all of these are, however, desirable nor correct. As such, further work will be done to add the necessary constraints to reduce the number of parse trees.

## 5.3 Test-suite

### 5.3.1 Sentences

A test-suite of 107 A-NOT-A sentences<sup>8</sup> was created and used as part of the implementation, in order to assess the parse coverage of grammar implementation, based on its ability to parse a variety of A-NOT-A sentences. At present, the test-suite contains 43 grammatical as well as 64 ungrammatical sentences, which is used to assess the parser’s ability to both parse grammatical sentences as well as reject ungrammatical ones.

These sentences and their respective grammaticality judgements come from native sources, as well as from the literature investigated earlier. As much as possible, I made use of the existing lexical items in the lexicon, without the need to include new items, and sentences sourced from the literature were modified accordingly to fit into this requirement wherever possible.

I have also included sentences which fall outside the scope of the present study. These are, therefore, not expected to be parse-able by the implementation in the present study. However, as the A-NOT-A structure will continue to be worked on after the conclusion of this present analysis, it is expected (or at least hoped) that these sentences will eventually be accounted for by the grammar implementation.

### 5.3.2 Coverage

A preliminary run of the present account through the test-suite was performed, with the following results:

|            |             |               |
|------------|-------------|---------------|
|            | Grammatical | Ungrammatical |
| (55) Items | 43 / 107    | 64 / 107      |
| Parses     | 26 / 43     | 22 / 64       |

As shown, the parser is able to parse — with the most updated account — 26 out of 43 (~60%) grammatical sentences, as well as reject 22 of the 64 (~34%) ungrammatical sentences. These successful parses are mostly similar to the ones in (50a), (50b) and (52). These numbers appear to be satisfactory for a preliminary test. However, despite the coverage, it does not guarantee that each sentence has a valid parse, nor that each sentence has only one parse — some of them might have more than one parse tree, which might not be desirable. Again, ongoing work will increase the number of parses while improving the accuracy.

<sup>8</sup>Due to space concerns as well as the possibility of the test-suite file and its sentences being changed during development, the sentences are not included in this paper. To view the latest, actual test-suite file — as plain text — kindly visit [https://github.com/nonochnein/testsuite-anota/blob/master/A-not-A\\_testsuite.txt](https://github.com/nonochnein/testsuite-anota/blob/master/A-not-A_testsuite.txt)

Additionally, as the test-suite is diagnostic in nature, it contains sentences which are created or selected from various literature sources, and are therefore “curated” in a sense. While an increase in coverage is a goal, it does not necessarily reflect how it might perform should it be used on “real-world” sentences. As such, a companion test will be to perform the parses on corpora, and this will be performed in the future.

## 5.4 Generation

Apart from parsing sentences, the ZHONG [1] grammar is also able to *generate* sentences based on a particular phenomenon of interest. An early attempt at generation for the A-NOT-A structure was performed using the current implementation, and the grammar was able to generate grammatically correct sentences and not produce sentences which violate the rules for the A-NOT-A structure. Note that this generation is done based on a basic sentence template, and more complex sentences could lead to over-generation. As with the rest of the implementation stage, further attempts will be made as the grammar implementation increases in complexity and accuracy.

## 6 Limitations and Ongoing Investigation

As part of ongoing efforts made on the development of the ZHONG [1] grammar, coverage of the analysis and implementation of the A-NOT-A structure will continue to be expanded after the conclusion of this paper. This includes the A-*méi*-A forms of the structure, as well as the analysis and coverage of sentences of increased complexity. It is also expected that the non-interrogative variant of the A-NOT-A structure will be accounted for and implemented. The accuracy of the parses will also be improved as development continues, and the number of (wrong) parses will be reduced as work is done to improve accuracy. Also, while the current account is being implemented, there is also a possibility that the “monolithic” approach (see §4.7) will be explored further.

Also, as previously mentioned, a proper account of the AB-NOT-AB structure remains out of the reach of the present analysis and system, and might remain so unless certain fundamental changes are permitted, such as the ability to concatenate and pass up strings of the daughter nodes. As this change concerns the fundamentals of the computational grammar as well as the nature of HPSG itself, further discussion with the HPSG linguists and developers will be required to come up with a consensus.

The assessment of parse coverage will also be extended to that of the various Chinese corpora, such as the NTU Multilingual Corpus (NTUMC) (Tan and Bond, 2012), the Penn Chinese Treebank (Xue et al., 2005), among others. Unlike the controlled, curated nature of the test-suite sentences, corpora sentences are more reflective of actual usage, and as such provide a more diverse assessment of the coverage (or lack thereof) in the current and future implementations. Given the writing-based nature of such corpora, it is believed that the number of A-NOT-A sentences could be relatively low, as A-NOT-A questions are used significantly more in spoken language. As such, corpora which contain more spoken discourse — such as those from text messages, Twitter or Sina Weibo — would also be useful.

## 7 Conclusion

In this paper, I have looked at the various accounts of the A-NOT-A structure in Mandarin Chinese, as well as provide an HPSG and MRS account of this structure. At the same time, the groundwork and implementation for the integration of the structure into the ZHONG [1] grammar has also been performed.

The analysis of the A-NOT-A structure has also brought to fore some of the limitations of the present system. The inability to properly account for AB-NOT-AB at present is indeed an unfortunate setback, but this could be changed in the future.

## References

- Cheng, L. L.-S., Huang, J. C.-T., and Tang, J. C.-C. (1996). In Black, J. R. and Motapanyane, V., editors, *Microparametric Syntax and Dialect Variation*, pages 41–78. John Benjamins, Amsterdam.
- Copestake, A., Flickinger, D., Pollard, C., and Sag, I. A. (2005). Minimal recursion semantics. *Research on Language & Computation*, 3(4):281–332.
- Dai, J. X.-l. (1990). Some issues on A-not-A questions in Chinese. *Journal of Chinese Linguistics*, 18(2):285–318.
- Ernst, T. (1994). Conditions on Chinese A-Not-A questions. *Journal of East Asian Linguistics*, 3(3):241–264.
- Gasde, H.-D. (2001). Yes/no questions in Mandarin Chinese revisited. *ZAS Papers in Linguistics*, 24:47–101.
- Gunlogson, C. (2004). *True to Form: Rising and Falling Declaratives as Questions in English*. Routledge.
- Hara, Y. (2014). Semantics and pragmatics of Cantonese polar questions: an inquisitive approach. In Aroonmanakun, W., Boonkwan, P., and Supnithi, T., editors, *Proceedings of the 28th Pacific Asia Conference on Language, Information and Computation*, pages 605–614.
- Huang, C.-T. J., Li, Y.-H. A., and Li, Y. (2009). *The Syntax of Chinese*. Cambridge University Press.
- Huang, J. C.-T. (1991). Modularity and Chinese A-Not-A questions. In Georgopoulos, C. and Ishihara, R., editors, *Interdisciplinary Approaches in Linguistics: Essays in Honor of Yuki Kuroda*, pages 305–332. Kluwer Academic Publishers.
- Huang, R.-h. R. (2008). Deriving VP-Neg questions in Mandarin Chinese: A unified analysis of A-Not-A syntax. *Taiwan Journal of Linguistics*, 6(1):1–54.
- Law, A. (2001). A-not-A questions in cantonese. *UCL Working Papers in Linguistics*, 13:295–318.
- Law, P. (2006). Adverbs in A-not-A questions in Mandarin Chinese. *Journal of East Asian Linguistics*, 15:97–136.
- Levine, R. D. and Meurers, D. W. (2006). Head-Driven Phrase Structure Grammar Linguistic Approach, Formal Foundations and Computational Realization. In Brown, K., editor, *Encyclopedia of Language and Linguistics*, pages 237–252. Elsevier.

- Li, C. N. and Thompson, S. A. (1989). *Mandarin Chinese: A Functional Reference Grammar*. University of California Press.
- Li, Y. (1990). On V-V compounds in Chinese. *Natural Language & Linguistic Theory*, 8:177–207.
- Liing, W.-J. (2014). *How to ask questions in Mandarin Chinese*. PhD thesis, City University of New York, Graduate Center.
- Liu, H. (2010). What is A in Mandarin A-not-A questions? In Clemens, L. E. and L, L. C.-M., editors, *Proceedings of the 22nd North American Conference on Chinese Linguistics (NACCL-22) and the 18th International Conference on Chinese Linguistics (IACL-18)*, volume 2, pages 287–304.
- McCawley, J. D. (1994). Remarks on the syntax of Mandarin yes-no questions. *Journal of East Asian Linguistics*, 3(2):179–194.
- Mueller, S. and Lipenkova, J. (2009). Serial Verb Constructions in Chinese: A HPSG Account. In Mueller, S., editor, *Proceedings of the 16th International Conference on Head-Driven Phrase Structure Grammar*, pages 234–254. CSLI Publications.
- Pollard, C. and Sag, I. A. (1994). *Head-Driven Phrase Structure Grammar*. The University of Chicago Press, Chicago, IL.
- Radzinski, D. (1990). Unbounded syntactic copying in Mandarin Chinese. *Linguistics and Philosophy*, 13(1):113–127.
- Tan, L. and Bond, F. (2012). Building and annotating the linguistically diverse NTU-MC (NTU-multilingual corpus). *International Journal of Asian Language Processing*, 22(4):161–174.
- Tseng, K. W.-H. (2009). A post-syntactic approach to the A-Not-A questions. *USTWPL*, 5:107–139. One of two versions found (the other is for two conferences).
- Wu, J. (1997). A model-theoretic approach to A-not-A questions. *University of Pennsylvania Working Papers in Linguistics*, 4:273–289.
- Xue, N., Xia, F., Chiou, F.-D., and Palmer, M. (2005). The Penn Chinese TreeBank: phrase structure annotation of a large corpus. *Natural Language Engineering*, 11(2):207–238.
- Yang, C. (2007). *Expert Systems for Pragmatic Interpretations of ziji and Quantified Noun Phrases in HPSG*. PhD thesis, Shanghai International Studies University.
- Yuan, M. and Hara, Y. (2013). Questioning and asserting at the same time: the L% tone in A-not-A questions. In Aloni, M., Franke, M., and Roelofsen, F., editors, *Proceedings of the 19th Amsterdam Colloquium*, pages 265–272.

Zhang, Y., Wang, R., and Chen, Y. (2011). Engineering a Deep HPSG for Mandarin Chinese. In *Proceedings of the 9th Workshop On Asian Language Resources*, Chiang Mai, Thailand.