An empirical investigation on
the distribution of A-not-A questions in Chinese literary texts

Jun Da

Department of Foreign Languages and Literatures
Middle Tennessee State University
jda@mtsu.edu

Abstract: In this study, we examine the distribution of A-not-A questions using a large collection of Chinese literary texts. Our results indicate that A-not-AB is preferred over AB-not-AB in Modern Chinese, whereas AB-not-AB is preferred over A-not-AB in Classical Chinese. We suggest that such a contrast in distribution provides support to a reductionist account of the A-not-A questions.

1. Introduction

A-not-A questions have received intensive studies in Chinese linguistics. In the past, most researches have focused on formal accounts of the structure from both syntactic and semantic/pragmatic perspectives. In an A-not-A question, the element A is simply 'a label for several predicative categories such as verb, adjective, modal and others' (Gasde 2004) as shown in Examples (1) to (3) below.

(1) V-neg-V
　　张三　　去 不 去?
　　Zhangsan qu bu qu
　　Zhangsan go not go
　　Is Zhangsan going or not?

(2) Adj.-neg-Adj
　　张三　　高 不 高?
　　Zhangsan gao bu gao
　　Zhangsan tall not tall
　　Is Zhangsan tall or not?

(3) M(odal)-neg-MV(O)

    张三　　能 不 能　来?

    Zhangsan neng bu neng lai

    Zhangsan can   not can   come

    Can Zhangsan come or not?

While a typical A-not-A question contains two identical elements on both sides of the negator, the structure has been found to have four surface variations when the element A is made up of more than one grammatical constituents. Examples (4) to (7) illustrate the four possibilities when, for instance, the element A is made up of a Verb-Object structure.

(4) V-neg-VO

    张三　　喝　不 喝　　酒?

    Zhangsan he   bu he   jiu

    Zhangsan drink not drink wine

    Does Zhangsan drink wine or not?

(5) VO-neg-VO

    张三　　喝　酒 不 喝　　酒?

    Zhangsan he   wine bu he   jiu

    Zhangsan drink wine not drink wine

    Does Zhangsan drink wine or not?

(6) VO-neg-V

    张三　　喝　酒 不 喝?

    Zhangsan he   wine bu he

    Zhangsan drink wine not drink

    Does Zhangsan drink wine or not?
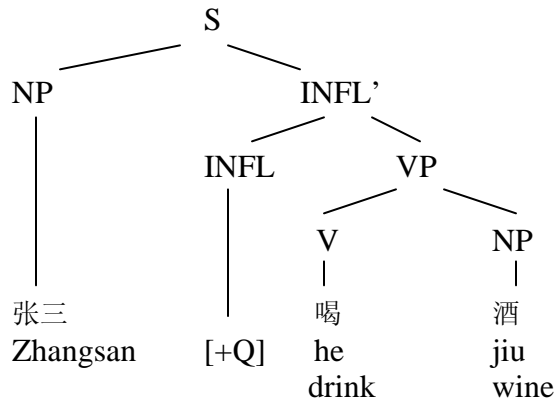
(7) VO-neg

    张三　　喝　酒 不?

    Zhangsan he   wine bu

    Zhangsan drink wine not

    Does Zhangsan drink wine or not?

In the rest of the paper, we will use A-not-AB, AB-not-AB, AB-not-A and AB-not to denote the four surface forms as shown above, where A is made up of such lexical categories such as verb, modal, adjective, co-verb or preposition and B some other grammatical constituent.

Various derivational accounts of A-not-A questions can be categorized into two different schools, depending on their assumptions about the D-structure of A-not-A questions. On the one hand, there are analyses such as Huang (1991) and Law (2001), etc. that claim that A-not-A questions of the A-not-AB type are derived from a D-structure that contains an interrogative ([+Q]) INFL constituent and a single element A (c.f. Example 8).

(8) D-structure in A-not-A questions

```
                        S
            ┌───────────┴───────────┐
           NP                     INFL'
            │              ┌─────────┴─────────┐
            │            INFL                  VP
            │              │           ┌───────┴───────┐
            │              │           V              NP
            │              │           │               │
           张三            │           喝              酒
         Zhangsan        [+Q]         he              jiu
                                      drink           wine
```

According to this school of analysis, the interrogative ([+Q]) INFL constituent is phonetically realized by a reduplication rule that copies a sequence immediately following INFL and inserting the negative morpheme *bu* between the original and the copy. We note that analysis along this line is mainly focused on observed structures such as V-neg-VO.

On the other hand, there are other analyses that treat the constituent structure A-not-A as a morphological word in D-structure, whose surface forms such as V-neg-VO are said to result from coordination reduction, anaphoric ellipsis (c.f., Li and Thompson 1981, Schaffar 2000 and Hsie 2001, among others) or movement (Gasde 2004). In Gasde's (2004) analysis, for example, under the assumption that Chinese is a typologically SOV language, the entire A-not-A structure is taken as a morphological word, where the neg-stem can be treated as a semi-suffix that can move along with the stem or be left behind. Examples (9) and (10) illustrate the derivation of A-not-A questions in Gasde's framework.

(9) V-not-VO
　　　张三　　喝　不 喝　酒?
　　　Zhangsan he　　bu he　jiu
　　　Zhangsan drink not drink wine
　　　Does Zhangsan drink wine or not?

D-structure of V':
[$_{V'}$ jiu he-bu-he]
S-structure
[$_{V'}$ he-bu-he$_1$ [$_{V'}$ jiu t$_1$]]

(10) VO-not-V
张三　　喝　酒　不　喝?
Zhangsan he　　wine bu　he
Zhangsan drink wine not drink
Does Zhangsan drink wine or not?
D-structure of V':
[$_{V'}$ jiu he-bu-he]
S-structure
[$_{V'}$ he$_1$ [$_{V'}$ jiu t$_1$-bu-he]]

In terms of functions, A-not-A questions have been regarded as either ordinary disjunctive/alternative questions (Li and Thompson 1981, Schaffar 2000, Law 2001 and Gasde (2004)) or WH-questions (Huang 1991). They are used in a (presuppositionally) neutral context (Li and Thompson 1981 and Zhang 1997), or as Zhang (1997) puts it: 'The V of A-not-A questions cannot be presupposed'. To compute its semantic meaning, the answer to an A-not-A question will involve picking one out of two mutually exclusive and jointly exhaustive values for a two-valued variable (Huang 1991, Wu 1997 and Law 2001). In Wu's (1997) analysis, for example, he assumes that the negator in an A-not-A question functions as an adverbial and proposes that an A-not-A question is a nonempty partition of the possible states of affairs into two mutually exclusive and jointly exhaustive cells. The answer to an A-not-A question thus involves a function that picks one of the cells as true and rejects the other as false.

Given those competing derivational accounts about A-not-A questions, however, there are still some issues that remain unanswered. In the case of A-not-AB vs. AB-not-A questions, for example, some researchers such as Huang and Hsie have to offer different accounts of the derivation of the two surface forms, where A-not-AB is said to be derived in terms of reduplication and AB-not-A syntactic reduction. In contrast, Gasde (2004), citing previous analyses by Shao (1996) and Chen (1990), refers to them as historical differences. Another remaining issue is the observation that a single AB element (whether containing two bound morphemes or two distinct grammar constituents) can occur in four surface variations at the same time, that is, A-not-AB, AB-not-AB, AB-not-A and AB-not are possible and valid A-not-A questions. None of the above formal analyses is ready to specify when and how to apply its rules to generate a particular surface form and exclude the other forms. Lastly, in the majority of the literature, only such grammatical categories such as verbs, modals and adjectives are said to occupy the

A position in an A-not-A structure. One wonders if other grammatical categories are also possible candidates for the A position.

In this study, we investigate in particular two empirical issues concerning A-not-A questions that have received little attention in previous formal studies: 1) Among the four possible surface variations of AB-not-AB questions, is there a preference of any particular form or forms in real language use? 2) Do A-not-A questions always function as ordinary Yes-No or WH-questions that require a response involving choosing one value over the other(s) for a question variable? We believe that results from this study may shed lights on the adequacy of various formal accounts proposed so far.

2. This study

2.1 Data collection and pre-processing

The main objective of our study is to discover the distribution patterns of A-not-A questions, especially the contrast among the four different surface variations. In order to discover those distribution patterns, a search was conducted on a very large corpus of Classical and Modern Chinese literary texts. We chose literary texts as a close approximation to real spoken language use because of the lack of accessible spoken data. The literary texts included in our study were collected between 1997 and 2003 from various online sources that fall into two different categories: 1) Digitized texts that were originally published in printed format. Examples of this type of texts include Classical and Modern Chinese texts published before 1995; 2) Original literary texts that are written and published on the Internet and intended for online readers.

As far as text registers are concerned, the Classical Chinese texts in our corpus include novel, prose, drama, historical records and commentaries, where the majority of the texts are novels written in the Ming and Qing dynasties. Our Modern Chinese corpus also contain fiction, prose, drama, biography, reportage, literary reviews and notes, etc. that have been published since 1912.

The majority of the texts collected are HTML-tagged web pages that contain Simplified Chinese texts encoded in either the GB2312 or GBK standard. Those web pages were downloaded automatically from various online sources using tools such as *w3mir*[1]. Before we conducted searches on those texts for A-not-A patterns, HTML tags were automatically removed from those texts. At the same time, any text strings used for website navigation or as webpage footers were also removed. In all, our text corpus contains more than 300 millions of characters.

---

[1] c.f. http://langfeldt.net/w3mir/

2.2 Identification of A-not-A patterns

To facilitate an automatic search of large text corpus, we define an A-not-A question as a text string that contains the character 不 (*not)* and ends with a question mark. Note that while other negative morphemes such as 没 (*not* (*yet*)) can also occupy the negator position, the character 不 (*not)* is presumably the most common negator found in A-not-A questions. Further, our search is restricted to those A-not-A questions where A is rendered as a two-character text string that corresponds to either two bound morphemes that form one grammatical category (e.g., 喜欢 (*like*)) or two lexical constituents such as verb and noun, etc. (e.g., 喝酒 (*drink wine*)) that form a VO structure.

Identification of those A-not-A patterns from our text corpus took two steps: A computer program was used to first extract all possible A-not-A patterns from the corpus. Once those A-not-A patterns were identified, a further search was conducted on those strings to pick out AB-not-AB patterns where AB is a two-character text string occurring in a question sentence (that is, marked with a question mark). It should be pointed here that while AB-not-AB questions are not limited to those two-character pattern, it is nevertheless sufficient for us to examine those particular cases with two-character strings that serve as a good indication of overall A-not-A question distribution in Chinese literary texts.

2.3 Results and discussions

2.3.1 The distribution of A-not-A questions

A total of 118,622 text segments were identified from our corpus of Classical and Modern Chinese texts that contain the character 不 (*not*) and end with a question mark [2]. Among those text segments, we are able to find the same two-character AB string occurring in the four surface variations of A-not-A questions, namely, AB-not-AB, A-not-AB, AB-not-A and AB-not. Examples (11) to (18) are the cases of 知道 (*know*) and 喝酒 (*drink wine*)[3] in Classical and Modern Chinese, respectively.

Classical Chinese
(11) A-not-AB
"潘将军丢失了一串玉念珠，不知你**知不知道**？"(李昉：《太平广记》)

---

(12) AB-not-AB
你来的时侯老太太**知道不知道**？(曹雪芹：《红楼梦》)

(13) AB-not-A
你头里过这边来，他大娘**知道不知**？(《金瓶梅 (崇祯版) 》)

(14) AB-not
独孤丈人又说："《凉州》一曲，你吹到第十三叠误入水调，你自己**知道不**？"(李昉：《太平广记》)

Modern Chinese
(15) A-not-AB
弯腰老头回身 …，含笑问道："小哥**喝不喝酒**？"
第三十章　剑劈四凶: 《东方第一剑》 东方玉

(16) AB-not-AB
许年华问："你现在还**喝酒不喝酒**？"
《官场》刘震云

(17) AB-not-A
郭襄道："雕大哥，咱们一起去罢。我请你吃好东西，你**喝酒不喝**？"
《神雕侠侣》 :金庸: 三十六回

(18) AB-not
老板热情地问："同志，**喝酒不**？"
《井冈风云录》王运朝

From those 118,622 text segments, 815 unique text strings are identified in Modern Chinese that contain identical two-character AB at both sides of the negator 不 (*not*). In Classical Chinese, 241 such unique records are identified. Table 1 lists frequency counts of the four surface variations of AB-not-AB structure where AB is a two-character text string.

Table 1 Distribution of the four possible variations of AB-not-AB questions

	AB-not-AB	A-not-AB	AB-not-A	AB-not
Modern Chinese	815	312	57	154
Classical Chinese	241	10	10	9

In Table 1, we notice the distribution asymmetry among the four possible variations. There is the tendency in both Classical and Modern Chinese that for the same two-character string AB occurring in the AB-not-AB form, it is less likely to find the other three corresponding surface forms, namely, A-not-AB, AB-not-A and AB-not. This is especially true in Classical Chinese where those 241 AB-not-AB questions have very few cases of the other three variations. A random manual examination of our search results indicates that the distribution pattern holds true whether AB is made up of two bound morphemes (e.g., 喜欢 (*like*)) or form a grammatical construction such as Verb-Object (e.g., 喝酒 (*drink wine*)) or Modal-V (敢看(*can/dare to look*).

However, there is a difference between Classical Chinese and Modern Chinese when AB-not-AB and A-not-AB are concerned: In cases where the same two-character string AB is found to occur in both A-not-AB and AB-not-AB forms, it occurs more frequently in AB-not-AB than in A-not-AB in Classical Chinese (c.f. Figure 1). One-tailed z-test shows that the difference between the two forms is significant at the $\rho < 0.01$ level. In contrast, there are only 301 cases in Modern Chinese where the same AB is found to occur in both A-not-AB and AB-not-AB forms (c.f. Figure 2). In those 301 instances, we notice that the A-not-AB form occurs at a higher frequency than the AB-not-AB form. One tailed z-test shows that the difference is also significant at the $\rho < 0.01$ level[4].

Figure 1 A-not-AB and AB-not-AB distribution in Classical Chinese

---
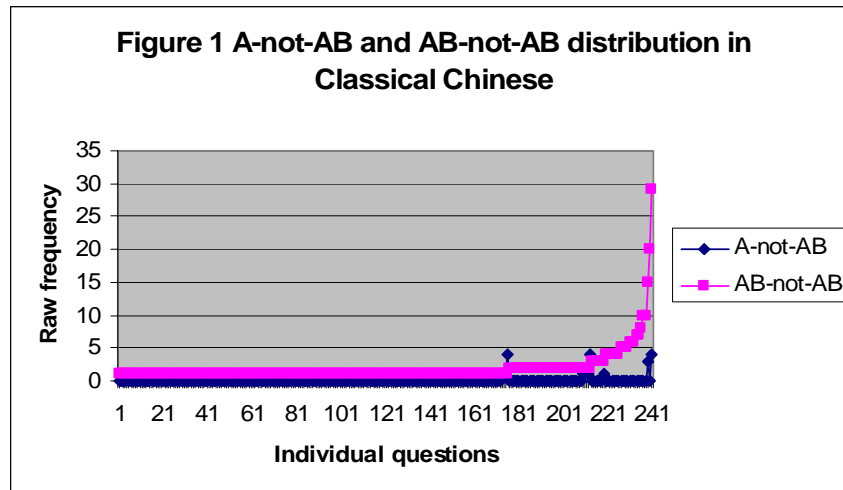
[4] We also notice that there are about 20 A-not-AB questions with very high frequency, where AB can contain a single verb with two bound morphemes or two lexical constituents.

Figure 2 A-not-AB and AB-not-AB distribution in Modern Chinese
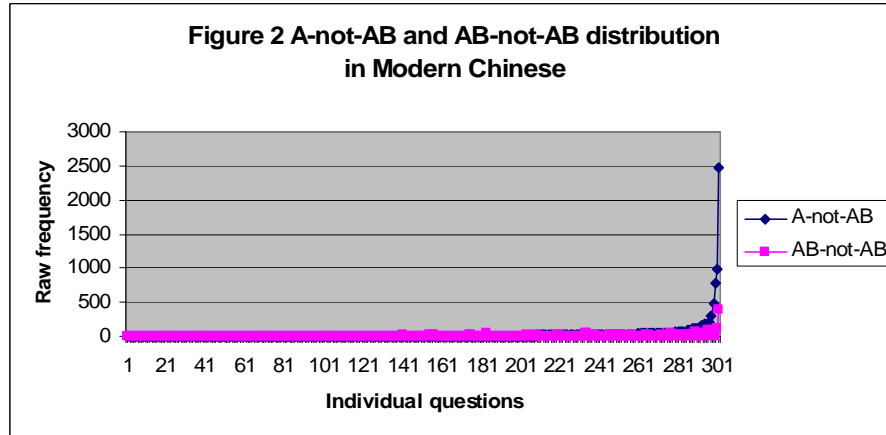
Given the fact that both the A-not-AB and AB-not-AB forms can occur in the same context, the contrast between A-not-AB and AB-not-AB in Classical (where AB-not-AB is preferred) and Modern Chinese (where A-not-AB is preferred) seems to suggest that there is an ongoing language change from the AB-not-AB form to the A-not-AB form. We think that this change is very possible as part of the disyllabification evolution from Classical to Modern Chinese (c.f., Feng 1995 and 2004). If it is true, the above observed different distribution patterns of A-not-A questions in Classical and Modern Chinese will favor reductionist proposals such as those by Gasde (2004) and Hsie (2001) about an underlying A-not-A structure as a morphological word. When A-not-A is taken as a morphological word in D-structure, its various four surface forms can be better accounted for in terms of prosodic domain and other morphological constraints as suggested by Dai (1990). There will be no need to propose two different syntactic accounts for the derivation of the A-not-AB and AB-not-A forms.

2.3.2 Constituency of A in A-not-A questions

While most of the AB element in AB-not-AB questions in our search result are made up of two bound morphemes (e.g., 喜欢 *(like)*) or form a grammatical construction such as Verb-Object (e.g., 喝酒 *(drink wine)*) or M-V (敢看 *(can/dare to look)*), we also found a few two-character AB strings in AB-not-AB structures that are nouns, as shown in Examples (19) to (21).

(19) 颜璧笑道："咱们只是游览，管它什么**鞑子不鞑子**？不犯王法，害怕什么？"（梁羽生《风云雷电》）

(20) "… 众人来到灵州，为的就是要做驸马，倘若不听公主吩咐，她势必不肯召见，见都见不到，还有什么**驸马不驸马**的？只怕要做驸牛驸羊也难。" (金庸《天龙八部》)

(21) 赵大嫂子笑着打断她的话： "啥**主子不主子**的？你这还是旧脑瓜。" (周立波《暴风骤雨》)

We notice that rather than functioning as Yes-No or WH-questions, the above three examples are actually rhetorical questions where the AB element functions as objects. Given that rhetorical questions have the illocutionary force of a strong assertion that does not involve the selection of one value over the other(s) for a question variable, we wonder if the above examples represent a further type of A-not-A questions that requires additional formal account.

3. Concluding remarks

In this study, we conducted a corpus-based study of A-not-A questions in Chinese literary texts. Our results indicate that when both the A-not-AB and AB-not-AB forms are possible for the same two-character AB string, A-not-AB is preferred over AB-not-AB in Modern Chinese, whereas the reverse is true in Classical Chinese. We suggested that such distribution differences are simply a reflection of a language change in progress and can be better accounted for following a reductionist approach.

Note that there are some limitations in this study. First, we have only examined A-not-A questions where the negator is rendered as 不 (*not*). It would be more interesting if we also looked at other negators such as 没 (*not yet*) in such structures as 看没看过, etc. Secondly, because of some technical constraints, we did not consider the case that A-not-AB patterns do not have the corresponding AB-not-AB form. There is always the possibility that in Classical Chinese the A-not-AB form outnumbers the AB-not-AB form. The other possibility is that all A-not-AB forms always have a corresponding AB-not-AB form either in Classical or Modern Chinese. Lastly, again because of some computer programming issues, we did not count those A-not-A patterns where the element A is made up of more than two characters. To validate our conjecture that the A-not-AB form simply results from the AB-not-AB form as part of the disyllabification evolution from Classical to Modern Chinese, it is necessary that we also look at A-not-A questions where the A element contains more than two characters. It is hoped that future research can take those deficiencies into consideration.

References

Dai, John Xiang-ling. 1990. Some issues on A-not-A questions in Chinese. Journal of Chinese Linguistics. 18,2,286-317

Feng, Shengli. 1995. Prosodic Structure and Prosodically Constrained Syntax in Chinese. The University of Pennsylvania Ph.D. Dissertation

Feng, Shengli. 2004. Prosodic structure and its implications for teaching Chinese as a second language. Journal of the Chinese Teachers Association. 39,1,1-24

Gasde, Horst-Dieter. 2004. Yes/no questions and A-not-A questions in Chinese revisited. Linguistics. 42,2,293-326

Hsieh, Miao-Ling. 2001. Form and Meaning: Negation and Question in Chinese. Ph.D. Dissertation, University of Southern California.

Huang, C.-T. James. 1991. Modularity and Chinese A-not-A questions. In Georgopoulos, C. and R. Ishihara (eds.). Interdisciplinary approaches to language: Essays in honor of S.-Y. Yuroda. 305-332. Dordrecht: Kluwer

Law, Ann. 2001. A-not-A questions in Cantonese. UCL Working Papers in Linguistics 13. 295-318

Li, Charles N. and Sandra A. Thompson. 1981. Mandarin Chinese: A functional reference grammar. Berkeley: University of California Press

Schaffar, Wolfram. 2000. Typology of Yes-No questions in Chinese and Tai languages. Talk given at ZAS, December 2000. (Online version at http://www.zas.gwz-berlin.de/mitarb/homepage/webfest/Schaffar.pdf. Last checked: 2004-05-04)

Wu, Jianxin. 1997. More on A-not-A questions: A model-theoretic approach. The Proceedings of the Sixteenth West Coast Conference on Formal Linguistics 16: 463-477

Zhang, Zheng-sheng. 1997. Focus, presupposition and the formation of A-not-A questions in Chinese. Journal of Chinese Linguistics. 25,2,227-257