A Unification-based Approach to Mandarin Questions

Yu-Ling Shiu* **& Chu-Ren Huang****

* BTC R&D Center
2F, No. 28, R&D Road II
Science-Based Industrial Park
Hsinchu, Taiwan, R.O.C.

** Institute of History and Philology
Academia Sinica
Nankang, Taipei, R.O.C.

ABSTRACT

This paper provides unification-based GPSG and LFG analyses of Mandarin questions. First, we briefly introduce four kinds of Mandarin questions, namely, **WH-questions**, **A-not-A questions**, **disjunctive questions**, and **particle questions**. Their different interrogative messages are adequately encoded with different feature-value pairs. Then, the compatibility of these interrogative information in simple sentence is investigated. Both GPSG and LFG can provide straightforward account for their mutual exclusiveness. Finally, the scope of percolation of Mandarin interrogative information is examined. It is suggested that the matrix verb of a complex sentence is responsible for the scope of interrogative information in its complement sentence. According to our observations, Mandarin verbs should be divided into at least three classes. We provide preliminary analyses of this topic. The GPSG analysis relies on the **Foot Feature Principle (FFP)** and the LFG analysis relies on **functional uncertainty**. It is shown that the transmitting of Mandarin interrogative information can also be adequately accounted for in GPSG and LFG.

0. Introduction

In contrast to a purely formal concern of whether a string is generatable by the grammar of a certain language, recently an informational approach to linguistic phenomena presents linguists' renewed perspective of regarding language as a system for encoding and transmitting ideas (see Kay (1986)). This approach requires grammar formalisms representing how language convey information. Such requirement is accomplished by associating strings with their informational domain of well-structured set of feature-value pairs. Grammar formalisms derived from this design choice are capable of encoding various kinds of information, which is especially important in the research community of natural language understanding and generation. Thus, in this paper we attempt to study Mandarin questions from an informational point of view.

Traditionally, Mandarin questions are divided into four main types, namely, **WH-questions, A-not-A questions, disjunctive questions** and **particle questions.**[1] Unlike English, which always involves Subject-Aux inversion or WH-word fronting in question formation, Mandarin Chinese does not have any characteristic syntactic constructions to mark interrogatives. Except for intonation,[2] which is beyond our syntactic consideration in this paper, declarative and interrogative counterparts in Mandarin may just differ in the existence of a crucial element, such as **a WH-word, an A-NOT-A construction, a disjunctive conjunction**, or **an interrogative sentential clitic**. This is illustrated as follows:[3]

(1) Yijing pa lauhu.
Yijing fear tiger
' Yijing is afraid of tiger. '

(2) ***Shei*** *pa lauhu* ? **(WH question)**
Who fear tiger
' Who is afraid of tigers ? '

(3) *Yijing* ***pa-bu-pa*** *lauhu* ? **(A-NOT-A question)**
Yijing fear-not-fear tiger
' Is Yijing afraid of tigers or not ? '

(4) *Yijing pa lauhu* ***haishr*** *pa shrtz* ? **(Disjunctive question)**
Yijing fear tigers or fear lions
' Is Yijing afraid of tigers or afraid of lions ? '

---

1 This classification is adopted mainly from Tang (1981), in which tag questions are not regarded as a separate type. Discussions of tag questions can be found in Tang (1981: 20-21) and Li & Tompson (1981: 546).

2 It is always possible to turn a Mandarin statement into a question by using a rising intonation.

3 The Romanization system adopted in this paper is Mandarin Phonetic Symbols II (MPS II), which is formally announced by the Ministry of Education R.O.C. in 1986.

(5) *Yijing pa lauhu ma* ? **(Particle question)**
Yijing fear tiger MA
' Is Yijing afraid of tigers ? '

Different kinds of interrogative elements may co-occur within a sentence, and their conditions on compatibility and environments of their co-occurrences seem rather intriguing. In addition, different kinds of interrogative elements encode different kinds of interrogative information and have different kinds of semantic implications. Taking the informational approach, we provide a systematic and straightforward solution to this problem and a preliminary study of the encoding and transmitting of Mandarin interrogative information. In particular, the compatibility nature and the scope of percolation of these interroative information will be carefully investigated. Since the flow of information is much more explicitly formulated in unification-based formalisms, and **Generalized Phrase Structure Grammar (GPSG)** and **Lexical Functional Grammar (LFG)** are two of the linguistically best-established frameworks using this approach, we will adopt them in subsequent discussions.[4] Accounts in either frameworks are independently motivated. Their mutual compatibility and validity, however, lend support to Shieber's (1986, 87) advocation of unification as an underlying grammar formalism.

I. The Encoding of Mandarin Interrogative Information

1.1. A GPSG Analysis

As mentioned previously, Mandarin questions are marked solely by the existence of interrogative elements. In GPSG, this phenomena may raise problems on semantic interpretation. Adopting the basic concept of Montague Grammar, syntax and semantics in GPSG are separate but parallel components, in which every syntactic structure is directly paired with a semantic interpretation. Since Mandarin declaratives and interrogatives do not differ in their syntactic structures, their semantic denotations could also be indistinguishable. As a consequence, syntactic specifications which are semantically interpreted have to be introduced to encode different kinds of interrogative information.[5]

A. WH Questions

In Mandarin, WH-questions are formed by simply replacing the elements questioned with appropriate WH-words. Thus, the presence of a WH-word is the sole marker of a WH-question. Since syntactic categories in GPSG are taken to be sets of feature-value pairs and each pair encodes a piece of linguistically significant information, a feature-value pair [QTYPE

[4] Readers are referred to Sells (1985) for a general overview of the GPSG and LFG frameworks, to Gazdar et al. (1985) for the most complete description of GPSG, and to Bresnan (1982) for a collection of important LFG literatures. In-depth discussion of unification can be found in Shieber (1986), Sag et al. (1986), and works cited therein.

[5] For more detailed discussion on how syntax and semantics interact in GPSG, please see Gazdar et al. (1985: 182-244).

WH] is hence postulated to encode the interrogative messages of WH-words. Accordingly, the typical WH-word *shei* will be listed in lexicon as shown in (6):

(6) < ***shei***, [N +], [V –], **[QTYPE WH]**, ...] ... >

One point worth noting is that the interrogative information is crucially related to sentence type. Thus, although the interrogative specifications are encoded in the lexical entry of WH-words, they must be semantically interpreted at a sentential level. A natural solution to this problem in GPSG is to assign the feature QTYPE to the class of FOOT features. In GPSG, features, according to their percolation properties, are divided into three classes; namely, HEAD features, FOOT features, and LOCAL features. Foot features distributions obey the Foot Feature Principle (FFP) :

(7) **FOOT Feature Principle (FFP)** :

> The FOOT feature specifications that are instantiated on a mother category in a tree must be identical to the unification of the instantiated FOOT feature specifications in all of its daughter categories.
> (Gazdar et al. (1985: 82))

The basic operation underlying FFP is unification. Based on such mechanism, specifications will be "passed up" from a phrasal daughter to a mother. Thus, interrogative information in GPSG can be locally specified in lexicon, while be checked and percolated (if unification is successful) unbounded up the tree.

B. A-NOT-A Questions

Traditionally, an A-NOT-A question is considered as the result of identical elements deletion from a full coordinate structure which is formed by an affirmative sentence and its negative counterpart. However, this analysis is not appropriate here because there are no transformations in GPSG at all. An alternative approach is to regard a A-NOT-A question as involving a morphological copying process. Thus, we assume that the whole A-NOT-A construction, after some kind of morphological process, encodes a specification **[QTYPE A-NOT-A]**.

C. Disjunctive Questions

Most linguistic articles analyze A-NOT-A questions on a par with disjuntive questions. Both of them explicitly present the respondent with a choice of some possible answers. But syntactically, disjunctive questions have less restrictions on their conjuncts.[6] Thus, in GPSG, we must assume the disjunctive conjunction *haishr* independently bears a kind of interrogative information [QTYPE DJ] in its lexicon. The lexical entry of *haishr* is given below:

(8) < ***haishr***, [..., **[QTYPE DJ]**, ...] ... >

---

[6] The conjuncts of an A-NOT-A question must be an affirmative predicate (or predicate phrase) and its negative counterpart. That is, the number of them is limited to two, and the syntactic category of them must be a predicate. But disjunctive questions do not have such restrictions on their conjuncts.

D. Particle Questions

According to Shiu (1989), *ma* is the most typical interrogative sentential clitic in Mandarin,[7] and it functions to turn a statement into a yes-no question. So, the lexicon of *ma* is presented in (9):

(9) < ***ma***, [[CLIT MA], **[QTYPE YN]**, ...] ... >

1.2. An LFG Analysis

In LFG, since semantic interpretation is derived from the attribute-value matrix representations of f-structures, we also have to properly introduce different feature-value pairs to encode interrogative information . Here, we also assume that the presence of the feature QTYPE marks a sentence as a question and the value of this feature further specifies which kind of question the sentence is. Thus, Mandarin interrogative elements are represented in lexicon as (10):

(10) **Lexicon**

ma	CLIT	(↑LAST) = + **(↑QTYPE) = YN**
pa-bu-pa	V	(↑PRED) = 'FEAR<(↑SUBJ)(↑OBJ)>' **(↑QTYPE) = A-NOT-A**
shei	N	**(↑QTYPE) = WH** (↑QTYPE) = body bottom (↑PRED) = 'PRO' (↑HUMAN) = +
haishr	CONJ	**(↑QTYPE) = DJ** (↑QTYPE) = body bottom

Again, this interrogative feature QTYPE should be interpreted at the matrix level in f-structure. But instead of general feature percolation principles as in GPSG, the LFG mechanism of functional equations explicitly specify how the functional information contained in lexicon or on a node in c-structure participates in f-structure. That is, the flow of information in LFG is governed by independent functional equations. The lexical entry of *ma* has been discussed in Shiu (1989). The treatment of A-NOT-A construction is similar to that of GPSG. We assume the whole A-NOT-A construction is the output of a morphological process and encodes an equation '(↑QTYPE)= A-NOT-A'. The WH word *shei* and the disjunctive conjunction *haishr* encode an equation (↑QTYPE)=WH and (↑QTYPE)=DJ respectively.

[7] Zwicky (1985) has investigated the grammatical status of clitics and particles. It is suggested that 'clitic' is a theoretical construct which belongs to a level between 'word' and 'affix', while 'particle' is a redundant cover term which should be eliminated. Following this line of approach, Huang (1985) explicitly points out that Mandarin sentential particles are indeed sentential clitics.

The equation '(↑QTYPE)=body bottom' encoded on both of them indicates a device of **functional uncertainty** (proposed in Kaplan & Zaenen (in press)), which will be discussed in detail in section IV.

Given the above GPSG and LFG analyses, every kind of Mandarin interrogative information can be adequately encoded and appropriately interpreted. These analyses will be further supported in the next two sections.

II. The Compatibility Nature of Mandarin Interrogative Information

In this section, we will briefly discuss how the interrogative information in Mandarin interacts within simple sentences. Let us consider the following sentences:

*(11) ***Shei pa-bu-pa*** *lauhu* ?
(WH word & A-NOT-A construction)

*(12) ***Shei*** *pa lauhu* ***haishr*** *pa shrtz ?*
(WH word & disjunctive conjuction)

(13) Yijing ***pa-bu-pa*** *lauhu* ***haishr*** *shrtz ?*
(A-NOT-A construction & disjunctive conjuction)

*(14) ***Shei*** *pa lauhu* ***ma*** *?*
(WH word & sentential clitic *ma*)

(15) Yijing ***pa-bu-pa*** *lauhu* ***ma*** *?*
(A-NOT-A construction & sentential clitic *ma*)

(16) Yijing pa lauhu ***haishr*** *shrtz* ***ma*** *?*
(disjunctive conjunction & sentental clitic *ma*)

From the above sentences, we can conclude that different kinds of interrogative elements cannot co-occur within simple sentences. Based on the analyses proposed in the previous section, we will provide adequate and straightforward accounts for this phenomenon.

2.1. A GPSG Analysis

Notice that syntactic categories in GPSG are partial functions from features to values. Defining categories this way has a natural consequence that no well-formed syntactic category may have different specifications for the same feature. Thus, the mutual exclusiveness of different kinds of interrogative information can be accounted for in GPSG by assuming each kind of interrogative element encodes one kind of specification of the feature QTYPE. Summarizing our encoding of Mandarin interrogative information in GPSG, the feature QTYPE and the set of its possible values are indicated below:

(17) feature value range feature class

QTYPE { YN, WH, DJ, A-NOT-A } FOOT

According to this analysis, the grammaticality of (11)-(16) can be nicely captured by FFP and unification. Owing to FFP, different kinds of interrogative specifications in a sentence will all percolate up to the matrix node and result in feature clash. Thus, all these sentences are ruled out as ungrammatical because of failure of unification.

2.2. An LFG Analysis

Taking a similar approach to GPSG, we attribute all kinds of interrogative information to the feature QTYPE. The encoding of this feature in different kinds of questions is summarized below:

(18) **Lexicon**

ma	CLIT	**(↑QTYPE)= YN**
pa-bu-pa	V	**(↑QTYPE)= A-NOT-A**
shei	N	**(↑QTYPE)= WH** **(↑QTYPE)= body bottom**
haishr	CONJ	**(↑QTYPE)= DJ** **(↑QTYPE)= body bottom**

So, the LFG account of the grammaticality of (11)-(16) is similar to that of GPSG in that they both resort to unification.

Thus, it is suggested that the seemingly complicated phenomena of the compatibility of Mandarin interrogative information can be straightforwardly accounted for with our analyses in unification-based formalisms.

III. The Scope of Percolation of Mandarin Interrogative Information

With an understanding of the compatibility of Mandarin interrogative information, we will further examine their behaviors within Mandarin complex sentences.

Consider the pair of contrasting sentences below:

(19) *Tamen **shiwang** [**shei** pa lauhu]* ?
they hope who fear tigers
' Who do they hope is afraid of tigers ? '

(20) *Tamen **tauluen** [**shei** pa lauhu].*
they discuss who fear tigers
' They discuss the topic that who is afraid of tigers. '

Although both sentences contain a WH-word *shei*'who', yet (19) must be interpreted as a direct question, and (20) must be interpreted as a statement taking an indirect question. The difference between (19) and (20) reveals an interesting phenomenon concerning the scope of percolation of Mandarin interrogative information. Again, we will discuss this topic within the GPSG and LFG frameworks.

3.1. A GPSG Analysis

As mentioned previously, the FFP in GPSG requires that all the FOOT feature specifications instantiated on a daughter be instantiated on its mother in any given local tree. Since our proposed interrogative features are all FOOT features, without additional stipulations, the interrogative messages should be passed to the top matrix node, rather than be limited in the embedded clause. But this prediction is contradictory to the empirical fact shown in (20).

According to Grimshaw (1979), it is suggested that the matrix verb of a sentence is responsible for the scope of interrogative information in its complement sentence.[8] Different kinds of verbs will result in different kinds of percolation of information. This idea has been widely adopted among researches on interrogatives. Here, we will following this line of approach and make a crucial use of the feature SUBCAT in our GPSG analysis.[9] In this section, we just take verbs ***shiwang***'hope', ***tauluen*** 'discuss' and ***jrdau***'know' as illustrative samples. Three ID rules are postulated as shown in (21):[10]

(21) **a. VP —> V[11], S – [QTYPE A-NOT-A]**

b. VP —> V[12], S[QTYPE]

c. VP —> V[13], S([QTYPE])

First, let us discuss the verb *shiwang*. We assume it is listed in lexicon as (22):

(22) < ***shiwang***, [N –], [V +], **[SUBCAT 11]**, ...] HOPE′ >

Consider the following sentences:

(23) *Tamen shiwang [Yijing pa lauhu].*
They hope Yijing fear tigers
' They hope that Yijing is afraid of tigers. '

*(24) *Tamen shiwang [Yijing pa lauhu ma] ?*
[QTYPE YN]

---

8 For ease of description, we use the term 'verbs' to stand for predicates in Mandarin.

9 The use of the feature SUBCAT is an important mechanism in GPSG whereby the relevant subclasses of a preterminal symbol can be matched with the ID rules that introduce it.

10 V[11] is just an abbreviation for V[SUBCAT 11].

(25) *Tamen shiwang [Yijing pa lauhu]* ***ma*** *?*
[QTYPE YN]

(26) *Tamen shiwang [****shei*** *pa lauhu] ?*
[QTYPE WH]

(27) *Tamen shiwang [Yijing pa lauhu* ***haishr*** *pa shrtz] ?*
[QTYPE DJ]

(28) Tamen shiwang [Yijing* ***pa-bu-pa* *lauhu] ?*
[QTYPE A-NOT-A]

*(29) ***Shei*** *shiwang [Yijing pa lauhu]* ***ma*** *?*
[QTYPE WH] [QTYPE MA]

*(30) ***Shei*** *shiwang [Yijing* ***pa-bu-pa*** *lauhu] ?*
[QTYPE WH] [QTYPE A-NOT-A]

(31)* ***Shei* *shiwang [Yijing pa lauhu* ***haishr*** *pa shrtz] ?*
[QTYPE WH] [QTYPE DJ]

(23) shows that *shiwang* can take a statement as its complement. The contrasting pair (24) and (25) show that the interrogative sentential clitic *ma* can only attach to a matrix sentence instead of an embedded sentence. This phenomenon has been discussed and accounted for in Shiu (1989: 33-41).[11] With the GPSG analyses proposed in Shiu (1989), *ma* will always function to form a direct question, and the specification [QTYPE YN] will be always interpreted at the level of matrix sentence. (26) and (27) show that although the [QTYPE WH] and [QTYPE DJ] specifications are introduced in the embedded sentences, they will percolate up to the matrix sentences by FFP, and make the whole sentences interpretated as direct questions. However, it is shown in (28) that [QTYPE A-NOT-A] cannot appear in the complement of *shiwang*. This fact can be nicely captured because *shiwang* is introduced by ID rule (20)a, in which the specification (– [QTYPE A-NOT-A]) is explicitly stipulated, and thus complements containing [QTYPE A-NOT-A] will be ruled out because of feature clash. (28) shows that if the matrix sentence has encoded one kind of interrogative message, the attachment of *ma* will cause unification of incompatible information and thus (29) is ungrammatical. Finally, in (30)-(31), both the matrix sentences and embedded sentences bear some kind of interrogative information. In these cases, except [QTYPE A-NOT-A],

[11] The GPSG analyses of *ma* proposed in Shiu (1989) are summarized below:
(i) lexicon
< *ma*, [[CLIT MA], [+ LAST], [QTYPE YN],...]>
(ii) ID rule
S′ —> S, [CLIT α]*
(iii) LP statement
x < [+ LAST]

other interrogative specifications encoded in embedded sentences will percolate up to matrix sentences and merge with the ones encoded in matrix sentences. Since a feature can have only one value, the grammaticalities of (30)-(31) will also be nicely accounted for.

Next, consider the verb *tauluen*. We assume this verb is listed in lexicon as (32):

(32) < ***tauluen***, [N -], [V +], [**SUBCAT 12**], ...] DISCUSS′ >

Let us consider the following sentences:

*(33) *Tamen (tzai) tauluen [Yijing pa lauhu].*
They (be Ving) discuss Yijing fear tigers

*(34) *Tamen (tzai) tauluen [Yijing pa lauhu]* ***ma*** *?*
[QTYPE YN]

(35) *Tamen (tzai) tauluen [****shei*** *pa lauhu].*
[QTYPE WH]

(36) *Tamen (tzai) tauluen [Yijing pa lauhu* ***haishr*** *pa shrtz].*
[QTYPE DJ]

(37) *Tamen (tzai) tauluen [Yijing* ***pa-bu-pa*** *lauhu].*
[QTYPE A-NOT-A]

(38) ***Shei*** *(tzai) tauluen [Yijing* ***pa-bu-pa*** *lauhu] ?*
[QTYPE WH] [QTYPE A-NOT-A]

(39) ***Shei*** *(tzai) tauluen [Yijing pa lauhu* ***haishr*** *pa shrtz].*
[QTYPE WH] [QTYPE DJ]

It is worth noting that *tauluen* obligatorily takes a question as its complement, as exemplified in (33)-(37). This can be achieved by the SUBCAT feature of *tauluen* and the ID rule in (21)b. As indicated earlier, the percolation of FOOT features in GPSG is manipulated by the FFP. But notice that the FFP governs only instantiated FOOT feature specifications.[12] Since the FOOT feature QTYPE in ID rule (21)b is inherited rather than instantiated, its behavior is not regulated by the FFP. As a consequence, all the QTYPE specifications encoded in embedded sentences will not be passed up to matrix sentences but rather be terminated within the embedded sentences. Thus, (35)-(37) are interpreted as indirect questions instead of direct questions. Further, (38)-(39) are not counterexamples to the proposals in previous section because no interrogative information will flow up from the embedded sentences and incompatible specifications do not co-occur in any categories in matrix sentences.

Last, let's turn to the verb *jrdau*. Its lexicon is shown in (40).

(40) < ***jrdau***, [N -], [V +], [**SUBCAT 13**], ...] KNOW′ >

[12] Readers are referred to (7) for the definition of the FFP.

We need to account for the following sentences with *jrdau*:

(41) *Tamen jrdau [Yijing pa lauhu].*
They know Yijing fear tigers
' They know that Yijing is afraid of tigers. "

(42) *Tamen jrdau [**shei** pa lauhu].*
[QTYPE WH]

(43) *Tamen jrdau [Yijing pa lauhu **haishr** pa shrtz].*
[QTYPE DJ]

(44) *Tamen jrdau [Yijing **pa-bu-pa** lauhu].*
[QTYPE A-NOT-A]

(45) *Shei jrdau [Yijing **pa-bu-pa** lauhu] ?*
[QTYPE WH] [QTYPE A-NOT-A]

(46) *Shei jrdau [Yijing pa lauhu **haishr** pa shrtz]?*
[QTYPE WH] [QTYPE DJ]

The verb *jrdau* can take either a statement or an indirect question as its complement. Thus we introduce it by (21)c, in which an optional QTYPE feature is specified. When *jrdau* takes a statement as its complement, the feature QTYPE is absent, but when it takes a question as its complement, the feature QTYPE is present. Thus, the grammaticality of (42)-(46) is accounted for in a way as we just discussed with *tauluen.*

Generally speaking, all the Mandarin verbs can be divided into these three classes, therefore the scope of percolation of Mandarin interrogative information is successfully accounted for in GPSG.

3.2. An LFG Analysis

Recall the LFG treatment of interrogative markers in previous section. We repeat the lexicon of these interrogative markers in (47):

(47) **Lexicon**

ma	CLIT	(↑LAST) = + (↑QTYPE) = YN
pa-bu-pa	V	(↑PRED) = 'FEAR<(↑SUBJ)(↑OBJ)>' (↑QTYPE) = A-NOT-A

shei	N	(↑QTYPE) = WH (↑QTYPE) = body bottom (↑PRED) = 'PRO' (↑HUMAN) = +
haishr	CONJ	(↑QTYPE) = DJ (↑QTYPE) = body bottom

Notice that both *shei* and *haishr* lexically encode an equation '(↑QTYPE)=body bottom'. This equation indicates a functional uncertainty device which is recently developed in LFG. The mechanism of functional uncertainty, explicated in Kaplan & Zaenen (in press), is originally proposed to account for long-distance dependencies in natural languages, such as topicalization and English WH questions. The basic idea of this mechanism is that long-distance dependencies are in fact functionally conditioned, and this kind of relation should be captured by a direct link between functions rather than through the mediation of local dependencies.[13]

The general rule of functional dependencies is formally expressed in Kaplan & Zaenen (in press), as shown in (48):

(48) $S' \rightarrow$ **Ω** **Σ**

(↑DF) = ↓ ↑ = ↓

(↑DF) = (↑body bottom)

[where **Ω** is a maximal phrasal category, **Σ** is some sentential category, DF is taken from the set of discourse functions (TOPIC, FOCUS, etc.), and body must be a regular expression.[14]]

The equation (↑DF)=(↑body bottom) in (48) is a functional uncertainty path in which any language can impose its own specific conditions on the functions of the body and the bottom only if the body is a regular expression.

This approach to long-distance dependencies is well supported by the study of Icelandic, English, and Japanese data. Huang et al. (1989), based on Mandarin topicalization and relative clauses, also suggests that functional uncertainty can provide an elegant solution to long-distance dependencies in Mandarin. In this paper, we use a reverse kind of functional uncertainty in resolving the percolation of interrogative information.

Mandarin interrogatives in fact do not involve overt long-distance dependencies. Unlike English WH questions, no gap-filler pairs can be found in any type of Mandarin questions. But as pointed out earlier, in some cases the existence of an interrogative element will turn the whole sentence into a direct question regardless of how deeply embedded the bearer of interrogation is. Thus, some bears of interrogation should be able to link to a f-structure

[13] The COMP to COMP movements in Transformational Grammar (TG) and the SLASH feature in GPSG are devices which try to account for long-distance dependencies through the mediation of local dependencies.

[14] A regular expression involves only the use of the Kleene closure operator, designated by '*', or the positive Kleene closure operator, designated by '+', on sets.

many layers up and theoretically there is no limit to the distance of such linking. In LFG, functional uncertainty is the mechanism to capture this kind of unbounded relation. But notice that there are two basic differences between the ordinary long-distance dependencies, such as topicalization, and the dependencies discussed in this section. First, as we have pointed out, Mandarin questions do not involve the so-called gap-filler relations, thus the functional uncertainty equations for them are not to specify the associations between the gap functions and the filler functions, but to ensure the interrogative feature QTYPE to be interpreted at the right places at f-structure. Second, Mandarin questions are characterized by the existence of bears of interrogation, but these interrogative elements do not occupy a specific position at surface structure, such as the sentence initual clause-external position for topic, therefore it is not appropriate to encode the functional uncertainty equations at c-structure rules such as (48). On the contrary, intuitively the functional uncertainty equations for Mandarin questions should be encoded in the lexicon of interrogative markers. Since the interrogative sentential clitic *ma* never occur in embedded sentences, no functional uncertainty path should be posed on it. As for the A-NOT-A construction, it is observed that its interrogative information never percolates to higher sentences, so no functional uncertainty path on this construction is necessary.[15] However, WH questions and disjunctive questions are not interpreted wholly locally. For example, consider the following sentences:

(49) *Dashiung jiuede tamen shiwang* ***shei*** *pa lauhu ?*
Dashiung feel they hope who fear tigers
' Who does Dashiung feel that they hope is afraid of tigers ? '

(50) *Dashiung jiuede tamen shiwang Yijing pa lauhu* ***haishr*** *pa shrtz ?*
Dashiung feel they hope Yijing fear tigers or fear lions
' Does Dashiung feel that they hope Yijing is afraid of tigers or is afraid of lions ? '

Though the WH word *shei* and the disjunctive conjunction *haishr* are encoded in embedded sentences, they turn the whole matrix sentences into direct questions. This phenomenon prompts us to propose a reverse kind of functional uncertainty equations which are encoded in the lexicon of WH words and *haishr* and can characterize the unbounded upward association between interrogative specifications. The general form of such equations is given in (51):

(51) **(↑QTYPE) = (body bottom)**

According to our observation, the **bottom** of the uncertainty path is the feature QTYPE, and the **body** of the path is a regular expression of the metavariable '↑'. The metavariable '↑' refers to the grammatical function represented by the mother node. Since the grammatical functions in LFG form a finite set, the **body** defined in this way is still a regular set.

[15] An apparent exception concerns a particular set of verbs, such as *tsai*'guess',and *shiang*'think', etc. Tang (1981,1983) call them "the semantically bleached verbs". These verbs cannot form A-NOT-A constructions, but if their complement sentences containing A-NOT-A constructions, the whole sentences are interpreted as direct questions. However, this type of verbs exhibit several other syntactic idiosyncrasies, such as their non-co-occurrence with aspect markers, their inability of constructing condensed answers by itself, etc. Since properties of this kind of verbs are not clear to us at this moment, the analyses of them are left open in this paper.

No particular difficulty will arised in solving the verification problem and the satisfactory problem of this kind of functional uncertainty.[16] Thus, the unbounded nature of Mandarin WH questions and disjunctive questions can be specified by the uncertainty equation given in (52):

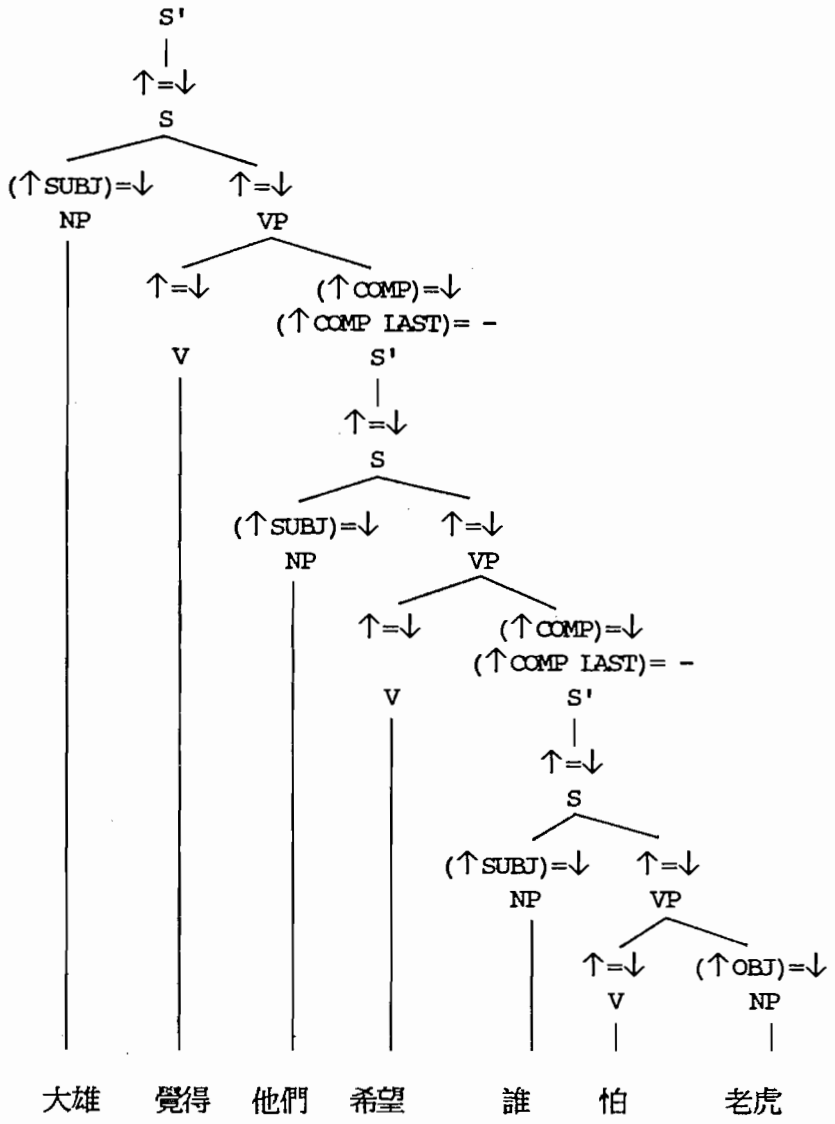
(52) **(↑QTYPE) = ({↑}* QTYPE)**

Under this approach of Mandarin interrogative information, the WH question in (49) and the disjunctive question in (50) will have correct c-structure and f-structure pairs as shown in (53) and (54) respectively.

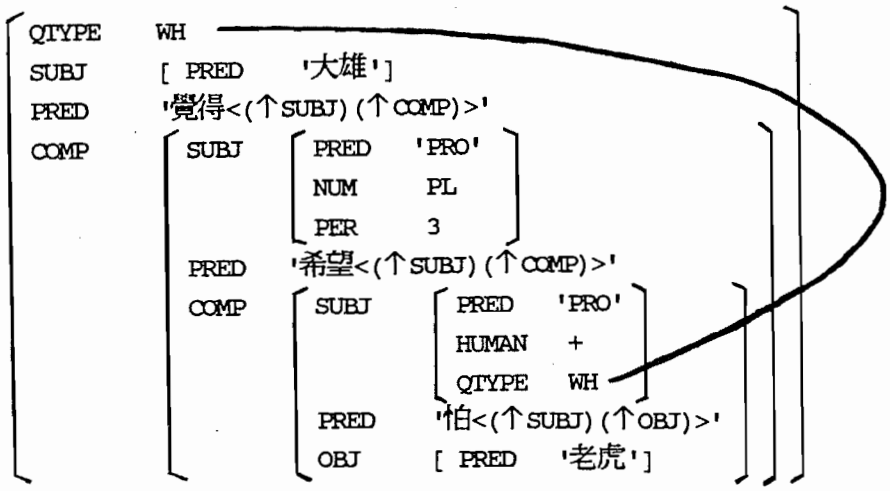
[16] An efficient algorithm for the verification and the satisfiability of functional uncertainty is proposed in Kaplan & Maxwell (1988 a).

(53) (for (49))

a. c-structure

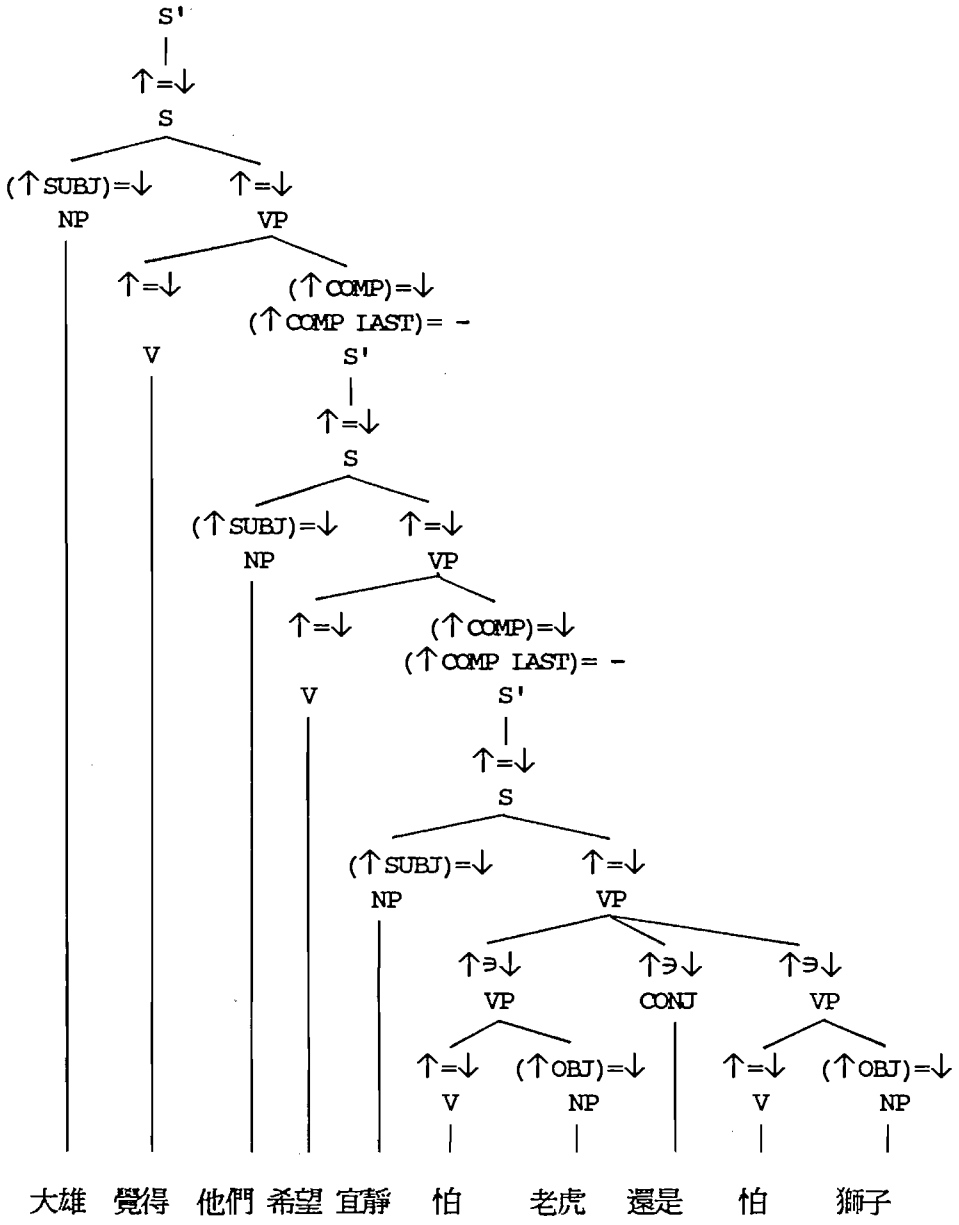


b. f-structure

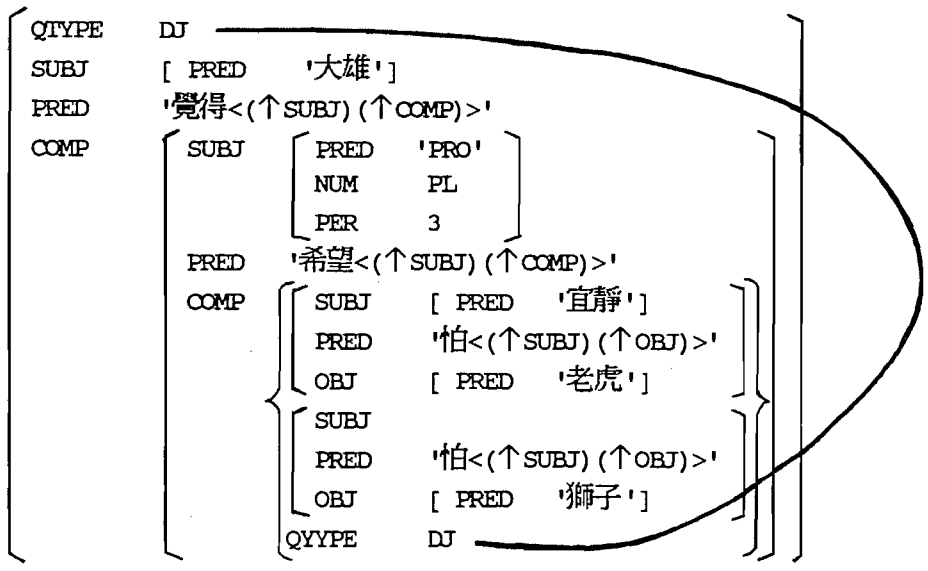


(54) (for (50))

a. c-structure



b. f-structure



However, the LFG analysis of Mandarin interrogative information above might appear to be still too general. As mentioned earlier, Mandarin verbs may impose their specific requirements on the sentence types of their complement sentences. Thus, the unbounded linking we proposed in (52) should be subject to the conditions encoded on verbs. Based on the same data as presented in section II A, we assume the verbs *shiwang*, *tauluen*, and *jrdau* encode different kinds of constraints as shown below:

(55) ***shiwang*** – **(| COMP QTYPE)**

(56) ***tauluen*** **(| COMP QTYPE)**

(57) ***jrdau*** **((| COMP QTYPE))**

(55) states that the feature QTYPE cannot be present at the complement function of the verb *shiwang*. Thus, the interrogative specifications encoded in embedded sentences must be linked to higher f-structures. On the other hand, the verb *tauluen* encodes an existential constraint which will ensure the presence of the feature QTYPE at the function of its complement sentence. Thus, *tauluen* must take a question as its complement, and this question is an indirect question because the feature QTYPE is just interpreted at embedded level. As for the verb *jrdau*, it can take either a statement or an indirect question as complement. Hence, an optional constraint is imposed on it.

In conclusion, we have successfully shown that the scope of interrogative information can also be adequately managed in LFG.

IV. Conclusion

This paper investigates the interrogative information of Mandarin questions. It is suggested that the compatibility and the scope of percolation of different kinds of interrogative information can be adequately and straightforwardly accounted for in GPSG and LFG. The GPSG analysis relies on the Foot Feature Principle (FFP) and the LFG analysis on functional uncertainty. However, from the comparative study we presented in this paper, readers may have noticed that the analyses in GPSG and LFG are quite similar. One important reason for their simility is that they are both unification-based formalisms. They agree with each other in taking feature-value pairs as their basic linguistic objects and in adopting unification as their basic operation. Owing to their simility, we are able to extract and compare the main concepts in them. Further, it is also easier to adopt ideas from the other theories to solve problems in their own. These merits of unification-based grammar formalisms have led many researchers to adopt this approach in their theoretical models as well as in their computational implementations.[17] Owing to the brievity of this paper, we just provide a preliminary unification-based study for Mandarin questions, but promising results on this topic can be expected along this line of research.

[17] Unification-based formalisms of theory type consist of LFG, GPSG, HPSG (Head-driven Phrase Structure Grammar), etc. and of tool type consist of PATR-II, FUG (Functional Unification Grammar), and DCG (Definite-Clause Grammar), etc.

References

Alleton, Viviane (1981). "Final Particles and Expression of Modality in Mandarin Chinese." Journal of Chinese Linguistics 9, 91-115.

Bresnan, Joan (1982) Ed. The Mental Representation of Grammatical Relations. Cambridge : MIT Press.

Chao, Yuen-Ren (1968). A Grammar of Spoken Chinese. Berkely : University of California Press.

Chinese Knowledge Information Processing (CKIP, 中文詞知識庫計劃) (1988). 國語的詞類分析(修訂版). 中研院計算中心: 台北 南港.

Chu, Chauncy C. (1983). "Ambiguities in Mandarin Verb Phrases — Cases with *Le* and *Shi-De*." Presented at Linguistic Conference on East Asian Languages. Univ. of Southern California.

—— (1987). "Semantics and Pragmatics of Modality in Mandarin." Manuscript. Univ. of Florida.

Falk, Yehuda N. (1983). "Constituency, Word Order, and Phrase Structure Rules." Linguistic Analysis 11, 331-360.

Gazdar, Gerald, Ewan Klein, Geoffrey Pullum, and Ivan Sag (1985). Generalized Phrase Structure Grammar. Oxford : Blackwell.

Grimshaw, Jane (1979). "Complement Selection and Lexicon." Linguistic Inquiry 10, 279-326.

Halvorsen, Per-Kristian (1982). "Lexical-Functional Grammar and Order-Free Semantic Composition." COLING 82, 115-120.

Huang, Cheng-Teh James (1982 a). "Move WH in a Language without WH Movement." Linguistic Review 1, 41-80.

—— (1982 b). Logical Relations in Chinese and the Theory of Grammar. MIT. dissertation.

Huang, Chu-Ren (1985). "Chinese Sentential Particles : A Study of Cliticization." LSA Annual Meeting, Dec. 1985.

—— (1987). Mandarin Chinese NP *de* : A Comparative Study of Current Grammatical Theories. Cornell Univ. Dissertation. Published as Special Publications No. 93. Taipei : Institute of History and Philology, Academic Sinica.

—— (1988 a). "A Unification-Based LFG Analysis of Lexical Discontinuity." The Fourth International Workshop on East Asia Languages and Linguistics, Paris, June 1988.

—— (黃居仁) (1988 b). " 聯併(Unification): 語法理論與剖析." Proceedings of ROCLING I, 29-54. Academia Sinica.

—— (1988 c). " 再析國語「領屬主語」結構 — 概化詞組結構語法(GPSG)及詞彙功能語法(LFG)之比較研究. " 漢學研究, 第六卷 第二期, 109-134.

—— (1988 d). "Towards a Morphosyntactic Account of Taiwanese Question Particle *Kam*." To appear in 史語所集刊 第五十九本.

——, Keh-Jann Chen, Wan-Pei Chen, & Tzu-Ying Hu (1989). "Resolution of Long-Distance Dependencies in Mandarin Chinese with an Algorithm Based on Functional Uncertainty." To appear in Proceedings of 1989 International Conference on Computer Processing of Chinese and Oriental Languages.

Kaplan, Ronald M. & John T. Maxwell (1988 a). "An Algorithm for Functional Uncertainty." COLING 88, 297-302.

——, & John T. Maxwell (1988 b). "Constituent Coordination in Lexical-Functional Grammar." COLING 88, 303-305.

——, & Annie Zaenen (in press). "Long-Distance Dependencies, Constituent Structure, and Functional Uncertainty." To appear in Alternative Conceptions of Phrase Structure. Eds. M. Baltin and A. Kroch. Chicago : University of Chicago Press.

Kay, Martin (1985). "Parsing Functional Unification Grammar." In Natural Language Parsing : Psychological, Computational, and Theoretical Perspectives. Eds. David Dowty, Lauri Karttunen, and Arnold M. Zwicky. 251-278. Cambridge Univ. Press.

Klavans, Judith (1982). Some Problems in a Theory of Clitics. Dissertation. Univ. College London. [Reproduced by IULC]

—— (1985). "The Independence of Syntax and Phonology in Cliticization." Language 61, 95-120.

Li, Charles N. and Sandra A. Thompson (1981). Mandarin Chinese : A Functional Reference Grammar. Berkeley : Univ. of California Press.

Lu, Jian-Ming (陸儉明) (1982). "由'非疑問形式+呢'造成的疑問句." 中國語文 6, 435-438.

—— (1984). " 關於現代漢語裡的疑問語氣詞." 中國語文 5, 331-337.

Nevis, Joel A. (1985). "A Syntactic Account of Second Position Clitics." In Proceedings of the Second Eastern States Conference on Linguistics. Eds. S. Choi, D. Danvitt, W. Janis, T. Mcloy, and Z.-S. Zhang. 186-197. Columbus, Ohio State Univ.

Sag, Ivan A., Gerald Gazdar, Thomas Wasow, and Steven Weisler (1984). "Coordination and How to Distinguish Categories." CSLI Report No. CSLI-84-4. Also in Natural Language and Languistic Theory 3.

——, Ronald Kaplan, Lauri Karttunen, Martin Kay, Carl Pollard, Stuart Shieber, and Annie Zaenen (1986). "Unification and Grammatical Theory." Proceedings of the Fifth West Coast Conference on Formal Linguistics. Stanford : Stanford Linguistic Association.

Sells, Peter (1985). Lectures on Contemporary Syntactic Theories. CSLI Lecture Notes, No. 3, Stanford : CSLI, Stanford Univ.

Sheu, Ying-Yu (1987). "Chinese Morphosyntax." Paper presented at the TwentiethInternational Conference on Sino-Tibetan Languages and Linguistics. Vancouver, Canada.

Shieber, Stuart M. (1986). An Introduction to Unification-based Approaches to Grammar. CSLI Lecture Notes, No. 4, Stanford : CSLI, Stanford Univ.

—— (1987). "Separating Linguistic Analyses from Linguistic Theories." In Linguistic Theory and Computer Applications. Eds. P. Whitelock, M. M. Wood, H. L. Somers, R. Johnson, & P. Bennett, 1-36. NewYork : Academic Press.

Shiu, Yu-Ling Una (1989). "Mandarin Sentential Particles and Particle Questions." Unpublished thesis. National Tsing Hua Univ.

—— & Chu-Ren Huang (1988). "Unification-based Analysis and Parsing Strategy of Mandarin Particle Questions." Proceedings of the International Computer Symposium, 38-43.

Tang, Ting-Chi (湯廷池) (1981). " 國語疑問句的研究." 師大學報 26, 219-277. also in Tang (1988 a), 241-311.

—— (1983). " 國語疑問句研究續論." 師大學報 29, 381-436. also in Tang(1988 a), 313-399.
—— (1984). " 國語裡「移動α」的邏輯形式規律." 教學與研究 6, 79-114. also in Tang (1988 a), 401-448.
—— (1988 a). 漢語詞法句法論集. 台北, 台灣學生書局.
—— (1988 b). " 普遍語法與漢英對比分析. " 第二屆世界華語文教學研討會論文集, 1309-1334.
—— (1989) " 普遍語法與英漢對比分析 第一章 「X標槓理論」與詞組結構." Manuscript. Tsing Hua Univ.
Wescoat, Michael T. (1987). "Practical Instructions for Working with the Formalism of Lexical Functional Grammar." Manuscript. Stanford Univ.
Zwicky, Arnold M. (1985). "Clitics and Particles." Language 61, No. 2, 283-305.
Zwicky, Arnold M., and Geoffrey K. Pullum (1983). "Cliticization vs. Inflection : English N'T." Language 59, No. 3, 502-513.