Using A Semantic Classification in Parsing Chinese: Some Preliminary Results

Kok Wee GAN
Department of Computer Science
Hong Kong University of Science & Technology
Clear Water Bay Road, Kowloon, Hong Kong
gankw@cs.ust.hk

Abstract

This paper describes a semantic classification of Chinese and compares its performance with CKIP's word-class classification on the task of identifying the attributive dependency relation and the unmarked coordinative dependency relation of word bigrams made up from nouns, verbs, and adjectives.

1 Introduction

In parsing Chinese, the practice of deriving the semantic structure of a sentence from word-class information faces two main challenges: (i) it is difficult to determine word classes due to the lack of inflections in Chinese; (ii) there is a lot of ambiguities in the mapping of word classes to the grammatical roles and to the semantic structure. Table 1 summarizes the types of dependency relation that could exist in word bigrams composed from nouns, verbs, and adjectives.

- Actant1Rel

The actant1 relation, denoted as Actant1Rel, connects between a lexeme whose meaning is a functor (the head) and its first argument (the dependent). Examples include:[1]

Actant1Rel(同意 'agree', 你 'you') Actant1Rel(大 'big', 氣量 'tolerance')

Actant1Rel(沒用 'useless', 說 'say') Actant1Rel(是 'is', 勤儉 'deligent')

- Actant2Rel

The actant2 relation, denoted as Actant2Rel, connects between a lexeme whose meaning is a functor (the head) and its second argument (the dependent). Examples include:

Actant2Rel(喝 'drink', 茶 'tea') Actant2Rel(感到 'feel', 親切 'cordial')

Actant2Rel(喜歡 'like', 游泳 'swim')

- MotionStateRel

The motion-state relation, denoted as MotionStateRel, exists between two predicates, the dependent (a predicate) expresses the result of the motion conveyed by the other predicate (the head).

MotionStateRel(拿 'take', 走 'walk') MotionStateRel(鬧 'make', 大 'big')

MotionStateRel(坐 'sit', 端正 'upright')

- AttrRel

The attributive relation, denoted as AttrRel, covers all kinds of modifiers:

AttrRel(杯子 'cup', 玻璃 'glass') AttrRel(字典 'dictionary', 袖珍 'pocket-size')

AttrRel(建設 'construct', 工業 'industry') AttrRel(正規 'regular', 非 'not')

AttrRel(方法 'method', 訓練 'train') AttrRel(延期 'extend', 停留 'stay')

- CoordRel

The coordinative relation discussed here, denoted as CoordRel, focuses only on unmarked coordinate constructions without any conjunction. The left-hand conjunct is assumed as the head. Examples are:

CoordRel(木桌 'wood table, 木椅 'wood chair') CoordRel(下雨 'rain', 刮風 'wind blow')

CoordRel(生動 'lively', 活潑 'vigorously')

In this paper, we propose a semantic classification of Chinese. We compare its performance against the

[1] In this paper, a dependency relation between two words is represented as: relation(head, dependent).

traditional approach on the task of identifying the attributive and the unmarked coordinative relations in word bigrams made up from nouns, verbs, and adjectives.

2 The Semantic Classification of Chinese

We propose a semantic classification of Chinese. This classification is derived by adopting the taxonomy of linguistic meanings proposed in the Meaning-Text approach (Polguere, to appear) and developing it further based on (Dong, 1994; 中文詞知識庫, 1993). Figures 1, 2, 3, 4, and 5 shows the details.

The object name and predicate classifications are adapted from EMICS (Dong, 1994). The difference is that our classifications are not as fine-grained as EMICS. We prefer to verify the adequacy of our classification with actual applications before working on a finer classification. The quantifier and operator classifications are adapted from (中文詞知識庫, 1993). The most general concept of quantifier is used here. Words of measure, for example, are interpreted as quantifiers. Their first actant being a lexical meaning. It is the quantified element as well as the dominant node of the utterance meaning. The second actant is an utterance meaning containing the first actant. Figure 6 shows an example of the graphical representation of a quantifier. Operators are words having utterances as the actancts. Figure 7 shows an example.

3 Word-Class Distributions

Eight articles on the theme of travelling from the Sinica Corpus V1.0 (中文詞知識庫, 1995) are used in the study. They comprises a total of 5882 Chinese characters. There are 46 classes in the CKIP tagset. This tagset is derived from (中文詞知識庫, 1993). In this paper, it is sufficient to simply state that tags which begin with "N" are noun classes; tags which begin with "V" are verb classes; non-predicative adjective is "A". Details about the CKIP tagset can be found in (中文詞知識庫, 1995). Word bigrams with word-class equals to noun, verb, and adjective are extracted. These bigrams are manually examined. Those with either an attributive relation or coordinative relation are retained.

In Chinese, the vast majority of adjectives may function as the head of a verb phrase (Li and Thompson, 1981). This type of adjectives has been classified as verbs in CKIP. Adjectives which cannot be the head of a verb phrase are called non-predicative adjective (非謂形容詞) in CKIP, and they are denoted by the tag "A". Examples are: 大型 'big-scale', 陽性 'positive', 新興 'newly properous', 真正 'truly', 永久 'permanent', 全盛 'most properous', and 上好 'best'. For verbs which could be used in a position that expects a nominal, they are not regarded as having two word classes: verb and noun. CKIP treats them as verbs. When they appear in a nominal position, they are tagged with an additional grammatical feature "+nom". For example, in the following bigrams:
(1) 參觀／VC+nom 門票／Na ("visit ticket")
(2) 管理／VC+nom 辦法／Na ("manage method")
In (1) and (2), the verbs 參觀 'visit' and 管理 'manage' are used as modifiers to a noun. They are therefore tagged with an addition grammatical feature '+nom'. Though CKIP has classified adjectives which may function as the head of a verb phrase as verbs, this class of verbs is handled differently with respect to the issue of nominalization. For example, in the following bigrams:
(3) 殘障／VH 人士／Na ("handicap people")
(4) 友好／VH 國家／Na ("friendly country")
the verbs 殘障 'handicap' and 友好 'friendly' are not tagged with the feature '+nom' . These verbs would traditionally be regarded as adjectives. From the different treatment of verbs of this class (VH) from verbs in the previous class (VC) when both function as modifiers to nouns, CKIP does make an implicit distinction between them.

Based on the word-class distributions of bigrams with attributive and coordinative relations, the following observations are noted.

- An attributive relation or a coordinative relation could exist in a word bigram with word class equals to 'VH VA'.

CoordRel(乖巧/VH 'clever', 盡職/VA 'dedicate') AttrRel(旅遊/VA 'tour', 深入/VH 'deep into')

- It is not clear why verbs such as 露營 'camp' and 憩息 'rest' in the following bigrams are tagged with the grammatical feature '+nom' while verbs such as 旅行 'tour' and 觀光 'sightseeing' are not, even though all of them have the same word class 'VA' and all of them appear in bigrams with an attributive relation. Similar cases are also observed for verbs in the VJ and VK class.

AttrRel(證/Na 'certificate', 旅行/VA 'tour') AttrRel(採購團/Na 'purchase team', 觀光/VA 'sightseeing')

AttrRel(車/Na 'car', 露營/VA+nom 'camp') AttrRel(需要/Na 'need', 憩息/VA+nom 'rest')

- A difficult issue in automatic word class tagging lies in nominalization, as exemplifed in the following examples.

AttrRel(門票/Na 'ticket', 參觀/VA+nom 'visit') Actant2Rel(參觀/VA 'visit', 博物館/Nc 'musuem')

AttrRel(車種/Na 'car type', 駕駛/VA+nom 'drive') Actant2Rel(駕駛/Va 'drive', 廂型車/Na 'van')

It is not clear how statistical information such as word ,cooccurrence frequency and word-class cooccurrence frequency are sufficient to decide whether nominalization exists in cases like 參觀 'sightseeing' and 駕駛 'drive', since there is no inflection in Chinese.

- It is not clear how unmarked coordinative relation could be identified with word-class information alone. We observed that a coordinative relation could exist in a word bigram with two different subcategories of verb, such as VA VCL, VA VH, VC VA, VH VA, and VL VC. However, it is also possible for an attributive relation to exist in a bigram with two different subcategories of verb, such as VA VF, VA VB, VH VA, VH VC, VH VCL, and VH VK.
- A parser to identify both the attributive and coordinative relations in our list of word bigrams based on word-class information alone will have an approximate baseline recognition rate of 45.6% (41/90) and 21.2% (3/14) respectively. These figures are derived by considering only those cases without ambiguity. For example, all bigrams made up of (A Na), (A Nc) and (VH *) are attributive relations. All bigrams with its constituents having the same word-class are coordinative relations.

4. Semantic-Class Distributions

The same list of word bigrams described in Section 3 are manually tagged with semantic classes. The following observations are noted.

- The constituents of word bigrams with a coordinative relation have the same semantic class.
- All attributive relations can be categorized as belonging to one of the following cases:

  a. Property Value-X: The first constituent belongs to the semantic class 屬性值 "property value", which describes some property value of the second constituent. The second constituent is labelled as "X", which means that there are no restrictions on the semantic class of the second constituent. There are 33 such cases. Examples include:
  優雅—女子 'elegant-lady' ， 大—有 'big-have' ､ 直接—受理 'direct-process'，新婚—夫妻 'newly married-couple'，賣力—引導 'diligently- guide'，好—消息 'good-news'，優惠—折扣 'preferential-discount'，順利—取得 'smoothly-obtain'，雙人—套房 'double-bedroom'，短暫—時間 'short-time' ，主要—行程 'main-journey'.

  b. X-Property: The second constituent describes some property of the first constituent. There are no restrictions on the semantic class of the first constituent. There are 8 such cases. Examples include:
  自身—安全 'self-safely'，服務—品質 'service-quality'，獵遊—收費 'hunting tour-charge' ，購物—折扣 'shopping-discount'.

  c. State-X: The first constituent belongs to the semantic class 狀態 "State" and there are no restrictions on the semantic class of the second constituent. There are 10 such cases. Examples include:
  休息—營區 'rest-camp site'，退除役—官兵 'retired-official'，懷舊—色彩 'remember old time-color'，認可—程度 'agree-degree'.

  d. Relation-X: The first constituent belongs to the semantic class 關系 "Relation" and there are no restrictions on the semantic class of the second constituent. There are 6 such cases. Examples include:
  不同—價格 'different-price' ，相關—證件 'relevant-document'，駐外—機構 'station out-organization'.

  e. Change of Relation-X or Change of State-X: The first constituent either belongs to the semantic class 變關系 "Change of Relation" or 變狀態 "Change of State". There are no restrictions on the semantic class of the second constituent. There are 11 such cases. Examples include:
  越區—申請 'cross border-apply' ，洗衣—服務 'washing-service'，查詢—電話 'enquiry-telephone'，購物—折扣 'shopping-discount'.

  f. The last category has the modifiers specifying the scope or range of the head. The most generic sense of scope and range is used here. There are 17 cases here.
  野餐—活動區 'camping-activity area'，觀光—採購團 'visit-purchase team'，露營—帳蓬

'camping-tent'，申請一事宜 'apply-incident'，行駛一路徑 'drive-route'，管理一部門 'manage-department'，管理一辦法 'manage-method'，護照一簽名 'passport-signature'，市區一遊覽 'city-tour'，國際一通用 'international-valid'，資料一供應 'information-supply'，客房一服務 'guest room-service'.

- When all the word bigrams with attributive and coordinative relations are considered together with all the other bigrams with the Actant1Rel and Actant2Rel relations (total of 256 bigrams), only two cases of ambuities are found.

Actant2Rel(夜宿／住 'night stay/living', 園區／空間 'campsite/space')	AttrRel(營區／空間 'campsite/space', 休息／住 'rest/living')
Actant2Rel(申請／做 'apply/do', 簽證／讀物 'visa/reading')	AttrRel(文件／讀物 'document/reading', 申請／做 'apply/do')

- Conceptually, it is possible that an Actant1Rel exists in a bigram of the class 'X-Property'. Let us assume that a parser is not able to identify all AttrRel relations in this class because of such ambiguity. Assuming further that the parser is also unable to identify AttrRel relations where the modifiers are generic scope/range of the head. The parser would still be able to achieve a recognition rate of 72.2% (65/90). Assuming that the simplistic strategy of assigning a coordinative relation if the constitutents of a bigram have the same semantic class. The accuracy of identifying coordinative relations in our sample will be 100%.

5. Conclusion

In this paper, we reported some preliminary results of using semantic classes to derive two dependency relations in Chinese word bigrams: the attributive and coordinative relations. We compared the results with the word classes used in the Sinica corpus. The results have suggested that a drastic improvement in parsing accuracy could be achieved with the semantic classification. Our second step now is to develop a statistical parsing model for Chinese based on this semantic classification.

Acknowledgements

This work has benefited from many discussions with Prof. Dong Zhendong. We would like to express our thanks to him.

References

Dong, Zhendong (1994) A Categorization System Of Chinese Movement Concepts. International Conference on Chinese Computing '94; page 145-149.

Li, Charles N., Thompson, Sandra A. (1981) Mandarin Chinese: a functional reference grammar. University of California Press, 1981.

Polguere, Alain (to appear) Meaning-Text Semantic Networks As A Formal Language. To appear in Current Issues in Meaning-Text Linguistics; Wanner Leo (Ed.)

中文詞知識庫（1993）中文詞類分析（三版）；中央研究院資訊科學研究所技術報告 93-05。

中文詞知識庫（1995）中央研究院平衡語料庫的內容與說明：中央研究院資訊科學研究所技術報告 95-02。

Table 1: Examples of dependency relations

Dependency Relation	Word-Class	Examples
Actant1Rel	Noun Verb	你同意 'you agree'
	Verb Verb	說沒用 'say useless'
	Noun Adjective	氣量大 'tolerance big'
	Adjective Verb	勤儉是 'diligent is'
Actant2Rel	Verb Noun	喝茶 'drink tea'
	Verb Verb	喜歡游泳 'like swim'
	Verb Adjective	感到親切 'feel cordial'
MotionStateRel	Verb Verb	拿走 'take walk'
	Verb Adjective	鬧大 'make big' 、坐端正 'sit upright'
AttrRel	Noun Noun	玻璃杯子 'glass cup'
	Noun Verb	工業建設 'industry construct'
	Verb Noun	訓練方法 'train method'
	Verb Verb	停留延期 'stay extend'
	Adjective Noun	袖珍字典 'pocket-size dictionary'
	Adjective Adjective	非正規 'not regular'
	Adjective Verb	認真學習 'serious learn'
CoordRel (unmarked)	Noun Noun	木桌木椅 'wood table wood chair'
	Verb Verb	下雨刮風 'rain wind-blow'
	Adjective Adjective	生動活潑 'lively vigorously'

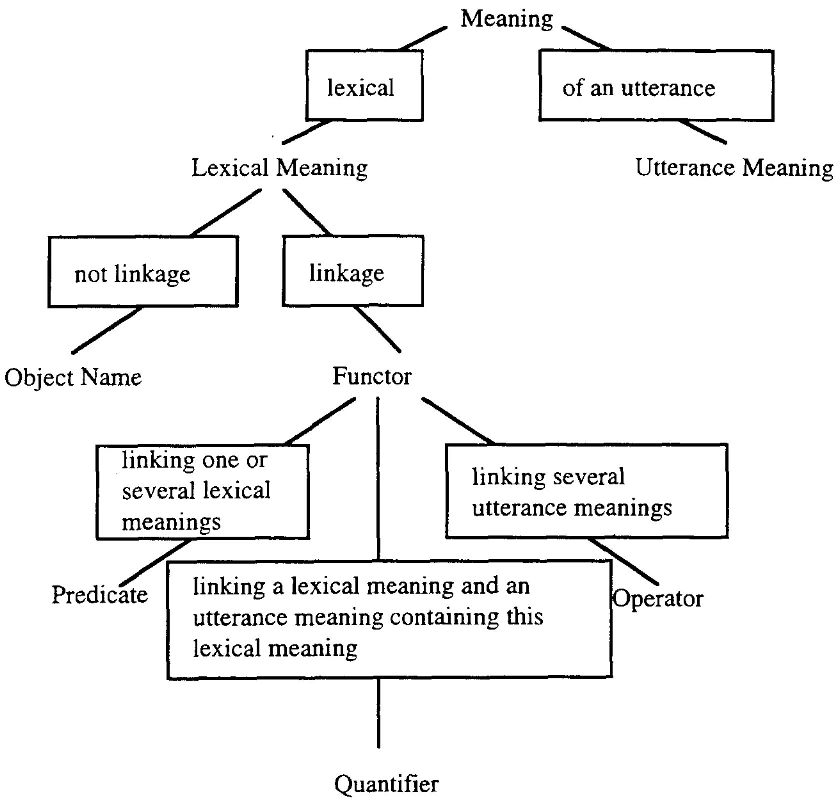
Meaning
lexical
of an utterance
Lexical Meaning
Utterance Meaning
not linkage
linkage
Object Name
Functor
linking one or several lexical meanings
linking several utterance meanings
Predicate
linking a lexical meaning and an utterance meaning containing this lexical meaning
Operator
Quantifier

Figure 1: Taxonomy of linguistic meanings

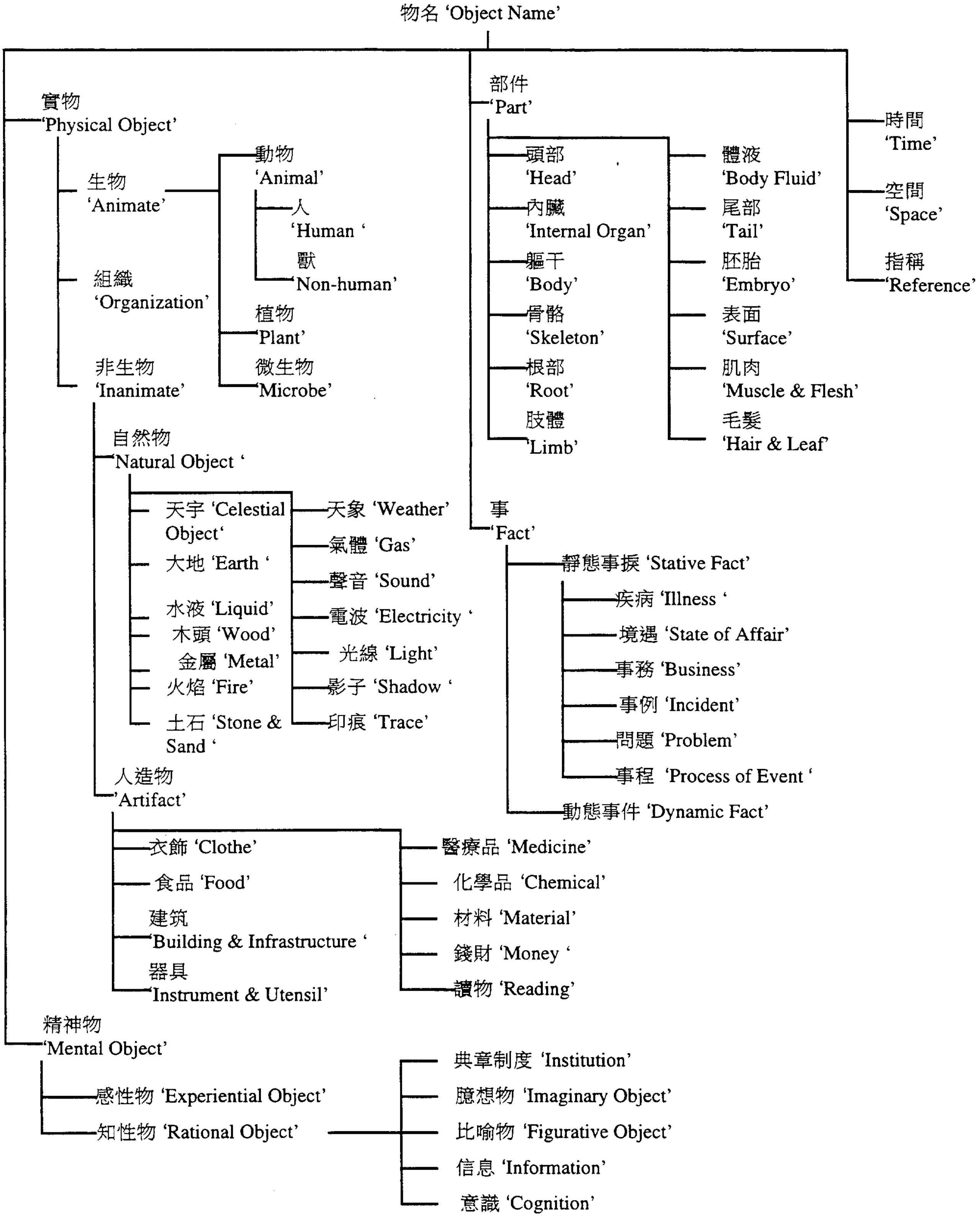

Figure 2: Object Name Classification

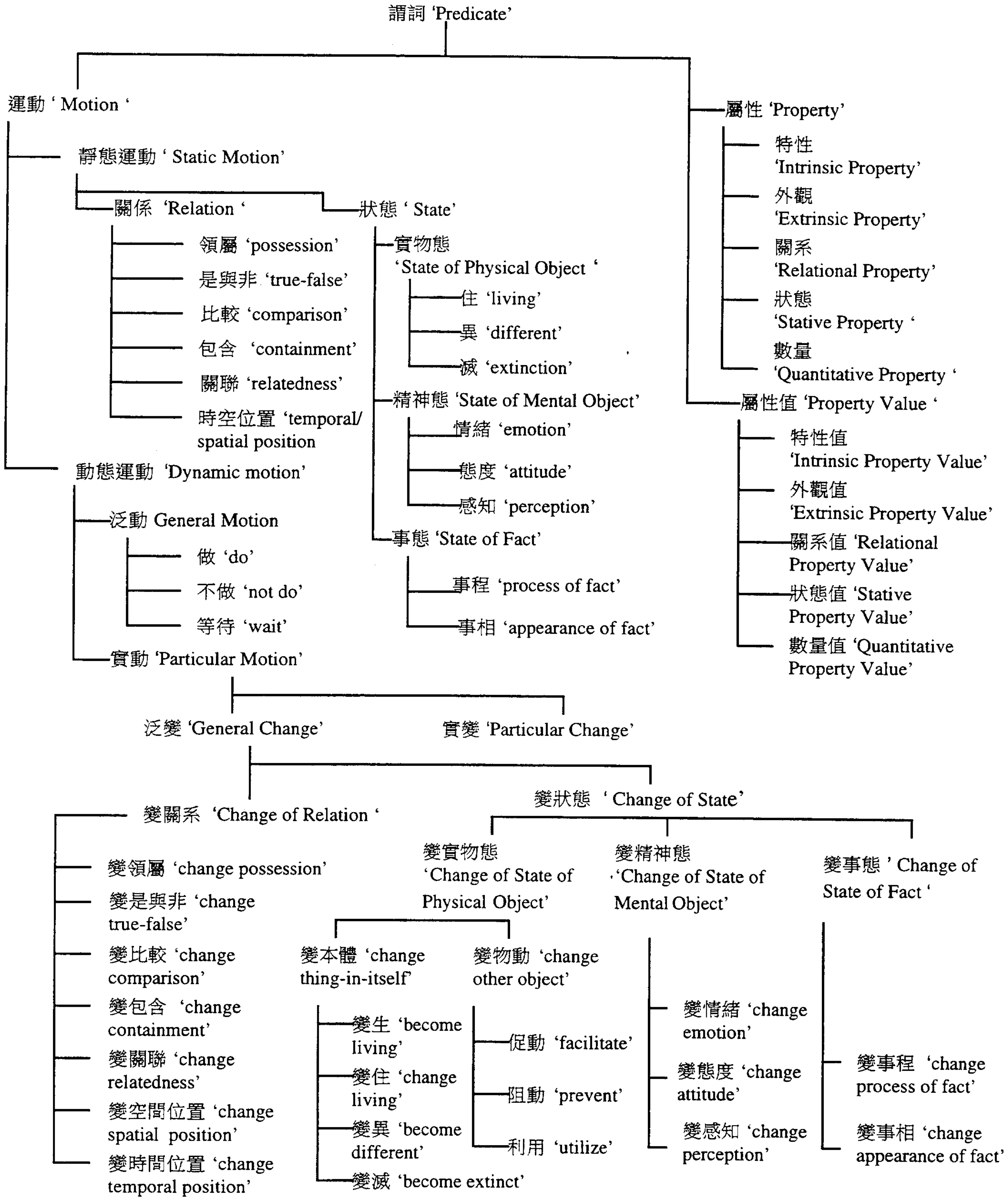
謂詞 'Predicate'
運動 ' Motion '
靜態運動 ' Static Motion'
關係 'Relation '
領屬 'possession'
是與非 'true-false'
比較 'comparison'
包含 'containment'
關聯 'relatedness'
時空位置 'temporal/ spatial position
狀態 ' State'
實物態 ' State of Physical Object '
生 'living'
異 'different'
滅 'extinction'
精神態 'State of Mental Object'
情緒 'emotion'
態度 'attitude'
感知 'perception'
事態 'State of Fact'
事程 'process of fact'
事相 'appearance of fact'
屬性 'Property'
特性 'Intrinsic Property'
外觀 'Extrinsic Property'
關系 'Relational Property'
狀態 'Stative Property '
數量 'Quantitative Property '
屬性值 'Property Value '
特性值 'Intrinsic Property Value'
外觀值 'Extrinsic Property Value'
關系值 'Relational Property Value'
狀態值 'Stative Property Value'
數量值 'Quantitative Property Value'
動態運動 'Dynamic motion'
泛動 General Motion
做 'do'
不做 'not do'
等待 'wait'
實動 'Particular Motion'
泛變 'General Change'
實變 'Particular Change'
變關系 'Change of Relation '
變領屬 'change possession'
變是與非 'change true-false'
變比較 'change comparison'
變包含 'change containment'
變關聯 'change relatedness'
變空間位置 'change spatial position'
變時間位置 'change temporal position'
變狀態 ' Change of State'
變實物態 'Change of State of Physical Object'
變本體 'change thing-in-itself'
變生 'become living'
變住 'change living'
變異 'become different'
變滅 'become extinct'
變物動 'change other object'
促動 'facilitate'
阻動 'prevent'
利用 'utilize'
變精神態 'Change of State of Mental Object'
變情緒 'change emotion'
變態度 'change attitude'
變感知 'change perception'
變事態 ' Change of State of Fact '
變事程 'change process of fact'
變事相 'change appearance of fact'

Figure 3: Predicate Classification

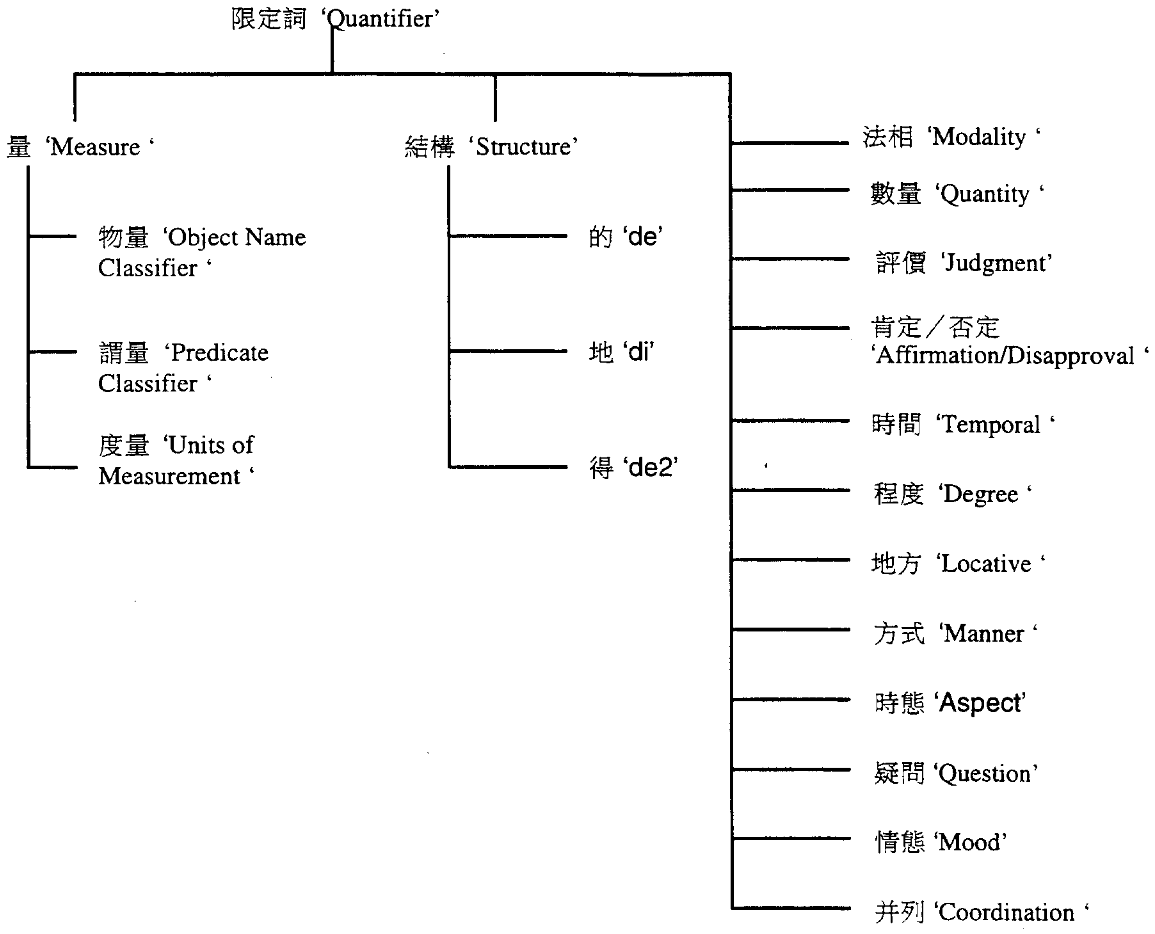

Figure 4: Quantifier Classification

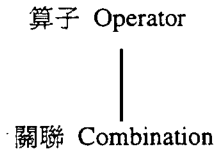

Figure 5: Operator Classification

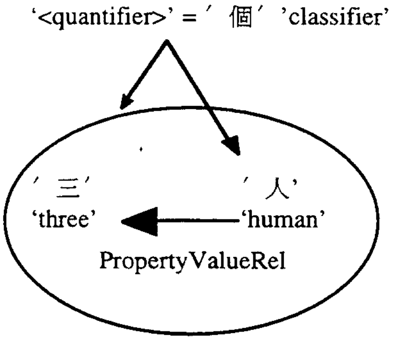

Figure 6: An example of the graphical representation of quantifiers

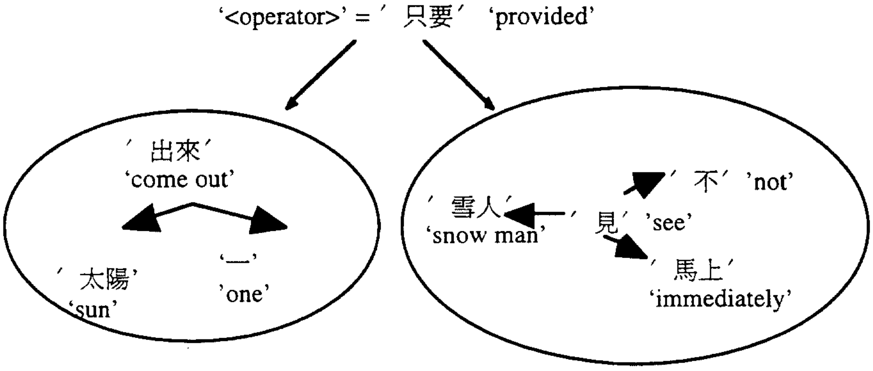

Figure 7: An example of the graphical representation of operators