

Taxonomy of Fine-grain Semantic Roles for Nominal Modifiers

Su-chu Lin, Shu-Ling Huang, Keh-Jiann Chen
CKIP, Institute of Information Science, Academia Sinica, Taipei
{jess, josieh}@hp.iis.sinica.edu.tw, kchen@iis.sinica.edu.tw

Abstract

To assign semantic roles in building Treebanks, there is a need for annotators having a guideline in determining semantic relations between phrasal head and its modifiers or arguments. Semantic roles are hard to have clear-cut definitions. It is not always easy to determine thematic relations between two concepts. This paper aims to introduce an integrated nominal modifier system. Basically we adopt other scholars' incisive idea in analyzing semantic roles that modify general nouns. We use the approach of building a fine-grain taxonomy of role system. The taxonomy of fine-grain thematic roles makes the role determination easier for human annotators, since the meaning of a fine-grain semantic role is self explanatory and a higher-level semantic role is described by its hyponyms. The proposed taxonomy has been attested during construction of Sinica TreeBank and HowNet definitions of nominal concepts and proven to be more applicable than conventional flat structures.

1 Introduction

To assign semantic roles in building Treebanks, there is a need for annotators having a guideline in determining semantic roles. Semantic roles establish the semantic relations between phrasal head and its modifiers or arguments. The verbal modifiers describing the semantic roles of an event are much discussed in various kinds of research, and FrameNet is the most representative among them. However discussions regarding semantics roles modifying nouns are scattered.

Semantic roles are hard to have clear-cut definitions. For instance, it is not always easy to determine thematic relations between two concepts. Dowty (1991) proposed criteria for determining proto-agent and proto-patient which means sometimes it is hard even to differentiate between agent role and patient role. We use the approach of building a fine-grain taxonomy of role system. The taxonomy of fine-grain thematic roles makes the role determination easier for human annotators, since the meaning of a fine-grain semantic role is self explanatory and a higher-level semantic role is described by its hyponyms. Another important advantage of taxonomic approach is that any flat fine-grain role system suffers the problem of completeness of role coverage, but a taxonomic role system can always find a coarse-grain hypernym role if a proper fine-grain role doesn't exist in the system.

2 Background

According to Pustejovsky(1991, 1995), the defining attributes of an object is named as the qualia structure. Pustejovsky argues that the qualia structure of a word tells us the concept it carries in a language. The qualia is further specified into four aspects: constitutive, formal, telic and agentive. He defines constitutive "the relation between an object and its constituents". Formal in qualia distinguishes the object within a larger domain. The role telic shows the purpose and function of the object. And agentive are factors involved in how the denotation of the word came into being. The structured representation which gives relational force of a lexical item is well-discussed in qualia by Pustejovsky(1995).

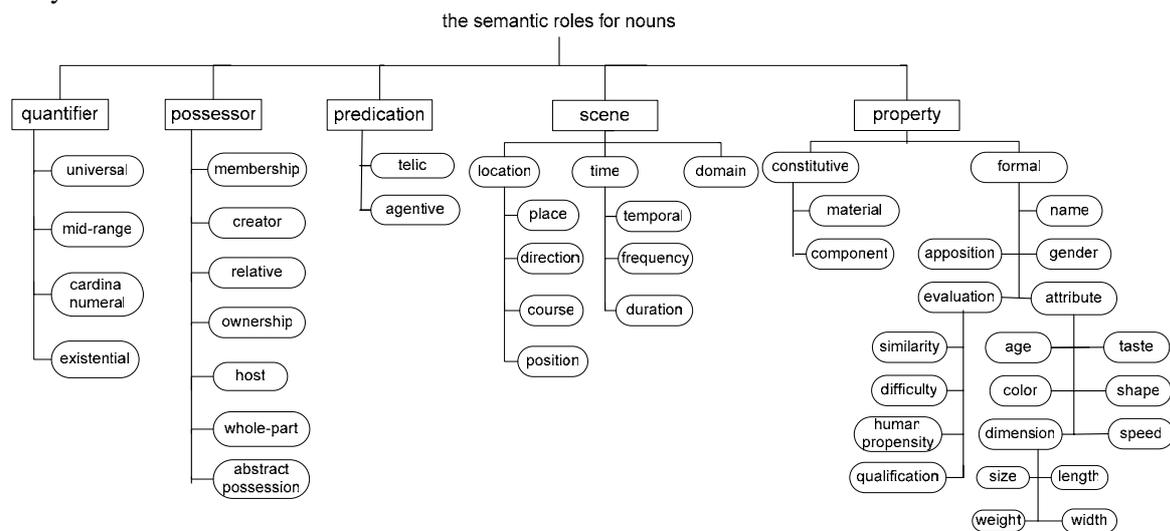
Barker(1998) put forth 20 noun modifier relationships (NMRs) in his semi-automatic program¹. However the individual NMR in this parallel framework tends to seem incomplete or fragmented. The

¹ The NMRs in Braker (1998) include "Agent (agt), Beneficiary (benf), Cause (caus), Container (ctn), Content (cont), Destination (dest), Equative (equa), Instrument (inst), Located (led), Location (loc), Material (matr), Object

firmness of the NMRs is to be challenged since one can add or delete a NMR easily. This is resulted from the scattered elements. Thus we decide to adopt a hierarchical order in our taxonomy of fine-grain semantic role for nominal modifiers.

3 Our analysis

At top level of taxonomy, we divide the semantic role for nouns into five categories: predication, quantifier, possessor, scene, and property. The following is our preliminary framework of fine-grain semantic role for nominal modifiers. The five categories are observable in real language and our data would not stray from these five. The subcategories in each role may not be exhausted or perfect, but they demonstrate a clearer picture on how the five roles function. We will introduce the skeleton of our analysis in the sections followed.



3.1 Possessor

Within the five, possessor and property are the two roles we get confused easily. We encounter some difficulties in assigning the semantic roles. The idea of possession seems to have less problem in examples like, *wo de ya sui qian* 我的壓歲錢 'my spring festival money gift' or *zi ji de yi fu* 自己的衣服 'one's own clothes'. The possessees in the two examples are separable from the possessors and the concept of owning something is vivid. Nonetheless, we hesitate to give the possessor role on some data like *zu fu de yang zi* 祖父的樣子 'the form of grandfather' and *yu fu de shang huo* 漁夫的生活 'the fisherman's life'. Does the grandfather own the form and does the fisherman have the life?

Heine (1997) compares many previous studies about possession in human languages. He uses criteria as alienable and inalienable, temporary and permanent, animate and inanimate, and abstract possession to subcategorize possession into seven. Heine's analyses are clear and ordered. But for us, some of the criteria are not so important as to be distinguished from the top. For instance, the physical possession, temporary possession, and permanent possession² in his analysis are combined to ownership in ours. The ownership means the alienable possession with an inanimate possessee. Heine leaves the

(obj), Possessor (poss), Product (prod), Property (prop), Purpose (purp), Result (resu), Source (src), Time (time), Topic (top)".

² According to Heine the definition of the three is as follows. Physical possession: the possessor and the possessee are physically associated with one another at reference time. Temporary possession: the possessor can dispose of the possessee for a limit time but cannot claim ownership to it. Permanent possession: the possessee is the property of the possessor, and typically the possessor has a legal title to the possessee.

inalienable possession³ to a single category. However, we consider the body-parts and relative relation which are described as inalienable possession in his study to be different; on the contrary, we think that body-parts (inalienable possession) and part-whole relationship (inanimate inalienable possession⁴) are similar, since they denote parts of their respective substances. Thus we have whole-part relation, head noun being a part of the modifier as well as relative in our study. In Heine's discussion, possession includes the inanimate alienable possession, such as the sentence *The tree has crows on it*. Yet we settle this group of data in scene for "the tree" is actually a position and the possessing connection between the two is not so close.

3.2 Property

We subdivide the property into constitutive and formal, following Pustejovsky's work. The constitutive includes the material and component parts of the noun. The phrase *yu mi nong tang* 玉米農湯 'corn puree' is one of the constitutive examples. Formal is describing the target noun's evaluation, attribute, apposition, gender or name. For instance, in *kun nan de ren wu* 困難的任務 'difficult task', *kun nan* is a difficulty evaluation of formal. We refer to Dixon (1982) on the features of attributes and evaluations of nouns⁵.

3.3 Quantifier

Gil (2001) distinguishes the internal and external quantification. In his studies, the external quantifiers is in the scope of syntax and thus beyond our discussion. His analysis of internal quantification is briefed as follows:

1. mass vs. count quantifiers
2. existential vs. universal quantifiers
3. numeral
4. others.

Similarly our quantifier is subdivided into four, including existential, universal, mid-range quantifiers and numerals. Existential quantifiers mean that there exists something or someone to make the proposition real. The word *you ge* 有個 'some, certain' can be an existential quantifier which modifies *nian tou* 念頭 'idea'. Universal quantification refers to the range within the group, as the word *quan bu* 全部 'all' modifying *sai cheng* 賽程 'game'. The massive degree in-between existential and universal is called mid-range quantification, like *shao shu* 少數 'few', *da duo* 大多 'most' and others. Then the last type of the quantification is the numeral quantifiers. As a classifier language, the numeral quantifiers in Mandarin usually denote exact numbers along with classifiers as *si tiao* 四條 'four' is a possible numeral quantifier of *yu* 魚 'fish'.

3.4 Scene

Lu (2001) uses "scene" in depicting the eventive case roles⁶. We adopt Lu's idea of scenic description and diminish it to adjust the nominal modifiers. The scenic description of location, time and domain are three self-explaining roles of nominal modifiers. The word *Taipei* is a locative use in *Taipei de jie yun* 台北的捷運 'Taipei's MRT', while the *san xiao shi* 三小時 'three hours' describes the duration of *dian ying* 電影 'movie'. Meantime, in *xuan mei huo dong* 選美活動 'beauty contest', we see that *xuan mei* is the domain of *huo dong*.

³ Inalienable possession: the possessee is conceived of typically as being inseparable from the possessor, e.g. body-parts, relative..

⁴ Inanimate inalienable possession: part-whole relationship, inanimate possessor. The possessee and the possessor are conceived of as being inseparable

⁵Dixon (1982) classifies English adjectives as dimension, age, value, color, physical property, human propensity, speed, difficulty, similarity, qualification.

⁶See Lu (2001) for further discussion.

3.5 Predication

The predication denotes the head noun would play an argument role rather than adjunct. The predication is an eventive modifier of the head noun. Pustejovsky's agentive aspect of qualia falls in the predication role of our system. Examples of predication are as below: *ni zuo de huo gou* 你做的火鍋 'the hot pot you made'. We can see that *ni zuo de* 你做的 'you made' shows how the *huo guo* 火鍋 'hot pot' come into being, and also 'the hot pot' is the object in the clause of 'you made'. As to the telic use of predication, usually means that there is an omitting verb between the modifier and the head noun, such as *shui guo pan* 水果盤 actually means the plate used for placing fruits. The simplified phrase is without the predicate "used for placing". We arrange this kind of modifiers in telic use. Thus form the two predication use of the head noun.

4 Future work

We consult previous studies on each subcategory of the semantic roles for nouns, look for evidence form corpus and then modify the framework to fit Mandarin. The proposed taxonomy has been attested during construction of Sinica TreeBank and HowNet definitions of nominal concepts and proven to be more applicable than conventional flat structures. In the future we expect to make a complete fine-grain role system for not only nominal but also eventive semantic roles. Furthermore, we would like to compare and contrast even unify our work with other similar projects, such as FrameNet and HowNet.

References

- Barker, Ken & Stan Szpakowicz. 1998. Semi-Automatic Recognition of Noun Modifier Relationships. *Proceedings of COLING-ACL '98*. Montréal: 96-102.
- Chappell and McGregor. 1996. *The grammar of inalienability: A typological perspective on body part terms and the part-whole relation*. Berlin, New York: Mouton de Gruyter.
- Dixon, R.M.W. 1982. *Where have all the adjectives gone?* Mouton Publishers, Berlin.
- Dowty, David R. 1991. Thematic Proto-roles and Argument Selection. *Language*, 67(3):547-619.
- Gil, David. 2001. Quantifiers. In *Language typology and language universals : an international handbook*, edited by Martin Haspelmath et al.
- Heine, Bernd. 1997. *Possession: cognitive sources, forces and grammaticalization*. Cambridge: Cambridge University Press.
- Li, N Charles and Sandra A. Thompson. 1989. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley: University of California Press.
- 魯川 (Lu, Chuan). 2001. 漢語語法的語意網絡. 北京：商務印書館.
- Nastase, Vivi and Stan Szpakowicz. 2003. Exploring Noun-Modifier Semantic Relations. *Proceedings of International Workshop on Computational Semantics*. Tillburg, Netherlands.
- Nichols. 1988. On alienable and inalienable possession. In *In honor of Mary Haas: from the Haas festival conference on native American Linguistics*, ed. by Shipley. Berlin: Mouton de Gruyter.
- Pustejovsky, James. 1995. *The Generative Lexicon*. Cambridge: MIT Press.