LEXICAL INFORMATION AND BEYOND: CONSTRUCTIONAL INFERENCES IN SEMANTIC REPRESENTATION

Mei-Chun Liu, National Chiao Tung University
Chu-Ren Huang, Academia Sinica
Ching-Yi Lee, National Tsing Hua University

Mei-Chun Liu (劉美君)
Department of Foreign Languages and Literatures
National Chiao Tung University, Hsinchu 300, Taiwan
Email: mliu@cc.nctu.edu.tw

Abstract

This paper aims to show that semantic representation of verbs may require the inclusion of constructional inferences in addition to lexical specifications. By examining the transitive pattern of the Mandarin verb *GAN* (趕) 'rush' , we found that verbal semantics can only be adequately represented if constructionally coerced information is taken into consideration. The construction [*GAN* + NP] renders specific interpretations that are not directly derived from the lexical meaning of either the verb or the object NP. The construction itself carries salient information for the appropriate interpretation. Besides outlining a compositional approach with Qualia Structure (Pustejovsky 1995) , this paper attempts to account for the construction-triggered meanings from the perspective of Construction Grammar and to explain their interrelationship with cognitive mechanisms for sense extension.

1. INTRODUCTION

Knowledge representation of the verbal system has always been one of the core issues in both theoretical and computational linguistics. In the early generative paradigm, verbs are considered to be the structural head of the sentence since it subcategorizes its arguments and hence dictates the form of the sentence. Verb meanings are treated as general tendencies in selectional preferences, and the semantic details of individual verbs are largely neglected. Recent development in lexical research has shifted the focus to investigating the interdependency between verb meanings and syntactic behavior. Most lexical semanticists share a common assumption that *the syntactic behavior of a verb, especially its argument expression, is determined by the meaning of the verb* (cf. Levin, Song and Atkins 1996, Pustejovsky 1995, Levin 1993, Atkins and Levin 1991, Atkins et al. 1988, etc.). Levin (1993) tries to categorize English verbs into semantically distinct classes on the basis of their argument alternation patterns. Pustejovsky (1995)'s program on generative lexicon moves one step further in spelling out a multi-layered representational scheme for lexical information that includes Argument Structure, Event Structure, Qualia Structure, and Inheritance Structure. His goal is to fully represent the interaction of word meaning and its composition.

Another commonly shared belief among lexical semanticists is that the traditional 'sense enumerative lexicon', as Pustejovsky puts it, is inadequate to capture the real picture of natural

language semantics. Specifically, such an approach cannot account for three lexically relevant phenomena: the creative use of words, the permeability of word senses, and the expression of multiple syntactic forms with a single word (Pustejovsky 1995: 39). It is argued that a more powerful compositional mechanism with predicative strength is needed to allow coercion of lexical meanings into more complicated or innovative senses.

In view of the fact that lexical knowledge involves more than just the listing of different senses or simple combination of word meanings, we would like to present in this paper an interesting case in Mandarin where a single verb gives rise to various contextualized meanings when combined with a complement NP. The issue under study is the transitive use of *GAN* (趕) 'rush' in the pattern of [*GAN* + NP], such as *GAN gongche* 'the bus' or *GAN baogao* 'a paper'. The appropriate interpretation of the pattern requires incorporation of semantic components beyond the purely lexically-specified information. It is proposed that to fully represent the semantics of the pattern, constructional inferences should be included in the knowledge representation of lexical information.

2. THE DATA

The main body of data for the analysis comes from the Sinica Balanced Corpus of both written and spoken contemporary Mandarin, containing a total of 5 million word, developed by the CKIP group in Academia Sinica, Taiwan. The total number of occurrences of *GAN* in the corpus is 54, and the transitive use occupies 32 % of the data.

3. INITIAL OBSERVATIONS

3.1 Temporal Reference in the Use of *GAN*

The characterization of *GAN* is highlighted when it is compared with its peer synonym *ZHUI* (追). We found that one typical use of *GAN* requires a nominal sense extension as well as a temporal reference. Compare the following examples, where *ZHUI* and *GAN* can both take 'bus' as its direct object but render completely different interpretations:

(1) a. 追公車   'chase the bus'
   ZHUI gongche
   chase bus

  b. 趕公車   'rush to catch the bus'
   GAN gongche
   rush bus

Example (1a) is a prototypical instance of transitive predication where the verb *ZHUI* may take any movable, spatial-physical entity as its direct object, and the object itself is the goal for the action of chasing. On the other hand, (1b) represents a rather peculiar and semi-fixed pattern in the sense that the nouns *GAN* may take as direct object do not fulfill the typical Goal-role and the meaning associated with the pattern are highly specified. The NPs following *GAN* are nominals that invoke an event reading and are contextually associated with a time frame. Therefore, in (1b), *GAN gongche* should not be interpreted as a simple event of spatial movement;

instead, it describes "an action of rushing in order to get to the bus stop by a certain TIME".
Such a contrast between the different roles played by the direct object NP is captured as follows:

(2) Difference in the Semantic Role of the Direct Object NP

	Role of the Direct Object NP
ZHUI	Spatial Goal
GAN	Temporal Target

3.2 Event-Evoking Nominals in the use of *GAN*

A further contrast between the use of *ZHUI* and *GAN* leads us to the conclusion that *GAN* not only requires a temporal reference, but it is followed only by nominals that evoke an event reading.   Consider the following pair of examples where the verb is followed by a measured path:

(3) a. 追了　兩百公尺
ZHUI le liang-bai gongchi
chase LE two-hundred meter
'chased (someone) for two hundred meters'

  b. 趕了　兩百公尺的路/*兩百公尺
GAN le liang-bai gongchi de lu/*liang-bai gongchi
rush LE two-hundred meter DE journey / two-hundred meter
'rushed two-hundred meters to finish the journey.'

While it is perfectly fine to add a measured length ('two hundred meters') with *ZHUI*, it is not allowed with *GAN*.   But, when the measure is embedded as a modifier to describe the length of an undertaken journey as shown in (2b), it can then follow *GAN*.   The reason why the reading of 'making a specific journey' is required with the use of *GAN* is, we suspect, that *GAN* has to be associated with some kind of an ACTIVITY, evoked or implicated by the complement NP.   In other words, [*GAN*+NP] actually describes an under-specified EVENT that involves a temporally-framed activity.   Only if a nominal is capable of evoking an event reading may it follow the verb *GAN*.

This observation can be further supported by the fact that the utterance *GAN le liang-bai gongchi* (趕了兩百公尺) 'rushed two hundred meters' can actually be accepted if it refers to an activity that can be 'measured out' (Tenny 1992).   A context of making ropes/strings would be perfectly suitable for such an utterance, as provided below:

(3) A: 你繩子做好了嗎？
ni shezi zuo hao le ma
you rope make done LE Q

  B: 我已經趕了兩百公尺了
wo yijing gan le liang-bai gongzhi le.
I already GAN LE two-hundred meters PERF

A: 'Have you finished your rope-making'
B: ' I have speeded up and finished two hundred meters (of rope).'

-29-

The point we try to make here is that the range of nominals that *GAN* may take, compared with that in the use of *ZHUI*, is rather limited and highly specified. In the following sections, we will concentrate specifically on the semantic properties associated with the transitive use of *GAN*.

4. CONSTRUCTIONAL INFERENCES IN THE CASE OF *GAN*

As shown above, the verb *GAN* is rather restrictive in its argument selection with respect to the transitive pattern [*GAN* + NP]. In the corpus, there are two major types of NPs associated with the pattern, and each facilitates a unique interpretation of the construction.

4.1 Types of NPs in [*GAN* + NP]

The NPs following *GAN* can be either Animate or Inanimate. Animate NPs refer mostly to humans or animals, as shown in (4). Inanimate NPs can be further divided into four sub-groups, as shown in (5):

(4) Animate NPs:
趕 人/蚊子/麻雀/牛/羊
GAN ren/wenzi/maque/niu/yang
'Drive away people/mosquitoes/sparrows/cows/sheep'

(5) Inanimate NPs:
a. Scheduled special events:
趕 集/考/廟會/演講
GAN ji/kao/miaohui/yanjiang
'Rush to take part in the market/exam/religious celebration/public speech'

b. Vehicles running on a fixed schedule:
趕 公車/飛機
GAN gongche/feiji
'Rush to catch the bus/airplane'

c. Lexically specified (overt) time expressions:
趕 時間/三點半/進度
GAN shijian/san-dian-ban/jindu
'Rush to save time/to get to the bank by 3:30/to catch up with projected progress'

d. Artifacts to be produced by a deadline:
趕 報告/作業/課/衣服/貨
GAN baogao/zuoye/ke/yifu/huo
'Rush to finish writing a paper/writing homework/teaching classes/making clothes/manufacturing products'

The two types of NPs share the same surface form - [*GAN* + NP], but render quite different interpretations.

4.2 Interpretational Distinction

As discussed above, the construction [*GAN* + NP] describes a telic event, that is, the event is bounded with a target state. With Animate object-NPs, the verb *GAN* encodes a purposeful action that aims to 'move the object IN or OUT of a designated area'. In the example of *GAN ren* (趕人), the agent is performing an action to drive *'people'* out of a certain place; while in the example of *GAN yang* (趕羊), it's more likely that the agent is driving the sheep into an area (e.g. the pen). The action of *GAN* results in a spatial movement or locational change on the part of the object-NP.

With Inanimate object-NPs, however, [*GAN* + NP] describes a bounded event whose interpretation is much more complicated. The interpretation reads like: 'to achieve a STATE by a certain TIME through engaging in an ACTIVITY'. It involves at least three meaning components: an ACTIVITY performed by the agent, a STATE associated with the object, and a TIME FRAME. In the example of *GAN baogao* (趕報告), the underspecified activity is 'writing a paper', the goal is to get the paper done, and the time frame is the deadline for submitting the paper. What's interesting is that all the information about the 'ellipsed' activity/goal/time is missing from the purely lexically-specified meaning of either the verb or the nominal. The interpretational distinction between the two types of [*GAN* + NP] can be captured as follows:

(6) Animacy and Interpretational Difference in x [*GAN* + NP(y)]

	Interpretation
With Animate NP	Target State = (in-or-out (y, Loc))
With Inanimate NP	Target State = (x reaches State (y) by Time (t) through Activity (w))

4.3 Other Salient Information with Inanimate NPs

Given the interpretational complexity in the pattern [*GAN* + Inanimate NP], more has to be said concerning the details of its semantic representation. We found that the semantics of the inanimate pattern should also include the following:

● **Presupposition**: Normal pace of performing the activity is not enough (already being late with respect to the TIME frame).
● **Manner**: with accelerating pace
● **Telicity**: The event is bounded with a projected endpoint (the target state).
● **Agent-control**: The activity is completely under the agent's control.

This last feature captures the fact that the agent of *GAN* has to be the actor undertaking the implicated activity. For example, *GAN yifu* (趕衣服) cannot be taken as 'rushing someone to finish making the clothes', but can only be said of a dressmaker or tailor.

The issue here is that all the above semantic specificities and the information about the "ellipsed" activity/goal/time in the [*GAN* + Inanimate NP] pattern cannot be obtained directly

from the lexicon.   Only when combining the verb *GAN* with a potentially event-evoking inanimate nominal is the information automatically generated.

5. INFORMATION BEYOND THE VERB: EXPLANATORY FRAMEWORKS

How can we explain and represent the fact that the construction [*GAN* + NP] contains more information than what is specified in the lexicon for the individual word.   There are two ways of dealing with it.   One is the construction-based approach, taking [*GAN* + NP] as a form-meaning pair.   The other is the compositional approach that allows coercion of lexical information.

5.1 Construction-Based Approach

From the perspective of Construction Grammar (Goldberg 1995, Fillmore and Kay 1993), a 'construction' can be viewed as a meaning-bearing unit, i.e. the syntactic configuration itself 'contributes semantic content above and beyond that contained in the constituent lexical items' (Jackendoff 1997:553).   A construction (or constructional idiom) represents a form-meaning association, that functions, in a broader sense, just like a lexical item.   The meaning of a construction is learnable in the same way word meanings are learned.   Taking a semi-fixed VP pattern [*GAN* + NP] as a constructional idiom allows as to specify the unique semantic features associated with the pattern which go beyond the semantics of its constituent words.

Given the fact that two types of NPs may occur with *GAN* and each is associated with a peculiar interpretation, we may propose that there are two distinct Constructions:

(7)
a.   Construction 1
    Form: [*GAN* + Animate NP]
    Meaning: a spatially bounded event [to reach a Target State by causing the NP to move In or Out of a designated space].
    Example: 趕人 '*GAN*-people'

b.   Construction 2
    Form: [*GAN* + Inanimate NP]
    Meaning: a temporally bounded event [to reach a Target State (associated with the NP) · through speeding up in an Activity (agent-control) with a Temporal Reference (contextually defined or world knowledge)]
    Example: 趕作業 '*GAN*-homework'

An alternative solution is to propose that a Constructional Polyseme be recognized, as specified and represented as follows:

(8)   Constructional Polyseme :
Form: [*GAN* + NP]
a.   Sense Association 1
    Meaning: a spatially bounded event [to reach a Target State by causing the NP to move In or Out of a designated space].
    Specification on NP slot: animate, self-movable entities

Instantiation: 趕人'*GAN*-people'

b. Sense Association 2
   Meaning: a temporally bounded event [to reach a Target State through speeding up in
         an Activity (agent-control) with a Temporal Reference (contextually defined or
         world knowledge)]
   Specification on NP slot: Event-evoking NPs (Activity Nominals or Time
         References that stand for a default activity/event.)
   Instantiation:
   - 趕作業 '*GAN*-homework' (NP-Nominal Activity)
      ACTIVITY [writing the homework] by TIME [deadline]
   - 趕三點半 '*GAN*-3:30' (NP-Time)
      ACTIVITY [rushing to the bank] by TIME [3:30]

   Although the associated meanings are quite distinct, the construction is considered to be
polysemous in the sense that there seems to be a conceptual link between the two senses.   Sense
1 highlights a intended movement in space, which is utilized to describe the progression of event
in time.   This confirms the general cognitive principle that spatial movement usually provides
the conceptual basis for expressing temporal or eventive processes (The Localist Hypothesis,
Lyons 1977).

5.2   Compositional Approach: Qualia Structure

   Pustejovsky (1995) proposed a generative, multi-leveled framework to represent lexical
information.   The four levels in the structured representation are Argument Structure (for the
representation of adicity information for functional elements), Event Structure (for the
representation of information related to Aktionsarten and event type, in the sense of Vendler
1967), Qualia Structure (for the representation of the defining attributes of an object), and
Inheritance Structure (for the representation of the relation between the lexical item and others in
the lexicon).   Among the four levels, Qualia Structure is responsible for the relational force of a
lexical item. It specifies four essential aspects of a word's meaning:

- Constitutive: the relation between an object and its constituent parts;
- Formal: that which distinguishes it within a larger domain;
- Telic: its purpose and function;
- Agentive: factors involved in its origin or 'bringing it about'.

   As Pustejovsky (1995: 87-88) made it clear that Qualia Structure, as defined above, not only
characterizes our knowledge of words, but also suggests interpretations of words in context. The
English verb *enjoy* is used as an illustration.
Consider the examples below:

   (9)   a. *Mary enjoyed the movie last night.* (watching)
        b. *John quite enjoys his morning coffee.* (drinking)

The different contextual meanings of *enjoy* or the 'ellipsed' activities are supplied by information
from the complement.   More specifically, the Telic roles for *movie* and *coffee* project the
activities of *watching the movie* and *drinking coffee*, respectively, to the interpretation of the VP.

-33-

In other words, Qualia structure provides a compositional means for meaning coercion based on characterizations of the four different roles.

The analysis of English *enjoy* parallels the case of Mandarin *GAN* with Inanimate NPs in that [*GAN* + Inanimate NP] also involves an ellipsed activity whose information is provided by the object-NP. Through coercion of one of the qualia roles, the information about the Activity or Time Frame can be integrated into the interpretation of the verb phrase, as illustrated below:

(10) Qualia Representaion
  a. *GAN gongche* 趕公車 'rush to catch the bus'
    Bus [Telic = running on a fixed schedule]
  b. *GAN baogao* 趕報告 'rush to finish the paper'
    Paper [Agentive = *writing*]

What is interesting here is that Qualia Structure can also solve potential ambiguity problems in some cases. Take 趕比賽 '*GAN-Ball Games*' as an example, which allows two different interpretations, each with a different qualia role:

(11) Two Different Interpretations with 趕比賽 '*GAN-Ball Games*'
  趕了三場比賽.
  GAN le san-chang bisai.
  a. 'Rush to play three games'
  b. 'Rush to watch three games.'

(12) Different Roles for 比賽 '*(Ball) Games*'
  a. [Telic = entertaining/watching]
  b. [Agentive = playing]

The distinction between Telic and Agentive role is also crucial in deriving the correct interpretation of [*GAN* + NP] where the NP is associated with verbs of creation, i.e. the NP comes into being through verbs of creation, such as '(*writng*) *a book/song*'. In the examples below, only the Agentive interpretation is allowed:

(13) Agentive interpretation
  a. 趕書
    GAN shu
    'rush to write/*read the book'

  b. 趕歌
    GAN ge
    'rush to write/*sing the song'

  c. 趕戲
    GAN xi
    'rush to produce/*watch the show'

With the specified attributes or roles in the Qualia Structure, contextual meanings or

information beyond individual lexical items can be facilitated through semantic coercion, without the cost of additional explanatory or representational mechanisms.

6. OTHER RELATED CASES

The semantics of [*GAN* + NP] requires incorporation of constructional inferences that are not inherent in the lexical semantics of its constituent words. In a similar vein, there are other interesting cases of transitive verbs whose object-NPs suggest different "activities or events" that are not explicitly expressed but deemed important in semantic representation. Below we list two of such verbs:

(14) 搶 'rob, vie for'
a. 搶 錢   'rob money or rush to earn money'
b. 搶 金牌   'trying the best to win the gold medal'
c. 搶 位子   'rush to occupy a seat'
d. 搶 股票   'rush to buy stock'
e. 搶 時間   'rush to earn time'

(15) 玩 'play, enjoy'
a. 玩遊戲   'playing a game'
b. 玩吉他   'playing the guitar'
c. 玩橋牌   'playing the bridge
d. 玩女人   'womenizing'

There must be more Mandarin verbs displaying a similar behavior, in addition to the two listed above. A unified approach, once adopted, can be applied to all these verbs.

7. CONCLUSION

This paper presents an interesting case for semantic representation of Mandarin verbs. The appropriate interpretation of the construction [*GAN* + NP] involves information beyond lexical semantic specification of its constituent words and thus requires additional mechanisms for semantic integration. To fully represent the information, two equally applicable approaches are suggested in this paper. The first is based on proposals of Construction Grammar, in which "constructions" are treated as lexical items which are uniquely associated with a given "meaning". In this view, [*GAN* + NP] will be listed in the lexicon as a meaning-bearing unit with all the information needed. The second approach is introduced in Pustejovsky (1995), where a Qualia Structure is used to specify four different meaning facets (Constitutive, Formal, Telic and Agentive attributes) for each lexical item, and contextual interpretation can be borne out through attribute coercion. As a pilot work on Mandarin lexical semantic research, this study has the following implications:

● Mandarin lexical semantic studies are advancing but are still in a primitive stage. More comprehensive investigation is needed to identify a set of crucial semantic attributes as well as compositional principles that have syntactic consequences.
● Constructional inferences as discussed above call for an unified and systematic treatment

when dealing with semantic representation of Mandarin.

8. REFERENCES

[1] Atkins, B.T. and Levin, B. Admitting Impediments in.Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon, ed. by U. Zernik. Hillsdale, NJ: Lawrence Erlbaum Associates. 1991.

[2] Atkins, B.T., J. Kegl, and B. Levin. Anatomy of a verb entry: from linguistic theory to lexicographic practice. International Journal of Lexicography 1: 84-126, 1988.

[3] CKIP group (Chen, Keh-jiann, Li-li Chang, Chu-Ren Huang, and Ching-Chun Hsieh). A classification of Chinese verbs for language parsing." Proceedings of 1988 International Conference on Computer Processing of Chinese and Oriental Language (ICCPCOL), 414-17.Toronto, Canada. 1988.

[4] Croft, William. Possible verbs and the structure of events. In meanings and Prototypes: Studies in Linguistic Categorization, ed. by S. Tsohatzidis. London/New York: Routledge. 1990.

[5] Dowty, David. Thematic proto-roles and argument selection. Language 67:547-619, 1991.

[6] Goldberg, Adele. Constructions: A Construction Grammar Approach to Argument Structure. Chicago: The Univ. of Chicago Press. 1995.

[7] Givon, T. English Grammar: a Function-based Introduction, Vol. 1. Amsterdam/Philadelphia: John Benjamins Publishing Co. 1993.

[8] Her, One Soon. Grammatical Functions and Verb Subcategorization in Mandarin Chinese. Taipei: Crane Publishing Co. 1990.

[9] Huang, Chu-ren. and Tsai, Mei-chih. From near synonyms to event structure: Corpus-based studies of Mandarin verbal semantics. Paper presented at the Mini-Conference on Lexical Semantics. Graduate Institute of Linguistics. National Chung Cheng University. 1997.

[10] Huang, Shuanfan. Emegent lexical semantics. Paper presented at the Mini Conference on Lexical semantics. Graduate Institute of Linguistics. National Chung-Cheng University. 1997.

[11] Jackendoff, R. Twistin' the night away. Language 73 (3): 534-559, 1997.

------. Semantic Structures. Cambridge: MIT Press. 1990.

[12] Lakoff, George. Women, Fire, and Dangrous Things. Chicago: U of Chicago Press. 1987.

[13] Langacker, R. Foundations of Cognitive Grammar, vol 1. Stanford: Stanford U Press. 1987.

------. Foundations of Cognitive Grammar, vol. 2. Stanford U, Press. 1991.

[14] Levin, Beth. Verb Classes and Alternation. Chicago: U of Chicago Press. 1993.

[15] Levin, Beth, Grace Song, and B.T.S Atkins. Making sense of corpus data: A case study of verbs of sound, in International Journal of Lexicrophy. 1996.

[16] Liu, Meichun. To appear. Lexical Meaning and Discourse Patterning - the three Mandarin cases of 'build'. In the 3rd Volume of Conceptual Structure, Discourse, and Language. Stanford: CSLI.

------. When endpoint meets endpoint: a corpus-based lexical semantic study of Mandarin verbs of 'throwing'. Paper presented in IACL-7/NACCL-10, Stanford University. 1998.

------. Conceptual basis and categorial structure: a study of Mandarin V-R compounds as a radial category. In Chinese Languages and Linguistics, Vol. IV: 425-51. Taipei: Academia Sinica. 1997.

------. Conceptual manipulation and semantic distinction: the case of Mandarin postverbal DE complement. In Conceptual Structure, Discourse, and Language, ed. by Adele Goldberg, 333-46. Stanford: CSLI. 1996a.

------. A Pilot Study on Chinese Verb Classes and Alternations. NSC Project Report (NSC85-2418-H-009-003). 1996b.

[17] Lyons, John. Semantics. Cambridge: Cambridge Unversity Press. 1977.

[18] Pustejovsky, James. The Generative Lexicon. Cambridge: The MIT Press. 1995.

------. The syntax of event structure. Cognition 41:47-81. Smith, C. S. 1991. The Parameter of Aspect. Dordrecht: Kluwer Academic Publishers. 1991.

[19] Teng, Shou-hsin, ed. Chinese Synonyms Usage Dictionary. Taipei: Crane Publishing Co. 1994.

[20] Tenny, Carol. The aspectual interface hypothesis. Center for Lang. & Information. Stanford, CA; xx, 328 pp, 1992.

[21] Tsai, Mei-chi, Chu-ren Huang, and Keh-jiann Chen. You jinyici bianyi biaozhun kan yuyi, qufa zhi hudong. (From near-synonyms to the interaction between syntax and semantics), paper presented at IsCLL-5, Taipei, Taiwan. 1996.

[22] Tsai, Mei-chi, Chu-Ren Huang, Keh-jiann Chen, Kathleen Ahrens. Towards a representation of verbal semantics - an appraoch based on near synonyms. Proceedings of ROCLING X, 34-48, Hsinchu, Taiwan. 1997.

[23] Tsao, Feng-fu. On verb classification in Chinese. Paper presented in the 6th North American Conference on Chinese Linguistics (NACCL VI). 1994.