A Comparative Study of English and Chinese Synonym Pairs: An Approach based on The Module-Attribute Representation of Verbal Semantics

Kathleen Ahrens
Graduate Institute of Linguistics
National Taiwan University
Taipei, Taiwn
ahrens@ms.cc.ntu.edu.tw

Chu-Ren Huang
Institute of Linguistics
Academia Sinica
Nankang, Taipei, Taiwan 115
churen@gate.sinica.edu.tw

1   Introduction

The Module-Attribute Representation of Verbal Semantics (MARVS) is a theory of the representation of verbal semantics that is developed based on Mandarin Chinese data (Huang et al., 2000). This theory proposes two types of modules: event structure modules and role modules, as well as two sets of attributes: event-internal attributes and role-internal attributes which are linked to the event structure module and role module respectively. These module-attribute semantic representations have associated grammatical consequences.

Studies in MARVS (e.g. Biq 2000) found that the composition of an event modules and its attested lexical semantic attribute(s) could be generalized to a natural semantic class. For example, the contrast between *bai3* ('set') and *fang4* ('put') in Mandarin Chinese to do with the fact that *bai3* has a role-internal feature of [design] attached to the location role, while *fang4* does not. Moreover, this contrast can be generalized across the semantic class of verbs that involve design on the focussed location role (i.e. *hua4* 'to paint') and those that do not (i.e. *tu2* 'to cover with paint, to doodle'). What we would like to determine in this paper is if similar contrasts be found in near synonyms of other languages. In particular, are event modules in other languages (such as English) organized along similar conceptual lines?

In this paper, we examine this theory in light of the English data. In particular, we will look at the near synonym contrast of the verbs 'put' and 'set' based on data from the two million word sampler of the British National Corpus (BNC). To preview our results, we find that the event structure in English is slightly different from that of Chinese. In Chinese, *bai3* has a role-internal feature of [design] attached to the location role, while *fang4* does not. However, in English, 'put' has the roles of Agent, Theme and Location in its event structure, while 'set' has the three roles of Agent, Theme, and Proposition. Thus, conceptualizations of 'set' and 'put' in English and Mandarin have different semantic and syntactic entailments.

2   Methodology

Our methodology in analyzing near synonym pairs contrasts with that of Levin (1993), who concentrates on analyzing the range of possible syntactic alternations of a single verb (or a single verb class), and extracting semantic information from syntactic behavior. However, there are two reasons

we have not followed Levin in examining the relationship between a verb alternation and its associated semantics. First, although the work done by Levin (1993) in this area is impressive (determining 50 different types of alternations and over 125 different semantic classes of verbs), the sheer number of possible permutations of alternations makes analysis difficult. In addition, when comparing verbs of very different meanings it becomes hard to determine the relevant area of semantic difference. For example, in order to attain the generalization concerning "cut" and "break", Levin had to look at two other verbs ('touch' and 'hit') and their respective diathesis alternations, as well as look at other verbs that could fit into those alternations (cf. 1993, pp. 5-8). If she had picked different verbs instead of 'touch' and 'hit' or different diathesis alternations from the three that she did, she might not have been able to come up with a generalization at all. These factors may be contributing to the fact that there is currently no unified theory of lexical knowledge based on verb alternations because the scope of the undertaking is so vast.

If we agree, however, that syntactic behavior can shed light on the relevant semantics of a verb; what other type of behavior is available? Our methodology focuses on delimiting the lexical semantic distinctions between near-synonym pairs that differ slightly in both their syntactic behavior and in their semantics. However, even in the cases where there is a difference in meaning, what we are looking for is the collateral differences in both syntax and semantics -- that is, along what semantic lines do these two words differ, and how is this difference related to their syntactic behavior (and vice versa)?

How do we look for these collateral differences? First, we examine these near synonym pairs by first combing a corpus for all relevant examples of the words in question. These examples are then categorized according to their syntactic function. Third, each instance is classified into its argument structure type. Fourth, the aspectual type associated with each verb is determined, and fifth, the sentential type for each verb is also determined. We find that near-synonyms usually have several cases of complementary distribution of syntactic functions. It is often these cases of complementary distribution that allow us to formulate a hypothesis concerning the relevant nature of the semantic differences.

In the MARVS representation, lexical knowledge is classified into two types: structural information is represented with the composition of atomic modules, while content information is represented with attributes attached to these modules. The overall shape of event structure is defined by the composition of five event modules. The roles that participate in the event are represented in the role modules. The semantic attributes pertaining to the whole event are called the event-internal attributes and are attached to the event modules. The semantic attributes pertaining to each role are termed role-internal attributes and attached to the appropriate role within the role module. It is important to note that the eventive information is attached to the sense of a verb. Verbs that have different senses will have different eventive information.

3   Analysis of 'Put' and 'Set'

The sense that we will be looking at for 'put' and 'set' is to "move an object to a certain place", loosely following the classification and definition of WordNet 5.0. These objects and locations may or may not be abstract. We assume as a null hypothesis that both the abstract and concrete cases are part of a single sense, with the abstract cases an extension from the concrete cases. However, we will discuss the concrete and abstract cases separately before making a unified analysis. We do this because the

prepositions used with each verb in the concrete and abstract cases are noticeably different which may lead to different semantic generalizations.

The English verbs 'put' and 'set' seem synonymous and interchangeable in most contexts (1).

(1) a. Put/set the book on the table.
   b. He set/put the pin on the cushion.

Moreover, the distributional differences that exist for Mandarin verbs of putting *bai3* and *fang4* (such as co-occurence with a progressive to describe a process (2a), taking a resultant object (3a), and being modified with an orientational adjunct (4a)) do not exist in English (see examples 2b-4b respectively).

(2) a. Ta zhengzai fang/*bai shu zai suo-shang
       s/he DURATIVE   book at table-top

   b. She is putting/setting the book on the table.

(3) a. Ta *fang/bai-le   yi   zuo cai
       s/he   -LE one table dish

   b. He *put/*set a table of food. (cf. He set/put food on the table.)

(4) a. ta   ba   yizi   chao tong *fang/bai
       s/he BA chair face   east

   b. She put/set the chair facing east.

Differences between 'put' and 'set' do exist, however. We found the following contrasts based on one hundred randomly extracted samples of 'put' and 'set' from BNC.   First, in the cases of concrete objects and locations, 'put' had the theme being placed into a location seven of nine times (cf. examples (5)a-f with examples (6)a-b).

(5) a. put into one prison vehicle
    b. put it into a urine container
    c. put their heads into a bucket of urine
    d. put the Sandinistas in the dock
    e. put the muzzle of his weapon into Gruson's mouth
    f. They put us back into the van
    g. put more money into Hong Kong

(6) a. put more pressure on the pedals
    b. put men on the moon

It is important to note that the location in all of the cases is a contained space.

Turning to the case of 'set', however, graspable objects were set 'alongside, 'down' or 'in' (as 'in a formation' -- not inside something) as in (7).

(7) a. set alongside his Evlangean law doctorate
   b. set it down between Elisabeth and himself
   c. set down Miss Danziger's breakfast herself
   d. set down a tureen for one before her
   e. set them (chairs) in a semi-circle round the platform

The theme is not necessarily mentioned (but it is almost always mentioned in the 'put' cases), and in addition, the location is not necessarily a specific, fixed or definite space. Moreover, in cases when the theme is not graspable (as in the case of a house or garden) it is ellipsed and moreover, the location is also not a definite, containing space, as is demonstrated in (8).

(8)a. It was set in a trim garden
   b. set beneath the dramatic limestone
   c. set in acres (of maintained comfort)
   d. set admist flower gardens
   e. set in acres of garden
   f. set against the cottage stone

From the contrasts drawn above with the BNC data, we can observe that 'put' has the roles <Agent, Theme, Location> with Location having the role-internal attribute [bounded]. This contrast can be seen in (9).

(9)a. He put water in the bathtub.
   b. *He set water in the bathtub.

Example (9a) means that he filled the bathtub with water. Example (9b) does not allow that interpretation. It cannot be interpreted because the theme 'water' is not given any definite position in the location. In other words, the example in (7b) indicates that although a bounded space is not required for 'set', a position of theme is needed. In all the examples in (7) the position of the theme in relation to an actor is being described. In (8) the position of the theme is being described in relation to the landscape. Thus, the roles for 'set' are <Agent, Theme, Location> and the theme role carries the feature of [positioned].

This analysis explains why, in example (10a), the message can be either sitting on top of the bulletin board, or it can be stuck to the bulletin board, or it can be on an electronic bulletin board; since these are the three possible interpretations of 'bulletin board' as a contained location. However, in (10b), the only reading possible is one where the note is physically sitting on top of the bulletin board, since this is the only available possible interpretation.

(10)a. She put the note on the bulletin board.
     b. She set the note on the bulletin board.

In the 'put' case the only requirement is that the note appear on some part of the contained space of the bulletin board (whether or not that space is physical or ephemeral). In the 'set' case, the position of the theme in relation to the location is critical and only the positional reading is allowed.

How do these analyses account for the abstract objects and locations? In the case of 'put' the analysis holds nicely. In all the cases in (11) below, the locative role is a metaphorical extension to an abstract concept such as 'quandary' or 'trial.'

(11) a. put flesh and blood on the skeleton structure of a possible united Europe
     b. put them all in a real quandary
     c. put on trial
     d. put the best face on the evident lack of progress

Moreover, the locative role still has the role-internal attribute of [bounded]. In all of the above cases, the locative role has a definite boundary (i.e. the quandary is a particular situation, as is the trial, the lack of progress, and the structure of a united Europe.) Each one of these situations has a definite boundary.

The analysis of 'set' with abstract objects and locations must be divided up into different scenarios. First, in example (12), we can see that locative role has been extended to time. The theme (i.e. the event in question) must be positioned in relation to a period of time, although once again, the exact position (on a timeline, for example) is not necessary. Thus, the event structure that we have proposed for 'set' can stand for this first scenario.

(12) a. set in the early 50's
     b. set in a historical context

But in the second scenario, agents, themes, and propositions are involved (see (13) and (14)). Moreover, the propositions in (14) are resultatives.

(13) a. set crisis management operations in train
     b. setting in motion the sequence of European conflicts
     c. set in train a flood of protest
     d. set in motion a big spy swap

(14) a. set her pulse racing
     b. set fire to banks and offices
     c. set alight two offices
     d. set our boats on fire

In addition, in the cases in (13) and (14), there is no reason to postulate that the theme is in a particular position. It is not necessary, for example, that the offices in (14c) are in a particular position.

31

However, the propositional phrase assigns a property to the theme, i.e. there is a predicative relation. Thus, the roles in the event structure for 'set' with an abstract theme and location is: <Agent, Theme, Proposition>, with the predicative relation stipulated. In MARVS, this information is represented in the Role Module. Event structures with different role modules are considered different event structures. Since MARVS assumes that each unique event structure is assigned a verbal sense, and each verbal sense has a unique event structure, our analysis of 'set' with non-locational complement indicates that it has a different sense from the locational 'set'.

In conclusion, Levin (1993) assumes that: "....the behavior of a verb, particularly with respect to the expression and interpretation of its arguments, is to a large extent, determined by its meaning. Thus, verb behavior can be used effectively to probe for linguistically relevant pertinent aspects of verb meaning" (p.1). We agree with this assumption. But we argue that the place to look for verb behavior is not among differences in verb class because the distinctions at this level are not fine-grained enough. Instead we propose that near-synonym pairs be used. By doing so we can compare syntactic differences and extrapolate the subtle differences in semantic meaning. In addition, we argue that intuition will not be enough to distinguish between the near-synonyms. Instead, data extracted from a corpus will be able to show syntactic patterns of distribution that distinguish among the near-synonym pairs. We can use these patterns to postulate the underlying meanings of verbs.

References

Biq, Yung-O. (2000). Ed. Special Issue on Chinese Verbal Semantics. *Computational Linguistics and Chinese Language Processing*, Volume 5.1.

Levin, Beth. (1993). *English Verb Classes and Alternations: A Preliminary Investigation.* Chicago: University of Chicago Press.

Huang, Chu-Ren and Kathleen Ahrens, Li-li Chang, Keh-Jiann Chen, Mei-Chun Liu, Mei-Chi Tsai. (2000). The Module-Attribute Representation of Verbal Semantics: From Semantics to Argument Structure. *Computational Linguistics and Chinese Language Processing*, 5.1, 19-46.