A Semantic-Based Approach to Noun-Noun Compound Interpretation

You-shan Chung and Keh-Jiann Chen

Institute of Information Science

Academia Sinica

Taipei, Taiwan

{yschung, kchen}@iis.sinica.edu.tw

Abstract

In this paper, we propose a method to identify the possible readings of Chinese noun-noun compounds (NNCs).To avoid problems such as vagueness of interpretation, limited or sporadic coverage, and arbitrariness of semantic relation classification, we considereda large number ofnoun-noun compounds from our Prefix-Suffix System and identified their semantic relations with two semantic networks: FrameNet (https://framenet.icsi.berkeley.edu/fndrupal/home) and E-HowNet(http://ehownet.iis.sinica.edu.tw/ehownet.php). We found that N1 and N2 are either linked by semantic roles assigned by events (complex relations) or by static relations (simple relations) including meronymy, conjunction, and the host-attribute-value relation. Furthermore, for both types of relations, the possible readings of the resulting compoundare limited enough for computational implementation.

Regarding simple relations, conjunctionhas limited productivity; meanwhile, the two components usually belong to the same semantic type, such as 鐘錶 *zhong-biao*'clock and watch.'The limited productivity makes it possible for E-HowNet to include a large proportion of such pairs in its dictionary, while similar semantic types can be identified through its taxonomy. Likewise, the host-attribute-value combinationsarealso a type of simple relations which involves combinations of hosts, attributes, and values in certain orders, such as Value-Host (e.g. 鐵桌 *tie-zhuo*'iron table/desk'), Host-Attribute (e.g. 車速 *che-su*'car speed'), and Value-Attribute (e.g. 法式 *fa-shi*'French-style'). Themeronymic relations are the third type of simple relations, which are annotated by semantic roles like 'part of' and 'whole' (Part-Whole: e.g. 雙底船 *shuang-dichuan*'double-bottom'; Whole-Part: e.g. 車輪 *che-lun* 'car tire')

As for NNCs involving complex relations,the component nouns are the arguments of an event that bridges them and by which they are assigned semantic roles. For example, in 家長費 *jiazhang-fei*'parental fee,'N1 and N2 are bridged by events such as'buying' or 'paying'denoted by verbs like 'buy' and 'pay,'which assign

N1 and N2 the semantic roles of, respectively, 'Buyer' and 'Money.'Each component's semantic role, along with the events that assign these roles, can be figured out through mappings to, respectively, frame elements (henceforth FEs) and lexical units (henceforth LUs)in frames that represent the concept a NNC conveys based on the semantic category of the head, i.e. N2.

Two instances of mappings are as follows; one is for the simple and the other for the complex type. The former is exemplified by NNCs derived from N2s denoting 'food,' which correspond to the frame FOOD. We found that most of the N1s of these NNCs correspond to one of the frame's FEs, which is 'Constituent_parts.'Examples of nouns assuming the FE are underlined: banana with a thick peel(FrameNet's original example); 蔥油餅 cong-you bing'Chinese spring onion pancake'(example of mapped Chinese NNC). That such NNCs involve simple relations are supported by the absence of verbal LUs that evoke them. By contrast, NNCs derived from N2s denoting 'money' sometimes involve complex relations, as shown by the FE 'Money' taking part in frameslike'COMMERCE_SELL' and 'COMMERCE_BUY,'which FrameNet deems as evoked by event-denoting LUs such as 'sell' and 'buy' and having FEs like 'Buyer,''Seller,'and "Goods.'Under the assumption that N1 and N2 are bridged in an event where they are assigned semantic roles, we mapped some of the money-derived NNCs to these FEs.For example, we mapped 書款 shu-kuan'money for buying books' and 家長費 jiazhang-fei'parental fee'respectively to the FE pairs of Goods-Money and Buyer-Money in the frame COMMERCE_BUY, which is evoked by LUs like 'buy' and 'purchase.'While there are usually various possible FEs in a frame, such mappings reduce the possible readings to a manageable range, facilitating computational implementation.

So far, we have applied such mappings to nine productive N2 categories with moderate success. We think the approach is worth extending to more categories.