The Polysemy of *Da3*: An ontology-based lexical semantic study*

Hong, Jia-Fei [a], Chu-Ren Huang [b], Kathleen Ahrens [a],

[a] National Taiwan University, Graduate Institute of Linguistics
No. 1, Sec. 4, Roosevelt Road 106, Taipei, Taiwan R.O.C
[b] Institute of Linguistics Academia Sinica
No. 128, Section 2, Academia Road 115, Taipei, Taiwan R.O.C
{jiafei, churen}@gate.sinica.edu.tw,
kathleenahrens@yahoo.com

Abstract. In this study, we explore the polysemy of *da3* through the ontological conceptual structure found in SUMO. First, we divide several different senses for *da3*, clustering physical event senses and metaphorical event senses. In here, we only focus on physical event senses of *da3*. From the physical event senses of *da3*, we divide them into two main categories: 1) *hit* and 2) *pump*. We then use SUMO ontological concepts to identify these physical senses. Finally, we can observe the common patterns of the "hit" sense group and the "pump" sense group for *da3*.

Keywords: *da3*, Polysemy, ontology, lexical semantics

1. Introduction

In this study, we explore all possible concepts for physical event senses of *da3* through the SUMO ontological concept system (Huang et al. 2004). According to previous work (Gao 2001), *da3* is a basic verb in the large domain of physical action verbs in Chinese, as 1) it refers to the most basic action of the hand; and 2) at the same time it can refer to a wide-range of actions or events that involve physical contact of one kind or another. We will compare her analysis with the analysis we provide based on SUMO.

First, we collect our data from Sinica Corpus and check their senses from Chinese Wordnet. Next, we take these physical event senses of *da3* into SUMO concept system (Huang et al. 2004) to find all possible concepts and distinguish them into different categories. Finally, we analyzed these concepts for semantic features which can help us to compare our analysis with the analysis in Gao's study (2001).

2. Previous research

Regarding verb studies, previous research has focused on VV compound verbs in Modern Chinese (Hong and Huang, 2004), or on near synonyms in Modern Chinese (Chief et al, 2000; Huang et al. 2000; Liu 2002; Tsai, 2002; Huang and Hong, 2005). Also, some scholars have worked on *da3* polysemy analyses. *Da3* is one if the most frequently used verbs, being ranked 16 in the list of most frequently used verbs in Chinese (Bei and Zhang, 1988). Specifically, Gao

(2001) explored the semantic properties of *da3* and its prototypical meaning and categorized its semantic representations to show the systematic patterning of its meaning extensions.

3. Motivation and Goals

Language knowledge representation is a manifestation of the systematic contrasts found in human communication, which defies conventional description. Take modal verbs as examples. Modal verbs have similar semantic functions and cannot be easily distinguished in terms of their lexical senses. Therefore, they are considered to be interchangeable. Nevertheless, it is not uncommon that this kind of polysemy always has contrasts in usage, as we can see the contrast between *hui4* (know) and *hui4* (can) below:

(1a) 臺灣廠商到德國開商展時，非常需要**會**德語和中文的人，做為溝通橋樑。
Tai2 wan1 chang3 shang1 dao4 de2 guo2 kai1 shang1 zhan3 shi2, fei1 chang2
Taiwan factory to Germany exhibit time, so

xu1 yao4 hui4/ * neng2 de2 yu3 han4 zhong1 wen2 de5 ren2, zuo4 wei2
need could German and Chinese MOD persons, to

gou1 tong1 qiao2 liang1.
communicate bridge

"When Taiwan factories exhibit in Germany, they so need some persons who could speak German and Chinese to communicate with other persons."

(1b) 在上課的時候，他不**會**講德語或中文和學生溝通，因為他怕自己沒辦法完整表達意思。
Zai4 shang4 ke4 de5 shi2 hou4, ta1 bu2 hui4 jiang3 de2 yu3 huo4
In class MOD time, he will not speak German or

zhong1 wen2 han4 xue2 sheng1 gou1 tong1, yin1 wei4 ta1 pa4 zi4ji3
Chinese with student communicate, because he worry himself

mei2 ban4 fa3 wan2 zheng3 biao3 da2 yi4 si1.
no way complete express meaning.

"In the class time, he can't communicate with his students in German or Chinese, because he worries that he can't completely express his meaning."

Considering the lexical sense of modal verb polysemy and its natural language use, *hui4* means both "know" or "can". As defined, we notice that they differ from each other, even though they share similar concept.

This paper will investigate the lexical semantic relations between each sense of *da3* polysemy (excluding metaphorical senses) by studying their sense distinction, word formation collocation, and distribution pattern.

4. SUMO

In this study, we use Suggested Upper Merged Ontology (SUMO) to analyze all concepts for *da3*. We find out all possible concepts and divide different them into different categories.

WordNet is inspired by current psycholinguistic and computational theories of human lexical memory (Fellbaum (1998), Miller et al. (1993)). English nouns, verbs, adjectives, and adverbs are organized into synonym sets, each representing one underlying lexicalized concept.

Different semantic relations link the synonym sets (synsets). The version of WordNet that Sinica BOW implemented is version 1.6, with nearly 100,000 synsets.

In Sinica BOW, ach English synset was given up to 3 most appropriate Chinese translation equivalents. In cases where the translation pairs are not synonyms, their semantic relations are marked (Huang et al. 2003). The bilingual WordNet is further linked to the SUMO ontology. We use the semantic relations in bilingual resource to expand and predict domain classification when it cannot be judged directly from a lexical lemma.

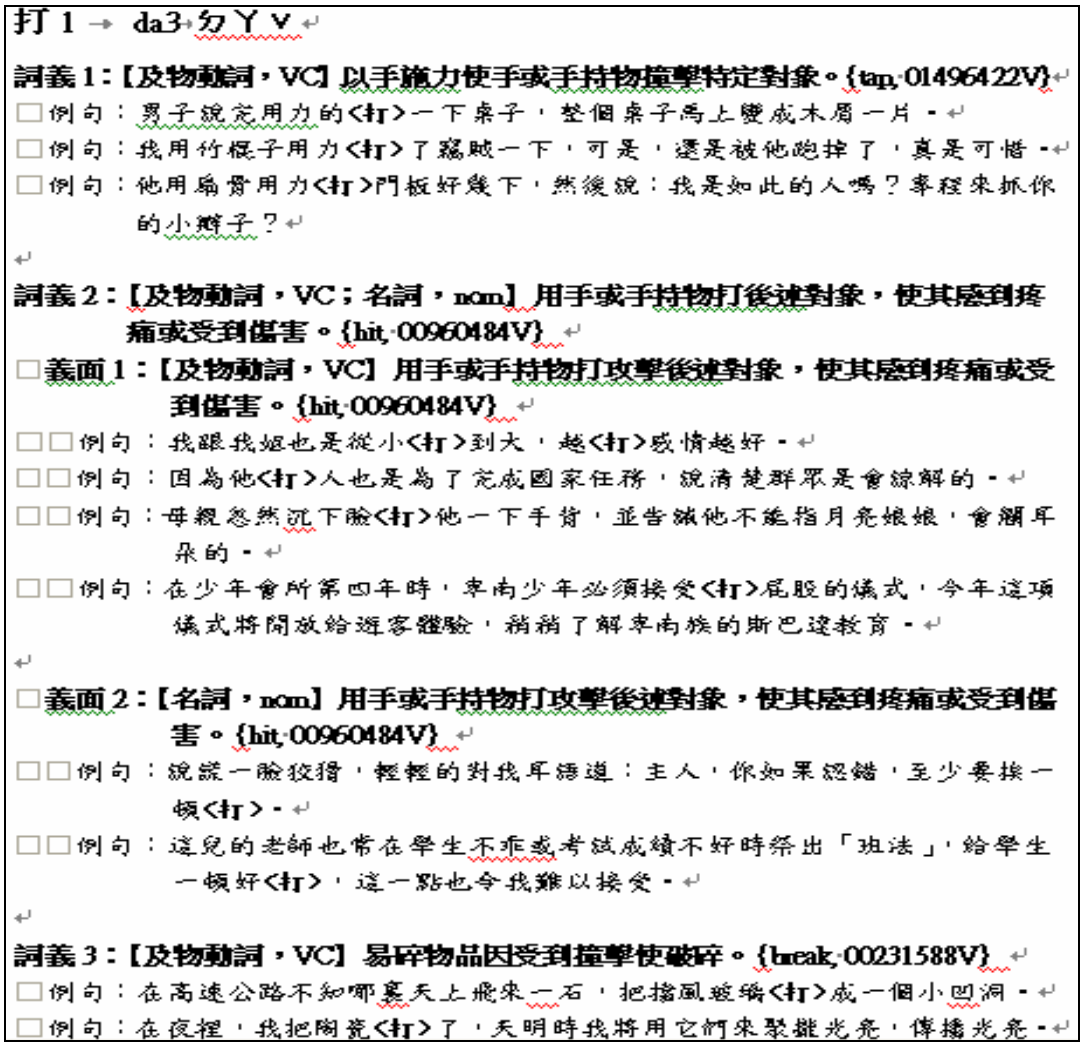
5. Data collection

From Sinica Corpus and Gigaword Corpus, we find out several patterns for *da3*. According to the Chinese Wordnet Group analysis (Huang et al., 2003), there are 114 senses which include physical activity senses, metaphor, metonymy and extension senses. In this study, we want to focus on physical activity senses, but not metaphor, metonymy and extension senses. *Da3* has 35 physical event senses listed, along with 79 additional senses. We will take focus on the physical activity senses.

6. Data analysis

The analysis, based on the Sinica Corpus, Gigaword Corpus and the criteria proposed by Huang et al. (2003) to differentiate the lexical meaning, presents several different senses for *da3*.

Table 1**:** The analysis of *da3* from Chinese Wornet Group



打 1 → da3 ㄉㄚˇ

詞義 1：【及物動詞，VC】以手施力使手或手持物撞擊特定對象。{tap,01496422V}

例句：男子說完用力的<打>一下桌子，整個桌子馬上變成木屑一片。

例句：我用竹棍子用力<打>了竊賊一下，可是，還是被他跑掉了，真是可惜。

例句：他用扁骨用力<打>門板好幾下，然後說：我是如此的人嗎？專程來抓你的小辮子？

詞義 2：【及物動詞，VC；名詞，nom】用手或手持物打後連對象，使其感到疼痛或受到傷害。{hit,00960484V}

義面 1：【及物動詞，VC】用手或手持物打攻擊後連對象，使其感到疼痛或受到傷害。{hit,00960484V}

例句：我跟我姐也是從小<打>到大，越<打>感情越好。

例句：因為他<打>人也是為了完成國家任務，說清楚群眾是會諒解的。

例句：母親忽然沉下臉<打>他一下手背，並告誡他不能指月亮娘娘，會爛耳朵的。

例句：在少年會所第四年時，卑南少年必須接受<打>屁股的儀式，今年這項儀式將開放給遊客體驗，稍稍了解卑南族的斯巴達教育。

義面 2：【名詞，nom】用手或手持物打攻擊後連對象，使其感到疼痛或受到傷害。{hit,00960484V}

例句：說話一臉狡猾，輕輕的對我耳語道：主人，你如果認錯，至少要挨一頓<打>。

例句：這兒的老師也常在學生不乖或考試成績不好時祭出「班法」，給學生一頓好<打>，這一點也令我難以接受。

詞義 3：【及物動詞，VC】易碎物品因受到撞擊使破碎。{break,00231588V}

例句：在高速公路不知哪裏天上飛來一石，把擋風玻璃<打>成一個小凹洞。

例句：在夜裡，我把陶瓷<打>了，天明時我將用它們來聚雖光亮，傳播光亮。

Among these 35 physical activity senses of *da3*, we divide two main physical event senses for *da3*: 1) *hit* and 2) *pump* such as below:

(2) a. 母親忽然沉下臉**<打>**他一下手背，並告誡他不能指月亮娘娘，會爛耳朵的。

Mu3 qin1 hu1 ran2 chen2 xia4 lian3 da3 ta1 yi2 xia4 shou3 bei4,
Mother suddenly sink down face hit he one time hand,

bing4 gao4 jie4 ta1 bu4 neng2 zhi3 yue4 liang4 niang2 niang5,
and warn he can't point moon queen ,

hui4 lan4 er3 duo1 de5.
will decayed ear MOD.

"His mother suddenly becomes long-faced and hits him on the back of his hand, then warned him that he will get rotting ears if he points to the moon."

b.這時每一鞭都如<打>在她的身上一般痛楚。

Zhe4 shi2 mei3 yi1 bian1 dou1 ru2 da3 zai4 ta1 de5

This time every whipped all like whip she MOD

shen1 shang4 yi1 ba1 tong4 chu3

body general pain.

"At this moment, every whipped whip pains her like she is being whipped."

(3)為了防止車輛陷進沙地，不要把輪胎氣<打>得太足。

Wei4 le5 fang4 zhi3 che1 liang4 jin4 sha1 di4, bu2 yao4 ba3

For prevent cars s tuck into sand, don't let

lun4 tai1 qi4 da3 de2 tai4 zu2.

tires gas pump too full.

"Do not pump the tires too full to avoid the cars being stuck in the sand."

Moreover, in "hit" sense of *da3*, we also can divide two different categories: 1) *hand* and *hand holdings* and 2) *force* and *impact*. Then, in the second category, we can divide *force* and *impact* in additional sub-categories: 1) direct contact and 2) contact by injection.

(4)他用力<打>門板好幾下，然後說：我是如此的人嗎？

Ta1 yong4 li4 da3 men2 ban3 hao3 ji3 xia4, ran2 hou4 shuo1:wo3

He use force beat door plank several times, then say: I

shi4 ru3 ci3 de5 ren2 ma5?

is this MOD man.

"He beat forcefully the door plank several times and then said: Am I a person like this?"

(5)當靜脈注射毒癮者<打>毒品時，通常是不會馬上把毒品立刻注射進去，而是將針頭插在血管上。

Dang1 jing4 mai4 zhu4 she4 du2 yin3 zhe3 da3 du2 pin3 shi2,

When intravenous injection drug addiction person inject drugs time,

tong1 chang2 shi4 bu2 hui4 ma3 shang4 ba3 du2 pin3 li4 ke4 zhu4 she4

usually will not at once let drugs immediately inject

jin4 qu4, er2 shi4 jiang1 zhen1 tou2 cha1 zai4 xie3 guan3 shang4.

into, instead use needle head insert in blood vessel above.

"When drug addicts take drugs, they usually rest the needle in the vein instead of injecting the drug directly."

(6)一個滿眼夢想的快樂女孩，因為受傷開刀，要<打>釘子進脊椎，像副衣架一般把彎彎的骨頭撐直。

Yi2 ge4 man3 yan3 meng4 xiang3 de5 kuai4 le4 nyu3 hai2, yin1 wei4

One full eyes dream MOD happy girl, because

shou4 shang1 kai1 dao1, yao4 da3 ding1 zi5 jin4 ji2 zhui1, xiang4 fu4
hurt operate, will put nail to spine, like set

yi1 jia4 yi1 ban1 ba3 wan1 wan1 de5 gu3 tou2 cheng1 zhi2.
clothes stand in general let bend MOD bone prop straight.

"A happy girl full of dreams due to an injury happened, has to undergo an operation that puts nails into her spine to straighten up the wiggled spine, which is very similar to clothing being hanged on hangers."

We follow these categories to explore the common features for these physical event senses of *da3* in the SUMO concept system.

7. Data analysis

According to Gao's study (2001), she based on sense division principle to analyze *da3*, generalized the patterns and features of the polysemy of *da3*, and obtained five major categories.

Gao (2001) mentioned the prototypical meaning of *da3*. She talked about in the prototypical case the most central part of the meaning of *da3* is the physical contact between an agent's hands and a concrete item. In her paper, she also mentioned that there were three different semantic elements for *da3* such as 1) *hand*, *hand holdings* or *instrument*; 2) *force direction* and 3) *impact*.

From all our senses of *da3*, we can divide two main categories: 1) physical event senses such as *da3 zhuo1 zi5* (to tap the table), *da3 shou3 bei4* (to hit the back of a hand), *ba3 wan3 da3 po4* (to break a bowl)... and so on and 2) metaphorical event senses such as *da3 jiao1 dao4* (to develop the interpersonal relationship/ to come into contact with), *da3 dian4 hua4* (to call), *da3 ke1 shui4* (to nod)... and so on. However, in this study, we just focus on physical event senses. Based on Gao's analysis (2001), we know that physical event senses of *da3* include these features such as hand, hand holdings, instrument; force and impact, so we thoroughly examine our physical event senses by SUMO concept system. In here, we can observe that there are several concept of SUMO for physical event senses of *da3* such as below table:

Table 2: SUMO concept for physical event senses of *da3*

The phrase of *da3* in Chinese	**Translation in English**	**SUMO concept**
da3 shou3 bei4	to *hit* the back of a hand	impacting
da3 zhuo1 zi5	to *tap* the table	touching
ba3 wan3 *da3* po4	to *break* a bowl	impacting
lun4 tai1 *da3* qi4	to *pump* gas into tire	putting
da3 shi1 li4 kang1	to *inject* silicon	putting

We follow SUMO concept system to obtain these concepts of *da3*. These concepts are such as impacting, touching, putting. We also know that the SUMO concept identifications correspond with the divisions and definitions in WordNet. For this reason, we need make sure the WordNet definitions of these concepts for physical event senses of *da3*.

Table 3: WordNet definition and SUMO concept for physical event senses of *da3*

The English lemma of *da3*	**WordNet definition**	**SUMO concept**
hit	deal a blow to, either with the hand or with an	impacting

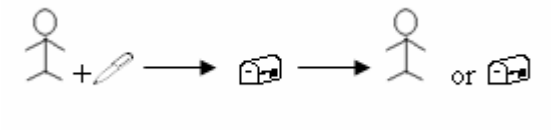
	instrument	
tap	a light touch or stroke	touching
break	destroy the integrity of; usually by force; cause to separate into pieces or fragments	impacting
pump	deliver forth	putting
inject	force or drive (a fluid or gas) into by piercing	putting

In this way, we obtain the common semantic elements from SUMO concepts for physical event senses of *da3*. The semantic features are hand, instrument, and force. In addition, we can detect when we do these actions, the manners are impact, direct contact or contact by injection. We may visual this as:

(1) Agent + hand, hand holding or instrument --> Patient or Object

(2) Agent --> impact (direct contact) --> Object

(3) Agent --> force (contact by injection) --> Patient or Object



From the SUMO concept system for the physical event senses of *da3*, we see these concepts imply the following semantic features: hand, hand holdings, instrument; force and impact (direct contact or contact by injection). We use SUMO concept system to find out all possible concepts for the physical event senses of *da3*, while Gao (2001) used semantic features to analyze and explain physical actions of *da3*. Following our analyses, explanations, comparison and demonstrations, we discover that our analyses correspond with Gao's study result for the physical event senses of *da3*.

8. Conclusion

In this study, we explore all possible concepts for physical event senses of *da3* through the SUMO concept system. These concepts imply some semantic features: hand, hand holdings, instrument; force and impact. We use concept-based approach, while Gao (2001) took a semantic-feature-based approach to examine the physical event senses of *da3*. We also compare our analysis with Gao's (2001) and find our results are very similar. This leads us to propose that a concept-base approach is a viable one when exploring the sense of polysemous verbs in Chinese.

References

Bei, Guqin and Xuetao Zhang. eds., 1988. Hanzi Pindu Tognuji – Sucheng Shudu Youxuan Biao (statistics of Chinese Word Frequency – A First Priority Lift for Quick Literacy Reading). Beijing: Dianzi Gongye Chubanshe (Electronics Industry Publishing House).

Chang, Li-li, Chen, Keh-Jiann and Huang, Chu-Ren. 2000. *A Lexical-Semantic Analysis of Mandarin Chinese Verbs: Representation and Methodology. Computational Linguistics and Chinese Language Processing.* Vol.5, No. 1, February 2000, 1-18.

Chief, Lian-Cheng, Chu-Ren Huang, Keh-Jiann Chen, Mei-Chih Tsai and Li-Li Chang. 2000. *What Can Near Synonyms Tell Us. International Journal of Computational Linguistics and Chinese Language Processing.* 5 (1)47-60.

Fellbaum C.. WordNet: An Electronic Lexical Database. Cambridge: MIT Press 1998

Gao, H. H. 2001. Da Polysemy. *The physical foundation of the patterning of physical action verbs, Lund University Press.* 157-204.

Hong, Jia-Fei, Xiang-Bing Li and Chu-Ren Huang**.** *2004. Ontology-based Prediction of Compound Relations: A study based on SUMO. Presented at PACLIC18*. December 8-10. 151-160. Tokyo: Waseda University.

Huang, Chu-Ren, and Ru-Yng Chang. Sinica BOW (Bilingual Ontological Wordnet): Integration of Bilingual WordNet and SUMO". *Presented at the 4th International Conference on Language Resources and Evaluation (LREC2004)*. Lisbon. Portugal. 26-28 May .2004.

Huang, Chu-Ren. Elanna I. J. Tseng, Dylan B. S. Tsai, Brian Murphy. Cross-lingual Portability of Semantic relations: Bootstrapping Chinese WordNet with English WordNet Relations. Languages and Linguistics. 4.3. 2003. 509-532

Huang, Chu-Ren, Kathleen Athens, Li-Li Chang, Keh-Jiann Chen, Mei-Chun Liu, Mei-Chih Tsai. 2000. *The Module-Attribute Representation of Verbal Semantics: From Semantics to Argument Structure*. In International Journal of Computational Linguistics & Chinese Language Processing 5.1: 19-46.

Liu, Mei-Chun. 2002. *Mandarin Verbal Semantic*s: A Corpus-based Approach. 2nd ed., Taipei, Taiwan: Crane Publishing Co.

Miller G. A., R. Beckwith, C. Fellbaum, D. Gross and K. Miller. "Introduction to WordNet: An On-line Lexical Database," *In Proceedings of the fifteenth International Joint Conference on Artificial Intelligence. Chambéry, France*. 28 August- 3 September .1993.

蔡美智. 2002。 講「清楚」、說「明白」—漢語動詞近義、多義、詞義劃分研究。第三屆中文詞彙語意學研討會。台北，南港：中央研究院。

Website Resources

Chinese Word Sketch Engine: http://wordsketch.ling.sinica.edu.tw/

English Word Sketch Engine: http://www.sketchengine.co.uk/

Lexical Data Consortium. 2005. Chinese Gigaword Corpus 2.5.: http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2005T14

Sinica Corpus. http://www.sinica.edu.tw/SinicaCorpus/