Verbal Semantics as Eventive Information: The Case of Verb "制服" and "制約" in Chinese Gigaword Corpus

Abstract. Using Chinese Gigaword Corpus, this study aims to explore the similarities and differences between two Chinese verbs "制服" (zhì fú) and "制約" (zhì yuē), so as to understand the lexicosemantics of verbs with and without an endpoint. First, the distributional patterns show that the two verbs occur mainly in the genre of story of journalism (98.4% for "制服" zhì fú and 99.1% for "制約"). Second, based on the MARVS theory and the collocations with the tense and aspect markers ("著" zhe、"了" liǎo、"過" guò), it is found that "制服" is telic (accomplishment), representing an event module with an endpoint, that is BOUNDARY (.), while "制約" is atelic and a PROCESS activity (///). Third, in contrast to the uni-directional verb "制服", it is found that "制約" can be framed bi-directionally between the agent and the patient, using the adverbals such as "相互" (xiāng hù) and "互相" (hù xiāng). As a continuous PROCESS verb, "制約" can be modified by: a) degree adverbials such as "嚴重" (yán zhòng), b) duration /continuous adverbials such as "一直" (yī zhí)、"繼續" (jì xù), and c) starting-relevant adverbials such as "開始" (kāi shǐ)。The study sheds light on teaching and learning verb meanings through lexicosemantics in context, particularly the event features suggested by the MARVS theory.

Keywords: corpus linguistics, semantics, tense and aspects, MARVS, coercion

1   Introduction

Understanding how meaning is conveyed in and transmitted through language is an important task for language learners. Researchers try to decode sense and meaning from a variety of perspectives, including for example, syntactically (Chomsky, 1965), innately (Pinker, 2003), and statistically (Ellis, 2002). With the development of computers and internet, data are accumulated at an incredibly fast speed. We now can easily access to large language database (corpora), and many of which are annotated with useful part-of-speech tagging, making corpus-based analysis for understanding patterns of sense and meaning feasible. This study reports a corpus-based analysis, using a tagged corpus and a Sketch Engine called Chinese Gigaword, consisting of about 600 million words from traditional and simplified Chinese news media (Huang, 1996; Kilgarriff et al., 2005).

1.1   Verb sense and meaning

Verbs and nouns are the two basic elements in a language that help us experience and express the world. Nouns stand for entity as they often prime concrete objects and verbs

represent event and state. Verbs are eventual so that sense of verb meaning can be determined by the event kind and formed through its conceptualization in context (Huang et al. 2000). Such a conceptualization regarding the lexicosemantic relations of verbs and their meaning can be understood from Fillmore's (1982, 2008) frame theory and Goldberg's (1995, 2003) Caused-Motion Construction theory, concerning the thematic relations of agents, patients, themes as well as the placement of verbs in the "construction grammar". In addition to the constructional grammar approaches (e.g., the application of FrameNet, Fillmore & Baker, 2000), the theory of Module-Attribute Representation of Verbal Semantics (MARVS) provides a perspective for understanding lexico-semantics by conceptualizing the event features following the Event Modules, Event Internal Attributes, Role Module and Role-Internal Attribute (Ahrens et al., 2003; Huang et al. 2000).

1.2 The MARVS theory

The event module proposed by the MARVS theory (Ahrens et al., 2003; Huang et al., 2000), which represents the information on the event structure of verbs, includes five distinguishable atomic event modules:

BOUNDARY .
PUNCTUALITY /
PROCESS ///
STATE _____
STAGE ^^^^^^

These event modules with their conceptual graphs indicate whether an event has an endpoint (BOUNDARY .) and whether it is a continuous process (PROCESS ///) among others. The benefits of having the MARVS system is to identify patterns in verbs that make them distinguishable to each other, which has important implications for teaching, learning, and information retrieval purposes. Therefore, we base our analysis on the MARVS theory (Ahrens et al., 2003; Huang et al., 2000).

2   Methodology

This study focuses on two Chinese verbs, "制服" and "制約" for anlaysis because they have both meaning overlaps and differences, which make them a good exmaple for employing courpus-based research to inform langauge teaching and learning. We first identify its sense and meaning from Chinese Verbnet. Then we explore the two verbs in Chinese gigawords corpus for analyzing their distribtuional patterns, tense and aspect collocations, and modifier patterns. Finally we discuss the results and draw a conclusion on our research findings.

3 Results

3.1 Basic meaning of "制服" and "制約" from Chinese WordNet

From Chinese WordNet dictionary (http://lope.linguistics.ntu.edu.tw/cwn2/query/), we know that the verb "制服" is a transitive verb (VC), which means *forcing someone to surrender*, and "制約" is a transitive verb (VC) too, which means *imposing restrictions on the current state or on the changes*.

Figure 1. Basic meaning.

3.2 Analysis from Chinese Gigaword Corpus

Frequency and distribution. We then use the Chinese Word Sketch (CWS) (https://wordsketch.ling.sinica.edu.tw/cws/) (Huang, 1996; Kilgarriff et al., 2005) for analyzing the frequency and distributional patterns of "制服" and "制約" in Chinese Gigaword Corpus, mainly composed of articles from Chinese journalism. First, it is found there are in total 5,496 instances of "制服", 496 of which are used as verbs, and 13,093 instances of "制約", all of which are used as verbs. Second, the verbs "制服" and "制約" mainly occur in the genre of story in journalism (98.4% for "制服" and 99.1% for "制約"). The results are presented in Table 1.

Table 1. Frequency and distribution in the Chinese Gigaword Corpus.

Genre Verb	Story	Multi	Other	Advis	Total
制服	486	4	4	0	494
%	98.4%	0.8%	0.8%	0.0%	100%
制約	12,972	76	44	1	13,093
%	99.1%	0.6%	0.3%	0.0%	100%

Story: of a particular topic or event;
Multi: describing a particular topic or event as in "news briefs in finance or sports";
Advis: what the news service addresses to news editors; not intended readers;
Other: none of the above; items such as sports scores, stock prices, temperatures report.

Collocations with tense and aspect markers. Using the CWS, we conducted searches on the collocations of the keywords "制服" and "制約" respectively with the five tense and aspect markers: "著" (zhe)、"了"(liǎo)、"過"(guò) (Keyword 制服 /制約 ＋ 著 /了/過), "在" (zài), and "正在" (zhèng zài) (在/正在 ＋ Keyword 制服 /制約). The purpose is to observe whether the two keywords are telic (e.g., accomplishment and achievement verbs like "build" and "retire") or atelic (e.g., state and activity verbs like "know", and "walk") based on the internal temporal constituency of a situation (Comrie, 1976; Smith, 1991, 1997), so that we can use the MARVAS theory to judge the internal event modules of the two verbs— whether they are verbs with boundary, punctuality, process, state, or stage (Ahrens et al., 2003; Huang et al., 2000). The results are presented in Table 2.

Table 2. Collocations with tense and aspect markers in the Chinese Gigaword Corpus (incidences)

	＋ 著 zhe	＋ 了 liǎo	＋ 過 guò	在 zài ＋	正在 zhèng zài ＋
制服	0	259	2	32	1
制約	786	1,562	1	24	7

"正在 zhèng zài +" is also included in "在 zài+"

Collocations with modifiers. Using the CWS, we conducted search on the collocations of the keywords "制服" and "制約" with modifiers (modifier + Keyword 制服 /制約) "相互" (xiāng hù)、"互相" (hù xiāng)、"彼此" (bǐ cǐ)、"共同" (gong tóng)、"嚴重" (yán zhòng)、"一直" (yī zhí)、"繼續" (jì xù)、and "開始" (kāi shǐ). The results are presented in Table 3. The results indicate that among the many modifiers, "制服" can only collocate with "共同" ("共同制服"). None of the other modifiers that work well with "制約" collocate with "制服". These modifiers work with "制約" to describe the bidirectional agent and object ("相互制約"、"互相制約"), the degree of severeness ("嚴重制約") , the duration relevance ("一直制約"、"繼續制約"、"開始制約").

It is because first of all, "制服" is uni-directional, so that the bidirectional agent and object ("相互"、"互相") do not collocate with it. Second, "制服" denotes an endpoint according to the MARVS theory (Ahrens et al., 2003; Huang et al., 2000), there is no need to put a degree and duration to "制服", which is the same as the accomplishment verbs such as "cross the road", "build a house". It is noted that due to the boundary that instantly make an endpoint (like semelfactive verbs such as "jump", "sneeze", "hit"), neither is it necessary for "制服" to collocate with starting-relevant adverbials such as "開始" (kāi shǐ). In contrast, "制約" with a continuous active course, can be modified by the duration or continuous adverbials, as well as degree adverbials and starting-relevant adverbials.

Table 3. Collocations with modifiers in the Chinese Gigaword Corpus (incidences)

	相互+ xiāng hù	互相+ hù xiāng	彼此+ bǐ cǐ	共同+ gòng tóng	嚴重+ yán zhòng	一直+ yī zhí	繼續+ jì xù	開始+ kāi shǐ
制服	0	0	0	16	0	0	0	0
制約	320	121	5	3	1,174	40	10	5

We observed from the modifier collocations in the Word Sketch Differences analysis that that the two transitive verbs "制服" and "制約" can have similar agent and patient regarding thematic relations (Fillmore, 1982), but they can also be modified by different adverbials, because the semantic event modules of "制服" has an endpoint while "制約" is a continuous process oriented verb. The results of Word Sketch Differences analysis are presented in Figure 2.
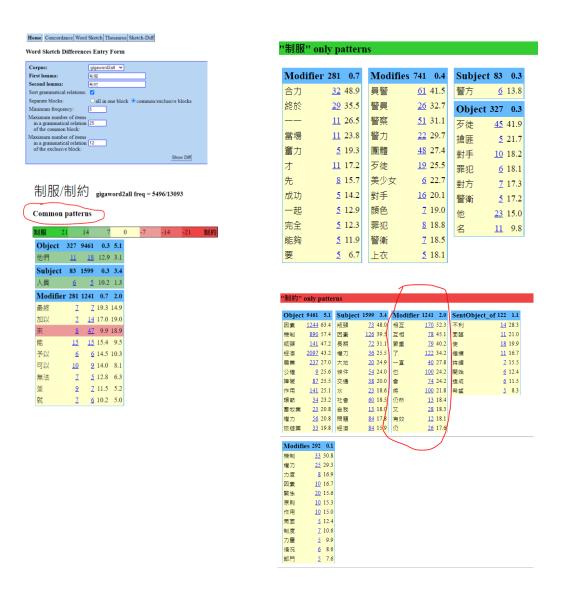
Word Sketch Differences Entry Form

Corpus: gigaword2all
First lemma: 制服
Second lemma: 制約
Sort grammatical relations: ☑
Separate blocks: ○ all in one block ● common/exclusive blocks
Minimum frequency: 5
Maximum number of items in a grammatical relation of the common block: 25
Maximum number of items in a grammatical relation of the exclusive block: 12
Show Diff

制服/制約 gigaword2all freq = 5496/13093

Common patterns

制服	21	14	7	0	-7	-14	-21	制約

Object	327	9461	0.3	5.1
他們	11	18	12.9	3.1

Subject	83	1599	0.3	3.4
人員	6	5	10.2	1.3

Modifier	281	1241	0.7	2.0
最終	7	7	19.3	14.9
加以	7	14	17.0	19.0
來	8	47	9.9	18.9
能	15	15	15.4	9.5
予以	6	6	14.5	10.3
可以	10	9	14.0	8.1
無法	7	5	12.8	6.3
並	9	7	11.5	5.2
就	7	6	10.2	5.0

"制服" only patterns

Modifier	281	0.7
合力	32	48.9
終於	29	35.5
一一	11	26.5
當場	11	23.8
奮力	5	19.3
才	11	17.2
先	8	15.7
成功	5	14.2
一起	5	12.9
完全	5	12.3
能夠	5	11.9
要	5	6.7

Modifies	741	0.4
員警	61	41.5
警員	26	32.7
警察	51	31.1
警力	22	29.7
團體	48	27.4
歹徒	19	25.5
美少女	6	22.7
對手	16	20.1
顏色	7	19.0
罪犯	8	18.8
警衛	7	18.5
上衣	5	18.1

Subject	83	0.3
警方	6	13.8

Object	327	0.3
歹徒	45	41.9
搶匪	5	21.7
對手	10	18.2
罪犯	6	18.1
對方	7	17.3
警衛	5	17.2
他	23	15.0
名	11	9.8

"制約" only patterns

Object	9461	5.1	Subject	1599	3.4	Modifier	1241	2.0	SentObject_of	122	1.1
因素	1244	63.4	瓶頸	73	48.0	相互	170	52.3	不利	14	28.3
機制	890	57.4	因素	126	39.5	互相	78	45.1	面臨	11	21.0
瓶頸	141	47.2	長期	72	31.1	嚴重	79	40.2	使	18	19.9
經濟	2097	43.2	權力	36	25.5	了	122	34.2	繼續	11	16.7
農業	237	27.0	大地	20	24.9	一直	40	27.8	持續	7	15.5
公權	9	25.6	條件	54	24.0	也	100	24.2	開始	6	12.4
障礙	87	25.5	交通	38	20.0	會	74	24.2	造成	6	11.5
作用	141	25.1	水	23	18.6	將	100	21.8	希望	5	8.3
環節	34	23.2	社會	60	18.5	仍然	13	18.4			
畜牧業	23	20.8	自我	13	18.0	又	28	18.3			
權力	56	20.8	問題	84	17.8	有效	12	18.1			
旅遊業	33	19.8	經濟	84	15.9	仍	26	17.6			

Modifies	292	0.1
機制	33	30.8
權力	25	29.3
力度	8	16.9
因素	10	16.7
關係	20	15.6
原則	10	15.3
作用	10	15.0
局面	5	12.4
制度	7	10.6
力量	5	9.9
情況	6	8.6
部門	5	7.6

Figure 2. Word Sketch Differences analysis

Additional analysis of "制服". From the analysis of "制服" and "制約" in the CWS, we also observed the phenomenon of "coercion", extending the basic meaning of "制服" from "making someone to surrender" to a metaphorical meaning of "controlling disease" and "controlling inflation by the government". In that sense, the object of "制服" is not necessarily a human ("someone") but can be problems such as "diseases" and "inflations". We summarize the results of in Figure 3.

Figure 4. From basic meaning of "制服" to the metaphorical meaning of "controlling"

4   Discussion and Conclusion

In this section, the results are summarized regarding the verb similarities and differences. There are probably three main similarities between the verbs "制服" and "制約": 1) both are transitive activity verbs that may have an object, 2) both have concrete animated agent and patient; 3) both can be modified by adverbials such as "finally". There are also significant differences between the two verbs. First, according to the MARVS theory, "制服" represents an event module with an endpoint, that is BOUNDARY (.) while "制約", a continuous PROCESS (\\\). Second, "制服" is uni-directional, and has telic aspect (like an accomplishment), so it cannot be modified by: a) bi-directional adverbial, b) degree adverbial, c) duration adverbial, and d) starting-relevant adverbial.

5   Implications

From the analysis of lexicosemantic of the verbs "制服" and "制約" in this article, it is argued that the MARVS theory (Huang et al., 2000; Chung & Ahrens, 2008) can be used to decode and understand verb meanings. The event features described in the Module-Attribute representations can be employed to conceptualize verb senses and meanings. Therefore, it is suggested that teachers and students can use the MARVS theory and the large language corpora such as the 600 million Chinese Gigaword corpus for gaining contextualized meaning understanding of individual words, distributional patterns, collocations, as well as meaning differences among words.

References

1. Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press.
2. Comrie, B. (1976). *Aspect: An introduction to the study of verbal aspect and related problems* (Vol. 2). Cambridge university press.
3. Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in second language acquisition, 24*(2), 143-188.
4. Pinker, S. (2003). *The language instinct: How the mind creates language*. Penguin UK.
5. Fillmore, C. J. (1982). Frame semantics. In Linguistic Society of Korea (Ed.) *Linguistics in the morning calm* (pp. 111-137). Seoul: Hanshin.
6. Fillmore, C. J. (2008). Frame semantics. In *Cognitive linguistics: Basic readings* (pp. 373-400). De Gruyter Mouton.
7. Fillmore, C. J., & Baker, C. F. (2000). FrameNet: Frame semantics meets the corpus. In *Proceedings of the 74th annual meeting of the linguistic society of America*.
8. Ahrens, K., Huang, C. R., & Chung, Y-H. (2003). Sense and meaning facets in verbal semantics: A MARVS perspective. *Language and Linguistics 4*, 468. http://www.ling.sinica.edu.tw/files/publication/j2003_3_02_8458.pdf
9. Huang, C-R, Chen K-J & Chang, L-L. (1996). Segmentation standard for Chinese natural language processing. In Jun'ich Tsujii (Program Chair), *Proceedings of the 16th Conference on Computational Linguistics*, vol. 2 (pp. 1045-1048). Stroudsburg: Association for Computational Linguistics. https://doi.org/10.3115/993268.993362
10. Huang, C. R., Ahrens, K., Chang, L. L., Chen, K. J., Liu, M. C., & Tsai, M. C. (2000). The module-attribute representation of verbal semantics: From semantic to argument structure. In *International Journal of Computational Linguistics & Chinese Language Processing, 5*(1), 19-46.
11. Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press
12. Smith, C. S. (1991/1997). The parameter of aspect. Dordrecht, Kluwer.
13. Kilgarriff, A., Huang, C. R., Rychlý, P., Smith, S., & Tugwell, D. (2005). Chinese word sketches. In *Proc ASIALEX 2005: Words in Asian cultural context*. Singapore.