

Exploring the role of verb frames in assessing semantic difficulty

Meichun Liu, Zhuo Zhang

Department of Linguistics and Translation, City University of Hong Kong, China
<https://doi.org/10.36505/ExLing-2021/0039/000512>

Abstract

This paper explores the role of verb frames in semantic difficulty. We built a textbook corpus based on textbooks in primary schools, and the grades of the textbooks were treated as semantic difficulty levels. The difficulty levels were further mapped onto the performance of verb frames. The study found that the verb-frame diversity is highly correlated with semantic difficulty level with statistical significance and that the distribution of verb frames may correspond to the skill theory (Fischer & Corrigan, 1981) in cognitive language development. The skill theory could be used to group verb frames by tiers and roughly predict their distributional trends. Thus, verb frames could be helpful as a semantic reference for textbook planning and readability assessment.

Keywords: verb frames; semantic difficulty; textbooks; readability; skill theory

Introduction

In readability assessment, lexical and syntactic cues are widely used, but semantic cues are quite limited. The newly emergent frame-based features look promising, though they have not yet been incorporated into a working model. Lee, Liu & Cai (2020) first proposed that frame-based features can potentially be utilized in Chinese readability assessment. However, the study just focused on ten frequent verbs across three frames and used less than 1,600 manually annotated sentences to find the potential indicators.

This paper aims to find out how verb frames contribute to semantic difficulty through a quantitative corpus-based study. Two hypotheses are proposed.

H1: There is a strong correlation between verb-frame diversity and difficulty level with statistical significance.

H2: The distributions of verb frames mostly vary in line with the three tiers—sensory-motor, representational, and abstract—proposed in the skill theory of cognitive development (Fischer, 1980; Fischer & Corrigan, 1981).

Literature review

Speakers understand the meaning of a word by knowing its background frame (Fillmore and Atkins, 1992:76-77). The Archi-frames indicate broad semantic domains distinguished by self-contained conceptual schema (Fillmore, 1982).

Language learning is based on actual usage, and language complexity comes from the interaction of cognition and use (Bybee, 2010). As some verbs are learned first in the L1 acquisition of verb-centered constructions (Jing-Schmidt, 2019: 13-31), complex texts may contain more diverse verb frames (H1). This skill theory regards cognitive development as "the construction of hierarchically ordered collections of specific skills," which can theoretically be divided into three tiers: sensory-motor tier, representational tier, and abstract tiers with gradually increased complexity (Fischer, 1980; Fischer & Corrigan, 1981). Therefore, frames from the lower tier should decrease while higher-tier frames should increase as the text becomes more advanced (H2).

Methodology

We propose the verb-centered frame-based model of semantic complexity to map verbs with semantic difficulty and use this model to study how verb frames may influence semantic difficulty. This part mainly introduces the materials and two experimental designs corresponding to the research questions.

Materials

A corpus with about a million characters was built based on nine sets of Chinese textbooks at the primary school level, and the grade levels of the textbooks were used as semantic difficulty levels. Then, verb frames were annotated automatically based on Mandarin VerbNet (<http://mega.lt.cityu.edu.hk/~yufechen/#/>), a Chinese-approximate semantic database with annotated verb frames.

Experimental designs

This section introduces the designs of two experiments.

Experiment#1 We randomly selected 1,000 sentences from textbooks at each grade level, divided them into 50 groups of 20 sentences each, and calculated the average number of the different verb frames (verb-frame diversity) in each group to minimize the effect of sentence length. We then repeated the random selection ten times and calculated the Spearman correlation between the diversity of verb frames and the grade level of the text.

Experiment#2 The proportion of each frame at each difficulty level was calculated, and frames were categorized into the three tiers guided by the skill theory. Linear regression models were employed to explore the frame distribution trends towards difficulty level regarding skill-based frame tiers.

Results and discussion

The results show a strong correlation between the diversity of verb frames and the difficulty level, $r = 0.89 (> 0.8)$, $p < 0.001 (< 0.05)$, with verb frames becoming more diverse as difficulty increases. Furthermore, the distribution of

the frames largely corresponds with the skill theory of cognitive language development (Fischer & Corrigan, 1981), and the distributions of most dominant frames by difficulty fit the regression models with high coefficients of determination. The tier-specific trends of Archi-frame distributions are displayed in Table 1. As the semantic difficulty level increases, the proportions of Tier 1 frames (COMMUNICATION, SELF MOTION & PERCEPTION) decrease, whereas Tier 3 Frame COGNITION increases.

Table 1. Analysis of dominant Archi-frames distribution by three tiers.

Archi-frame	Typical verbs	Tier	Distribution trend
COMMUNICATION	说 (<i>shuō</i> , 'say') 谈 (<i>tán</i> , 'talk')	1: Sensory-motor	Downward
SELF MOTION	来 (<i>lái</i> , 'come') 走 (<i>zǒu</i> , 'walk')	1: Sensory-motor	Downward
PERCEPTION	听 (<i>tīng</i> , 'hear') 触 (<i>chù</i> , 'touch')	1: Sensory-motor	Downward
EXISTENCE	有 (<i>yǒu</i> , 'have') 具有 (<i>jùyǒu</i> , 'possess')	2: Representational	No obvious trend (Slightly upward)
TRANSFERENCE	给 (<i>gěi</i> , 'give') 学习 (<i>xuéxí</i> , 'learn')	2: Representational	No obvious trend (Slightly upward)
COGNITION	相信 (<i>xiāngxìn</i> , 'believe') 决定 (<i>juéding</i> , 'decide')	3: Abstract	Upward

The findings agree with the empirical knowledge of language usage that the more abstract meaning is considered to be more difficult. We further explored other less dominant frames and found similar trends in tiers. Also, even in the same tier, the frames displaying upward trends are usually considered more difficult and easier to be used in a metaphorical way. For example, CAUSED-POSITION and CAUSED-MOTION all belong to Tier 2, but the former displays a downward trend, so the verbs within the frames are generally less difficult than those within the latter frame. For example, CAUSED-POSITION verbs such as 摆放 (*bǎifàng*, 'place'), 搁置 (*qēzhì*, 'put aside'), 镶嵌 (*xiāngqiàn*, 'embed') care more about the action of putting something somewhere whereas CAUSED-MOTION verbs such as 背负 (*bèifù*, 'bear'), 搬动 (*bāndòng*, 'move') and 提携 (*tíxié*, 'support') are grounded in the physical action of carrying something somewhere, which can be easily used in a metaphorical way like the examples below.

1. 背负 压力 / *bèifù yālì* / bear pressure / to bear pressure
2. 用 科学的 力量 搬动 这座 大山
use scientific power move this mountain
yòng kēxuéde lìliàng bāndòng zhèzuò dàshān
to move this mountain with the power of science
3. 互相 提携 / *hùxiāng tíxié* / mutual support / to support each other

Conclusion

As a pioneering quantitative investigation on Chinese verb-frame behaviors by difficulty, the study shows that the diversity of verb frames is highly correlated with semantic complexity with statistical significance. The skill theory could be applied to categorize verb frames by tiers, compare the general difficulty of verb frames, and even roughly predict the distributional trends of several dominant frames. Thus, verb-frame diversity and distribution can potentially serve as a semantic reference for textbook planning and readability assessment.

Acknowledgments

We gratefully acknowledge the grant support (Project#9360163) of the Hong Kong Institute of Data Science, City University of Hong Kong.

References

- Lee, J.S., Liu, M., Cai, T. 2020. Using verb frames for text difficulty assessment. Proceedings of the International FrameNet Workshop 2020: Towards a Global, Multilingual FrameNet, 56–62.
- Fillmore, C.J., Atkins, B. T. S. 1992. Toward a frame-based lexicon: The semantics of RISK and its neighbors. In Lehrer, A., & Kittay, E. (Eds.), *Frames, fields, and contrasts: New essays in semantic and lexical organization* (pp. 75–102). Hillsdale: Erlbaum.
- Fillmore, C.J. 1982. Frame semantics. In the Linguistic Society of Korea (Ed.) *Linguistics in the morning calm* (pp. 111–137). Seoul: Hanshin Publishing Co.
- Bybee, J. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review*, 87, 477–531.
- Fischer, K.W., Corrigan, R. 1981. A skill approach to language development. In R. E. Stark (Ed.) *Language behavior in infancy and early childhood* (pp. 245–273). Elsevier: North Holland.
- Jing-Schmidt Z. 2019. Corpus and Computational Methods for Usage-Based Chinese Language Learning: Toward a Professional Multilingualism. In: Lu X., Chen B. (eds) *Computational and Corpus Approaches to Chinese Language Learning*. Chinese Language Learning Sciences. Springer, Singapore. https://doi-org.ezproxy.cityu.edu.hk/10.1007/978-981-13-3570-9_2