A corpus-based comparative study of light verbs in three Chinese speech communities

Benjamin K. Tsou[1], Ka-Fai Yip[2]

[1]City University of Hong Kong, [1]The Hong Kong University of Science and Technology; [2]Yale University

[1]btsou99@gmail.com, [2]kafai.yip@yale.edu

Light verbs vs. lexical verbs

*1. **give** a pull*   vs. ***give** a book*

Others: *make, take, have …*

2. 做研究 "to do research"   vs. 做蛋糕 "make a cake"

3. 給予解釋 "give an explanation"   vs. 給他一支筆 "give him a pen"

4. 打造光輝未來 "create a bright future"   vs. 打造兵器 "forge a weapon"

Grammatical properties: abstract nouns vs. concrete nouns

Semantic properties: "bleached" meaning vs. lexical meaning

Research focus

1. An understudied yet very common light verb: *da* 打
   - Only several *da* compound verbs are briefly discussed (Wang 1985, Ren 2013)
   - *Cf. jingxin* 進行, *zou* 做... (Lü 1980, Zhu 1985, Diao 2004, Wang 2004 , Huang et al. 2014, *i.a.*)

2. **Metaphorization** and semantic bleaching
   - Key process in light verb development
   - Mainly grammatical studies in the literature

3. New approach to measure **Variations** in Chinese language
   - Mainly Mainland China and Taiwan in the literature
   - Hong Kong?

Roadmap

- Corpus base LiVaC
- Metaphorization and word structure
- Variations in three Chinese speech communities
- Implications

Corpus base LiVaC

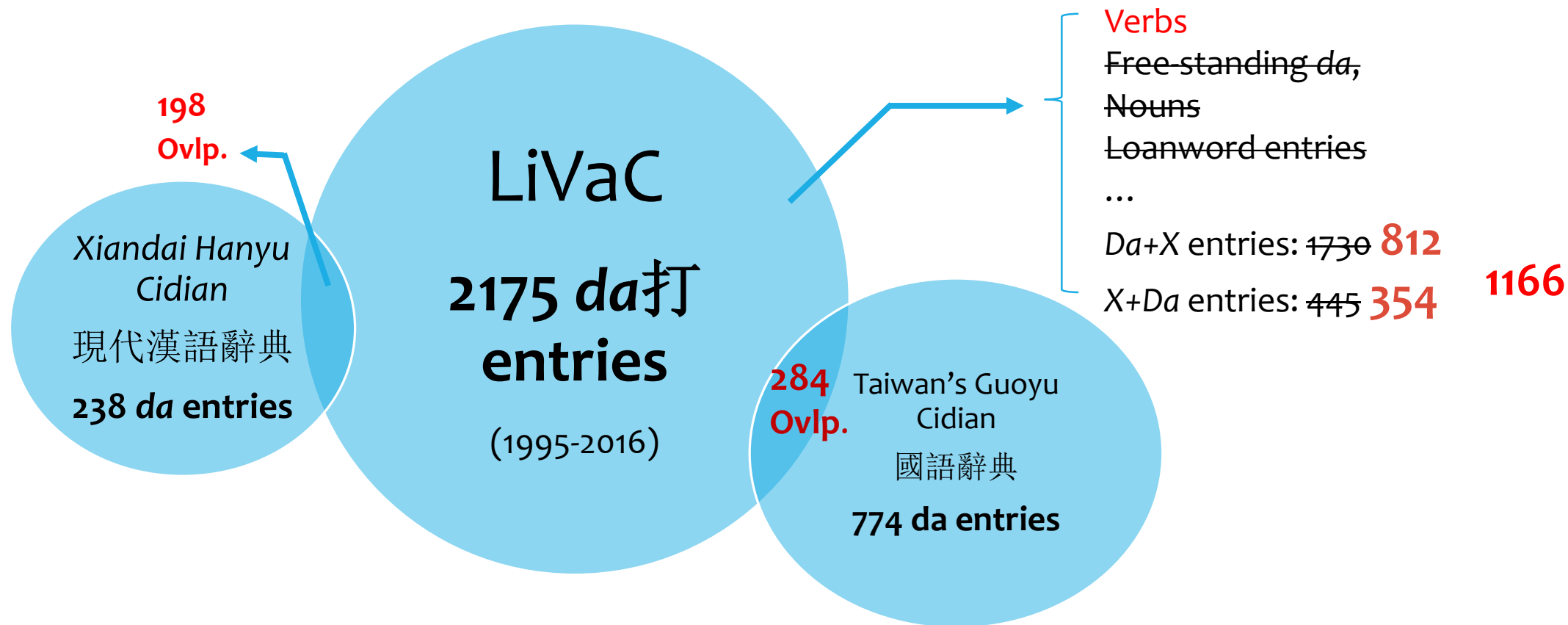# Corpus base: LiVaC *(Linguistic Variation in Chinese Speech Communities)*

Pan-Chinese synchronous database **LIVAC**

(https://en.wikipedia.org/wiki/LIVAC_Synchronous_Corpus)

since 1995

More than 600 million characters of news media texts, rigorously curated with assisted verification, from Pan-Chinese communities:

Beijing, (Guangzhou,) Hong Kong, Macau, Shanghai, (Shenzhen,) Singapore, Taiwan, etc.

Corpus base: LiVaC da entries

198 Ovlp.

Xiandai Hanyu Cidian

現代漢語辭典

238 *da* entries

LiVaC

2175 *da*打 entries

(1995-2016)

284 Ovlp.

Taiwan's Guoyu Cidian

國語辭典

774 da entries

Verbs

~~Free-standing *da*,~~

~~Nouns~~

~~Loanword entries~~

...

Da+X entries: ~~1730~~ **812**

X+Da entries: ~~445~~ **354**

1166

Metaphorization and word structure

Metaphorization

- Major process in grammaticalization (e.g. *is going to*, Bybee & Pagliuca 1985)
- Involves an understanding of one conceptual domain in terms of another (Lakoff & Johnson 1980)

- Working definition for *da* 打:
- **Literal** meaning: physical domain
  - *daji bangqiu* 打擊棒球 "hit a baseball"
- **Metaphorical** meaning: other conceptual domains, e.g. emotion
  - *daji xinqing* 打擊心情 "(lit.)hit-heart status, emotionally striken"

Three types of metaphorization in usage

Type I (Literal):
◦ Only literal meaning, e.g. *dasi* 打死 "beat to death, (lit.) hit-die", *daquan* 打拳 "fist boxing, (lit.) hit-fist"

Type II (Incipient Metaphorization):
◦ Could be used both literally (Type I) and metaphorically (Type III), e.g. dazao打造 "(lit.) fabricate (furniture)" vs. "(metaph.) forge (bright future)", daci打氣 "(lit.) pump air" as in wei luntai daqi 為輪胎打氣 "pump air into tires" vs. "(metaph.) cheer on" as in wei qiuyuan daqi為球員打氣 "cheer the players on".

Type III (Metaphorical):
◦ Only metaphorical meaning, e.g. daya打壓 "suppress, (lit.) hit-press", dajia打假 "crack down on counterfeit, (lit.) hit-falsehood"

Distribution of 3 types by word structure

	entries	Type I	Type II	Type III
X-da打	354	**87%**	8%	5%
da打-X	812	31%	24%	**45%**

◦ *ji**da**擊打* "(lit.) strike-hit"　　　vs.　　*da**ji**打擊* "(lit.) hit-strike"

◦ *Ji**da*** (Type I): *Only* allows literal meaning, e.g. *ji**da** luogu*擊打鑼鼓 "strike gongs and drums"

◦ *Da**ji*** (Type II): Allows metaphorical meaning, e.g. *da**ji** zuian*打擊罪案 "strike at crimes".

Word structure

- *X-da*: Modifier-Verb 偏正/狀中 (e.g. *jida*擊打 "(lit.) strike-hit")
- *Da-X:*
  - (i) Verb-Verb 並列 (e.g. *dazao*打造 "(lit.) forge, fabricate")
  - (ii) Verb-Object 動賓 (e.g. *daci*打氣 "(lit.) pump air")
  - (iii) Verb-Complement 動補 (e.g. *dajin*打進 "compete to enter")

- **Take home message**:
- **Metaphorization is sensitive to word structure**

Variations in three Chinese speech communities

BEIJING, HONG KONG, TAIWAN

Distribution of *da*-X entries

Distribution	Types (%)	Tokens (%)
LIVAC	812 (100)	318,095 (100)
BJ, HK or TW	672 (83)	179,182 (56)
BJ, HK and TW	151 (19)	170,245 (54)

Type II entries: 48 (32%) types, 102,277 (60%) tokens
→ Predominant

Sampling Type II sentences

- Dual usage → must examine authentic sentences

- 1st period: 1995-2000
- 2nd period: 2011-2016

- At most 50 sentences per year for each word in each community
- Total 16955 sentences have been obtained and tagged
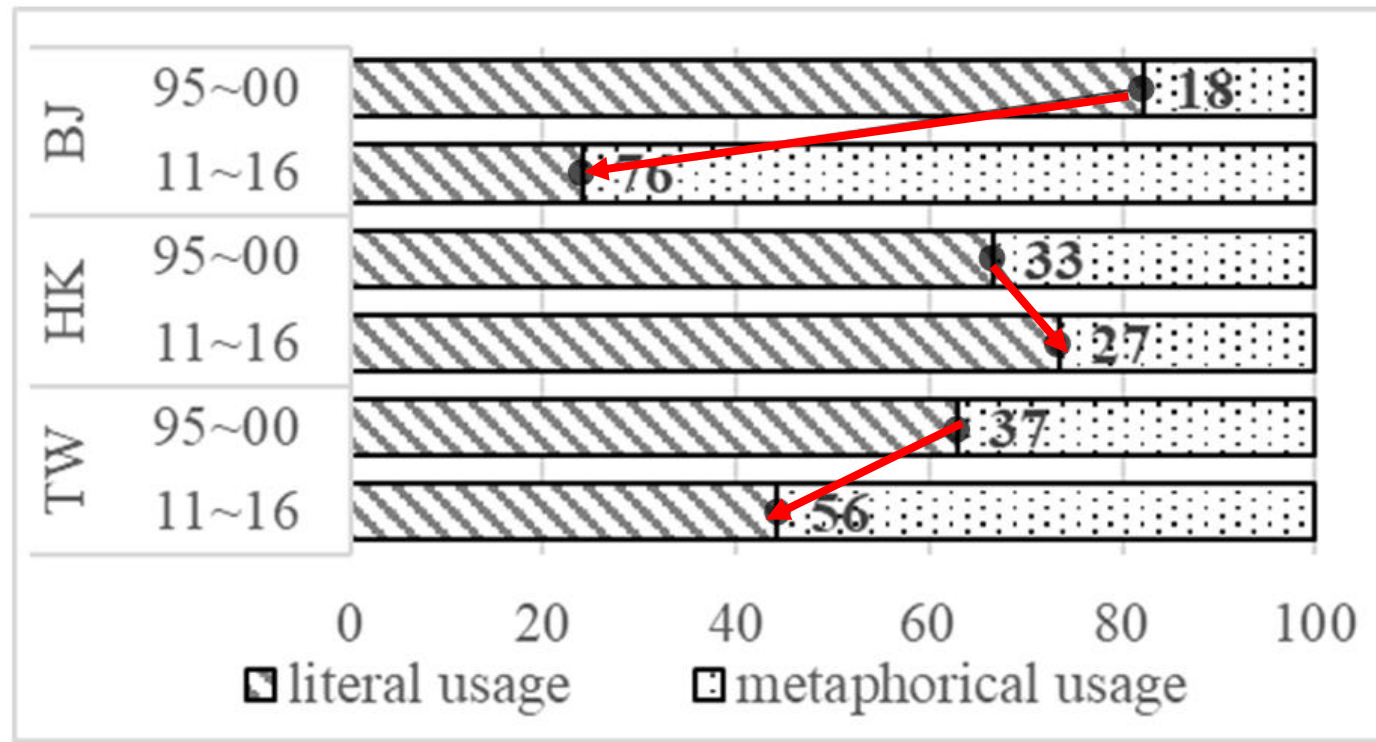- Percentage of metaphorical usage

Metaphorization Index (**MI**)

$$MI = \frac{C + (B_1 + B_2 \ldots + B_n)}{A + B + C} \times 100\%$$

(where A, B, C = total tokens of Types I, II & III respectively, and B_n = tokens of metaphorical usage of a Type II entry)

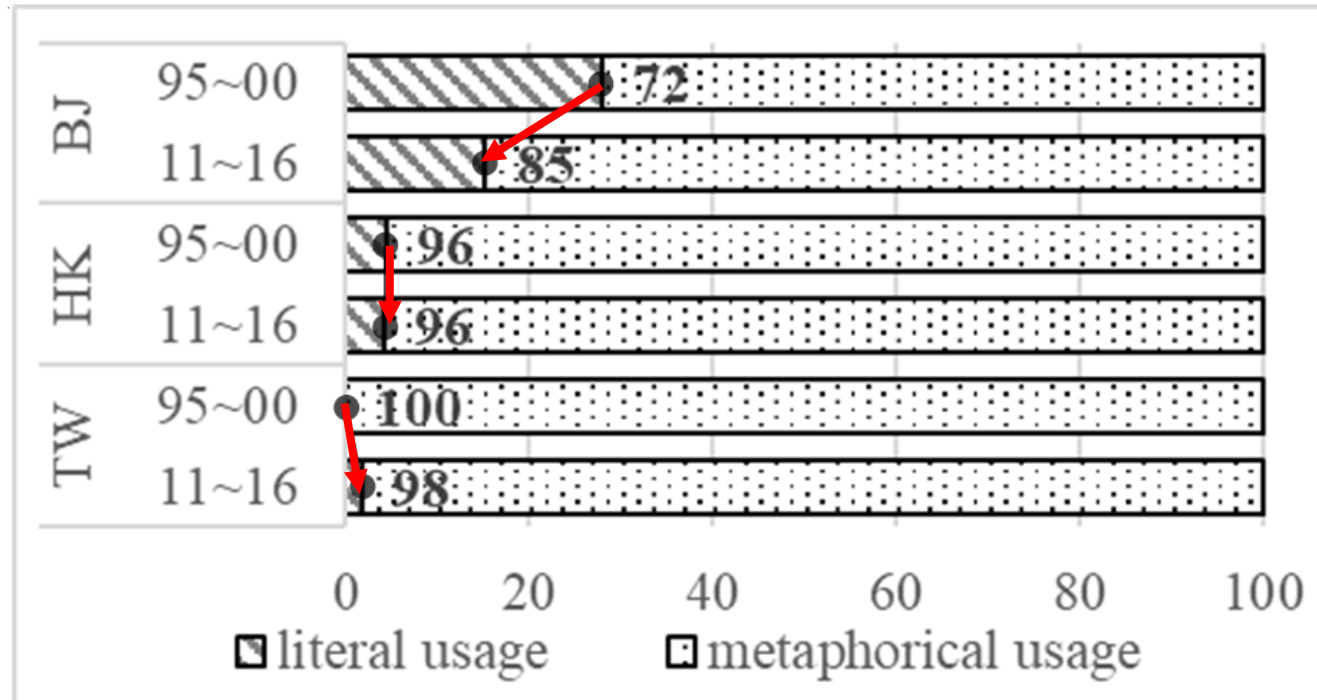- ◦ Aggregated for each community
- ◦ Enable quantitative comparison

Longitudinal variations: *datong* 打通 "hit-through"

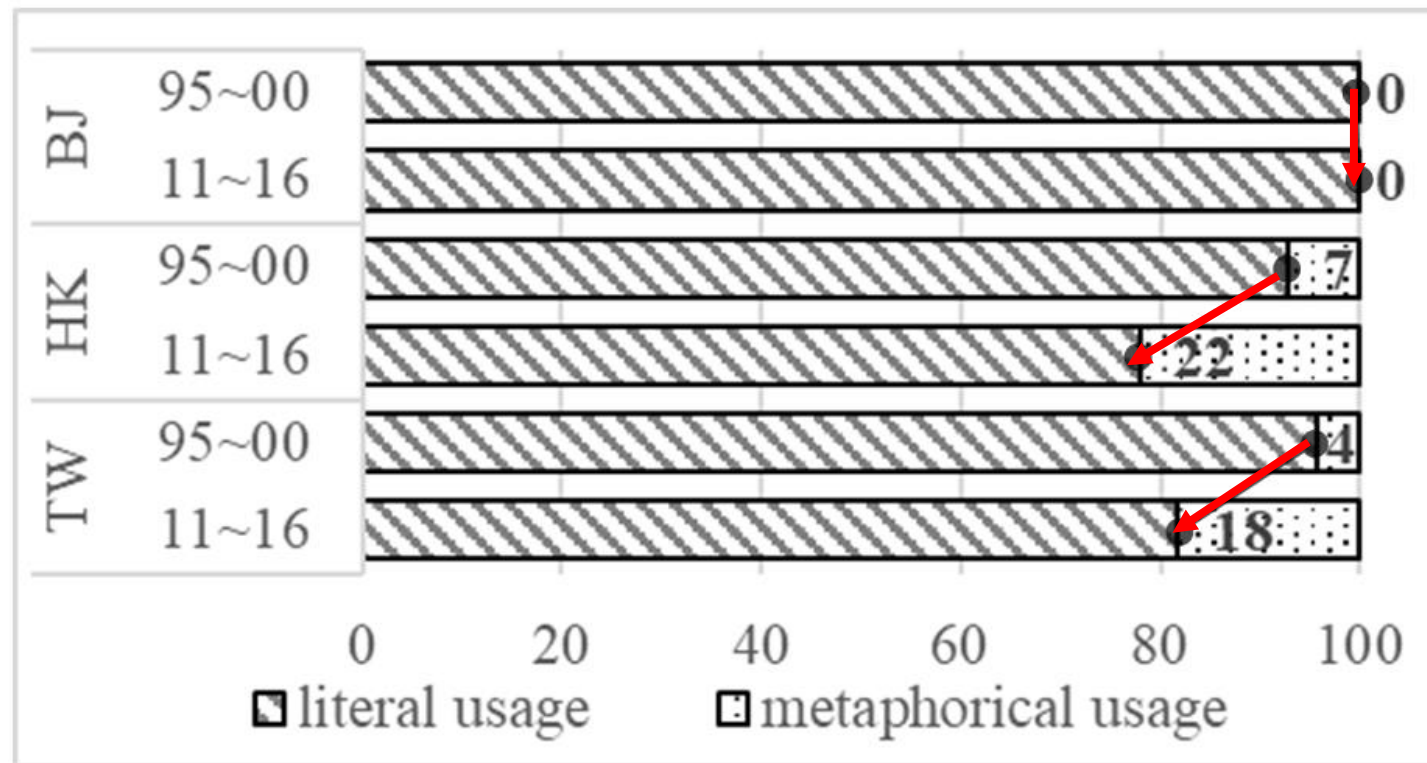◦ BJ shows rapid metaphorization, followed by TW, with HK inert

Longitudinal variations: *daxiang* 打響 "hit-loud"

◦ Metaphorization is almost completed in HK and TW, but is still ongoing in BJ

Longitudinal variations: *dazhang* 打仗 "hit-conflict"

◦ Metaphorization is initiated in HK and TW, but not in BJ

Incipient regional variations:
Top 20 Type II *da*

○ Top percentage of metaphorical usage among three communities

|	
<u>BJ (12/20)</u>:	*dazao*打造 "hit-make", *dapo*打破 "hit-break", *dakai*打開 "hit-open", *dachu*打出 "hit-out", *daxia*打下 "hit-down", *datong*打通 "hit-through", *dajia*打架 "hit-fight", *dajiaodao*打交道 "hit-contact-road", *dasao*打掃 "hit-sweep", *daduan*打斷 "hit-break", *dadiao*打掉 "hit-drop", *dazhong*打中 "hit-at"
<u>HK (6/20)</u>:	*daji*打擊 "hit-strike", *daru*打入 "hit-in", *daqi*打氣 "hit-air", *dacheng*打成 "hit-become", *dazhang*打仗 "hit-conflict", *dadao*打倒 "hit-collapse"
<u>TW (4/20)</u>:	*dajin*打進 "hit-enter", *daqi*打氣 "hit-air", *daxiang*打響 "hit-loud", *dazhong*打中 "hit-at"

Overall incipient regional variations
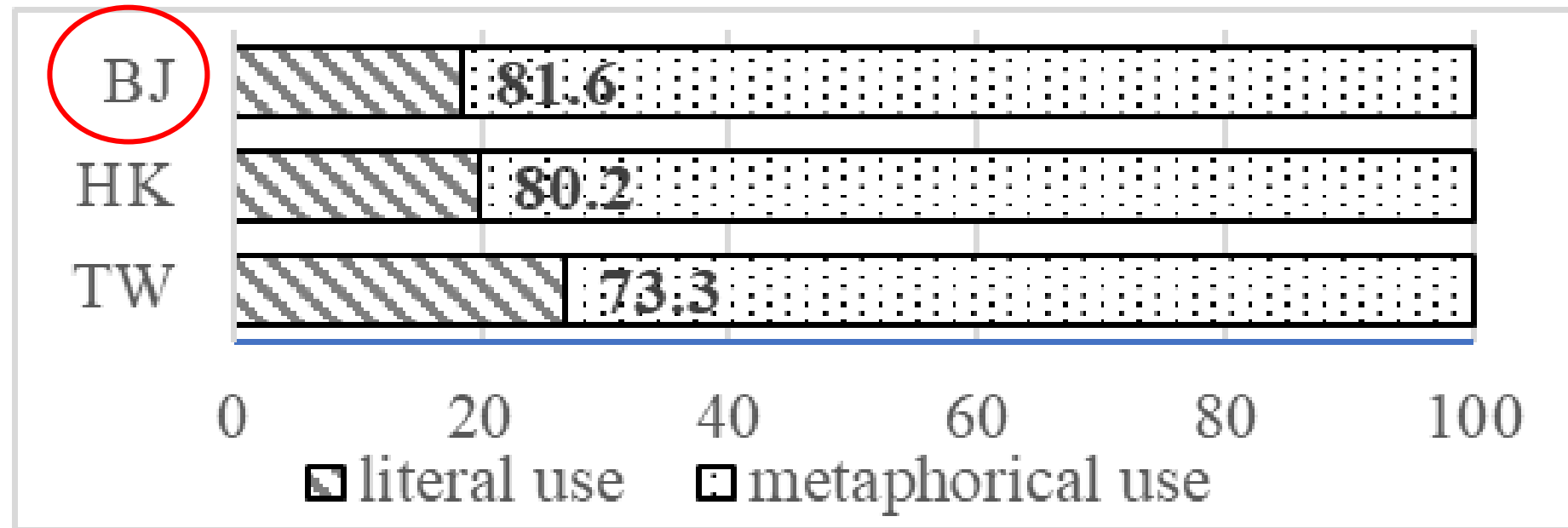## Metaphorical usage of top 20 Type II entries

12 years: 95-00,11-16
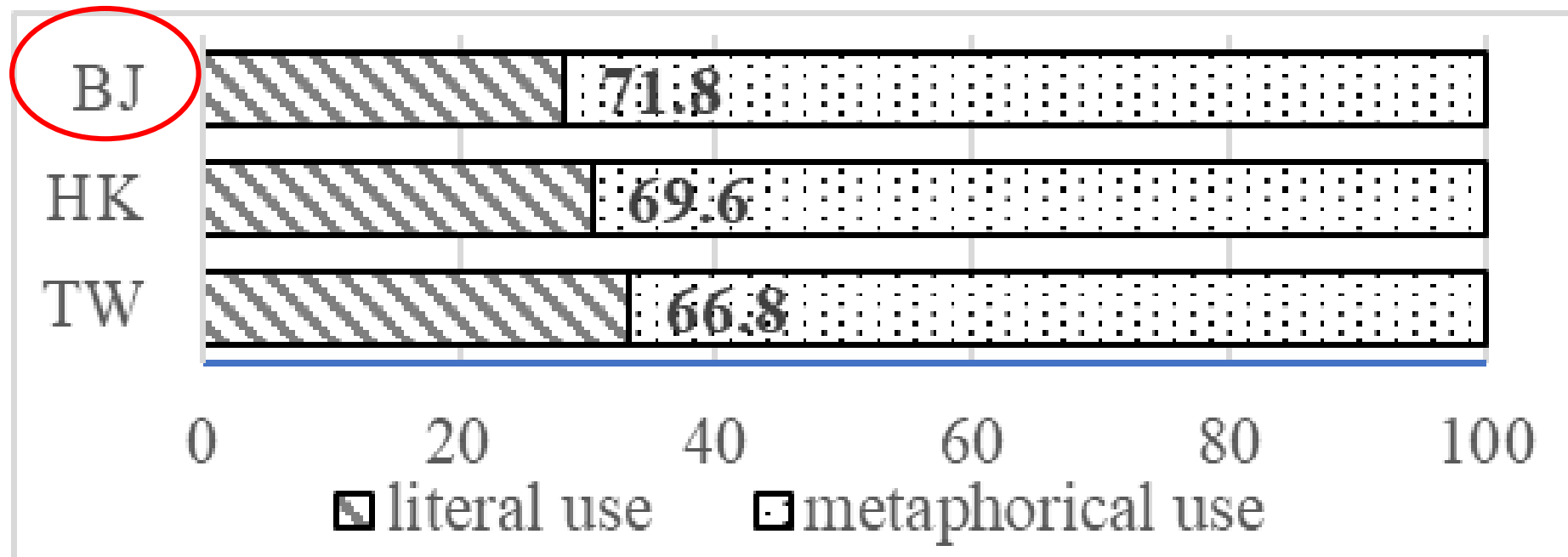
Variations in metaphorization: Aggregated MI for top 50 common da-X verbs

22 years: 95-16

Predominant local entries

◦ Top most frequency (≥90%) local verbs with metaphorical usage, among top 200 entries in each community

BJ (5.5%)	① dafen打分 "to score, (lit.) hit-points", ② daxin打新 "staging (to buy initial round of publicly offered stocks), (lit.)hit-new", ③ danian打蔫 "being tired, (lit.) hit-wither", etc.
HK (19.5%)	① dajiu打救 "to rescue, (lit.) hit-save", ② dazhengqihao打正旗號 "under the name of, (lit.) hit-right-flag", ③ dashuding打書釘 "to read at bookstores without purchase, (lit.) hit-book-nail", etc.
TW (10.5%)	① dalian打臉 "to offend or upset someone/ give a slap in the face, (lit.) hit-face", ② dapi打屁 "to chit-chat, (lit.) hit-fart", ③ dashu打書 "to promote new books, (lit.) hit-book", etc.

Variations in three Chinese speech communities

◦ **Take home message**:

◦ Beijing and Hong Kong show a higher degree of metaphorical shift in different ways

 ◦ BJ: items common to the three communities

 ◦ HK: items unique to their own community

  ◦ could be a source of new metaphorical usage when the items enter other communities through internal language contact

◦ Taiwan: the most conservative in terms of metaphorization

Implications

Correlation between semantic and grammatical properties

◦ For light verbs, the two seem to be correlated: semantic bleaching & subcategorization

◦ (1) Metaphorization is sensitive to word structure

◦ (2) Across Chinese speech communities, TW is the most conservative in terms of metaphorization, AND grammatical properties

  ◦ Jiang et al. (2016): Taiwan light verbs are more transitive ("verbal") than Mainland, and thus are less grammaticalized

Other light verbs:
Jinxing 進行 **"proceed"** & *Zuo* 做 **"make"**

◦ Show significant latitudinal and longitudinal variations in LiVaC (tokens)
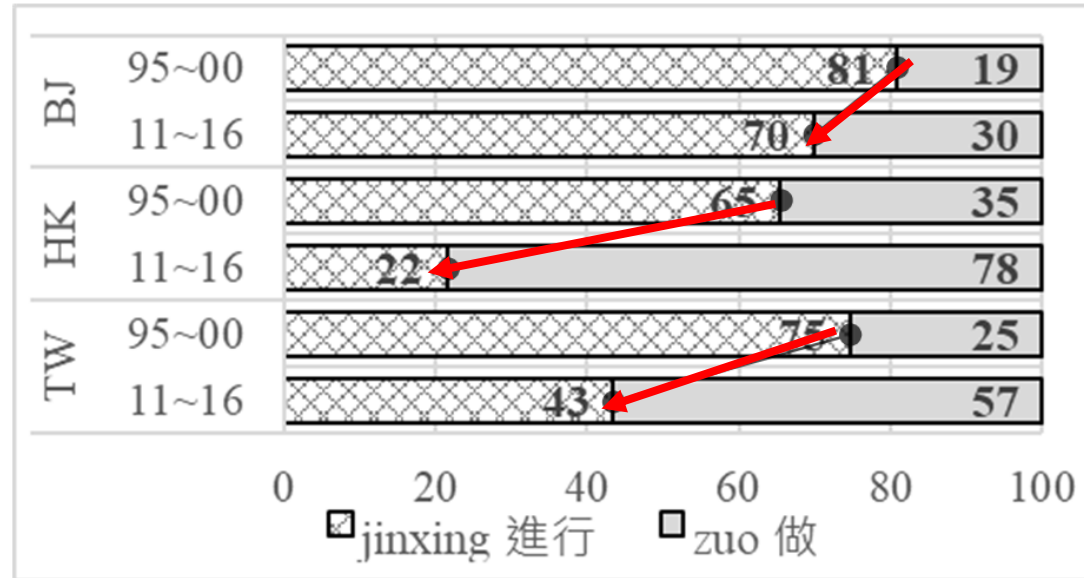
◦ How about metaphorization?
  ◦ E.g. 做研究 "to do research" vs. 做蛋糕 "make a cake"

Concluding remarks

◦ Filling a gap: Light verb *da*

◦ General mechanism of metaphorization and grammatical developments

◦ Dynamics of language variation in Chinese language

References

Cattell, Ray. 1984. *Composite Predicates in English*. North Ryde, New South Wales: Academic Press Australia.

Chae, Hee-Rahk. 1996. Light verb constructions and structural ambiguity. *Proceedings of the 11th Pacific Asia Conference on Language, Information and Computation (PACLIC 11)*: 99-107.

Chen, Yongli. 2003 Xingshi dongci de fanwei, cilei ji tezheng [The scope, subcategories and features of dummy verbs]. *Academic Journal of Jinyang* 3: 92-94.

Diao, Yanbin. 2004. *Xiandai Hanyu Xuyi Dongci Yanjiu* [A Study on Delexical Verb in Modern Chinese]. Dalian: Liaoning Normal University Press.

Grimshaw, Jane, and Armin Mester. 1988. Light verbs and θ-marking. *Linguistic Inquiry* 19: 205-232.

Huang, Chu-Ren, Jingxia Lin, Menghan Jiang, and Hongzhi Xu. 2014. Corpus-based study and identification of Mandarin Chinese light verb variations. In Marcos Zampieri, Liling Tan, Nikola Ljubešic and Jörg Tiedemann, eds, *Proceedings of the First Workshop on Applying NLP Tools to Similar Languages, Varieties and Dialects*, 1-10. Dublin: Association for Computational Linguistics and Dublin City University.

Jespersen, Otto. 1954. *A Modern English Grammar on Historical Principles*. London: George Allen & Unwin & Copenhagen: Ejnar Munksgaard.

Jiang, Haiyan. 2020. Chouxiang dongzuo shijian yuyi fanchou yu Hanyu xingshi dongci [Abstract action event semantic category and Chinese dummy verbs]. Doctoral dissertation, Jilin University.

Jiang, Menghan, Dingxu Shi, and Chu-Ren Huang. 2016. Transitivity in light verb variations in Mandarin Chinese – a comparable corpus-based statistical approach. *Proceedings of the 30th Pacific Asia Conference on Language, Information and Computation (PACLIC 30)*: 459-468.

Lakoff, George, and Mark Johnson. 1980. *Metaphor We Live By*. Chicago and London: University of Chicago.

Li, Feng, and Yun Chai. 1995. "V+yi" lei xuhua dongci de binyu [The objects of "V+yi" dummy verbs]. *Journal of Xinjiang Education Institute* 4: 68-71.

References

Li, Hanlei. 2003. "Jiayi" de yuyong gongneng [The pragmatic functions of *jiayi*]. *Journal of Suzhou Education College* 1: 11-16.

Lin, Jingxia, Hongzhi Xu, Menghan Jiang, and Chu-Ren Huang. 2014. Annotation and classification of light verbs and light verb variations in Mandarin Chinese. In Takuya Nakamura, ed., *Proceedings of Workshop on Lexical and Grammatical Resources for Language Processing*, 75-82. Dublin: Association for Computational Linguistics and Dublin City University.

Lü, Shuxiang. 1980. *Xiandai Hanyu Babaici* [Modern Chinese 800 Words]. Beijing: Commercial Press.

Lü, Shuxiang. 1999. *Xiandai Hanyu Babaici (Zengdingben)* [Modern Chinese 800 Words (Enlarged Edition)]. Beijing: Commercial Press.

Mao, Hongyuan. 1997. Hua xingshi dongci he xingshihua dongci [On formal verbs and formalized verbs]. *Journal of Kashgar Teachers College* 4: 78-79.

Pham, Hoa. 1999. Cognate objects in Vietnamese transitive verbs. *Toronto Working Papers in Linguistics* 27: 227-246.

Ren, Fengmei. 2013. A grounding approach to the semantic meaning of the light verb *da*. In Pengyuan Liu and Qi Su, eds., *Chinese Lexical Semantics: 14th Workshop, CLSW 2013, Zhengzhou, China, May 10-12, 2013. Revised Selected Papers*, 88-96. Berlin & Heidelberg: Springer-Verlag.

Shen, Jiaxuan, and Jiangzhi Zhang. 2013. Ye tan xingshi dongci de gongneng [On the grammatical function of dummy verbs in Chinese]. *TCSOL Studies* 2: 8-17.

Su, Lei. 2009. Lun "da" de yufahua [On the grammaticized process of the word "fight"]. *Journal of Hubei University of Education* 5: 30-31.

Tsou, Benjamin K., and Oi Yee Kwong. 2015. LIVAC as a monitoring corpus for tracking trends beyond linguistics. *Journal of Chinese Linguistics Monograph Series* 25: 447-471.

Tsou, Benjamin K., Andy C. Chin, and Oi Yee Kwong. 2011. From synchronous corpus to monitoring corpus, LIVAC: The Chinese case. In Friedrich Laux and Lena Strömbäck, eds., *The Third International Conference on Advances in Databases, Knowledge, and Data Applications (DBKDA 2011)*, 175-180. St. Maarten: IARIA XPS Press.

Wang, Leslie Fu-mei. 2004. A corpus-based study of mandarin verbs of doing. *Concentric: Studies in Linguistics* 30.1: 65-85.

Wang, Li. 1985. *Zhongguo Xiandai Yufa* [Modern Chinese Grammar]. Beijing: Commercial Press.

References

Yan, Zhongsheng. 1998. Shuo "houxu dongcixing binyu dongci" [On "verbs taking verbal objects"]. *Journal of Hebei Normal University (Natural Science)* 2: 91-93.

Zhou, Gang. 1987. Xingshi dongci de cifenlei [Subdivision of dummy verbs]. *Chinese Language Learning* 1: 11-14.

Zhu, Dexi. 1985. Xiandai shumian hanyu li de xuhua dongci he mingdongci [Dummy verbs and nominal verbs in Modern Written Chinese]. *Journal of Peking University (Humanities and Social Sciences)* 5: 1-6.

Zhu, Jian-jun. 2004. "Da-V" zhi "da" de yufahua tanxi [The grammaticalization of *da* in the structure of *da-V*]. *Research in Ancient Chinese Language* 3: 38-44.

Zhuang, Huibin. 2014. Xiandai Hanyu Qingdongci "da" de laiyuan chuyi [On the etymology of the light verb *da* in Chinese]. *Language Teaching and Linguistic Studies* 3: 67-74.

Thank you!

ANY COMMENTS ARE WELCOME!