MSC Variations in Metaphorization among Pan-Chinese Speech Communities

Benjamin K. Tsou, Ka-Fai Yip

City University of Hong Kong/The Hong Kong University of Science and Technology; Yale University

btsou99@gmail.com; kafai.yip@yale.edu

Modern Standard Chinese (MSC) and its varieties

◦ Beijing, Hong Kong, Macau, Shanghai, Singapore, Taiwan, etc. (in alphabetical order)

◦ Different dialectal backgrounds: Mandarin, Cantonese, Taiwanese/Southern Min, Hakka…

◦ Different socio-cultural backgrounds

◦ Previously studies: mainly lexical variations, or *morpheme-based* syntactic constructions (Shi et al. 2006, Diao 2014, Lin et al. 2018, *i.a.*)

Metaphorization

◦ A semantic *process*

◦ Major process in grammaticalization (e.g. *is going to,* Bybee & Pagliuca 1985)
  ◦ → crucial in language change and language variations

◦ E.g. *daci* 打氣 "(lit.) pump air"
  ◦ *wei luntai daqi* 為輪胎打氣 "pump air into tires" vs.
  ◦ "(metaph.) cheer on": *wei qiuyuan daqi*為球員打氣 "cheer the players on"

◦ How the process varies among Pan-Chinese speech communities?

Light verbs vs. lexical verbs

1. ***give*** *a kick*   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

3. New approach to measure **covert variations** in Chinese language
   - A *process-based* approach
   - Differential and parallel developments

Roadmap

- Corpus base LiVaC
- Metaphorization in light verbs
- Variations in Pan-Chinese speech communities
- Concluding remarks

Corpus base LiVaC

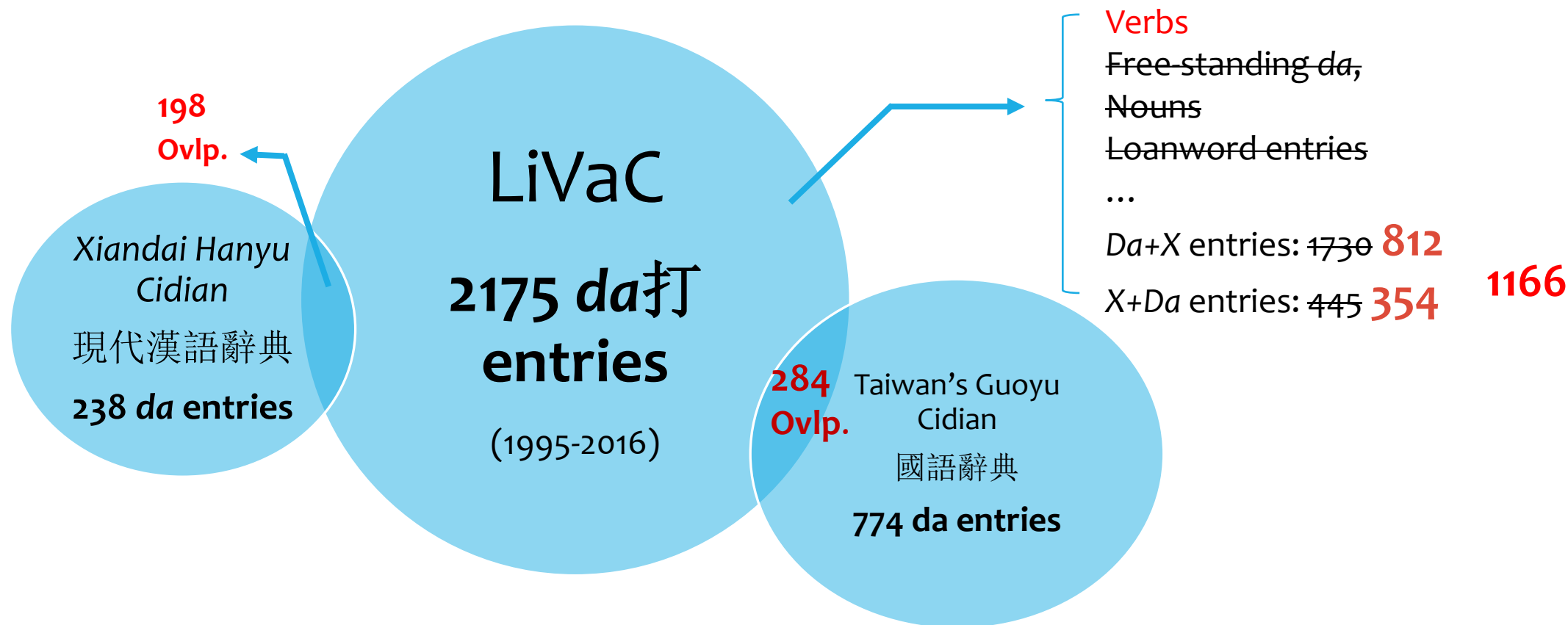# Corpus base: LiVaC *(Linguistic Variation in Chinese Speech Communities)*

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 in light verbs

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".
- **Metaphorization is sensitive to word structure**

Variations in Pan-Chinese speech communities

BEIJING, HONG KONG, MACAU, TAIWAN

Distribution of *da*-X entries

Distribution	Types (%)	Tokens (%)
LIVAC	812 (100)	318,095 (100)
BJ, HK, MC or TW	672 (83)	197,486 (62)
BJ, HK, MC and TW	124 (18)	186,286 (59)

Type II entries: 40 (32%) types, 115,006 (62%) tokens
→ Predominant

Morpheme-based comparison

○ Comparing the shared entries of *da-X* among communities

○ By word tokens (instead of word types)

○ Normalized tokens (Int%)

○ i.e. Word tokens divided by total tokens in that region

    ○ E.g. ***daji*** 打擊 "(lit.) hit-strike"

    ○ 20% in BJ;   17% in HK;   40% in MC;   14% in TW

Shared tokens

◦ Take BJ and TW as reference points

◦ HK and MC: both are a Cantonese-speaking society

◦ Shared normalized tokens (int%) among the top 10 common *da* verbs

Top10-int%	vs.BJ	vs.TW
HK	42	47
MC	69	65

◦ **BJ has more influence on MC than on HK**

◦ **TW has more influence on HK than on MC**

Process-based comparison

- Measure the developments of the metaphorization process

- **Type II *da-X* verbs:**
  - Metaphorical polysemy (i.e. covering both literal and metaphorical usage)
  - Still developing → a window to observe metaphorization development

- How the use (literal vs. metaphorical) varies over time in different communities

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

Metaphorization Index (MI)

◦ Measured for each word

◦ The degree of metaphorization of a type II verb

$$MI = \frac{M}{T} \times 100\%$$

（M：虛化句例數量、T：句例總數）

(M: metaphorical tokens)
(T: total tokens)

Overall Metaphorization Index (OMI)

◦ Aggregated for each community over time
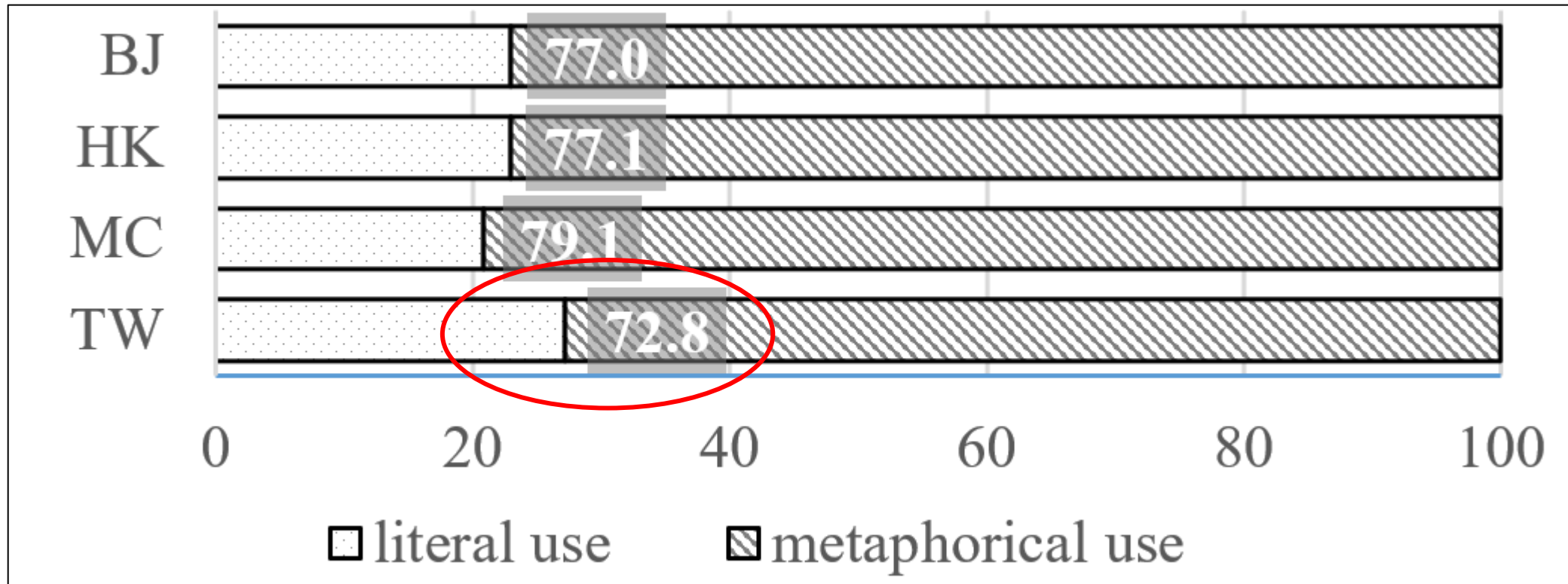◦ Enable quantitative comparison: regional variations and longitudinal variations

$$OMI = \frac{C + (Bm_1 + Bm_2 \ldots + Bm_n)}{A + B + C} \times 100\%$$

(A、B、C：第I、II、III類詞總詞標、Bm_n：任一第II類詞虛化用法總詞標)

(where A, B, C = total tokens of Types I, II & III respectively, and Bm_n = tokens of metaphorical usage of a Type II entry)
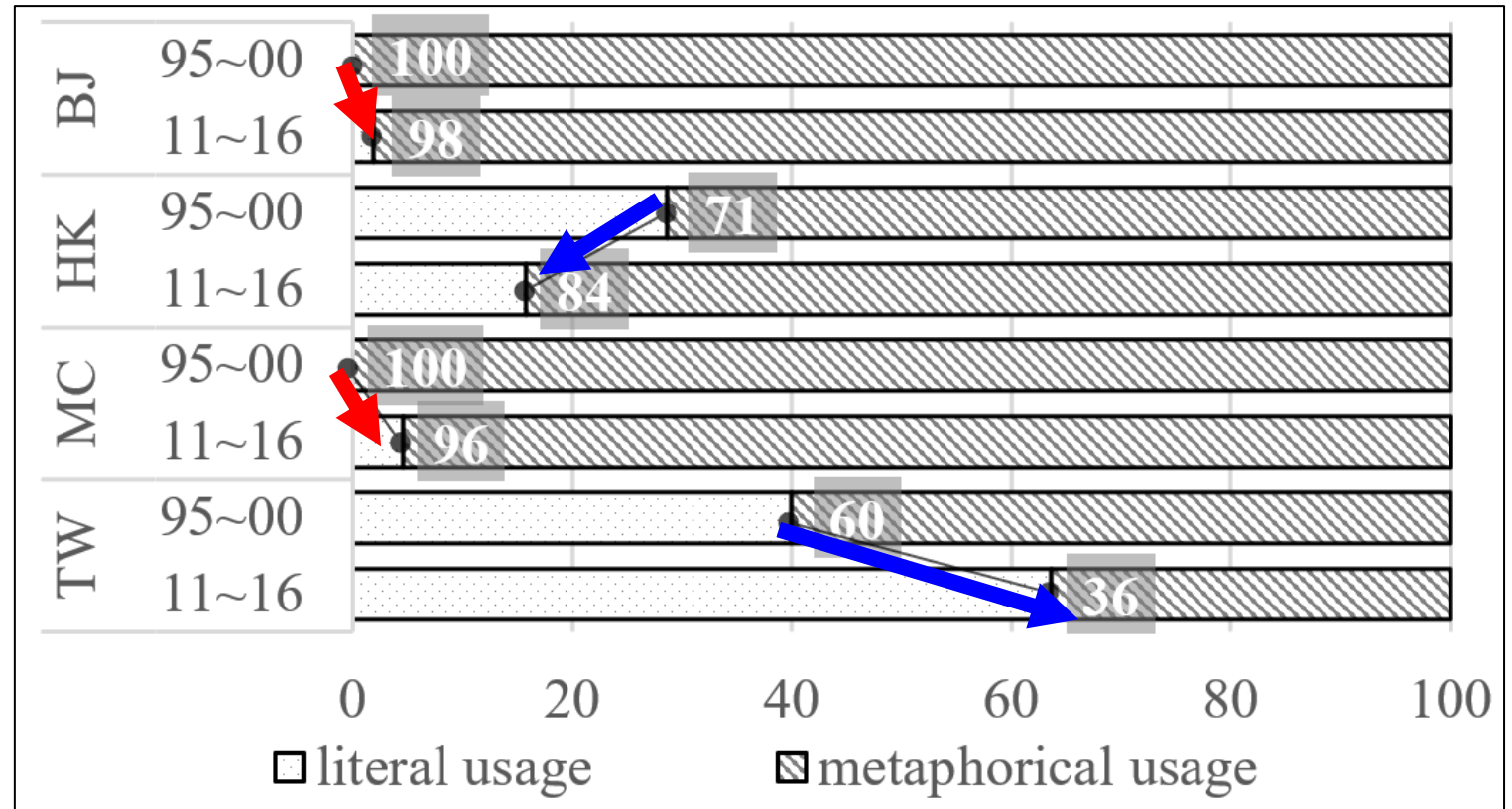
Variations in metaphorization:
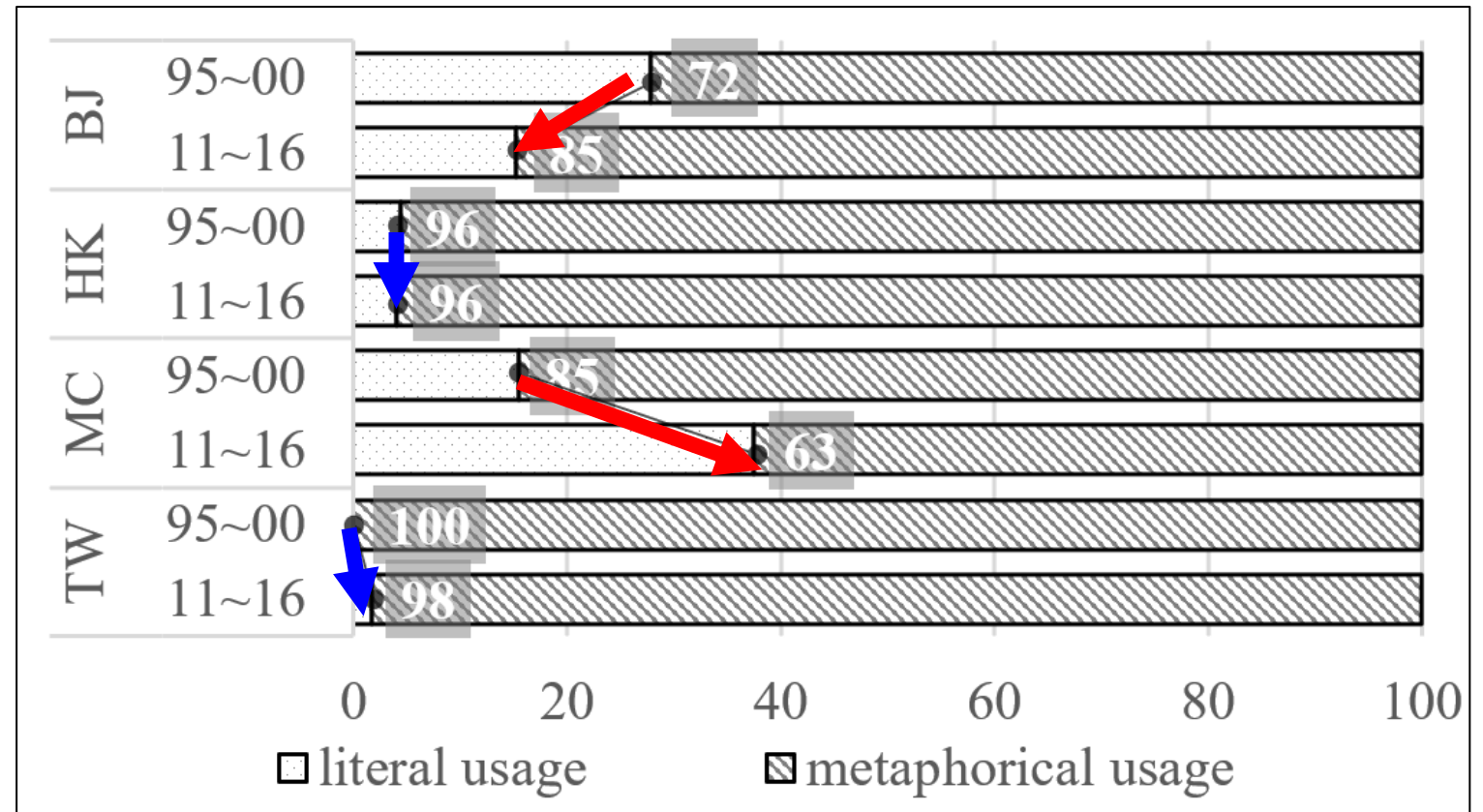## Overall MI for 124 shared da-X verbs

22 years: 95-16

Longitudinal variations:
dadiao 打掉 "hit-down"

◦ Metaphorization is almost completed in BJ and MC, but is still ongoing in HK, and even going backward in TW
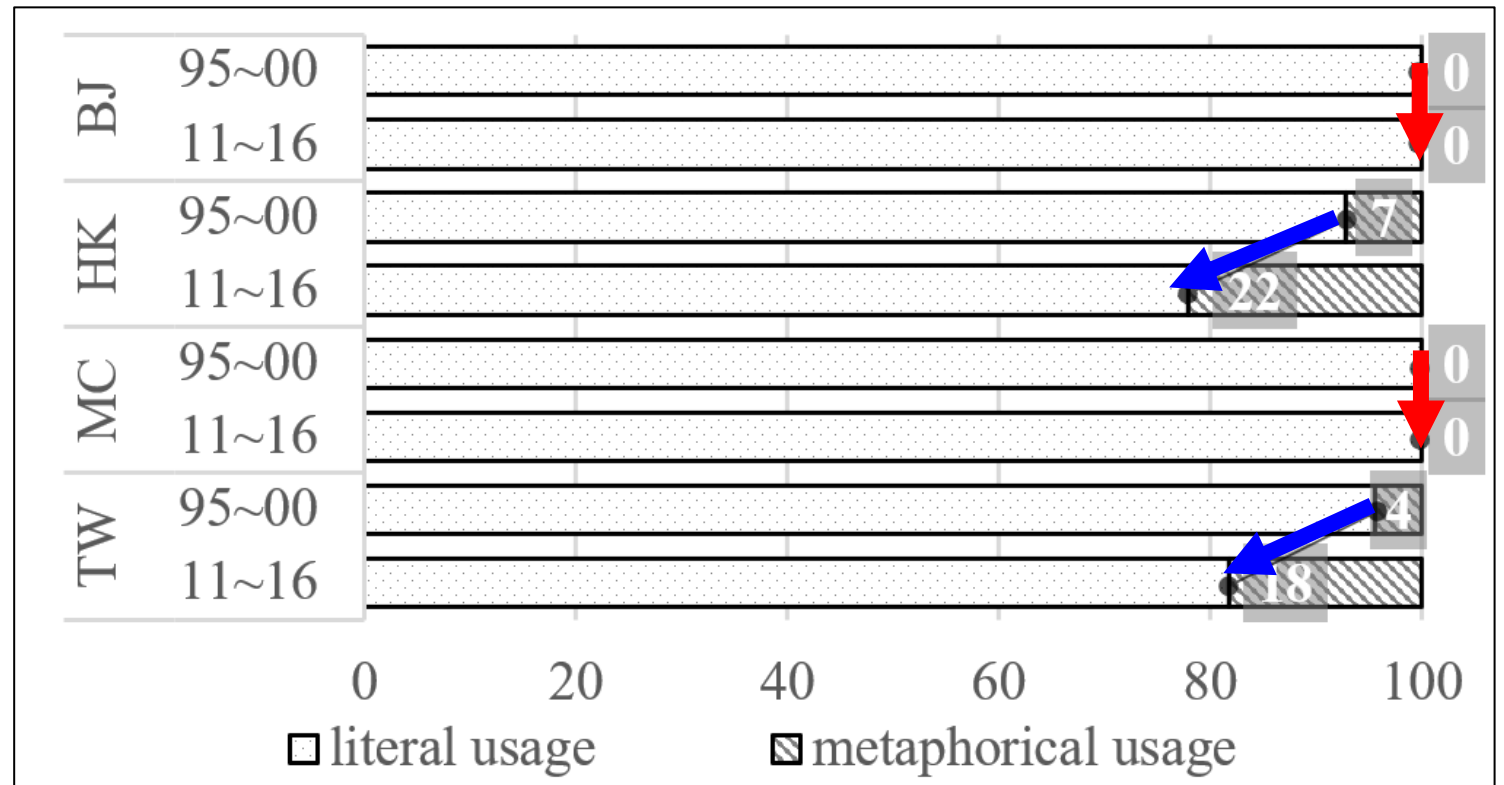
Longitudinal variations:
daxiang 打響 "hit-loud"

◦ Metaphorization is almost completed in HK and TW, but is still ongoing in BJ, and even going backward in MC

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

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

Parallel and differential developments

◦ Among the top 20 type II verbs, those with **parallel developments** are measured by normalized tokens (Int%)

Top20 (int%)	vs.BJ	vs.TW
HK	6	17
MC	10	8.5

◦ HK and TW: *dazhang* 打仗 'hit-conflict' (↑), *dajin* 打進 'hit-enter' (−), …

◦ MC and BJ: *datong* 打通 'hit-through' (↑), *daxia* 打下 'hit-below' (−), …

◦ Differential developments: HK//TW vs. MC//BJ

A closer look at *dazhang* 'hit-conflict' (HK)

- From '(military) war' to metaphorical 'compete/conflict'
- 林鄭：不覺與鄉議局◆打仗
- …政府推出政策，若遇反對意見，就埋首「◆打仗」，盲目捍衛政策方案 …

- …該劇被指拿來跟港視開台劇對撼◆打仗，祥仔認為不算是…

- …但稱未曾料及事件拖延一年半，「估唔到（余）要◆打仗打咁耐」…

- Conceptual domains: politics, entertainment, …
- Quotes: internal language contact with Cantonese

A closer look at *dazhang* 'hit-conflict' (TW)

○ From '(military) war' to metaphorical 'compete'

○ …文創政策形成多頭馬車，如此「一國兩制」，如何對外◆打仗？
○ …也是把在柯文哲競選時運用年輕人◆打仗的模式，「完全複製」到小英身上。…

○ Conceptual domain: mainly politics (including elections)
○ The role of socio-cultural factors in metaphorization developments

Variations in Pan-Chinese speech communities

◦ **Take home message**:

◦ While the overall extent of metaphorization is similar in Pan-Chinese communities (TW being the most conversative), the **internal developments of metaphorization** differ considerably

◦ **Parallel** and **differential** developments:
HK and TW vs. MC and BJ

◦ Next step: why? (linguistic and non-linguistic factors)

Concluding remarks

Concluding remarks

- A *dynamic* perspective:
  - how the development of a linguistic process varies in Chinese language
  - Covert variations (as compared to overt lexical/syntactic variations)

- Metaphorization and grammaticalization
  - Light verb: a grammatical category
  - See Jiang et al. (2016) for grammatical variations of light verbs in BJ and TW

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

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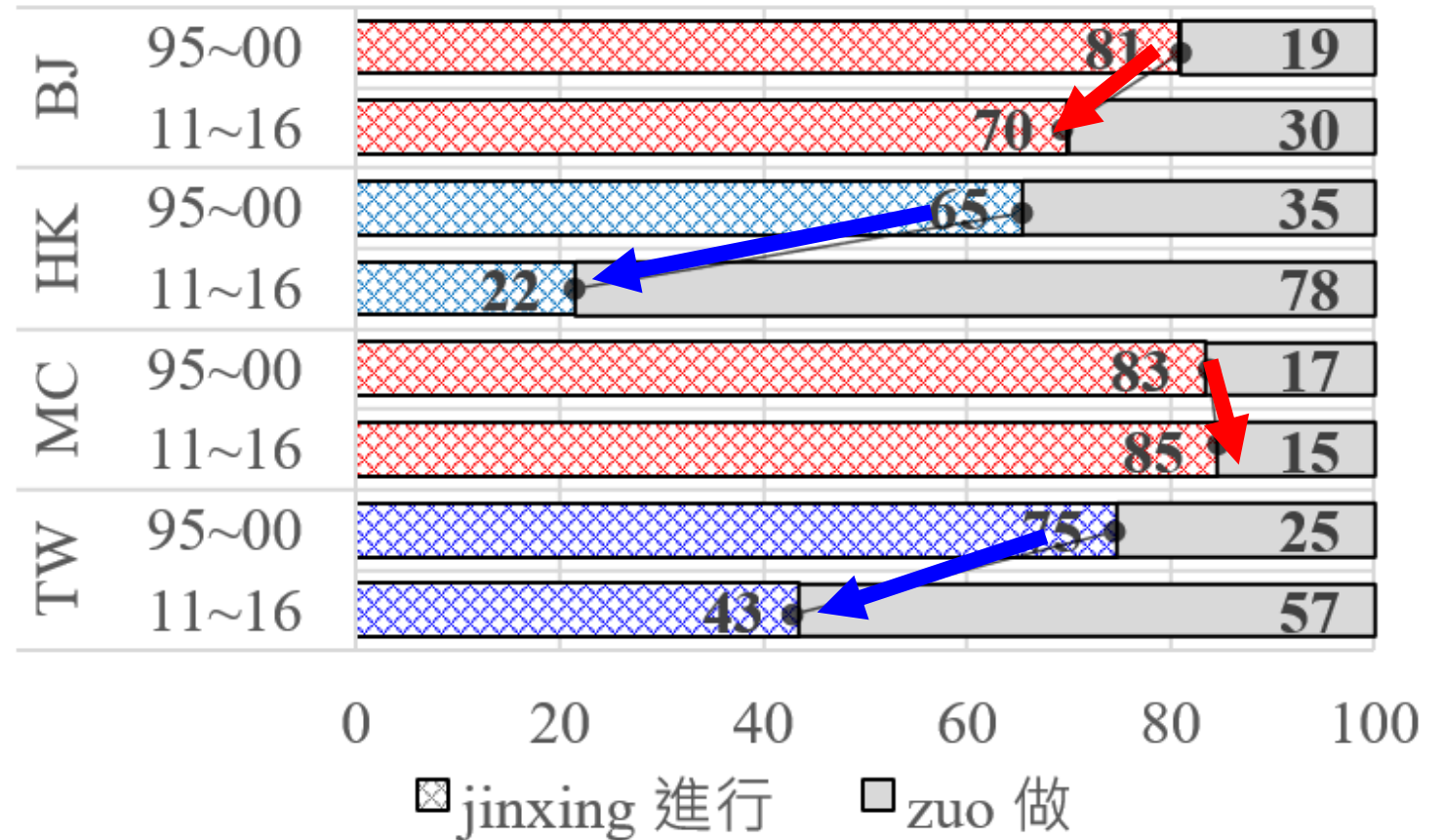
References

1. Cattell, Ray. 1984. *Composite Predicates in English*. North Ryde, New South Wales: Academic Press Australia.
2. 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.
3. 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.
4. Diao, Yanbin. 2004. *Xiandai Hanyu Xuyi Dongci Yanjiu* [A Study on Delexical Verb in Modern Chinese]. Dalian: Liaoning Normal University Press.
5. Grimshaw, Jane, and Armin Mester. 1988. Light verbs and θ-marking. *Linguistic Inquiry* 19: 205-232.
6. 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.
7. Jespersen, Otto. 1954. *A Modern English Grammar on Historical Principles*. London: George Allen & Unwin & Copenhagen: Ejnar Munksgaard.
8. 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.
9. 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.
10. Lakoff, George, and Mark Johnson. 1980. *Metaphor We Live By*. Chicago and London: University of Chicago.
11. 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

1. Li, Hanlei. 2003. "Jiayi" de yuyong gongneng [The pragmatic functions of *jiayi*]. *Journal of Suzhou Education College* 1: 11-16.
2. 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.
3. Lü, Shuxiang. 1980. *Xiandai Hanyu Babaici* [Modern Chinese 800 Words]. Beijing: Commercial Press.
4. Lü, Shuxiang. 1999. *Xiandai Hanyu Babaici (Zengdingben)* [Modern Chinese 800 Words (Enlarged Edition)]. Beijing: Commercial Press.
5. Mao, Hongyuan. 1997. Hua xingshi dongci he xingshihua dongci [On formal verbs and formalized verbs]. *Journal of Kashgar Teachers College* 4: 78-79.
6. Pham, Hoa. 1999. Cognate objects in Vietnamese transitive verbs. *Toronto Working Papers in Linguistics* 27: 227-246.
7. 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.
8. 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.
9. Su, Lei. 2009. Lun "da" de yufahua [On the grammaticized process of the word "fight"]. *Journal of Hubei University of Education* 5: 30-31.
10. 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.
11. 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.

References

1. Tsou, Benjamin K., and Ka-Fai Yip. 2020. A corpus-based comparative study of light verbs in three Chinese speech communities. In Minh Le Nguyen, Mai Chi Luong, and Sanghoun Song, eds., *Proceedings of the 34th Pacific Asia Conference on Language, Information and Computation (PACLIC-34)*, 302–311. Association for Computational Linguistics.

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

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

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

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

6. 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.

7. 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.

8. 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.

Acknowledgements

- For data preparation and annotation, we wish to thank Wing Fu Tsoi, Janice Chong, Clara Hui, Steffi Lo, Yuki Wong, and especially Renee Ji.

Thank you!

ANY COMMENTS ARE WELCOME!