

Bridging phonology, meaning, and written form across time:

Introducing a database of Chinese ideophones — CHIDEOD

Thomas Van Hoey
National Taiwan University
國立台灣大學

Arthur L. Thompson
The University of Hong Kong
香港大學

ILL-12
12th International Symposium on
Iconicity in Language and Literature
3-5 May 2019, University of Lund

Ideophones: a typologically widespread phenomenon

Dingemanse's definition

marked

words

that **depict**

sensory imagery (2011; 2012)

[and that belong to an **open lexical class**] (2019)

What we know about Chinese ideophones

Some cross-linguistically notable features of ideophones
(cf. a.o. Dingemanse & Akita 2016)

- prosodic foregrounding
- often less integrated in sentences
- often marked by reduplication
- expands on the phonological system of prosaic words
- gesture
- variable written forms

What we know about Chinese ideophones

Some cross-linguistically notable features of ideophones
(cf. a.o. Dingemanse & Akita 2016)

- prosodic foregrounding
- often less integrated in sentence
- often marked by reduplication
- expands on the phonological system of prosaic words
- gesture
- variable written forms

(1) 她 咻地一聲 跑過去了。
tā xiū=de=yì-shēng pǎo-guò-qù=le
she IDEO=ADV one-sound run-past-go=PFV
“Shoow, she ran by.”

(2) 飛機 咻咻咻 飛過去。
fēijī xiū xiū xiū fēi-guò-qù
airplane IDEO IDEO IDEO fly-over-go.
The planes whizzed over.

What we know about Chinese ideophones

Some cross-linguistically notable features of ideophones
(cf. a.o. Dingemanse & Akita 2016)

- prosodic foregrounding
- often less integrated in sentence
- often marked by reduplication
- expands on the phonological system of prosaic words
- gesture
- variable written forms

(1) 她 咻地一聲 跑過去了。
tā **xiū**=de=yì-shēng pǎo-guò-qù=le
she **IDEO**=ADV one-sound run-past-go=PFV
“Shoow, she ran by.”

(2) 飛機 咻咻咻 飛過去。
fēijī **xiū xiū xiū** fēi-guò-qù
airplane **IDEO IDEO IDEO** fly-over-go.
The planes whizzed over.



Actor Jackie Chan performing newly coined ideophone
duāng ‘very black, thick and smooth hair’ > ‘wow!’

成龍 成龍 Cheng Long ‘Jackie Chan’

What we know about Chinese ideophones

Chinese research on Chinese ideophones is mostly concentrated on onomatopoeia (ideophones that depict sound).

Zhao Aiwu 赵爱武 (2008)

Qiu Di 邱迪 (2018)

What we know about Chinese ideophones

Chinese research on Chinese ideophones is mostly concentrated on onomatopoeia (ideophones that depict sound).

Zhao Aiwu 赵爱武 (2008)
Qiu Di 邱迪 (2018)

However, there are some broader discussions of ideophones

- phonology
- Beijing dialect
- Southern Sinitic
- Cantonese (vs. Dagaare)
- Mandarin (vs. Japanese)
- Japanese (vs. Mandarin)
- reduplication in Old Chinese
- Middle Chinese

Mok (2001); Thompson (2018)
Meng (2012)
Wu (2014)
Bodomo (2006)
You (2015)
Lu (2006)
Sun (1999)
Van Hoey (2015)

What we know about Chinese ideophones

Chinese research on Chinese ideophones is mostly concentrated on onomatopoeia (ideophones that depict sound).

Zhao Aiwu 赵爱武 (2008)
Qiu Di 邱迪 (2018)

However, there are some broader discussions of ideophones

- phonology
- Beijing dialect
- Southern Sinitic
- Cantonese (vs. Dagaare)
- Mandarin (vs. Japanese)
- Japanese (vs. Mandarin)
- reduplication in Old Chinese
- Middle Chinese

Mok (2001); Thompson (2018)
Meng (2012)
Wu (2014)
Bodomo (2006)
You (2015)
Lu (2006)
Sun (1999)
Van Hoey (2015)

These studies often append data to their work (good!)
but the data is not standardized so not always reusable (less good).

How can we unify what we know about Chinese ideophones?

We need to centralize these data so they can be reused

- Dedicated studies
- Dictionaries
- (scattered examples)

Our answer: **CHIDEOD** — **the Chinese Ideophone Database**

Collecting data from these sources,
storing them in a user-friendly dataset and repository,
provide a number of formal and semantic variables that can be explored

Similar databases: BCCWJ's word profiler

Balanced Corpus of Contemporary Written Japanese (NINJAL 2016) has a word profiler (LWP)

The word profiler looks up words in the BCCWJ and provides sketch grammar-like statistics.

NINJAL-LWP for BCCWJ

fuwafuwa 絞り込み 元に戻す

all noun verb adj. conj. adv. mimetic
すべて 名詞 動詞 形容詞 連体詞 副詞 オノマトペ

見出し	読み	ローマ字表記	頻度
ふわふわ	フワフワ	fuwafuwa	414

The goal of CHIDEOD is to collect all TYPES,

which later could be used in a corpus study

Similar databases: MEJaM

The Multimedia Encyclopedia of Japanese Mimetics (Akita 2016)

bururu
[collocation]
verb
adjective
nouns

Body movement / 体の動き

ぶるぶる

[コロケーション]

動詞：震える / shiver

形容詞：寒い / cold

名詞：手 / hand、唇 / lip、身体 / body、眼 / eye

[Google画像]



buru~buru

CHIDEOD might eventually evolve into a Multimedia CHIDEOD

Which would include pictures and video clips to illustrate the depictive nature.

However, given the diachronic and synchronic scope, this may not be realizable for all items.

Similar databases: Quechua Real Words

Audiovisual ANTI-dictionary of expressive Quechua ideophones

(Nuckolls 2017; Nuckolls & Swanson 2019)

The goal is to study the multimodal interaction between Quechua ideophones and gesture through video clips

Subsets of data provided in CHIDEOD can aid in researching how multimodality(gesture) interacts with Chinese ideophones.



polan

CHIDEOD: why and where



- digitization of data
- centralization of data
- exploration
 - semantics
 - phonology
 - ortography
 - historical
- expandable research resource rather than the finalized tool
- type frequencies
(not yet token frequencies)

Open source project available at OSF
<https://osf.io/kpwgf/>

Available as online app
<https://simazhi.shinyapps.io/Chineseideophone/>

Also available as an R package
(see osf website)

CHIDEOD is structured in a *tidy* format

- Ever growing resource
- Modeled after the recent [Chinese Lexical Database](#) Sun et al. (2018)
- Lives mostly as a large *tidy dataframe*
 - R (but also other programming languages like python)
 - Export to csv, excel, pdf

1 character: 3,913
2 characters: 34,233
3 characters: 7,143
4 characters: 3,355
Total: 48,644
(>200 variables)

country	year	cases	population
Afghanistan	1999	31745	19987071
Afghanistan	2000	2666	20095360
Brazil	1999	31737	17200362
Brazil	2000	80488	174504898
China	1999	210258	127201272
China	2000	210766	128002583

variables

country	year	cases	population
Afghanistan	1999	31745	19987071
Afghanistan	2000	2666	20095360
Brazil	1999	31737	17200362
Brazil	2000	80488	174504898
China	1999	210258	127201272
China	2000	210766	128002583

observations

country	year	cases	population
Afghanistan	99	75	98071
Afghanistan	00	66	09360
Brazil	99	737	00362
Brazil	00	488	50898
China	99	258	272
China	00	766	02583

values

Wickham (2014);
Wickham & Grolemund (2016);
Forkel et al. (2018)

CHIDEOD (the online app version)

CHIDEOD Table About

Phonology

- Pinyin without tones
- Pinyin with tones
- Pinyin with numbers
- Middle Chinese
- Old Chinese

Orthography

- Traditional Chinese
- Simplified Chinese
- Traditional 1
- Traditional 2
- Traditional 3
- Traditional 4
- Simplified 1
- Simplified 2

Copy CSV Excel Search:

pinyinnone traditional Kroll

1	xu	吁	gasp, onom. of surprise, whew!, or dismay, oh no!; to sigh, ahh
2	fei	吠	onom. Dog's bark
3	zha	咤	onom. of eating noisily and with gusto, smacking the lips; onom. of exclamatory admiration or sighing regret
4	zi	咨	onom. of weariness, discouragement, heartache: ah me!
5	wa	哇	vomit, retch (onom.)
6	xi	唏	onom. for sighing in sadness or grief, syeeh!
7	kui	喟	onom. Sighing, heaving a sigh
8	xiao	嗥	onom. Bamboo flute
9	sou	嗾	onom. of sound by which one calls or commands a dog
10	xu	嘘	onom. of exhaling serenely, in a satisfied and unhurried manner; a sigh of ease and contentment

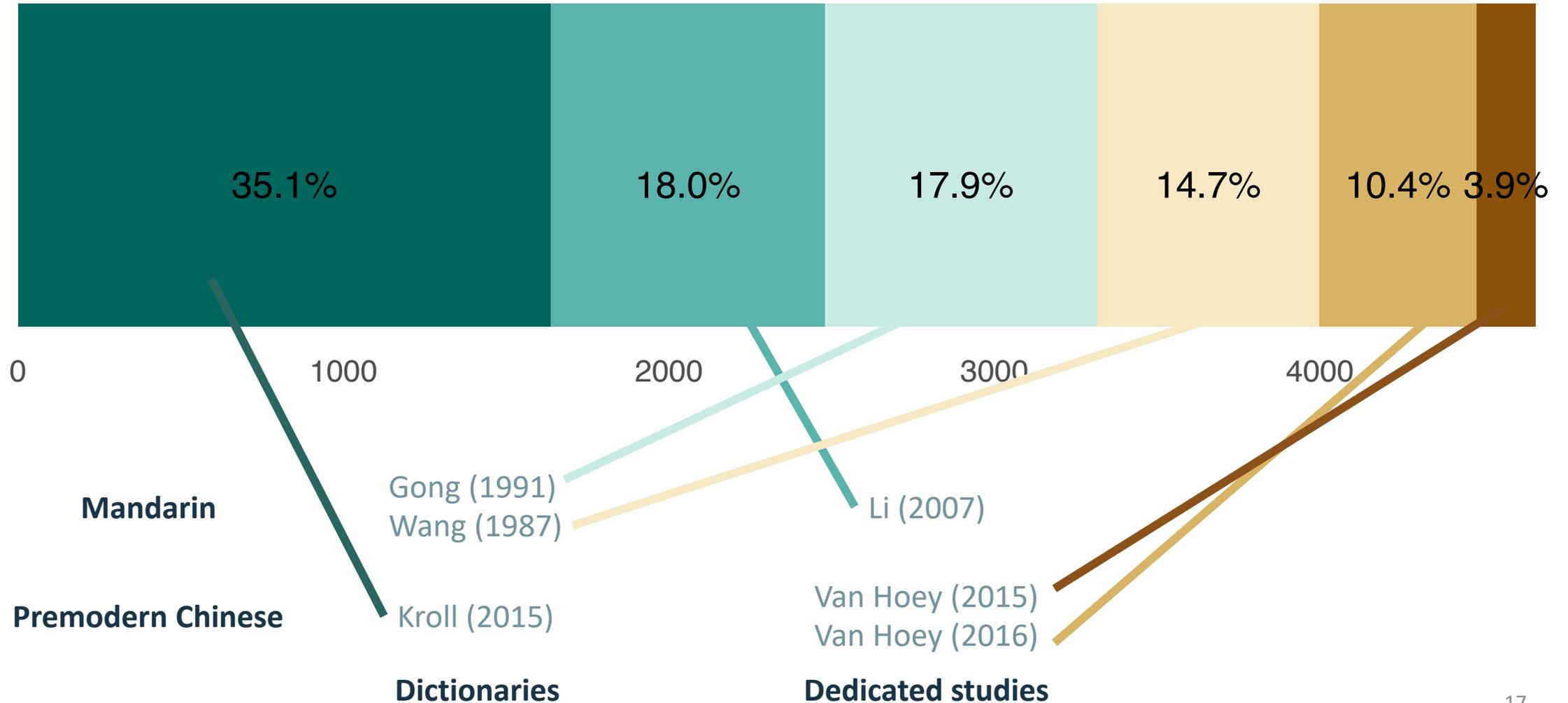
Showing 1 to 10 of 4,662 entries

Previous 1 2 3 4 5 ... 467 Next

4662 entries; 3452 distinct onomatopoeia and ideophones

Variables coded in CHIDEOD

Sources for CHIDEOD



Formal variables

Semantic variables

Other variables

phonology

pinyin tone,
pinyin num,
pinyin none

Middle Chinese (MC)
Old Chinese (OC)

word
level

Kroll dictionary
Handian (zdic)
Hanyu Da Cidian

variants

note

sensory modality

datasource

traditional
simplified

T1-T4

S1-S4
S1-S4.charfreq
S1-S4.famfreq

morphology

character
level

S1-S4.sem
S1-S4.semfreq
S1-S4.semfam

radical support

below-
character
level

S1-S4.phon
S1-S4.phonfreq
S1-S4.phonfam

orthography

variables in CHIDEOD

Abbreviations:

S = simplified,

T = traditional,

sem = semantic radical,

phon = phonetic radical,

freq = (token) frequency,

fam = (type) family frequency

Formal variables

phonology

pinyin tone,
pinyin num,
pinyin none

Middle Chinese (MC)
Old Chinese (OC)

traditional
simplified

T1-T4

S1-S4
S1-S4.charfreq
S1-S4.famfreq

S1-S4.sem
S1-S4.semfreq
S1-S4.semfam

S1-S4.phon
S1-S4.phonfreq
S1-S4.phonfam

morphology

radical support

word
level

character
level

below-
character
level

Semantic variables

Kroll dictionary
Handian (zdic)
Hanyu Da Cidian

sensory modality

Other variables

variants

note

datasource

variables in CHIDEOD

Abbreviations:
S = simplified,
T = traditional,
sem = semantic radical,
phon = phonetic radical,
freq = (token) frequency,
fam = (type) family frequency

orthography

Formal variables

phonology { pinyintone,
pinyinnum,
pinyinnone
Middle Chinese (MC)
Old Chinese (OC)

traditional
simplified

T1-T4

S1-S4
S1-S4.charfreq
S1-S4.famfreq

morphology

S1-S4.sem
S1-S4.semfreq
S1-S4.semfam

radical support

S1-S4.phon
S1-S4.phonfreq
S1-S4.phonfam

**word
level**

**character
level**

**below-
character
level**

Semantic variables

Kroll dictionary
Handian (zdic)
Hanyu Da Cidian
sensory modality

Other variables

variants
note
datasource

variables in CHIDEOD

Abbreviations:
S = simplified,
T = traditional,
sem = semantic radical,
phon = phonetic radical,
freq = (token) frequency,
fam = (type) family frequency

Formal variables

Traditional
character

關關

Simplified
character

关关

onomatopoeia

'cry of an osprey'

Formal variables

word level

Phonology

guān~guān

guan1~guan1

guan~guan

Middle Chinese

kwaen~kwaen

Old Chinese

*[k]ʰro[n]~[k]ʰro[n]

(Baxter & Sagart 2014; 2015)

Traditional
character

關關

Simplified
character

关关

onomatopoeia

'cry of an osprey'

Formal variables

word level

Phonology

guān~guān

guan1~guan1

guan~guan

Middle Chinese

kwaen~kwaen

Old Chinese

*[k]ʰro[n]~[k]ʰro[n]

(Baxter & Sagart 2014; 2015)

character level

T1 T2 T3 T4

關 關 NA NA

S1 S2 S3 S4

关 关 NA NA

Character frequency per million

976,716

Family frequency

90

Traditional character

關 關

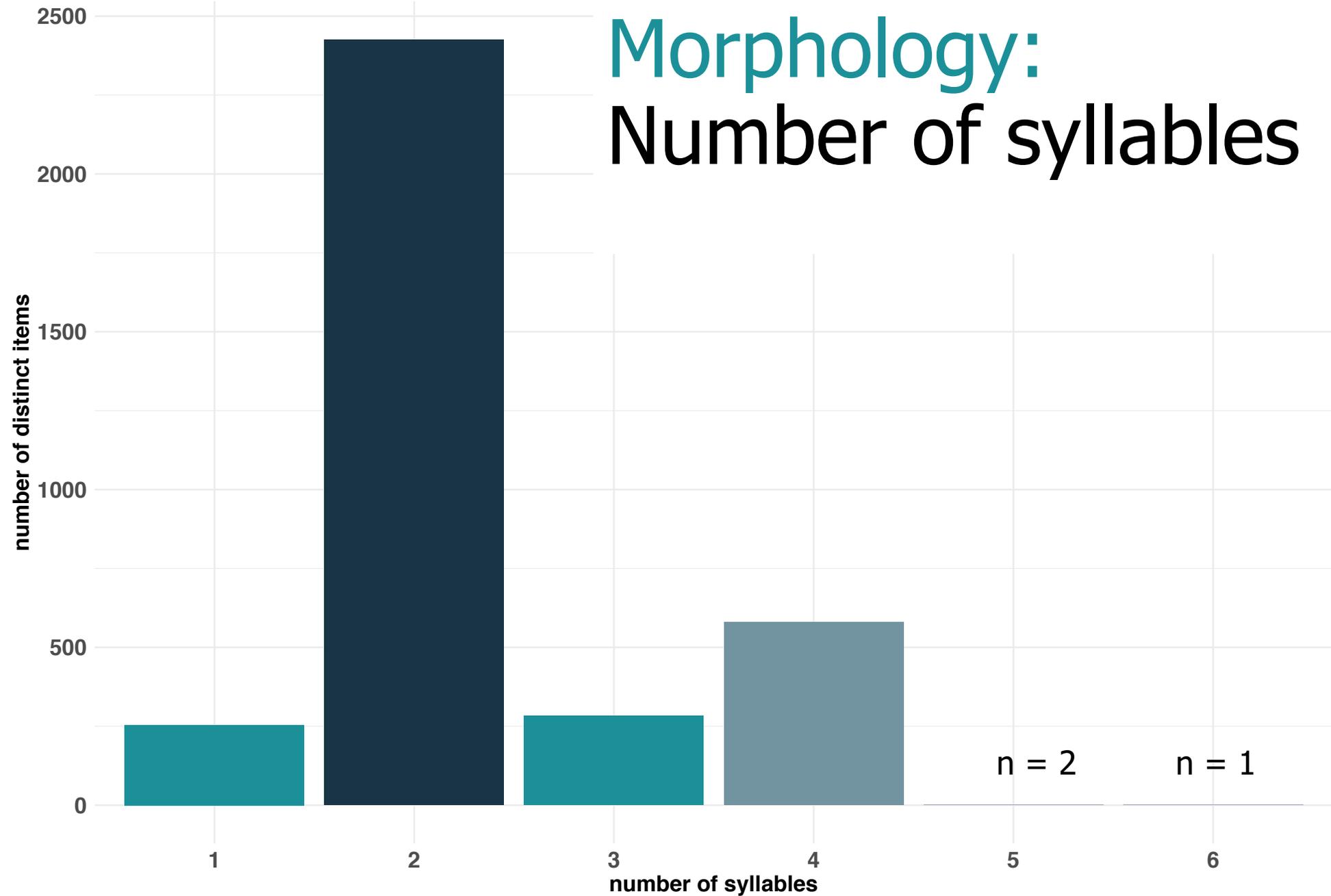
Simplified character

关 关

onomatopoeia

'cry of an osprey'

Morphology: Number of syllables



Based on identifying a

- **BASE**
- **REDUPLICANT**
- **extra elements**

(cf. Sun 1999)

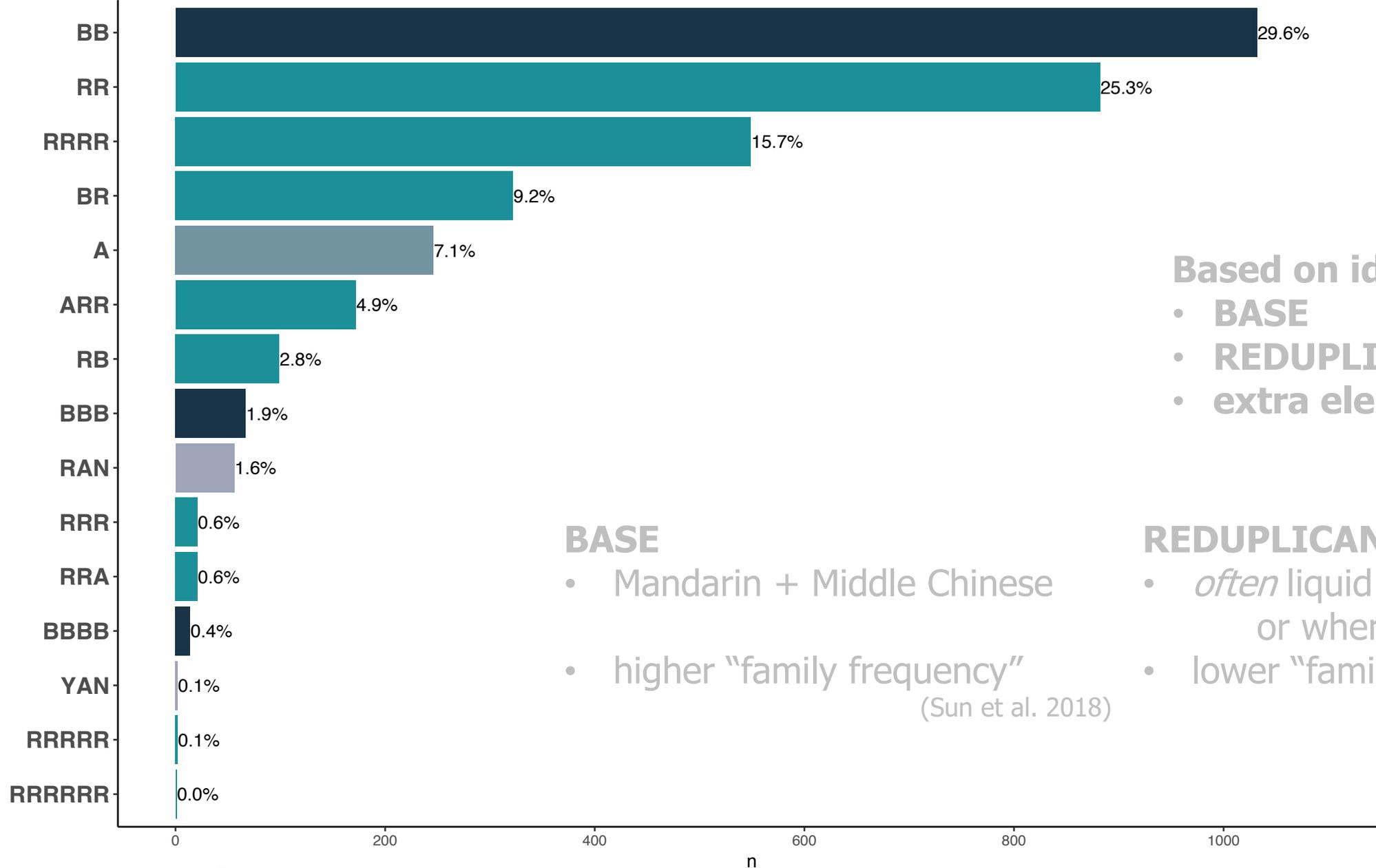
BASE

- Mandarin + Middle Chinese
- higher “family frequency”
(Sun et al. 2018)

REDUPLICANT

- *often* liquid (or MC reflex)
or when base unclear
- lower “family frequency”

Morphological categories and their frequencies



Based on identifying a

- **BASE**
- **REDUPLICANT**
- **extra elements**

(cf. Sun 1999)

BASE

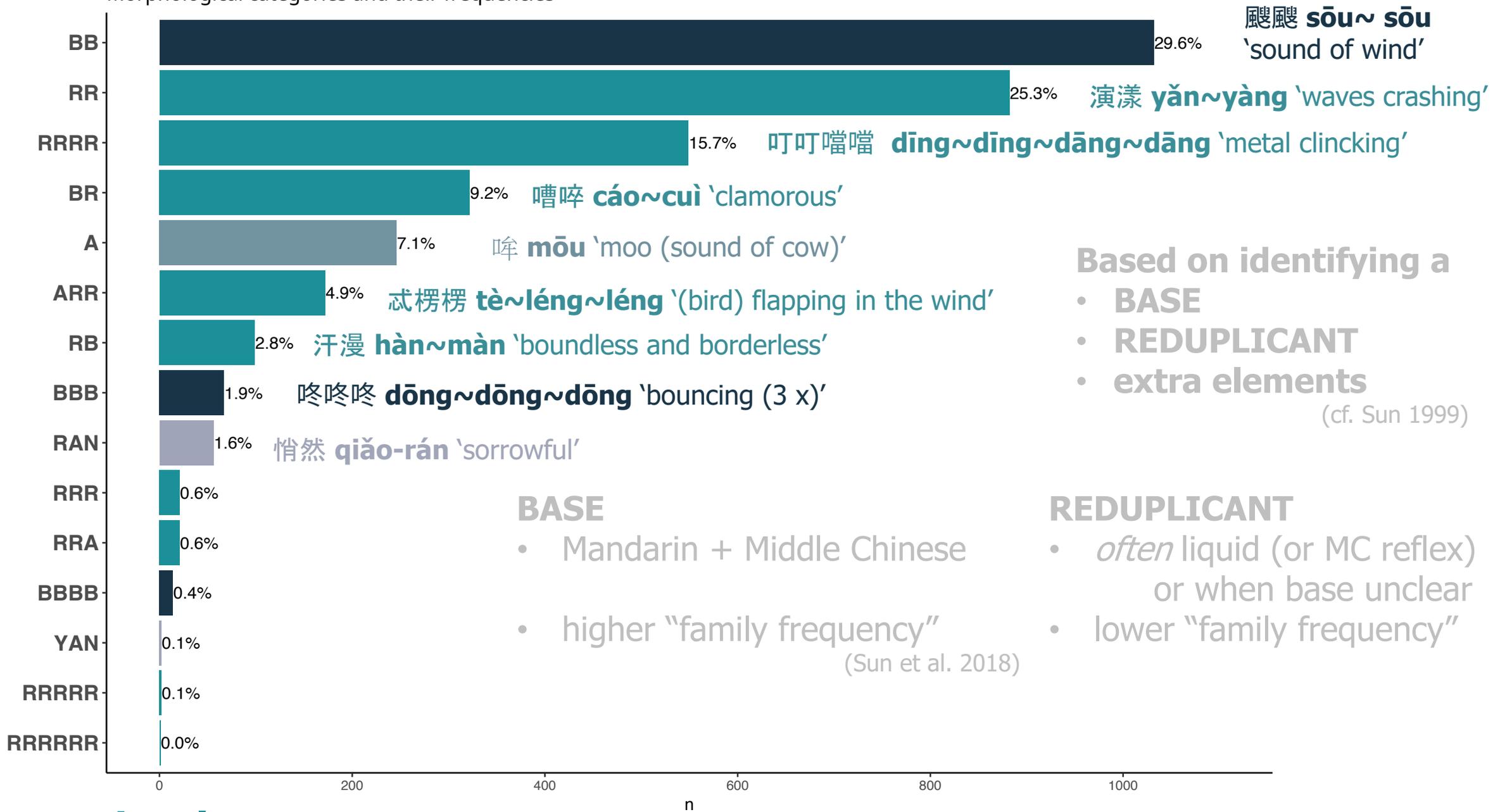
- Mandarin + Middle Chinese
- higher “family frequency”

(Sun et al. 2018)

REDUPLICANT

- *often* liquid (or MC reflex) or when base unclear
- lower “family frequency”

Morphological categories and their frequencies



Based on identifying a

- **BASE**
- **REDUPLICANT**
- **extra elements**

(cf. Sun 1999)

BASE

- Mandarin + Middle Chinese
- higher "family frequency"

(Sun et al. 2018)

REDUPLICANT

- *often* liquid (or MC reflex) or when base unclear
- lower "family frequency"

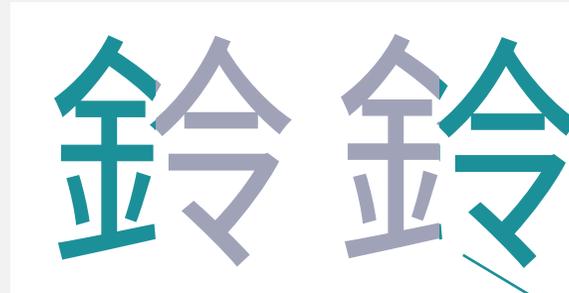
Formal variables: below-character level

About 72 % of Chinese characters are composed of a semantic radical and a phonetic part, based Chinese Lexical Database (Sun et al. 2018)

250 semantic radicals in the CLD

e.g.

- 口 MOUTH 286 characters
- 扌 HAND 255
- 艹 GRASS 244
- 木 WOOD 255
- 亻 PERSON 222



líng~líng
'onomatopoeia:
sound of small clapper bells
on carriages'

1079 phonetic radicals in the CLD

e.g.

- 非 *fei* 19 characters
- 客 *ke* 19
- 隹 *zhui* 18
- 包 *bao* 18
- 且 *ju* 18

In 鈴鈴 the semantic radical 金 (钅) indicates METAL

In 鈴鈴 the phonetic radical is 令 *ling*

The ideophone is motivated by orthography

The ideophone is motivated by phonology (BB type)

Phonological support: Sound correspondences the orthographic forms (of ideophones)

Full reduplication (BB+)

Partial reduplication

- BR
- RB
- RR
- RRA
- ARR
- RR+

Single A

Compositional RAN/YAN

懊惱

ào~nǎo
'vexed'

奧 *ao*

惱 *nao*

42.9% of CHIDEOD (1483 items) have this kind of phonological support.

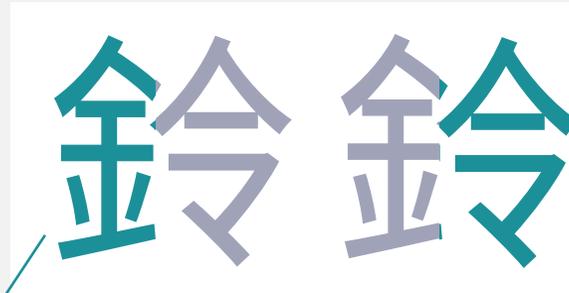
Formal variables: below-character level

About 72 % of Chinese characters are composed of a semantic radical and a phonetic part, based Chinese Lexical Database (Sun et al. 2018)

250 semantic radicals in the CLD

e.g.

- 口 MOUTH 286 characters
- 扌 HAND 255
- 艹 GRASS 244
- 木 WOOD 255
- 亻 PERSON 222



líng~líng
'onomatopoeia:
sound of small clapper bells
on carriages'

1079 phonetic radicals in the CLD

e.g.

- 非 /fei/ 19 characters
- 客 /ke/ 19
- 隹 /zhui/ 18
- 包 /bao/ 18
- 且 /ju/ 18

In 鈴鈴 the semantic radical 金 (钅) indicates METAL

The ideophone is motivated by orthography

In 鈴鈴 the phonetic radical is 令 /ling/

The ideophone is motivated by phonology (BB type)

Formal below: semantic radical support

Radical support:

reduplication in the orthographic form
(of ideophones)

Different ontological domains

- speaking
- nature
- human
- ...

Whether this differs significantly from prosaic words is still undetermined.

Most interesting in **partially reduplicated** forms.

radical support		morphological pattern frequencies				
radical	meaning	A	BR	RB	RR	RRRR
口	MOUTH	129	8		196	266
氵	WATER		50	7	51	7
山	MOUNTAIN		21		21	
艹	GRASS		11		21	
足	FOOT				18	
王	JADE				15	5
卜	HEART1		8	6	14	
糸	SILK				12	
扌	HAND		5		10	
心	HEART2				9	
金	METAL				9	6
女	WOMAN		5		8	
辵	WALK				8	
NA		98	154	45	374	254

What is more important in partial?

Subset: `datasource = Kroll + binomes (two characters) [morphology = RR|RB|BR]`

With these two parameters

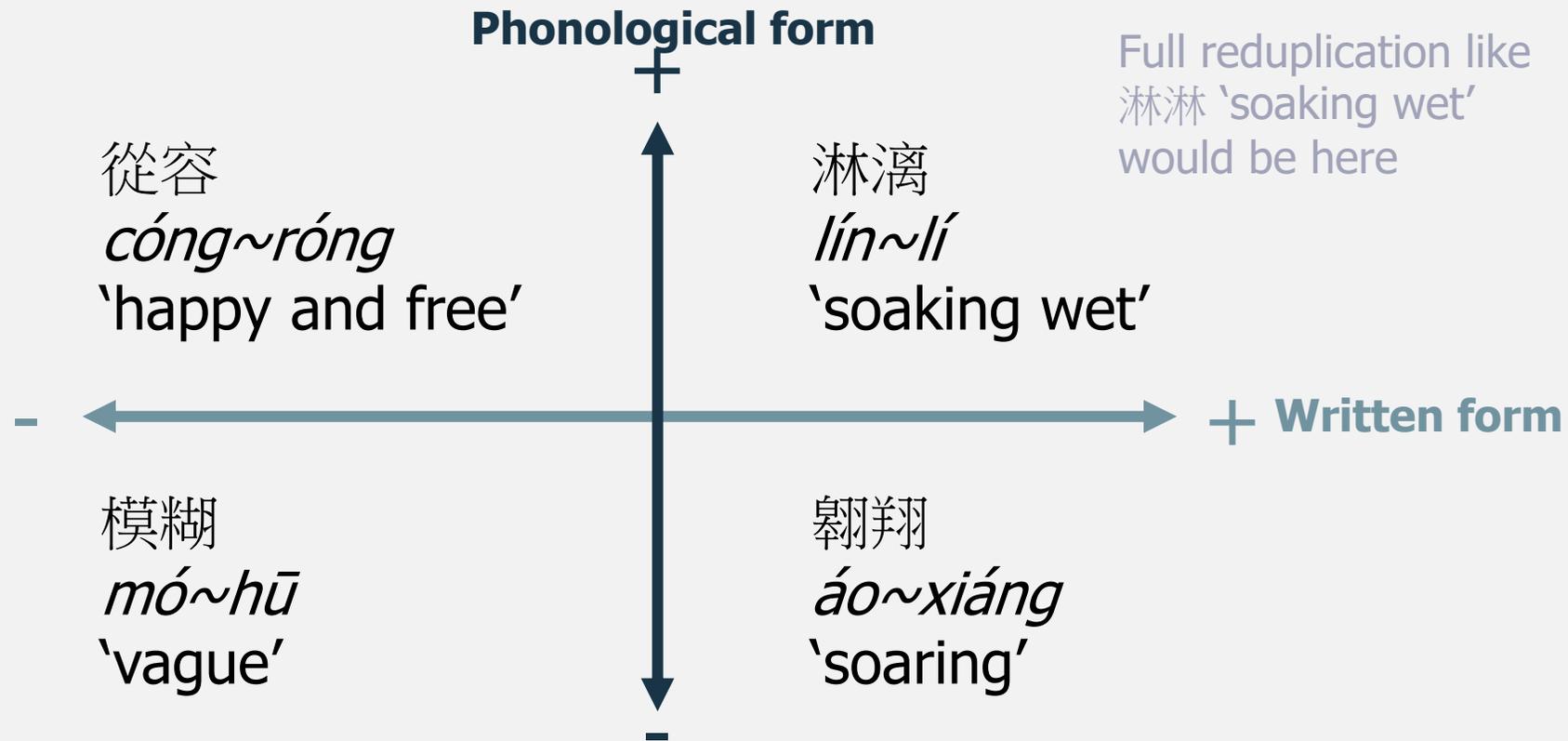
- **Phonological form** — Morphophonological motivation / markedness
- **Written form** — Ortographic motivation / markedness

What is more important in partial?

Subset: datasource = Kroll + binomes (two characters) [morphology = RR | RB | BR]

With these two parameters

- **Phonological form** — Morphophonological motivation / markedness
- **Written form** — Orthographic motivation / markedness

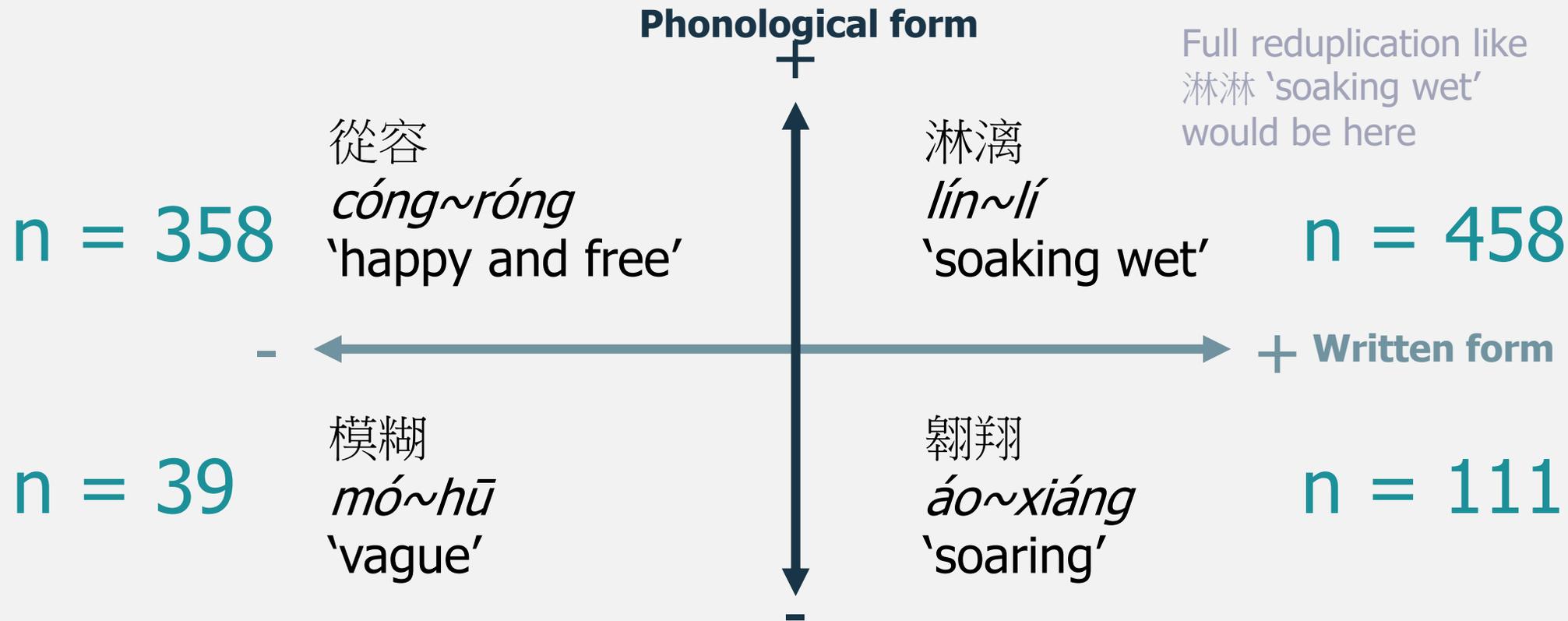


What is more important in partial?

Subset: `datasource = Kroll + binomes (two characters) [morphology = RR|RB|BR]`

With these two parameters

- **Phonological form** — Morphophonological motivation / markedness
- **Written form** — Orthographic motivation / markedness



$\chi^2 = 15.99$
 $p = 6.369e-05$

Formal variables

phonology {
pinyin tone,
pinyin num,
pinyin none

Middle Chinese (MC)
Old Chinese (OC)

traditional
simplified

T1-T4

S1-S4
S1-S4.charfreq
S1-S4.famfreq

S1-S4.sem
S1-S4.semfreq
S1-S4.semfam

S1-S4.phon
S1-S4.phonfreq
S1-S4.phonfam

morphology

radical support

Semantic variables

Kroll dictionary
Handian (zdic)
Hanyu Da Cidian

sensory modality

Other variables

variants

note

datasource

word
level

character
level

below-
character
level

variables in CHIDEOD

Abbreviations:

S = simplified,

T = traditional,

sem = semantic radical,

phon = phonetic radical,

freq = (token) frequency,

fam = (type) family frequency

Semantic variables:

3 dictionaries define most ideophones

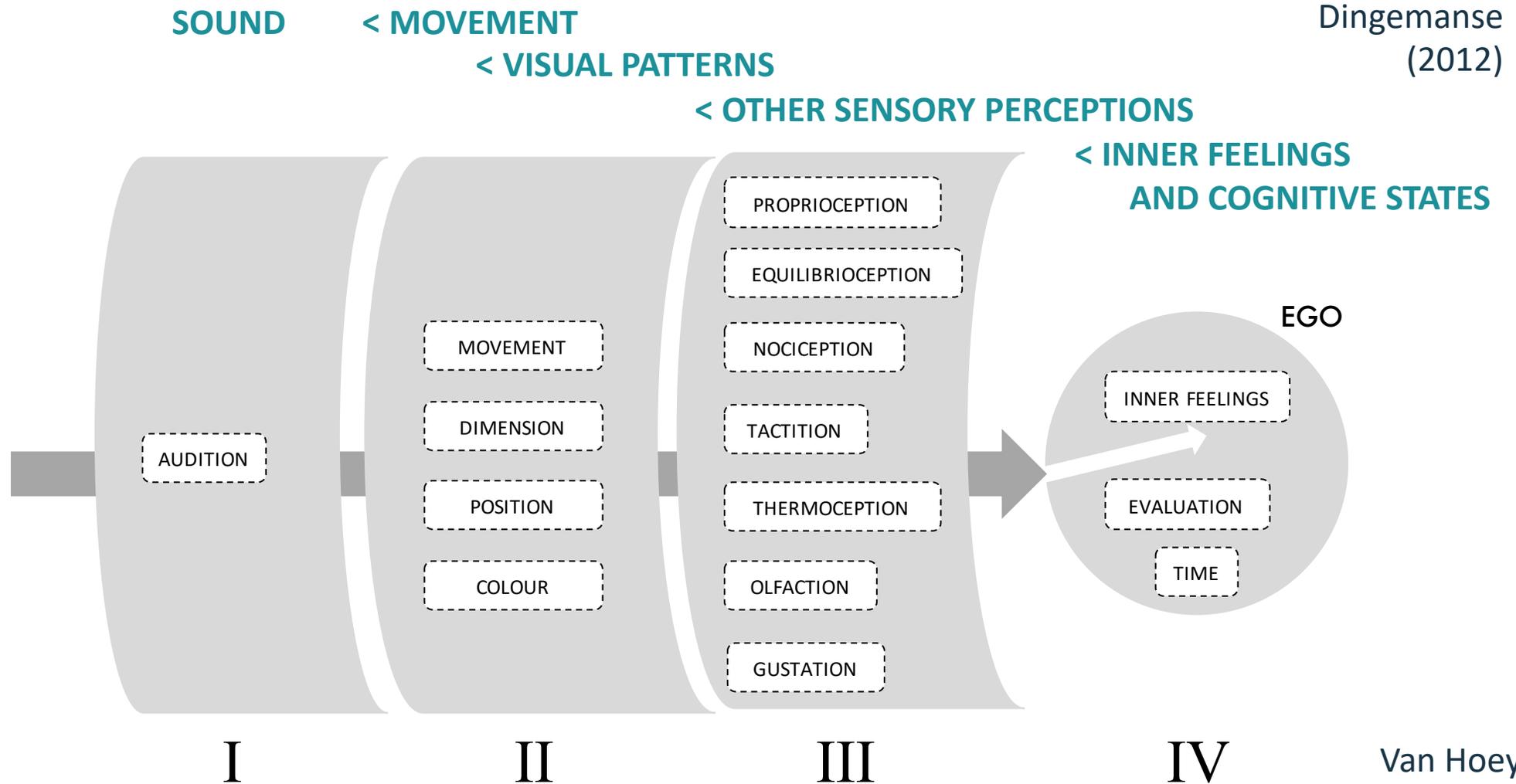
comprehensive dictionary	Classical Chinese	Mandarin dictionary	n
Hanyu Da Cidian	Kroll	zdic (Handian)	
✓	✓	✓	971
✓	×	✓	710
✓	✓	×	423
✓	×	×	464
×	✓	×	178
×	×	✓	46
×	✓	✓	15
×	×	×	987

2807

This is why in-depth studies are important.

- BBB BBBB RRR RRRR types
- obscure forms

Semantic variables: Sensory domains in-depth studies



Dingemans
(2012)

Van Hoey
(2016, in prep.)

Formal variables

Semantic variables

Other variables

phonology

pinyin tone,
pinyin num,
pinyin none

Middle Chinese (MC)
Old Chinese (OC)

word
level

Kroll dictionary
Handian (zdic)
Hanyu Da Cidian

variants

note

sensory modality

datasource

traditional
simplified

T1-T4

S1-S4
S1-S4.charfreq
S1-S4.famfreq

morphology

character
level

S1-S4.sem
S1-S4.semfreq
S1-S4.semfam

radical support

below-
character
level

S1-S4.phon
S1-S4.phonfreq
S1-S4.phonfam

orthography

variables in CHIDEOD

Abbreviations:

S = simplified,

T = traditional,

sem = semantic radical,

phon = phonetic radical,

freq = (token) frequency,

fam = (type) family frequency

Other variables: variant written forms

xiāng~yáng

'wandering and wavering'

form radical support

- 僂佯 PERSON 亻
- 徬徉 WALKING 彳
- 相羊 NA

Future exploration: differences in the conceptualization between these different orthographic forms with different radical support.

cf.

máng~máng

'stretching farther than they eye can see'

- 芒芒 (GRASS 艹)
- 茫茫 (GRASS 艹 + WATER 氵)

Over time the grass+water variant became more popular +
took over the non-water variant when used in relation to bodies of water (Van Hoey 2019)

A short application of CHIDEOD

Vowel alternation in partially reduplicated syllables

Vowel alternation (ding~dang)

Group 1:

- SOUND ideophones in Kroll's (2015) dictionary of Classical and Medieval Chinese
- Partial reduplication types

Group 2:

- SOUND ideophones in 3 onomatopoeia data sources of Mandarin Chinese (Wang 1987, Gong 1991; Li 2007)
- Partial reduplication types

Vowel alternation (ding~dang)

Group 1:

- SOUND ideophones in Kroll's (2015) dictionary of Classical and Medieval Chinese

```
[datasource == Kroll]  
[sensory modality == SOUND]
```

- Partial reduplication types

```
[morphology == BR, RB, RR, RR+]
```

Group 2:

- SOUND ideophones in 3 onomatopoeia data sources of Mandarin Chinese (Wang 1987, Gong 1991; Li 2007)

```
[datasource == Wang|Gong|Li]  
[sensory modality == SOUND]
```

- Partial reduplication types

```
[morphology == BR, RB, RR, RR+]
```

Vowel alternation (ding~dang)

Group 1:

- SOUND ideophones in Kroll's (2015) dictionary of Classical and Medieval Chinese

```
[datasource == Kroll]  
[sensory modality == SOUND]
```

- Partial reduplication types

```
[morphology == BR, RB, RR, RR+]
```

n = 167

23/167 with vowel alternation (13.77%)

Group 2:

- SOUND ideophones in 3 onomatopoeia data sources of Mandarin Chinese (Wang 1987, Gong 1991; Li 2007)

```
[datasource == Wang|Gong|Li]  
[sensory modality == SOUND]
```

- Partial reduplication types

```
[morphology == BR, RB, RR, RR+]
```

n = 1421

983/ 1421 with vowel alternation (69.2%%)

Vowel alternation (ding~dang)

Group 1:

- SOUND ideophones in Kroll's (2015) dictionary of Classical and Medieval Chinese

```
[datasource == Kroll]  
[sensory modality == SOUND]
```

- Partial reduplication types

```
[morphology == BR, RB, RR, RR+]
```

n = 167

23/167 with vowel alternation (13.77%)

Group 2:

- SOUND ideophones in 3 onomatopoeia data sources of Mandarin Chinese (Wang 1987, Gong 1991; Li 2007)

```
[datasource == Wang|Gong|Li]  
[sensory modality == SOUND]
```

- Partial reduplication types

```
[morphology == BR, RB, RR, RR+]
```

n = 1421

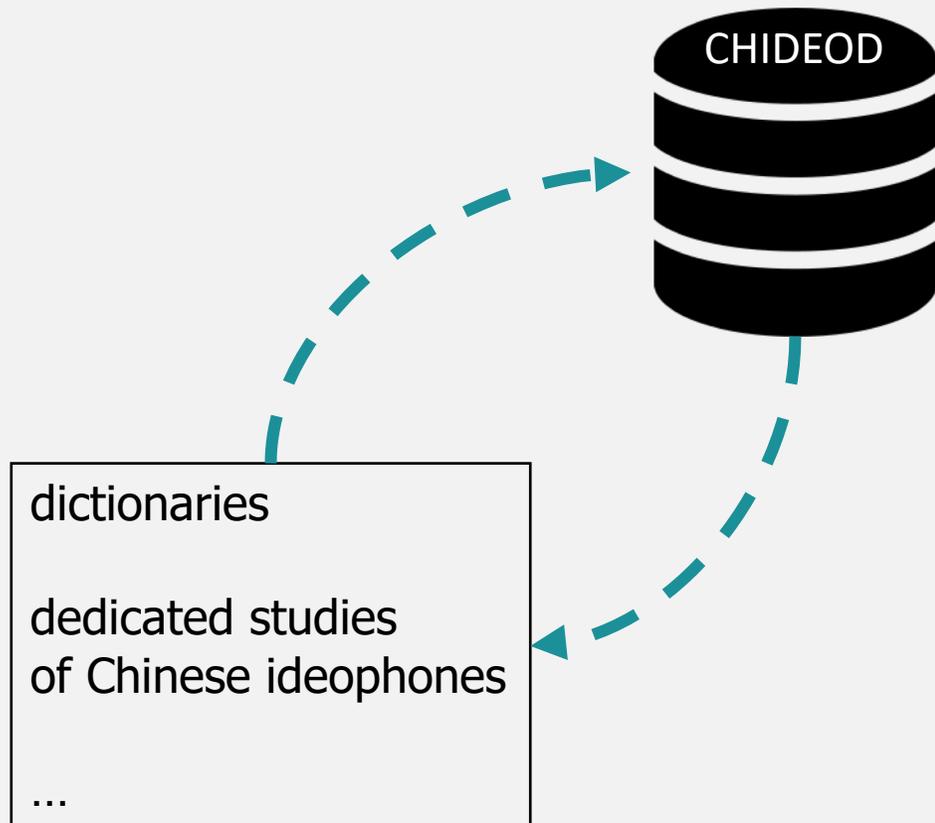
983/ 1421 with vowel alternation (69.2%%)

$\chi^2 = 195.1928,$
 $p < 0.001,$
 $\phi = 0.35$

**Either this is a recent development (not very likely)
or it has been underdocumented (more likely).**

Summary

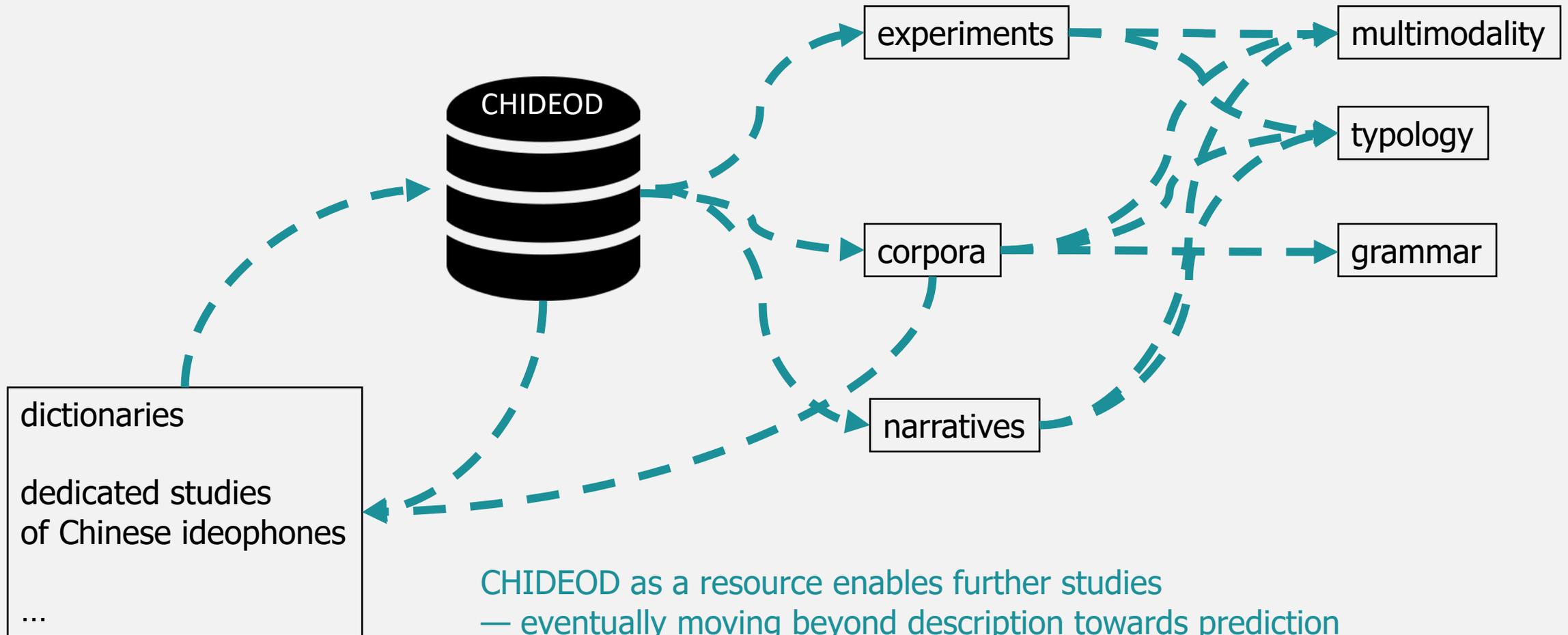
CHIDEOD: future applications



In future versions of the database:

- ❖ more sources
- ❖ other variables:
 - ❖ other Sinitic languages such as Cantonese, Taiwanese
 - ❖ token frequencies based on corpora
- ❖ ...

CHIDEOD: future applications



References

- Akita, Kimi. 2016. A multimedia encyclopedia of Japanese mimetics: A frame-semantic approach to L2 sound-symbolic words. In Kaori Kabata & Kiyoko Toratani (eds.), *Cognitive-functional approaches to the study of Japanese as a second language* (Studies on Language Acquisition volume 46), 139–167. Boston: De Gruyter Mouton.
- Baxter, William Hubbard & Laurent Sagart. 2014. *Old Chinese: a new reconstruction*. Oxford ; New York: Oxford University Press.
- Baxter, William Hubbard & Laurent Sagart. 2015. Baxter-Sagart Old Chinese reconstruction (v. 13 october 2015).
<http://ocbaxtersagart.lsa.umich.edu/BaxterSagartOC2015-10-13.xlsx>.
- Bodomo, Adams. 2006. The structure of ideophones in African and Asian languages: The case of Dagaare and Cantonese. In John Mugane, John P. Hutchison & Dee A. Worman (eds.), *Selected Proceedings of the 35th Annual Conference on African Linguistics*, 203–213. Somerville, MA: Cascadilla Proceedings Project.
- Dingemans, Mark. 2011. *The meaning and use of ideophones in Siwu*. Nijmegen: Radboud University Nijmegen dissertation.
- Dingemans, Mark. 2012. Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass* 6(10). 654–672.
- Dingemans, Mark. 2019. “Ideophone” as a comparative concept. In Kimi Akita & Prashant Pardeshi (eds.), *Ideophones, mimetics and expressives* (Iconicity in Language and Literature, ILL 16), 13–33. Amsterdam : Philadelphia: John Benjamins Pub. Co.
- Flaticon: <https://www.flaticon.com/authors/smashicons>.
- Forkel, Robert, Johann-Mattis List, Simon J. Greenhill, Christoph Rzymiski, Sebastian Bank, Michael Cysouw, Harald Hammarström, Martin Haspelmath, Gereon A. Kaiping & Russell D. Gray. 2018. Cross-Linguistic Data Formats, advancing data sharing and re-use in comparative linguistics. *Scientific Data* 5. 180205. doi:[10.1038/sdata.2018.205](https://doi.org/10.1038/sdata.2018.205).
- Gong, Liangyu 龚良玉 (ed.). 1991. *Xiangshengci cidian 象声词词典 [Dictionary of onomatopoeia]*. Guiyang: Guizhou jiaoyu chubanshe.
- Kroll, Paul W. 2015. *A student's dictionary of Classical and Medieval Chinese* (Handbook of Oriental Studies: Section 4 China 30). Leiden: Brill.
- Li, Jing'er 李镜儿. 2007. *Xiandai Hanyu nishengci yanjiu 现代汉语拟声词研究 (Onomatopoeics in Modern Chinese)*. Shanghai: Xuelin chubanshe.
- Lu Chiarung 吕佳蓉. 2006. *Giongo, gitaigo no hiyuteki kakuchō no shosō: ninchi gengogaku to ruikeiron no kanten kara 擬音語・擬態語の比喩的拡張の諸相——認知言語学と類型論の観点から [Figurative extensions of mimetics: A Cognitive Linguistic and typological study]*. Kyōto: Kyōto University PhD dissertation.
- Meng, Chenxi. 2012. *A description of ideophonic words in Mandarin Chinese*. Leiden: Leiden University Research Master in Linguistics.
- Mok, Waiching Enid. 2001. *Chinese sound symbolism: A phonological perspective*. Hawai'i: University of Hawai'i PhD dissertation.

References

- NINJAL. 2016. Lago Word Profiler for the Balanced Corpus of Contemporary Written Japanese (NINJAL-LWP for BCCWJ). Database. <http://nlb.ninjal.ac.jp/search/> (25 March, 2019).
- Nuckolls, Janis B. & Tod D. Swanson. 2019. Quechua Real Words: An audiovisual ANTI-dictionary of expressive Quechua ideophones. <http://quechuarealwords-dev.byu.edu/index.php> (12 March, 2019).
- Nuckolls, Janis B., Tod D. Swanson, Diana Shelton, Alexander Rice & Sarah Hatton. 2017. Lexicography in-your-face: The active semantics of Pastaza Quichua ideophones. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 62(02). 154–172. doi:[10.1017/cnj.2017.9](https://doi.org/10.1017/cnj.2017.9).
- Qiu, Di 邱迪. 2018. Xiandai Hanyu zhong de “gezhong + nishengci” jiegou yanjiu 现代汉语中的“各种+拟声词”结构研究 [The construction “ALL KINDS OF + onomatopoeia” in Modern Chinese]. *Liaoning jiaoyu xingzheng xueyuan xuebao* 辽宁教育行政学院学报 4. 69–72.
- Sun, Ching Chu, Peter Hendrix, Jianqiang Ma & Rolf Harald Baayen. 2018. Chinese lexical database (CLD): A large-scale lexical database for simplified Mandarin Chinese. *Behavior Research Methods* 50(6). 2606–2629. doi:[10.3758/s13428-018-1038-3](https://doi.org/10.3758/s13428-018-1038-3).
- Sun, Jingtao. 1999. *Reduplication in Old Chinese*. Vancouver: University of British Columbia PhD dissertation.
- Thompson, Arthur Lewis. 2018. Are tones in the expressive lexicon iconic? Evidence from three Chinese languages. *PLOS ONE* 13(12). e0204270. doi:[10.1371/journal.pone.0204270](https://doi.org/10.1371/journal.pone.0204270).
- Van Hoey, Thomas. 2015. *Ideophones in Middle Chinese: A typological study of a Tang dynasty poetic corpus*. Leuven: KU Leuven Master thesis.
- Van Hoey, Thomas. 2016. Ideophones in Old Chinese: the case of the Shijing 詩經. *ISACG 9 [International Symposium on Ancient Chinese Grammar]*. Berlin: Humboldt University.
- Van Hoey, Thomas & Chiarung Lu. 2019. Lexical variation of ideophones in Chinese classics: Their implications in embodiment and migration. In Janice Fon (ed.), *Dimensions of diffusion and diversity* (Cognitive Linguistics Research 63), 195–226. Berlin; Boston: De Gruyter Mouton.
- Wang, Warren 王万仁 (ed.). 1987. *Xiang sheng ci li shi* 象声词例释 [Examples and explanations of onomatopoeia]. Nanning: Guangxi jiaoyu chubanshe: Guangxi xinhua shudian faxing.
- Wickham, Hadley. 2014. Tidy data. *Journal of Statistical Software* 59(10). 1–23. doi:[10.18637/jss.v059.i10](https://doi.org/10.18637/jss.v059.i10).
- Wickham, Hadley & Garrett Golemund. 2016. *R for data science: import, tidy, transform, visualize, and model data*. First edition. Sebastopol, CA: O’Reilly.

References

- Wu, Mengqi. 2014. *The structure of ideophones in Southern Sinitic*. Hong Kong: University of Hong Kong Master thesis.
- You, Weilun 游韋倫. 2015. *Nichū ryō gengo ni okeru giongo no imi to imi kakuchō: Furēmu imi-ron no kanten kara no apurōchi 日中両言語における擬音語の意味と意味拡張 —フレーム意味論の観点からのアプローチ— [The semantics and semantic extensions of onomatopoeia in Japanese and Chinese: A Frame Semantics approach]*. Kobe: Kobe University PhD dissertation.
- Zhao, Aiwu 赵爱武. 2008. Jin 20 nian Hanyu xiangshengci yanjiu zongshu 近20年汉语象声词研究综述 (Chinese onomatopoeia study in recent 20 years: summary). *Wuhan daxue xuebao 武汉大学学报 (renwen kexue ban 人文科学版)* 61(2). 180–185.

Demonstration

- <https://simazhi.shinyapps.io/Chineseideophone/>
- [**https://osf.io/kpwgf/**](https://osf.io/kpwgf/)

Acknowledgements

National Taiwan University

The University of Hong Kong

Chiarung Lu 呂佳蓉

Youngah Doh

Funding:

MOST-108-2922-I-002-198