Levels of structure within Chinese character constituents

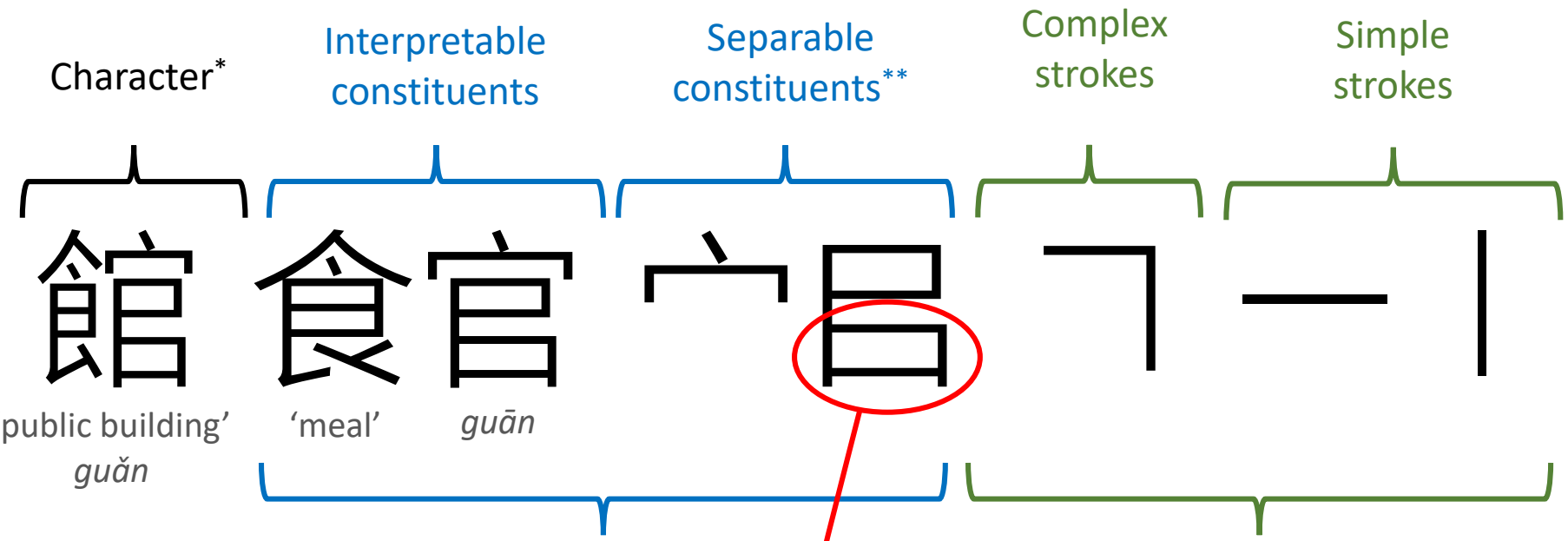
James Myers
National Chung Cheng University, Taiwan
http://personal.ccu.edu.tw/~lngmyers/

Abstract

- Character constituents are like morphemes, strokes are like segments
- In between are strokes groups, which act like syllables:
  - Target of stress-like prominence
  - Onset-nucleus-coda-like internal structure
  - Compete for space in accordance with Menzerath's law

Thanks to Taiwan's Ministry of Science and Technology (MOST 103-2410-H-194-119-MY3), my lab assistants, and anonymous reviewers. I absolve them of all responsibility for errors.

Levels of Chinese character structure



Constituents are psychologically real (Lee 2017; Li & Zhou 2007; Prün 1994), even if uninterpreted (Chen & Cherng 2013), like **morphemes** (Myers 2019)

Strokes are also psychologically real (Bohn 1998; Sze et al. 2014; Wang et al. 2020) and have distinctive features (Peng 2017; Wang 1983), like **segments** (Myers 2019)

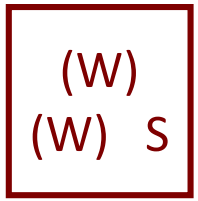
But even when not separated, the stroke group may behave as a distinct level of representation

* Examples here are traditional, but simplified system works virtually the same (Myers 2019)

** History no guide to modern *system*: formally related 月朋服青 formerly distinct 𠂉 青

Stroke groups as “prosodic” units

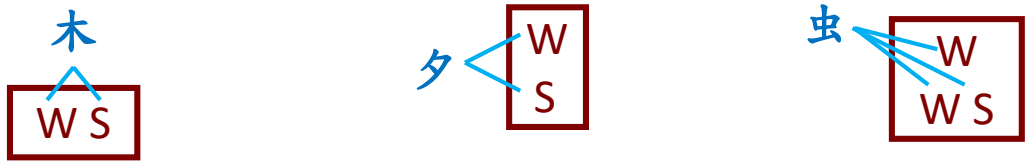
- Character prosody (Myers 2019)
  - Template for position-based patterns



Foot-like prosodic template, with Weak vs. Strong slots

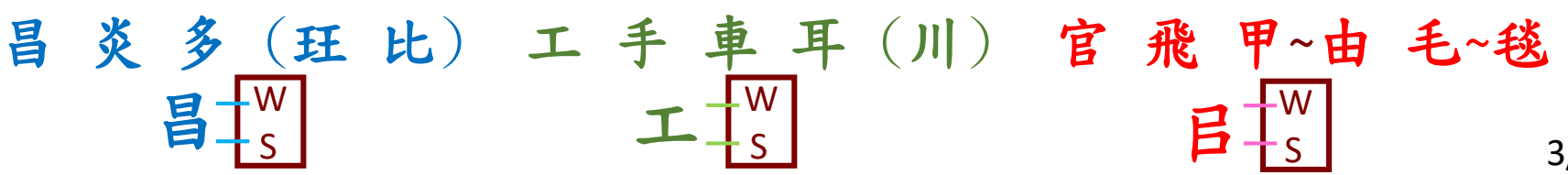
- Reduplication of constituents

林 *lín* ‘forest’ 多 *duō* ‘many’ 蟲 *chóng* ‘insects’



- Curving of strokes
  - Leftmost position, especially in tall, narrow constituents (Wang 1983)

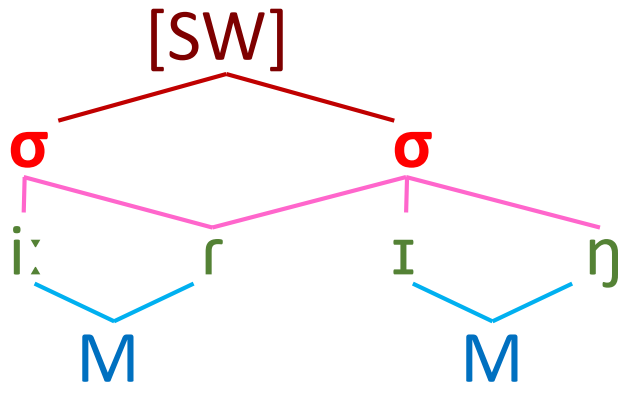
- Prominence (“stress”)
  - Bottommost (and rightmost) constituent, stroke, and ...???



What receives prominence?

- Like syllables, stroke groups form a representation parallel to constituents (morphemes) and strokes (segments)

‘eating’ (N. American English)



(feet)

(syllables)

(segments)

(morphemes)

'official' *guān*	constituents	prosodic templates	**stroke groups**	strokes
官	官	W S, W S, W S	官	官

Internal structure of stroke groups

- Stroke combinations favor certain interactions over others (as illustrated in two- & three-stroke characters; cf. Myers 2019)

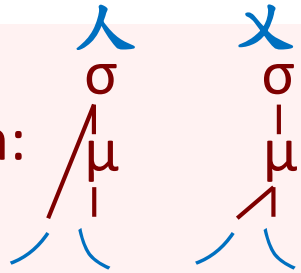
No contact	二三八小川么儿刁于亻乞与凡寸叉弋勺亡刃
Cross	十乂七力九丸又叉乜也廾寸弋子孑孓千干于才大尢丈女巾屮毛土士*
Chain	了子孑孓丫
Start at contact (ㄒ, ㄧ)	丁丌下于亻千干于才大尢卜人久入刀刃勺万乃厂几凡匚亡毛工上又叉口口尸巳己弓夕巾乞匕与
End at contact	上土士工凵山屮口口巳巳己尸么夕弓丫

- Contact at stroke start (its top/left point) is also seen when children copy simple line drawings (Ninio & Lieblich 1976)
- This is similar to coordination of gestures at syllable onsets (Browman & Goldstein 1988), as well as to favoring of onsets and disfavoring of codas (Prince & Smolensky 2004)

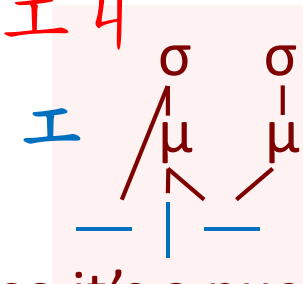
* Exceptional topmost prominence (see Myers 2019)

Structure, prominence, and curving

- Start on contact (least marked) = **O**nset-**N**ucleus: 丁卜人
  - Also complex strokes: **ON:** 刁乚 And chains: **ON+ON:** 了
- No contact (most marked) = **N (+ N + …):** 一二三八小川
- Cross (unmarked) = **NN:** 十乂
  - Unlike start contact, crossed strokes share location:



- End contact (bounded) = … **NC**oda
  - Box bottom stroke is not prominent: 口 **= ONNC** (一冂一)
- End contact (unbounded) = ambisyllabic **C+N:** 工刂



  - Prominence shows contactee is also a nucleus:
- Curving = ambisyllabic **N+O:** 冂冂
  - Width effect on curving shows 丨 in templatic slot, so it's a nucleus
- Each stroke interaction forms a separate stroke group:

冗 = ON_{curv}+ON 并 = NN_{curv}+NN_{prom} 目 = ONNCC (一冂一一)

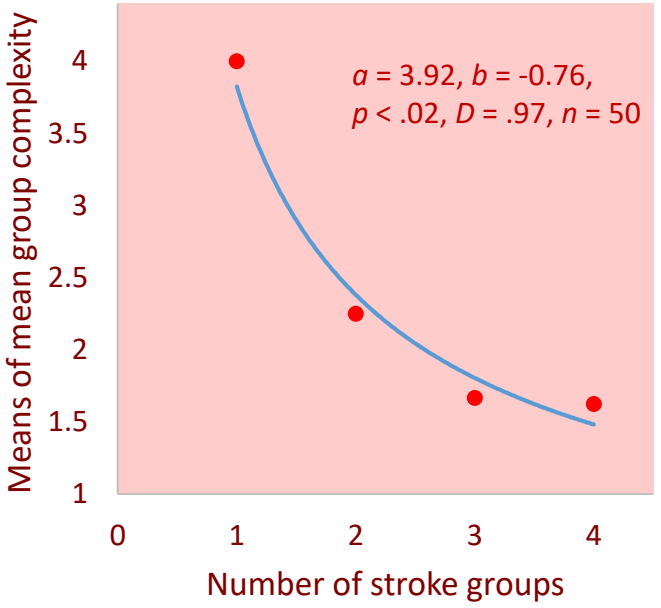
Competing for space

- The more Xs, the simpler their mean complexity Y
  - Menzerath-Altmann law: $y = ax^b$, $b < 0$ (Altmann 1980)
  - Applies to strokes in constituents (Bohn 1998)
  - Applies to constituents in characters (Prün 1994)
  - Suggests that strokes and constituents are genuine levels
- Stroke groups seem to be genuine for the same reason

Some three-stroke characters

	Stroke groups	Structure	Mean group complexity
口	1	ONNC	4
山	2	ONC+NC	2.5
巾	3	N+ON+NN	1.67
丸	4	N+NN+ON+N	1.5

All three-stroke characters



Open questions

- Can all constituents be analyzed consistently?
  - Same or different structures? 人 vs. 入　吕 vs. 吕
  - Scaling up? 龜 = ???
- Is any of this psychologically real?
  - Reduplication, prominence and curving are (Myers 2019)
  - For stroke groups, experimental evidence is still limited
- How far should the syllable analogy be taken?
  - Sign languages also seem to have syllables (Sandler 2008)
  - Or is sign structure more like that of segments (Channon 2002)?
- What about other writing systems?
  - Alphabetic writing also has syllables (Fuhrhop et al. 2011) and stress feet (Evertz 2018), but they directly interact with speech
- What do *you* think?

References

Altmann 1980. Prolegomena to Menzerath's Law. In Grotjahn (Ed.) *Glottometrika* 2. Bochum.

Bohn 1998. *Quantitative Untersuchungen der modernen chinesischen Sprache und Schrift*. Verlag Dr. Kovač.

Browman & Goldstein 1988. Some notes on syllable structure in articulatory phonology. *Phonetica* 45.

Channon 2002. *Signs are single segments*. University of Maryland, College Park.

Chen & Cherng 2013. The proximate unit in Chinese handwritten character production. *Frontiers in Psychology* 4.

Evertz 2018. *Visual prosody*. Walter de Gruyter.

Fuhrhop et al. 2011. The length hierarchy and the graphematic syllable. *Written Language & Literacy* 14.

Lee 2017. Sublexical processes for reading Chinese characters, neurolinguistic studies. In Sybesma et al. (Eds.), *Encyclopedia of Chinese language and linguistics*. Brill.

Li & Zhou 2007. Chinese character structure analysis based on complex networks. *Physica A* 380.

Myers 2019. *The grammar of Chinese characters*. Routledge.

Ninio & Lieblich 1976. The grammar of action. *Child Development* 47.

Peng 2017. Stroke systems in Chinese characters. *Semiotica* 218.

Prince & Smolensky 2004. *Optimality Theory*. Blackwell.

Prün 1994. Validity of Menzerath-Altmann's Law. *Journal of Quantitative Linguistics* 1.

Sandler 2008. The syllable in sign language. In Davis & Zajdó (Eds.) *The syllable in speech production*. Lawrence Erlbaum.

Sze et al. 2014. The Chinese Lexicon Project. *Behavior Research Methods* 46.

Wang 1983. *Toward a generative grammar of Chinese character structure and stroke order*. U. Wisconsin Ph.D. thesis.

Wang et al. 2020. Chinese character handwriting: A large-scale behavioral study and a database. *Behavior Research Methods* 52.