

NORBERT KORDEK

Złożoność pisma na przykładzie znaków chińskich

Abstract (Complexity of Script on the Example of Chinese Characters). This paper discusses two methods of measuring the graphical complexity of script, focusing on the Chinese characters. Main purpose is to establish the validity of the compositional (Altmann 2004) and intersectional (Peust 2006) methods for the analysis of Chinese script. The interpretation of the graphical elements of Chinese characters in the compositional method and the modification of the rules of calculation in the intersectional method will be proposed. The comparative analysis of the values of complexity obtained for the sample set of characters by the two methods with the number of strokes and basic strokes will also be introduced.

Abstrakt. Tematem artykułu jest krytyczne przedstawienie dwóch metod obliczania złożoności graficznej znaków pisma, ze szczególnym uwzględnieniem znaków chińskich. Głównym celem jest wstępna weryfikacja adekwatności metody kompozycyjnej Altmanna (Altmann 2004) i intersekcyjnej Peusta (Peust 2006) do analizy pisma chińskiego. Zaproponowana zostanie interpretacja elementów graficznych znaków chińskich w metodzie kompozycyjnej oraz modyfikacja reguł obliczania złożoności intersekcyjnej. Przedstawiona również będzie analiza porównawcza wyników uzyskanych za pomocą wspomnianych metod na próbie znaków z liczbą kresek i kresek podstawowych w znakach.

Językoznawstwo kwantytatywne zyskuje w ostatnich latach na znaczeniu, oferując unikatową perspektywę wielu zjawisk językowych, nawet tych lepiej poznanych. Jedną z propozycji badawczych, które ciągle czekają na rozwinięcie i szersze wyniki badań, jest przedstawiona przez G. Altmanna (2004) metoda parametryzacji i kwantyfikacja stopnia złożoności graficznej pisma. Jak dotąd, za wyjątkiem modyfikacji zaproponowanej przez Peusta (2006), pomysł badania pisma pod względem złożoności nie doczekał się implementacji badawczej.

Celem niniejszego artykułu jest ocena przydatności obu wspomnianych propozycji do analizy pisma chińskiego i zaproponowanie zmian prowadzących do uwzględnienia jego specyfiki.

1. Metody pomiaru złożoności pisma

W założeniu pomysłodawcy metoda obliczania złożoności graficznej powinna wychwytywać i wyrażać w skwantyfikowany sposób różnice w stopniu złożoności znaków. W większości wypadków różnice takie są intuicyjnie wyczuwalne. Dla przykładu, następujący ciąg znaków charakteryzuje się rosnącym stopniem złożoności: I, L, R, 申, 鸞. Altmann zaproponował warunki, jakie powinna spełniać poszukiwana metoda¹:

- stosowalna do dowolnego rodzaju pisma;
- prosta w użyciu;
- dostosowywalna do specyfiki indywidualnych rodzajów i stylów pisma.

Metoda obliczania stopnia złożoności pisma powinna być zatem uniwersalnym narzędziem badawczym, którego stosowalność jest niezależna od typów, rodzajów i stylów pisma – innymi słowy daje porównywalne wyniki dla dowolnego typu pisma alfabetycznego (np. łacińskiego i cyrylicy), sylabicznego czy ideograficznego, włączając różne rodzaje i style czcionek komputerowych oraz pisma odręcznego.

1.1. Metoda kompozycyjna²

Zaproponowana przez Altmanna metoda pomiaru oparta jest na dwóch komponentach – zestawie wartości liczbowych przyporządkowanych różnym typom elementów konstytuujących graficzne znaki pisma:

Tab. 1. Rodzaje elementów graficznych i ich wartości liczbowe³

	Punkt dowolnego rozmiaru	Linia prosta dowolnej długości i kierunku	Łuk dowolnego rozmiaru i kierunku
Wartość	1	2	3
Przykłady	• • ►	-/ \!	∪ ∪ ∩ ∪ ∩ ∩ ∩

oraz zestawie wartości liczbowych przyporządkowanych różnym typom połączeń tychże elementów:

¹ Altmann 2004: 68.

² Terminy 'kompozycyjna' oraz 'intersekcyjna' są zapożyczone z: Peust (2006).

³ *Ibid.*, s. 69.

Reguła 1.⁷: *Złożoność znaku to maksymalna liczba punktów przecięcia, które można osiągnąć za pomocą linii prostej.*

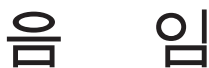
Jak widać, reguła jest prosta. Brakuje przyjętych wartości dla arbitralnie wyróżnionych typów składników znaku, jak to miało miejsce w wypadku metody kompozycyjnej. Sam proces obliczania złożoności również jest dość prosty:



Złożoność powyższych liter wynosi kolejno 3 i 4.

Peust wykazuje, że jego metoda lepiej oddaje intuicje dotyczące złożoności graficznej znaków poprzez przykładowe ciągi znaków o rosnącej złożoności, podczas gdy zastosowanie metody kompozycyjnej daje niekiedy rezultaty sprzeczne z intuicją.

System Peusta wydaje się funkcjonować dla stosunkowo nieskomplikowanego rodzaju pisma, jakim jest pismo oparte na alfabecie łacińskim. Peust, świadomy problemów, jakie mogą sprawić bardziej skomplikowane rodzaje znaków, podaje regułę, która pozwala rozwiązać jeden z problemów poważniejszych. Oto przykład z pisma koreańskiego⁸:



Nawet pobieżne przyjrzenie się zapisowi obu sylab pozwala stwierdzić, że graficznie różnią się one jedynie aranżacją przestrzenną tych samych elementów *jamo*. Intuicyjnie zapisy te nie różnią się między sobą złożonością, ale obliczenia z zastosowaniem reguły 1. nie wychwytyją tego faktu – pojedyncza linia przecinająca oba znaki daje odpowiednio 5 i 4 punkty przecięcia. Jako rozwiązanie Peust proponuje następującą regułę:

Reguła 2.⁹: *Złożoność zbitki graficznej niepołączonych z sobą elementów nie jest obliczana za pomocą pojedynczej linii prostej. Złożoność takiego znaku zdefiniowana jest jako suma złożoności poszczególnych elementów.*

Przy takim potraktowaniu struktury znaków otrzymamy identyczny wynik dla obu zapisów – w pierwszym wypadku nie ma znaczenia, czy użyta zostanie jedna linia czy osobna dla każdego elementu *jamo*, w drugim wypadku otrzymany wynik równy będzie temu, który uzyskujemy dla pierwszego zapisu:

⁷ *Ibid.*

⁸ *Ibid.*

⁹ *Ibid.*, s. 15.



złożoność wynosi 5

Reguła 2. stanowi rozwiązanie również dla analizy pisma chińskiego. Niestety, Peust nie poświęcił mu wiele miejsca, ograniczając się jedynie do kilku przykładów¹⁰, a wydaje się, że znaki chińskie stanowią wyzwanie, które wykracza poza powyższe reguły w przedstawionej formie. Dalsza część artykułu poświęcona będzie analizie złożoności pisma chińskiego.

2. Złożoność pisma chińskiego

W tradycji nauki o chińskich znakach istnieją uznane sposoby określania stopnia ich złożoności. Najczęściej stosowanym kryterium jest liczba kresek, za pomocą których znak jest zapisany. Kategorie złożoności, których wykładnikiem jest liczba kresek, służą co prawda innym celom niż badanie złożoności samej w sobie – np. klasyfikacji, porządkowaniu i sortowaniu znaków do celów leksykograficznych, ale niewątpliwie jest to przy okazji kategoryzacja pod względem stopnia złożoności. Metoda liczenia kresek stosowana w chińskiej leksykografii i np. nauce pisania oraz kaligrafii opiera się na nieprzerwanym kontakcie instrumentu piszącego z materiałem, na którym się pisze. Od razu należy zaznaczyć, że ze względu na oczywiste zróżnicowanie skomplikowania kresek liczonych w ten sposób bardziej miarodajne będzie odwołanie się do tzw. kresek podstawowych, które są elementami kresek bardziej złożonych¹¹.

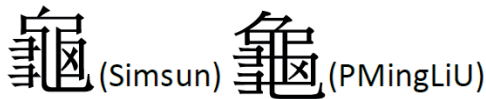
Metoda liczenia kresek jest niezależna od różnic w formie znaków, wynikających z użytej czcionki czy też z różnic w stylach pisma odręcznego. Metody kompozycyjna i intersekcyjna są wyjątkowo wrażliwe na tego typu różnice. Można powiedzieć, że wyniki analiz przeprowadzonych za ich pomocą odnoszą się jedynie do np. danego rodzaju czcionki, a w wypadku pisma odręcznego wręcz do pojedynczych znaków. Nie oznacza to, że metody te są wadliwe, a jedynie, że należy zdawać sobie sprawę, jak interpretować wyniki analiz. Przykładowo, znak 電 *diàn* może być wyświetlony na ekranie komputera bądź przedstawiony w druku za pomocą różnych glifów w zależności od użytej czcionki:

電 (Simsun) 電 (PMingLiU)

¹⁰ *Ibid.*, s. 14-15.

¹¹ Zagadnienie rozróżnienia typów kresek ma duże znaczenie dla omawianych tu problemów. Jednak ze względu na ograniczenie objętości tekstu nie będzie szerzej rozwinięte.

Część, którą znaki wyraźnie się różnią, w pierwszym z nich należy interpretować jako 4 linie proste (wartość kompozycyjna 8), w wypadku drugiego możliwa jest interpretacja uwzględniająca dwa łuki, linię prostą oraz punkt (wartość kompozycyjna 9), choć inna interpretacja też jest zapewne możliwa. W powyższym przykładzie różnice w formie nie wpływają na wyniki uzyskane metodą intersekcyjną. Przykładem, w którym zmiany formy wynikające z użycia różnych czcionek wpływają na wyniki uzyskane za pomocą obu metod, są glify reprezentujące grafem *guī* ‘żółw’:



W tym wypadku różnica występuje nie tylko w kształcie elementów, ale również ich rodzaju oraz liczbie¹². Pismo odręczne wykazuje w tym względzie jeszcze większe zróżnicowanie.

Wyniki obu metod są więc ograniczone do danego stylu pisma. Z drugiej strony, mogą być bez ograniczeń stosowane do dowolnego stylu, rodzaju czy systemu. Liczba kresek jest uznaną i precyzyjną metodą obliczania złożoności chińskich znaków. Jest zdecydowanie mniej wrażliwa na zmiany kształtu znaków, ale ma zastosowanie wyłącznie do znaków chińskich.

Zastosowanie obu metod do analizy pisma chińskiego stawia dwa osobne zestawy problemów do rozwiązania. Zostaną one pokrótce omówione, osobno dla każdej z metod.

2.1. Obliczanie stopnia złożoności znaków chińskich

Przedstawienie problemów związanych z zastosowaniem obu metod do obliczania złożoności znaków chińskich ma charakter przykładowy oraz zwięzły, nieomal hasłowy. Zaproponowane rozwiązania siłą rzeczy dotyczą tylko podanych przykładów i stanowią jedynie propozycje. Dla dodatkowego uproszczenia dyskusji pod uwagę nie są brane problemy z pismem odręcznym.

2.1.1. Metoda kompozycyjna a pismo chińskie

Własności graficzne pisma chińskiego sprawiają, że bezpośrednio zastosowanie metody Altmanna napotyka na przeszkody natury interpretacyjnej. Innymi słowy, konieczne okazuje się podejmowanie mniej lub bardziej arbitralnych decyzji co do typu elementów tworzących znak oraz liczby połączeń.

Pierwszy rodzaj problemów interpretacyjnych dotyczy przyporządkowania typów kresek stosowanych w piśmie chińskim do typów elementów w metodzie kompozycyj-

¹² Warto zauważyć, że w tym wypadku glify różnią się też liczbą kresek.

nej. W zasadzie tylko dwa rodzaje kresek nie nastrożają problemów interpretacyjnych: pozioma 一 (橫 *héng*)¹³ i pionowa | (豎 *shù*). Pozostałe – tak podstawowe, jak i złożone, stanowią bądź to zasadniczy problem interpretacyjny, bądź zależny od stylu czcionki albo wymuszony przez kształt całego znaku. Zasadniczy problem stwarzają np. kreska nazywana tradycyjnie kropką 丶 (點 *diǎn*), która graficznie może być interpretowana również jako łuk, czy też kreska złożona pionowa+haczyk 丿 (豎鉤 *shùgōu*), która może być interpretowana jako pojedynczy łuk, albo dwa elementy z punktem połączenia: dwie proste, prosta i kropka, a nawet prosta i łuk. Wydaje się, że traktowanie tradycyjnej ‘kropki’ jako punktu w metodzie kompozycyjnej (za wyjątkiem wypadków, kiedy jest ona typograficznie rozciągnięta i zdecydowanie bardziej przypomina łuk) jest rozwiązaniem najprostszym i intuicyjnie właściwym. Podobne argumenty przemawiają za traktowaniem kreski pionowa+haczyk jako łuku. Przykładem problemów wynikających ze stosowania różnych czcionek może być kreska wznosząca (提 *tí*), która jest dolnym elementem w grafemach¹⁴:



Na pierwszy rzut oka widać, że możliwa jest cała gama interpretacji.

Inny problem interpretacyjny stanowi liczba połączeń. Wydawałoby się, że prosta formuła $x-1$ (gdzie x jest liczbą łączących się elementów) jest niezawodnym i niekontrowersyjnym sposobem obliczania liczby połączeń w każdej sytuacji. Własności graficzne znaków chińskich powodują jednak, że o ile w większości wypadków formuła ta jest jednoznacznie właściwa, to w co najmniej dwóch rodzajach wystąpień wymagają podjęcia dość arbitralnej decyzji. Pierwszy z nich wynika z dużego stopnia przypadkowości kontaktu elementów. Nie wszystkie elementy struktury chińskich znaków i zależności przestrzennych między jego komponentami są uregulowane formalnymi zasadami. Dotyczy to również zakazów bądź nakazów wzajemnego kontaktu (dotyk, przecięcie) między elementami. Często mają one charakter dowolny i przez to ‘przypadkowy’. W przykładzie punkty wskazane strzałkami to miejsca, gdzie kontakt bądź jego brak jest kwestią nieuregulowaną zasadami kompozycji:



Tego rodzaju wypadki zasadniczo są łatwe do rozwiązania. Ponieważ metoda kompozycyjna i tak opisuje jedynie dany styl pisma, więc niezależnie, czy kontakt jest wymuszony czy nie, obliczenia złożoności powinny oddawać faktyczny stan rzeczy.

¹³ Chociaż nawet i ta w wypadku niektórych czcionek może ulec wygięciu do stopnia przypominającego łuk.

¹⁴ Dla ścisłości – nie są to samodzielne znaki pisma chińskiego, ale ich elementy składowe.

W powyższym wypadku widać jeden kontakt. Bardziej złożony problem stanowią połączenia 3 elementów, z których przynajmniej 2 są już połączone w ramach innego elementu tworzącego kompozycyjną całość, np.:

Z czysto ilościowego punktu widzenia we wskazanym miejscu łączą się 3 elementy. Powinny być zatem liczone 2 połączenia. Można jednak sytuację taką interpretować jako punkt połączenia dwóch komponentów znaku – w tym wypadku 丿 i 口, z których jeden w miejscu połączenia ma już połączenie. Wydaje się, że decyzja o wyborze interpretacji jest tu całkowicie arbitralna. Nie ma żadnych innych czynników uzasadniających taką, a nie inną wybór. Może za wyjątkiem prostoty rozwiązania, która sugerowałaby 1 połączenie we wskazanym miejscu.

2.1.2. Metoda intersekcyjna a pismo chińskie

Prosta w założeniach metoda intersekcyjna ignoruje typ elementów tworzących znaki pisma. Z tego powodu problemy opisane powyżej jej nie dotyczą. Jak się okazuje, znaki chińskie stanowią wyzwanie i wobec tej metody. Wydaje się, że dwie zasady zaproponowane przez Peusta spełniają założoną funkcję w wypadku:

- wyjątkowo prostych znaków (np. 人, 二, 干, 上);
- znaków złożonych o wyraźnie rozdzielonych prostych komponentach (np. 众, 仁, 垠).

W pozostałych wypadkach reguły raczej nie przynoszą zamierzonych wyników obliczeń złożoności, co zostanie zilustrowane kilkoma przykładami. Niedostatki reguł w oryginalnej formie najlepiej widać przy znakach o mniej lub bardziej skomplikowanej strukturze, bez wyraźnego podziału na komponenty:

凸 白 血 金 崩 鳳 龜 蠹

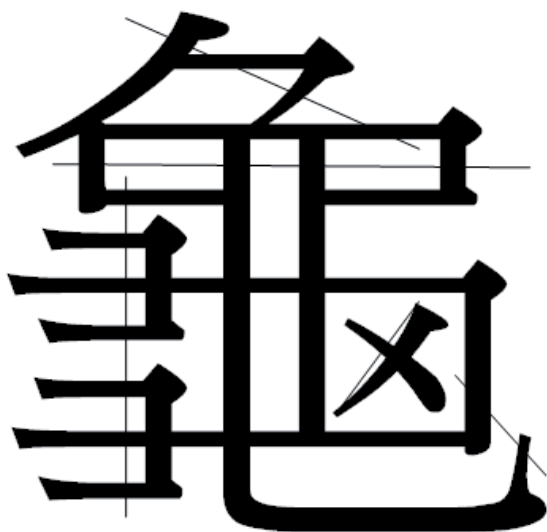
Na pierwszy rzut oka widać, że punkty przecięcia z pojedynczą linią, którą reguły kazały tu zastosować, mogą oddać jedynie część struktury znaków i, co za tym idzie, stopnia ich złożoności. Nawet w znakach prostych, gdzie reguły metody intersekcyjnej zdają się spełniać swoją funkcję:

Z drugiej jednak strony, łatwo zauważyć, że tak poprowadzone przecięcie oddaje dokładnie połowę struktury przykładowego znaku. Są zatem przesłanki, by sądzić, że poniższy sposób jest bardziej adekwatny:



Dodatkową przesłanką świadczącą o tym jest analiza znaków 二 i 匚. Złożoność wszystkich trzech znaków, obliczona za pomocą reguł Peusta, wyniesie 2. Za pomocą zaproponowanych tu zmodyfikowanych reguł odpowiednio 4., 2. i 3., co dokładnie oddaje intuicyjne różnice między nimi.

Ogólna idea kryjąca się za regułą 2. jest niezbędna. Nieuzasadnione jednak wydaje się definiowanie struktury znaków (które można analizować za jej pomocą) przez ‘niepołączone’ komponenty. Konieczne jest przeformułowanie reguły 2. w taki sposób, by obejmowała również znaki, jak powyższe. Przy pełnym braku lub niejasnym podziale na poszczególne komponenty uwzględnione być powinny jednokrotnie wszystkie znaczące fragmenty kompozycji i struktury znaków. Ta dość intuicyjna reguła (która, jak się wydaje, mogłaby być jedyną regułą) nie będzie tu zdefiniowana precyzyjniej. Objasniona zostanie na przykładzie i następnie zastosowana do analizy próbki znaków chińskich (tab. 4.). Znak 龜 jest jednym z najbardziej skomplikowanych znaków pisma chińskiego, w którym nie wyróżnia się komponentów w sensie tradycyjnym. Co prawda mierzenie złożoności znaku nie opiera się na tradycyjnej dekompozycji znaków, ale pomimo tego jest to dobry przykład do zilustrowania rozszerzonej zasady pomiaru:



Wszystkie fragmenty struktury znaku zostały uwzględnione jednokrotnie. Złożoność mierzona w opisany sposób wynosi 20. Maksymalna liczba punktów przecięcia z pojedynczą linią prostą wynosi 10 – różnica jest znaczna. Ten sposób obliczania złożoności intersekccyjnej przynosi bardziej miarodajne wyniki, ale kosztem częściowej utraty prostoty i zwiększenia arbitralności.

2.2. Porównanie metod

Ponieważ nauka o znakach chińskich oferuje uznane i sprawdzone metody badania stopnia skomplikowania znaków, uzasadnione jest ich porównanie z metodami omawianymi w tym artykule. Ze względu na wstępny charakter prowadzonych tu rozważań oraz ograniczenie miejsca porównanie oparte będzie na bardzo niewielkiej, niereprezentatywnej próbie chińskich znaków. Znaki zostały wybrane do analizy na podstawie liczby kresek – po dwa znaki dla każdej kategorii ilościowej w przedziale od 4 do 17 kresek. Wyniki analizy są przedstawione w tab. 4.

Analiza porównawcza ma dwa podstawowe cele:

- określenie zależności ilościowych między wynikami uzyskanymi za pomocą różnych metod;
- zbadanie związku między wzrostem złożoności mierzonym w liczbie kresek.

Tab. 4. Miary złożoności znaków chińskich

Znak	Typy elementów	Typ połączeń	Łączna złożoność kompozycyjna	Złożoność intersekccyjna	Liczba kresek	Liczba kresek podstawowych
升	2(2)+2(3)	1(2)+2(3)	18	4	4	4
手	2(2)+2(3)	1(2)+2(3)	18	4	4	5
玉	1(1)+4(2)	2(2)+3(1)	16	5	5	5
汀	2(1)+2(2)+2(3)	2(2)	16	6	5	7
耳	6(2)	7(2)+1(3)	29	6	6	6
成	1(1)+2(2)+4(3)	3(2)+2(3)	29	7	6	9
夾	1(2)+2(1)+4(3)	3(2)+3(1)	25	7	7	7
豆	1(1)+8(2)	4(2)	25	8	7	8
玥	7(2)+3(2)	8(2)+3(1)	39	9	8	10
囨	10(2)	11(2)+3(2)	48	10	8	10
契	6(2)+4(3)	3(2)+4(3)	42	11	9	11

Znak	Typy elementów	Typ połączeń	Łączna złożoność kompozycyjna	Złożoność intersekccyjna	Liczba kresek	Liczba kresek podstawowych
思	3(1)+6(2)+3(1)	8(2)+3(1)	37	11	9	12
桌	9(2)+3(2)	10(2)+3(1)	47	12	10	11
越	7(2)+4(3)	6(2)+3(3)	47	12	10	14
越	8(2)+4(3)	8(2)+2(3)	50	13	11	13
握	1(1)+9(2)+3(3)	6(2)+2(3)	46	15	11	15
握	1(1)+11(2)+2(3)	3(3)+8(2)	54	15	12	15
握	11(2)+3(3)	6(2)+9(3)	70	16	12	16
鼓	1(1)+12(2)+2(3)	8(2)+3(3)	56	13	13	15
膚	2(1)+11(2)+3(3)	14(2)+3(3)	70	18	13	18
熔	6(1)+5(2)+5(3)	10(2)	51	16	14	16
糖	2(1)+12(2)+2(3)	11(2)+5(3)	69	16	14	16
燎	2(1)+9(2)+5(3)	13(2)+2(3)	67	17	15	17
趾	5(1)+11(2)+1(3)	13(2)	56	17	15	18
賺	1(1)+14(2)+2(3)	14(2)+7(3)	84	19	16	17
葬	4(1)+7(2)+6(3)	7(2)+2(3)	56	19	16	19
養	2(1)+10(2)+6(3)	11(2)+5(3)	77	18	17	18
觥	1(1)+11(2)+9(3)	19(2)+3(3)	97	24	17	25

Zakładając, że liczba kresek podstawowych jest miarodajnym wykładnikiem złożoności, a przy tym jest bliższa koncepcji złożoności Altmanna niż liczba kresek, korelacja wyników uzyskanych omawianymi metodami z ich liczbą może być podstawą weryfikacji tych metod. Poniższy wykres przedstawia zestawienie krzywych uzyskanych dla poszczególnych wartości złożoności.

Rysunek 1. Korelacja wartości złożoności uzyskanych różnymi metodami



Jak widać, krzywe dla liczby kresek podstawowych oraz złożoności intersekcyjnej, liczonej zmodyfikowaną metodą, są mocno skorelowane. Wartości kompozycyjne są bardziej niezależne od liczb kresek podstawowych, ale tendencja wzrostowa wraz ze wzrostem liczby kresek jest zachowana – wydaje się, że obie metody oddają stopień złożoności znaków chińskich. Weryfikacja powinna zostać przeprowadzona na większej liczbie znaków. Reprezentatywna próba pozwoliłaby też na sformułowanie ogólniejszych zależności między liczbą kresek (podstawowych, ale nie tylko) i cechami strukturalnymi znaków a złożonością kompozycyjną i intersekcyjną. Innym sposobem weryfikacji może też być porównanie zmian złożoności znaków po ich uproszczeniu (par znak tradycyjny – znak uproszczony). Większa ilość danych umożliwiłaby też obliczenie wartości średnich dla całej próby oraz poszczególnych kategorii liczby kresek i porównanie ich z sobą.

Bibliografia

- Altmann, G. 2004: Script complexity. *Glottometrics* 8. 68-74.
- 陳學志 (Chen Hsueh-Chih), 張樂勻 (Chang Li-yun), 邱郁秀 (Chiou Yu-Shiou), 宋曜廷 (Sung Yao-Ting), 張國恩 (Chang Kuo-En). 2011: 中文部件組字與形構資料庫之建立及其在識字教學的應用 [Chinese Orthography Database and Its Application in Teaching Chinese Characters]. *教育心理學報* 43卷閱讀專刊 [Bulletin of Educational Psychology (Special Issue on Reading)] 43. 269-290.
- Bishop, T., Cook., R. 2007: A Character Description Language for CJK. *Multilingual* 18.7. 62-68 (<http://www.wenlin.com/cdl/MLC-CDL.pdf>).

- Cook, Richard S. 2001: The Extreme of Typographic Complexity: Character Set Issues Relating to Computerization of The Eastern Han Chinese Lexicon “說文解字” *Shuowenjiezi. 18th International Unicode Conference*. (<http://linguistics.berkeley.edu/~rscook/pdf/IUC18-SWJZZ.wp3-600.pdf>).
- 陈原 (Chen Yuan) 主编 (ed.) 1989: *现代汉语定量分析 [Quantitative Analysis of Modern Chinese]*. 上海教育出版社 (Shanghai: Jiaoyu Chubanshe).
- 莊德明 (Chuang Der-ming), 鄧賢瑛 (Teng Hsian-ying). 2009: 漢字構形資料庫的研發與應用 [Research and Development of Chinese Characters Information Database and Its Application] (<http://cdp.sinica.edu.tw/service/documents/T090904.pdf>).
- 费锦昌 (Fei Jinchang). 1997 (2002). 汉字部件分解的原则 [Principles of Decomposition of Chinese Characters]. In: Su Peicheng 2002a (ed.). 90-101.
- 高更生 (Gao Gengsheng). 1999: *汉字研究 (The Chinese Characters Study)* 济南: 山东教育出版社 (Jinan: Shandong Jiaoyu Chubanshe).
- 韩布新 (Han, Buxin). 1994: 汉字部件信息数据库的建立 – 部件和部件组合频率的统计分 [Development of Database of Chinese Constituents Information – Statistical Analysis of the Frequency of the Constituents and Their Combination]. *心理学报 [Acta Psychologica Sinica]* 26.2. 147-52.
- 韩布新 (Han, Buxin). 1995: 部件组合——潜在的汉字结构层次 [Combination of Chinese character constituents-A latent structural unit]. *中文信息学报 [Journal of Chinese Information Processing]* 9.3. 27-32.
- Hsieh, C. C., Chang, C. T., Huang, J. K. T. 1990: On the Formalization of Glyph in the Chinese Language. *Journal of Library & Information Science* 16.1. 1-26.
- Johnson, K. 2008: *Quantitative Methods in Linguistics*. Blackwell.
- Peust, C. 2006: Script complexity revisited. *Glottometrics* 12. 11-15.
- Rogers, H. 2005: *Writing Systems: A Linguistic Approach*. Oxford: Blackwell.
- Sampson, G. 1985: *Writing systems: a linguistic introduction*. London: Hutchinson.
- Sampson, G. 1994: Chinese script and the diversity of writing systems. *Linguistics* 32. 117-32.
- 沈克成, 沈迦 (Shen Kelong, Shen Jia). 1998: *汉字部件学 [Theory of Chinese Characters Composition]*. 北京: 机械工业出版社 (Beijing: Jixie Gongye Chubanshe).
- Stalph, J. 1989: *Grundlagen einer Grammatik der sinojapanischen Schrift*. Wiesbaden: Harrasowitz Verlag.
- 苏培成 (Su Peicheng). 2001: *现代汉字学纲要 [The Outline of the Modern Chinese Characterology]*. 北京大学出版社 (Beijing: Beijing Daxue Chubanshe).
- 苏培成 (Su Peicheng). 2002a: (ed.). *现代汉字学纲要. 参考资料 [The Outline of the Modern Chinese Characters Study. Reference Materials]*. 北京大学出版社 (Beijing: Beijing Daxue Chubanshe).
- 苏培成 (Su Peicheng). 2002b: 汉字的性质 [On the nature of Chinese characters]. In: Su Peicheng 2002a (ed.). 7-23.
- Yin, B., Rohsenow, J. S. 1994: *Modern Chinese Characters*. Beijing: Sinolingua.
- 张玉金 (Zhang Yujin). 2000: 论汉字的部件拆分和字符拆分 [On the Decomposition of Chinese Characters into Components and Symbols]. *辽宁师范大学学报 (社会科学版) [Journal of Liaoning Normal University (Social Sciences)]* 4. 67-69.

