

UNIVERZITA PALACKÉHO V OLOMOUCI

Filozofická fakulta

Katedra asijských studií



MAGISTERSKÁ DIPLOMOVÁ PRÁCE

**Aplikace Menzerath-Altmanova zákona na současnou
psanou čínštinu**

An Application of the Menzerath-Altman Law to Contemporary
Written Chinese

Tereza Motalová a Lenka Spáčilová

Vedoucí diplomové práce: Mgr. Ondřej Kučera

OLOMOUC 2013

Podklad pro zadání DIPLOMOVÉ práce studenta

PŘEDKLÁDÁ:	ADRESA	OSOBNÍ ČÍSLO
MOTALOVÁ Tereza	Holandská 5, Znojmo	F111013

TÉMA ČESKY:

Aplikace Menzerath-Altmanova zákona na současnou psanou čínštinu

NÁZEV ANGLICKY:

An Application of the Menzerath-Altman Law to the Contemporary Written Chinese

VEDOUCÍ PRÁCE:

Mgr. Ondřej Kučera - ASH

ZÁSADY PRO VYPRACOVÁNÍ:

- 1) Stanovení hypotézy
- 2) Rešeršní fáze
- 3) Získání základních informací, vzhled do problematiky tématu
- 4) Zvolení výběrového souboru pro analýzu
- 5) Stanovení jednotek pro segmentaci výběrového souboru
- 6) Segmentace výběrového souboru
- 7) Kvantifikování výběrového souboru
- 8) Výpočet parametrů Menzerath-Altmanova zákona
- 9) Testování statistickými metodami
- 10) Interpretace výsledků
- 11) Sepsání diplomové práce

SEZNAM DOPORUČENÉ LITERATURY:

1. BEDNAŘÍKOVÁ, Božena. Slovo a jeho konverze. Olomouc: Univerzita Palackého v Olomouci, 2009, 253 s. ISBN 978-80-244-2220-6.
2. ČERMÁK, František. Lexikon a sémantika. Praha: Lidové noviny, 2010, 357 s. ISBN 978-807-4220-203.
3. HŘEBÍČEK, Luděk. Lectures on text theory. Prague: Oriental Institute, 1997, 191 s. ISBN 80-85425-26-2.
4. HŘEBÍČEK, Luděk. Vyprávění o lingvistických experimentech s textem. Praha: Academia, 2002, 195 p. ISBN 80-200-0973-6.
5. CHANG, Hung-nien; LIU, Sze-yun; SHI, Lilin. A practical Chinese grammar. Hong Kong: Chinese University Press, 1994, 520 p. ISBN 96-220-1595-6.
6. PACKARD, Jerome Lee. New approaches to Chinese word formation: morphology, phonology and the lexicon in modern and ancient Chinese. New York: Mouton de Gruyter, 1998, 386 p. ISBN 31-101-5109-X.
7. PACKARD, Jerome Lee. The morphology of Chinese: a linguistic and cognitive approach. Cambridge: Cambridge University Press, 2006. ISBN 978-052-1026-109.
8. ROSS, Claudia; MA, Jing-heng Sheng. Modern Mandarin Chinese grammar: a practical guide. New York: Routledge, 2006. ISBN 04-157-0010-8.
9. ŠVARNÝ, Oldřich; UHER, David. Hovorová čínština: úvod do studia hovorové čínštiny. Olomouc: Univerzita Palackého, 2001, 190 s. ISBN 80-244-0298-X.
10. WIMMER, Gejza. Úvod do analýzy textov. Bratislava: Veda, 2003, 344 s. ISBN 80-224-0756-9.
11. XING, Janet Zhiquan. Studies of Chinese linguistics: functional approaches. Hong Kong: Hong Kong University Press, 2009, 235 p. ISBN 96-220-9964-5.

Podpis studenta: Tereza Motolová

Datum: 5.3.2013

Podpis vedoucího práce: [Signature]

Datum: 5.3.2013

Seř!

Podklad pro zadání DIPLOMOVÉ práce studenta

PŘEDKLÁDÁ:	ADRESA	OSOBNÍ ČÍSLO
SPÁČILOVÁ Lenka	Bezejmená 134, Dub nad Moravou	F111010

TÉMA ČESKY:

Aplikace Menzerath-Altmanova zákona na současnou psanou čínštinu

NÁZEV ANGLICKY:

An Application of the Menzerath-Altman Law to the Contemporary Written Chinese

VEDOUcí PRÁCE:

Mgr. Ondřej Kučera - ASH

ZÁSADY PRO VYPRACOVÁNÍ:

1. Stanovení hypotézy
2. Rešeršní fáze
3. Získání základních informací, vzhled do problematiky tématu
4. Zvolení výběrového souboru pro analýzu
5. Stanovení jednotek pro segmentaci výběrového souboru
6. Segmentace výběrového souboru
7. Kvantifikování výběrového souboru
8. Výpočet parametrů Menzerath-Altmanova zákona
9. Testování statistickými metodami
10. Interpretace výsledků
11. Sepsání diplomové práce

SEZNAM DOPORUČENÉ LITERATURY:

1. BEDNAŘÍKOVÁ, Božena. Slovo a jeho konverze. Olomouc: Univerzita Palackého v Olomouci, 2009, 253 s. ISBN 978-80-244-2220-6.
2. ČERMÁK, František. Lexikon a sémantika. Praha: Lidové noviny, 2010, 357 s. ISBN 978-807-4220-203.
3. HŘEBÍČEK, Luděk. Lectures on text theory. Prague: Oriental Institute, 1997, 191 s. ISBN 80-85425-26-2.
4. HŘEBÍČEK, Luděk. Vyprávění o lingvistických experimentech s textem. Praha: Academia, 2002, 195 p. ISBN 80-200-0973-6.
5. CHANG, Hung-nien, Sze-yun LIU a Lilin SHI. A practical Chinese grammar. Hong Kong: Chinese University Press, 1994, 520 p. ISBN 96-220-1595-6.
6. PACKARD, Jerome Lee. New approaches to Chinese word formation: morphology, phonology and the lexicon in modern and ancient Chinese. New York: Mouton de Gruyter, 1998, 386 p. ISBN 31-101-5109-X.
7. PACKARD, Jerome L. The morphology of Chinese: a linguistic and cognitive approach. Cambridge: Cambridge University Press, 2006. ISBN 978-052-1026-109.
8. ROSS, Claudia a Jing-heng Sheng MA. Modern Mandarin Chinese grammar: a practical guide. New York: Routledge, 2006. ISBN 04-157-0010-8.
9. ŠVARNÝ, Oldřich a David UHER. Hovorová čínština: úvod do studia hovorové čínštiny. Olomouc: Univerzita Palackého, 2001, 190 s. ISBN 80-244-0298-X.
10. WIMMER, Gejza. Úvod do analýzy textů. Bratislava: Veda, 2003, 344 s. ISBN 80-224-0756-9.
11. XING, Janet Zhiqun. Studies of Chinese linguistics: functional approaches. Hong Kong: Hong Kong University Press, 2009, 235 p. ISBN 96-220-9964-5.

gestl

Podpis studenta:

Lenka Galiová

Datum: *5.3.2013*

Podpis vedoucího práce:

Jiří Štěrba

Datum: *5.3.2013*

Seal

Prohlášení

Prohlašujeme, že jsme diplomovou práci vypracovali samostatně a uvedly veškeré použité prameny a literaturu.

V Olomouci dne Podpis:

Podpis:

Anotace

Jména a příjmení autorů:	Tereza Motalová a Lenka Spáčilová
Název katedry:	Katedra asijských studií
Název fakulty:	Filozofická fakulta
Název diplomové práce:	Aplikace Menzerath-Altmanova zákona na současnou psanou čínštinu
Název diplomové práce v angličtině:	An Application of the Menzerath-Altmann Law to Contemporary Written Chinese
Vedoucí diplomové práce:	Mgr. Ondřej Kučera
Rozsah práce:	188
Počet příloh:	4

Práce si klade za cíl testovat platnost Menzerath-Altmanova zákona na současných čínských textech psaných zjednodušenými znaky a různými jazykovými styly, tj. vědeckém článku a článku z blogu. Práce je rozdělena do pěti částí. První kapitola seznamuje čtenáře s kvantitativní lingvistikou a jejími důležitými vývojovými mezníky. V této kapitole je také blíže popsán Menzerath-Altmanův zákon. Druhá kapitola nastiňuje metodologii výzkumu. Další dvě kapitoly, které představují hlavní část práce, jsou zaměřeny na interpretaci získaných výsledků a diskuzi. Získané závěry jsou v páté kapitole komparovány s výsledky výzkumu, který byl provedený v roce 2012 v rámci projektu Studentské grantové soutěže. Závěr sumarizuje výsledná zjištění.

Klíčová slova

Kvantitativní lingvistika, Menzerath-Altmanův zákon, psaná čínština, zjednodušené znaky, jazykové jednotky, segmentace

Poděkování

Touto cestou bychom rády poděkovaly Mgr. Ondřeji Kučerovi nejen za jeho podporu a vstřícný přístup, ale především za inspiraci, diskuze a poskytování velmi cenných rad a připomínek při vedení naší magisterské diplomové práce.

Velké poděkování na tomto místě rovněž patří Mgr. Martině Benešové, Ph.D. za její vstřícnost a trpělivost, s kterými vedla naše konzultace nad danými problematikami tématu, a zároveň za přínosnou spolupráci a podporu, kterou nám po celou dobu vypracování diplomové práce poskytovala.

Obsah

Anotace	7
Obsah	9
Seznam tabulek	11
Seznam obrázků	15
Ediční poznámka	17
Úvod (Motalová, Spáčilová)	18
1 Kvantitativní lingvistika (Motalová, Spáčilová)	22
1.1 Úvod	22
1.2 Dějinný přehled – hlavní mezníky ve vývoji kvantitativní lingvistiky	23
1.3 Menzerath-Altmannův zákon	25
2 Metodologie (Motalová, Spáčilová)	28
2.1 Stanovení kritérií pro volbu výběrových souborů a jejich odůvodnění	28
2.2 Volba výběrových souborů	35
2.3 Stanovení a definování jazykových jednotek a úrovní	35
2.4 Segmentace a kvantifikace výběrových souborů	48
2.5 Testování spolehlivosti modelu pomocí statistických metod	48
2.6 Interpretace získaných dat	49
3 Aplikace Menzerath-Altmannova zákona na vědecký článek (Motalová)	51
3.1 Stanovení kritérií pro volbu výběrového souboru a jejich odůvodnění	51
3.2 Volba výběrového souboru	52
3.3 Stanovení a definování jazykových jednotek a úrovní	53
3.4 Segmentace a kvantifikace výběrového souboru	54
3.5 Testování spolehlivosti modelu pomocí statistických metod	59
3.6 Interpretace získaných dat	60

3.6.1	Jazyková úroveň U4	61
3.6.2	Jazyková úroveň U3	71
3.6.3	Jazyková úroveň U2	79
3.6.4	Jazyková úroveň U1	84
3.6.5	Shrnutí	91
4	Aplikace Menzerath-Altmanova zákona na článek z blogu (Spáčilová)	94
4.1	Stanovení kritérií pro volbu výběrového souboru a jejich odůvodnění	94
4.2	Volba výběrového souboru	96
4.3	Stanovení a definování jazykových jednotek a úrovní	100
4.4	Segmentace a kvantifikace výběrového souboru	103
4.5	Testování spolehlivosti modelu pomocí statistických metod	104
4.6	Interpretace získaných dat	104
4.6.1	Jazyková úroveň U4	105
4.6.2	Jazyková úroveň U3	110
4.6.3	Jazyková úroveň U2	122
4.6.4	Jazyková úroveň U1	128
4.6.5	Shrnutí	133
5	Komparace výběrových souborů (Motalová, Spáčilová)	136
5.1	Komparace výběrových souborů na jazykové úrovni U4 (Motalová)	137
5.2	Komparace výběrových souborů na jazykové úrovni U3 (Motalová)	139
5.3	Komparace výběrových souborů na jazykové úrovni U2 (Spáčilová)	142
5.4	Komparace výběrových souborů na jazykové úrovni U1 (Spáčilová)	144
6	Závěr (Motalová, Spáčilová)	147
7	Resumé	152
8	Seznam použité literatury	153
9	Seznam příloh	164
	Přílohy	165

Seznam tabulek

Tabulka 1 Periodizace vývoje čínského jazyka _____	29
Tabulka 2 Ukázka vývoje vybraných znaků _____	32
Tabulka 3 Ukázka klasifikace základních tahů podle kaligrafických příruček na čínském znaku 永 _____	37
Tabulka 4 Klasifikace znaků: jednoduché tahy _____	38
Tabulka 5 Klasifikace znaků: jednoduché tahy s hákem _____	39
Tabulka 6 Komparace znaků demonstrující oscilaci počtu prvků (P_p) v závislosti na příslušném fontu _____	41
Tabulka 7 Vybraná interpunkční znaménka, platná pro hranice parcelátů _____	43
Tabulka 8 Vybraná interpunkční znaménka, neplatná pro hranice parcelátů _____	45
Tabulka 9 Jazykové úrovně U_i , x_i konstrukt, y_i konstituent ($i = 1, 2, 3, 4$) _____	48
Tabulka 10 Jazykové úrovně U_i , x_i konstrukt, y_i konstituent ($i = 1, 2, 3, 4$) _____	54
Tabulka 11 Alternativní jazyková úroveň U_i , x_i konstrukt, y_i konstituent _____	55
Tabulka 12 Ukázka dvou typů zápisu desetinných čísel a jejich rozdílnosti demonstované na počtu prvků a tahů _____	57
Tabulka 13 Jazyková úroveň U4: znak (měřený v prvcích) – prvek (měřený v průměrném počtu tahů) _____	61
Tabulka 14 Ukázka osmi a devíti prvkových znaků a jejich frekvence v rámci Frekvenčního seznamu znaků programu 文林 Wenlin Software _____	62
Tabulka 15 Parametr A , parametr b a koeficient determinace R^2 pro matematický model související s pozorováními uvedenými v Tabulce 13 _____	63
Tabulka 16 Příklad zjednodušených znaků tvořených jedním prvkem o osmi tahách _____	64
Tabulka 17 Příklady zjednodušených znaků tvořených různými počty prvků a tahů a jejich průměrných délek _____	65
Tabulka 18 Příklad tradičního znaku tvořeného 17 prvky a 32 tahy _____	66
Tabulka 19 Příklad tradičního znaku tvořeného 5 prvky a 30 tahy _____	66
Tabulka 20 Experiment 4A – Jazyková úroveň: znak (měřený v prvcích) – prvek (měřený v průměrném počtu tahů) _____	67

Tabulka 21 Parametr A , parametr b a koeficient determinace R^2 pro matematický model související s pozorováními uvedenými v Tabulce 20	69
Tabulka 22 Jazyková úroveň U3: parcelát (měřený ve znacích) – znak (měřený v průměrném počtu prvků)	71
Tabulka 23 Parametr A , parametr b a koeficient determinace R^2 pro matematický model související s pozorováními uvedenými v Tabulce 22	73
Tabulka 24 Jazyková úroveň U_i se slovem na pozici konstruktů x_i a jazyková úroveň U_i se slovem na pozici konstituentů y_i	75
Tabulka 25 Frekvence znaků podle počtu prvků	77
Tabulka 26 Jazyková úroveň U2: souvětí (měřené v parcelátech) – parcelát (měřený v průměrném počtu znaků)	79
Tabulka 27 Parametr A , parametr b a koeficient determinace R^2 pro matematický model související s pozorováními uvedenými v Tabulce 26	80
Tabulka 28 Ukázka 1 souvětí tvořeného třemi parceláty s jejich výslednou průměrnou délkou v počtu prvků	82
Tabulka 29 Ukázka 2 souvětí tvořeného třemi parceláty s jejich výslednou průměrnou délkou v počtu prvků	82
Tabulka 30 Ukázka 3 souvětí tvořeného třemi parceláty s jejich výslednou průměrnou délkou v počtu prvků	83
Tabulka 31 Jazyková úroveň U1: odstavec (měřený v souvětích) – souvětí (měřené v průměrném počtu parcelátů)	84
Tabulka 32 Parametr A , parametr b a koeficient determinace R^2 pro matematický model související s pozorováními uvedenými v Tabulce 31	85
Tabulka 33 Experiment 1A – Jazyková úroveň U1: odstavec (měřený v souvětích) – souvětí (měřené v průměrném počtu parcelátů)	87
Tabulka 34 Experiment 2A – Jazyková úroveň U2: souvětí (měřené v parcelátech) – parcelát (měřený v průměrném počtu znaků)	87
Tabulka 35 Parametr A , parametr b a koeficient determinace R^2 pro matematický model související s pozorováními uvedenými v Tabulce 33	89
Tabulka 36 Parametr A , parametr b a koeficient determinace R^2 pro matematický model související s pozorováními uvedenými v Tabulce 34	90
Tabulka 37 Současní čínští blogeři	97
Tabulka 38 Použitá interpunkční znaménka	100

Tabulka 39 Jazykové úrovně U_i , x_i konstrukt, y_i konstituent, ($i = 1,2,3,4$) _____	103
Tabulka 40 Jazyková úroveň U4: znak (měřený v prvcích) – prvek (měřený v průměrném počtu tahů) _____	105
Tabulka 41 Hodnoty parametru A , parametru b a koeficientu determinace R^2 pro matematický model související s pozorováními uvedenými v Tabulce 40 _____	106
Tabulka 42 Příklad znaků s jedním, devíti a deseti prvky _____	108
Tabulka 43 Jazyková úroveň U3: parcelát (měřený ve znacích) – znak (měřený v průměrném počtu prvků) _____	110
Tabulka 44 Hodnoty parametru A , parametru b a koeficientu determinace R^2 pro matematický model související s pozorováními uvedenými v Tabulce 43 _____	112
Tabulka 45 Přehled frekvencí znaků podle počtu prvků _____	113
Tabulka 46 Zastoupení všech znaků po odebrání duplicitních hodnot v rámci frekvenčních pásem _____	116
Tabulka 47 Zastoupení všech různých znaků podle počtu prvků v rámci frekvenčních pásem a mimo ně _____	117
Tabulka 48 Experiment 3B – Jazyková úroveň U3: slovo (měřené ve znacích) – znak (měřený v průměrném počtu prvků) _____	119
Tabulka 49 Hodnoty parametru A , parametru b a koeficientu determinace R^2 pro matematický model související s pozorováními uvedenými v Tabulce 48 _____	120
Tabulka 50 Jazyková úroveň U2: souvětí (měřené v parcelátech) – parcelát (měřený v průměrném počtu znaků) _____	122
Tabulka 51 Hodnoty parametru A , parametru b a koeficientu determinace R^2 pro matematický model související s pozorováními uvedenými v Tabulce 50 _____	123
Tabulka 52 Experiment 2C – Jazyková úroveň U2: souvětí (měřené v parcelátech) – parcelát (měřený v průměrném počtu znaků) _____	125
Tabulka 53 Hodnoty parametru A , parametru b a koeficientu determinace R^2 pro matematický model související s pozorováními uvedenými v Tabulce 52 _____	126
Tabulka 54 Jazyková úroveň U1: odstavec (měřený v souvětích) – souvětí (měřený v průměrném počtu parcelátů) _____	128
Tabulka 55 Hodnoty parametru A , parametru b a koeficientu determinace R^2 pro matematický model související s pozorováními uvedenými v Tabulce 54 _____	129
Tabulka 56 Experiment 1C – Jazyková úroveň U1: odstavec (měřený v souvětích) – souvětí (měřený v průměrném počtu parcelátů) _____	130

Tabulka 57 Hodnoty parametru A , parametru b a koeficientu determinace R^2 pro matematický model související s pozorováními uvedenými v Tabulce 56	132
Tabulka 58 Hodnoty parametrů A , b a koeficientů determinace R^2 pro matematický model související s empiricky získanými pozorováními	138
Tabulka 59 Hodnoty parametrů A , b a koeficientu determinace R^2 pro matematický model související s empiricky získanými pozorováními	140
Tabulka 60 Hodnoty parametrů A , b a koeficientu determinace R^2 pro matematický model související s empiricky získanými pozorováními	143
Tabulka 61 Hodnoty parametrů A , b a koeficientu determinace R^2 pro matematický model související s empiricky získanými pozorováními	145

Seznam obrázků

Obrázek 1 Znázornění klesající tendence křivky zobrazující vztah mezi délkou konstruktů x a délkou konstituentu y definovaný MALEM	26
Obrázek 2 Znázornění vztahu mezi konstruktem x (horizontální osa) a konstituenty y (vertikální osa) Menzerath-Altmannova zákona pro tři různé absolutní hodnoty záporného parametru b při $A = 15$	27
Obrázek 3 Jazykové jednotky	36
Obrázek 4 Přehled jazykových jednotek I	55
Obrázek 5 Grafická vizualizace pozorování uvedených v Tabulce 13	63
Obrázek 6 Grafická vizualizace pozorování uvedených v Tabulce 20	69
Obrázek 7 Grafická vizualizace pozorování uvedených v Tabulce 22	72
Obrázek 8 Přehled jazykových jednotek II	75
Obrázek 9 Grafická vizualizace frekvence znaků podle počtu prvků, uvedené v Tabulce 25, oddíl A	77
Obrázek 10 Grafická vizualizace frekvence znaků podle počtu prvků, uvedené v Tabulce 25, oddíl B	78
Obrázek 11 Grafická vizualizace pozorování uvedených v Tabulce 26	80
Obrázek 12 Grafická vizualizace pozorování uvedených v Tabulce 31	85
Obrázek 13 Grafická vizualizace pozorování uvedených v Tabulce 33	89
Obrázek 14 Grafická vizualizace pozorování uvedených v Tabulce 34	89
Obrázek 15 Grafická vizualizace pozorování uvedených v Tabulce 40	106
Obrázek 16 Grafická vizualizace pozorování uvedených v Tabulce 43	111
Obrázek 17 Grafická vizualizace frekvencí znaků podle počtu prvků (s odebranými duplicitními hodnotami), data získaná z Tabulky 45	114
Obrázek 18 Grafická vizualizace frekvencí znaků podle počtu prvků (veškeré hodnoty), data získaná z Tabulky 45	114
Obrázek 19 Grafická vizualizace pozorování uvedených v Tabulce 48	120
Obrázek 20 Grafická vizualizace pozorování uvedených v Tabulce 50	123
Obrázek 21 Grafická vizualizace pozorování uvedených v Tabulce 52	126
Obrázek 22 Grafická vizualizace pozorování uvedených v Tabulce 54	129
Obrázek 23 Grafická vizualizace pozorování uvedených v Tabulce 56	131

Obrázek 24 Grafická vizualizace vztahu mezi znakem a prvkem, novinový článek _____	137
Obrázek 25 Grafická vizualizace vztahu mezi znakem a prvkem, povídka _____	137
Obrázek 26 Grafická vizualizace vztahu mezi znakem a prvkem, vědecký článek	137
Obrázek 27 Grafická vizualizace vztahu mezi znakem a prvkem, článek z blogu	137
Obrázek 28 Grafická vizualizace vztahu mezi parcelátem a znakem, novinový článek _____	139
Obrázek 29 Grafická vizualizace vztahu mezi parcelátem a znakem, povídka _____	139
Obrázek 30 Grafická vizualizace vztahu mezi parcelátem a znakem, vědecký článek _____	139
Obrázek 31 Grafická vizualizace vztahu mezi parcelátem a znakem, článek z blogu _____	139
Obrázek 32 Grafická vizualizace vztahu mezi souvětím a parcelátem, novinový článek _____	142
Obrázek 33 Grafická vizualizace vztahu mezi souvětím a parcelátem, povídka _____	142
Obrázek 34 Grafická vizualizace vztahu mezi souvětím a parcelátem, vědecký článek _____	142
Obrázek 35 Grafická vizualizace vztahu mezi souvětím a parcelátem, článek z blogu _____	142
Obrázek 36 Grafická vizualizace vztahu mezi odstavcem a souvětím, novinový článek _____	144
Obrázek 37 Grafická vizualizace vztahu mezi odstavcem a souvětím, povídka – varianta 1 _____	144
Obrázek 38 Grafická vizualizace vztahu mezi odstavcem a souvětím, povídka – varianta 2 _____	144
Obrázek 39 Grafická vizualizace vztahu mezi odstavcem a souvětím, vědecký článek _____	145
Obrázek 40 Grafická vizualizace vztahu mezi odstavcem a souvětím, článek z blogu _____	145

Ediční poznámka

V celé práci je důsledně používána zjednodušená podoba čínského znakového písma s výjimkou příkladů ilustrující tradiční podobu znaků. Pro přepis znakového písma jsme zvolily čínskou standardizovanou transkripční abecedu pinyin (pīnyīn, 拼音字母). Každý termín nejprve uvedený v podobě bez tónu doprovází závorka s pinyinem s tóny a znaky. Pinyin s tónem je pro přehlednost uváděn v jiném fontu – Courier New. Názvy čínských zdrojů jsou doprovázeny pinyinem bez tónu a českým překladem, jména autorů pouze pinyinem bez tónu.

Bibliografická citace vychází z normy ISO 690: 2011 – forma harvardského systému. Odkaz každé citace je vsunut přímo do textu a je zaznamenán metodou prvního prvku a data, kde první prvek představuje jméno autora, příp. název díla (u čínských zdrojů uváděný v pinyinu bez tónů), za nímž následuje rok vydání publikace (popřípadě i číslo strany). Při citování více děl jsou bibliografické citace seřazeny abecedně, u více děl od stejného autora jsou seřazeny chronologicky od nejstaršího po nejnovější. Co se týče přímých citací zahraničních děl, původní text je vždy vložen do poznámky pod čarou.

Seznam použité literatury je uveden v závěru práce, přičemž díla jsou zde primárně řazena abecedně dle příjmení autorů, v případě více děl od stejného autora jsou sekundárně řazena dle roku vydání.

Poznámky vztahující se k textu jsou uvedeny pod čarou na téže stránce a jsou označeny číselným indexem.

Úvod

Kvantitativní lingvistika je se svou zemí původu spíše doménou evropského kontinentu. Ačkoliv se tato vědní disciplína utváří teprve přes jedno století, během tohoto krátkého období si vydobyla nejen své místo mezi ostatními vědními disciplínami, ale i prestiž.

„Velmi rychlého vývoje dosáhla kvantitativní lingvistika v Německu a východní Evropě. Většina současných představitelů tohoto oboru pochází z Německa, Rakouska a zemí východní Evropy. Nejznámějším z nich je profesor z německé Ruhrské univerzity v Bochumi – Gabriel Altmann“¹ (Liu a Huang, 2012, s. 180), který je považován za jednoho ze zakladatelů moderní kvantitativní lingvistiky. Luděk Hřebíček dokonce uvádí, že „skutečná lingvistika druhé poloviny 20. století je altmannovskou lingvistikou.“ (Hřebíček, 2008, s. 488).

Jelikož se „badatelé zapojení do kvantitativně lingvistického výzkumu koncentrují zejména v Německu a východní Evropě, hlavními oblastmi výzkumu jsou indoevropské jazyky, slovanské jazyky a hláskové písmo. Výzkumy zaměřující se na čínštinu a čínské písmo jsou jak z hlediska rozsahu výzkumu, tak i jeho hloubky stále nedostačující. ... Čínština a čínské písmo se svou formou navíc liší od indoevropských jazyků a latinky. Zda jsou teorie (zákonitosti) a metody odvozené z indoevropských jazyků a latinky aplikovatelné na kvantitativně lingvistický výzkum čínštiny, vyžaduje další bádání.“² (Liu a Huang, 2012, s. 182)

S ohledem na výše nastíněnou situaci bude hlavním předmětem této diplomové práce (DP) právě kvantitativně lingvistický výzkum současné psané čínštiny. Protože jedním ze zásadních přínosů G. Altmanna do oblasti kvantitativní lingvistiky byla formulace Menzerath-Altmanova zákona (MAL³), práce se konkrétně zaměří na aplikaci tohoto jazykového zákona, jehož prostřednictvím se budou testovat současné čínské texty psané zjednodušenými znaky a různými

¹ „此后，计量语言学在德国和东欧得到了快速发展。目前这一领域的主要代表人物大多来自德国、奥地利及东欧国家，其中最著名的是德国波鸿大学的 Gabriel Altmann 教授。“ (Liu a Huang, 2012, s. 180)

² „从事计量语言学研究的学者主要集中在德国与东欧地区，研究对象主要是印欧语、斯拉夫语和拼音文字。以汉语和汉字为对象的计量语言学研究无论从研究范围还是研究深度来说都还很不够。... 再加上汉语、汉字在形式上与印欧语、拉丁文字有较大差别，那些从印欧语和拉丁文字等语言材料中发现的计量语言学理论（定律）和方法是否适用于汉语和汉字的研究，仍需进一步检验。“ (Liu a Huang, 2012, s. 182)

³ zkratka pochází z anglického termínu Menzerath-Altman law

jazykovými styly. Práce svými analýzami přispěje nejen do oblasti kvantitativně lingvistického výzkumu, ale zároveň navýší množství zkoumaných výběrových souborů a zpracovaných dat.

Výše nastíněné téma se bude opírat o předchozí výzkum provedený v rámci dokončeného a obhájeného projektu Studentské grantové soutěže 2012 *Segmentace pro testování Menzerath-Altmanova zákona a hypotéz s ním souvisejících*, číslo projektu FF_2012_035. Získané výsledky budou publikovány ve vědeckém časopise *Czech and Slovak Linguistic Review* 1/2013 pod názvem *An Application of the Menzerath-Altman Law to Contemporary Written Chinese*.

Tímto výzkumem realizovaným v roce 2012 jsme testovaly hypotézu, že na současném psaném čínském jazyce lze pomocí Menzerath-Altmanova zákona testovat platnost vzájemných vztahů jazykových jednotek na různých jazykových úrovních na základě grafického hlediska. Za výběrové soubory jsme si z důvodu kontrastnosti stylů zvolily novinový článek a povídku. Jednotlivé jazykové jednotky jsme vymezily na základě přesně daných kritérií, v našem případě dle grafiky textu s minimálním nevyhnutelným přihlédnutím k syntaktickému hledisku. V průběhu experimentu jsme operovaly s šesti jazykovými jednotkami: tah – prvek – znak – parcelát – souvětí – odstavec, které jsme daly do vzájemných vztahů a tím vytvořily čtyři jazykové úrovně, na nichž jsme ověřovaly platnost MALu. Výsledky analýz při daných způsobech segmentace ukázaly, že se MAL prokázal jako adekvátní a dobře sedící model na hladinách: znak – prvek a souvětí – parcelát. Ačkoliv se tendence definována tímto zákonem ukázala také na hladině parcelát – znak, výsledná shoda s matematickým modelem MALu byla minimální. Na hladině odstavec – souvětí se MAL se neprojevil vůbec.

Z důvodu návaznosti na tuto fázi výzkumu (2012) jsme se diplomovou prací rozhodly vypracovat formou týmového projektu a pokračovat tak v předchozím společném výzkumu. Dalším důvodem, který nás vedl k volbě vypracování DP touto formou, je možnost komparace výsledků experimentu provedeného v rámci výše uvedeného projektu s výsledky experimentu realizovaného v této diplomové práci. Metodologie obou výzkumů je navolena tak, aby umožnila výsledky vzájemně porovnat a následně na základě této komparace stanovenou hypotézu ověřit či vyvrátit, příp. přinést hypotézy nové. Posledním důvodem byla statistická náročnost výpočtů.

Tendence nastíněné výsledky experimentu 2012 jsme se rozhodly dále testovat na větším množství výběrových souborů, abychom ověřily následující hypotézy. Na jazykové úrovni znak – prvek předpokládáme, že se platnost MALu potvrdí a empiricky získaná pozorování opět vykáží s matematickým modelem velmi vysokou shodu. Naší tezí je, že platnost tohoto jazykového zákona a výsledná extrémní shoda na jazykových stylech výběrových souborů nezávisí, neboť utváření vzájemných vztahů jednotek na této úrovni ovlivňuje stále stejný faktor – grafické pole.

Na úrovni parcelát – znak předpokládáme, že se platnost MALu naopak nepotvrdí, příp. empiricky získaná pozorování vykáží s matematickým modelem MALu minimální shodu, čímž rovněž potvrdí, že MAL na této úrovni nepředstavuje adekvátní a dobře sedící model. Faktorů, které by mohly narušit vztah těchto jednotek, je hned několik. Prvním z nich by mohl být rozdílný charakter jazykových jednotek: parcelát představuje jednotku s proměnlivou délkou, kdežto znak se svojí pevně danou strukturou jednotku s délkou neměnnou. Zásadní vliv by zde mohla mít i reforma písma, která u 2 236 znaků snížila počty tahů a prvků, a tím i rozdíly mezi znaky z hlediska počtu prvků. Protože i tyto faktory jsou pravděpodobně nezávislé na jazykových stylech, předpokládáme, že výběrové soubory testované v této diplomové práci vykáží velmi podobné tendence.

Domníváme se, že na úrovni souvětí – parcelát se platnost MALu potvrdí, neboť se jedná o jazykové jednotky, jejichž délku tvoří autor sám v souladu se svými požadavky a v závislosti na svých vyjadřovacích schopnostech. Na této úrovni se totiž jeví, že zde nepůsobí žádný faktor, který by mohl vzájemný vztah těchto jednotek narušit.

Na úrovni odstavec – souvětí naopak předpokládáme, že by se potenciální vztah mezi těmito jazykovými jednotkami nemusel projevit z důvodu krátké tradice používání interpunkce v čínských textech psaných ve znacích. Interpunkční znaménka, která vytváří hranice souvětí a parcelátu, představovala cizí import a pro čínské texty znamenala jistou inovaci. Tyto jednotky by proto nemusely být ustálené a jednoznačné. Dalším důvodem neplatnosti MALu by mohla být nízká četnost výskytu pozorování, která byla prokázána již v předchozím experimentu (2012). Abychom mohly co nejpřesněji porovnat výběrové soubory i s předchozím experimentem, je nutné zvolit texty s podobnou délkou. Proto očekáváme, že četnost výskytu bude mít na platnost MALu obdobný vliv.

Diplomová práce je rozdělena do pěti částí. V první, teoretické části, nastíníme důležité mezníky ve vývoji kvantitativní lingvistiky a představíme významné zástupce této vědní disciplíny. Zvláštní podkapitola bude věnována samotnému Menzerath-Altmannovu zákonu. Metodologie výzkumu a její jednotlivé kroky budou detailně představeny v kapitole druhé. Třetí a čtvrtá kapitola bude uvádět výsledky experimentů, které zde budou zároveň interpretovány. Poslední část je věnována komparaci výsledků získaných jak z první fáze (2012), tak i z druhé fáze výzkumu (2013). Na konci práce budou získané závěry sumarizovány. Autorství jednotlivých kapitol a podkapitol je uvedeno v obsahu u jejich názvů.

Ačkoliv se na práci podílely dvě autorky, závěrečná redakce a editace proběhla ve spolupráci z důvodu ucelenosti a jednotného stylu.

1 Kvantitativní lingvistika

1.1 Úvod

Kvantitativní lingvistika představuje dílčí disciplínu matematické lingvistiky, která vznikla na přelomu padesátých a šedesátých let. Za oficiální počátek matematické lingvistiky se považuje rok 1957, kdy se konal osmý mezinárodní lingvistický kongres v Oslu uspořádaný mezinárodním komitétem („Comité international permanent des Linguistes“, CIPL). (Černý, 1996, s. 248) a (Havránek, Horálek, 1958, s. 47-52)

Předmětem matematické lingvistiky je výzkum přirozeného i umělého jazyka za využití matematických, popř. logických metod (Těšitelová, 1987, s. 7). Jak ve své knize uvádí Jiří Černý, tento vědní obor také kromě kvantitativní lingvistiky zahrnuje lingvistiku algebraickou, která využívá kvalitativních metod, a strojovou (počítačovou), která využívá metody jak kvantitativní, tak i kvalitativní. Ačkoliv je vznik matematické lingvistiky datován do druhé poloviny 20. století, matematické metody pronikaly do jazykovědy již od konce století předchozího. Protože tyto metody svým charakterem odpovídaly kvantitativním metodám, dnes se na ně spíše pohlíží jako na počátky kvantitativní lingvistiky. Algebraická a strojová lingvistika se naopak tak dlouhé tradici netěší, neboť vznikly přibližně o šedesát let později v souvislosti s počátky moderní logiky a výpočetní techniky. (Černý, 1996, s. 248)

Kvantitativní lingvistika se zabývá výzkumem jazykových jevů a jejich vztahů pomocí vhodných kvantitativních metod. Je důležité zdůraznit, že „... kvantitativní údaje, formule apod. nejsou cílem kvantitativní lingvistiky, ..., nýbrž prostředkem, popř. kontrolou našeho poznání. Z hlediska lingvistické je nezbytné podávat lingvistickou interpretaci zjištěných statistických dat, tříd, formulí apod.“ (Těšitelová, 1987, s. 9)

1.2 Dějinný přehled – hlavní mezníky ve vývoji kvantitativní lingvistiky

Hlavní mezníky ve vývoji kvantitativní lingvistiky

Jak již bylo zmíněno výše, kvantitativní lingvistika se začala vyvíjet od konce 19. století v souvislosti s pronikáním matematiky do dalších vědních oborů. Za první matematicko-jazykovědnou spolupráci by se dala považovat diskuze nad využíváním kvantitativních metod v jazykovědě, kterou do jazykovědy zavedl již v roce 1847 ruský matematik **Viktor Jakovlevič Buňakovskij**. Statistiku při svých úvahách o jazyce zohledňoval i německý mladogramatik **Herman Paul**, který považoval jazyk za statistický průměr všech jazykových projevů jednotlivců. K dalšímu využití matematiky v jazykovědě došlo v 80. letech, kdy český matematik a fyzik **August Seydler** pomocí počtu pravděpodobností ověřoval pravost tzv. Rukopisů – královedvorského a zelenohorského.

Jedním z nejdůležitějších pojmů kvantitativní lingvistiky byla od jejího počátku frekvence. Prvním lingvistou, který se tímto pojmem zabýval, byl Američan **William Dwight Whitney**. V 70. letech 19. století se zaměřil na výzkum frekvence anglických hlásek. Frekvence se také využívala v oblasti stenografie. Americký vynálezce **Samuel Morse** přisuzoval symboly anglickým písmenům podle jejich frekvence – nejjednodušší znaky přiřadil nejfrekventovanějším písmenům, nejsložitější nejméně frekventovaným písmenům. Německý stenograf **Friedrich Wilhelm Käding** se v závěru 19. století zabýval frekvencí slov – svá pozorování publikoval v „Slovníku četnosti výskytu německého jazyka“, který je prvním slovníkem četnosti vůbec.

Ve 20. století přispěli v k výraznému rozvoji kvantitativní lingvistiky ruský matematik **Andrey Andreyevich Markov** a americký lingvista a psycholog německého původu **George Kingsley Zipf**.

A. A. Markov na počátku 20. století objevil zvláštní jev – v mluveném jazyce se k již vysloveným jazykovým jednotkám připojují jednotky nové, a to podle relativní frekvence závazné pro daný jazyk. Tento jev byl nazván *Markovův proces*. Nejenže se jednalo o první důsledné využití statistiky v lingvistice, ale zároveň tato teorie dala impuls k rozvoji teorie pravděpodobnosti, matematické statistice a teorie informace.

Dalším podobným výzkumem se ve 20. a 30. letech zabýval **G. K. Zipf**. Jeho největším přínosem byly matematicky formulované zákonitosti slovní zásoby přirozených jazyků, které jsou všeobecně známy jako *Zipfovy zákony*. První Zipfův zákon řeší vztah mezi pořadím (rankem) slova při uspořádání slov od nejvyšší frekvence po nejnižší a jeho frekvencí. Vztahem poměru mezi frekvencí slova a počtem různých slov, kterou tuto frekvenci mají, se zabývá druhý Zipfův zákon. Třetí Zipfův zákon se týká vztahu mezi frekvencí slova a počtem jeho významů.

Tyto a další informace o vývoji kvantitativní lingvistiky viz (Černý, 1996) a (Těšitelová, 1987).

Hlavní mezníky ve vývoji kvantitativní lingvistiky v českých zemích

Počátky kvantitativní lingvistiky se u nás kladou na začátek 20. století a jsou spojeny s pedagogikou a rozvojem těsnopisů. O souvislý vývoj této vědní disciplíny se zasloužili zejména zástupci Pražské školy **Vilém Mathesius**, **Bohumil Trnka** a **Josef Vachek**, kteří ve 30. letech 20. století zkoumali fonologickou a lexikální oblast jazyka. Spolu s nimi byli velkým přínosem pro kvantitativní lingvistiku také **Jiří Krámský** a **Roman Jakobson**. Kromě nich se kvantitativní lingvistikou zabývali pedagog **Václav Příhoda** a bohemista **Vladimír Šmilauer**, kteří podnítily vznik českého frekvenčního slovníku.

V následujících letech stála kvantitativní lingvistika na okraji zájmu, k jeho obnově došlo až v 60. letech. Zásadní prací v tomto období se stal již výše zmíněný frekvenční slovník češtiny, první svého druhu u nás. Jeho autory byli **Jaroslav Jelínek**, **Josef Václav Bečka** a **Marie Těšitelová**. Na počátku 60. let také došlo k institucionalizaci kvantitativní lingvistiky založením Oddělení matematické lingvistiky při Ústavu jazyka českého. Výsledky za 25letou existenci tohoto oddělení byly shrnuty v monografii M. Těšitelové, která se stala významnou představitelkou české kvantitativní lingvistiky.

V 80. letech se v kvantitativní lingvistice začínají uplatňovat pravděpodobnostní modely jako základní nástroje výzkumu. Těmito modely se jako první zabýval matematický lingvista **Jan Králík**.

Na závěr je třeba uvést další významnou osobnost – orientalistu a textologa **Lud'ka Hřebíčka**, který se zabývá kvantitativní teorií textu a Menzerath-Altmanovým zákonem.

Tyto a další informace o vývoji kvantitativní lingvistiky na našem území viz (Benešová, 2011) a (Uhlířová, 2005).

1.3 Menzerath-Altmanův zákon

Známým jazykovým zákonem kvantitativní lingvistiky se ve 20. století stal Menzerath-Altmanův zákon, který se považuje za mezník ve vývoji vnímání jazyka a který přispěl ke vzniku dílčích lingvistických disciplín.

Zatímco před-menzerathovská jazykověda zkoumala jazykové jevy izolovaně, německý lingvista Paul Menzerath si jako první všiml určitého vztahu mezi slabikou a slovem. Na základě rozboru německých slov došel v roce 1928 k závěru, že čím delší je slovo v počtu slabik, tím kratší je v průměru slabika. Toto pozorování však zůstalo po dlouhou dobu mimo zájem lingvistů. (Altmann, 1980, s. 124), (Andres et al, 2012, s. 1), (Hřebíček, 2002, s. 53), (Hřebíček, 2007, s. 84) a (Hřebíček, 2008, s. 490)

Teprve až v osmdesátých letech na myšlenky P. Menzeratha navázal Gabriel Altmann zavedením pojmů *konstrukt* a *konstituent*. Konstrukt představuje jazykovou jednotku určité jazykové úrovně a jeho konstituent jazykovou jednotku úrovně bezprostředně nižší. G. Altmann vztahy mezi těmito jazykovými jednotkami na různých jazykových úrovních testoval a prokázal tak, že mezi jednotlivými jazykovými jednotkami existuje následující vzájemný vztah (Altmann, 1980, s. 124):

Čím delší je jazykový konstrukt, tím kratší jsou jeho konstituenty.

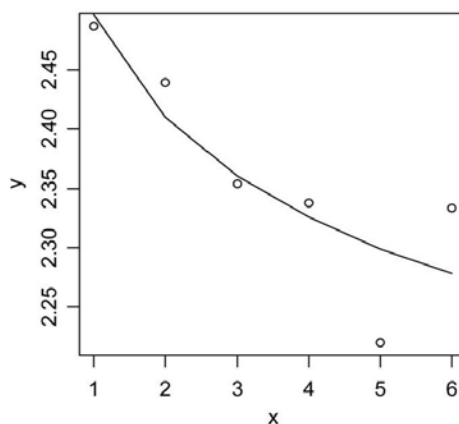
Pomocí těchto dvou nových jazykových pojmů P. Menzerathovo pozorování zobecnil a povýšil na jazykový zákon, pro který navrhl název *Menzerathův zákon*. „Hypotéza vyslovená Menzerathem naznačuje, jaký by měl být mezi těmito dvěma veličinami vztah: čím větší je konstrukt, tím menší je konstituent. To je vztah nepřímé úměrnosti čili nepřímé proporcionality. ... V Altmannově uvažování je tato hypotéza, původně formulovaná Menzerathem, doplněna předpokladem přímé

úměrnosti mezi nekonečně malým přírůstkem velikosti konstituentů a danou velikostí konstituentů: čím větší je konstituent, tím větší je jeho přírůstek – to vše uvažováno jako nějaká střední hodnota čili průměr.“ (Hřebíček, 2002, s. 54) G. Altmann pro tento vztah rovněž odvodil matematickou podobu ve formě mocninného zákona. Jeho jednoduchá verze zní:

$$y = Ax^{-b}$$

kde x je délka konstruktu měřená v jeho konstituentech, y je průměrná délka jeho konstituentů měřená v jednotkách na nejbližší nižší úrovni jazyka a A , b jsou kladné reálné parametry.

Délka konstruktů je měřená v konstituentech a představuje tak vždy celé číslo. Zatímco délka konstituentů je měřená v průměrných hodnotách zjištěných délek a nejčastěji se tedy jedná o desetinné číslo. Ze vzorce je patrné, že s rostoucí velikostí konstruktů klesá velikost konstituentů. (Hřebíček, 2007, s. 84) Při grafické vizualizaci vztahů mezi těmito dvěma veličinami bude mít výsledná křivka klesající tendenci (viz Obrázek 1).

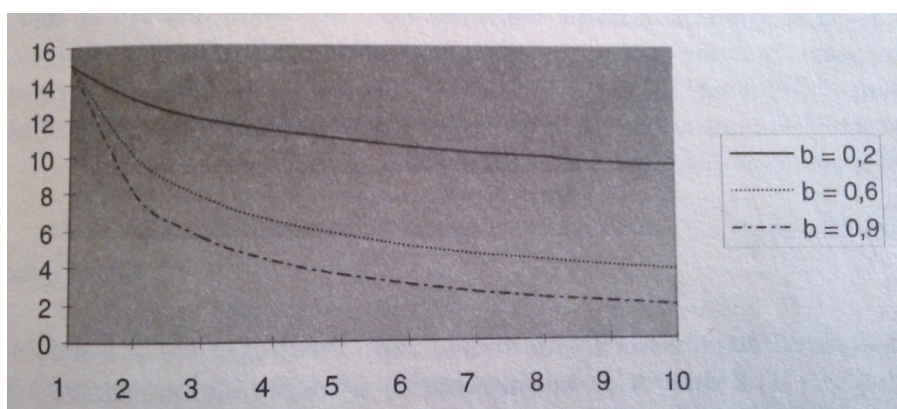


Obrázek 1 Znárodnění klesající tendence křivky zobrazující vztah mezi délkou konstruktů x a délkou konstituentů y definovaný MALEM

(Zdroj: Andres et al, 2012, s. 15)

Aby byla prokázána platnost MALu, parametr b musí být kladné reálné číslo, neboť v tomto případě vyjadřuje míru přímé úměrnosti. Ve vzorci přibírá záporné znaménko a mění se tak na zápornou hodnotu, která reprezentuje nepřímou

úměrnost. „ ... parametr b představuje onu zmíněnou společnou míru proporcionality, která sjednocuje dvě hypotézy o proporcionalitě uvažovaných veličin.“ (Hřebíček, 2002, s. 55-56) V grafické interpretaci to znamená, že křivka je klesající a konvexní, viz Obrázek 2. Parametr A udává posun na ose y , tedy vzdálenost křivky od osy x . Dá se říct, že se jedná o počáteční hodnotu dané křivky. (Kelih, 2010, s. 71) Hodnoty obou parametrů jsou získány pomocí prostředků statistického softwaru.



Obrázek 2 Znázornění vztahu mezi konstruktem x (horizontální osa) a konstituenty y (vertikální osa) Menzerath-Altmanova zákona pro tři různé absolutní hodnoty záporného parametru b při $A = 15$

(Zdroj: Hřebíček, 2002, s. 55)

Z důvodu přínosu G. Altmanna se pro tento zákon vžil název Menzerath-Altmanův zákon.

Tyto a další informace viz (Altmann, 1980), (Andres et al, 2012), (Hřebíček, 1997), (Hřebíček, 2002), (Hřebíček, 2007), (Hřebíček, 2008) a (Kelih, 2010, s. 71).

2 Metodologie

Cílem tohoto experimentu je testování platnosti vzájemných vztahů jazykových jednotek na určitých jazykových úrovních na dalších výběrových souborech psaných rozdílným jazykovým stylem a získání dalších výsledků, které rozšíří předchozí výzkum. Proto jsme i v tomto případě pro testování výběrových souborů zvolily kvantitativní metodu, a to Menzerath-Altmanův zákon. Experiment jsme realizovaly v následujících krocích:

1. Stanovení kritérií pro volbu výběrových souborů a jejich odůvodnění
2. Volba výběrových souborů
3. Stanovení a definování jazykových jednotek a úrovní
4. Segmentace a kvantifikace výběrových souborů
5. Testování spolehlivosti modelu pomocí statistických metod
6. Interpretace získaných dat

2.1 Stanovení kritérií pro volbu výběrových souborů a jejich odůvodnění

Volba výběrových souborů je z hlediska jejich vhodnosti pro tento experiment fundamentálním krokem, jenž vyžaduje precizní stanovení kritérií. Proto jsme tedy nejprve musely určit podmínky, které budou platné pro výběr obou vzorků. Pro účely našeho experimentu jsme zvolily následující kritéria:

- A. Současná moderní čínština
- B. Psaná podoba – zjednodušené znaky
- C. Jazykové styly
- D. Aktuálnost
- E. Souvislost výběrových souborů
- F. Délka výběrových souborů
- G. Renomé autora či instituce

Ke každému výše uvedenému kritériu věnujme nyní několik slov.

A. Současná moderní čínština

„Čínština je jedním z mála současných jazyků, jejichž historie je zdokumentovaná v nepřerušené tradici sahající do druhého tisíciletí př. n. l.“⁴ (Norman, 2012, s. ix) Čínský jazyk prošel během tohoto dlouhého vývoje několika jazykovými změnami. Na základě analýz dochovaných písemných památek lze jeho vývoj rozdělit do několika etap, které však nelze z důvodu komplikovaného vývoje jednoznačně časově vymezit, více viz (Švarný, 1967) a (Vochala, Hrdličková, 1985). Jednou z alternativ periodizace čínštiny se nabízí např. návrh čínského lingvisty Wang Liho⁵, viz (Wang Li, 2004), který se především opíral o změny gramatické a fonetické a částečně i o změny slovní zásoby. Podle něj by se vývoj čínského jazyka mohl dělit na následující období:

Tabulka 1 Periodizace vývoje čínského jazyka

Období	Znaky	Pinyin	Datace	Poznámky
Starověké	上古期	Shànggǔqī	do 3. st. n. l.	3. – 4. st. přechodná fáze
Středověké	中古期	Zhōnggǔqī	4. – 12. st.	12. – 13. st. přechodná fáze
Předmoderní	近代	Jìndài	13. – 19. st.	od Opiové války roku 1840 do Hnutí 4. května r. 1919 přechodné období
Moderní	现代	Xiàndài	20. st.	od Hnutí 4. května

Zdroj: Wang Li, 2004, s. 43-44; vlastní zpracování

⁴ “Chinese is only one of a very few contemporary languages whose history is documented in an unbroken tradition extending back to the second millennium BC.” (Norman, 2012, s. ix)

⁵ Wang Li 王力 (1900 – 1986) představuje jednoho z významných čínských lingvistů 20. století. Během svého života napsal 40 knih a přes více než 200 akademických článků. Jeho lingvistická díla jsou zaměřena především na experimentální fonetiku a gramatiku klasické a moderní čínštiny. Mimo tyto dvě hlavní oblasti se také věnoval standardizaci a výuce jazyka a struktúře poezie. Byl prvním, kdo při svém studiu čínské gramatiky užil vlastních termínů, neboť nebyl příliš nakloněn slepému využívání konceptů anglické gramatiky v čínském jazyce. (LaPolla, 2005, s. 514-515)

Kromě Wang Liho se můžeme setkat i s dalšími typy periodizací, které se však rozcházejí v datování jednotlivých fází z důvodu rozdílných přístupů. Uvedme si zde několik příkladů. Čínský lingvista a profesor Ping Chen se při periodizaci opírá pouze o gramatické hledisko. I když ve své knize *Modern Chinese: History and Sociolinguistics* používá stejnou terminologii jako Wang Li, u časového vymezení jednotlivých fází se s tímto autorem značně rozcházejí, podrobněji viz (Chen, 1999). Americký sinolog a lingvista Jerry Norman ve své knize *Chinese* uvádí periodizaci švédského sinologa Bernharda Karlgrena, který ji publikoval již v letech 1915-1916. B. Karlgren staví periodizaci čínského jazyka pouze na fonologickém aspektu a používá nejen rozdílné datování, ale i odlišnou terminologii, podrobněji viz (Norman, 2012).

Kvůli rozdílnosti jednotlivých vývojových fází čínského jazyka by se dalo mluvit jako o několika různých jazycích, neboť každý z nich se od ostatních určitým způsobem vymezuje a odlišuje. Proto je nutné přesně časově stanovit, jakou vývojovou fází se bude daný výzkum zabývat. Co se týče našeho experimentu, pak jsme jej zaměřily na poslední vývojové období – moderní čínštinu. Protože se však i během tohoto poměrně krátkého časového období čínština dramaticky měnila, je třeba přesně vymezit, do jaké jazykové fáze budou spadat námi zvolené výběrové soubory. Protože klademe důraz na synchronní aspekt tohoto výzkumu, naším prvním kritériem bylo vybrat texty zachycující standardní formu současné moderní čínštiny, známou jako putonghua (pǔtōnghuà, 普通话).

B. Psaná podoba – zjednodušené znaky

Nejstarší písemné památky pocházejí z doby Shangské dynastie (17. st. – 11. st.). Jedná se o věšebné nápisy na želvích krunýřích a hovězích lopatkách (jiǎgǔwén, 甲骨文) z období mezi 14. a 11. st. př. n. l. Paralelně s nimi vznikaly nápisy na bronzových nádobách (jīnwén, 金文). Obě písma vychází z piktogramů a jsou již značně vyspělá. „Jejich vyspělost napovídá, že musela projít poměrně dlouhým vývojem, než dosáhla této úrovně. Naneštěstí se z doby před 14. st. př. n. l.

dochovalo pouze několik málo exemplářů, což nedovoluje provést rekonstrukci této fáze vývoje.“⁶ (Norman, 2012, s. 58)

Nápisy na bronzových nádobách se ve větším množství začaly objevovat zejména až za vlády Západních Zhouů (1046 – 771 př. n. l.) a v prvním období vlády Východních Zhouů, tj. v období Jar a podzimů (771 – 476 př. n. l.). Písmo této doby je ještě stále velmi podobné věštebným nápisům na želvích krunýřích a hovězích lopatkách, neboť si stále uchovává piktografickou podobu.

V této a zejména pak v následující fázi Válčících států (475 – 221 př. n. l.), kdy došlo k rozdělení říše na sedm samostatných států, se vývoj čínského písma začal s ohledem na politickou situaci diverzifikovat. Postupně se začalo objevovat velké pečetní písmo (dàzhuàn, 大篆) psané na hedvábí a bambusové destičky. Jak uvádí L. Zádrapa, definovat tento typ písma není snadné. Někteří badatelé za něj považují všechny druhy písma období Východních Zhouů, jiní zase písmo státu Qin coby předchůdce menšího pečetního písma. (Zádrapa a Pejčochová, 2009, s. 138-9) Podle J. Vochaly a J. Normana se jedná o styl písma Západních Zhouů, nazývaného též zhouwen (zhòuwén, 籀文), více viz (Vochala et al, 1975) a (Norman, 2012).
















V roce 221 př. n. l. se k moci dostal první svrchovaný císař Qin Shi Huangdi, který poprvé sjednotil Čínu. Za své vlády provedl řadu reforem, z nichž jedna se týkala i reformy zápisu jazyka. Jejím cílem byla simplifikace čínských znaků a jejich následné zavedení v celé zemi. Tím byla poprvé sjednocena nejen Čína, ale i písmo, kterému se dnes říká malé pečetní (xiǎozhuàn, 小篆). Souběžně s ním se také užívalo písmo úřednické (lìshū, 隶书), které bylo zjednodušenou obdobou velkého pečetního písma a které využívali nižší úředníci. Přestože se tento typ písma používal již za dynastie Qin, tradičně je spojován až s následující dynastií Han (206 př. n. l. – 220 n. l.), která ho povýšila na oficiální formu jazykového zápisu. Z tohoto důvodu se úřednické písmo dělí na dva druhy: starší úřední písmo (neboli qínské úřední písmo; qínlì, 秦隶) a nové úřední písmo (neboli hanské úřední písmo; hàn lì 汉隶). Na konci dynastie Han se zároveň utvářel další typ písma, písmo vzorové (kāishū, 楷书), které vycházelo ze stylu nového úředního písma a užívá se až

⁶ “The maturity of this early script has suggested to many scholars that it must have passed through a fairly long period of development before reaching this stage, but the few examples of writing which precede the fourteenth century are unfortunately too sparse to allow any sort of reconstruction of this development.” (Norman, 2012, s. 58)

dodnes. Pro názornou ukázkou uvádíme v Tabulce 2 několik vybraných znaků ilustrujících jednotlivé vývojové fáze čínského písma.

Tyto a další informace viz (Kučera et al, 2005), (Norman, 2012), (Qiu, 2000), (Vochala et al, 1975), (Vochala, Hrdličková, 1985) a (Zádrapa a Pejčochová, 2009).

Tabulka 2 Ukázka vývoje vybraných znaků

Význam	Typ znakového písma				
	Věšební nápisy 甲骨文	Nápisy na bronzích 金文	Pečetní písmo 小篆	Úřednické písmo 隶书	Vzorové písmo 楷书
Slon					
Děšť					
Srdce					

Zdroj: (Xiang de shufa, © 2004 – 2012), („Xiang“ zi de jiben xinxi, © 2004 – 2012), (Xin de shufa, © 2004 – 2012), („Xin“ zi de jiben xinxi, © 2004 – 2012), (Yu de shufa, © 2004 – 2012), („Yu“ zi de jiben xinxi, © 2004 – 2012); vlastní zpracování

„Od písma na věšebních destičkách až po vzorové písmo a jeho kursivní modifikace je převládající tendencí ve vývoji čínského znakového písma jeho grafická formalizace, která se projevuje ve zjednodušení tvaru znaků. Permanentním činitelem, který v tomto směru ovlivňoval vývoj znaků, byl požadavek snadnějšího psaní tohoto graficky příliš náročného písma.“ (Vochala et al, 1975, s. 26) Tendence zjednodušení je typická zejména pro neoficiální sféru, kde se tyto nestandardní zjednodušené znaky tzv. *suti* (sútǐ, 俗体) těšily velké oblibě. „Značné množství těchto znaků bylo vytvořeno a široce používáno běžným lidem pro psaní např. knih, zástavních lístků, lékařských předpisů, scénářů pro opery a

pro určité formy neoficiální literatury. Dokonce i literáti se uchýlovali k používání těchto neoficiálních a praktických forem znaků v osobní korespondenci a pro přepisování materiálů pro osobní použití.“⁷ (Norman, 2012, s. 80) Naopak v oficiální sféře bylo na tyto zjednodušené znaky nahlíženo jako na vulgární a zejména od Tangů až po začátek 20. století panoval konzervativní přístup, který se stavěl proti jakékoli inovaci, a proto se grafická podoba vzorového písma příliš neměnila. Více informací viz (Norman, 2012) a (Vochala et al, 1975).

Proces zjednodušování vyvrcholil v polovině 20. století, kdy se vládní orgány Čínské lidové republiky (ČLR) rozhodly zjednodušit tradiční čínské písmo a tím ho zpřístupnit co největšímu počtu lidí. V roce 1956 a 1964 proběhla reforma písma, která byla realizována ve dvou fázích. V první fázi bylo zjednodušeno více než 500 a v druhé fázi více než 2 000 znaků. Obě fáze reformy vycházely z přirozeného procesu zjednodušování znaků, který probíhal v předchozích staletích. Tyto a další informace viz (Chen, 1999), (Motalová et al, 2013), (Norman, 2012) a (Zádrapa a Pejčochová, 2009).

I přes prosazování zjednodušených znaků vládou a akademickým prostředím se tradiční sada znaků stále používá, a to zejména v publikacích zaměřených na dějiny jazyka a písma; v neoficiální sféře s ní lze přijít do kontaktu např. na různých nápisech. (Zádrapa a Pejčochová, 2009, s. 33-34) Na Taiwanu a v dalších oblastech, které nespádají pod správu ČLR (např. Hong Kong, Macao), stále přetrvává použití znaků tradičních. Další informace viz (Chen, 1999), (Motalová, 2013), (Norman, 2012) a (Zádrapa a Pejčochová, 2009).

Stejně jako i v předešlém případě vývoje čínského jazyka je třeba si stanovit, jakým typem písma se bude experiment zabývat. S ohledem na již výše zmíněný synchronní aspekt výzkumu bude předmětem našeho experimentu zjednodušená podoba vývojově nejmladší formy písma kaishu (vzorové písmo).

⁷ “... a large number of popular simplified characters were created and used widely among the common people for writing such things as account books, pawn tickets, medicinal prescriptions, operatic scripts and certain forms of vernacular literature. Even members of the literati employed these non-official but convenient forms in personal correspondence and for copying materials for private use; ...”

C. Jazykové styly

S ohledem na návaznost na předchozí výzkum (2012), který se zaměřil na kvantitativní analýzu novinového a literárního stylu, jsme se rozhodly podrobit analýze další dva jazykové styly – vědecký a umělecký styl. Hlavním důvodem tohoto rozhodnutí je možnost komparace nově získaných výsledků s výsledky dosavadními.

D. Aktuálnost

Jak již bylo zmíněno výše, experiment se soustředí na analýzu textů psaných v současné standardní čínštině. Aby texty odrážely současný jazyk, je nutné zvolit co nejaktuálnější výběrové soubory. S ohledem na tento požadavek jsme stanovily nejstarší akceptovatelnou dobu vydání na rok 2003 – texty nesmí být v době jejich výběru starší více než deset let.

E. Souvislost výběrových souborů

Jako páté kritérium jsme stanovily souvislost textu, na kterou upozorňuje L. Hřebíček. Ten ve své práci uvádí, že text, který má být podroben analýze, musí být souvislým útvarem. Za souvislý útvar L. Hřebíček považuje nepřerušenu sekvenci jazykových jednotek na všech různých jazykových úrovních, která má zřetelný začátek a konec a není ničím přerušena, např. obrázkem, grafem apod. (Hřebíček, 2002, s. 43)

F. Délka výběrových souborů

Dalším kritériem z hlediska reprezentativnosti výběrových souborů byla délka. Při jejím určování jsme musely brát ohled na to, aby výběrové soubory nebyly ani příliš krátké, ani příliš dlouhé. U krátkých textů se totiž nemusí vzájemné vztahy jazykových jednotek, formulované MALEM, projevit. V případě dlouhého textu se

naopak mohou vytrátit. Na základě námi již dříve provedených experimentů jsme rozmezí délky empiricky stanovily na 2 500 až 3 500 znaků.

G. Renomé autora či instituce

Naším posledním kritériem je autorovo renomé či prestiž instituce, pod kterou je práce publikována. Autor díky své oblíbenosti a vyhledávanosti širokými vrstvami působí na utváření současné podoby jazyka. Slova a slovní obraty, které při psaní textů volí, ovlivňují skladbu slovní zásoby čtenářů a tím i frekvencovanost slova znaků. Autorův slovník představuje jeden z faktorů utváření aktuální podoby čínského jazyka. U instituce je důležitá její uznávanost. Svou prestiží se dostává do širokého povědomí lidí a tím potenciálně zvyšuje čtenost prací, které pod svou záštitou publikuje.

2.2 Volba výběrových souborů

Na základě uvedených kritérií byly vybrány dva výběrové soubory reprezentující různé jazykové styly – vědecký a umělecký.

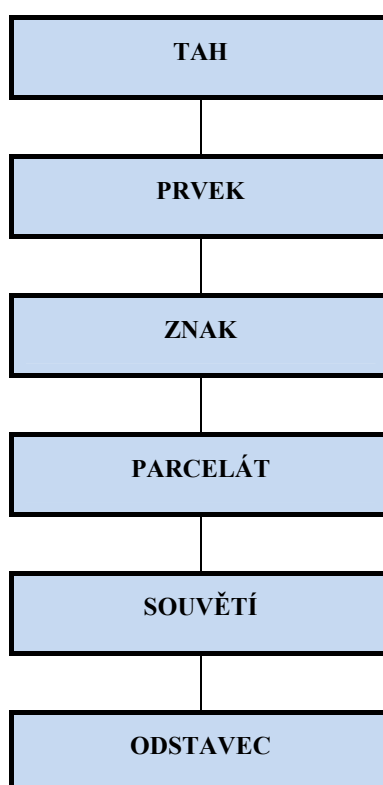
2.3 Stanovení a definování jazykových jednotek a úrovní

Jakmile jsme dokončily volbu výběrových souborů, mohly jsme přistoupit k dalšímu kroku – stanovení jazykových jednotek. Segmentace na základě přesně stanovených jednotek je zásadním krokem pro úspěšný experiment, proto musí být jednotky stanoveny precizně.

V případě tohoto experimentu jsme se při výběru jazykových jednotek primárně neopřely o běžné lingvistické definice, ale zvolily jsme alternativní přístup. Protože se naše experimenty týkají psaného jazyka, jako nejvhodnější alternativa se při stanovení jazykových jednotek nabízela aplikace grafického principu. Kromě tohoto hlavního kritéria, jsme u jazykových jednotek definovaných interpunkcí

musely rovněž přihlédnout k syntaktickému hledisku. V případě volby těchto principů jsme se rozhodly stejně jako v již provedeném výzkumu (2012) zvolit primárně grafický princip. Toto rozhodnutí vycházelo z požadavku komparace výsledků získaných z těchto experimentů s výsledky již výše zmíněného výzkumu. Aby výsledky mohly být porovnány, musí experimenty probíhat za identických podmínek.

Na základě těchto principů jsme stanovily šest jazykových jednotek, z nichž nejnižší představuje tah a nejvyšší odstavec (viz Obrázek 3).



Obrázek 3 Jazykové jednotky

Jednotlivým jazykovým jednotkám a jejich vymezením se budeme věnovat níže. Při popisu jazykových jednotek budeme postupovat od nejnižší jazykové jednotky k nejvyšší, tj. od tahu k odstavci.


Tah (bǐhuà, 笔画)

Nejmenší grafickou jednotkou čínského znaku je tah, který představuje nepřerušenu linii, tj. píše se najednou. Díla, která se touto jednotkou zabývají, se v mnoha případech rozcházejí nejen v klasifikaci a počtu tahů náležející do příslušných tříd, ale i v jejich terminologii. Různé přístupy se objevují jak v čínských, tak i v zahraničních publikacích. (Vochala, 1986, s. 14)

Úkolem této podkapitoly není podat vyčerpávající přehled o přístupech ke klasifikaci tahů, ale pouze představit základní podoby této jazykové jednotky a pro ilustraci uvést několik způsobů jejich třídění. Druhou kategorií – odvozenými tahy – se v této práci zabývat nebudeme, jelikož se většinou jedná pouze o různé modifikace tahů základních. (Zádrapa a Pejčochová, 2009, s. 67)

V kaligrafických příručkách bývá často rozlišováno osm základních tahů. Tradičně bývají ilustrovány na čínském znaku znamenajícím „věčnost“ (yǒng, 永), který je z těchto osmi základních tahů složen, viz Tabulka 3. Tímto způsobem jsou rozlišeny základní tahy např. v knihách *Říše čínských znaků* od Cecilie Lindqvist (2010), *Čínská kaligrafie* od Qu Leilei (2005) nebo v *Chinese Running Script Calligraphy for Beginners* od Wang Xianchuna (2007).

Tabulka 3 Ukázka klasifikace základních tahů podle kaligrafických příruček na čínském znaku 永

Znak	Kategorie tahů zastoupené v tomto znaku jsou:
	1 Tečka: Cche nebo Tien
	2 Vodorovný tah: Lej nebo Cheng
	3 Svislý tah: Nu nebo Kou
	4 Háček: Tchi nebo Kou
	5 Tah vzhůru: Cche nebo Tchiao
	6 Tah stranou: Lue nebo Pchie
	7 Tah vzhůru: Cche nebo Tchiao
	8 Tah dolů doprava: Če nebo Na

Zdroj: Qu, 2005, s. 33; vlastní zpracování

Co se týče výukových materiálů, rozdělení základních tahů není jednotné. Např. v *Učebnici čínských znaků* od Ondřeje Kučery a kol. (2005) je rozlišeno pět základních tahů, ostatní typy jsou označeny pouze jako jejich modifikace, a proto jsou zařazeny mezi tahy odvozené. Oproti tomu v učebnici *Integrated Chinese* od Liu Yuehua a kol. (2009) se uvádí, že existuje 11 základních tahů, odvozené tahy zde uvedené nejsou.

Materiály zabývající se obecně čínským jazykem se v počtu základních tahů rovněž rozcházejí. Například v publikacích *Čínské písmo* od Lukáše Zádrapy a Michaely Pejčochové (2009) a *Úvod do hovorové čínštiny I* od Oldřicha Švarného (1967) autoři rozlišují šest základních tahů. Wang Ning (2002) v knize *Studie o utváření čínských znaků* (Hànzi gòuxíng xué jiǎngzuò, 汉字构形学讲座) zmiňuje základních tahů sedm. Naproti tomu Leon Wieger (1965) ve své práci *Chinese Characters: Their origin, etymology, history, classification and signification* uvádí devět základních tahů.

Stejně jako v minulém experimentu (2012) představíme základní tahy podle klasifikace navržené Jaromírem Vochalou (1986). Tento český sinolog vymezuje ve své práci *Chinese Writing System* 11 základních tahů, které navíc rozděluje do dvou subkategorií – jednoduché tahy a jednoduché tahy s hákem (viz Tabulka 4 a 5).

Tabulka 4 Klasifikace znaků: jednoduché tahy

Tah	Pinyin	Znaky	Český ekvivalent
一	héng	横	Vodorovný tah
丨	shù	竖	Svislý tah
㇇	piě	撇	Levý tah
㇇	nà	捺	Pravý tah
㇇	tí	提	Stoupavý tah
丶	diǎn	点	Levý bodový tah
丶	diǎn	点	Pravý bodový tah

Zdroj: (Vochala, 1986, s. 30), (Bihua, © 2013), vlastní zpracování

Tabulka 5 Klasifikace znaků: jednoduché tahy s hákem

Tah	Pinyin	Znaky	Český ekvivalent
→	hénggōu	横钩	Vodorovný tah s hákem
↓	shùgōu	竖钩	Svislý tah s hákem
㇇	wāngōu	弯钩	Svislý zahnutý tah s hákem
㇏	xiéngōu	斜钩	Pravý tah s hákem

Zdroj: (Vochala, 1986, s. 30), (Bihua, © 2013), vlastní zpracování

Další typy tahů a jejich modifikace viz (Vochala, 1986).

Tah nebo kombinace tahů tvoří další jazykovou jednotku – *prvek*.

Prvek (bùjiàn, 部件)

Prvek představuje jednotku, která je tvořena určitým počtem tahů a která se podílí na vytváření struktury znaku. O přesné a jednotné definici prvku však mluvit nemůžeme. Ačkoliv existují různé přístupy, které prvek definují, jejich aplikovatelnost je obtížná právě z důvodu jejich nepřesnosti. Obecně řečeno se za prvek považuje strukturální jednotka, která je vyšší než tah, ale nižší než znak. Precizní determinace prvků ale často naráží na problém přesně vymezit, které kombinace tahů vytváří prvek a které nikoliv. Nehledě na fakt, že se při segmentaci znaků na prvky setkáváme s řadou výjimek – obecné tvrzení, že prvek je vyšší než tah, vyvrací existence znaků tvořených jedním tahem. Stejně tak prvky, které mohou vystupovat jako samostatné znaky, poukazují na neúplnost druhé části obecné definice, která tvrdí, že prvek je nižší jednotkou než znak.

Jak již bylo řečeno výše, přístupy k determinaci prvku se různí. Např. v čínské encyklopedii Baidu Baike jsou zmíněna následující kritéria (Hanzi bujian, © 2013):

1. První kritérium se opírá o počet tahů. V tomto případě lze prvky rozdělit do dvou kategorií – první kategorii představují prvky tvořené jedním tahem (např. 一 a 乙), druhou prvky složené ze dvou a více tahů (např. 士 a 重).
2. Druhé kritérium kategorizuje prvky do dvou skupin na základě jejich samostatnosti. Do první z nich náleží ty prvky, které vytváří znak kombinací s jinými prvky a zároveň jsou samostatným znakem (např. 吉→口 a 和→可). Druhá skupina zahrnuje prvky, které jsou pouze součástí znaků, jako samostatné znaky nevystupují (např. 同→冂 a 病→疒).
3. Třetí kritérium dělí prvky do dvou skupin na základě jejich dekompozice. Do první náleží základní prvky, které představují nejmenší a dále nedělitelné součásti znaku (např. 男→田 a 力). Naopak v druhé skupině jsou prvky složené, tj. prvky, které lze zároveň rozložit na dva a více menších prvků (např. 想→相 a 心, které zde tvoří první „vrstvu“ a 相→木 a 目, které tvoří „vrstvu“ druhou).

S ohledem na diferenciaci přístupů k prvku jsme se rozhodly pro naše účely vybrat takovou metodu segmentace znaku na prvky, která bude výhradně respektovat formalizační a grafický aspekt. Z tohoto důvodu jsme tuto jazykovou jednotku vymezily na základě doteků tahů. Prvek v našem případě představuje tzv. ostrov, tedy tu část znaku, která je složena z jednoho tahu či skupiny tahů navzájem se dotýkajících či protínajících a která je zároveň zřetelně oddělaná od ostatních částí (= prvků) znaku. Kombinace těchto ostrovů vytváří vyšší jazykovou jednotku – *znak*.

Aplikace této koncepce však v předchozím výzkumu (2012) odhalila, že různé fonty v případě identického znaku vytváří rozdílné hranice prvků. Celkový počet prvků tak variuje v závislosti na příslušných fontech. V Tabulce 6 uvádíme pro ilustraci příklady těch znaků, u nichž počet prvků fluktuuje nejvíce.

Tabulka je rozdělená do pěti sloupců. První z nich obsahuje názvy osmi zvolených fontů, které mají dané kolísání počtu prvků demonstrovat. Do každého ze zbývajících sloupců je vložen znak, jehož počet prvků se mění s fontem, do kterého je převeden. První řádek obsahuje příklady vybraných znaků s červeným označením těch prvků, jejichž hranice se různí. Zbylé řádky uvádí vybrané znaky v příslušných fontech spolu s počtem prvků (P_p) v dolní části buňky.

Tabulka 6 Komparace znaků demonstrující oscilaci počtu prvků (P_p) v závislosti na příslušném fontu

Fonty		Čínské znaky			
		翻	各	麻	新
1.	Simsun	翻	各	麻	新
		P_p 6	P_p 1	P_p 3	P_p 5
2.	DF Kai-SB	翻	各	麻	新
		P_p 10	P_p 2	P_p 6	P_p 4
3.	Han ding jiankaiti (汉鼎简楷体)	翻	各	麻	新
		P_p 10	P_p 2	P_p 4	P_p 7
4.	Mingliu	翻	各	麻	新
		P_p 8	P_p 1	P_p 7	P_p 4
5.	Fangsong	翻	各	麻	新
		P_p 5	P_p 2	P_p 3	P_p 6
6.	Meiryo	翻	各	麻	新
		P_p 6	P_p 1	P_p 3	P_p 2
7.	Jhenghei	翻	各	麻	新
		P_p 9	P_p 2	P_p 7	P_p 4
8.	SimHei	翻	各	麻	新
		P_p 6	P_p 2	P_p 4	P_p 6

Zdroj: (Motalová et al, 2013)

S ohledem na tento fakt jsme se rozhodly vybrat pouze jeden font, který bude aplikován na oba výběrové soubory. V případě první fáze experimentu bylo klíčovým kritériem pro tuto volbu zachování zdrojového fontu, ve kterém byly texty napsány, tj. SimSunu. V druhé fázi se primárně jednalo o dodržení stejného segmentačního přístupu za účelem komparace výsledků obou fází výzkumů. U výběrových souborů však nebylo nutné čínské znaky převádět do námi požadovaného fontu, neboť pro oba texty byl použit stejný zdrojový font jako u novinového článku a povídky, tedy SimSun.

Znak (hànzì, 汉字)

Znak je základní jednotkou čínského znakového písma, která odpovídá v mluveném projevu jedné slabice⁸. Každý znak se zapisuje do jednoho a vždy stejně velkého grafického pole bez ohledu na počet tahů. Grafické pole má tvar čtverce nebo obdélníku, jehož výška není o moc větší než šířka. (Švarný, 1967, s. 31) Jednotlivá grafická pole k sobě těsně přiléhají a nejsou tedy oddělena mezerami. Proto nemůžeme v čínských textech identifikovat ortografické slovo.

Kromě čínských znaků mohou také čínské texty obsahovat arabské číslice, které lze zapsat ve dvojím formátu. V případě prvního každá arabská číslice zabírá jedno grafické pole, a proto je považována za jednotku rovnu znaku. Pokud je však arabská číslovka naformátována druhým způsobem, jednotlivé číslice k sobě těsně přiléhají, a proto se za jednotku ekvivalentní znaku považuje buď samostatně v textu stojící číslice, nebo skupina číslic, tedy jak celé číslo, tak i racionální.

Skupina znaků vytváří další jazykovou jednotku – *parcelát*.

Parcelát

Za parcelát považujeme část souvětí, kterou vymezují vybraná interpunkční znaménka. Tuto jednotku jsme stanovily již v předešlém výzkumu, v němž bylo

⁸ V čínských textech existuje pouze jedna výjimka, kdy dva znaky reprezentují jednu slabiku. Jedná se o případ erizace, např. slovo “chvíle”(huǐr, 会儿).

nutné najít jazykovou jednotku vyšší než znak a nižší než souvětí. Současné čínské texty jsou po grafické stránce rozděleny interpunkcí na určité segmenty. Proto jsme se rozhodly opřít o interpunkční znaménka, na jejichž základě budou stanoveny hranice této jednotky. Pro účely první fáze experimentu (2012) jsme tuto jazykovou jednotku nazvaly parcelát.

S ohledem na variabilitu čínských interpunkčních znamének a rozdílnost jejich funkcí bylo nutné určit, která z nich budou pro hranice parcelátů platná (viz Tabulka 7). Tímto krokem jsme vytvořily hierarchizaci interpunkčních znamének a kromě grafického hlediska tak přihlédly i k hledisku syntaktickému.

Tabulka 7 Vybraná interpunkční znaménka, platná pro hranice parcelátů

Interpunkční znaménko	Pinyin	Znaky	Český ekvivalent
。	jùhào	句号	Tečka
?	wèn hào	问号	Otazník
!	tàn hào	叹号	Vykřičník
,	dòu hào	逗号	Čárka
;	fēn hào	分号	Středník
:	mào hào	冒号	Dvojtečka

Zdroj: čínská terminologie převzata z (Biaodian fuhao, © 2013), vlastní zpracování

Kromě výše uvedených znamének se v čínském textu také vyskytují interpunkční znaménka se specifickou funkcí. Protože se jejich použitím nenaruší syntaktická struktura, tj. S-V-O (podmět – přísudek – předmět), nezohlednily jsme tato interpunkční znaménka jako hranice parcelátů. Jedná se o:

1. Čárka pro výčet 、

Tento typ čárky se používá mezi souřadně spojenými větnými členy nebo se píše za znakem pro číslo v číselném seznamu. (GB/T 15834 – 2011, 2012) a (Dunhao, © 2013)

2. Uvozovky “ ” [] , ‘ []

Do uvozovek se vkládají přímé citace, přímé řeči a určité části textu, na které chce autor upozornit. Dělí se dvou kategorií: dvojitě uvozovky a jednoduché

uvozovky, které mají jak evropskou, tak i čínskou alternativu zápisu.(GB/T 15834 – 2011, 2012) a (Yinhao, © 2013)

3. Elipsa

Značí vynechání určité části v citované pasáži, přerušovanou promluvu nebo také může zastupovat vynechané položky v seznamu. (GB/T 15834 – 2011, 2012) a (Shengluohao, © 2013)

4. Zdůrazňující tečka ·

Upozorňuje na důležité části textu (znak, slovo či větu), kterým by měl čtenář věnovat zvláštní pozornost. Tečka bývá zpravidla umístěna pod textem. (GB/T 15834 – 2011, 2012) a (Zhuozhonghao, © 2013)

5. Závorky pro tituly 《》 〈〉

Do tohoto typu závorek se vkládají různé názvy a jména, jako jsou např. názvy knih, periodik, článků, dokumentů a dalších literárních děl, dále také názvy divadelních her, písní, obrazů, filmů a televizních pořadů. I tato interpunkční znaménka mají dvojitou a jednoduchou variantu.(GB/T 15834 – 2011, 2012) a (Shuminghao, © 2013)

6. Oddělovací tečka ·

Odděluje části slovních spojení, které jsou ve vzájemném vztahu. Vkládá se např. mezi vlastní jména osob či názvy knih, které nejsou čínského původu. Vymezují tak jednotlivé entity daných jmen. (GB/T 15834 – 2011, 2012) a (Jianghao, © 2013)

7. Pomlčka — —— - ~

Vyjadřuje spojení vzájemně souvisejících částí. Stejně jako v západních jazycích existují dvě varianty této pomlčky – krátká a dlouhá. (GB/T 15834 – 2011, 2012) a (Lianjiehao, © 2013)

8. Podtržítka vlastních jmen _____

Vyznačuje speciální kategorii vlastních jmen, která se objevují v historických dílech. Používá se k označení osobních jmen, místních názvů, názvů dynastií, jmen etnických skupin, států a institucí. (GB/T 15834 – 2011, 2012) a (Zhuanminghao, © 2013)

Tabulka 8 Vybraná interpunkční znaménka, neplatná pro hranice parcelátů

Interpunkční znaménko	Pinyin	Znaky	Český ekvivalent
、	dùnhào	顿号	Čárka pro výčet
“ ” 『 』	yǐnhào	引号	Uvozovky
“ ” 「 」			
……	shěnglüèhào shānjié hào	省略号 删节号	Elipsa
.	zhúzhòng hào	着重号	Zdůrazňující tečka
《 》	shūmíng hào	书名号	Závorky pro tituly
〈 〉			
.	jiàngé hào	间隔号	Oddělovací tečka
—~	liánjié hào	连接号	Pomlčka
——	zhuānmíng hào sīmíng hào	专名号 私名号	Podtržítka vlastních jmen

Zdroj: čínská terminologie převzata z (Biaodian fuhao, © 2013), vlastní zpracování

Určitými přechodnými interpunkčními znaménky mezi těmito dvěma skupinami jsou závorky () [] { } (kuò hào, 括号) a zvláštní typ pomlčky — (pòzhé hào, 破折号), u kterých se s ohledem na syntax muselo určit, v jakých případech budou považována za hranice parcelátu a v jakých nikoli.

1. Závorky () [] { }

Obsahují poznámku vloženou přímo do textu. Dodatečné informace k tématu, které se do nich vkládají, mohou být vyjádřeny jak větným členem, tak i celou větou. (GB/T 15834 – 2011, 2012) a (Kuohao, © 2013) Pokud se jedná o první případ, pak závorky za hranici parcelátu považovány nejsou. Zatímco v druhém případě hranice parcelátu tvoří.

2. Pomlčka —

Používá se pro vysvětlení a doplnění výkladu, dále značí změnu konverzačního tématu, významu či způsobu vyjádření, užívá se rovněž za onomatopoiemi na znamení pokračování daného zvuku či hlasu, nebo také slouží jako odrážka před každou položkou v seznamu. Za touto pomlčkou může následovat jak

větný člen, tak i celá věta. (GB/T 15834 – 2011, 2012) a (Pozhehao, © 2013)
Z tohoto důvodu jsme pro tuto pomlčku jakožto hranici parcelátu stanovily následující pravidla:

Toto interpunkční znaménko nezohledňujeme jako hranici parcelátu v následujících případech:

- a) U vysvětlení a doplnění výkladu formou větného členu nebo jejich výčtu.
- b) Za onomatopoiemi na znamení pokračování daného zvuku či hlasu.
- c) Před každou položkou v seznamu, která je vyjádřena jedním větným členem.

Toto interpunkční znaménko zohledňujeme jako hranici parcelátu v následujících případech:

- a) U vysvětlení a doplnění výkladu formou věty, tj. věty, která má podmět, přísudek a předmět.
- b) Při změně konverzačního tématu, významu či způsobu vyjádření.
- c) Před každou položkou v seznamu, která je vyjádřena větou.

V čínských textech se také mohou vyskytovat znaménka západní interpunkce, a to v těch případech, kdy autor pro zápis čísel používá na místo čínských znaků arabské číslice. Znaménka, která tato čísla doprovází (např. desetinná tečka), nevytváří hranice žádné jazykové jednotky.

Parcelát nebo skupina parcelátů vytváří další jazykovou jednotku – *souvětí*.

Souvětí (fùjù, 复句)

Hranice této jazykové jednotky stejně jako v předchozím případě tvoří interpunkční znaménka. Na rozdíl od parcelátu se však jedná pouze o tečku 。 (jùhào, 句号), otazník ? (wèn hào, 问号) a vykřičník ! (gǎntànhào, 感叹号). Ostatní interpunkční znaménka jsou platná pouze pro nižší jazykovou jednotku – parcelát. Tímto výběrem jsme opět zohlednily syntaktické hledisko.

Souvětí nebo skupina souvětí vytváří další jazykovou jednotku – *odstavec*.

Odstavec (duànluò, 段落)

Vymezení této jazykové jednotky vychází čistě z grafického principu. Každý odstavec začíná na novém řádku odsazeném od okraje a zároveň je oddělený od ostatních odstavců vloženým prázdným řádkem.

Jazykové úrovně

Jak uvádí L. Hřebíček „dvě jazykové jednotky různého druhu mohou být uvažovány jako vztah dvou vzájemně hierarchizovaných úrovní (...), čili mohou být uvažovány jako konstrukt a konstituent. Když však přemýšlíme o textu, je omezení na vztah právě dvou úrovní nedostatečné; okamžitě se ptáme, co se v textu děje s ostatními úrovněmi. ... Jinak řečeno, každá jazyková entita vůči všem vyšším jazykovým úrovním je konstituentem a vůči všem nižším úrovním je konstruktem.“ (Hřebíček, 2002, s. 59)

Spojením výše uvedených, v hierarchii bezprostředně sousedících jazykových jednotek do vzájemných vztahů jsme získaly čtyři jazykové úrovně. Nejnižší jazykovou úroveň U4 reprezentuje znak (měřený v prvcích), který je na této úrovni konstruktem, a prvek (měřený v průměrném počtu tahů), jenž je jeho konstituentem. Na jazykové úrovni U3 je konstruktem parcelát (měřený ve znacích) a konstituentem znak (měřený v průměrném počtu prvků). Další jazykovou úroveň U2 tvoří souvětí (měřené v parcelátech) coby konstrukt a parcelát (měřený v průměrném počtu znaků) coby konstituent. Nejvyšší jazyková úroveň U1 je reprezentovaná odstavcem (měřeným v souvětích), který je zde konstruktem, a souvětím (měřeným v průměrném počtu parcelátů), jenž je jeho konstituentem. Pro přehlednost uvádíme tyto jazykové úrovně v níže uvedené Tabulce 9:

Tabulka 9 Jazykové úrovně U_i , x_i konstrukt, y_i konstituent ($i = 1, 2, 3, 4$)

Jazyková úroveň	Konstrukt x_i ; konstituent y_i		Délka
U4	x_4	znak	v prvcích
	y_4	prvek	v průměrném počtu tahů
U3	x_3	parcelát	ve znacích
	y_3	znak	v průměrném počtu prvků
U2	x_2	souvětí	v parcelátech
	y_2	parcelát	v průměrném počtu znaků
U1	x_1	odstavec	v souvětích
	y_1	souvětí	v průměrném počtu parcelátů

2.4 Segmentace a kvantifikace výběrových souborů

Stanovení jazykových jednotek a jazykových úrovní umožnilo přistoupit k dalšímu kroku – segmentaci výběrových souborů.

Po dokončení segmentační fáze jsme následně provedly kvantifikaci textů. Texty bylo nutné převést do řeči čísel, abychom z nich mohly získat tabulky pro všechny jazykové úrovně, které budou obsahovat délky daných veličin, tj. délku konstruktů x_i , jejich frekvence z_i a délku konstituentů y_i . Segmentaci a následnou kvantifikaci jsme zpracovávaly v programu Microsoft Excel.

2.5 Testování spolehlivosti modelu pomocí statistických metod

Pro výpočet parametrů A a b potřebných pro jednoduchou verzi MALu jsme použily statistický program R⁹. Pomocí metody lineární regrese¹⁰ byly v tomto

⁹ Software R je programovacím jazykem a prostředím pro statistickou a grafickou analýzu dat. (Whatis R?: Introduction to R, 2013)

programu zpracovány v Excelu vytvořené tabulky, z nichž jsme získaly hodnoty parametrů pro každou jazykovou úroveň. Pro ověření spolehlivosti modelu pro jednotlivé jazykové úrovně bylo rovněž nutné vyextrahovat pomocí tohoto softwaru koeficient determinace R^2 , jehož hodnota nám určuje míru této spolehlivosti. „Interval, ve kterém se koeficient determinace může pohybovat, je $0 \leq R^2 \leq 1$. Čím blíže koeficient k 1, tím lépe model sedí, Hodnoty R^2 větší nebo rovny 0,7 mohou prokazovat adekvátní a dobře sedící model v kvantitativní lingvistice.“ (Benešová, 2011, s. 77) Hodnoty koeficientu determinace jsou v této práci uváděny v procentech.

Pozorování s nízkou četností výskytu (tzv. extrém), která jsou v porovnání s ostatními zanedbatelná, byla statisticky ošetřena metodou vynechání. Hodnoty získané jejich odebráním jsou vždy uvedeny v poznámce pod čarou vztahující se k tabulce uvádějící parametry A , parametr b a koeficient determinace R^2 .

2.6 Interpretace získaných dat

Poslední krok této kvantitativní analýzy představoval interpretování získaných výsledků a jejich grafických vizualizací. Tomuto kroku je věnovaná následující interpretační část, která bude rozdělena do dvou sekcí. Autorkou první z nich bude Tereza Motalová, která se zaměří na interpretaci dat získaných kvantitativní analýzou vědeckého textu. Druhá část, jejíž autorkou je Lenka Spáčilová, se bude zabývat interpretací výsledků získaných kvantitativní analýzou textu uměleckého.

¹⁰ „Vztah mezi závislou a nezávislou proměnnou je také možno vyjádřit graficky pomocí regresní přímky. Regresní přímka je pak znázorněním lineárního trendu v datech. ... Regresní přímka je sestrojena způsobem, který zajišťuje, aby druhé mocniny vzdáleností všech datových bodů od ní byly menší než od jakékoliv jiné přímky. ... Při sestrojování regresní přímky se proto mluví o metodě nejmenších čtverců, nebo přesněji o metodě nejmenší sumy čtverců. ... Odchylky datových bodů od regresní přímky se měří svisle, tj. paralelně s osou závislé proměnné. Říká se jim rezidua a je zřejmé, že čím menší jsou rezidua, tím přesnější predikci hodnot bude přímka poskytovat. Těsnosti, s jakou se datové body k přímce přimykají, se říká kvalita proložení. ... Budou-li všechny body ležet na regresní přímce, pak od ní nebudou mít žádné odchylky“ (Volín, 2007, s. 208-209)

KVANTITATIVNÍ ANALÝZA SOUČASNÉ PSANÉ ČIŠTINY

Aplikace Menzerath-Altmannova zákona
na vědecký styl

Autor: Tereza Motalová

3 Aplikace Menzerath-Altmanova zákona na vědecký článek

Cílem této kapitoly je provést kvantitativní analýzu výběrového souboru reprezentujícího vědecký styl. Na text bude aplikován Menzerath-Altmanův zákon, jehož pomocí se bude testovat existence vzájemných vztahů jazykových jednotek na příslušných úrovních.

Tato část je rozdělena do šesti podkapitol, které postupně z důvodu ucelenosti a přehlednosti představí jednotlivé kroky experimentu. Důraz bude kladen zejména na ty kroky, které se v souvislosti s vědeckým textem vyznačují určitými specifiky. Ostatním krokům, uváděným pouze z důvodu zachování posloupnosti postupu experimentu, bude věnována jen zmínka s odkazem na podkapitolu výše, ve které jsou podrobně rozepsány. Největší prostor bude věnován poslední podkapitole zabývající se interpretací získaných dat, neboť představuje gró této části práce. Data, která budou získána analýzou každé jazykové hladiny, budou jednotlivě analyzována a interpretována v dílčích podkapitolách. Výsledná zjištění budou sumarizována ve shrnutí uvedeném v závěru této části.

3.1 Stanovení kritérií pro volbu výběrového souboru a jejich odůvodnění

Před zahájením samotné kvantitativní analýzy a aplikace Menzerath-Altmanova zákona bylo nejprve nutné stanovit kritéria pro volbu výběrového souboru. Jak jsme již se spoluautorkou Lenkou Spáčilovou uvedly v kapitole pojednávající o metodologii výzkumu (viz podkapitola 2.1), aby určitý výběrový soubor mohl být vybrán, musí nejprve splnit několik námi stanovených podmínek.

Fundamentem volby výběrového souboru bylo stanovení jazykového stylu, v němž měl být daný vzorek napsán. Platnost (příp. neplatnost) vzájemných vztahů jazykových jednotek na určitých jazykových úrovních jsem se rozhodla testovat na vědeckém článku. Získaná data tak doplní výsledky předchozího výzkumu (2012), jenž analyzoval novinový a literární styl, a výsledky experimentu spoluautorky Lenky Spáčilové, která testuje styl umělecký. Aby byla v co největší možné míře zaručena vědeckost, článek musela publikovat významná vědecká instituce. Dále se

nesměl opomenout požadavek textu psaného ve zjednodušených znacích současnou standardní čínštinou a zároveň vydaného od roku 2003 včetně. V případě délky bylo nevyhnutelné přihlédnout k charakteru vědeckého stylu, jehož články se obecně vyznačují větší délkou, a horní hranici posunout o 2 000 znaků. Vybraný vzorek tedy nesměl přesáhnout počet 5 500 znaků.

3.2 Volba výběrového souboru

Všechna tato kritéria splnil vědecký článek s názvem *Ekonomický aspekt jazyka* (*Rènshì yǔyán de jīngjìxué shǔxìng*, 认识语言的经济属性). Článek byl publikován roku 2012 v akademickém periodiku *Aplikovaná lingvistika*¹¹ (*Yǔyán wénzì yìngyòng*, 语言文字应用) a roku 2013 zveřejněn na oficiálních webových stránkách *Čínské akademie společenských věd*¹² (*Zhōngguó shèhuì kēxuéyuàn*, 中国社会科学院). Jeho autorem je Li Yuming¹³ (Lǐ Yǔmíng, 李宇

¹¹ **Aplikovaná lingvistika** je pro lingvistiku a čínský jazyk klíčovým akademickým periodikem vydávaným od roku 1992 Institutem aplikované lingvistiky při Ministerstvu školství („Yuyan wenzi yingyong“ Bianjibu (Yingyong yuyanxue yanjiu zhongxin), 2005).

¹² **Čínská akademie společenských věd** (ČASV) je nejvyšší akademickou institucí a výzkumným centrem v Čínské lidové republice, které se zaměřuje na výzkum v oblasti filozofie a společenských věd. ČASV vznikla v květnu roku 1977 s cílem podpořit rozvoj filozofie a společenských věd. Tehdy měla tato výzkumná instituce 14 výzkumných ústavů a zaměstnávala více než 2 200 lidí. V současné době ČASV operuje s šesti odděleními, mezi něž patří *Oddělení pro literaturu a filozofii* (Wénzhé xuébù, 文哲学部), *Oddělení pro historii* (Lìshǐ xuébù, 历史学部), *Oddělení pro ekonomii* (Jīngjì xuébù, 经济学部), *Oddělení pro společenské vědy, politiku a právo* (Shèhuì zhèngfǎ xuébù, 社会政法学部), *Oddělení pro mezinárodní výzkum* (Guójì yánjiū xuébù, 国际研究学部) a *Oddělení pro studium marxismu* (Mǎkèsīzhǔyì yánjiū xuébù, 马克思主义研究学部). Tato oddělení dohromady pod sebou sdružují 37 výzkumných ústavů a 45 výzkumných center, které provádí výzkum v oblasti přibližně 120 vědních disciplín a zaměstnávají více než 4 200 lidí. Post prezidenta ČASV v současnosti vykonává Chen Kuiyuan (Chén Kuíyuán, 陈奎元). V roce 2011 se ČASV stala předním „think tankem“ v celé Asii, mezi 30 institucemi zaujala první místo. Tyto a další informace viz (Wo yuan gaikuang, 2010), (Yuan jigou, 2011), (Zheng, 2012) a (Zhongguo shehui kexueyuan, © 2013).

¹³ **Li Yuming** se narodil v roce 1955 v provincii Henan v Miyangu. Roku 1981 úspěšně ukončil studium na Katedře čínského jazyka na Zhengzhouské univerzitě a následně nastoupil na magisterský program oboru Moderní čínština na Central China Normal University, kde po absolutoriu v roce 1984 začal také učit a v roce 1993 obhájil titul profesora. O tři roky později byl jmenován děkanem fakulty humanitních věd a v roce 1998 dokonce zástupcem rektora. Od roku 2001 zastával různé funkce v institutech při Ministerstvu školství (např. Institut aplikované lingvistiky). V současné době vykonává pozici tajemníka Pekingské univerzity jazyka a kultury. Li Yuming se specializuje na oblast teorie lingvistiky, moderní čínštiny, psycholingvistiky, jazykového plánování a zejména pak na gramatiku, dětskou řeč, obnovu řeči u hluchých dětí ad. Co se týče Liovy publikační činnosti, vydal několik knih a přes tři sta článků a zároveň se stal šéfredaktorem časopisu Aplikovaná lingvistika. Za svůj přínos získal mnohá ocenění. Tyto a další informace viz (Li Yuming, © 2013) a (Beiyu jiaoshou: Li Yuming jiaoshou, © 2006).

明), profesor *Univerzity jazyka a kultury v Pekingu* (*Běijīng yǔyán dàxué*, 北京语言大学). Článek je napsaný ve zjednodušených znacích současnou moderní čínštinou a obsahuje 5 155 znaků (vlastní stať). Zvolený výběrový soubor je se všemi svými náležitostmi zpřístupněn v přílohách této diplomové práce, kde je označen jako vzorek A (viz Příloha 1).

3.3 Stanovení a definování jazykových jednotek a úrovní

Před samotným testováním existence vzájemných vztahů jazykových jednotek na určitých jazykových úrovních bylo nejprve nutné vymezit, s jakými jednotkami bude experiment po celou dobu operovat. V našem případě se jednalo o šest jazykových jednotek, které byly precizně stanoveny již v první fázi výzkumu (2012). Protože jsou jejich jednotlivé definice podrobně popsány v podkapitole 2.3, uveďme si nyní pouze jejich výčet:

TAH – PRVEK – ZNAK – PARCELÁT – SOUVĚTÍ – Odstavec.

Určením vzájemných vztahů mezi dvěma bezprostředně ležícími jazykovými jednotkami jsme získaly čtyři jazykové úrovně (viz Tabulka 10). Na každé z nich budeme testovat platnost (příp. neplatnost) vzájemných vztahů těch jednotek, kterými je daná hladina právě tvořena.

Jednotlivé jazykové úrovně jsou značeny indexem i , $i = 1$ značí hladinu odstavec – souvětí, $i = 2$ souvětí – parcelát, $i = 3$ parcelát – znak a $i = 4$ znak – prvek.

Tabulka 10 Jazykové úrovně U_i , x_i konstrukt, y_i konstituent ($i = 1, 2, 3, 4$)

JAZYKOVÁ ÚROVEŇ U4		JAZYKOVÁ ÚROVEŇ U2	
x_4	znak měřený v prvcích	x_2	souvětí měřené v parcelátech
y_4	prvek měřený v průměrném počtu tahů	y_2	parcelát měřený v průměrném počtu znaků

JAZYKOVÁ ÚROVEŇ U3		JAZYKOVÁ ÚROVEŇ U1	
x_3	parcelát měřený ve znacích	x_1	odstavec měřený v souvětích
y_3	znak měřený v průměrném počtu prvků	y_1	souvětí měřené v průměrném počtu parcelátů

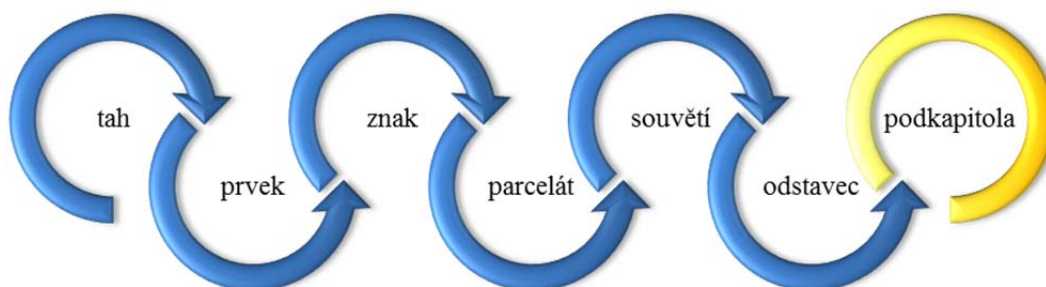
3.4 Segmentace a kvantifikace výběrového souboru

Po stanovení a definování jazykových jednotek a jazykových úrovní následoval krok další – segmentace výběrového souboru.

Nejprve však několik slov ke struktuře vědeckého článku. Článek je rozdělen do několika částí. Do prvního oddílu náleží nadpis článku *Ekonomický aspekt jazyka* (*Rènshì yǔyán de jīngjìxué shǔxìng*, 认识语言的经济属性), jméno autora Li Yuming (Lǐ Yǔmíng, 李宇明), autorovo stručné představení (*zuòzhě jiǎnjiè*, 作者简介), datum publikování a zdroj. Poté následuje část s bilingvním abstraktem (*nèiróng tíyào*, 内容提要) a klíčovými slovy (*guānjiàncí*, 关键词). Třetí oddíl začíná vlastní statí vědeckého článku, která je podnadpisy rozčleněna na čtyři „podkapitoly“. V závěru článku autor uvádí vysvětlivky (*zhùshì*, 注释) a bibliografické údaje použitých zdrojů (*cānkǎo wénxiàn*, 参考文献).

Ověřování platnosti (příp. neplatnosti) vzájemných vztahů jazykových jednotek bylo třeba testovat na souvislém a uceleném textu. Z tohoto důvodu jsem segmentaci a následnou kvantitativní analýzu provedla pouze na vlastní statí vědeckého článku; ostatní náležitosti textu (nadpis, úvodní informace, dvojjazyčný abstrakt, klíčová slova, vysvětlivky a odkazy na literaturu) jsem analýze nepodrobila, neboť se jedná o samostatné ucelené segmenty, které na sebe obsahově nenavazují.

Co se týče členění statě prostřednictvím podnadpisů, pak ani tyto nebyly zohledněny. Jejich používání ve struktuře textu by však nemělo být opomenuto. Segmenty textu vymezené podnadpisy by mohly představovat další jazykovou jednotku, zde pouze pracovní nazvanou jako „podkapitola“ či „podnadpisí“, viz Obrázek 4.



Obrázek 4 Přehled jazykových jednotek I

Spojením této jednotky a jednotky bezprostředně nižší by vznikla další jazyková úroveň – podkapitola (měřená v odstavcích), která by zde byla konstruktem, a odstavec (měřený v průměrném počtu souvětí), jenž by byl jejím konstituentem, viz Tabulka 11.

Tabulka 11 Alternativní jazyková úroveň U_i , x_i konstrukt, y_i konstituent

JAZYKOVÁ ÚROVEŇ U_i	
x_i	podkapitola měřená v odstavcích
y_i	odstavec měřený v průměrném počtu souvětí

Pro testování vzájemných vztahů jazykových jednotek na této nejvyšší úrovni však daný výběrový soubor není svou délkou příliš vhodný – nejenže by empiricky získaných pozorování bylo málo, ale i jejich frekvence by byly velmi nízké. Délka tohoto vědeckého textu, cíleně námi limitovaná z důvodu komparace jednotlivých výběrových souborů reprezentující různé jazykové styly, tedy příliš neumožňuje provést testování existence vzájemných vztahů těchto dvou jazykových jednotek. Zároveň však ani nevylučuje možnost podrobit tuto jazykovou úroveň kvantitativnímu rozboru v případě dalšího experimentu, jenž by testoval delší výběrový soubor.

K vymezení předmětu segmentace je nutné na závěr zmínit, že se ve vlastní stati článku vyskytla latinka. U zahraničního jména autor uvedl kromě čínské verze také jeho alternativu psanou v latině „弗朗斯瓦·格林 (François Grin)“ (Li, 2013). Protože tato alternativa vložená do závorky opakovala již zmíněnou informaci, a tudíž nepředstavovala obsahově neoddělitelnou část parcelátu, nebyla zohledněna.

Tím, že jsme si vymežily objekt kvantitativní analýzy, jsme mohly přistoupit k dalšímu kroku, kterým byla segmentace. Před jejím zahájením však bylo ještě potřebné stanovit, jaký segmentační přístup zvolit v případě určitých náležitostí, které se v textu objevily a které neměly ryze čínský charakter. Jednalo se zejména o čísla zapsaná arabskými číslicemi. U těchto numerických zápisů bylo nutné vymežit, ke kterým jazykovým jednotkám čísla a číslice použité v textu náleží. Na rozdíl od čínského znaku čísla ani číslice nezabíraly grafické pole, tzn., že se číslice k sobě těsnaly. Z tohoto důvodu se na číslo (celé i racionální) nahlíželo jako na jeden znak. Dané číslice pak byla jeho prvky o určitém počtu tahů. Pro názornou ukázkou si uveďme například číslo 2009, které se rovná jednomu znaku tvořenému čtyřmi prvky.

Využívání arabských číslic v čínském textu nejspíše souvisí s ekonomizací jazyka. Použití tohoto typu číselného zápisu představuje oproti čínským znakům rychlejší a snadnější způsob, nejen v případě psaní rukou, ale i v případě elektronického zápisu jazyka.

Nyní věnujme několik slov zápisu desetinných čísel. Jak již bylo zmíněno výše, autor pro záznam celých i racionálních čísel využívá arabské číslice. Co se týče desetinného symbolu, který odděluje celé číslo od jeho zlomkové části, pak i v tomto případě nahrazuje čínský znak dian (diǎn, 点) jednodušší a rychlejší forma – desetinná tečka. Pro segmentaci výběrového souboru to znamená, že na desetinnou tečku se nahlíží jako na jeden z prvků desetinného čísla, který je tvořen jedním tahem. Za využíváním této formy zápisu by opět mohla být ekonomizace jazyka. Použití arabských číslic a desetinné tečky totiž psaní jednoznačně zjednoduší a urychlí, jak lze ostatně demonstrovat na příkladu výrazných rozdílů mezi počty prvků a tahů obou podob desetinných čísel, viz Tabulka 12. Obě formy zápisu desetinného symbolu jsou v tabulce pro přehlednost zaznačeny červeně.

Tabulka 12 Ukázka dvou typů zápisu desetinných čísel a jejich rozdílnosti demonstrované na počtu prvků a tahů

Zápis desetinných čísel – arabské číslice			Zápis desetinných čísel – čínské znaky		
Číslo	Počet prvků	Počet tahů	Číslo	Počet prvků	Počet tahů
5.68	4	5	五 ^点 六八	12	19
6.33	4	4	六 ^点 三三	15	19
1.43	4	7	一 ^点 四三	10	14

Zdroj: (Li, 2013), vlastní zpracování

Číslicemi autor nahradil i čínské výrazy používané pro výčet. Místo potenciálních čínských ekvivalentů, jako jsou např. di-yi (dì-yī, 第一) a shouxian (shǒuxiān, 首先) pro výraz „poprvé“, di-er (dì-èr, 第二) a qici (qíci, 其次) pro výraz „zadruhé“ a di-san (dì-sān, 第三) pro výraz „začtřetí“, použil arabské číslice opatřené tečkou (tj. 1. , 2. a 3.). U segmentace se na tuto formu nahlíží opět jako na jeden znak, který tvoří dva prvky – číslice a tečka. I tento typ zápisu nás přivádí k myšlence, že by se mohlo jednat o proces ekonomizace jazyka.

Poslední poznámka, která s čísly a číslicemi souvisí, patří procentům. V čínském jazyce existuje pro procento výraz bai fenzhi (bǎi fēnzhī, 百分之), po němž následuje číselné vyjádření. Od použití tohoto výrazu autor upustil a nahradil ho snadnější a rychlejší formou zápisu – symbolem pro procento %. Ačkoliv tento symbol grafické pole nezabírá, pokládá se v tomto experimentu za jeden znak tvořený třemi prvky. V praxi to znamená, že se celý výraz pro vyjádření procentuální míry bude skládat ze dvou znaků – čísla a symbolu pro procento. Ekonomizace jazyka by zde mohla hrát opět velmi důležitou roli.

Jisté anomálie se objevily i u použité interpunkce. Jak již bylo napsáno výše, pro hranice jazykové jednotky souvětí platí vybraná interpunkční znaménka (。 ? !). Ve dvou případech však souvětí v závěru odstavce zakončovala dvojtečka a text, který následoval za ní, začínal jako nový odstavec (tj. oddělený vloženým prázdným řádkem a odsazený od okraje), viz níže uvedený příklad (Li, 2013):

„人类的经济活动与语言密不可分，而且在某些领域，语言和语言知识已经成为重要的经济资源。这可以从以下几个方面来看 。

首先，语言能力是劳动力的重要构成要素。“

Za chybu lidského faktoru tyto odchylky považovat nelze, neboť použití dvojtečky se v obou případech vyskytlo za identických podmínek. Za dvojtečkou zakončeným souvětím, resp. odstavcem následoval výčet určitých aspektů. Každému z nich autor věnoval vždy jeden odstavec, který začínal čínskými výrazy používanými právě pro výčet (zprvé, zadruhé atd.; čínsky: shǒuxiān, 首先/ qícì, 其次 atd.). Protože jsme stanovení jazykových jednotek a následnou segmentaci stavěly primárně na grafickém hledisku, zvolila jsem přednostně tu alternativu segmentování, kdy odstavce zůstaly zachovány a dvojtečka se sekundárně stala hraničním bodem souvětí.

Zvláštní případ se vyskytl u souvětí, ve kterém název dokumentu uvedený v závorkách pro tituly (《》) obsahoval čárku (,), viz příklad níže (Li, 2013). Závorky pro tituly jsme se spoluautorkou Lenkou Spáčilovou zařadily do skupiny interpunkčních znamének, která nerozlišují hranice parcelátů (viz podkapitola 2.3, Tabulka 8). Oproti tomu čárka patří do skupiny, která hranice vytváří (viz podkapitola 2.3, Tabulka 7). Segmentace tohoto souvětí se primárně řídila námi stanovenými pravidly – závorky pro tituly nebyly zohledněny. Protože by jejich nezohledněním daná čárka narušila syntaktickou strukturu věty, výjimečně nebyla spolu se závorkami považována za hranici parcelátu.

„在 2010、2011 年北京市“两会”上，北京市人大代表贺宏志先生连续提出《关于发展我市语言产业的建议》和《加强语言文化建设 促进语言产业发展》的建议，语言经济的话题首次提到了地方人民代表大会的论坛上。“

Poslední nutnou záležitostí, kterou je třeba ve spojitosti se segmentací zmínit, je šest kategorií interpunkčních znamének, které ze syntaktického hlediska nedisponují kompetencí dělit souvětí na obsahově odlišné segmenty (v našem

případě parceláty) a které se objevily ve vědeckém článku. Nejčastěji se v textu vyskytujícím interpunkčním znaménkem bez segmentační způsobilosti stala čárka pro výčet 、 (dùnhào, 顿号). Poměrně často byly využívány také dvojité uvozovky evropského typu “ ” (yǐnhào, 引号) a dvojité závorky pro tituly 《 》 (shūmíng hào, 书名号). Další interpunkční znaménka – elipsa (shěnglüè hào, 省略号), pomlčka —— (pòzhéhào, 破折号) a závorky () (kuò hào, 括号) – se objevila spíše sporadicky.

Po dokončení segmentační fáze jsem přistoupila k dalšímu kroku, jenž měl za cíl vědecký článek kvantifikovat a vypočítat délky veličin x_i (délka konstruktů), z_i (frekvence) a y_i (délka konstituentů), kde $i = 1, 2, 3, 4$, více viz podkapitola 2.4.

3.5 Testování spolehlivosti modelu pomocí statistických metod

Kvantifikací získaná data jsem vložila do statistického softwaru R, který za pomoci lineární regrese (viz podkapitola 2.5, poznámka pod čarou č. 10) vypočítal pro každou jazykovou úroveň zbývající potřebné výstupy, a to parametry A , b a koeficient determinace R^2 (více viz podkapitola 2.5), a ke každé úrovni rovněž vytvořil grafickou vizualizaci těchto výsledků.

Statistické metody byly také využity pro nejméně frekventovaná pozorování, tzv. extrémů. Pozorování s nízkou četností výskytu, která byla v porovnání s ostatními zanedbatelná, byla statisticky ošetřena metodou vynechání. Pro přehlednost jsou tyto extrémů zaznačeny v příslušných tabulkách šedou barvou. Hodnoty získané jejich odebráním – parametry A , b a koeficient determinace R^2 – jsou uváděny v poznámkách pod čarou.

Co se týče mnou provedeného experimentu, statisticky jsem ošetřila pouze pozorování získaná na jazykové hladině znak – prvek. U ostatních úrovní, příp. dílčích experimentů, nebylo potřebné empiricky získaná pozorování s nejnižší frekvencí odebrat.

3.6 Interpretace získaných dat

Tímto posledním krokem přicházíme k samotnému jádru této části diplomové práce. Následující čtyři podkapitoly budou věnovány námi zvoleným jazykovým hladinám U4 – U1. V každé z nich bude uvedena tabulka s délkami veličin x_i (délka konstruktů), z_i (frekvence) a y_i (délka konstituentu) a grafická vizualizace těchto pozorování spolu s komentáři, za nimiž bude následovat diskuze. Závěry interpretací a diskuzí budou shrnuty na konci této podkapitoly.

3.6.1 Jazyková úroveň U4

Znak – prvek

Tabulka 13 uvádí empirické hodnoty získané kvantifikací vědeckého článku. První sloupec náleží délkám znaků měřených v prvcích x_4 , sousední sloupec četnostem jejich výskytu z_4 a poslední sloupec průměrným délkám prvků měřených v tazích y_4 . Výplň buněk šedé barvy značí ta pozorování, jejichž četnost výskytu ($z_4 \leq 8$) je v porovnání s ostatními zanedbatelná a která jsou proto statisticky ošetřena. Hodnoty získané jejich odebráním (parametry A , b a koeficient determinace R^2) jsou uvedené v poznámce pod čarou č. 17.

Tabulka 13 Jazyková úroveň U4: znak (měřený v prvcích) – prvek (měřený v průměrném počtu tahů)

x_4	z_4	y_4
1	789	4,7136
2	1240	3,3056
3	1285	2,4999
4	995	2,0766
5	540	1,6815
6	134	1,9701
7	163	1,3681
8	8	1,5000
9	1	1,7778

Výběrový soubor operuje celkem s devíti délkami konstruktů. Ty jsou měřeny v konstituentech, jejichž průměrná délka se pohybuje v rozmezí $\langle 1,37; 4,71 \rangle$ tahů. Z tabulky je patrné, že čím větší je znak v počtu prvků, tím menší jsou v průměru jeho prvky v počtu tahů. Vztah nepřímé úměrnosti mezi délkami veličin definovaný MAlem tedy na této jazykové úrovni platí. Přesto klesající tendenci průměrných délek narušily šesti ($y_4 = 1,97$), osmi ($y_4 = 1,50$) a devíti ($y_4 = 1,78$) prvkové znaky. Tuto odchylku by u osmi a devíti prvkových znaků mohla pravděpodobně zapříčinit nízká četnost výskytu. Podle uvedené frekvence můžeme také konstatovat, že se v textu vyskytovaly nejvíce dvou a tři ($z_4 \geq 1240$) prvkové znaky a nejméně znaky osmi a devíti prvkové ($z_4 \leq 8$), viz Tabulka 13.

V tabulce níže vidíme přehled osmi a devíti prvkových znaků. Každý z nich je doprovázen počtem prvků, rankem¹⁴ přiděleným programem 文林 Wenlin¹⁵ na základě frekvence, frekvenčním pásmem¹⁶ a četností výskytu ve výběrovém souboru. Z ranků a pásem je patrné, že většina z nich náleží ke znakům s nižší frekvencí.

Tabulka 14 Ukázka osmi a devíti prvkových znaků a jejich frekvence v rámci Frekvenčního seznamu znaků programu 文林 Wenlin Software

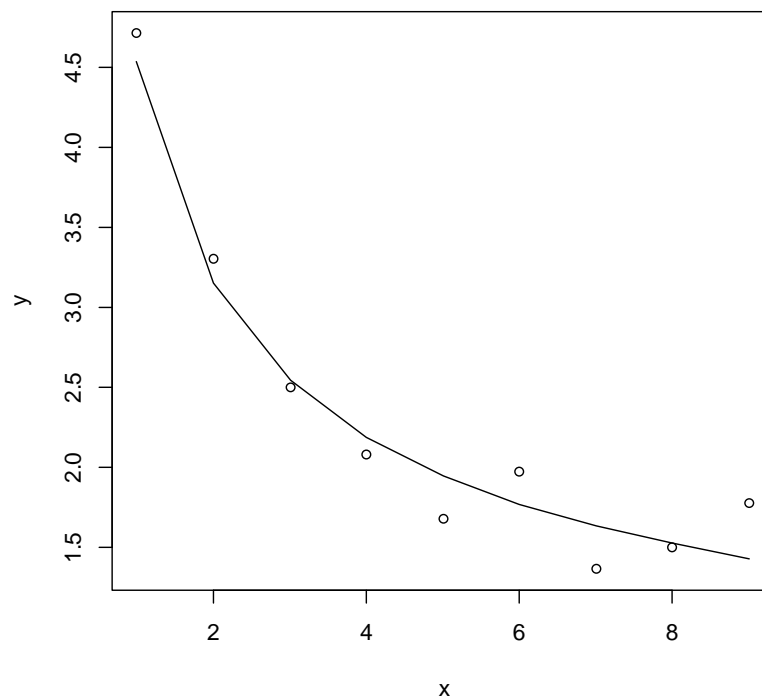
Znak	Počet prvků	Rank	Frekvenční pásmo	Četnost výskytu
感	8	336	1. pásmo	1
德	8	401	1. pásmo	1
测	8	1121	2. pásmo	2
综	8	1553	2. pásmo	2
墨	8	1680	2. pásmo	1
韵	8	2126	3. pásmo	1
飙	9	×	×	1

Zdroj: 文林 Wenlin Software, vlastní zpracování

¹⁴ Čím nižší je hodnota uváděných ranků, tím vyšší je frekvence daných znaků.

¹⁵ Frekvenční seznam znaků z programu 文林 Wenlin Software for Learning Chinese. Version 4.0.2.

¹⁶ 3 000 nejvíce frekventovaných znaků uvedených ve frekvenčním seznamu Wenlinu jsou pro lepší orientaci rozděleny podle ranků do tří frekvenčních pásem: první pásmo zahrnuje nejfrekventovanější znaky, tj. znaky s rankem od 1 do 1 000, druhé pásmo znaky s rankem od 1 001 do 2 000 a třetí pásmo nejméně frekventované znaky, tj. znaky s rankem od 2 001 do 3 000. U znaků, které kvůli nízké četnosti výskytu svůj rank nemají, je v tabulce uveden křížek.



Obrázek 5 Grafická vizualizace pozorování uvedených v Tabulce 13

Z grafické vizualizace vidíme, že předpoklady pro platnost MALu byly na této úrovni splněny. Parametr b (viz Tabulka 15) má kladnou reálnou hodnotu, a proto je křivka, zobrazující vztah mezi délkou znaku a délkou prvku, klesající a konvexní. Je zjevné, že empiricky získaná pozorování vykazala s matematickým modelem vysokou shodu, jak ostatně dokazuje koeficient determinace R^2 (viz Tabulka 15). Procento shody dosáhlo extrémní hodnoty – 90 %.

Tabulka 15 Parametr A , parametr b a koeficient determinace R^2 pro matematický model související s pozorováními uvedenými v Tabulce 13¹⁷

Parametr A	Parametr b	Koeficient determinace R^2 (%)
4,5406	0,5259	90,61

¹⁷ Po vynechání nejméně frekventovaných pozorování ($z_4 \leq 8$): parametr $A = 4,8117$; parametr $b = 0,5974$; koeficient determinace $R^2 = 95,19$ %.

Diskuze

Je na místě si položit otázku, co by tuto extrémní shodu mohlo zapříčinit. Na úvod věnujme několik slov charakteristice konstruktů této úrovně. Oldřich Švarný uvádí, že „znaky tvoří vždy přibližně stejně velké grafické jednotky, bez ohledu na počet tahů, z nichž se znak skládá; tahy jsou uspořádány do čtverce nebo spíše do obdélníka (o málo větší výšce než šířce).“ (Švarný, 1967, s. 31) Tím, že si grafické pole zachovává bez ohledu na počet tahů ve znaku stále stejnou velikost, vytváří na strukturu znaku velmi silný tlak. Uvnitř svého limitovaného prostoru pole formuje systém pravidel, kterým řídí uspořádání tahů a prvků ve znaku, čímž vytváří s tahy a prvky jistý subordinační vztah. Za hlavní faktor široké shody na této úrovni bychom tedy mohli považovat samo grafické pole, resp. jeho konstantní velikost.

Aplikujeme-li MAL na tuto úroveň, jeho slovní formulace bude následovná:

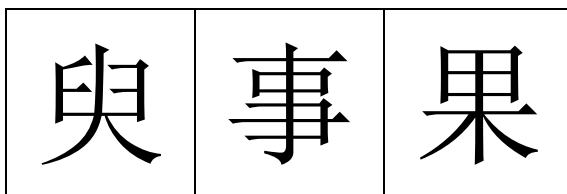
ČÍM VÍCE MÁ ZNAK PRVKŮ, TÍM MÉNĚ MÁ PRVEK TAHŮ.

V opačném případě pak platí:

ČÍM MÉNĚ MÁ ZNAK PRVKŮ, TÍM VÍCE MÁ PRVEK TAHŮ.

Zastavme se nejprve u poslední zmíněné obměny MALu, kterou demonstrováme na příkladu jedno prvkových znaků. Znak tvořený jedním prvkem má potenciál nabývat složitější strukturu, aniž by byla narušena jejich čitelnost, neboť mohou disponovat prostorem celého pole. Z důvodu absence dalších prvků jsou prvky těchto znaků složeny z většího počtu tahů (viz Tabulka 16), neboť prostorový tlak pole je zde minimální.

Tabulka 16 Příklad zjednodušených znaků tvořených jedním prvkem o osmi tazích



U více prvkových znaků jsou tyto možnosti limitované. Jestliže se znak skládá z více prvků, bude každý z nich oproti jednoprvkovým znakům okupovat v grafickém poli mnohem menší prostor. Jinými slovy, čím více prvků znak má, tím

menší je prostor, kam se mohou dané prvky zapsat. Tento fakt vyvíjí na strukturu znaku velký tlak. Čím menší je velikost segmentu grafického pole, tím větší je z důvodu zachování čitelnosti požadavek jednoduchosti na prvek. Aby si prvky zachovaly čitelnost, nesmí být příliš složité. Z tohoto důvodu jsou více prvkové znaky tvořeny menším počtem tahů.

Zároveň je třeba mít na paměti, že při větším počtu prvků ve znaku se v rámci grafického pole zmenšuje jejich velikost. Protože se více tahový prvek může zmenšit jen do určité míry, aniž by hrozila ztráta jeho čitelnosti, nutí grafické pole z důvodu zachování čitelnosti využívat především prvky jednodušší, tj. prvky o menším počtu tahů.

Čím větší počet prvků tedy znak má, tím menší počet tahů budou prvky mít (viz Tabulka 17). Tímto se dostáváme k Menzerath-Altmanovu zákonu. Dá se říct, že se jedná o určitou míru ekonomizace prostoru grafického pole, která sílí v závislosti na větším počtu prvků ve znaku a zabezpečuje tak jeho čitelnost.

Tabulka 17 Příklady zjednodušených znaků tvořených různými počty prvků a tahů a jejich průměrných délek

	Počet prvků	Počet tahů ve znaku	Průměrný počet tahů prvků
潮	5	15	3
撰	6	15	2,5
影	7	15	2,1

Při sledování struktur prvků a tahů ve znaku si můžeme všimnout, že zde existuje ještě jedna tendence působící opačným směrem. Jestliže je znak tvořen velkým počtem prvků, jeho prvky inklinují k tomu, že se k sobě těsnají a navzájem se dotýkají, neboť grafické pole jim neposkytuje dostatečně velký prostor, aby si zachovaly svoji samostatnost. Prázdný prostor mezi prvky se eliminuje a z hlediska počtu tahů se vytvoří složitější prvek. Tato tendence má za výsledek jak redukcii

počtu prvků v rámci znaku, tak navýšení počtu tahů daného prvku. MAL se nám potvrzuje i tentokrát – čím méně má znak prvků, tím více tahů má prvek.

Tento systém pravidel pro vytváření struktur znaků uvnitř grafického pole nás dokonce přivádí k myšlence, že tato pravidla musí u tradičních znaků platit mnohem striktněji, neboť oproti zjednodušeným variantám disponují obecně větším počtem tahů a prvků. Tím se vracíme k našim předchozím tvrzením, že čím vyšší mají znaky počet prvků, tím nižší mají prvky počet tahů, viz Tabulka 18, a čím méně prvků znak má, tím více má prvek tahů, ať již z důvodu disponování větším prostorem uvnitř pole, nebo z důvodu tendence prvků se pro nedostatek prostoru spojovat a vytvářet prvky složitější, viz Tabulka 19.

Tabulka 18 Příklad tradičního znaku tvořeného 17 prvky a 32 tahy



Tabulka 19 Příklad tradičního znaku tvořeného 5 prvky a 30 tahy



Zda tedy tato pravidla platí v takové míře i u tradičních znaků, jsme zkusily ověřit experimentem na vědeckém článku transformovaném do tradičních znaků. Experiment je značen podle příslušné úrovně (tj. 4) a příslušného výběrového souboru (tj. A).

EXPERIMENT 4A – TRADIČNÍ ZNAKY

Platnost vzájemného vztahu mezi znakem a prvkem jsem ověřovala na identickém výběrovém souboru, který jsem pomocí programu 文林 Wenlin Software převedla do nezjednodušené podoby. Výsledná transformace se týkala 2 027 znaků z celkového počtu 5 155 (282 různých znaků z celkového počtu 761), zbylé v textu použité znaky (3 128; 479) si zachovaly svoji původní podobu. Následná segmentace transformovaných forem znaků striktně dodržovala námi zvolenou definici jazykové jednotky (viz ostrovy, podkapitola 2.3, s. 33).

Tabulka 20 uvádí pozorování empiricky získaná kvantifikací transformovaného vědeckého textu. První sloupec náleží délkám znaků měřených v prvcích x_4 , sousední sloupec četnostem jejich výskytu z_4 a poslední sloupec průměrným délkám prvků měřených v tazích y_4 .

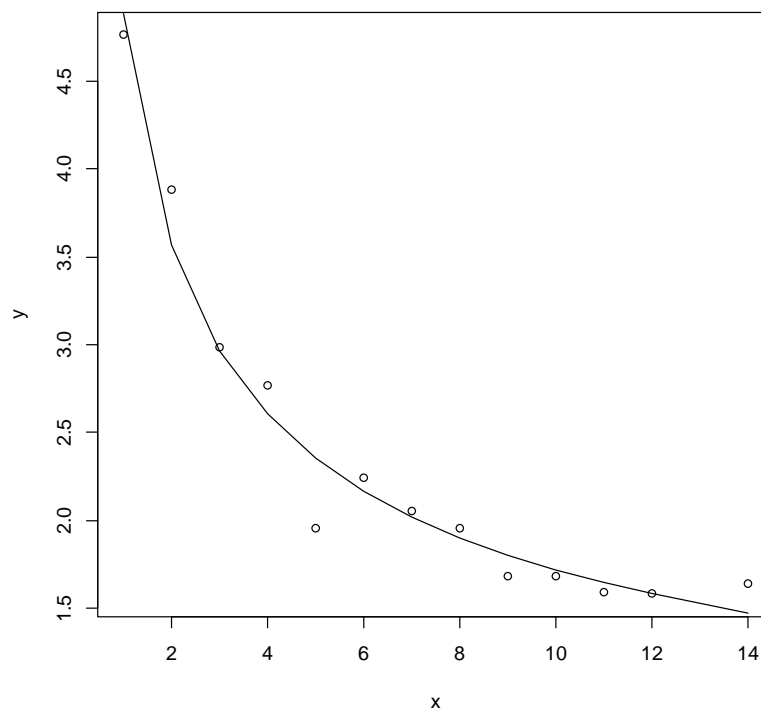
Tabulka 20 Experiment 4A – Jazyková úroveň: znak (měřený v prvcích) – prvek (měřený v průměrném počtu tahů)

x_4	z_4	y_4
1	739	4,7605
2	975	3,8795
3	1011	2,9829
4	607	2,7702
5	662	1,9541
6	638	2,2422
7	128	2,0569
8	63	1,9603
9	193	1,6845
10	134	1,6866
11	2	1,5909
12	1	1,5833
14	2	1,6429

V transformovaném výběrovém souboru se objevilo celkem třináct délek konstruktů. Průměrná délka jejich konstituentů se pohybuje v intervalu (1,58 – 4,76) tahů. Je zjevné, že s rostoucí délkou znaků klesá průměrná délka prvků; vztah nepřímé úměrnosti mezi konstruktem a konstituentem na této jazykové úrovni platí. Klesající tendence průměrných délek byla narušena celkem dvakrát. Nepatrný nárůst průměrné délky konstituentu vykázaly deseti ($y_4 = 1,69$) a čtrnácti ($y_4 = 1,64$) prvkové znaky. Poněkud zvláštním případem jsou znaky tvořené pěti prvky. Ačkoliv průměrná délka jejich konstituentů ($y_4 = 1,95$) projevila klesající tendenci, její hodnota výrazně narušuje pokles následujících průměrných délek a vytváří tak odchylku. Podle v tabulce uvedené frekvence můžeme také konstatovat, že se ve výběru vyskytly nejvíce znaky dvou prvkové a tří prvkové ($z_4 \geq 975$). Poměrně výrazné zastoupení mají i znaky tvořené jedním prvkem, dále pak čtyřmi, pěti a šesti

prvky. Frekvence těchto znaků se pohybuje v intervalu (607; 739). Nejnižší četnost výskytu vykazují jedenácti, dvanácti a čtrnácti prvkové znaky ($z_4 \leq 2$).

Porovnáme-li mezi sebou empiricky získané výsledky na této úrovni, zjistíme, že oproti původnímu textu psanému zjednodušenými znaky (viz Tabulka 13) disponuje transformovaný text větším počtem délek konstruktů. V textu se navíc vyskytly znaky tvořené deseti, jedenácti, dvanácti a čtrnácti prvky. V případě délek konstituentů je situace jiná. Přestože se průměrná délka prvků y_4 u obou variant výběrového souboru pohybuje přibližně ve stejném rozmezí – mezi jedním a pěti tahy, v případě transformované varianty textu je u jednotlivých délek znaků x_4 patrné mírné navýšení průměrných délek prvků měřených v tazích. Toto navýšení odkazuje na přítomnost transformovaných znaků, které se obecně vykazují větším počtem tahů. Vztah nepřímé úměrnosti – s rostoucí délkou znaků klesá průměrná délka prvků – se u obou variant prokázal. Obě dvě také zaznamenaly stejný počet odchylek klesající tendence průměrných délek (3). Nejvyšší četnost výskytu u obou variant souboru vykazaly dvou prvkové a tří prvkové znaky. V případě původní varianty to dále byly čtyř a jedno prvkové znaky. U transformovaného textu se konstruktem s třetí nejvyšší frekvencí staly znaky jedno prvkové, za něž se zařadily poměrně hojně používané znaky složené z pěti, šesti a čtyř prvků. Frekvence těchto více prvkových znaků (tzn. pět a více) se navýšila na úkor původní vysoké frekvence dvou a tří prvkových znaků (viz Tabulka 13). Co se týče nejméně frekventovaných pozorování, v obou případech se jedná o závěrečné délky konstruktů – osmi a devíti prvkové u první varianty, jedenácti, čtrnácti a dvanácti prvkové znaky u varianty druhé.



Obrázek 6 Grafická vizualizace pozorování uvedených v Tabulce 20

Nejen z vizualizace vztahů mezi znakem a prvkem, ale z i koeficientu determinace R^2 (Tabulka 21) vyplývá, že matematický model vykazuje s pozorováními získanými kvantifikací vědeckého textu transformovaného do tradiční podoby velmi vysokou procentuální shodu. Shoda dokonce dosáhla vyšší hodnoty, než jak je tomu v případě simplifikované varianty textu (viz Tabulka 15).

Tabulka 21 Parametr A , parametr b a koeficient determinace R^2 pro matematický model související s pozorováními uvedenými v Tabulce 20¹⁸

Parametr A	Parametr b	Koeficient determinace R^2 (%)
4,8822	0,4530	95,68

Nejen potvrzení platnosti MALu na této úrovni ověřené na výběrovém souboru transformovaném do tradiční podoby, ale zejména výsledná shoda, která přesahuje 90 %, nás utvrzuje v tom, že grafické pole se svým vnitřním systémem

¹⁸ V případě této jazykové úrovně nebylo potřebné empiricky získaná pozorování statisticky ošetřit vynecháním nejméně frekventovaných pozorování.

pravidel platných pro formování struktury znaku hraje při vytváření vzájemných vztahů mezi jazykovými jednotkami – tahu, prvky a znaku – fundamentální roli.

Je však třeba zdůraznit, že empiricky získaná pozorování odráží nejen vzájemné vztahy mezi prvky a znaky převedenými do tradiční podoby, ale i vzájemné vztahy mezi prvky a znaky zachovávající si svou původní podobu. Transformace do tradiční formy totiž zahrnovala pouze 2 027 znaků z celkového počtu 5 155, zbylé znaky (téměř 61 %) si zachovaly svoji původní podobu. Svým většinovým zastoupením mají tyto znaky na výsledná data stále markantní vliv. Získaná pozorování a hodnoty proto nejsou komplexní a pouhou reflexí vzájemného vztahu mezi prvky a znaky, které lze převést do tradiční podoby. Získaná hodnota koeficientu determinace ($R^2 = 95,69$) by mohla napovídat, že dodržování výše popsaných pravidel grafického pole je u znaků, jejichž podoba byla simplifikována reformami minulého století, striktnější. Výsledná procentuální shoda s matematickým modelem MALu by v tomto případě mohla být dokonce vyšší než v námi testovaných případech a tím by se opět potvrdila klíčová role grafického pole. Tuto hypotézu je potřeba dále testovat, a proto plánuji uskutečnit další experimenty zacílené výhradně jen na tyto formy znaků.

3.6.2 Jazyková úroveň U3

Parcelát – znak

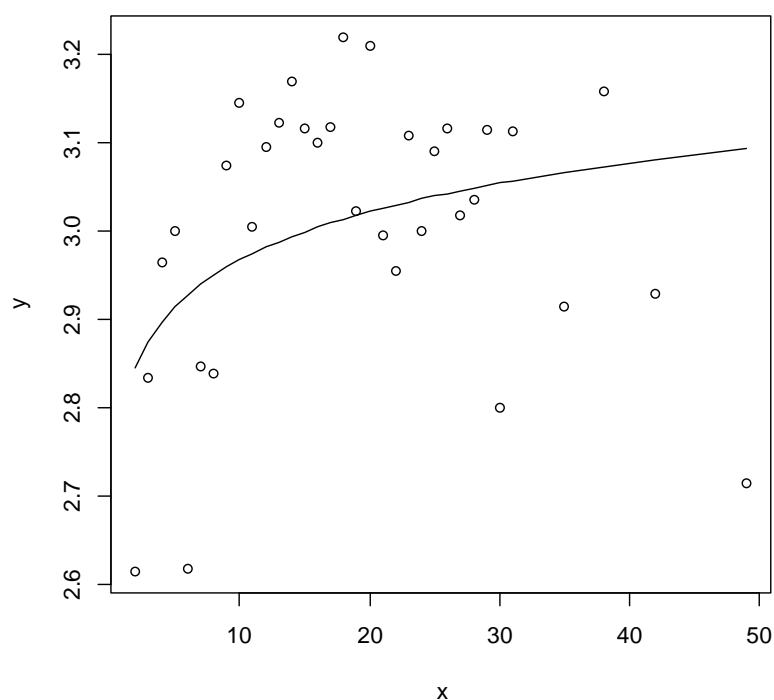
Tabulka 22 uvádí empirické hodnoty získané kvantifikací vědeckého článku. První sloupec obsahuje délky parcelátů měřených ve znacích x_3 , prostřední sloupec četnosti jejich výskytu z_3 a třetí sloupec průměrné délky znaků měřených v prvcích y_3 .

Tabulka 22 Jazyková úroveň U3: parcelát (měřený ve znacích) – znak (měřený v průměrném počtu prvků)

x_3	z_3	y_3
2	13	2,6154
3	2	2,8333
4	7	2,9643
5	4	3,0000
6	17	2,6176
7	14	2,8469
8	17	2,8382
9	15	3,0741
10	20	3,1450
11	17	3,0053
12	35	3,0952
13	15	3,1231
14	22	3,1688
15	24	3,1167
16	22	3,0994
17	19	3,1176
18	17	3,2190
19	18	3,0234
20	15	3,2100
21	10	2,9952
22	5	2,9545
23	6	3,1087
24	6	3,0000
25	4	3,0900
26	6	3,1154
27	2	3,0185

28	2	3,0357
29	3	3,1149
30	1	2,8000
31	2	3,1129
35	1	2,9143
38	1	3,1579
42	1	2,9286
49	1	2,7143

Ve zkoumaném článku se objevilo celkem 34 délek konstruktů. Průměrná délka jejich konstituentů se pohybuje v rozmezí $(2,62; 3,22)$ prvků. Protože se klesající tendence průměrných délek neprokázala, ba naopak silně kolísá ve výše uvedeném intervalu, vztah nepřímé úměrnosti mezi parcelátem a znakem se zde neprokázal. Co se týče četnosti výskytu konstruktů, frekvence nedosahují z důvodu poměrně vysokého počtu délek parcelátů příliš vysokých hodnot. Přesto můžeme konstatovat, že se v textu objevily nejčastěji parceláty tvořené 12 znaky ($z_3 = 35$). Parceláty s největší délkou ($x_3 = 30, 35, 38, 42$ a 49) se naopak vyskytly v textu vzácně ($z_3 \leq 2$).



Obrázek 7 Grafická vizualizace pozorování uvedených v Tabulce 22

Parametr b má v tomto případě zápornou hodnotu, což odporuje předpokladu platnosti MALu (viz Tabulka 23). Z vizualizace vztahů mezi parcelátem a znakem vidíme, že se v případě této jazykové úrovně projevila opačná tendence křivky, než jakou předpokládá MAL, tedy tendence stoupající. Z tohoto důvodu matematický model nevykázal s empiricky získanými pozorováními žádnou procentuální shodu (koeficient determinace R^2 uvedený v Tabulce 23 platí v tomto případě pro stoupající tendenci křivky, která je v rozporu s předpokladem platnosti MALu).

Tabulka 23 Parametr A , parametr b a koeficient determinace R^2 pro matematický model související s pozorováními uvedenými v Tabulce 22¹⁹

Parametr A	Parametr b	Koeficient determinace R^2 (%)
2,7933	-0,0262	14,35

Diskuze

V případě jazykové úrovně: parcelát – znak, se vzájemný vztah jazykových jednotek definovaný MALEM neprokázal. Je nutno zde poukázat na několik faktorů, které by mohly mít vliv na neplatnost jejich vzájemného vztahu.

Za první z nich bychom mohly považovat odlišný charakter jazykových jednotek této hladiny. Parcelát měřený v počtu znaků reprezentuje z hlediska délky variabilní konstrukt. Autor textu tvoří délku této jazykové jednotky sám v závislosti na své jazykové inteligenci a vyjadřovacích schopnostech. Jinými slovy, s trochou nadsázky můžeme říct, že parcelát může být nekonečně dlouhý, neboť není ničím limitovaný. Oproti tomu znak, jehož délka je měřená v prvcích, představuje konstituent s konstantní délkou. Nejenže se tato jazyková jednotka váže na určitý význam s danými pravidly pro jeho užití, čímž znemožňuje autorovi jednotlivé znaky svévolně zaměňovat za alternativy o nižším či vyšším počtu prvků, ale především má fixní a neměnnou strukturu, kterou autor měnit nemůže. Ve srovnání s parcelátem je znak z hlediska délky (tj. počtu prvků) na autorovi zcela nezávislou jazykovou

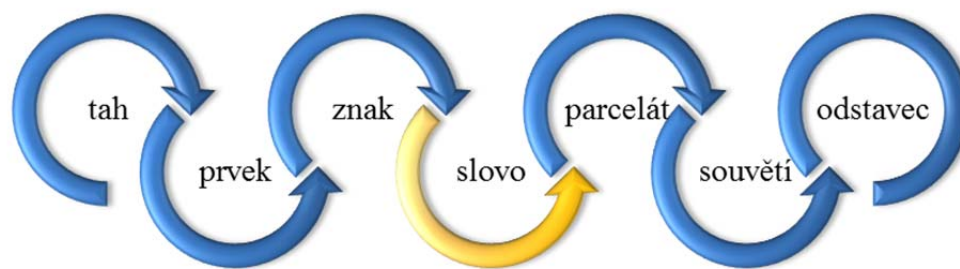
¹⁹ V případě této jazykové úrovně nebylo potřebné empiricky získaná pozorování statisticky ošetřit vynecháním nejméně frekventovaných pozorování.

jednotkou. Tento protiklad by mohl výrazně narušit vztah nepřímé úměrnosti mezi délkami daných veličin.

Nejen parcelát a znak, ale i prvek může mít markantní vliv na neadekvátnost matematického modelu MALu na této úrovni. Jak již bylo uvedeno v podkapitole 2.3, z důvodů absence obecně platné definice může být prvek určen více způsoby. Dokonce i zvolení příslušného způsobu segmentace znaku na prvky nemusí být definitivní. Může totiž dojít k jeho diverzifikaci na dílčí přístupy, jak již bylo demonstrováno u variability fontů (viz Tabulka 6). Ačkoliv se může na základě vybrané definice prvku platnost MALu na jazykové úrovni znak – prvek projevit, na úrovni parcelát – znak tomu tak být nemusí. Daná definice totiž nemusí pro tuto úroveň představovat adekvátně zvolený přístup. Vztah těchto jazykových jednotek proto nadále zůstane předmětem výzkumu a bude zkoumán na základě různých přístupů určení prvků, které budou ověřovány pomocí MALu.

Neplatnost vzájemného vztahu mezi veličinami této úrovně by rovněž mohla poukazovat na nedostatečné množství jazykových jednotek a hladin. Délku parcelátu tvoří určitý počet znaků, které jsou v tomto experimentu pojímány jako samostatně vystupující entity. Za důležité považují poznamenat, že se vznikem víceslabičných slov ztratily znaky logografický charakter. Významově se tak váží na znaky jiné, čímž se mezi nimi vytváří jistý vztah (= absentující jazyková hladina). Ten je však z důvodu primárního kritéria grafiky ignorován, neboť jej grafika čínského textu psaného ve znacích nereflektuje. Protože znak a parcelát tedy představují od sebe značně vzdálené jednotky, nemusel by u nich vzájemný vztah nepřímé úměrnosti existovat. Jejich spojením do jedné jazykové úrovně a vynecháním jazykové jednotky, která je vyšší než znak a nižší než parcelát, tak dochází ke ztrátě vzájemných závislostí struktur, tj. vzájemných vztahů nepřímé úměrnosti mezi délkami chybějící jednotky (konstrukt) a znaku (konstituent) a mezi délkami parcelátu (konstrukt) a chybějící jednotky (konstituent).

Jaká jazyková jednotka tedy mezi znakem a parcelátem chybí? Absentující jednotkou by pravděpodobně mohlo být *slovo*. Připojením slova bychom získaly následující jazykové jednotky (viz Obrázek 8) a úrovně (viz Tabulka 24).



Obrázek 8 Přehled jazykových jednotek II

Tabulka 24 Jazyková úroveň U_i se slovem na pozici konstruktů x_i a jazyková úroveň U_i se slovem na pozici konstituentu y_i

JAZYKOVÁ ÚROVEŇ U_i		JAZYKOVÁ ÚROVEŇ U_i	
x_i	slovo měřené ve znacích	x_i	parcelát měřený ve slovech
y_i	znak měřený v průměrném počtu prvků	y_i	slovo měřené v průměrném počtu znaků

Slovní formulace MALu by pro slovo na pozici konstruktů zněla takto:

**ČÍM DELŠÍ JE SLOVO V POČTU ZNAKŮ, TÍM KRATŠÍ JSOU JEHO PRVKY V POČTU
TAHŮ;**

a pro slovo na pozici konstituentu takto:

**ČÍM DELŠÍ JE PARCELÁT V POČTU SLOV, TÍM KRATŠÍ JSOU JEHO SLOVA V POČTU
ZNAKŮ.**

Pokud bychom slovo připustily jako další jazykovou jednotku, musely bychom opustit grafické hledisko, které bylo při stanovování jazykových jednotek pro obě fáze (2012 a 2013) našeho experimentu klíčové. Jestliže bychom chtěly grafický aspekt zachovat, musela by grafika čínského textu psaného ve znacích využívat mezi slovy mezery. To je však záležitostí paralelních vesmírů, a proto alternativa ortografického slova nepřichází v úvahu, neboť čínské texty takovou možností nedisponují. Při stanovování definice slova a následné segmentaci

parcelátů by se muselo zvolit jiné kritérium jako např. kritérium sémantické, syntaktické atd.

Protože je tuto hypotézu potřeba ověřit, provedla spoluautorka Lenka Spáčilová ve své části experiment (viz podkapitola 4.6.2). Na jazykové hladině slovo – znak testovala platnost vzájemného vztahu mezi syntaktickým slovem, které je na této úrovni konstruktem měřeným v počtu znaků, a znakem, jenž je jeho konstituentem měřeným v průměrném počtu prvků. Ačkoliv pozorování empiricky získaná z tohoto experimentu vykazala s matematickým modelem MALu minimální shodu, je potřeba dalších výzkumů, které navýší množství zkoumaných výběrových souborů a zpracovaných dat, jejichž kvantifikací získané výsledky budou ověřovat a precizovat nastavené definice slova.

Na neplatnost vzájemného vztahu parcelátu a znaku by mohla mít díky absenci jazykové jednotky velký vliv i četnost výskytu konstruktů nejnižší jazykové hladiny z_4 . Jak již bylo řečeno výše, průměrná délka znaků kolísá v intervalu $\langle 2,62; 3,22 \rangle$ prvků. Pro každou délku parcelátu tedy platí, že je v průměru tvořena dvou prvkovými či tří prvkovými znaky. Podíváme-li se na četnost výskytu délek těchto znaků (viz Tabulka 13), zjistíme, že se jedná o znaky s nejvyšší frekvencí. Co se týče procentuálního zastoupení, tyto znaky dohromady představují 48,98 % z celkového počtu všech v textu použitých znaků, tj. 2 525 z 5 155, a 50,07 % z celkového počtu znaků po odebrání duplicitních hodnot, tj. 381 ze 761 (viz Tabulka 25, Obrázek 9 a 10).

Širokým zastoupením těchto znaků na jedné straně a absencí jazykové jednotky a úrovně na straně druhé pravděpodobně dochází zprůměrováním délek znaků v každém parcelátu ke ztrátě jemných rozdílů. Ve výsledku pak průměrné délky znaků měřené v prvcích nabývají podobné hodnoty. V případě zařazení další jazykové jednotky by rozdíly mezi průměrnými délkami znaků mohly být výraznější a tím by se mohl projevit vzájemný vztah mezi znakem a chybějící jednotkou.

Tabulka 25 Frekvence znaků podle počtu prvků

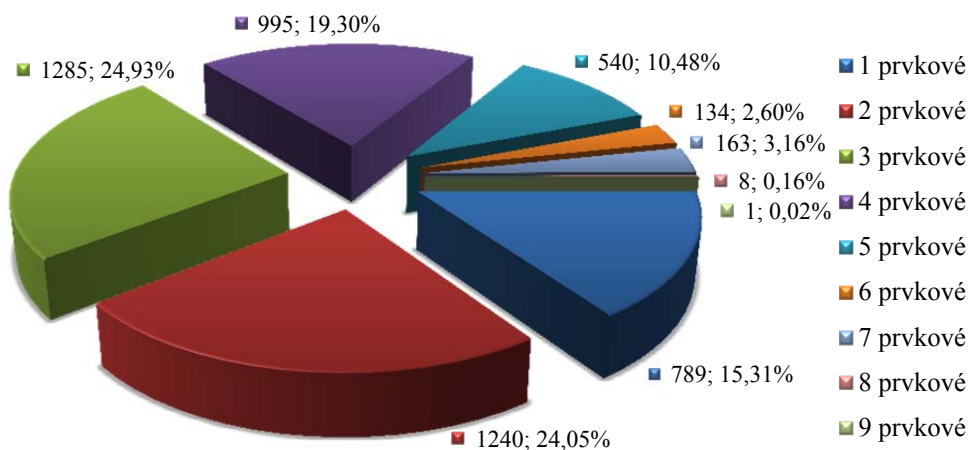
A) Frekvence všech v textu použitých znaků podle počtu prvků

Délka znaku v prvcích	Počet znaků	%
1 prvkové	789	15,3055
2 prvkové	1240	24,0543
3 prvkové	1285	24,9273
4 prvkové	995	19,3016
5 prvkové	540	10,4753
6 prvkové	134	2,5994
7 prvkové	163	3,1620
8 prvkové	8	0,1552
9 prvkové	1	0,0194
Celkem	5155	

B) Frekvence v textu použitých znaků podle počtu prvků (bez duplicit)

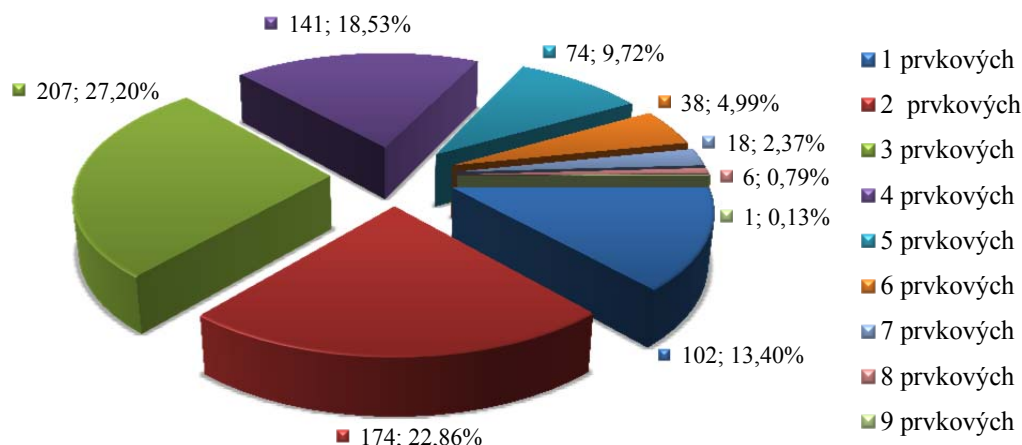
Délka znaku v prvcích	Počet znaků	%
1 prvkové	102	13,4034
2 prvkové	174	22,8647
3 prvkové	207	27,2011
4 prvkové	141	18,5283
5 prvkové	74	9,7240
6 prvkové	38	4,9934
7 prvkové	18	2,3653
8 prvkové	6	0,7884
9 prvkové	1	0,1314
Celkem	761	

Frekvence všech v textu použitých znaků podle počtů prvků



Obrázek 9 Grafická vizualizace frekvence znaků podle počtu prvků, uvedené v Tabulce 25, oddíl A

Frekvence v textu použitých znaků podle počtů prvků (bez duplicit)



Obrázek 10 Grafická vizualizace frekvence znaků podle počtu prvků, uvedené v Tabulce 25, oddíl B

Posledním faktorem, který by mohl narušit vztah nepřímé úměrnosti mezi znakem a parcelátem, by mohla být simplifikace čínského písma realizovaná v 50. a 60. letech minulého století. Opatření této reformy zredukovala počty tahů a následně i počty prvků téměř u 3 000 znaků, čímž do značné míry tento cílový soubor unifikovala. Do jaké míry ho unifikovala, je třeba zjistit komparací tradičních forem s jejich simplifikovanými variantami. Touto analýzou by se mohla zjistit míra redukce počtu tahů a počtu prvků u každého znaku zvlášť, čímž by se podařilo získat přesnou představu o rozsahu aplikovaných opatření. V otázce reformy jakožto jednoho z možných faktorů by pak bylo fundamentální zejména zjištění, jaké procentuální zastoupení tvoří v rámci této zjednodušené sady dvou prvkové a tři prvkové znaky. Testování vlivu reformy jakožto jednoho z faktorů neplatnosti vzájemného vztahu mezi parcelátem a znakem bude předmětem našeho dalšího výzkumu.

3.6.3 Jazyková úroveň U2

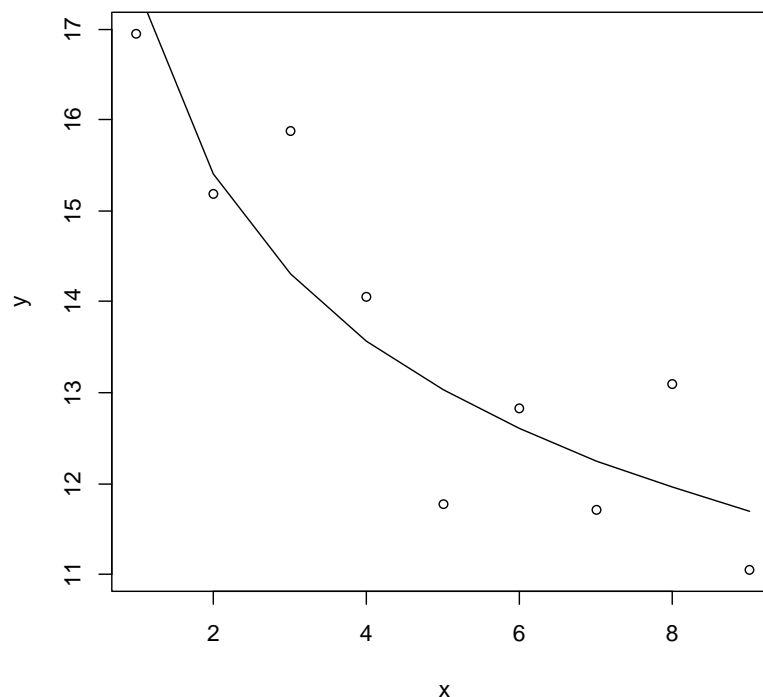
Souvětí – parcelát

Tabulka 26 uvádí empirické hodnoty získané kvantifikací vědeckého článku. První sloupec zahrnuje délky souvětí měřené v parcelátech x_2 , druhý sloupec četnosti jejich výskytu z_2 a třetí sloupec průměrné délky parcelátů měřených ve znacích y_2 .

Tabulka 26 Jazyková úroveň U2: souvětí (měřené v parcelátech) – parcelát (měřený v průměrném počtu znaků)

x_2	z_2	y_2
1	18	16,9444
2	27	15,1852
3	30	15,8778
4	16	14,0469
5	6	11,7667
6	6	12,8333
7	2	11,7143
8	5	13,1000
9	2	11,0556

Jak lze vidět z výše uvedené tabulky, celkem se na této úrovni vyskytlo devět délek konstruktů. Průměrné délky konstituentů se pohybují v rozpětí $\langle 11,06; 16,94 \rangle$ znaků a kromě tří případů vykazují s rostoucí délkou konstruktů klesající tendenci. Nárůst je pozorován u souvětí složeného ze tří ($y_2 = 15,88$), šesti ($y_2 = 12,83$) a osmi parcelátů ($y_2 = 13,10$). I přesto tyto odchylky vzájemný vztah nepřímé úměrnosti mezi souvětím a parcelátem definovaný MALEM platí. Na základě četností výskytu délek souvětí uvedených ve druhém sloupci lze konstatovat, že se v textu nejčastěji objevila dvou a tří parcelátová souvětí ($z_2 \geq 27$). Nejnižší frekvence je zaznamenána u souvětí tvořených sedmi a devíti ($z_2 = 2$) parceláty.



Obrázek 11 Grafická vizualizace pozorování uvedených v Tabulce 26

Na základě dat uvedených v Tabulce 27 můžeme konstatovat, že kladná hodnota parametru b splnila první předpoklad platnosti MALu na této úrovni. Podíváme-li se na grafickou vizualizaci vztahů mezi souvětím a parcelátem, zjistíme, že je křivka, znázorňující jejich vztah, klesající a konvexní. MAL tak na této úrovni představuje adekvátní a dobře sedící model. Jeho procentuální shoda s empiricky získanými pozorováními dosáhla vysokých hodnot, koeficient determinace R^2 činí téměř 79 % (viz Tabulka 27).

Tabulka 27 Parametr A , parametr b a koeficient determinace R^2 pro matematický model související s pozorováními uvedenými v Tabulce 26²⁰

Parametr A	Parametr b	Koeficient determinace R^2 (%)
17,4816	0,1827	78,62

²⁰ V případě této jazykové úrovně nebylo potřebné empiricky získaná pozorování statisticky ošetřit vynecháním nejméně frekventovaných pozorování.

Diskuze

Vraťme se nyní k nárůstu průměrných délek parcelátů. V případě souvětí tvořeného třemi parceláty lze stoupající tendenci spojovat s interpunkčními znaménky a jejich funkcemi. Jak lze vidět z Tabulky 26, souvětí s touto délkou se ve vědeckém textu vyskytlo celkem třicetkrát. Po vypočítání průměrné délky parcelátů u každého z nich se ukázalo, že 15 případů z 30 vykázalo vyšší průměrnou délku svých konstituentů y_2 , než jakou mají konstituenty souvětí tvořeného dvěma parceláty ($y_2 = 15,19$). Podíváme-li se na tato souvětí detailněji, zjistíme, že je můžeme na základě grafické struktury rozdělit do tří skupin. První skupinu představuje pět souvětí, u kterých nebyla nalezena žádná specifika. Do druhé náleží tři souvětí, u kterých se objevila čárka pro výčet. Protože se tento typ čárky používá pro vyjádření koordinačních vztahů mezi větnými členy (viz podkapitola 2.3), není mu přisouzena způsobilost vytvářet hranice mezi parceláty. Tímto opatřením se souřadně spojené větné členy stávají neoddělitelnou součástí parcelátů, čímž se jejich celková délka v počtu znaků zvyšuje. Největší vliv by však na nárůst průměrné délky konstituentů tří parcelátových souvětí měla mít třetí skupina, kam se řadí ta souvětí, která obsahovala uvozovky (“ ”), určené pro přímé citace, přímé řeči a důležité části textu, a závorky pro tituly (《 》), uvádějící názvy a jména. Obsahy těchto interpunkčních znamének značně navyšují počty znaků v parcelátech a tím i jejich celkové průměrné délky. Tato interpunkční znaménka se vyskytla celkem v sedmi případech, z nichž tři obsahovaly navíc čárku či čárky pro výčet.

Pro názornou ukázkou níže uvádím Tabulky 28, 29 a 30 s třemi příklady zmíněných souvětí, ve kterých se objevily kombinace výše uvedených interpunkčních znamének (čárka pro výčet, uvozovky a závorky pro tituly). V pravých částech tabulek jsou souvětí rozčleněna vždy na tři parceláty, nalevo jsou u každého z nich uvedeny počty znaků a průměrná délka. Znaménka jsou vyznačena červeně v rámečku stejné barvy.

Tabulka 28 Ukázka 1 souvětí tvořeného třemi parceláty s jejich výslednou průměrnou délkou v počtu prvků

Parceláty souvětí		Počet znaků
1.	2012 年 3 月 2 日，	6
2.	黄少安 □ 苏剑 □ 张卫国三位发表的 □《语言经济 学与中国的语言产业战略》□，	28
3.	基本上代表了我国学界在语言经济方面的认识。	20
Průměrná délka parcelátů		18

Zdroj: (Li, 2013), vlastní zpracování

Tabulka 29 Ukázka 2 souvětí tvořeného třemi parceláty s jejich výslednou průměrnou délkou v počtu prvků

Parceláty souvětí		Počet znaků
1.	国家语委全力支持山东大学 □ 南京大学 □ 广州 大学等高校的语言经济学研究，	31
2.	还于 2008 年 12 月 29 日支持商务印书馆成立了 □“中国语言资源开发应用中心”□，	30
3.	中心的宗旨是 □“致力于把语言及语言知识转化为 生产力和文化商品”□。	28
Průměrná délka parcelátů		29,67

Zdroj: (Li, 2013), vlastní zpracování

Tabulka 30 Ukázka 3 souvětí tvořeného třemi parceláty s jejich výslednou průměrnou délkou v počtu prvků

Parceláty souvětí		Počet znaků
1.	在 2010 年 2011 年北京市“两会”上，	10
2.	北京市人大代表贺宏志先生连续提出《关于发展我市语言产业的建议》和《加强语言文化建设，促进语言产业发展》的建议，	49
3.	语言经济的话题首次提到了地方人民代表大会的论坛上。	24
Průměrná délka parcelátů		27,67

Zdroj: (Li, 2013), vlastní zpracování

Co se týká nárůstu délky u šesti a osmi parcelátového souvětí, lze odchylku přisoudit jak výše zmíněným interpunkčním znaménkám, tak nízké frekvenci daných pozorování.

Na závěr je třeba uvést pravděpodobnou příčinu široké shody matematického modelu MALu s empiricky získanými pozorováními na této úrovni. V podkapitole věnované předchozí jazykové úrovni U3 jsme se zabývaly možnými příčinami malé procentuální shody s matematickým modelem MALu. Jako první příčinu jsme uváděly odlišný charakter jednotek, kde parcelát představuje jazykovou jednotku s proměnlivou délkou a znak jednotku s délkou konstantní, zcela nezávislou na uživateli čínského jazyka. U jazykových jednotek úrovně U2 – souvětí a parcelátu – se s tímto protikladem nesetkáváme. Délky obou jednotek jsou proměnlivé a uživatel je tvoří sám na základě své jazykové inteligence a vyjadřovacích schopností. Z důvodů eliminace konstantnosti délky konstituentu a stejného charakteru jazykových jednotek nedochází k narušení jejich vzájemného vztahu, což by mohlo mít pozitivní vliv na vysokou shodu s matematickým modelem na této úrovni.

3.6.4 Jazyková úroveň U1

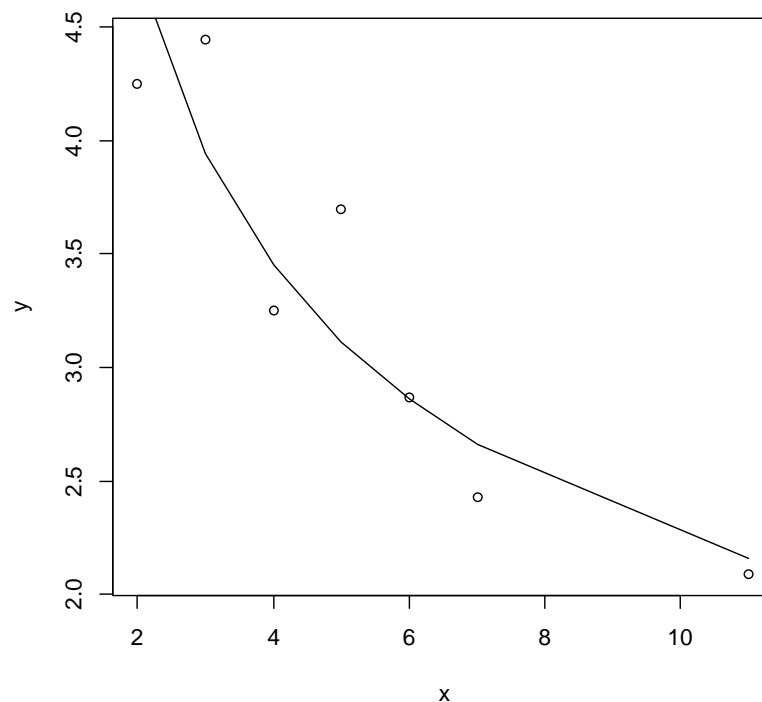
Odstavec – souvětí

Tabulka 31 uvádí empirické hodnoty získané kvantifikací vědeckého článku. První sloupec přísluší délkám odstavců měřených v souvětích x_1 , druhý sloupec četnostem jejich výskytu z_1 a třetí sloupec průměrným délkám souvětí měřených v parcelátech y_1 .

Tabulka 31 Jazyková úroveň U1: odstavec (měřený v souvětích) – souvětí (měřené v průměrném počtu parcelátů)

x_1	z_1	y_1
2	8	4,2500
3	3	4,4444
4	3	3,2500
5	4	3,7000
6	5	2,8667
7	2	2,4286
11	1	2,0909

Na této jazykové úrovni se vyskytlo celkem sedm délek konstruktů. Rozmezí, ve kterém se pohybují průměrné délky konstituentů, odpovídá intervalu $\langle 2,09; 4,44 \rangle$ parcelátů. Kromě dvou případů vykazují průměrné délky klesající tendenci. Nárůst se objevil pouze u odstavců tvořených třemi ($y_1 = 4,44$) a pěti ($y_1 = 3,70$) souvětími. I přesto můžeme konstatovat, že s rostoucí délkou odstavců klesá průměrná délka souvětí. Vztah nepřímé úměrnosti mezi konstruktem a konstituentem na této úrovni tedy platí. Co se týče frekvence, nejčastěji vědecký článek operoval s nejkratšími odstavci složenými ze dvou souvětí ($z_1 = 8$). Nejdelší odstavec o 11 souvětí má naopak četnost výskytu nejnižší – v textu se vyskytl pouze jednou.



Obrázek 12 Grafická vizualizace pozorování uvedených v Tabulce 31

I v případě poslední jazykové úrovně se kladná hodnota parametru b (Tabulka 32) projevila v grafické vizualizaci klesající konvexní křivkou. Je tedy zjevné, že vztah nepřímé úměrnosti mezi odstavcem a souvětím, předpokládaný MAlem, na této úrovni platí. Matematický model vykázal s empiricky získanými pozorováními širokou shodu – R^2 činí 85 % (viz Tabulka 32).

Tabulka 32 Parametr A , parametr b a koeficient determinace R^2 pro matematický model související s pozorováními uvedenými v Tabulce 31²¹

Parametr A	Parametr b	Koeficient determinace R^2 (%)
6,5604	0,4636	85,35

²¹ V případě této jazykové úrovně nebylo potřebné empiricky získaná pozorování statisticky ošetřit vynecháním nejméně frekventovaných pozorování.

Diskuze

Jak je patrné nejen z Tabulky 31, ale i z Obrázku 12, klesající tendenci průměrných délek souvětí y_1 narušila dvě pozorování. Jedná se o průměrné délky konstituentů u odstavců tvořených třemi a pěti souvětími. Z tohoto důvodu byla stejně jako u předchozí jazykové úrovně zjišťována průměrná délka jednotlivých souvětí u každého odstavce zvlášť, aby se ukázalo, která z nich způsobila nárůst průměrných délek konstituentů. U odstavců tvořených třemi souvětími hodnotu 4,25 – tedy hodnotu průměrné délky souvětí u odstavců tvořených o souvětí méně – přesáhly dva odstavce ze tří. Pro oba dva platilo, že obsahovaly středník. K podobnému zjištění došlo i u druhé odchylky. Ačkoliv hodnotu 3,25 (průměrnou délku souvětí u odstavců o čtyřech souvětích) převýšily všechny čtyři průměrné délky souvětí, dvě z nich s největším rozdílem mezi vlastní průměrnou délkou a hodnotou 3,25 se také vyznačovaly výskytem středníku.

Podle *Všeobecných pravidel pro použití interpunkce* středník představuje druh čárky, který se užívá buď pro oddělení těch vět, které jsou součástí souvětí a mezi nimiž existuje koordinační vztah, nebo se vkládá mezi věty prvních úrovní několikanásobného souvětí, které ve vztahu koordinace nejsou. (GB/T 15834 – 2011, 2012, s. 6) Čínská encyklopedie *Baidu Baike* uvádí středník také jako interpunkční znaménko, které stojí na pomezí čárky a tečky. (Fenhao, © 2013)

Zda by sporná funkce tohoto interpunkčního znaménka mohla mít vliv na nárůst průměrných délek souvětí v parcelátech, jsem se rozhodla ověřit dílčími experimenty, ve kterých bylo na středník nahlíženo jako na hranici souvětí. Tím, že souvětí vystupuje nejen coby konstituent jednotky bezprostředně vyšší, ale zároveň coby konstrukt jednotky bezprostřední nižší, bylo nutné provést experiment jak na jazykové úrovni odstavce – souvětí, tak i na hladině souvětí – parcelát. Experimenty jsou opět značeny dle příslušných úrovní (tj. 1 a 2) a příslušného výběrového souboru (tj. A). Získané výsledky uvádím níže.

EXPERIMENTY 1A A 2A – ZPŮSOBILOST STŘEDNÍKU ODDĚLOVAT SOUVĚTÍ

Tabulky 33 a 34 uvádí empirické hodnoty získané alternativní segmentací a kvantifikací vědeckého článku, ve kterém byl středník považován za hranici souvětí.

Tabulka 33 se týká dat získaných z jazykové úrovně odstavec – souvětí. V prvním sloupci jsou uvedeny délky odstavců měřených v souvětích x_1 , v druhém sloupci četnosti jejich výskytu z_1 a ve třetím sloupci průměrné délky souvětí měřených v parcelátech y_1 .

Tabulka 24 zahrnuje data získaná z jazykové úrovně souvětí – parcelát. První sloupec náleží délkám souvětí měřených v parcelátech x_2 , druhý sloupec četnostem jejich výskytu z_2 a třetí sloupec průměrným délkám parcelátů měřených ve znacích y_2 .

Tabulka 33 Experiment 1A – Jazyková úroveň U1: odstavec (měřený v souvětích) – souvětí (měřené v průměrném počtu parcelátů)

x_1	z_1	y_1
2	8	4,2500
3	1	3,6667
4	5	3,4000
5	2	3,5000
6	7	2,9762
7	2	2,4286
11	1	2,0909

Tabulka 34 Experiment 2A – Jazyková úroveň U2: souvětí (měřené v parcelátech) – parcelát (měřený v průměrném počtu znaků)

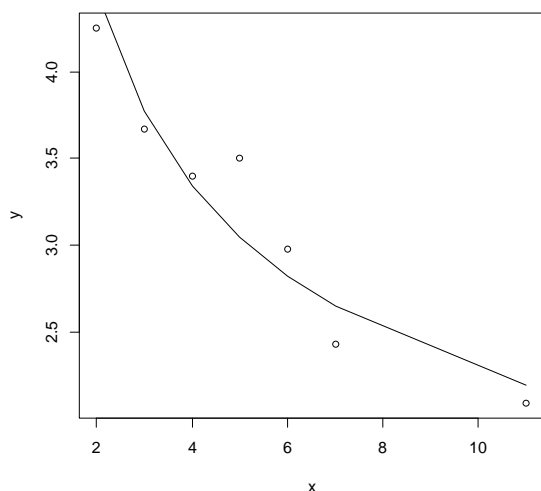
x_2	z_2	y_2
1	20	16,8500
2	36	14,8472
3	36	15,0741
4	20	13,4125
5	4	13,5500
6	5	12,8667
8	2	12,0000
9	2	11,0556

V případě jazykové úrovně U1 (odstavec – souvětí), jejíž empiricky získaná pozorování uvádí Tabulka 33, počet délek konstruktů zůstal stejný (7). U rozpětí průměrných délek konstituentů se změnila pouze nejvyšší hodnota. Protože se průměrná délka souvětí u odstavce tvořeného třemi souvětími alternativním způsobem segmentace snížila (ze 4,44 na 3,67). Konstituentem s nejvyšší průměrnou délkou se stalo souvětí u odstavce tvořeného dvěma souvětími ($y_1 = 4,25$). Ze sloupce náležejícího průměrným délkám konstituentů je patrné, že hodnoty vykázaly s rostoucí délkou konstruktů klesající tendenci téměř ve všech případech. K nárůstu došlo pouze u průměrné délky souvětí odstavce tvořeného pěti souvětími ($y_1 = 3,50$), a to pravděpodobně pod vlivem nízké četnosti výskytu. Na závěr lze

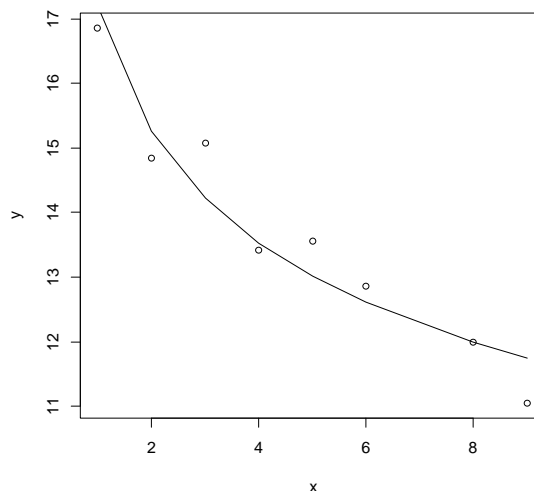
konstatovat, že i v případě alternativní segmentace se vzájemný vztah nepřímé úměrnosti mezi odstavcem a souvětím projevil.

Na jazykové úrovni U2 (souvětí – parcelát) došlo ke změnám jak u délek konstruktů, tak i u průměrných délek konstituentů. V prvním případě se rozdílnou segmentací eliminovalo souvětí tvořené sedmi parceláty, a proto se celkový počet délek konstruktů snížil o jednu délku. Nejvyšší průměrná délka parcelátů mírně poklesla (z 16,94 na 16,85), nejnižší zůstala stejná jako v případě výsledků získaných původním způsobem segmentace ($y_1 = 11,06$). Klesající tendence průměrných délek byla oproti původnímu experimentu narušena pouze dvakrát. Shoda byla v případě nárůstu průměrné délky parcelátu u souvětí tvořeného třemi parceláty. Jak již bylo řečeno výše, tento nárůst mohla způsobit interpunkční znaménka (uvozovky a zápornky pro tituly), neboť obsah, který uvádí, značně navyšuje délku parcelátů. Druhý nárůst je vyzorován u průměrné délky parcelátů u pěti parcelátového souvětí ($y_1 = 13,55$), původně se jednalo o průměrnou délku parcelátů u šesti parcelátového souvětí. Zde by mohl klíčovou roli sehrát značný pokles četnosti výskytu. Přesto se i na této úrovni vzájemný vztah – čím větší je souvětí v počtu parcelátů, tím menší jsou v průměru jeho parceláty v počtu znaků – projevil.

Oba způsoby segmentace, založené na rozdílných přístupech ke středníku, vykazaly na obou jazykových úrovních klesající tendence průměrných délek konstituentů, a tedy i vzájemné vztahy nepřímé úměrnosti mezi konstrukty a konstituenty. Odchylek od klesající tendence bylo v případě alternativního přístupu méně. Do jaké míry tedy ovlivnila způsobilost středníku oddělovat souvětí, uvidíme z následných grafických vizualizací a zejména z procentuální shody s matematickým modelem MALu vyjádřeným koeficientem determinace R^2 .



Obrázek 13 Grafická vizualizace pozorování uvedených v Tabulce 33



Obrázek 14 Grafická vizualizace pozorování uvedených v Tabulce 34

Oba parametry b mají kladnou hodnotu (viz Tabulky 35 a 36) a tím splňují předpoklad platnosti MALu. Klesající a konvexní tendence křivek, znázorňující vztah mezi délkami veličin (odstavec a souvětí; souvětí a parcelát) předpokládaný MAlem, se u obou jazykových úrovní projevyly, jak lze ostatně vidět z grafických vizualizací pozorování získaných alternativním způsobem segmentace. Co se týče shody s matematickým modelem, koeficienty determinace dosáhly vyšších hodnot než v původním experimentu. V případě jazykové úrovně U1 (odstavec – souvětí) se R^2 navýšil z původních 85,35 % na 90,41 %, u jazykové úrovně U2 (souvětí – parcelát) došlo dokonce k výraznějšímu nárůstu – z původních 78,62 % na 91,73 % (viz Tabulky 35 a 36).

Tabulka 35 Parametr A , parametr b a koeficient determinace R^2 pro matematický model související s pozorováními uvedenými v Tabulce 33²²

Parametr A	Parametr b	Koeficient determinace R^2 (%)
5,9661	0,4179	90,41

²² V případě této jazykové úrovně nebylo potřebné empiricky získaná pozorování statisticky ošetřit vynecháním nejméně frekventovaných pozorování.

Tabulka 36 Parametr A , parametr b a koeficient determinace R^2 pro matematický model související s pozorováními uvedenými v Tabulce 34²³

Parametr A	Parametr b	Koeficient determinace R^2 (%)
17,2152	0,1737	91,73

Z výsledků získaných segmentací na základě způsobilosti středníku oddělovat souvětí je patrné, že interpunkční znaménko svou funkcí spíše inklinuje k tečce. Otázkou je, zda je funkce středníku natolik sporná, nebo zda se nejedná o vliv samotné interpunkce, jejíž užívání má v čínských textech poměrně krátkou tradici.

Integrovanou součástí psaného jazyka se interpunkce stala až ve 20. století, do té doby se v textech objevovala sporadicky a mnohdy měla podobu specifických symbolů, jejichž funkce byla podobná funkcím západních interpunkčních znamének (více viz Biaodian fuhao, © 2013). Teprve až v druhé dekádě 20. století se začaly formovat snahy učenců zavést v čínských textech interpunkci, vycházející z interpunkce západní. První dílo využívající interpunkční znaménka bylo vydáno v roce 1919. Jednalo se o Nástin historie čínské filozofie (Zhōngguó zhéxuéshǐ dàgāng, 中国哲学史大纲) od čínského filozofa a spisovatele Hu Shiho (Hú Shì, 胡适), (Biaodian fuhao, © 2013). V roce 1951 vydaly vládní orgány dokument, který shrnuje pravidla pro používání interpunkčních znamének. Protože v následujících čtyřech dekádách prošla interpunkce ještě dalšími vývojovými změnami, byla Všeobecná pravidla pro interpunkci (Biāodiǎn fúhào yòngfǎ, 标点符号用法) z roku 1951 vládními orgány aktualizována a znovu vydána roku 1990. Více viz (Biaodian fuhao yongfa, © 2013).

Vliv interpunkce na vzájemné vztahy jazykových jednotek je nutné dále testovat na dalších výběrových souborech. Z tohoto důvodu navrhuji provést experimenty, které by se zejména zaměřily na vliv interpunkce v kontextu vývoje.

Stejně tak samotnou jazykovou jednotku odstavec je potřeba podrobit dalším kvantitativním analýzám především těch souborů, které by byly delší, než jsou námi testované texty.

²³ V případě této jazykové úrovně nebylo potřebné empiricky získaná pozorování statisticky ošetřit vynecháním nejméně frekventovaných pozorování.

3.6.5 Shrnutí

Shrňme si na závěr našeho experimentu výsledky získané kvantifikací a statistickou analýzou vědeckého článku. Platnost (příp. neplatnost) MALu jsme ověřovaly celkem na čtyřech jazykových úrovních. Nejnižší jazyková hladina, tvořená znakem na pozici konstruktů a prvkem na pozici konstituentů, potvrdila, že mezi znakem a prvkem existuje vzájemný vztah předpokládaný MALem. Nejenže je MAL na této úrovni adekvátním a dobře sedícím matematickým modelem, ale zejména jeho výsledná procentuální shoda s empiricky získanými pozorováními dosahuje extrémně vysokých hodnot, $R^2 = 90,61\%$ ²⁴. Oproti tomu následující jazyková hladina reprezentovaná parcelátem (konstruktem) a znakem (konstituentem) nevykázala shodu žádnou. MAL, resp. vztah mezi délkami veličin jím definovaný, v případě parcelátu a znaku neplatí. U souvětí a parcelátu, které jsou konstruktem a konstituentem další jazykové úrovně, se existence vztahu nepřímé úměry mezi délkami jazykových jednotek potvrdila. Matematický model vykázal s pozorováními na této úrovni širokou shodu. Koeficient determinace R^2 činí 78,62%. Jako adekvátní a dobře sedící model se oproti našemu očekávání MAL projevil i v případě nejvyšší a námi jako poslední testované jazykové úrovně odstavec (na pozici konstruktů) – souvětí (na pozici konstituentů). Nejenže byla potvrzena platnost vzájemného vztahu mezi těmito jednotkami, ale především procentuální shoda opět dosáhla velmi vysokých hodnot – $R^2 = 85,35\%$.

Platnost MALu byla potvrzena celkem na třech jazykových úrovních: znak – prvek, souvětí – parcelát a odstavec – souvětí. Jedinou jazykovou hladinou, jejíž empiricky získaná pozorování nevykázala s matematickým modelem tohoto jazykového zákona žádnou shodu, byl parcelát – znak. Náš předpoklad MALu jakožto adekvátního a dobře sedícího modelu na jazykových hladinách znak – prvek a souvětí – parcelát jsme experimentem realizovaným na vědeckém článku potvrdily. V případě teze, podle které se MAL neměl projevit na jazykových hladinách parcelát – znak a odstavec – souvětí, však došlo pouze k jejímu částečnému potvrzení. Analýza jazykové úrovně odstavec – souvětí u vědeckého článku prokázala, že mezi těmito jazykovými jednotkami existuje vzájemný vztah, čímž naši tezi v případě této

²⁴ Bez statistického ošetření

hladiny vyvrátila. Na jazykové úrovni parcelát – znak jsme naopak potvrdily, že mezi délkami těchto veličin vzájemný vztah neplatí.

V mé části, zabývající se kvantitativní analýzou vědeckého článku, jsem rovněž provedla v rámci jazykových úrovní dva dílčí experimenty. Jedná se jednak o analýzu tradičních znaků, a jednak o analýzu interpunkčního znaménka – středníku. Předmětem prvního experimentu, který jsem uskutečnila na nejnižší jazykové hladině znak – prvek, se stal identický výběrový soubor, jehož zjednodušené varianty znaků byly transformovány do tradiční podoby. Následná kvantifikace a statistická analýza opět potvrdily existenci vzájemného vztahu mezi těmito jazykovými jednotkami a tím i platnost MALu. Procentuální shoda s matematickým modelem tohoto zákona dosáhla z celého experimentu nejvyšší hodnoty – $R^2 = 95,38$. Předmětem druhého experimentu, který byl proveden na úrovni souvětí – parcelát a odstavec – souvětí, se stal rovněž stejný článek, u něhož však došlo ke změnám původně zvoleného segmentačního kritéria jazykové jednotky souvětí – interpunkčnímu znaménku středník byla přisouzena způsobilost oddělovat souvětí. Výsledky získané touto alternativní segmentací platnost MALu nevyvrátily, ba naopak se dokonce ukázalo, že se tato alternativa jeví jako vhodnější. U obou úrovní koeficient determinace přesáhl 90 %.

KVANTITATIVNÍ ANALÝZA SOUČASNÉ PSANÉ ČIŠTINY

Aplikace Menzerath-Altmannova zákona
na umělecký styl

Autor: Lenka Spáčilová

4 Aplikace Menzerath-Altmanova zákona na článek z blogu

Jak již napovídá název kapitoly, tato část práce se bude zabývat kvantitativní analýzou uměleckého stylu, který v tomto případě bude reprezentován článkem z blogu. Stejně jako spoluautorka Tereza Motalová budu pro ověřování hypotéz aplikovat Menzerath-Altmanův zákon.

Tento úsek práce bude rozdělen do šesti částí. V prvních pěti podkapitolách uvedu jednotlivé kroky experimentu. Nejprve si podle daných kritérií zvolím vhodný výběrový soubor a odůvodním jeho výběr. Z důvodu posloupnosti experimentu zde dále představím jazykové jednotky a jazykové úrovně, které vznikly uspořádáním jednotek do vzájemných vztahů. Pro každou stanovenou jazykovou úroveň vytvořím tabulky, ze kterých získám potřebné hodnoty. Tyto hodnoty budou sloužit k výpočtu parametrů a koeficientů determinace a také k sestrojení grafů. Uvedené kroky provedu za pomoci statistického softwaru R. Poslední šestá podkapitola se zabývá interpretací získaných dat a představuje jádro této kapitoly.

4.1 Stanovení kritérií pro volbu výběrového souboru a jejich odůvodnění

V podkapitole 2.1 již bylo řečeno, že při volbě výběrového souboru je nutné dodržet kritéria, která jsou stanovena pro výběr vhodných textů. Je to hlavně z toho důvodu, aby experimenty v rámci této diplomové práce (DP) probíhaly stejnou formou a jejich výsledky bylo možné porovnat. Stanovená kritéria je však nutné dodržet nejen z důvodu porovnání výsledků mezi sebou navzájem, ale také kvůli možnosti komparace s výsledky, které jsme se spoluautorkou Motalovou již získaly na základě provedených experimentů v rámci projektu Studentské grantové soutěže IGA (FF_2012_035), na kterém jsme pod odborným vedením Mgr. Martiny Benešové Ph.D. spolupracovaly a na který tato práce úzce navazuje.

Požadavkem, na který je kladen hlavní důraz a ze kterého vychází i další kritéria, je zaměření experimentu na zkoumání čínského jazyka ze synchronního pohledu. V následujících odstavcích krátce připomenou, o která kritéria se jedná. Pro přehlednost je uvádím v pořadí, jak je uvedeno v podkapitole 2.1.

1. Současná moderní čínština

Moderní čínština je poslední vývojovou fází čínského jazyka. Její počátek bývá datován do druhé dekády 20. století (po Májovém hnutí v roce 1919). Přestože se jedná o poslední vývojovou fází čínštiny, toto období zahrnuje dlouhý časový úsek, ve kterém se čínština stále vyvíjela a prošla mnoha změnami. Proto by bylo neúplné tvrdit, že experiment bude zaměřen na moderní čínštinu. Je nutné přesně vymežit, jakým vývojovým stadiem se bude výzkum zabývat. S ohledem na synchronní hledisko výzkumu je prvním aplikovaným kritériem zvolit text, který bude psán standardní formou současné moderní čínštiny.

2. Psaná čínština – zjednodušené znaky

Cílem experimentu je ověřit platnost případně neplatnost MALu na psanou čínštinu, a proto je potřebné, aby výběrový soubor byl napsán v čínských znacích. I v tomto bodě je možné volit mezi více variantami (viz podkapitola 2.1). Opět je však nutné přihlédnout k synchronnímu aspektu experimentu a zvolit vývojově nejmladší typ znaků – tzn. zjednodušené znaky. Dalším kritériem tedy je vybrat text psaný ve zjednodušených znacích. Z toho rovněž vyplývá, že autor musí pocházet z pevninské Číny a používat zjednodušenou sadu znaků.

3. Jazykové styly

Volba jazykových stylů nám nechává poměrně širokou možnost výběru, neboť je možné zvolit jakýkoli současný jazykový styl. V rámci výzkumu této diplomové práce je ovšem důležité vybrat dva texty psané různými jazykovými styly, aby mohla být provedena vzájemná komparace výsledků. V předešlém výzkumu (2012), na který tato DP navazuje, jsme analýze podrobily novinový a literární styl. Proto je žádoucí zvolit takové jazykové styly, které se od nich budou lišit. Hlavním důvodem je možnost ověřit hypotézy na větším množství výběrových souborů. Stále je však třeba dbát, aby vybraný text splňoval ostatní kritéria.

4. Aktuálnost

Další kritérium opět vychází ze synchronního zaměření experimentu. Výběrové soubory, které budou analyzovány, musí být co nejaktuálnější a odrážet tak soudobý jazyk. Text by tedy měl být co nejsoučasnejší a neměl by být starší deseti let, tzn. vydaný od roku 2003.

5. Souvislost výběrových souborů

Výběrové soubory musí mít zřetelný začátek a konec a nesmí být přerušeny např. obrázkem, grafem nebo tabulkou. Pouze takovéto texty můžeme označit za souvislé útvary a zvolit je jako vhodný výběrový soubor, viz (Hřebíček, 2002).

6. Délka výběrových souborů

Dalším kritériem je vybrat takové texty, které budou mít vhodnou délku. Výběrové soubory nesmí být příliš krátké ani příliš dlouhé. Na základě již provedených experimentů byla délka textů empiricky stanovena v rozmezí 2500 – 3500 čínských znaků.

7. Renomé autora

Poslední kritérium souvisí s renomé autora. Populárnější autoři jsou více čtení, a proto mají větší vliv na současný jazyk. Tím, že některá slova používají častěji, mohou ovlivňovat slovní zásobu čtenářů a ve výsledku i celkovou frekventovanost slov. Proto je třeba vybrat autora, o kterém převládá obecné povědomí v kruzích laické veřejnosti.

4.2 Volba výběrového souboru

Při hledání výběrového souboru bylo nejprve nutné zvolit, jakým literárním stylem bude výběrový soubor napsán. Mým úmyslem bylo vybrat text psaný v takovém jazykovém stylu, který je čtený širokými vrstvami, a zároveň se jedná o současný text. Na základě těchto podmínek jsem se k výzkumu rozhodla použít článek z blogu.

Prvním krokem bylo zvolit si současného čínského blogera. Při výběru nejvhodnějšího autora jsem si stanovila několik podmínek. Nejednalo se tedy o zcela náhodný výběr, ale k volbě jsem dospěla až po zvážení více hledisek. Hledala jsem takového autora, jehož slovní zásoba je co nejméně ovlivněná cizími jazyky. Proto jsem brala v úvahu, zda autor neabsolvoval studium v zahraničí nebo tam dlouhodobě nepobýval. Další podmínkou byl, aby jeho povolání nesouviselo s žurnalistikou, což by mohlo ovlivňovat styl jeho psaní. Cílem bylo najít takový styl, který se co nejvíce blíží stylu uměleckému.

Jako zdroj pro obecný přehled mi posloužil seznam čínských bloggerů na stránkách wikipedie (Category: Chinese bloggers, 2013). O každém uvedeném autorovi jsem si na čínské encyklopedii *baidubaike* (bǎidùbǎikē, 百度百科) nejprve vyhledala základní údaje, a pokud zkoumaný autor vyhovoval kritériím, která jsem si stanovila, prohlédla jsem si také jeho blog, u nějž jsem zjišťovala návštěvnost. Bližšímu průzkumu jsem podrobila blogery, kteří jsou uvedeni v následující tabulce (viz Tabulka 37):

Tabulka 37 Současní čínští blogeři

Jméno blogera	Znaky	Návštěvnost blogu²⁵	Odkaz
Ai Weiwei	艾未未	4 467 128	http://www.bullogger.com/blogs/aiww/
Han Han	韩寒	590 523 080	http://blog.sina.com.cn/s/blog_4701280b0102e7er.html
Kong Qingdong	孔庆东	79 035 813	http://blog.sina.com.cn/u/198367585
Li Chengpeng	李承鹏	7044 788	http://www.weibo.com/lichengpeng
Murong Xuecun	慕容雪村	3 393 579	http://blog.sina.com.cn/hawkking
Muzi Mei	木子美	2 477 349	http://muzimeiriji.blog.sohu.com
Ran Yunfei	冉云飞	2 336 987	http://tufeilaoran.blog.163.com/
Rao Xueman	饶雪漫	1 608 917	http://weibo.com/raoxueman
Liu Mangyan	流氓燕	3 881 102	http://blog.tianya.cn/blogger/blog_main.asp?BlogID=19329

²⁵ Počet zhlédnutí blogu k 17. dubnu 2013 (18:40)

Sima Nan	司马南	632 202	http://weibo.com/simanan
Vivibear / Zhang Weiwei	张薇薇	28 440763	http://blog.sina.com.cn/vikimgbear333
Wang Keqin	王克勤	7 521701	http://wangkeqin.blog.sohu.com/
Xu Jinglei	徐静蕾	312017085	http://blog.sina.com.cn/xujinglei
Zeng Jinyan	曾金燕	–	http://zengjinyan.wordpress.com/
Zhou Shuguang	周曙光	–	https://www.zuola.com/

Zdroj: vlastní zpracování

Z uvedených autorů se jako nejvhodnější jevil blogger Han Han (Hán Hán, 韩寒), protože splňuje všechna daná kritéria a jeho blog navíc navštívilo nejvíce čtenářů (o autorovi viz níže). Za výběrový soubor jsem se tedy rozhodla zvolit jeden z článků uveřejněných na jeho blogu.

Co se týče článku, při jeho výběru jsem vycházela z dalšího kritéria, které souvisí s aktuálností výběrového souboru. Bylo nutné vybrat co nejsoučasnější text, proto jsem při volbě postupovala od nejnověji přidaných článků a zároveň jsem se zaměřovala na takové články, které našemu experimentu vyhovují svou délkou (tzn. 2500 až 3500 znaků). Mimo tato kritéria jsem brala na zřetel počet zhlédnutí jednotlivých příspěvků a snažila se zvolit článek s vysokým počtem zhlédnutí.

S ohledem na dané podmínky jsem za výběrový soubor zvolila článek *Život, jak mu rozumím* (wǒ suǒ lǐjiě de shēnghuó, 我所理解的生活), (Han, 2012), neboť splňuje všechna stanovená kritéria. Článek vyšel 20. června 2012, je psán ve zjednodušených znacích a obsahuje 2 641 znaků. Text má jasně vyznačený začátek i konec a nejsou do něj vloženy žádné obrázky či grafy, proto jej můžeme označit za souvislý. Podle informací uvedených na blogu k 22. březnu 2013 tento text již vidělo 925 631 čtenářů. Tento výběrový soubor je vložen do příloh pod označením vzorek B (viz Příloha 2).

Han Han (Hán Hán, 韩寒)

Han Han je oblíbený čínský spisovatel, blogger, profesionální automobilový závodník, hudebník a byl také zakladatel a hlavní editor literárního časopisu *Soubor sólistů* (Dúchàng tuán, 独唱团), který začal vycházet v roce 2010, avšak do

prodeje se dostalo pouze první číslo (vydání druhého čísla se dlouho odkládalo a nakonec k němu vůbec nedošlo (Duchang tuan, ©2013).

Han Han se narodil 23. září 1982 v Šanghaji a již na druhém stupni základní školy začal publikovat svá první díla (Han Han, ©2013). Na literární scénu prorazil v roce 2000 s bestsellerem *Trojitá brána* (*Sānchóng mén*, 三重门), (Elegant, 2009). V roce 2005 založil svůj blog, který se v roce 2009 stal nejpopulárnějším v Číně (Elegant, 2010). Han Han na svůj blog přidává uštěpačné komentáře, které kritizují vládu a současné poměry v Číně, jako je např. korupce, cenzura a zneužívání moci. Avšak snaží se vyhýbat přímému útoku na vládu, který by pro něj mohl znamenat cenzuru. Při kritice používá spíše jemné narážky a satiru (Abrahamsen, 2012). Han Han se řadí ke skupině spisovatelů mladé generace označované jako mládež, která se narodila po 80. letech (the post-80s generation). Vydal mnoho prozaických sbírek a románů. Jeho díla obvykle vyvolávají velký zájem a bouřlivé diskuze. V roce 2010 byl časopisem *Time* zařazen mezi sto nejvlivnějších osobností planety. Jeho blog doposud zhlédlo přes 590 milionů návštěvníků a tím patří k nejčtenějším blogům nejen v Číně, ale také na světě (Pilling, 2012).

Život, jak mu rozumím (wǒ suǒ lǐjiě de shēnghuó, 我所理解的生活)

Článek *Život, jak mu rozumím* je předmětem tohoto výzkumu, proto je na místě nejprve přiblížit jeho strukturu a další zvláštnosti, které by mohly ovlivňovat výsledky experimentu. Jak již bylo řečeno výše, tento článek je souvislý text a kromě hlavního nadpisu, jména autora a internetového odkazu se v článku nevyskytují žádná přerušení, jako jsou například obrázky či grafy.

Článek se skládá z 2 641 znaků (údaj s interpunkčními znaménky). Je organizován do 12 odstavců, které jsou v textu jasně vyznačeny. Každý odstavec začíná odskokem od okraje a jednotlivé odstavce jsou navíc mezi sebou odděleny prázdným řádkem.

Autor pochází z pevninské Číny, text je proto napsán v zjednodušených znacích. V současné době se na čínsky psaných blozích a fórech často objevují anglické výrazy a jejich autoři používají k zápisu latinku (o tomto tématu např. viz Sikora, 2011). V případě tohoto článku však Han Han nepoužívá žádná cizí slova a k použití čínských znaků se přiklání dokonce i v tom případě, kdy je možné použít

variantní zápis formou latinských písmen. Např. vulgární slovo „snobský bastard“ (zhuāngbī, 装逼) píše pomocí dvou čínských znaků místo variantního výrazu 装 B (čtvrtý odstavec), stejně tak je tomu i v případě dalšího vulgárního slova „zasraně skvělý“ (niúbī, 牛逼), které lze zapsat jako 牛 B (desátý odstavec).

Co se týče čísel, autor se ve většině případů uchyluje k používání čínských znaků a arabské číslice volí jen zřídka. Jedinou výjimku tvoří číslo 30, které je zapsáno pomocí arabských číslic (v desátém odstavci) a je použito ve spojení „třicet let“ (sānshí nián, 30 年). V tomto případě je číslo 30 formátováno druhým způsobem, u kterého se číslice k sobě těsnají (viz podkapitola 2.3), proto je chápáno jako jeden znak.

V textu se vyskytují pouze 4 druhy čínských interpunkčních znamének, a to čárka, tečka, vykřičník a dvojtečka (viz Tabulka 38). Z článku se jeví, že se autor pravděpodobně snaží volit jednodušší strukturu textu, a proto používá pouze základní interpunkci. Dokonce i v částech, kde popisuje rozhovor se svým kamarádem, nepoužívá uvozovky, které by jasněji vyznačovaly přímou řeč.

Tabulka 38 Použitá interpunkční znaménka

Český název	Interpunkční znaménko	Pinyin	Znaky
Čárka	,	dòuhào	逗号
Tečka	。	jùhào	句号
Otazník	?	wèn hào	问号
Dvojtečka	:	màohào	冒号

Zdroj: čínská terminologie převzata z (Biaodian fuhao, © 2013); vlastní zpracování

4.3 Stanovení a definování jazykových jednotek a úrovní

Precizní stanovení jazykových jednotek je fundamentálním krokem pro každý podobný experiment. Jelikož náš výzkum navazuje na experiment (2012), ve kterém jsme již jazykové jednotky přesně vymezily, nebylo žádoucí určovat nové jazykové jednotky. Důležité ovšem bylo, aby jejich vymezení bylo v obou experimentech důsledně dodržováno, a to z důvodu možné komparace. Přesné určení jednotlivých jazykových jednotek je uvedeno v podkapitole 2.3. Z důvodu poslušnosti a

ucelenosti experimentu níže opět stručně zopakují, o jaké jednotky se jedná. Zmíním zejména ta vymežující pravidla, která se týkají mnou zkoumaného článku. Stejně jako v ostatních výběrových souborech byly použity následující jazykové jednotky:

tah – prvek – znak – parcelát – souvětí - odstavec

1. Tah (bǐhuà, 笔画)

Tah představuje nejmenší grafickou jednotku čínského znaku. Každý znak se skládá z jednoho nebo více tahů. Existují různá dělení této jazykové jednotky, co se týče její charakteristiky nebo terminologie, avšak pro naši segmentaci toto dělení nemá příliš velký vliv. Zásadní vlastností tahu pro náš experiment je, že se vždy píše najednou (tj. bez přerušení) a tím lze odlišit od ostatních vyšších jazykových jednotek.

2. Prvek (bùjiàn, 部件)

Prvek v případě našeho experimentu představuje tzv. ostrůvek, což je prostorem oddělená část znaku. Může být tvořen jedním nebo více navzájem se dotýkajícími tahy.

3. Znak (hànzì, 汉字)

Znak je základní jednotkou čínského písma. Každý znak se píše do jednoho grafického pole nehledě na počet tahů. Ve zkoumaném článku je za jeden znak považována i arabská číslice.

4. Parcelát

Parcelát je jazyková jednotka, kterou jsme zvolily a nazvaly pro účely předchozího experimentu (2012). Její hranice vymežují vybraná interpunkční znaménka (viz podkapitola 2.3). Při určování této a následující jazykové jednotky tedy bylo nutné kromě grafického principu aplikovat i syntaktické hledisko. V mém případě hranice parcelátů tvoří interpunkční znaménka uvedená v Tabulce 38.

5. Souvětí (fùjù, 复句)

Souvětí představuje tu část textu, která se skládá z jednoho a více parcelátů. Hranice souvětí tvoří stejně jako v předchozím případě vybraná interpunkční znaménka. V článku z blogu jsou těmito znaménky: tečka 。 (jùhào, 句号) a otazník ? (wèn hào, 问号). Ostatní znaménka jsou platná pro nižší jazykovou jednotku – parcelát.

6. Odstavec (duànluò, 段落)

Poslední jazykovou jednotkou určenou výhradně na základě grafické podoby je odstavec. Odstavec je ta část textu, jejíž začátek je tvořen odskočením od okraje stránky. Jednotlivé odstavce jsou mezi sebou odděleny prázdným řádkem.

Propojením výše uvedených jazykových jednotek získáme čtyři jazykové úrovně (hladiny), na kterých budeme zkoumat platnost MALu. Jazykovou úroveň označíme jako i , kde $i = 1$ představuje jazykovou úroveň odstavec – souvětí, $i = 2$ jazykovou úroveň souvětí – parcelát, $i = 3$ jazykovou úroveň parcelát – znak a $i = 4$ jazykovou úroveň znak – prvek; i tedy může nabývat hodnot $i = 1, 2, 3, 4$. Dolní index označující číslo úrovně spolu s indexem j značí alternativu experimentu, $j = B, C$, kdy $j = B$ značí vzorek B (článek z blogu) a $j = C$ značí vzorek C (delší článek z blogu). Označení dolního indexu j je uvedeno v příslušných podkapitolách.

Tabulka 39 Jazykové úrovně U_i , x_i konstrukt, y_i konstituent, ($i = 1,2,3,4$)

Jazyková úroveň U_i	Konstrukt x_i ; konstituent y_i		Délka
U4	x_4	znak	v prvcích
	y_4	prvek	průměrná délka měřená v tazích
U3	x_3	parcelát	ve znacích
	y_3	znak	průměrná délka měřená v prvcích
U2	x_2	souvětí	v parcelátech
	y_2	parcelát	průměrná délka měřená ve znacích
U1	x_1	odstavec	v souvětích
	y_1	souvětí	průměrná délka v parcelátech

4.4 Segmentace a kvantifikace výběrového souboru

Předmětem kvantifikace této části práce bude článek z blogu *Život, jak mu rozumím*. Kvantifikace bude provedena na celém textu s výjimkou hlavního nadpisu a náležitostí jako jsou údaje o vydání, jméno autora a internetový odkaz. Ve srovnání se spoluautorkou, jejíž výběrový soubor má složitější strukturu, článek z blogu není dále vnitřně členěn. Skládá se pouze z jasně oddělených odstavců dále nedělených a kromě již zmíněné číslovky 30, která je zapsána arabskými číslicemi, se v textu neobjevila žádná odchylka.

Na základě stanovených jednotek jsem v programu Microsoft Excel zvolený článek segmentovala a ve stejném programu jsem poté přistoupila ke kvantifikaci textu. Tímto krokem jsem získala tabulky s potřebnými daty pro x_i (délka konstrukt) z_i (frekvence) a y_i (délka konstituentu). Segmentace a kvantifikace proběhla na všech zmíněných jazykových úrovních U4 – U1.

4.5 Testování spolehlivosti modelu pomocí statistických metod

Hodnoty z tabulek, která jsem získala po kvantifikaci článku, jsem použila jako vstupní data v statistickém softwaru R. Na jejich základě bylo možné vypočítat parametr A , parametr b a koeficient determinace R^2 a sestavit grafy právě za využití tohoto softwaru. Pro výpočet parametru A , parametru b a koeficientu determinace R^2 jsem zvolila metodu jednoduché lineární regrese (viz podkapitola 2.5, poznámka pod čarou č. 10), která byla aplikována na zkrácenou formuli MALu.

Pozorování, která jsou z důvodu nízké četnosti jejich výskytu v porovnání s ostatními zanedbatelná (tzv. extrém), jsou na každé jazykové úrovni statisticky ošetřena. Hodnoty získané jejich odebráním jsou vždy uvedeny v poznámce pod čarou vztahující se k tabulce uvádějící parametr A , parametr b a koeficient determinace R^2 .

4.6 Interpretace získaných dat

V následující části budu interpretovat získaná kvantitativní data, tabulky a grafické vizualizace. Jedná se o velmi důležitou část experimentu, a proto jí bude věnován největší prostor diplomové práce. Při interpretaci budu postupovat od nejnižší jazykové úrovně U4 (znak – prvek). Výsledky a závěry získané ze všech čtyř jazykových úrovní budou shrnuty na konci této podkapitoly.

4.6.1 Jazyková úroveň U4

Znak – prvek

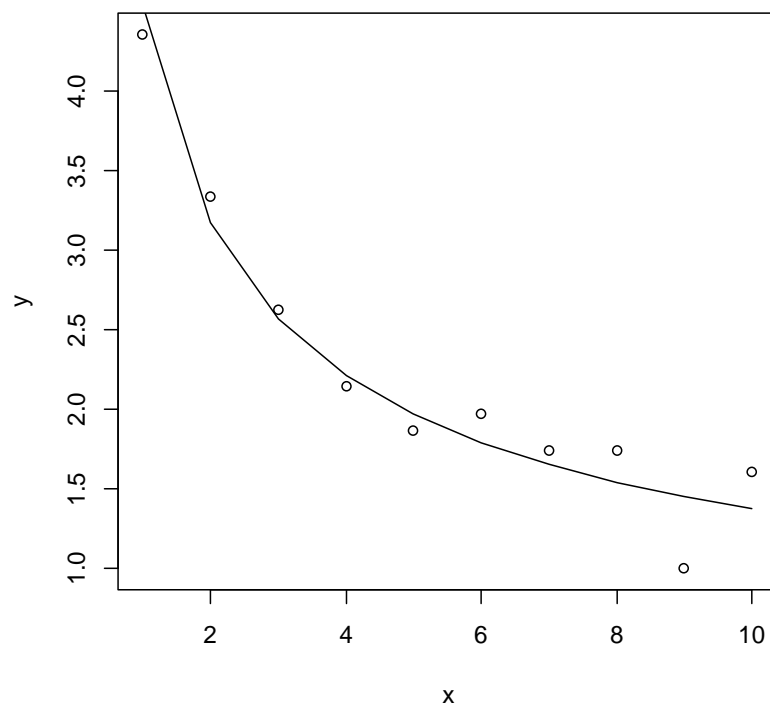
Tabulka 40 uvádí empirické hodnoty získané kvantifikací článku z blogu. Konstrukt x_4 představuje délku znaku (měřenou v prvcích), z_4 je jejich frekvence a konstituent y_4 představuje průměrnou délku prvků (měřenou v tazích). Empirická pozorování s nízkou četností výskytu ($z_4 = 1$) jsou v tabulce vyznačena šedým pozadím.

Tabulka 40 Jazyková úroveň U4: znak (měřený v prvcích) – prvek (měřený v průměrném počtu tahů)

x_4	z_4	y_4
1	522	4,3544
2	670	3,3366
3	479	2,6200
4	323	2,1416
5	224	1,8652
6	58	1,9684
7	69	1,7329
8	14	1,7411
9	1	1,0000
10	5	1,6000

Jak je patrné z Tabulky 40, ve výběrovém souboru se vyskytují znaky (konstrukty) složené z 1 až 10 prvků. Průměrná délka prvků (konstituentů) se pohybuje v rozmezí $(1,00; 4,35)$. Z uvedených dat je očividné, že s rostoucí délkou konstruktů klesá průměrná délka konstituentů, tím je naplněn předpoklad MALu o nepřímé závislosti konstruktů a jeho konstituentů.

Nejfrekventovanější znaky jsou tvořeny dvěma prvky, dále to jsou znaky složené z jednoho a třech prvků. Čtyř prvkové znaky jsou také v poměrně velkém zastoupení. Co se týče znaků složených z pěti a více prvků, jejich frekvence výskytu má klesající tendenci. Nejnižší četnost výskytu vykazovala pozorování $x_4 = 9$ a v Tabulce 40 je vyznačeno šedým pozadím.



Obrázek 15 Grafická vizualizace pozorování uvedených v Tabulce 40

Z grafické vizualizace je zřejmé, že klesající tendence formulovaná MAlem se na této jazykové úrovni projevila a matematický model navíc vykazuje velmi vysokou shodu, přesahuje 85 %. Předpoklad pro splnění MALu, že parametr b má být kladné reálné číslo, je naplněn. Jeho kladná hodnota se v grafické vizualizaci projevuje klesající a konvexní křivkou. Přesné hodnoty parametru A , parametru b a koeficientu determinace R^2 jsou uvedeny v Tabulce 41.

Tabulka 41 Hodnoty parametru A , parametru b a koeficientu determinace R^2 pro matematický model související s pozorováními uvedenými v Tabulce 40

Článek z blogu ²⁶	Parametr A	Parametr b	Koeficient determinace R^2 (%)
	4,5461	0,5208	86,98

Z Tabulky 40 a grafické vizualizace je zřejmé, že jediné pozorování, které se odchyluje z tendence, kterou definuje zákon, lze spatřit u znaku tvořeného devíti prvky. Jedná se o znak 洲 (zhōu). Tento odskok je pravděpodobně způsoben nízkou

²⁶ Po vynechánínejméně frekventovaného pozorování $x_4 = 9$ parametr $A = 4,3123$; parametr $b = 0,4554$; koeficient determinace $R^2 = 97,19\%$

četností jeho výskytu – ve výběrovém souboru se objevil pouze jednou a ani žádný jiný znak s devíti prvky, který by ovlivňoval průměrnou délku devíti prvkových znaků, se v textu nevyskytl. Toto pozorování je proto možné zanedbat. Pokud bychom toto jediné pozorování odebraly, vhodnost modelu by se zvýšila na 97 %. Hodnota parametru A by v tomto případě byla 4,3123 a parametru b 0,4554.

Ze získaných dat a vizualizací je patrné, že MAL na jazykové úrovni U4 platí. Jeho slovní formulace v tomto případě zní:

Čím více má znak prvků, tím méně tahů má v průměru prvek.

nebo také:

Čím méně má znak prvků, tím více tahů má v průměru prvek.

Diskuze

Čím je však způsobena tak vysoká shoda s matematickým modelem MALu? Jak je možné, že stavba znaků dodržuje striktně tendenci definovanou MALem? Jako jedno z možných vysvětlení se nabízí, že významnou roli ve stavbě znaků hraje grafické pole, do kterého se znaky zapisují. Jak již bylo uvedeno výše (viz podkapitola 2.3), grafické pole je neměnný prostor, do kterého se zapisují znaky nehledě na počet jejich tahů. Grafické pole tedy ovlivňuje uspořádání tahů a prvků ve znaku.

Pokud se znak skládá z menšího počtu prvků, prvky se mohou skládat z více tahů, protože mohou zabrat více prostoru a není narušena jeho rozpoznatelnost. Jeden znak tvořený jedním prvkem (tzn., že znak je totožný s prvkem) může mít komplikovaný prvek, protože má dostatek prostoru, aby byl tento prvek rozpoznatelný. Grafické pole v případě znaků s nízkým počtem prvků limituje stavbu znaku jen minimálně.

V případě znaků s velkým počtem prvků je třeba, aby každý prvek byl od dalšího oddělen. Z důvodu menšího prostoru a požadavku čitelnosti se tedy prvky musejí skládat z menšího počtu tahů. Čím více prvků znak obsahuje, tím méně

prostoru na každý prvek připadá a tím se také snižuje počet tahů, ze kterého se prvek skládá. Vliv grafického pole je v tomto případě velmi silný.

Pro názornost jsou v Tabulce 42 uvedeny příklady znaků s menším a větším počtem tahů.

Tabulka 42 Příklad znaků s jedním, devíti a deseti prvky

	Počet prvků	Počet tahů	Průměrný počet tahů v prvku
夏	1	10	10
着	1	11	11
洲	9	9	1
憾	10	16	1,6

Z výše uvedených Tabulek 40 a 42 vyplývá, že jedno prvkové znaky jsou ve většině případů složeny z vyššího počtu tahů. V případě článku z blogu je průměrný počet tahů u jedno prvkových znaků 4,3544. Tyto prvky mohou vystupovat i v rámci jiných znaků, avšak jejich výskyt bude pravděpodobně klesat s rostoucím počtem prvků ve znaku, jelikož by ve více prvkovém znaku byly obtížně rozpoznatelné.

Co se týče znaků tvořených více prvky, průměrný počet tahů na jeden prvek je v uvedeném případě výrazně nižší (viz Tabulka 42). Z Tabulky 40 vyplývá, že znaky složené z pěti a více prvků mají v průměru méně než dva tahy na jeden prvek. Takovéto prvky se mohou vyskytovat jako složky dalších více prvkových znaků bez větších omezení, jelikož se dají snadno odlišit.

V návaznosti na dosud zjištěné údaje, jsme došly k závěru, že tyto zákonitosti musí platit ještě ve větší míře u tradičních znaků, které byly v rámci reformy provedené v letech 1956 a 1964 zjednodušeny na současnou formu zápisu. Tato reforma snížila počet tahů u 2 236 znaků a tím zjednodušila jejich strukturu. Oproti zjednodušené formě by tradiční znaky měly vykazovat vyšší shodu s MAlem,

jelikož je jejich původní podoba složitější a tlak grafického pole by se u nich měl projevit ještě výrazněji. Stejně jako v případě textů psaných zjednodušenými znaky, i tady by mělo platit: čím více mají znaky prvků, tím je průměrná délka prvků měřená v tazích kratší.

S kolegyní jsme se rozhodly ověřit, do jaké míry se odráží MAL v textech psaných v tradičních znacích. Výběrový soubor, který spoluautorka použila pro experiment DP (vzorek A, viz Příloha 1), byl převeden do tradiční podoby a následně kvantifikován. Předpokladem bylo, že shoda s MALEM by měla být ještě vyšší, protože některé znaky, které bude převedený text obsahovat, budou mít složitější strukturu, a to by se mělo odrazit v celkových výsledcích.

Na základě dat získaných z dílčího experimentu 4A, který provedla spoluautorka Motalová (viz podkapitola 3.6.1), jsme zjistily, že procentuální shoda s MALEM se dle očekávání ještě navýšila. Tím se potvrdilo, že tlak grafického pole je u složitějších struktur znaků výraznější.

Výsledky získané na této jazykové úrovni potvrzují, že konstantní velikost grafického pole hraje důležitou roli v případě utváření vzájemných vztahů mezi znaky a prvky. Takto vysoká shoda s matematickým modelem MALu může být způsobena také tím, že prvky ve složitějších strukturách mají k dispozici menší prostor, a proto se k sobě musí více „lepít“. V případě, že se dotýkají, vytváří společně jeden nový prvek. Tento prvek se tedy skládá z více tahů. MAL na této úrovni platí a zní následovně:

čím méně má znak prvků, tím je vyšší průměrná hodnota prvků měřených v tazích.

4.6.2 Jazyková úroveň U3

Parcelát – znak

Tabulka 43 uvádí empirické hodnoty získané kvantifikací článku z blogu. Konstrukt x_3 představuje délku parcelátu (měřenou ve znacích), z_3 je jejich frekvence a konstituent y_3 představuje průměrnou délku znaků (měřenou v prvcích). Empirická pozorování s nízkou četností výskytu ($z_3 \leq 2$) jsou v tabulce vyznačena šedým pozadím.

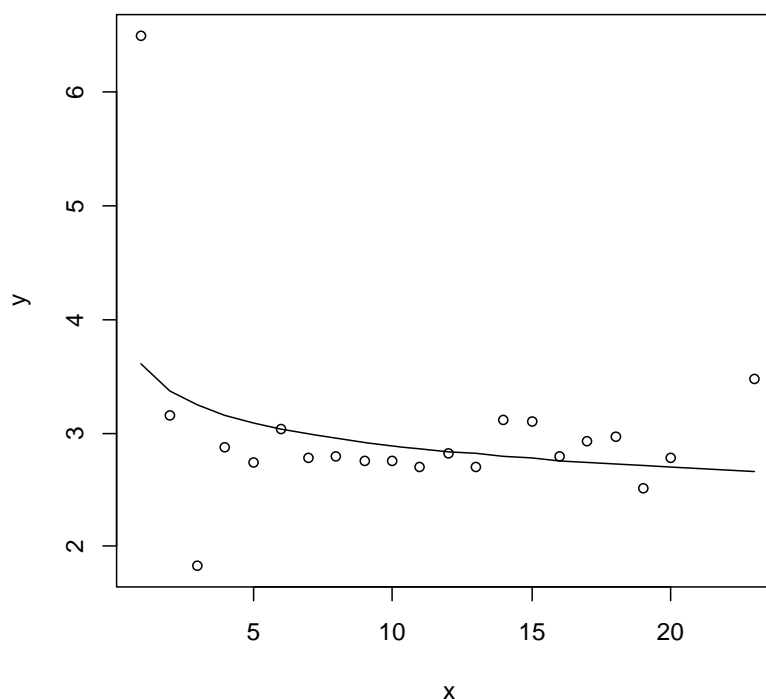
Tabulka 43 Jazyková úroveň U3: parcelát (měřený ve znacích) – znak (měřený v průměrném počtu prvků)

x_3	z_3	y_3
1	2	6,5000
2	10	3,1500
3	4	1,8333
4	35	2,8714
5	20	2,7400
6	30	3,0333
7	38	2,7744
8	24	2,7917
9	23	2,7536
10	14	2,7571
11	12	2,7045
12	14	2,8155
13	10	2,6920
14	8	3,1161
15	7	3,0952
16	6	2,7917
17	5	2,9294
18	2	2,9722
19	4	2,5132
20	6	2,7833
23	2	3,4783

Z Tabulky 43 lze vidět, že ve zkoumaném článku se vyskytlo 21 parcelátů s různou délkou. Nejmenší parcelát je složen z 1 znaku a nejdelší parcelát jich obsahuje 23. Průměrná délka znaků na této jazykové úrovni evidentně nekoresponduje s tvrzením MALu, že s rostoucím počtem parcelátů klesá průměrná délka znaků. Průměrná délka naopak ve většině případů osciluje v pásmu hodnot dvou a tří prvků. Její celkové rozmezí se pohybuje v intervalu $\langle 1,83; 6,5 \rangle$. Co se týče

frekvence parcelátů, parceláty tvořené 1, 18 a 23 znaky ($x_3 = 1; 18; 23$) se vyskytují s nejmenší četností. Tato empirická pozorování jsou v tabulce vyznačena šedým pozadím.

Vztah formulovaný MAlem, který by na této jazykové úrovni U3 mohl znít *čím více má parcelát znaků, tím méně prvků má v průměru znak*, se evidentně neprojevil.



Obrázek 16 Grafická vizualizace pozorování uvedených v Tabulce 43

Z grafické vizualizace je patrné, že shoda s matematickým modelem je minimální (procentuální shoda je 13,16 %), avšak klesající tendence formulovaná MAlem byla naznačena. Nízká hodnota parametru b naznačuje, že křivka bude mít konstantní tendenci. Viz Tabulka 44.

Z konstantní tendence vybočují hlavně pozorování $x_3 = 1$ a $x_3 = 23$. Tato pozorování se však vyskytují s nejmenšími frekvencemi, a proto se jedná o deviace, které je možno zanedbat. V případě $x_3 = 1$, tzn. parceláty složené z jednoho znaku, je tato odchylka zajímavá z toho důvodu, že se jedná o dva parceláty složené pouze z jednoho znaku 操 (cāo) a 滾 (gǔn), které mají pět a osm prvků. Ve srovnání s delšími parceláty, jejichž znaky jsou složeny průměrně ze dvou a tří prvků, mají tyto dva znaky vyšší počet prvků a tím odpovídají MALu:

čím méně má parcelát znaků, tím více prvků má v průměru znak.

Jedná se ovšem pouze o dvě pozorování, proto nelze říci, že parceláty tvořené jedním znakem mívají obecně komplikovanější znaky. Z tohoto důvodu je třeba provést více pozorování.

Tabulka 44 Hodnoty parametru A , parametru b a koeficientu determinace R^2 pro matematický model související s pozorováními uvedenými v Tabulce 43

Článek z blogu ²⁷	Parametr A	Parametr b	Koeficient determinace R^2 (%)
	3,6055	0,0965	13,16

Diskuze

Nízká shoda s matematickým modelem a rozptýlenost pozorování na této jazykové úrovni mohou být způsobeny několika možnými důvody. Prvním z nich by mohl být fakt, že parcelát (konstrukt) a znak (konstituent) jsou velmi odlišné jazykové jednotky. Konstrukt je jednotka s proměnlivou délkou a konstituent jednotka s neměnnou délkou. Délka parcelátu může být ovlivněna dle přání pisatele, zatímco délku znaku libovolně měnit nelze. Znak má svou pevně danou formu a nelze nijak upravovat. Pisatel má sice možnost zvolit jiný znak s podobným významem, avšak tím může pozměnit i smysl vyjádření. A co víc, v mnoha případech nelze volit jinou alternativu, jelikož se jedná o specifické výrazy. Odlišný charakter jednotek je proto možnou příčinou, proč se vztah formulovaný MALem na této úrovni projevil s tak nízkou shodou.

Míra shody s matematickým modelem MALu by mohla být také ovlivněna různými přístupy ke stanovení prvku. Jelikož prvkem nemá svou přesnou definici, může být určen více způsoby. Různé stanovení prvku a s ním související segmentace znaku na prvky zůstane předmětem dalšího výzkumu. Přístupy k určení prvku budou ověřovány pomocí MALu.

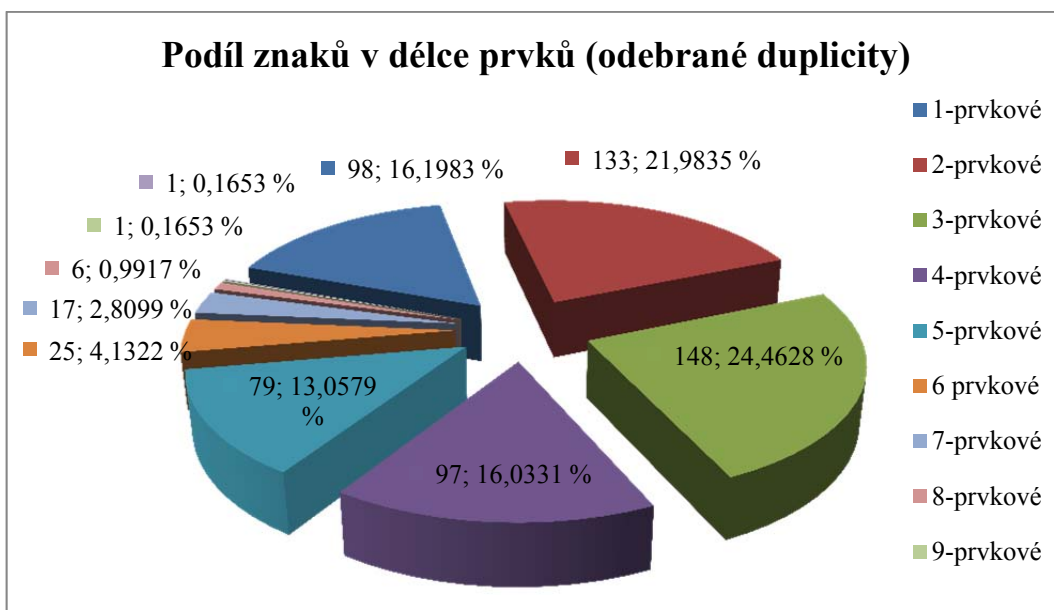
²⁷ Po vynechání nejméně frekventovaných pozorování $x_3 = 1; 18; 23$ parametr $A = 2,5571$; parametr $b = -0,0364$; koeficient determinace $R^2 = 4,11$ %

Dalším možným důvodem nízké shody mohou být strukturní změny znaků zapříčiněné centrálně plánovanou reformou provedenou v letech 1956 a 1964. Reforma měla za cíl zjednodušit čínské znakové písmo, aby bylo přístupnější širší veřejnosti a zvýšilo gramotnost čínských obyvatel. U 2 236 znaků byl snížen počet tahů a tím byly zredukovány i průměrné délky prvků ve znacích (viz podkapitola 3.6.1). Reformy také pravděpodobně zapříčinily, že se průměrná délka u většiny znaků měřená v počtu prvků snížila a u nejfrekventovanějších znaků se jeví, že se ustálila na dva a tři prvky. V případě tradičních znaků je délka znaků (v prvcích) rozmanitější (viz podkapitola 3.6.1, Tabulka 20), a proto by se u textů psaných tradičními znaky mohl MAL projevit s vyšší shodou, což už částečně naznačily výsledky experimentu A1. Tuto hypotézu je však třeba dále prozkoumat na větším vzorku textů psaných tradičními znaky. Počty prvků v souvislosti s frekvencí daných délek znaků mají na této úrovni pravděpodobně zásadní vliv, proto se na ni zaměřím níže.

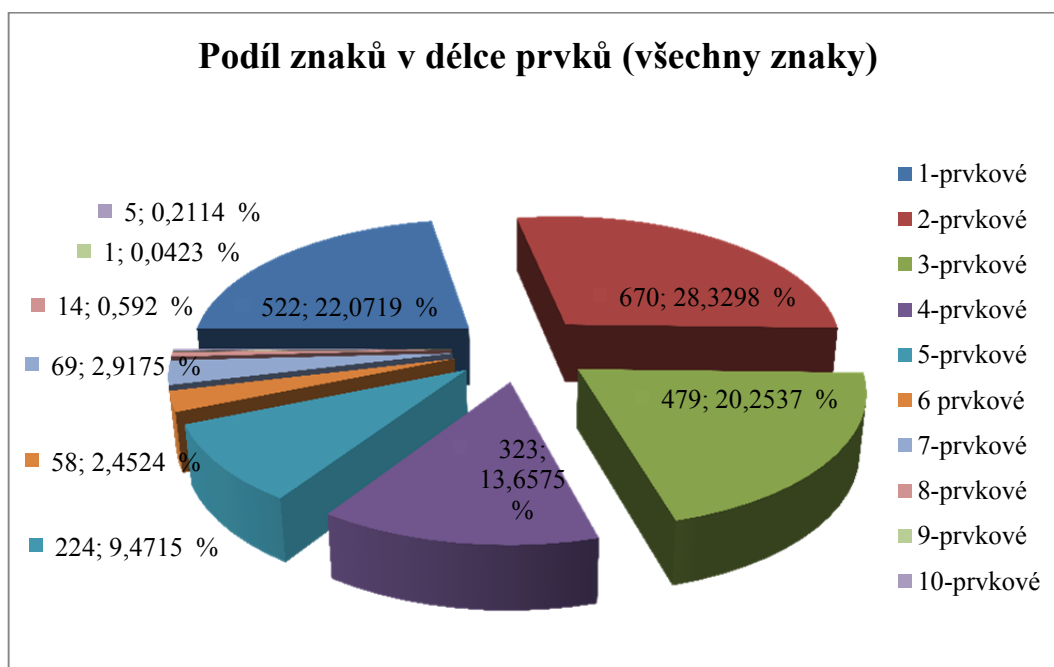
Na jazykové úrovni znak – prvek jsme zjistily, že nejfrekventovanější znaky jsou tvořeny dvěma a třemi prvky. Jejich vysoká frekventovanost proto může být dalším možným důvodem neprojevení se MALu. Přesné údaje o zastoupení frekvencí znaků je uvedeno v následující Tabulce 45 a graficky zobrazeno na Obrázcích 17 a 18:

Tabulka 45 Přehled frekvencí znaků podle počtu prvků

Délka znaků (v prvcích)	Všechny různé znaky v textu (odebrané duplicity)		Všechny znaky v textu (s duplicitami)	
	Počet znaků	%	Počet znaků	%
1-prvkové	98	16,1983	522	22,0719
2-prvkové	133	21,9835	670	28,3298
3-prvkové	148	24,4628	479	20,2537
4-prvkové	97	16,0331	323	13,6575
5-prvkové	79	13,0579	224	9,4715
6 prvkové	25	4,1322	58	2,4524
7-prvkové	17	2,8099	69	2,9175
8-prvkové	6	0,9917	14	0,592
9-prvkové	1	0,1653	1	0,0423
10-prvkové	1	0,1653	5	0,2114
Celkem	605		2 365	



Obrázek 17 Grafická vizualizace frekvencí znaků podle počtu prvků (s odebranými duplicitními hodnotami), data získaná z Tabulky 45



Obrázek 18 Grafická vizualizace frekvencí znaků podle počtu prvků (veškeré hodnoty), data získaná z Tabulky 45

Na základě dat uvedených v Tabulce 45 a grafických vizualizací lze konstatovat, že dvou a tří prvkové ve zkoumaném článku představují přes 46 % všech různých znaků (tj. 281 z 605 různých znaků), v případě všech znaků použitých v textu (tzn. i s opakováním) je to přes 48 % (1 149 z 2 365 všech znaků). Dalšími vysoce frekventovanými znaky jsou znaky složené z jednoho nebo čtyř prvků.

Jelikož dvou a tří prvkové znaky zaujímají téměř polovinu všech znaků, ovlivňují stavbu parcelátu ze všech znaků nejvíce. Je možné, že právě z důvodu malé variability znaků (co se týče počtu prvků, ze kterých se znaky skládají), se MAL na této jazykové úrovni projevuje s tak nízkou shodou. Pokud jsou parceláty z velké části složené ze znaků s průměrnou délkou jeden až pět prvků, potom nezáleží, jak je parcelát dlouhý, průměrná délka znaků zůstane konstantní, tzn. okolo dvou a tří prvků. Nízká frekvence více prvkových znaků (šest a více) způsobuje, že je na této úrovni narušen vztah nepřímé úměrnosti mezi parcelátem (konstruktem) a znakem (konstituentem).

Dvou a tří prvkové znaky se vyskytovaly s vysokou frekvencí nejenom v případě tohoto článku, ale také i v předešlém experimentu (2012). Zda tento typ znaků patří k frekventovaným znakům v rámci frekvenčního seznamu programu Wenlin²⁸, zkusíme prověřit v následujícím výzkumu, ve kterém se zaměříme na frekvenční analýzu znaků obsažených v článku z blogu.

Frekvenční analýza

U všech různých znaků použitých v článku z blogu budu zjišťovat jejich pořadí (rank) v rámci frekvenčního seznamu programu Wenlin. Čím nižší bude hodnota ranků, tím vyšší bude frekvence daných znaků. Na základě hodnoty ranku vytvořím tři pásma. První pásmo bude zahrnovat znaky s rankem 1-1000, druhé pásmo 1001-2000 a třetí pásmo 2001-3000 nejfrekventovanějších znaků. Klíčové bude zejména zjištění, jaké zastoupení mají dvou a tří prvkových znaků v rámci jednotlivých pásem.

Tabulka 46 uvádí zastoupení všech znaků článku z blogu po odebrání duplicitních hodnot, tj. celkem 605 znaků. Znaky s přiděleným rankem jsou dle jeho

²⁸ Zdroj: Frekvenční seznam znaků programu 文林 Wenlin Software. 4.0.2

hodnoty roztríděny do tří frekvenčních pásem. Znaky, které rank nemají, tvoří zvláštní skupinu, jež leží mimo tato pásma.

Tabulka 46 Zastoupení všech znaků po odebrání duplicitních hodnot v rámci frekvenčních pásem

Frekvenční pásmo	Všechny znaky v textu (odebrané duplicity)	
	Počet znaků	%
První frekvenční pásmo (1-1000)	472	78,02
Druhé frekvenční pásmo (1001-2000)	89	14,71
Třetí frekvenční pásmo (2001-3000)	32	5,29
Znaky mimo frekvenční pásma	12	1,98
Celkem	605	100

Zdroj: Frekvenční seznam znaků z programu Wenlin Software for Learning Chinese Version 4.0.2; vlastní zpracování

Většina znaků použitých v článku z blogu se řadí mezi 1000 nejfrekventovanějších znaků (tj. do prvního pásma nejfrekventovanějších znaků). Ze všech různých znaků, které se v textu vyskytly, tvoří přes 78 %.

V návaznosti na již zjištěné údaje je dále třeba zjistit, jakou část zaujímají v rámci každého frekvenčního pásma znaky tvořené dvěma a třemi prvky.

Tabulka 47 Zastoupení všech různých znaků podle počtu prvků v rámci frekvenčních pásem a mimo ně

Počet prvků	1. frekvenční pásmo		2. frekvenční pásmo		3. frekvenční pásmo		Znaky mimo frekvenční pásma	
	Počet znaků	%	Počet znaků	%	Počet znaků	%	Počet znaků	%
1-prvkové	92	19,49	6	6,74	-	-	-	-
2-prvkové	116	24,58	14	15,73	2	6,25	1	8,33
3-prvkové	105	22,25	30	33,71	10	31,25	3	25,00
4-prvkové	80	16,95	10	11,24	4	12,50	3	25,00
5-prvkové	52	11,02	17	19,10	6	18,75	4	33,33
6-prvkové	9	1,91	10	11,24	5	15,63	1	8,33
7-prvkové	14	2,97	-	-	3	9,38	-	-
8-prvkové	3	0,64	2	2,25	1	3,13	-	-
9-prvkové	1	0,21	-	-	-	-	-	-
10-prvkové	-	-	-	-	1	3,13	-	-

Zdroj: Frekvenční seznam znaků z programu Wenlin Software for Learning Chinese Version 4.0.2; vlastní zpracování

V prvním frekvenčním pásmu (1-1000 nejfrekventovanějších znaků) mají největší zastoupení dvou prvkové znaky – 24,58 %, tj. 116 ze 472 znaků, a následně znaky tři prvkové – 22,25 %, tj. 105 ze 472 znaků. V druhém frekvenčním pásmu (1001-2000 nejfrekventovanějších znaků) zaujímají největší část znaky složené z tří prvků – 33,71 %, tj. 30 z 89 znaků. Stejně tak je tomu i ve třetím frekvenčním pásmu (2001-3000 nejfrekventovanějších znaků), kde tři prvkové znaky tvoří 31,25 %, tj. 10 z 32 znaků. Teprve znaky, které nespádají do pásem tří tisíc nejfrekventovanějších znaků, jsou nejčastěji tvořeny pěti prvky – 33,33 %, tj. 4 z 12 znaků. V případě těchto znaků však bylo analyzováno jen málo pozorování, a proto je třeba provést další experimenty. Nejfrekventovanější znaky v rámci každého frekvenčního pásma jsou v Tabulce 47 označeny šedým pozadím. Co se týče tohoto článku, celkově by se dalo shrnout, že čím se číslo ranku zvyšuje, tím se zároveň navyšují počty prvků ve znacích. Dvou nebo tří prvkové znaky procentuálně zaujímají v každém frekvenčním pásmu největší část. Je velmi pravděpodobné, že výsledky získané na této jazykové

úrovni nejsou ojedinělé, podobná situace by měla nastat i u jiných výběrových souborů.

V rámci dalšího výzkumu se proto zaměříme na výzkum všech 3 000 nejfrekventovanějších znaků, u kterých budeme zjišťovat počet jejich prvků. Následně budou vytvořeny skupiny znaků se stejnou délkou, u nichž bude sledováno jejich procentuální zastoupení – jak v rámci těchto 3 000 znaků, tak i v rámci jednotlivých frekvenčních pásem.

Na místě je otázka, proč mají dvou a tří prvkové znaky tak vysoké zastoupení? Příčinou může být fakt, že čínské znakové písmo bylo v procesu vývoje ovlivněno ekonomizací, které se projevilo v reformních změnách minulého století. Dalo by se říct, že na vývoj čínského písma působí dvě protikladné síly: jednou z nich je požadavek jednoduchosti a druhou požadavek srozumitelnosti. Dva a tři prvky v tomto případě tvoří průnik minimální nutné potřeby z hlediska odlišnosti a ekonomičnosti. I z toho důvodu jsou dvou a tří prvkové znaky tlačeny do popředí a také ve frekvenčním seznamu by měly zaujímat přední místa, což ověříme v rámci dalšího výzkumu.

Posledním důvodem, který by mohl zapříčinit tak nízkou shodu s MALEM na této úrovni, je absence další jazykové jednotky, která by pravděpodobně narušila konstantní tendenci. Tato jazyková jednotka musí být vyšší než znak a zároveň menší než parcelát. Nejpravděpodobněji by touto jednotkou mohlo být slovo. Při jejím stanovení se však již není možné opírat o grafický princip, protože v čínských textech psaných ve znacích neexistuje slovo založené na ortografickém principu, neboť čínské znaky k sobě těsně přiléhají a nejsou od sebe odděleny mezerami. Z grafického hlediska tedy nelze určit hranice slova. Proto je třeba zvolit jiný způsob určení této jednotky. Jelikož jsme v první i v druhé fázi experimentu (2012, 2013) přihlédly při stanovení jazykové jednotky parcelát k syntaktickému hledisku, proto je i v zde nejvhodnější dodržovat sekundárně použité segmentační kritérium.

V případě zařazení nové jazykové jednotky by hierarchie jazykových jednotek měla následující podobu:

tah – prvek – znak – **SLOVO** – parcelát – souvětí – odstavec

Pokud bychom tuto jazykovou jednotku zařadily do uvedené hierarchie, získaly bychom další jazykové úrovně: slovo – znak a parcelát – slovo. Na první uvedené jazykové úrovni by slovo představovalo konstrukt a znaky by byly jeho konstituenty. Na vyšší jazykové úrovni (uvedené na druhém místě) by potom slova tvořila konstituenty parcelátu, který by byl na této úrovni konstruktem.

Experiment 3B

Abychom ověřily hypotézu, že v hierarchii jazykových jednotek chybí jazyková jednotka slovo, rozhodly jsme se provést experiment. Experiment je značen podle příslušné úrovně (tj. 3) a příslušného výběrového souboru (tj. B).

Vstupní data pro experiment 3B poskytl opět článek z blogu. Segmentace slova byla provedena dle pravidel syntaxe uvedených v knize *Úvod do studia hovorové češtiny* (Švarný, 2001). Níže jsou uvedeny výsledky získané po kvantifikaci dat.

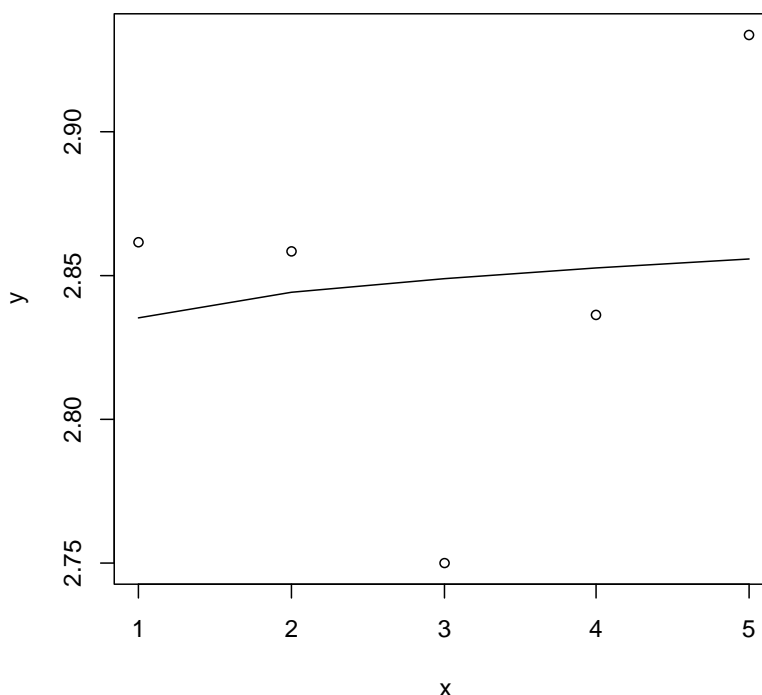
Tabulka 48 uvádí empirické hodnoty získané kvantifikací článku z blogu. Konstrukt x_{3B} představuje délku slova (měřenou ve znacích), z_{3B} je jejich frekvence a konstituent y_{3B} představuje průměrnou délku znaků (měřenou v prvcích). Empirická pozorování s nízkou četností výskytu ($z_{3B} \leq 3$) jsou v tabulce vyznačena šedým pozadím.

Tabulka 48 Experiment 3B – Jazyková úroveň U3: slovo (měřené ve znacích) – znak (měřený v průměrném počtu prvků)

x_{3B}	z_{3B}	y_{3B}
1	490	2,8612
2	652	2,8582
3	144	2,7500
4	31	2,8362
5	3	2,9333

Z výše uvedené Tabulky 48 lze vidět, že v článku z blogu se vyskytují slova složená z jednoho až pěti znaků. Slova tvořená dvěma znaky mají nejvyšší četnost, poté jsou to slova složená z jednoho a tří znaků. Je evidentní, že nezávisle na délce slova (měřeného ve znacích) se průměrná délka znaků (měřená v prvcích) u všech

pozorování pohybuje opět v pásmu hodnot dvou a tří prvků, přesněji v intervalu $\langle 2,75; 2,93 \rangle$. Rozdíly v délce znaků (měřených v prvcích) jsou minimální. Slova složená z pěti znaků se vyskytla s nejmenší četností, a proto mohou být zanedbána. Toto empirické pozorování je v tabulce vyznačeno šedým pozadím.



Obrázek 19 Grafická vizualizace pozorování uvedených v Tabulce 48

Také z grafické vizualizace je zřejmé, že klesající tendence křivky zobrazující vztah mezi slovem a znaky předpokládaná MALem se neprojevila. Předpoklad pro splnění MALu, že parametr b má být kladné reálné číslo, není naplněn (viz tabulka 49). Shoda s matematickým modelem MALu je pouhých 1,47 %.

Tabulka 49 Hodnoty parametru A , parametru b a koeficientu determinace R^2 pro matematický model související s pozorováními uvedenými v Tabulce 48

Článek z blogu ²⁹	Parametr A	Parametr b	Koeficient determinace R^2 (%)
	2,8352	-0,0044	1,47

²⁹ Po vynechání nejméně frekventovaných pozorování $x_3 = 5$ parametr $A = 2,8618$; parametr $b = 0,0158$; koeficient determinace $R^2 = 26,12$ %

Na základě provedeného experimentu lze konstatovat, že ani při zařazení další jazykové jednotky slovo (měřené ve znacích) se shoda s MALem neprojevila. Důvodů neprojevení se shody může být několik. Prvním z nich může být nepřesné stanovení jazykové jednotky (slovo), případně nevhodně zvolený princip jejího stanovení, tzn. syntaktické hledisko. V dalším výzkumu se proto zaměříme na jiné způsoby určení hranic této jazykové jednotky.

Další důvod souvisí s délkou znaků (měřených v prvcích). Jak je uvedeno výše, většina znaků je složena ze dvou a tří prvků. Z toho důvodu bude i většina slov složena právě ze dvou a tří prvkových znaků. Proto nezáleží z kolika znaků je slovo složeno, průměrná délka znaků by se vždy měla pohybovat přibližně okolo stejných hodnot. Délka slov měřená ve znacích bude mít konstantní tendenci.

4.6.3 Jazyková úroveň U2

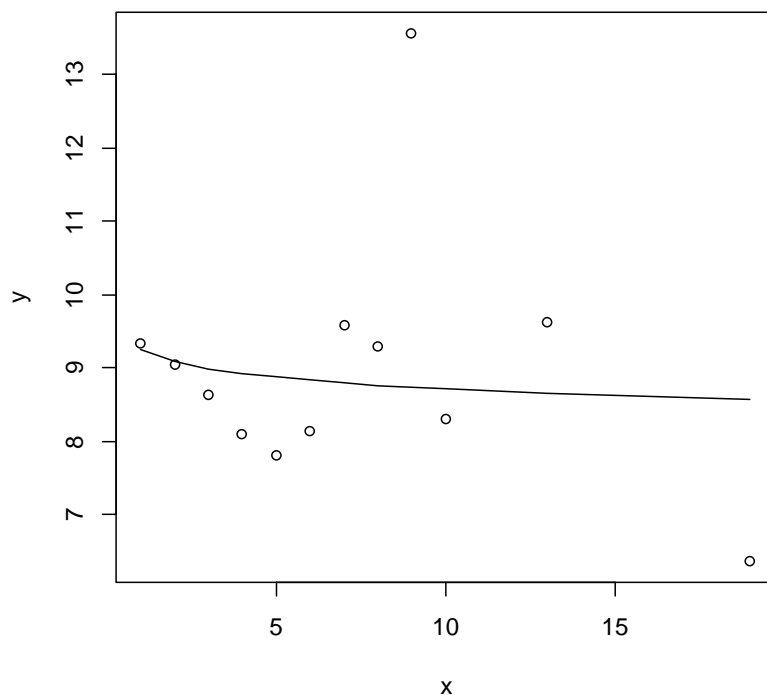
Souvětí – parcelát

Tabulka 50 uvádí empirické hodnoty získané kvantifikací článku z blogu. Konstrukt x_2 představuje délku souvětí (měřenou v parcelátech), z_2 je jejich frekvence a konstituent y_2 představuje průměrnou délku parcelátu (měřenou ve znacích). Empirická pozorování s nízkou četností výskytu ($z_2 \leq 1$) jsou v tabulce vyznačena šedým pozadím.

Tabulka 50 Jazyková úroveň U2: souvětí (měřené v parcelátech) – parcelát (měřený v průměrném počtu znaků)

x_2	z_2	y_2
1	9	9,3333
2	11	9,0455
3	10	8,6333
4	15	8,1000
5	5	7,8000
6	5	8,1333
7	1	9,5714
8	4	9,2813
9	1	13,5556
10	2	8,3000
13	1	9,6154
19	1	6,3684

Z Tabulky 50 je zřejmé, že ve výběrovém souboru se vyskytlo 12 různých délek konstruktů (souvětí měřených v parcelátech) složených v rozmezí z 1 až 19 parcelátů. Průměrné délky parcelátů v případě všech empiricky získaných pozorování nekorespondují s tvrzením MALu, že s rostoucí délkou konstruktů (souvětí) klesá průměrná délka jeho konstituentů (parcelátů). Průměrná délka všech parcelátů spíše kolísá v intervalu $\langle 6,37; 13,56 \rangle$. Klesající tendenci však lze pozorovat u hodnot $x_2 = 1; 2; 3; 4; 5$, které jsou navíc v textu zastoupené s nejvyšší četností. Nejméně frekventovaná jsou souvětí složená ze sedmi, devíti, třinácti a devatenácti parcelátů, $x_2 = 7; 9; 13; 19$. Tato souvětí se ve výběrovém souboru vyskytla pouze jednou, jejich frekvence výskytu je tedy v poměru k ostatním frekvencím nízká, proto je možné tato pozorování zanedbat. V tabulce jsou vyznačena šedým pozadím.



Obrázek 20 Grafická vizualizace pozorování uvedených v Tabulce 50

Na základě grafické vizualizace můžeme tvrdit, že vztah nepřímé závislosti souvětí a parcelátů formulovaný MAlem se na této úrovni sice projevil, ale s minimální shodou. Matematický model vykazuje s empiricky získanými pozorováními shodu pouhých 1,44 %.

Podíváme-li se však na prvních pět pozorování, zjistíme, že se u nich klesající tendence křivky projevila. Z toho plyne, že MAL v tomto pásmu platí.

Tabulka 51 Hodnoty parametru A , parametru b a koeficientu determinace R^2 pro matematický model související s pozorováními uvedenými v Tabulce 50

Článek z blogu ³⁰	Parametr A	Parametr b	Koeficient determinace R^2 (%)
	9,2452	0,0258	1,444

³⁰ Po vynechání nejméně frekventovaných pozorování $x_2 = 7; 9; 13; 19$ parametr $A = 9,0852$; parametr $b = 0,0434$; koeficient determinace $R^2 = 23,33$ %

Diskuze

Hypotéza, že s rostoucím počtem souvětí měřených v parcelátech klesá průměrná délka parcelátů (měřená ve znacích) se na této úrovni neprojevila s velkou shodou z několika možných důvodů.

Prvním z nich byl zmíněn již v předchozí podkapitole; pravděpodobnou příčinou neprojevení se MALu je absence jazykové jednotky, která je větší než znak a menší než parcelát. Touto jednotkou by pravděpodobně mohlo být slovo, které by spolu s parcelátem vytvořilo další jazykovou úroveň parcelát – slovo. Parcelát by na této jazykové úrovni představoval konstrukt a slova jeho konstituenty. Ověření této hypotézy a stanovení hranic jazykové jednotky *slovo* a bude předmětem našeho dalšího výzkumu.

Dalším důvodem nízké shody s MALem může být nedostatečný počet pozorování na této jazykové úrovni, který by mohl způsobit neprojevení se vzájemných vztahů konstruktů a jeho konstituentů. V rámci ověření předpokladu, že se MAL projeví až v případě delšího článku, provedeme experiment, jehož předmětem bude testovat platnost MALu na delším výběrovém souboru. Minimální délka je stanovena na 3 500 znaků, což je horní hranice délky výběrového souboru určené pro obě fáze experimentu (2012, 2013). Za výběrový soubor poslouží další článek z blogu od stejného autora (tj. od Han Hana). Pro účely tohoto dílčího experimentu jsme zvolily článek *Můj otec Han Renjun a jeho dílo* (Wǒ de fùqin Hán Rénjūn yǐjǐ tā de zuòpǐn, 我的父亲韩仁均以及他的作品) – vzorek C (Příloha 3), který splňuje všechna kritéria stanovená pro volbu výběrového souboru. Článek obsahuje 9254 znaků a na blogu byl zveřejněn 27. ledna 2012. Experiment je značen podle příslušné úrovně (tj. 2) a příslušného výběrového souboru (tj. C).

Experiment 2C

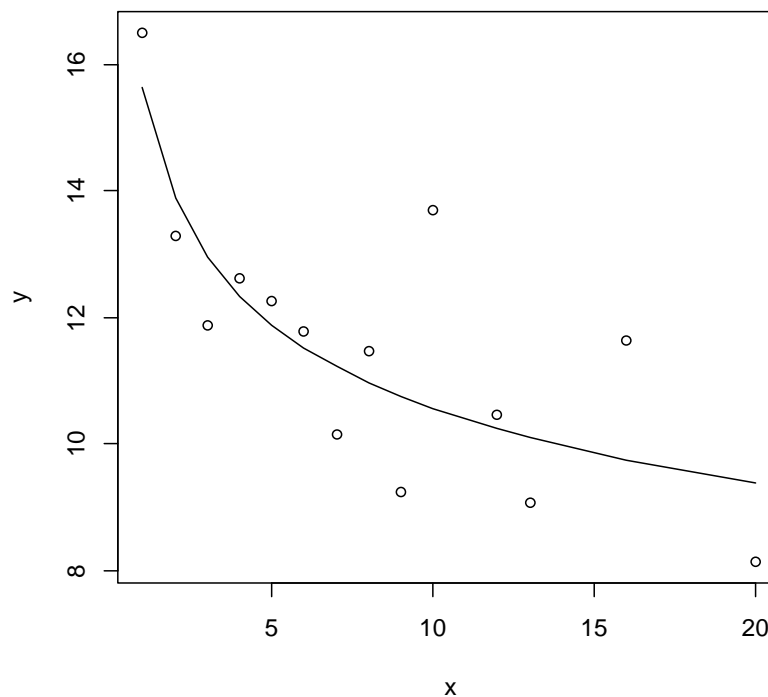
Tabulka 52 uvádí empirické hodnoty získané kvantifikací článku z blogu. Konstrukt x_{2C} představuje délku souvětí (měřenou v parcelátech), z_{2C} je jejich frekvence a konstituent y_{2C} představuje průměrnou délku parcelátů (měřenou

ve znacích). Empirická pozorování s nízkou četností výskytu ($z_{2c} \leq 1$) jsou v tabulce vyznačena šedým pozadím.

Tabulka 52 Experiment 2C – Jazyková úroveň U2: souvětí (měřené v parcelátech) – parcelát (měřený v průměrném počtu znaků)

x_{2c}	z_{2c}	y_{2c}
1	41	16,4878
2	52	13,2885
3	40	11,8750
4	26	12,6250
5	13	12,2615
6	10	11,7667
7	6	10,1429
8	6	11,4583
9	4	9,2500
10	1	13,7000
12	2	10,4583
13	1	9,0769
16	1	11,6250
20	1	8,1500

Ve zkoumaném článku se vyskytlo celkem 14 různě dlouhých souvětí (měřených v parcelátech), přitom souvětí složené z 10, 13, 16 a 20, tj. $x_2 = 10; 13; 16; 20$ parcelátů, se v textu vyskytlo pouze jednou, a proto je možné tato měření zanedbat. Tato empirická pozorování jsou v tabulce vyznačena šedým pozadím. Průměrná délka parcelátů se pohybuje v intervalu $\langle 8,15; 16,49 \rangle$. Z tabulky lze vypočítat klesající tendenci průměrných délek parcelátů definovanou MAlem.



Obrázek 21 Grafická vizualizace pozorování uvedených v Tabulce 52

Z vizualizace vztahů mezi znakem (konstruktem) a prvkem (konstituentem) můžeme konstatovat, že se klesající tendence křivky zobrazující vztah mezi délkou konstruktů (znak) a délkou konstituentů (prvek), kterou definuje MAL, projevila. Podmínka, že parametr b má být kladné reálné číslo, je splněna. Matematický model vykázal mnohem vyšší shodu s MALEM než je tomu v případě kratšího článku (viz Tabulka 53).

Tabulka 53 Hodnoty parametru A , parametru b a koeficientu determinace R^2 pro matematický model související s pozorováními uvedenými v Tabulce 52

Článek z blogu ³¹	Parametr A	Parametr b	Koeficient determinace R^2 (%)
	15,6188	0,1699	59,31

Z dat a vizualizací získaných kvantifikací článku s větší délkou je patrné, že na jazykové úrovni U2 MAL platí. Jeho slovní formulace pro tuto hladinu zní:

Čím více má souvětí parcelátů, tím méně znaků má v průměru parcelát.

³¹ Po vynechání nejméně frekventovaných pozorování $x_2 = 10; 13; 16; 20$ parametr $A = 15,7993$; parametr $b = 0,1897577$; koeficient determinace $R^2 = 80,81\%$

nebo také:

Čím méně má souvětíparcelátů, tím více znaků má v průměru parcelát.

Předpoklad, že vztah formulovaný MALem se projeví až u většího množství pozorování, byl v tomto případě správný. V budoucnu je však třeba provést více experimentů, které předpoklad potvrdí.

4.6.4 Jazyková úroveň U1

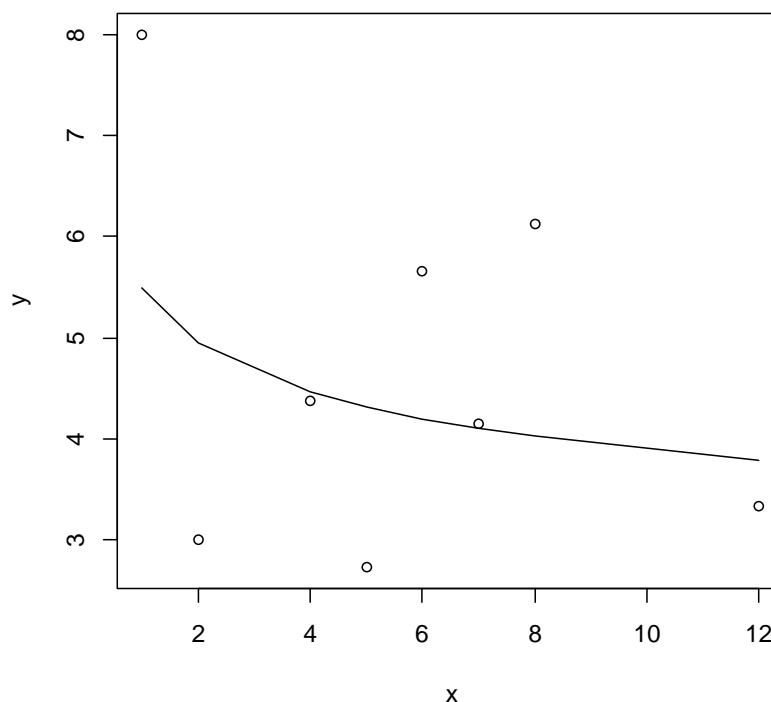
Odstavec – souvětí

Tabulka 54 uvádí empirické hodnoty získané kvantifikací článku z blogu. Konstrukt x_1 představuje délku odstavců (měřenou v souvětích), z_1 je jejich frekvence a konstituent y_1 představuje průměrnou délku souvětí (měřenou v parcelátech).

Tabulka 54 Jazyková úroveň U1: odstavec (měřený v souvětích) – souvětí (měřený v průměrném počtu parcelátů)

x_1	z_1	y_1
1	1	8,0000
2	1	3,0000
4	2	4,3750
5	3	2,7333
6	2	5,6667
7	1	4,1429
8	1	6,1250
12	1	3,3333

Jak je uvedeno v Tabulce 54, ve výběrovém souboru se vyskytlo 8 různých délek odstavců měřených v souvětích. Nejkratší odstavec je složen z jednoho souvětí a nejdelší z dvanácti. Frekvence odstavců je u všech pozorování velmi nízká, nejčastěji se vyskytl odstavec tvořený z pěti souvětí, avšak i tato frekvence představuje pouhé tři výskyty. Proto není možné žádné pozorování zanedbat. Jak vyplývá z uvedených dat, u průměrných délek souvětí měřených v parcelátech není klesající tendence definovaná MALEM patrná; průměrné délky spíše oscilují v intervalu $\langle 2,73; 8,00 \rangle$.



Obrázek 22 Grafická vizualizace pozorování uvedených v Tabulce 54

Z grafické vizualizace je patrné, že klesající tendence křivky, která vystihuje vztah jazykových jednotek formulovaný MALEM, je na této úrovni pouze naznačena. Vykázala se jen velmi nízká shoda s matematickým modelem (procentuální shoda je 10,22 %). Parametr b svou kladnou hodnotou splňuje podmínku splnění MALu.

Tabulka 55 Hodnoty parametru A , parametru b a koeficientu determinace R^2 pro matematický model související s pozorováními uvedenými v Tabulce 54

Článek z blogu	Parametr A	Parametr b	Koeficient determinace R^2 (%)
	5,4888	0,1493	10,22

Diskuze

I když je na této úrovni nízká frekvence pozorování, minimální shoda by také mohla být zapříčiněna například interpunkcí. Interpunkce, která stanovuje hranice konstituentů, tj. souvětí, a také parcelátů, v nichž jsou měřeny průměrné délky

souvětí, byla do čínských textů zavedena poměrně nedávno. Z tohoto důvodu nemusí být použití interpunkčních znamének ustálené a vymezení jazykových jednotek se nemusí řídit vždy identickými podmínkami (blíže viz podkapitola 3.6.4 a Motalová et al, 2013).

Dalším faktorem by stejně jako na jazykové úrovni U2 mohla být nízká frekvence pozorování, která zapříčinila neprojevení se vztahu nepřímé úměrnosti mezi konstruktem a konstituentem na této jazykové úrovni. Zda má délka textu vliv na projevení se vztahu formulovaného MALEM opět zkusíme ověřit experimentem, ve kterém bude analýze podroben delší text. Stejně jako v předešlém experimentu 2C na úrovni U2 jako výběrový soubor poslouží článek *Můj otec Han Renjun a jeho dílo* (Wǒ de fùqīn Hán Rénjūn yǐjǐ tā de zuòpīn, 我的父亲韩仁均以及他的作品) s délkou přesahující 9 200 znaků. Experiment je značen podle příslušné úrovně (tj. 1) a příslušného výběrového souboru (tj. C).

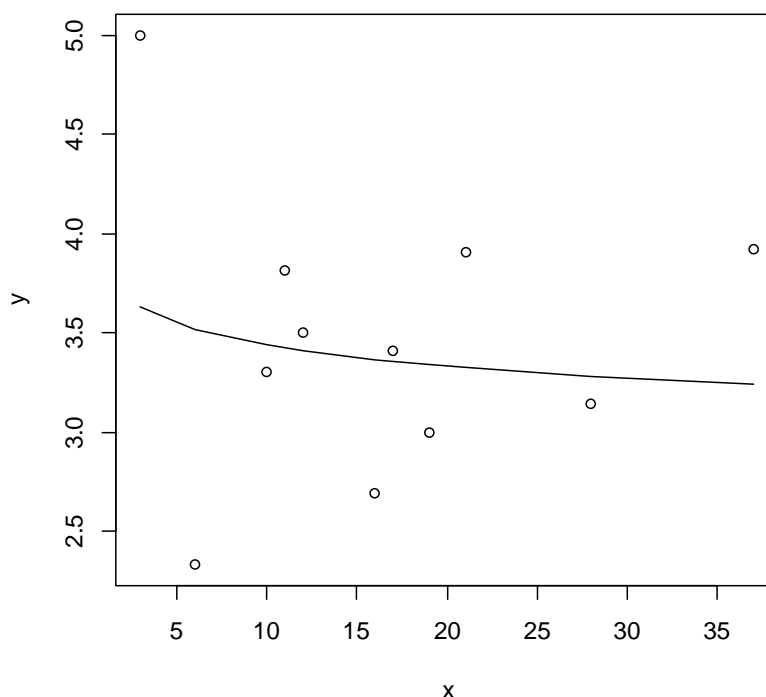
Experiment 1C

Tabulka 56 uvádí empirické hodnoty získané kvantifikací delšího článku z blogu. Konstrukt x_{1C} představuje délku odstavců (měřenou v souvětích), z_{1C} je jejich frekvence a konstituent y_{1C} představuje průměrnou délku souvětí (měřenou v parcelátech).

Tabulka 56 Experiment 1C – Jazyková úroveň U1: odstavec (měřený v souvětích) – souvětí (měřený v průměrném počtu parcelátů)

x_{1C}	z_{1C}	y_{1C}
3	1	5,0000
6	1	2,3333
10	1	3,3000
11	1	3,8182
12	3	3,5000
16	1	2,6875
17	1	3,4118
19	1	3,0000
21	1	3,9048
28	1	3,1429
37	1	3,9189

Ze získaných empirických dat uvedených v Tabulce 56 jsme zjistily, že delší článek operuje celkem s 11 různými délkami odstavců (měřených v souvětích). Nejkratší odstavec je složen z 3 souvětí a nejdelší z 37. Frekvence pozorování je však stále minimální. Pouze odstavec sestavený z 12 souvětí ($x_1 = 12$) se vyskytl ve třech případech. Klesající tendence průměrných délek souvětí není z tabulky patrná a tím pravděpodobně ani v případě delšího článku není naplněn předpoklad MALu, že s rostoucí délkou odstavců (konstruktů) klesá průměrná délka souvětí (jeho konstituentů).



Obrázek 23 Grafická vizualizace pozorování uvedených v Tabulce 56

Na základě grafické vizualizace můžeme konstatovat, že se klesající tendence křivky, znázorňující vztah mezi délkami veličin (odstavec a souvětí) předpokládaný MALem, neprojevila s velkou shodou ani v případě delšího článku. Parametr b sice splňuje podmínku kladného reálného čísla, avšak shoda s matematickým modelem je pouhých 2,46 % (viz tabulka 57). Tato shoda je dokonce nižší, než v případě kratšího článku (viz Tabulka 55).

Tabulka 57 Hodnoty parametru A , parametru b a koeficientu determinace R^2 pro matematický model související s pozorováními uvedenými v Tabulce 56

Článek z blogu	Parametr A	Parametr b	Koeficient determinace R^2 (%)
	3,8167	0,0455	2,46

Můžeme shrnout, že slovní definování MALu, které by na této úrovni mohla znít: *čím delší je odstavec v počtu souvětí, tím kratší je průměrná délka souvětí v počtu parcelátů*, se v případě článku z blogu neprojevila ani u kratšího článku z blogu složeného z 2641 znaků, ani u delšího článku obsahujícího 9254 znaků.

Neprojevení se MALu na této jazykové úrovni, (případně projevení se s tak malou shodou) může být způsobeno nedostatečným množstvím vstupních dat, tzn. nedostatečnou délkou článku a následnou nízkou četností výskytu odstavců. I když byl proveden experiment, který zkoumal více než trojnásobně dlouhý článek (tedy s délkou přesahující 9 200 znaků), ani po jeho analyzování jsme nezískaly dostatek vstupních dat. Navrhujeme proto provést další experimenty, jejichž výběrové soubory budou obsahovat více než 5 000 znaků, což je přibližná délka vědeckého článku, na kterém byla platnost MALu na této úrovni potvrzena.

4.6.5 Shrnutí

V této podkapitole byly interpretovány data a grafické vizualizace získané na základě kvantifikace článku z blogu. Ověřování platnosti (či neplatnosti) MALu bylo provedeno na čtyřech jazykových úrovních U4, U3, U2 a U1. Každá úroveň byla tvořena dvěma v hierarchii vedle sebe sousedícími jazykovými jednotkami. V tomto shrnutí budou zopakovány výsledky získané na každé jazykové úrovni a budou nastíněny možné důvody projevení se nebo neprojevení se shody s MAlem.

Na nejvyšší úrovni U4 znak – prvek se shoda s matematickým modelem MALu projevila s nejvyšší shodou ($R^2 = 86,98\%$)³². Takto vysoká shoda je pravděpodobně způsobena strukturou samotných znaků, na kterou působí konstantní velikost grafického pole. Na základě této teze jsme došli k závěru, že v případě textů psaných v tradiční podobě znaků se musí tlak grafického pole projevovat ještě výrazněji. Kolegyně provedla experiment, který zkoumal platnost MALu na textu převedeném do tradičních znaků. Výsledky prokázaly, že po transformaci výběrového souboru do tradiční podoby znaků, se shoda s matematickým modelem ještě zvýšila.

Na další úrovni U3 parcelát – znak byla tendence formulovaná MAlem pouze naznačena. Náznak klesající tendence byl způsoben zejména díky málo četnými pozorováními, které způsobily odchylky od konstantní tendence. Z výsledné vizualizace vztahů mezi konstruktem a konstituentem lze vyzorovat spíše konstantní tendenci křivky, tzn., že se průměrné délky znaků pohybují v pásmu hodnot dvou a tří prvků. Po odebrání nejméně četných pozorování by tendence křivky byla konstantnější a procento shody by se ještě snížilo. Minimální shoda s MAlem může být způsobena více možnými příčinami. První z nich je odlišný charakter jazykových jednotek. Parcelát je jednotka s proměnlivou délkou, zatímco znak je neměnná jazyková jednotka. Dalším důvodem velmi nízké shody mohou být strukturální změny znaků zapříčiněné reformami v minulém století, které zredukovaly počty tahů u znaků a tím způsobily, že se u mnoha znaků změnil také počet prvků. S tímto důvodem souvisí fakt, že téměř polovina znaků je tvořena dvěma a třemi prvky, a proto také z velké části ovlivňují stavbu parcelátů, které se ze znaků skládají. Posledním možným důvodem je absence další jazykové jednotky (větší než znak a menší než parcelát), kterou by pravděpodobně mohlo být *slovo*. Na základě

³² Bez statistického ošetření

tohoto předpokladu jsem provedla experiment, který stanovil další jazykovou jednotku „slovo“, dle syntaktických pravidel uvedených v *Úvodu do hovorové čínštiny* (Švarný, 2001). Stanovením této jazykové jednotky vznikly další jazykové úrovně: slovo (měřené ve znacích) – znak (průměrná délka měřená v prvcích) a parcelát (měřený ve slovech) – slovo (průměrná délka měřená ve znacích). Dílčí experiment byl proveden na jazykové úrovni slovo – znak. Výsledky výzkumu přinesly zjištění, že i se zvyšující se délkou slova ve znacích je průměrná délka znaků (složených z prvků) téměř konstantní, tj. průměrná délka znaků se pohybuje v intervalu (2,75; 2,93). Shoda s MAlem se tedy neprojevila ani po zařazení další jazykové jednotky „slovo“. Proto se v dalších výzkumech zaměříme na její preciznější stanovení.

Námi předpokládaná vzájemná nepřímá závislost mezi konstruktem a konstituentem se téměř neprojevila ani na úrovni U2 souvětí – parcelát. Minimální shoda může být stejně jako v případě předchozí jazykové úrovně U3 způsobena absencí další jazykové jednotky, která je větší než znak a menší než parcelát. Dalším možným důvodem je nedostatečný počet pozorování. Z tohoto důvodu jsem se opět rozhodla provést experiment, jehož výběrovým souborem bude delší článek. Za výběrový soubor jsme zvolily článek z blogu od autora Han Hana *Můj otec Han Renjun a jeho dílo* s délkou 9 254 znaků. Po kvantifikaci tohoto článku jsme zjistily, že se tendence klesající křivky projevila, a také shoda s matematickým modelem MALu byla poměrně vysoká ($R^2 = 59,31\%$). Předpoklad, že se při nedostatečné délce výběrového souboru neprojeví vzájemné závislosti vztahů formulované MAlem, byl v tomto případě správný.

Na poslední nejvyšší jazykové úrovni U1 odstavec – souvětí se klesající tendence křivky projevila, avšak vykazovala se jen velmi nízká shoda s matematickým modelem. Jedním z možných důvodů projevení se pouze malé shody může být interpunkce (stanovující hranici souvětí), která nemá v čínských textech dlouhou tradici. Dalším důvodem může být nedostatečný počet empirických pozorování. Proto byl na této úrovni opět proveden experiment, který zkoumal delší výběrový soubor – článek *Můj otec Han Renjun a jeho dílo*. Klesající tendence křivky znázorňující vztah mezi délkami veličin (odstavec a souvětí) předpokládaná MAlem se téměř neprojevila ani v případě delšího článku a procentuální shoda byla minimální. Velký vliv na výsledky však stále mají nedostatečné četnosti výskytu pozorování konstruktů, tj. odstavců.

Náš předpoklad, že MAL je adekvátní a dobře sedící model na jazykových úrovních U4 znak – prvek a U2 souvětí – parcelát, jsme experimentem realizovaným na článku z blogu potvrdily pouze na úrovni U4. Na jazykové hladině U2 se prokázala pouze minimální shoda s matematickým modelem MALu. Teze, že se MAL neprojeví (případně projeví s minimální shodou) na jazykových hladinách U3 parcelát – znak a U1 odstavec – souvětí byla v obou případech naplněna.

5 Komparace výběrových souborů

V této části budeme komparovat výsledky (grafické vizualizace, parametry A , parametry b a koeficienty determinace R^2) získané kvantifikací čtyř výběrových souborů: novinového článku a povídky, které byly předmětem první fáze výzkumu realizované v roce 2012, s vědeckým článkem a článkem z blogu, které byly analyzovány v druhé fázi výzkumu realizované v roce 2013.

Předtím, než zahájíme komparaci výběrových souborů, je třeba uvést, že výsledky první fáze výzkumu budou publikovány v časopise *Czech and Slovak Linguistics Review*. K empiricky získaným hodnotám této první fáze jsme se z důvodů detekce hlavních tendencí vzájemných vztahů veličin rozhodly přistupovat statisticky. Proto jsme získané hodnoty s nejnižší četností výskytu, tzv. extrémny, ošetřily statistickou metodou vynechání.

Druhá fáze si kladla za cíl navýšit množství zpracovaných dat analyzováním dalších výběrových souborů, které byly napsány jinými jazykovými styly. V případě vědeckého článku a článku z blogu jsme se rozhodly aplikovat lingvistický přístup, tzn. zachovat všechna pozorování. Výsledky experimentu jsou v této diplomové práci uváděny i s pozorováními, které mají nízkou četnost výskytu.

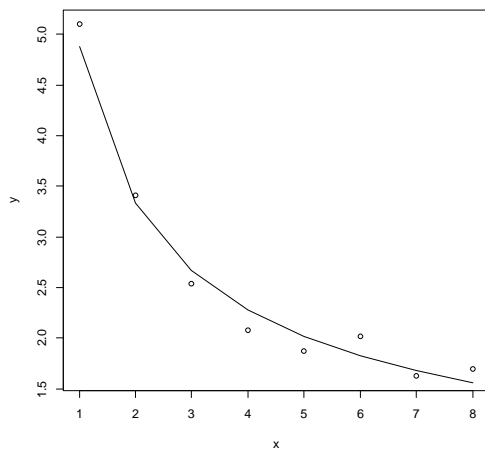
Abychom mohly výsledky vzájemně porovnat, musely jsme zvolit jeden z aplikovaných přístupů. Z důvodu sledování tendence jazykových stylů na jednotlivých úrovních jsme se rozhodly porovnávat data bez odebraných pozorování.

Statisticky ošetřené hodnoty parametru A , parametru b a koeficientu determinace R^2 uváděné v článku *An Application of the Menzerath-Altmann Law to Contemporary Chinese* budeme pro srovnání uvádět v poznámce pod čarou. Tyto údaje jsou pro přehlednost doplněny rovněž statisticky ošetřenými daty z druhé fáze experimentu (2013).

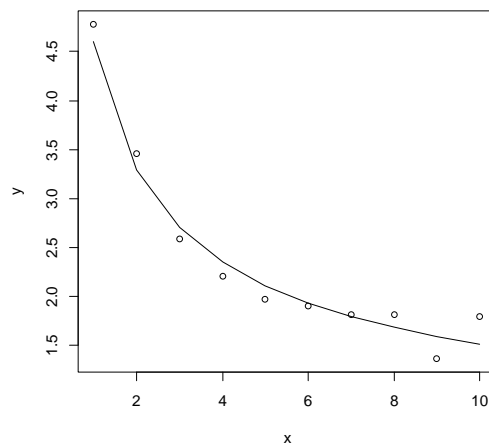
U každé komparace příslušných jazykových úrovní nejprve uvádíme grafické vizualizace jednotlivých výběrových souborů a tabulky s hodnotami parametrů a koeficientů determinace. Následně jsou daná data interpretována. Tabulky obsahující pozorování získaná kvantifikací novinového článku a povídky (viz Motalová et al, 2013) jsou vloženy v přílohách (viz Příloha 4).

5.1 Komparace výběrových souborů na jazykové úrovni U4

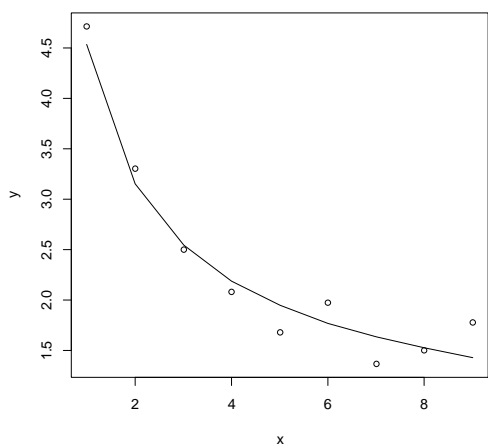
Znak – prvek



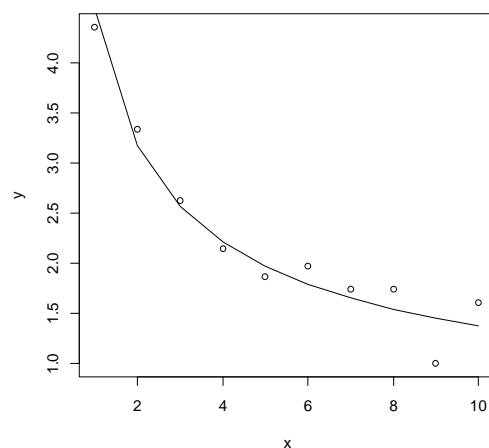
Obrázek 24 Grafická vizualizace vztahu mezi znakem a prvkem, novinový článek



Obrázek 25 Grafická vizualizace vztahu mezi znakem a prvkem, povídka



Obrázek 26 Grafická vizualizace vztahu mezi znakem a prvkem, vědecký článek



Obrázek 27 Grafická vizualizace vztahu mezi znakem a prvkem, článek z blogu

Tabulka 58 Hodnoty parametrů A , b a koeficientů determinace R^2 pro matematický model související s empiricky získanými pozorováními³³

Výběrový soubor	Parametr A	Parametr b	Koeficient determinace R^2 (%)
Novinový článek ³⁴	4,8774	0,5482	96,58
Povídka ³⁵	4,5992	0,4828	93,89
Vědecký článek ³⁶	4,5406	0,5259	90,61
Článek z blogu ³⁷	4,5461	0,5208	86,98

Ve všech čtyřech případech je parametr b kladné reálné číslo (viz Tabulka 58), čímž je první předpoklad splnění MALu naplněn. Z grafické vizualizace je patrné, že tendence křivky zobrazující vztah nepřímé úměrnosti mezi konstruktem a konstituentem se u všech souborů projevila (viz Obrázek 24, 25, 26 a 27). Shoda u všech výběrových souborů dosahuje extrémních hodnot. Nejvyšší hodnoty dosáhla u novinového článku (96,58 %) a nejnižší u článku z blogu (86,98 %). V případě zbylých dvou se pohybovala přes 90 % (viz Tabulka 58).

Vztahy mezi tahem, prvkem a znakem jsou neměnné. Struktura každého znaku se utváří pod vlivem identického faktoru, tzn. pod vlivem konstantní velikosti grafického pole. Proto se u každého znaku projevují stejné tendence. Bez ohledu na jazykový styl či délku textu matematický model vykázal ve všech čtyřech případech širokou shodu s empiricky získanými pozorováními, což by mohlo být způsobeno pravděpodobně díky této tendenci (viz podkapitola 3.6.1 a 4.6.1).

³³ viz Příloha 4

³⁴ Po vynechání nejméně frekventovaných pozorování parametr $A = 4,9724$; parametr $b = 0,5738$; koeficient determinace $R^2 = 97,17$ %

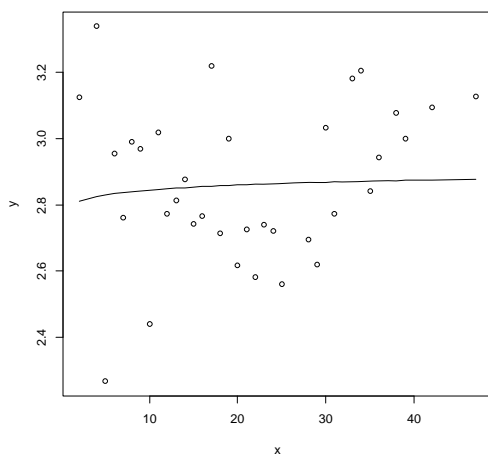
³⁵ Po vynechání nejméně frekventovaných pozorování parametr $A = 4,6546$; parametr $b = 0,4941$; koeficient determinace $R^2 = 97,67$ %

³⁶ Po vynechání nejméně frekventovaných pozorování parametr $A = 4,8117$; parametr $b = 0,5974$; koeficient determinace $R^2 = 95,19$ %

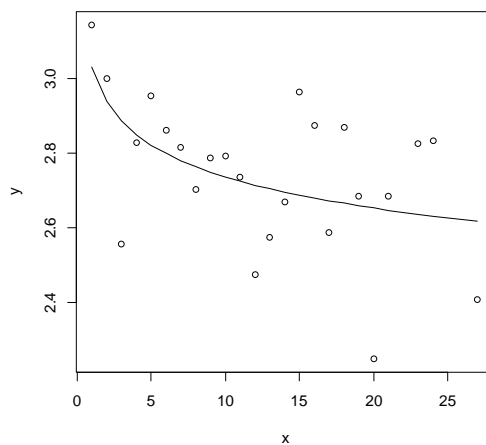
³⁷ Po vynechání nejméně frekventovaných pozorování parametr $A = 4,3123$; parametr $b = 0,4554$; koeficient determinace $R^2 = 97,19$ %

5.2 Komparace výběrových souborů na jazykové úrovni U3

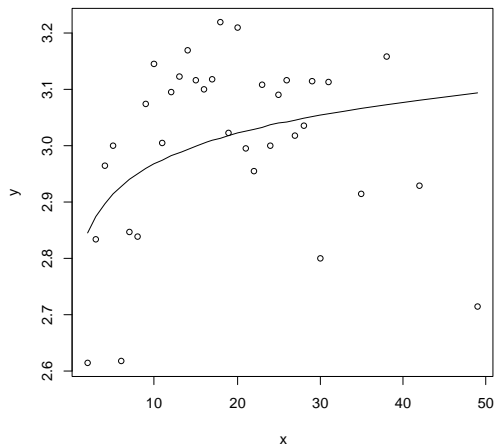
Parcelát – znak



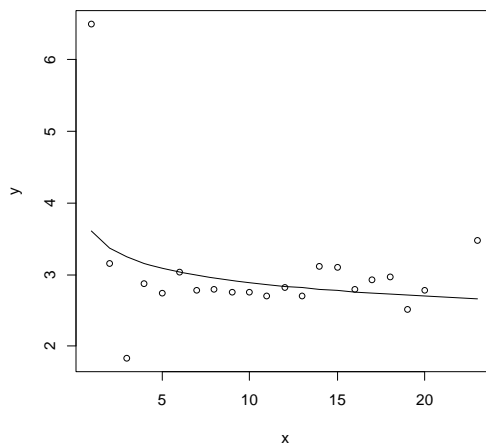
Obrázek 28 Grafická vizualizace vztahu mezi parcelátem a znakem, novinový článek



Obrázek 29 Grafická vizualizace vztahu mezi parcelátem a znakem, povídka



Obrázek 30 Grafická vizualizace vztahu mezi parcelátem a znakem, vědecký článek



Obrázek 31 Grafická vizualizace vztahu mezi parcelátem a znakem, článek z blogu

Tabulka 59 Hodnoty parametrů A , b a koeficientu determinace R^2 pro matematický model související s empiricky získanými pozorováními³⁸

Výběrový soubor	Parametr A	Parametr b	Koeficient determinace R^2 (%)
Novinový článek ³⁹	2,7968	-0,0074	0,43
Povídka ⁴⁰	3,0303	0,0444	24,70
Vědecký článek ⁴¹	2.7933	-0,0262	14,35
Článek z blogu ⁴²	3,6055	0,0965	13,16

Parametry b u novinového a vědeckého článku mají zápornou hodnotu a tím nespĺňujú prvú podmítku platnosti MALu (viz Tabulka 59). Naopak povídka a článek z blogu tuto podmítku naplnily, jejich parametry b mají kladnou hodnotu (viz Tabulka 59). Podíváme-li se na grafické vizualizace, můžeme konstatovat, že klesající tendence křivek jsou naznačeny pouze u povídky a článku z blogu (viz Obrázek 29 a 31). U novinového a vědeckého článku se neprojeví vůbec, neboť zobrazená empiricky získaná pozorování oscilují v intervalu $\langle 2,4; 3,2 \rangle$ (viz Obrázek 28 a 30). U článku z blogu je náznak klesající tendence způsoben prvním pozorování. Četnost jeho výskytu je však velmi nízká ($z_3 = 2$). Pokud bychom ho statisticky ošetřili metodou vynechání, tendence křivky by byla konstantní. Klesající tendenci nejlépe vystihují pozorování získaná kvantifikací povídky. Přesto procentuální shoda s matematickým modelem není vysoká, ostatně jako i v případě ostatních výběrových souborů (viz Tabulka 59). Protože procentuální shoda nepřesáhla ani v jednom případě 70 %, MAL se na této úrovni neprokázal jako adekvátní a dobře sedící model. (Benešová, 2011, s. 77)

Z výsledku se jeví, že vzájemný vztah nepřímé úměrnosti mezi parcelátem a znakem je narušen. Toto narušení by mohlo být způsobeno několika faktory. Prvním z nich by mohl být rozdílný charakter jazykových jednotek na této úrovni. Dále se

³⁸ viz Příloha 4

³⁹ Po vynechání nejméně frekventovaných pozorování parametr $A = 3,0945$; parametr $b = 0,0396$; koeficient determinace $R^2 = 10,34$ %

⁴⁰ Po vynechání nejméně frekventovaných pozorování parametr $A = 3,0442$; parametr $b = 0,0488$; koeficient determinace $R^2 = 34,97$ %

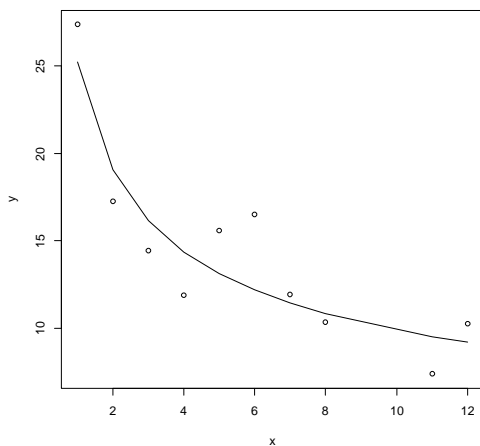
⁴¹ Po vynechání nejméně frekventovaných pozorování parametr $A = 2,6651$; parametr $b = -0,0484$; koeficient determinace $R^2 = 46,15$ %

⁴² Po vynechání nejméně frekventovaných pozorování parametr $A = 2,5571$; parametr $b = -0,0364$; koeficient determinace $R^2 = 4,10$ %

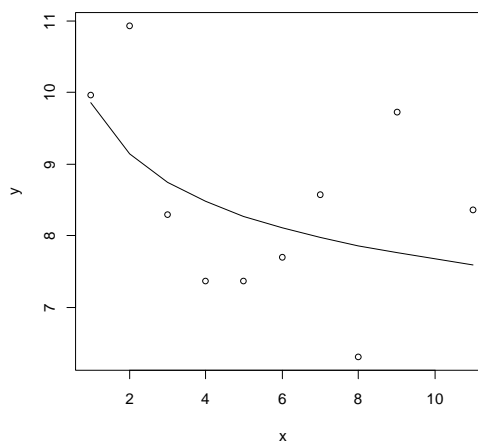
také může jednat o různé přístupy k vymezení prvku, absentující jednotku (příp. úroveň) a v neposlední řadě rovněž redukce variability znaků z hlediska počtu prvků, kterou zapříčinila reforma písma v 50. a 60. letech minulého století. Podrobněji viz podkapitoly 3.6.2 a 4.6.2.

5.3 Komparace výběrových souborů na jazykové úrovni U2

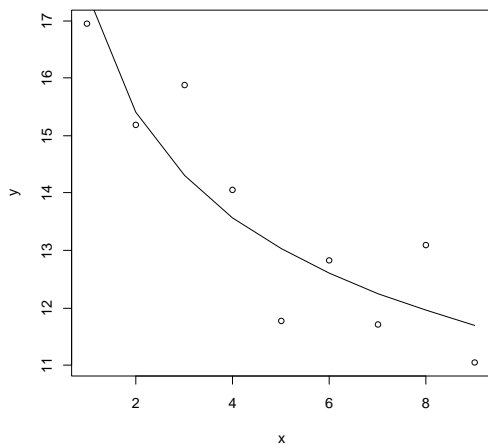
Souvětí – parcelát



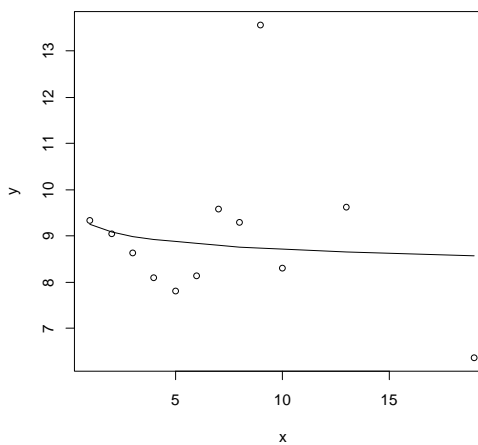
Obrázek 32 Grafická vizualizace vztahu mezi souvětím a parcelátem, novinový článek



Obrázek 33 Grafická vizualizace vztahu mezi souvětím a parcelátem, povídka



Obrázek 34 Grafická vizualizace vztahu mezi souvětím a parcelátem, vědecký článek



Obrázek 35 Grafická vizualizace vztahu mezi souvětím a parcelátem, článek z blogu

Tabulka 60 Hodnoty parametrů A , b a koeficientu determinace R^2 pro matematický model související s empiricky získanými pozorováními⁴³

Výběrový soubor	Parametr A	Parametr b	Koeficient determinace R^2 (%)
Novinový článek ⁴⁴	25,2439	0,4067	77,02
Povídka ⁴⁵	9,8600	0,1090	24,36
Vědecký článek ⁴⁶	17,4816	0,1827	78,62
Článek z blogu ⁴⁷	9,2452	0,0258	1,44

Jak je vidět z výše uvedené Tabulky 60, ve všech případech je hodnota parametru b kladné reálné číslo. Předpoklad MALu je tedy naplněn. Nejlépe je klesající tendence křivky patrná u novinového a vědeckého článku, kdežto u povídky je pouze naznačena a u článku z blogu je téměř konstantní (viz Obrázek 32, 33, 34 a 35). Ukázalo se, že na této úrovni matematický model sedí pro novinový a vědecký článek. Procentuální shoda u nich dosáhla více než 77 % (viz Tabulka 60). V případě povídky a článku z blogu je shoda výrazně nižší, u povídky činí 24,36 % a u článku z blogu pouhých 1,44 %.

Ze zatím provedených experimentů na této úrovni vyplývá, že texty, které musí dodržovat formální strukturu typickou pro daný jazykový styl (např. novinový a vědecký článek), lépe odrážejí vztah mezi veličinami formulovanými MALem. Co se týče textů volnějších jazykových stylů, ze statisticky neošetřených získaných dat se jeví, že matematický model s nimi vykazuje malou shodu.

Zaměříme-li se na nejméně frekventovaná pozorování (tzv. extrém), zjistíme, že by mohla mít na míru těsnosti matematického modelu zásadní vliv. Jejich chaotické rozptýlení totiž zastírá klesající tendenci, která je patrná u více frekventovaných pozorování. Protože se po odebrání těchto extrémů procentuální shoda výrazně navýší (viz hodnoty uvedené v poznámce pod čarou č. 44 a 46), navrhuje se provést výzkumy výběrových souborů dalších jazykových stylů.

⁴³ viz Příloha 4

⁴⁴ Po vynechání nejméně frekventovaných pozorování parametr $A = 24,7192$; parametr $b = 0,4077$; koeficient determinace $R^2 = 84,78$ %

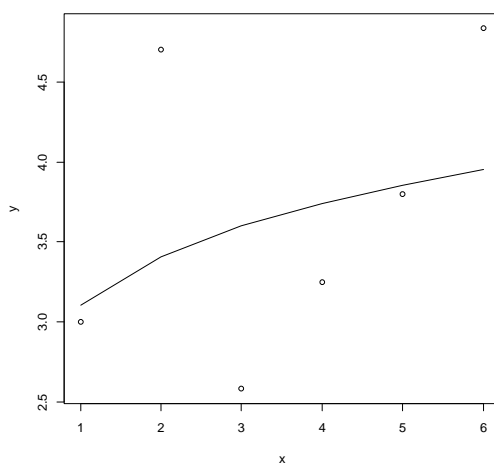
⁴⁵ Po vynechání nejméně frekventovaných pozorování parametr $A = 10,6873$; parametr $b = 0,2091$; koeficient determinace $R^2 = 68,7$ %

⁴⁶ Po vynechání nejméně frekventovaných pozorování parametr $A = 17,1221$; parametr $b = 0,1542$; koeficient determinace $R^2 = 71,67$ %

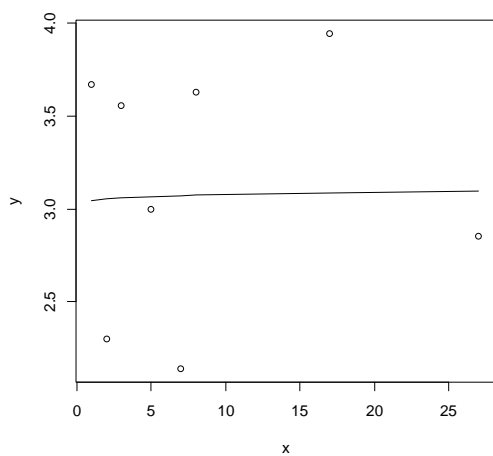
⁴⁷ Po vynechání nejméně frekventovaných pozorování parametr $A = 9,0852$; parametr $b = 0,0434$; koeficient determinace $R^2 = 23,33$ %

5.4 Komparace výběrových souborů na jazykové úrovni U1

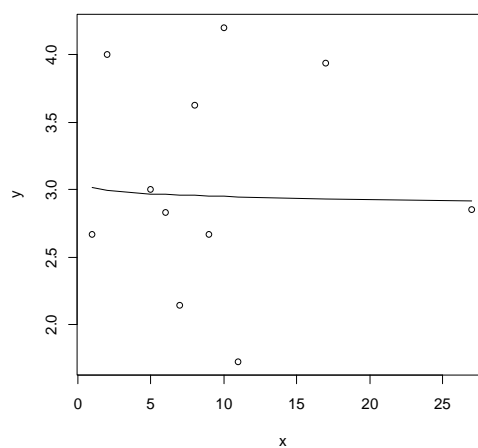
Odstavec – souvětí



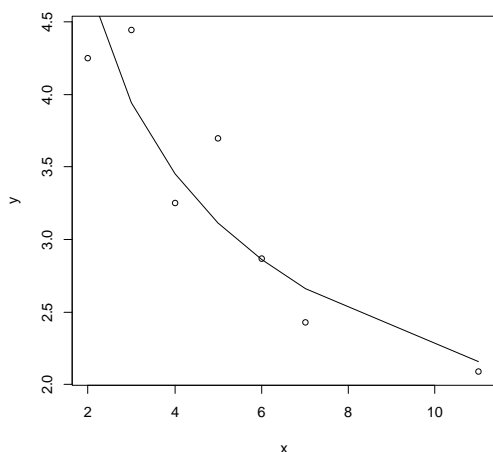
Obrázek 36 Grafická vizualizace vztahu mezi odstavcem a souvětím, novinový článek



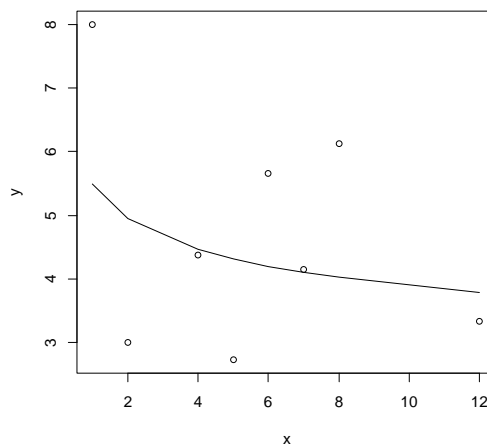
Obrázek 37 Grafická vizualizace vztahu mezi odstavcem a souvětím, povídka – varianta 1



Obrázek 38 Grafická vizualizace vztahu mezi odstavcem a souvětím, povídka – varianta 2



Obrázek 39 Grafická vizualizace vztahu mezi odstavcem a souvětím, vědecký článek



Obrázek 40 Grafická vizualizace vztahu mezi odstavcem a souvětím, článek z blogu

Tabulka 61 Hodnoty parametrů A , b a koeficientu determinace R^2 pro matematický model související s empiricky získanými pozorováními⁴⁸

Výběrový soubor	Parametr A	Parametr b	Koeficient determinace R^2 (%)
Novinový článek	3,1068	-0,1342	12,56
Povídka ⁴⁹	Vzorek A	-0,0051	0,06
	Vzorek B	0,0101	0,11
Vědecký článek	6,5604	0,4636	85,35
Článek z blogu	5,4888	0,1493	10,22

Porovnáme-li parametry b , zjistíme, že záporné hodnoty se projeví u novinového článku a povídky (vzorek A). U ostatních výběrových souborů je hodnota parametru b kladná (viz Tabulka 61). Z grafické vizualizace lze vypožorovat, že se klesající tendence křivek na této jazykové úrovni projevila u vědeckého článku, u článku z blogu byla pouze naznačena a u ostatních se neprojevila vůbec (viz Obrázky 36, 37, 38, 39 a 40). Co se týče matematické shody

⁴⁸ viz Příloha 4

Na této jazykové úrovni nebyl empiricky získán vysoký počet pozorování, a proto data nebyla v tomto případě statisticky ošetřena.

⁴⁹ V případě povídky nebyly odstavce navzájem od sebe graficky jednoznačně rozděleny. Proto jsme zvolily dvě varianty segmentace. První reprezentuje vzorek A a druhá vzorek B. Bližší informace viz (Motalová et al, 2013).

MALu s empiricky získanými pozorováními, u vědeckého článku dosáhla velmi vysoké hodnoty – 85,35 %. Důvodem této vysoké shody by mohla být výrazně větší délka tohoto výběrového souboru, díky níž mají daná pozorování v porovnání s ostatními soubory vyšší frekvenci. U ostatních výběrových souborů byla shoda minimální nebo žádná (viz Tabulka 61).

6 Závěr

Experiment realizovaný v rámci této diplomové práce byl zaměřen na kvantitativní analýzu současné psané čínštiny, kterou reprezentovaly dva výběrové soubory psané ve zjednodušených znacích a různými jazykovými styly. Pro analýzu zaměřenou na aplikaci Menzerath-Altmanova zákona jsme zvolily vědecký článek, který byl publikován v akademickém periodiku *Aplikovaná lingvistika* a následně zveřejněn *Čínskou akademií společenských věd*, a článek z blogu, jehož autorem je známý čínský bloger Han Han.

Cílem tohoto kvantitativního výzkumu bylo nejen testovat platnost tohoto jazykového zákona na námi zvolených výběrových souborech, ale zároveň získané výsledky porovnat se závěry experimentu provedeného v rámci projektu Studentské grantové soutěže 2012 *Segmentace pro testování Menzerath-Altmanova zákona a hypotéz s ním souvisejících* (číslo projektu FF_2012_035).

Pro testování platnosti MALu jsme jako v předchozí fázi výzkumu (2012) zvolily tyto jazykové jednotky: tah, prvek, znak, parcelát, souvětí a odstavec. Jejich spojením jsme získaly čtyři jazykové úrovně: znak – prvek, parcelát – znak, souvětí – parcelát a odstavec – souvětí. Naším předpokladem bylo, že se na úrovni znak – prvek MAL potvrdí a matematický model tohoto zákona vykáže s empiricky získanými pozorováními opět velmi vysokou shodu. Na hladině parcelát – znak jsme předpokládaly, že MAL nebude i nadále na této úrovni adekvátním a dobře sedícím modelem, jeho platnost se buď neprokáže, nebo výsledná shoda dosáhne pouze minimálních hodnot. Co se týče jazykové úrovně souvětí – parcelát, naší tezí bylo, že se MAL na této hladině potvrdí. V případě poslední jazykové úrovně odstavec – souvětí jsme naopak opět předpokládaly jeho neplatnost.

S ohledem na námi stanovené jazykové jednotky jsme výběrové soubory nejprve segmentovaly a následně kvantifikovaly. Abychom mohly naši hypotézu ověřit, bylo zapotřebí kvantifikací získaná data dále podrobit statistické analýze. Těmito kroky jsme dospěly k následujícím závěrům.

Co se týče vědeckého článku, platnost MALu byla překvapivě potvrzena na třech jazykových úrovních. V případě úrovní znak – prvek a souvětí – parcelát se námi předpokládaná platnost a vysoká shoda tohoto jazykového zákona s empiricky

získanými pozorováními potvrdily. Naopak nečekané výsledky vykázala hladina odstavec – souvětí. Naše původní teze, že se platnost MALu v tomto případě neprokáže, byla vyvrácena, a to dokonce extrémní shodou. U zbývajících hladiny parcelát – znak se vzájemný vztah jazykových jednotek předpokládaný tímto jazykovým zákonem neprojevil, jak jsme ostatně očekávaly.

I u článku z blogu jsme došly k překvapivým zjištěním. Shodu s matematickým modelem MALu vykázala pouze nejnižší jazyková hladina znak – prvek, u níž se nám potvrdil jak předpoklad platnosti MALu, tak i extrémní shoda, kterou s empiricky získanými pozorováními tento jazykový zákon vykázal. Naše teze týkající se jazykové úrovně parcelát – znak se nám rovněž potvrdila. Z výsledků se jeví, že zde MAL neplatí, neboť křivka má spíše konstantní tendenci. Neočekávané pro nás byly výsledky na úrovni souvětí – parcelát. Klesající tendence zde byla pouze naznačena a výsledná procentuální shoda dosáhla minimálních hodnot. Naš předpoklad tak byl vyvrácen. Podobnými výsledky se vykázala i poslední jazyková úroveň odstavec – souvětí. Oproti předchozímu případu jsme však dané závěry očekávaly.

Komparací výsledků těchto výběrových souborů se závěry první fáze experimentu (2012) jsme zjistily, že jazykové úrovně znak – prvek a parcelát – znak vykázaly podobné tendence, zatímco výsledky jazykových úrovní souvětí – parcelát a odstavec – souvětí se lišily v závislosti na výběrových souborech.

Extrémní shodu s matematickým modelem na úrovni znak – prvek, která se projevila u všech výběrových souborů, pravděpodobně zapříčinila konstantní velikost grafického pole a její vliv na strukturu znaku. Získané výsledky nejen akcentují její roli, ale především potvrzují, že jazykové styly nemají na platnost MALu na této úrovni žádný vliv.

V případě jazykové úrovně parcelát – znak se MAL neprojevil u novinového a vědeckého článku. Ačkoliv byla u povídky klesající tendence křivky naznačena, výsledná shoda s matematickým modelem MALu je minimální. U článku z blogu má dokonce křivka spíše konstantní tendenci. Celkově lze konstatovat, že MAL pro tuto úroveň nepředstavuje adekvátní a dobře sedící model. Důvodů této minimální těsnosti může být hned několik. Jako první se nabízí rozdílný charakter délek jednotek: parcelát představuje jazykovou jednotkou s proměnlivou délkou, kdežto znak jednotku s délkou konstantní. Za další důvody neplatnosti MALu, příp. jeho nízké shody s empiricky získanými pozorováními, bychom mohly považovat různé

přístupy ke stanovení prvku a absenci jazykové jednotky vyšší než znak a nižší než parcelát a následně absenci jazykové úrovně. Posledním důvodem může být zjednodušení čínského znakového písma provedené v rámci reformy v 50. a 60. letech minulého století. U 2 236 znaků byl zredukován počet tahů a prvků, a tím i jejich variabilita.

Výsledky dvou zbývajících úrovní diferencují v závislosti na výběrovém souboru. Zatímco na úrovni souvětí – parcelát širokou shodu vykázaly novinový a vědecký článek, u povídky a článku z blogu byla shoda s matematickým modelem MALu minimální. Tento fakt poukazuje na to, že vyšší shodu s MALEM pravděpodobně vykazují jazykové styly s korigovanou strukturou. V úvahu však přicházejí i další důvody. Stejně jako v případě předchozí jazykové úrovně může být příčinou neprojevení se MALu absence jazykové jednotky vyšší než znak a nižší než parcelát a rozdílná délka výběrových souborů.

Na úrovni odstavec – souvětí se široká shoda projevila pouze u vědeckého článku, u článku z blogu byla tato shoda minimální a u zbylých výběrových souborů není pozorována vůbec. Rozdílné výsledky mohla zapříčít jednak interpunkce, která byla pod vlivem západní tradice zavedena v čínských textech poměrně nedávno, a jednak nízká četnost pozorování, tzn. nedostatečná délka výběrových souborů na této úrovni.

Klíčové faktory, které by mohly mít zásadní vliv na shodu s MALEM, jsme se rozhodly testovat dílčími experimenty.

V případě jazykové úrovně znak – prvek jsme se rozhodly aplikovat MAL na vědecký článek transformovaný do tradičních znaků. Podle naší hypotézy by konstantní velikost grafického pole měla mít na tradiční znaky větší vliv, neboť obecně disponují složitější strukturou. Výsledná shoda, která měla být podle našeho předpokladu vyšší než u zjednodušených znaků, se nám tímto experimentem potvrdila, stejně jako zásadní role grafického pole.

Testování předpokladu, že mezi parcelátem a znakem chybí další jazyková jednotka – *slovo*, bylo cílem našeho dalšího experimentu. Při určování hranic parcelátu a souvětí jsme musely kromě grafiky nevyhnutelně přihlídnout k syntaktickému hledisku. Z důvodu zachování stejného segmentačního kritéria jsme se rozhodly otestovat vzájemný vztah mezi syntaktickým slovem a znakem. Výsledky experimentu však platnost MALu na této úrovni nepotvrdily. Ukázalo se, že nezávisle na délce slova (měřeného ve znacích) se průměrná délka znaků (měřená

v prvcích) u všech pozorování pohybovala v pásmu hodnot dvou a tří prvků stejně jako na hladině parcelát – znak.

Další dílčí experimenty byly provedené na úrovni souvětí – parcelát a odstavec – souvětí. První z nich se týkal interpunkce použité ve vědeckém článku. Středníku byla s ohledem na jeho diskutabilní funkci přisouzena způsobilost oddělovat mezi sebou souvětí. Hypotézu, že svou funkcí inklinuje spíše k tečce, potvrdila extrémní shoda s matematickým modelem na obou úrovních.

Druhý experiment realizovaný na úrovni souvětí – parcelát se zaměřil na kvantitativní analýzu delšího výběrového souboru. Experiment ověřoval tezi, podle které by se měly vzájemné vztahy mezi konstrukty a jejich konstituenty v případě uměleckého stylu projevit až ve výběrovém souboru tvořeném větším počtem znaků (více než 3 500 znaků). Analýze tak byl podroben další článek z blogu avšak tentokrát s délkou 9 254 znaků. Procentuální shoda s matematickým modelem MALu se výrazně navýšila, čímž se náš předpoklad potvrdil.

Co se týče experimentu na jazykové úrovni odstavec – souvětí, ani po kvantifikaci delšího uměleckého článku se MAL neprojevil jako adekvátní a dobře sedící model, naopak procentuální shoda byla ještě nižší.

Výsledky získané z těchto experimentů přináší řadu otázek a nabádají tak k realizaci dalších výzkumů. Na jazykové úrovni znak – prvek se jedná o kvantitativní analýzy výběrových souborů psaných v tradičních znacích. Jako alternativy se nabízejí tradiční texty vydané před simplifikací čínského písma a současné texty publikované na Taiwanu, příp. v Hong Kongu a Macau. Tyto výzkumy by si především kladly za cíl porovnávat získané výsledky se závěry již provedených experimentů a ověřovat působení grafického pole na vzájemný vztah jazykových jednotek na této hladině. Klíčovým zjištěním by zároveň bylo, jaký vliv mají tradiční znaky na úroveň parcelát – znak.

Další výzkumy je potřeba uskutečnit na úrovni parcelát – znak. Protože se vztah těchto jednotek může měnit v závislosti na různém definování prvku, navrhuje aplikovat jiné přístupy k segmentaci znaku na prvky a následně je ověřovat pomocí MALu.

S ohledem na fakt, že průměrná délka znaků osciluje v rozmezí dvou a tří prvků, chtěly bychom provést analýzu 3 000 nejfrekventovanějších znaků z hlediska počtu prvků a zjistit, jaké procentuální zastoupení zde představují dvou a tří prvkové znaky. V této souvislosti by bylo rovněž vhodné provést výzkum, který by se týkal

komparace tradičních forem s jejich simplifikovanými variantami. Výsledky by přinesly zjištění, do jaké míry reforma zredukovala rozdíly mezi délkami zjednodušených znaků měřených v prvcích. Procentuální zastoupení dvou a tří prvkových znaků v rámci zjednodušené sady znaků by se rovněž stalo předmětem této analýzy.

Neprojevení se vztahů na této úrovni by také mohlo souviset s absencí jazykové jednotky vyšší než znak a nižší než parcelát. Naší tezí je, že by touto jazykovou jednotkou mohlo být slovo. Z tohoto důvodu je potřeba provést výzkumy, které se zaměří na precizaci této jednotky a její chápání v čínském jazyce.

V případě úrovně souvětí – parcelát je nutné provést větší množství experimentů z důvodu testování vzájemných vztahů těchto jazykových jednotek v souvislosti s délkou výběrových souborů. Předpokládáme, že by se jejich existence měla potvrdit spíše u textů s větším počtem znaků (3 500 a více).

Zejména u nejvyšší úrovně odstavec – souvětí je nutné podrobit kvantitativní analýze dlouhé texty, aby se výrazně navýšila jejich četnost výskytu a tím bylo možné sledovat vztahy mezi příslušnými jednotkami.

Neadekvátnost MALu, která se ukázala u většiny výběrových souborů, může být také zapříčiněna grafickým členěním na základě interpunkce, která byla ovlivněna západním stylem a nemá v čínském textu dlouhou tradici. Proto se předmětem výzkumu stane i interpunkce, u níž se bude sledovat vliv na jazykové jednotky v kontextu vývoje.

Zároveň je nutné provést experimenty, které by se zaměřily na sledování vzájemných vztahů jazykových jednotek na úrovních souvětí – parcelát a odstavec – souvětí v souvislosti s různými jazykovými styly.

Posledním námi navrhovaným výzkumem je kvantitativní analýza dostatečně dlouhého výběrového souboru, která by operovala s jazykovou jednotkou „podkapitola“.

7 Resumé

The objective of this thesis is to verify the validity of the Menzerath-Altmann law applied to the standard form of contemporary written Chinese. The validity is verified using current texts written in simplified characters and in two different styles – the scientific article and the blog article. The thesis is divided into five parts. The first chapter introduces quantitative linguistics and its important milestones, it also mentions Menzerath-Altmann law. The second chapter focuses on the methodology of the research. The next two chapters (the main parts) analyse and interpret the data acquired from this experiment. This data and the data obtained from the experiment realized in 2012 are compared in the fifth chapter. The last part of this thesis summarizes the results, then draws conclusions and suggests several possibilities for further research.

Key words

Quantitative Linguistics, Menzerath-Altmann law, written Chinese, simplified characters, language units, segmentation

8 Seznam použité literatury

Monografie

1. BENEŠOVÁ, Martina (2011). *Kvantitativní analýza textu se zvláštním zřetelem k analýze fraktální*. Olomouc. 112 s. Disertační práce. Univerzita Palackého v Olomouci. Vedoucí práce Jan Andres.
2. CHEN, Ping (1999). *Modern Chinese: history and sociolinguistics*. New York: Cambridge University Press. 229 s. ISBN 05-216-4572-7.
3. ČERNÝ, Jiří (1996). *Dějiny lingvistiky*. Olomouc: Votobia. 517 s. ISBN 80-85885-96-4.
4. HŘEBÍČEK, Luděk (1997). *Lectures on Text Theory*. Prague: Academy of Sciences of the Czech Republic. 191 s. ISBN 80-85425-26-2.
5. HŘEBÍČEK, Luděk (2002). *Vyprávění o lingvistických experimentech s textem*. Praha: Academia. 195 s. ISBN 80-200-0973-6.
6. KUČERA, Ondřej et al (2005). *Učebnice čínských znaků I*. Olomouc: Univerzita Palackého. 237 s. ISBN 80-244-1118-0.
7. LINDQVIST, Cecilia (2010). *Říše čínských znaků*. Praha: Nakladatelství Lidové noviny. 427 s. ISBN 978-80-7422-054-8.
8. LIU, Yuehua et al (2009). *Integrated Chinese: Textbook Simplified Characters, Level 1 Part 1*. 3. vydání. Boston: Cheng & Tsui Company. 335 s. ISBN 978-0-88727-644-6.
9. NORMAN, Jerry (2012). *Chinese*. New York: Cambridge University Press. 292 s. ISBN 05-212-9653-6.
10. QIU, Xigui (2000). *Chinese writing*. Berkeley: The Society for the Study of Early China. 547 s. ISBN 15-572-9071-7.

11. QU, Lei Lei (2005). *Čínská kaligrafie*. Brno: CP Books. 127 s. ISBN 80-251-0711-6.
12. SIKORA, Michal (2011). *Specifika čínského lexika na internetových diskuzích*. Olomouc. Diplomová práce. Univerzita Palackého v Olomouci. Vedoucí práce Mgr. Ondřej Kučera. 133 s.
13. ŠVARNÝ, Oldřich et al (1967). *Úvod do hovorové čínštiny I*. Praha: SPN. 458 s.
14. ŠVARNÝ, Oldřich a UHER, David (2001). *Hovorová čínština: úvod do studia hovorové čínštiny*. 2. vyd. Olomouc: Univerzita Palackého. 190 s. ISBN 80-244-0298-X.
15. TĚŠITELOVÁ, Marie (1987). *Kvantitativní lingvistika*. Praha: SPN. 187 s.
16. VOCHALA, Jaromír; NOVÁK, Miroslav a PUCEK, Vladimír (1975). *Úvod do čínského, japonského a korejského písma I.: Vznik a vývoj*. Praha: SPN. 132 s.
17. VOCHALA, Jaromír a HRDLÍČKOVÁ, Věna (1985). *Úvod do studia sinologie: část filologická*. Praha: SPN. 265 s.
18. VOCHALA, Jaromír (1986). *Chinese Writing System: Minimal Graphic Units*. Praha: Univerzita Karlova. 143 s.
19. VOLÍN, Jan (2007). *Statistické metody ve fonetickém výzkumu*. Praha: Epoque. 343 s. ISBN 978-808-7027-547.
20. WANG, Li 王力 (2004). *Hanyu shigao 《汉语史稿》 Historický nástin čínského jazyka*. Beijing: Zhonghua shuju. 714 s. ISBN 7-101-04199-X
21. WANG, Ning 王宁 (2002). *Hanzi gouxing xue jiangzuo 《汉字构形学讲座》 Studie o utváření čínských znaků*. Shanghai: Shanghai jiaoyu chubanshe. 104 s. ISBN 7-5320-8070-6
22. WANG, Xianchun (2007). *Chinese running script calligraphy for beginner*. Beijing: Foreign Languages Press. 159 s. ISBN 978-7-119-04860-4.

23. WIERGER, Leon (1965). *Chinese characters: Their origin, etymology, history, classification and signification*. 2nd ed., enl. and rev. according to the 4th French ed. New York: Paragon Book Reprint. 820 s. ISBN 04-862-1321-8.
24. ZÁDRAPA, Lukáš a PEJČOCHOVÁ, Michaela (2009). *Čínské písmo*. Praha: Academia. 297 s. ISBN 978-80-200-1755-0.

Články a příspěvky

1. ALTMANN, Gabriel (1980). Prolegomena to Menzerath's Law. *Glottometrika* 2, s. 124-129.
2. ANDRES, Jan; BENEŠOVÁ, Martina; KUBÁČEK, Lubomír a VRBKOVÁ, Jana (2012). Methodological Note on the Fractal Analysis of Texts. *Journal of Quantitative Linguistics*, Roč. 19, č. 1, s. 1-31.
3. HAVRÁNEK, Bohuslav a HORÁLEK, Karel (1958). Osmý mezinárodní kongres lingvistů v Oslo. *Slovo a slovesnost*. Roč. 19, č. 1, s. 47-52. Dostupné z: <http://sas.ujc.cas.cz/archiv.php?art=923>
4. HŘEBÍČEK, Luděk (2007). Sémantické slapy v textových strukturách. *Slovo a slovesnost*. Roč. 68, s. 83-90. Dostupné z: http://www.kb-old.upol.cz/data/soubor_kb_807.pdf
5. HŘEBÍČEK, Luděk (2008). Filologie versus lingvistika: Nadvětné textové konstrukty. *Vesmír* 87. s. 488-490. Dostupné z: <http://www.vesmir.cz/clanek/filologie-versus-lingvistika>
6. KELIH, Emmerich (2010). Parameter-Interpretation of Menzerath's Law: Evidence from Serbian. In: *Text and Language: Structures·Functions·Interrelations·Quantitative Perspectives*. Wien: Praesens, s. 71 – 78.
7. LAPOLLA, Randy J. (2005). Wang Li (1900 – 1986). In: *Encyclopedia of Language and Linguistics*. London: Elsevier, s. 514–515. ISBN 978-0-08-044299-0.

8. LIU, Haitao a HUANG, Wei (2012). Jiliang yuyanxue de xianzhuang, lilun yu fangfa 《计量语言学的现状、理论与方法》 Kvantitativní lingvistika: současný stav, teorie a metody. 《浙江大学学报（人文社会科学版）》 [online]. Roč. 42, č. 2 [cit. 2013-04-10]. ISSN 1008-942X. Dostupné z: <http://www.journals.zju.edu.cn/soc/CN/abstract/abstract10497.shtml>
9. MOTALOVÁ, Tereza; SPÁČILOVÁ, Lenka; BENEŠOVÁ, Martina a KUČERA, Ondřej (2013). An Application of the Menzerath-Altmann Law to Contemporary Written Chinese. *Czech and Slovak Linguistic Review*. Roč. 3, č. 1 (v tisku).
10. UHLÍŘOVÁ, Ludmila (2005). Quantitative linguistics in the Czech Republic. In: *Quantitative Linguistik/Quantitative Linguistics: Ein internationales Handbuch/An International Handbook*. Berlin: Walter de Gruyter, s. 129-135. ISBN 978-3-11-015578-5.

Internetové články

1. ABRAHAMSEN, Eric (2012). Han Han's U-Turn?. In: *IHT Global Opinion* [online]. [cit. 2013-04-16]. Dostupné z: <http://latitude.blogs.nytimes.com/2012/01/26/blogger-han-han-controversy-on-democracy-in-china/>
2. ELEGANT, Simon (2009). Han Han: China's Literary Bad Boy. In: *Time Magazine* [online]. [cit. 2013-04-16]. Dostupné z: <http://www.time.com/time/magazine/article/0,9171,1931619,00.html>
3. ELEGANT, Simon (2010). Han Han. In: *The 2010 TIME 100* [online]. [cit. 2013-04-16]. Dostupné z: http://www.time.com/time/specials/packages/article/0,28804,1984685_1984940_1985515,00.html
4. HAN, Han (2012). Wo suo lijie de shenghuo 《我所理解的生活》 Život, jak mu rozumím. In: <http://blog.sina.com.cn/twocold> [online]. [cit. 2013-04-16]. Dostupné z: http://blog.sina.com.cn/s/blog_4701280b0102e7er.html

5. HAN, Han (2012). Wo de fuqin Han Renjun yiji ta de zuopin 《我的父亲韩仁均以及他的作品》 Mŭj otec Han Renjun a jeho dílo. In: <http://blog.sina.com.cn/twocold> [online]. [cit. 2013-04-16]. Dostupné z: http://blog.sina.com.cn/s/blog_4701280b0102e0eu.html
6. LI, Yuming 李宇明 (2013). Renshi yuyan de jingjixue shuxing 《认识语言的经济属性》 Ekonomický aspekt jazyka. In: 《中国社会科学网: 中国社会科学院主办》 [online]. [cit. 2013-03-01]. Dostupné z: <http://www.cssn.cn/news/677346.htm>
7. PILLING, David (2012). Lunch with the FT: Han Han. In: *Financial Times* [online]. April 21 [cit. 2013-04-16]. Dostupné z: <http://www.ft.com/cms/s/2/3be0e84e-8896-11e1-a727-00144feab49a.html#axzz2QhVk2QHq>
8. ZHENG, Qingting 郑青亭 (2012). Zhongguo shehui kexueyuan zai yazhou zhiku paimingzhong weiju di-yi 《中国社会科学院在亚洲智库排名中位居第一》 Čínská akademie společenských věd zaujala první místo mezi asijskými think tanky. In: 《人民网: 国际频道》 [online]. [cit. 2013-03-17]. Dostupné z: <http://world.people.com.cn/GB/16922861.html>

Webové stránky a příspěvky

1. Ai, Weiwei 艾未未 (© 2006-2009). Ai Weiwei 《艾未未》 Ai Weiwei. In: *bullogger.com* [online]. [cit. 2013-04-17]. Dostupné z: <http://www.bullogger.com/blogs/aiww/>
2. Beiyu jiaoshou: Li Yuming jiaoshou 《北语教授: 李宇明 教授. 北京语言大学》 Profesori Univerzity jazyka a kultury v Pekingu: Profesor Li Yuming (© 2006). In: 《北京语言大学》 [online]. [cit. 2013-03-17]. Dostupné z: http://www.blcu.edu.cn/blcuWeb/chinese/professor_intro.asp?bh=130

3. Biaodian fuhao 《标点符号》 Interpunkční znaménka (© 2013). In: *Baidu baike* 《 百 度 百 科 》 [online]. [cit. 2013-03-26]. Dostupné z: <http://baike.baidu.com/view/31516.htm>
4. Biaodian fuhao yongfa 《标点符号用法》 Všeobecná pravidla pro použití interpunkce (© 2013). In: 《 百 度 百 科 》 [online]. [cit. 2013-04-01]. Dostupné z: <http://baike.baidu.com/view/564500.htm>
5. Bihua 《笔画》 Tahy čínských znaků (© 2013). In: *Baidu baike* 《 百 度 百 科 》 [online]. [cit. 2013-03-25]. Dostupné z: <http://baike.baidu.com/view/168365.htm>
6. Category: Chinese bloggers (2013). In: *Wikipedia* [online]. [cit. 2013-03-22]. Dostupné z: http://en.wikipedia.org/wiki/Category:Chinese_bloggers
7. Duchang tuan 《独唱团》 Soubor sólistů (© 2013). In: 《 百 度 百 科 》 [online]. [cit.2013-04-16]. Dostupné z: <http://baike.baidu.com/view/2926923.htm>
8. Dunhao 《顿号》 Čárka pro výčet (© 2013). In: *Baidu baike* 《 百 度 百 科 》 [online]. [cit. 2013-03-26]. Dostupné z: <http://baike.baidu.com/view/54969.htm>
9. Fenhao 《分号》 Středník (© 2013). In: 《 百 度 百 科 》 [online]. [cit. 2013-04-1]. Dostupné z: <http://baike.baidu.com/view/54972.htm>
10. HAN, Han 韩寒 (© 1996 - 2013). Han Han 《韩寒》 Han Han. In: *blog.sina.com.cn* [online]. [cit. 2013-04-17]. Dostupné z: <http://blog.sina.com.cn/twocold>
11. Han Han 《韩寒》 Han Han (© 2013). In: 《 百 度 百 科 》 [online]. [cit. 2013-04-16]. Dostupné z: <http://baike.baidu.com/view/5972.htm>
12. Hanzi bujian 《汉字部件》 Prvek čínských znaků (© 2013). In: 《 百 度 百 科 》 [online]. [cit. 2013-03-30]. Dostupné z: <http://baike.baidu.com/view/3151455.htm>

13. Jiangehao 《间隔号》 Oddělovací tečka (© 2013). In: Baidu baike 《百度百科》 [online]. [cit. 2013-03-26]. Dostupné z: <http://baike.baidu.com/view/55027.htm>
14. KONG, Qingdong 孔庆东 (© 1996 – 2013). Dongbo shuyuan - Kong Qingdong de boke 《东博书院——孔庆东的博客》 Klasická akademie Dongbo – Kong Qingdongův blog. In: *blog.sina.com.cn* [online]. [cit. 2013-04-17]. Dostupné z: <http://blog.sina.com.cn/u/1198367585>
15. Kuohao 《括号》 Závorky (© 2013). In: Baidu baike 《百度百科》 [online]. [cit. 2013-03-26]. Dostupné z: <http://baike.baidu.com/view/54999.htm>
16. LI, Chengpeng (© 1996-2013). Li Chengpeng 《李承鹏》 Li Chengpeng. In: *weibo.com* [online]. [cit. 2013-04-17]. Dostupné z: <http://www.weibo.com/lichengpeng>
17. Li Yuming 《李宇明》 Li Yuming (© 2013). In: 《百度百科》 [online]. [cit. 2013-03-17]. Dostupné z: <http://baike.baidu.com/view/682964.htm>
18. Lianjiahao 《连接号》 Pomlčka (© 2013). In: Baidu baike 《百度百科》 [online]. [cit. 2013-03-26]. Dostupné z: <http://baike.baidu.com/view/55019.htm>
19. LIU, Mangyan (© 1999 – 2013). Liu Mangyan yuanchuang boke 《流氓燕原创博客》 Původní blog Liu Mangyana. In: *blog.tianya.cn* [online]. [cit. 2013-04-17]. Dostupné z: http://blog.tianya.cn/blogger/blog_main.asp?BlogID=19329
20. MURONG, Xuecong (© 1996 – 2013). Hulu hulu 《葫芦葫芦》 Hulu hulu. In: *blog.sina.com.cn* [online]. [cit. 2013-04-17]. Dostupné z: <http://blog.sina.com.cn/hawking>
21. MUZI, Mei (© 2013). Muzi Mei 《木子美》 Muzi Mei. In: *sohu.com* [online]. [cit. 2013-04-17]. Dostupné z: <http://muzimeiriji.blog.sohu.com/>

22. Pozhehao 《破折号》 Pomlčka (© 2013). In: *Baidu baike* 《百度百科》 [online]. [cit. 2013-03-26]. Dostupné z: <http://baike.baidu.com/view/55016.htm>
23. RAN, Yunfei (©1997-2013). Ran Yunfei boke: feihua lianpian 《冉云飞博客：匪话连篇》 Ran Yunfeijǔv blog: slang zlodějǔ. In: *www.163.com* [online]. [cit. 2013-04-17]. Dostupné z: <http://tufeilaoran.blog.163.com/>
24. RAO, Xueman (© 1996-2013). Rao Xueman 《饶雪漫》 Rao Xueman. In: *weibo.com* [online]. [cit. 2013-04-17]. Dostupné z: <http://weibo.com/raoxueman>
25. Shengluhao 《省略号》 Elipsa (© 2013). In: *Baidu baike* 《百度百科》 [online]. [cit. 2013-03-26]. Dostupné z: <http://baike.baidu.com/view/55004.htm>
26. Shuminghao 《书名号》 Závorky pro tituly (© 2013). In: *Baidu baike* 《百度百科》 [online]. [cit. 2013-03-26]. Dostupné z: <http://baike.baidu.com/view/55021.htm>
27. SIMA, Nan (© 1996-2013). Sima Nan 《司马南》 Sima Nan. In: *weibo.com* [online]. [cit. 2013-04-17]. Dostupné z: <http://weibo.com/simanan>
28. WANG, Keqin (© 2013). Wang Keqin 《王克勤》 Wang Keqin. In: *sohu.com* [online]. [cit. 2013-04-17]. Dostupné z: <http://wangkeqin.blog.sohu.com/>
29. What is R?: Introduction to R (2013). In: *The R Project for Statistical Computing* [online]. [cit. 2013-03-21]. Dostupné z: <http://www.r-project.org/>
30. Wo yuan gaikuang 《我院概况》 O naší akademii (2010). In: 《中国社会科学网: 中国社会科学院主办》 [online]. [cit. 2013-03-17]. Dostupné z: <http://www.cssn.cn/news/140195.htm>
31. Xiang de shufa 《象的书法》 Kaligrafie znaku pro „slona“ (© 2004 – 2012). In: *Handian shufa* 《汉典书法》 [online]. [cit. 2013-03-21].

- Dostupné z:
<http://sf.zdic.net/shufa/1123/5537f8fb44fcfab5ac9a353d5b5be03.html#ks>
32. „Xiang“ zi de jiben xinxi 《“象”字的基本信息》 Základní informace k čínskému znaku pro „slona“ (© 2004 – 2012). In: Handian 《汉典》 [online]. [cit. 2013-03-21]. Dostupné z:
<http://www.zdic.net/zd/zi/ZdicE8ZdicB1ZdicA1.htm>
33. Xin de shufa 《心的书法》 Kaligrafie znaku pro „srdce“ (© 2004 – 2012). In: Handian shufa. 《汉典书法》 [online]. [cit. 2013-03-21]. Dostupné z:
<http://sf.zdic.net/shufa/0815/3419eb9cbb79b0bdd53bce2b45d4e1f0.html#ks>
34. „Xin“ zi de jiben xinxi 《“心”字的基本信息》 Základní informace k čínskému znaku pro „srdce“ (© 2004 – 2012). In: Handian 《汉典》 [online]. [cit. 2013-03-21]. Dostupné z:
<http://www.zdic.net/zd/zi/ZdicE5ZdicBFZdic83.htm>
35. XU, Jinglei (© 1996 – 2013). Xu Jinglei 《老徐》 Xu Jinglei. In: *blog.sina.com* [online]. [cit. 2013-04-17]. Dostupné z:
<http://blog.sina.com.cn/xujinglei>
36. Yin hao 《引号》 Uvozovky (© 2013). In: Baidu baike 《百度百科》 [online]. [cit. 2013-03-26]. Dostupné z:
<http://baike.baidu.com/view/54994.htm>
37. Yu de shufa 《雨的书法》 Kaligrafie znaku pro „déšť“ (© 2004 – 2012). In: Handian shufa 《汉典书法》 [online]. [cit. 2013-03-21]. Dostupné z:
<http://sf.zdic.net/shufa/0611/7062ebdb5d8db69266b442dd9e37d89b.html#ks>
38. „Yu“ zi de jiben xinxi 《“雨”字的基本信息》 Základní informace k čínskému znaku pro „déšť“ (© 2004 – 2012). In: Handian 《汉典》 [online]. [cit. 2013-03-21]. Dostupné z:
<http://www.zdic.net/zd/zi/ZdicE9Zdic9BZdicA8.htm>

39. Yuan jigou 《院机构》 Struktura akademie (2011). In: 《中国社会科学网: 中国社会科学院主办》 [online]. [cit. 2013-03-17]. Dostupné z: <http://www.cssn.cn/news/141052.htm>
40. „Yuyan wenzi yingyong“ Bianjibu (Yingyong yuyanxue yanjiu zhongxin) 《〈语言文字应用〉编辑部 (应用语言学研究中心)》 „Aplikovaná lingvistika“ – Redakční oddělení (při Výzkumném centru aplikované lingvistiky) (2005). In: 《教育部语言文字应用研究所》 [online]. [cit. 2013-03-17]. Dostupné z: <http://www.yys.ac.cn/80/publish.htm>
41. ZHANG, Weiwei (© 1996 – 2013). Vivibear de BLOG 《Vivibear 的 BLOG》 BLOG Viviber. In: *blog.sina.com* [online]. [cit. 2013-04-17]. Dostupné z: <http://blog.sina.com.cn/vikingbear333>
42. Zhongguo shehui kexueyuan 《中国社会科学院》 Čínská akademie společenských věd (© 2013). In: 《百度百科》 [online]. [cit. 2013-03-17]. Dostupné z: http://baike.baidu.com/view/33292.htm#refIndex_1_33292
43. Zhuanminghao 《专名号》 Podtržítka vlastních jmen (© 2013). In: Baidu baike 《百度百科》 [online]. [cit. 2013-03-26]. Dostupné z: <http://baike.baidu.com/view/297482.htm>
44. Zhuozhonghao 《着重号》 Zdůrazňující tečka (© 2013). In: Baidu baike 《百度百科》 [online]. [cit. 2013-03-26]. Dostupné z: <http://baike.baidu.com/view/55028.htm>

Normy

1. GB/T 15834 – 2011. *Zhongguo renmin gongheguo guojia biao zhun: Biaodian fuhao yongfa* 《中华人民共和国国家标准: 标点符号用法》 Státní normy Čínské lidové republiky: Všeobecná pravidla pro použití interpunkce (2012). Beijing: Zhongguo biao zhun chubanshe.

Databáze

1. Wenlin Institute, Inc. 文林 Wenlin Software for Learning Chinese [software].Version 4.0.2.Wenlin Institute, Inc. Copyright © 1997-2011.

9 Seznam příloh

Příloha 1: Vědecký článek – vzorek A _____	165
Příloha 2: Článek z blogu – vzorek B _____	175
Příloha 3: Článek z blogu – vzorek C _____	178
Příloha 4: Tabulky uvádějící pozorování empiricky získaná kvantifikací novinového článku a povídky _____	186

Přílohy

Příloha 1: Vědecký článek – vzorek A

认识语言的经济属性

李宇明

2013-2-27 15:47:01 来源：《语言文字应用》(京)2012年3期

【作者简介】李宇明，北京语言大学教授，主要研究领域为语言学理论、现代汉语、心理语言学和语言规划（北京 100083）。

【内容提要】语言与经济的关系非常紧密，但在以往的社会语言意识中，语言的经济属性并没有得到清晰的认识。在人口流动和信息化两大驱力的推动下，语言对经济的贡献越来越显著，以至于成为不容漠视的经济现象。有些国家，语言对经济的贡献度竟然达到 10%。我国语言经济学研究已经起步，而且具有光辉的发展前景。语言经济学的重要任务是，认识语言在经济活动中的作用，认识语言经济的运行规律，研究语言对社会的经济贡献度，研究语言政策的成本及其产生的经济效益，探讨促进语言经济发展的政策环境和各种举措，发展语言产业，培育语言职业，促进语言消费，使国家和个人充分赚取语言红利。

Language is closely tied to economy; however, the economic attributes of language, in the sociolinguistic awareness so far, have not been clearly understood. Population mobility and information technology have been two driving forces that make language an important contributing factor of economic growth. Language, therefore, becomes a critical economic matter. In some countries, language accounts for about 10% in economic growth. In China, the study of economic linguistics has just started and foresees a promising future. The important task of the economic linguistics is to understand the role of language in economic activities, and to find out the operational rules of language economy, and to study the economic contribution of language for the society as well as cost-effectiveness of language policies. This paper will explore the policies and methods of promoting the development of economic linguistics in the hope of developing language industry, nurturing language professions, and promoting

language consumption to the benefit of the country and individuals alike.

【关键词】 语言/经济 language/economy

现代语言规划十分关注的三个概念是语言意识、语言政策和语言实践。语言意识也称语言意识形态，是指社会对语言的认识和态度，是语言政策和语言实践的思想基础，有什么样的语言意识，才可能有什么样的语言政策，产生什么样的语言实践。在当今社会的语言意识中，必须意识到语言的经济属性，从而在制定语言政策时自觉进行经济学的考量，并制定出在经济活动中能够充分发挥语言作用的政策。

一、语言是经济活动中不可缺少的要素

人类的经济活动与语言密不可分，而且在某些领域，语言和语言知识已经成为重要的经济资源。这可以从以下几个方面来看：

首先，语言能力是劳动力的重要构成要素。语言是人的本质属性之一，是人类最为重要的交际工具和思维工具，语言能力与劳动能力总体上呈正相关。比如：1. 由于各种病理原因而失去语言能力的人，如失语症患者、聋哑人等，是劳动水平相对较低、适应性较差的弱劳动力。2. 文盲只有口语能力，没有书面语能力，在脑力劳动在人类的经济活动中的比例逐渐上升的时代，文盲已经成为质量较低的劳动力，很多经济活动都无缘参与。3. 人的语种能力总体上看与收入相关，具有单语能力的人，与具有双语能力、三语能力的人相比，其经济收入总体上要低。

在一些特殊的劳动阶层和经济领域，语言能力的地位会更加重要。比如工程设计、广告策划、劳动管理等，文盲等没有书面语言能力者，是无法承担的；语言艺术家、电视节目主持人、导游、导购、公司售后服务人员等，都需要较高的语言能力；同声传译、国际组织雇员等工作领域，与单语人无缘。

在以体力劳动为主、社会分工较粗的时代，对劳动者的语言能力要求不高；但是到了信息化时代，社会分工急剧加细，脑力劳动的比重急剧加大，跨地区、跨国家的经济活动急剧增多，劳动者培养的时间越来越长，语言能力在劳动力构成中的比重也急剧增大。在每年大学毕业季的用人单位的招工要求中，虽不见得出现语言能力的字眼，但是仔细分析，关于语言能力的要求其实成了用人单位的主要考虑因素。语言能力在今天已经成为劳动力的重要构成要素，教育部门和劳动培训机构对此应有清醒认识。

其次，经济活动需要通过语言来组织进行。语言是信息最为重要的载体，人类的生活须臾离不开语言，离开语言社会就将崩溃；经济活动是人类社会生活的重要部分，经济活动需要通过语言才能组织起来。正因如此，许多经济学家、语言学家都比较重视经济学术语的规范，重视经济学文本的修辞，并由此形成了最早的经济语言学。经济语言学主要关心的是经济活动中语言使用的得体与效率，当然也试图利用经济学规律来解释语言现象。

再次，语言和语言知识已经成为重要的经济资源。在一些特殊的经济活动领域，语言及其知识甚至具有“生产资料”的性质。比如语言教育活动，语言知识、语言教育方法是教师的资本，通过语言知识的传授和语言训练，学生获得语言知识，并内化出语言能力。语言教育活动还需要有一系列的保障和评估系统，例如语言教学设施、教科书、工具书、各种录放语言的设备、衡量学生学习水平的考试活动及有关证书等，这些保障和评估系统，都涉及语言教育的经济问题。当然，语言教育有事业和产业之分，语言教育事业需要成本投入，涉及经济问题；语言教育产业则主要是经济活动。

再如语言文字艺术活动，其基本凭借是语言和文字，其成果是语言文字艺术产品。语言文字艺术可以粗分为语言艺术和文字艺术，小说、诗歌、话剧、相声、评书等是语言艺术，它们通过对语言的艺术运用形成艺术语言。艺术语言一般都是有声语言，或可以成为有声语言，并可以同其他方式结合形成戏曲、电影等综合艺术。文字艺术最为典型的是书法。除此之外，还有其他文字艺术，如将文字变形而形成的别有韵味的“文字图画”“文字雕塑”；如用文字或汉语拼音设计的商标、图案等等；如将金银珠宝镂雕成文字形状，或是将文字及其变形刻附在金银珠宝上。语言文字艺术不仅具有艺术价值，而且也往往具有较高的经济价值，形成附加值很高的语言文字艺术产业。

最值得关注的是，现代语言技术的发展所形成的现代语言经济。“语言技术”这一概念提出的时间不长，但语言技术已经有悠久的发展历史。文字的创制是古代最为重要的语言技术，口头语言借此打破了时间和空间的限制，可以传后达远。之后的笔墨纸砚、印刷术、打字机、电报、电话、传真、留声机、录音机、广播等等，都是至今仍用的最为重要的语言技术。现代语言技术是用计算机处理语言文字所形成的一系列技术，它使语言知识及其应用有可能成为工业标准，成为语言技术产品，形成各种专利，比如语言文字的各种规范标准，各种语言数据库，各种键盘输入法，各种处理语言文字的软件和计算机字库等等。现代语言技术使人类语言知识成为十分重要的经济资源。

二、语言对社会的经济贡献

但是，由于语言与人、与经济活动的关系太密切，密切到人们很少关注语

言的经济学意义。正如空气、阳光是人类片刻不能离开的，然而在人类相当长的历史长河里，它们都没有成为商品；只有到了空气污染严重、阳光常被遮蔽之地，它们才可能具有商品价值，例如现在的一些楼盘推销商，会将楼盘所在地的空气质量、房间的向阳状况等包装为卖点，空气和阳光在这里具有了商品意义。

语言的经济学属性较早引起人们的有意识关注，是语言能力与个人收入之间的关系。随着国际经济一体化进程的不断加快，人口流动的半径加长、规模加大、频率提高，移民成为当今社会的重要现象，并引发出一系列问题。国外很多研究发现，移民的语言能力同其就业状况、经济收入水平呈正相关。此类研究开垦了语言经济学的处女地。20世纪末期，信息化浪潮奔涌而起，狂飙突进，语言是计算机处理的主要对象，语言信息化是信息化的基础，语言、语言知识、语言技术等等，成为新经济的重要资源和重要增长点，许多信息产业都可以视为语言产业。

人口流动和信息化，是促进语言经济发展的两大驱动力，也是促使人们关注语言的经济学属性的两个重要方面。当今世界，促进语言经济发展的这两大驱动力更加强劲，人口流动几乎成为社会常态，信息化以加速度的态势发展，可以预见，语言对社会的经济贡献将会持续加大，语言的经济学属性将会不断彰显，因此，社会语言意识中必有语言经济的一席之地。

当带着语言经济的眼光来观照人类生活时，蓦然发现语言经济对人类社会是如此的重要。日内瓦大学弗朗斯瓦·格林(Fran ois Grin)教授的研究表明，瑞士语言的多样性，为瑞士每年创造500亿瑞郎的收入，约占瑞士国内生产总值的10%。瑞士有德语、法语、意大利语、罗曼山地语四种语言，瑞士公民一般都能够掌握三种语言，综合院校的大学生需要掌握四种语言。瑞士还是联合国、欧盟以及很多国际组织的所在地，有许多语言资源可以利用。瑞士经济的发展，语言在里面占了很大的比重。10%的说法可靠与否，由经济学家去判断，但是瑞士的经济发展跟它丰富的语言资源确有关系。

世界许多国家都开始注意语言产业的发展，并不断有相关的报告问世。据报道，美国的命名产业1999年的年产值就达到了150亿美元。有人估计，全世界翻译市场年产值可达1万亿元人民币；全球英语教育市场，除大学和政府培训机构外，约有600亿美元的规模。《2009年欧盟语言行业市场规模报告》是笔者见到的当前最为全面的语言行业状况报告。该报告指出：2008年欧盟成员国的语言市场总产值达84亿欧元，其中语言技术领域的产值为5.68亿，电影字幕和配音领域为6.33亿，语言教学领域为16亿，会议组织中的多语言服务为1.43亿。该报告预测，欧盟2015年语言行业的实际产值可达到200亿欧元。

我国正处在人口大流动时期，上亿农村人口向城市流动，劳动力国外输出渐成规模，学生出国留学人数与日俱增，境外回内地、海外来中国的学习者、投资者、工作者也逐渐增多。我国语言教育产业、语言翻译产业拥有强大活力，据统计，仅英语学习市场年产值已超过 100 亿元人民币，翻译市场年产值约 120 亿元人民币。我国信息化的发展日新月异，2012 年网民即将突破 5 亿，手机用户 9 亿，其中手机上网用户 3.5 亿；信息产品的社会普及率速度惊人，语言信息产业具有巨大的发展潜力。

由于语言经济学刚刚兴起，人们对语言经济的业态状况还不怎么了解，语言经济的数据采集系统尚未建立，甚至缺乏有效的语言经济计算方法，因此，还无法对语言经济进行全面而科学的统计。但是仅上面的事例已经颇具魅力，由此已可窥见语言经济的巨大社会意义，感受到认识语言的经济学属性的巨大社会意义。

三、促进我国的语言经济研究

与语言经济的社会实践相比，我国对语言经济的认识显然滞后。但是近些年来，学界、业界和政界已开始关注语言经济问题，呈现出良好的发展势头。

山东大学黄少安教授是我国语言经济学研究的首倡者，他率先招收语言经济学的博士研究生，他的团队连续召开中国语言经济学论坛，并成立了语言经济研究中心。2012 年 3 月 2 日，黄少安、苏剑、张卫国三位发表的《语言经济学与中国的语言产业战略》，基本上代表了我国学界在语言经济方面的认识。南京大学中国语言战略研究中心徐大明教授，带领他的团队在研究语言政策、语言国情的同时，开展了语言经济学的理论与实践研究，成果引人注目。广州大学屈哨兵教授对语言服务保持着浓厚的研究兴趣，今年初还举办了“语言服务研究高级论坛”。

国家语委是专司语言事务的国家部门，近些年来在工作实践中发现语言经济问题十分重要，认识到语言不仅是国家的“软实力”，而且也是“硬实力”。国家语委全力支持山东大学、南京大学、广州大学等高校的语言经济研究，还于 2008 年 12 月 29 日支持商务印书馆成立了“中国语言资源开发应用中心”，中心的宗旨是“致力于把语言及语言知识转化为生产力和文化商品”。《国家语委“十二五”科研规划 2011 年项目指南》，首次把“语言经济与语言产业发展战略研究”列为重要的科研方向，表明语言经济问题已经开始进入国家的语言规划。在 2010、2011 年北京市“两会”上，北京市人大代表贺宏志先生连续提出《关于发展我市语言产业的建议》和《加强语言文化建设，促进语言产业发展》的建议，语言经济的话题首次提到了地方人民代表大会的议

坛上。

2010年9月28日，在国家语委支持下，北京市语委的研究基地——“北京语言产业研究中心”，在首都师范大学揭牌成立。这本来是一个平常的日子，但在中国语言规划史上也许会不平常，因为这是我国第一个以“语言产业”为专门研究对象的科研机构。该中心建立伊始，就着手调研北京语言产业的发展状况，建立北京语言产业数据库，思考语言产业的发展政策；据说还筹划举办我国第一届语言产业论坛，编辑出版《语言产业研究通讯》，翻译《2009年欧盟语言行业市场规模报告》，出版《北京语言产业的发展及政策研究》等。理论与实践并重，引进与创新同举，近期与长远共谋。令人欣喜的是，在北京语言产业研究中心成立仅一年多的时间里，就撰写了我国第一部专门研究语言产业的著作《语言产业导论》。

《语言产业导论》虽不能说字字珠玑，但却处处闪光。特别是关于语言产业的分类、语言产业的要素分析、语言产业的业态梳理及其案例征引等等，读起来引人入胜，不忍释卷。这部著作给人许多新理念、新知识、新数据，而且能够引发读者许多质疑与思考，诸如：语言真能生钱吗？语言产业真的存在吗？哪些职业是语言职业？怎样统计语言对社会的经济贡献度？哪些消费属于语言消费？语言消费与语言产业、语言职业的发展有何关系？个人如何收取语言收益？国家如何赚取语言红利？……只提供知识的著作，培养的是被动消费型的读者，只能给社会以精神填充；而能够引发质疑与思考的著作，培养的是主动创造型的读者，能够促进社会的精神再生产，甚至引发社会行动。

尽管有如上的进展，但我国的语言经济学研究还处在起步阶段，研究工作需要一个案例一个案例地收集，一步一步地前进。不过宏观上看，语言经济学的研究应当完成两项基本任务：

第一，研究语言对社会的经济贡献度。这需要有一个科学的研究框架，需要广泛搜集与语言相关的经济领域的的数据，通过大数据的统计分析得出结论。当前的最大困难是，我们还没有一个有效的计算语言经济的方法，哪些产业属于语言产业？如何计算语言产业的经济效益？在非语言产业中语言是否能够产生经济效益？如果能够产生经济效益的话，如何计算非语言产业中因语言而产生的经济效益？当然，当前也无法从国家现有的统计数据中得到语言经济方面的数据，因为统计口径中没有语言经济。

第二，研究语言政策的成本及其产生的经济效益，探讨促进语言经济发展的政策环境和各种举措。过去对语言政策的评估，主要侧重于社会效益；当认识到语言的经济属性之后，还应当对语言政策进行经济学的评估。比如我国民族地区的双语教育，从语言经济学的角度看，需要多大的经济投入？这一政

策将为民族地区的经济发展带来多大效益？掌握双语的学生的经济收入将会比单语学生有多大提高？进而可以改善多少民族家庭的经济状况？再如，我国有一系列促进经济发展的政策，对这些政策应当进行语言经济学的考察，看看是否有利于促进语言经济的发展，应当做哪些必要的政策补充，或是在这些政策环境中设计出促进语言经济发展的各种举措。

四、结束语

在当今社会，不包含语言的经济属性的意识，不是与时代契合的语言意识。在语言经济可能影响到 10% 的经济生活的今天，社会必须树立清晰的语言经济意识，仔细观察语言经济活动，全面收集语言经济数据，认识语言经济的运行规律，发展语言产业，培育语言职业，促进语言消费，使国家和个人充分赚取语言红利。

最后我想引用林毅夫先生的一段话：“这些年来我在各种场合经常讲中国经济现象是经济学研究的一个大金矿，研究中国的经济问题有可能产生一批世界级的经济学大师。我的信念源自经济理论的作用在于解释经济现象，其贡献的大小由所解释的现象的重要性决定。……经济学成为一门独立的社会科学以来，世界级的经济学家大多先后产生于作为世界经济中心的英国和美国。中国的经济规模很有可能在 21 世纪 30 年代超过美国，中国将有可能逐渐成为一个新的领导经济学思潮的国际中心。如果我的乐观预测是正确的，中国经济学界的第一个诺贝尔经济学奖获得者，最有可能是来自于从事制度经济学研究的经济学家。”^①如果林毅夫先生的推断是正确的话，我国在语言经济学的研究上也完全有可能后起居上，因为我国有十分丰厚的语言经济学研究资源。

本文原是《语言产业导论》的序言，这次发表在内容上作了一些补充。

注释：

①林毅夫《“当代制度分析前沿系列”总序》，见鲁宾斯坦（2004）。

【参考文献】

[1]薄守生. 语言规划的经济学分析[J]. 制度经济学研究, 2008, (2).

[2]蔡辉. 语言经济学：发展与回顾[J]. 外语研究, 2009, (4).

- [3]陈章太. 语言资源与语言问题[J]. 云南师范大学学报(哲学社会科学版), 2009, (4).
- [4]贺宏志, 陈鹏. 语言产业导论[M]. 北京: 首都师范大学出版社, 2012.
- [5]黄少安, 苏剑, 张卫国. 语言经济学与中国的语言产业战略[N]. 光明日报, 2012-03-02.
- [6]江桂英. 中国英语教育: 语言经济学的视角[M]. 厦门: 厦门大学出版社, 2010.
- [7]李现乐. 语言资源与语言经济研究[J]. 经济问题, 2010, (9).
- [8]李现乐. 语言服务与服务语言——语言经济视角下的语言应用研究[D]. 南京大学博士学位论文, 2011.
- [9]李宇明. 中国语言规划论[M]. 北京: 商务印书馆, 2010a.
- [10]李宇明. 中国语言规划续论[M]. 北京: 商务印书馆, 2010b.
- [11]李宇明. 语言也是“硬实力”[J]. 华中师范大学学报, 2011, (5).
- [12]林勇, 宋金芳. 语言经济学评述[J]. 经济学动态, 2004, (3).
- [13]刘国辉, 张卫国. 语言经济学在中国的发展: 2009(首届)中国语言经济学论坛综述[Z]<http://weiguozhang.blog.sohu.com/138597516.html>, 2010-04-07/2010-05-12.
- [14]刘建达. 试论语言中的经济现象[J]. 山东外语教学, 1989, (3).
- [15]鲁宾斯坦. 经济学和语言(钱勇、周翼译)[M]. 上海: 上海财经大学出版社, 2004.
- [16]马慈君. 语言经济学视野下的大学英语教育[J]. 云南民族大学学报(哲学社会科学版), 2010, (1).
- [17]宁继鸣. 汉语国际推广: 关于孔子学院的经济学分析与建议[D]. 山东大学博士论文, 2006.

- [18]屈哨兵. 语言服务研究论纲[J]. 江汉大学学报, 2007, (6).
- [19]宋金芳, 林勇. 语言经济学的政策分析及其借鉴[J]. 华南师范大学学报(社会科学版), 2004, (6).
- [20]汪丁丁. 语言的经济学分析[J]. 社会学研究, 2001, (6).
- [21]王世凯. 略论我国语言资源的开发与利用[J]. 云南师范大学学报(哲学社会科学版), 2010, (5).
- [22]王振顶. 汉语国际传播的语言经济学研究[J]. 云南师范大学学报(对外汉语教学与研究版), 2009, (6).
- [23]徐大明. 有关语言经济的七个问题[J]. 云南师范大学学报(哲学社会科学版), 2010, (5).
- [24]许其潮. 语言经济学: 一门新兴的边缘学科[J]. 外国语, 1999, (4).
- [25]杨依山. 语言经济学理论框架初探[J]. 山东社会科学, 2007, (10).
- [26]张卫国, 刘国辉, 陈屹立. 语言与收入分配关系研究评述[J]. 经济学动态, 2007, (7).
- [27]张卫国. 语言的经济学分析: 一个初步框架[D]. 山东大学博士学位论文, 2008.
- [28]张忻. 语言的经济学与大学英语教育[J]. 中南大学学报(社会科学版), 2008, (3).
- [29]Coulmas, F. *Language and Economy*, Oxford: Blackwell Publishers Ltd, 1992.
- [30]Grin, F. The economics of language: Match or mismatch? *International Political Science Review*, 1994, 15, 27~44.
- [31]Grin, F. The economics of language: Survey, assessment, and prospects, *International Journal of the Sociology of Language*, 1996, 121, 17~44.

[32]Grin, F. European research on the economics of language. Homepage of Languages and the Economy: Canada in an International Perspective ,2000,11/13/2000.

[33]Grin, F. Language planning and economics. Current Issues in Language Planning, 2003,1, 1~66.

[34]McManus Labor market costs of language disparity: An interpretation of Hispanic earnings differences, The American Economic Review, 1985,4, 818~827.

[35]Mettewie, L. & L. van Mensel Multilingualism at all costs: Language use and language needs in business in Brussels, Sociolinguistic 23: Language Choice in European Companies, 2009.

[36]Vaillan-court F. The economics of language and language planning. Language Problems and Language Planning, 1982,2,162~178.

责任编辑：隋萌萌

Příloha 2: Článek z blogu – vzorek B

我所理解的生活

多天前参加比赛，来了一个久未见面的朋友。他现在的工作是在给明星做经纪。整个周末他就在我们车队的帐篷里。周日分别，他对我说，其实我的自我开发做的并不好，形象管理有问题，如果有职业的经纪人可以打理一下，必然远不是今天的模样。这样，回去给你一个总结的邮件。

刚才他打来了电话，说你问题太多，邮件说不清楚，比如你在比赛那天一直双反你知道么？我当时就晕了，我只知道双规和单反，双反我真不知道。弄半天才明白，所谓双反，原来是衣服穿反了，而且内外和前后都反了。我说我出门太急，真没注意，也没人提醒我，难怪一整天觉得脖子有点勒。

朋友说这个问题不大，你本来就粗心，但是容易被人取笑，但要命的是，你在车队帐篷的沙发上乱睡觉，你睡觉的时候总共有十二个人来拍过你的睡姿，五个是挂记者证的，四个是车队成员，三个是其他车手，其中有两个是故意拍丑态大头照的。我查了一下，其中五个人发微博了。有一张照片很难看，影响形象，你身边也没人拦着人家拍照，这在我们这行里是绝对不允许的。我说这我也没办法，熬夜看欧洲杯，的确睡眠不够，你教我怎么才能睡的玉树临风？

朋友继续教育我，面部表情是其次，关键是你团着身子睡，手还一直塞在你自己的裆部，这个猥琐的动作绝对是破坏形象的，照片如果上传，有些网友看见了容易反感。我说我又没把手塞在那些网友的裆部，我碰了自己的鸟，关他们鸟事。反感就拉到呗。朋友说不是这样的，你是一个公众人物，现在又是微博的时代，谁都能随手拍，越夸张的传播越快，你要确保自己的每一张照片不能影响你的形象，比如你那个手放的位置不对，很容易被一下转发数千条。我说这我实在没办法，空调温度太低了，只要一冷，我就自动睡成捂裆派了，从小就这样。总不能我睡觉，雇几个保镖拦着不让拍照，这也太装逼了。

朋友还指出了一堆问题，比如随意让人合影，人家递过来什么都签，会留下隐患。我说不，人家如果递过来一百块人民币我就不签。朋友肯定道，不错，你还算有这个意识，我们行业里有明星在递过来的钞票上签字的，结果被网友骂死。破坏人民币肯定不好。我说不是的，是因为我不想把自己的名字和老毛放在一起。

朋友痛心疾首道，你看，你这种话又不能乱说，得罪的人太多。你在车队聊天也是这样，什么都说，而且常出脏话，你要知道，如果现场有一个不怀好

意的，把你说的那些用手机录下来，放到网上，是很大的负面新闻。你知道当时帐篷里多少人，十八个，你都认识么？我回答说有几个不认识。朋友听筒差点掉地上：有几个不认识你就那么说话？你考虑过后果么？你一睡醒就和人合影，有一撮头发翘的跟天线宝宝似的，人家还开着闪光灯，照片效果可想而知。你看你衣服的配色，是很乡土的，最关键，再不拘小节，裤子拉链还是要拉上的。总之，你太随便了，也没有一个专业点经纪人帮你，你如果不严格的对自己的形象进行系统的管理，就不能保持神秘感和名人的气质。你如果对自己有一个好的定位，有合适的人帮你运作和服务，调整一下你的社交圈子，你能赚的远比现在多很多。你告诉我，你打算怎么经营自己？你是怎么想的？

我说，整个周末只在想一个问题，我和对手差了零点三秒，我该怎么追回来？我能惦记着出门要穿衣服已经不错了，哪还顾得上搭配。

挂了电话，夜深人静，回想朋友所言，有些也对。我在帐篷里口无遮拦，对所有人没有设防，要是真有人偷拍偷录或者微博直播的，的确会惹麻烦。对陌生人的提防与否取决于你的出厂原始设定，我喜欢先把人设定成好人，再从中甄别坏人，有些人反之。但所谓的甄别方式其实就是被坑一次。我相信以诚相待，也相信倒霉认栽。

至于衣着，这个夏天我就买了十件白体恤，以往冬天我也就两件黑皮衣，鞋子就盯着那么一两双穿，我是去比赛的，又不是去比美的，赛车开砸了我在乎，衣服穿难看了我真不在乎。

我所理解的生活就是除了造谣以外，去造其他一切东西。我心中的造化，就是你创造了多少文化。既然 30 年前，我以一挑亿，跑赢了其他所有的精子，那么我必然生来就是牛逼的，我来到这个世上，总要留下点痕迹。我承认衣着光鲜，举止优雅也是一种对美好的创造，但这方面我不太拿手。我承认这个社会，很多人觉得你只要不说脏话，说点假话，空话，套话，造谣，大肆的造谣，疯狂的造谣都不算是道德败坏，但我觉得反之，并且还要对道貌岸然的傻逼们加一个操字，是的，这会让那些道德惊诧家们浑身颤抖，严厉批判，大呼小叫，满地打滚，然后突然起立，开始审判，解决之道就是再说一次，操。我就操了，但我既不操你也不操你全家，我操这世道，这世道觉得文绉绉的诬陷没问题，这世道让那些不说粗话但最缺德的人能做道德评判家，这世道让那些话不脏但心眼脏手段脏的人当道，这世道能任意颠倒黑白混淆是非，这世道觉得公众人物或者随便谁说一个操字就不应该，那就操翻这世道。

我所理解的生活就是做着自己喜欢的事情，养活自己，养活家人。生活它不是攀爬高山，也不是深潜海沟，它只是在一张标配的床上睡出你的身形。我也不觉得留有遗憾是一种缺憾美，相比之下，干砸了倒是一种美。我喜欢的事

情远不止写点东西和赛车，我还做很多事，有些做的不够好，有些做的很失败，和朋友聊天时，我直接告诉他们，这事我特喜欢，也干过，但我真的不适合，丢人了。我就最讨厌听见有人这么说，要是我去干这事，一定比某某某干的好。滚。你在台面上看见我成功一次，我在台面下就干砸十次，那又如何，我又没死，不停的干就行了，人们只会记住你成功的那一次。

我所理解的生活就是和自己喜欢的一切在一起，我曾经在快餐厅看上一个姑娘，犹豫五分钟，没敢去和人家说话，结果人家走了，我到现在都很遗憾。在那一刻，我就是白痴，我去了又如何，最坏的结果无非就是他男朋友从厕所里出来。哪天若要死了，遗憾这事没干，那事没干，还不如自吹这事干成了，自嘲那事干砸了。我现在干的事足够多，陪伴家人爱人和孩子，每年比赛接近二十场，又开始写新的小说和游记，除了偶然进棚拍杂志，其他时间真没有精力来捋饬自己，更没心思去考虑什么形象和定位的问题，觉得我观感欠佳的，挪步就是，我只负责制造作品，不负责用户体验，也没有售后服务，更不会根据大家的口味来改进。你若喜欢，便是晴天，你若讨厌，也是晴天。谢谢这位朋友给我的忠告和精心的设计，我知道我会为我的性格和生活方式吃无数亏，吞无数恶果，但至少大到理想，小到闪念，我几乎都没有放过，所以就算我的生活里充满挫败甚至后悔，但遗憾并不多。朋友，感谢你所说的一切，世间万千种宠爱，无数种人心，得之我幸，不得我也没什么不幸。但我只认可一点，就是出门再匆忙，裤子拉链还是得拉好。

Příloha 3: Článek z blogu – vzorek C

我的父亲韩仁均以及他的作品

今天一大早，我的父亲给我电话，说写了一篇文章来说明一下，问我这样写能不能把事情说清楚了。我觉得特别的凄凉。一开始，他们说我有团队，并重金鼓励网友举证，结果千万网友中没有人能举证出身边的亲朋好友属于我的写作团队，于是他们又说金波是我少年成名的推手，结果发现金波 98 年的时候还在河南的一个罐头厂工作。他们最后的一招就是把所有的脏水泼到了我父亲身上。说我的父亲替我写了我少年时候的文章，因为我少年时候的文章特别的老成，不可能是 17 岁的学生写的。这个非常可笑，我在很多的场合说过，我小时候喜欢阅读钱钟书梁实秋和很多民国作家，因为我觉得他们文字好。在一个人刚开始写文章的时候，你阅读谁，必然模仿谁。而了显得渊博和少年老成，我还摘录了很多典故或者英语，准备随时引用在文章里显摆。而我的父亲则对民国文人兴趣不大，所以我们两人的文字非常好辨认。如果这样去加罪文学作品，我在十七岁的时候还发表过两篇写大学生活的小说，十五岁的时候还发表过写成年人生活的散文，当时我非常得意于杂志社的编辑都不知道我的真实年龄，现在想来，这些岂不都是死罪。

我只想说明，这就是一个好作家的底子。我的作品里的“我”如果一直是一个初中生形象出现，那岂不是和我喜欢的作家钱钟书梁实秋相去甚远？只要文学作品里的“我”不符合我的实际身份，我就是在造假，就是有人代笔？真是可悲，这么说的人恐怕从来没有体会到文学的乐趣。对于一个少年来说，文学的快乐就是可以把自己藏起来。不过对于一个写作者来说，还要对人解释这些，的确更可悲。

我甚至发现还有伪科学证明法，忽悠了不少人。有理科生做了一个科学统计，选取了八九本书，其中有我父亲的书为 A，我的四本书 B，C，D，E，其他作家的书 G，F。分析书中关键词出现的次数，比如三个“的”“得”“地”，比如“因为”“所以”，结果发现，我父亲的书和我的那四本书出现这些词的总次数差不多，而另外两本书明显要数倍与 ABCDE。实验得出的结论是，我和我父亲的风格雷同。这篇文章充满了我看不懂的专业术语，很多试图证明我的书是我父亲代写的朋友也像捡到了宝一样兴奋不已转发不止。甚至其中还有科普人士，他们认为最终还是科学和数据来说话了。我真的不知道该说什么好，A，B，C，D，E 都是十万字，G 是三十万字，F 则超过了五十万字。所以结果当然是 ABCDE 的常用词出现次数差不多，而 G 是他们的三倍左右，F 则是五倍左右。像我这样的理科白痴也知道做这种对比的时候，抽样对象的字数得是一样的吧。我不明白这位理科生为什么这么恨我，虽然我一直对理科生特别有好感，我小

时候的梦想就是科学家。但是理科更需要严谨，能在完全缺乏条件的情况下依然把实验做的津津有味并得出结论，也算有本事。而文科生们更不能“我觉得他 17 岁写的文章太成熟了，所以必然是中年男人写的”来论证问题吧

好了，下面请欣赏我的父亲@韩仁均叔叔 写的一篇文章，文章有点长，希望大家能耐心的读完，再来判断我的文章是不是我的父亲代写的。最后还附有我的父亲以前曾经发表过的作品，我的父亲为人正直，可爱，他的文笔非常的淳朴流畅，根本就不像我十六岁那样，看似老道，其实做作。我父亲念书的时候没有初中高中，根本没学过也不会英语，其中又经历文革，构陷者居然能想象出我父亲帮我写出一本讲初高中生活的校园长篇《三重门》。也欢迎大家拿我父亲的文章和我的文章来对比。很多人劝我不要再与这些人纠缠，因为他们就是要找各种茬搞臭你，你说什么都没有用。他们的逻辑就是走街上看人不爽，上前一口咬定你十年内必然杀过一个人。如果你不能自证十年内每一分钟的去处，那么我枪毙你是天经地义。关键是居然真有看客似乎觉得他们很正义，是替天行道。我知道有些人是不会错过这个好机会的，但是我相信明白人如果能够看到现在，也都能明白是怎么回事了。感谢所有支持我的朋友。今天我希望为我父亲讨一个清白，不过，这真是一个荒谬的时代，加害者像个原告一样，大摇大摆，不需要任何证据，只需要想象力，就能够损毁你和你家人的名誉，而且什么罪名都敢往你头上加，而受害者却要像个被告一样，不停的出示着人证物证，也未必能还自己一个清白：

说说我自己

韩仁均

之所以想说说自己，是因为最近忽然有人发掘了我的超凡能耐，把我描述得能操控一切，既能写出《三重门》让韩寒“默写”，又能一手操控组织严密完整的全国新概念作文大赛，想买通谁就买通谁，想得什么奖就得什么奖，并且能一路代笔代思想，最起码有一个微型耳麦佩在韩寒耳边，可以随时指导这话怎么说，这事怎么做，而且能一路走红 10 多年。现在就差操控全国公务员考试的事还没来得及上提出来。当然说说我自己我也不会把自己贬得一钱不值，我只是把一个真实平凡的我告诉大家。

我生于 1957 年，韩寒之前说我生于 1958 年，那是他把她妈妈的年龄记成我了。小学中学都在村里念的，中学当时只有四年，叫做中一中二中三中四，没分初高中。读好四年中学后就在村里务农。1977 年恢复高考制度后，拼命复习，但理科外语都不懂，只得报考文科类的，后来被华东师范大学（当时好像叫上海师范大学）中文系录取。1978 年初入学。进校后，在学生名单上看到了好几个当时已经非常有名的作者，感觉以后要在这里和这么些同学一起度过几年的学习生涯很荣幸的。但事实上正经的课都没上过一节，第一个星期只是开会学

习劳动之类，还有新生身体复查。结果是 GPT100 多一点，肝功能不正常（当时指标 40 以下算正常），于是就住进了师大后门那儿的肝炎隔离病房。住进去后检查的范围更大，还查出是澳抗阳性（就是现在的大三阳小三阳之类），被定性为乙型肝炎。以后每隔一些时间查一次，一直没有全部正常。最后 10 个月后，好像是 1978 年 12 月份，做了最后一次检查，同时抽了三份血样，分送华山医院等三家医院检验，三份结果完全不同，一份正常，二份不正常。于是，被认为还没痊愈，就被退学。因为大家都谈肝色变，我在病房里也很识相，不出去接触人，所以还没来得及认识一个同学就离开了学校，回到了家乡亭新公社（乡和镇的叫法是后来的事）。因为多少算考取过大学了，所以回来就到了亭新公社的文化站工作。当时的文化站和现在的文体中心完全不是同一回事，就一间办公室，就我一个人。所做大部分的工作就是为公社机关服务，比如开会拍照，并自己冲印贴在公社门口的画廊里，布置会场等等。文化站里的上级业务指导对口单位是县文化馆和县文化局。

县里每年对各公社（乡镇）有业务考核指标，比如群众文艺创作、演出等等。为了应付这些考核，或者说是工作吧，各个公社乡镇就得组织人员进行创作。组织者自己当然也得写。这个时期，硬着头皮学写故事、表演唱、小散文等等。金山有故事创作演讲的传统，县里也经常组织培训等。后来就陆陆续续的写了些故事及散文等东西。发表在上海的解放日报市郊版，上海的故事会、故事大王及外省市的一些故事刊报刊上。因为我到文化站后就开始学写些小东西，但又感到自己的名字太过普通，就取了个笔名叫韩寒，当时韩寒还没有出生，但实际上没怎么用到，只在一二个小豆腐块上用了一下，后来觉得舍不得，而且本来发表东西又比较少，所以就决定把这个笔名作为我未来的儿子或者女儿的名字。所以 1982 年儿子出生后就叫韩寒了。韩寒出生那年国家又开始了高等教育自学考试，我因为受英语和理科的影响，也只能选择了文科，当时还是选择了华东师大中文专业的考试。自考每年考两次，每次最多可以报考 4 门课，大专在 10 门课左右，但还好工作时间比较空和自由，所以看书的时间比较多，晚上也是差不多全部时间花在读教材背题目上，当时也还不会打麻将，也没其他娱乐活动，电视机也只有一个黑白 14 英寸金星吧。我每次报的 4 门课一般都能通过 2 门，分数大都是 60 分多一点那种，只够及格，最高也只考过 70 几分吧。所以考了两年半，就完成了华东师大中文专业专科 10 门课程的学分，拿到了自学考试专科毕业证书。专科考过后，本科阶段还要考近 10 门课，觉得一些课太难了，像古代汉语、古代文学等等，所以后来就不考下去了（但韩寒多次在采访里把我说成是本科，因为他对这些专科本科的本来就没有什么概念）。但说实话，这种考试真的只为考试而考试，所学 10 门课在工作中根本用不到的，考过后也很快就忘掉了。但这个文凭也给我带来了好处，后来 1985 年文化站转体制时派到了用场，成为我转成事业单位编制的一个有利条件，户口又迁出农村到城镇，户口在当时的中国决定着好多东西。后来直到 1994 年底，当时时兴办区县报，金山也要办一份金山周报。当时文化局的一位领导调

到县委宣传部筹办这份报纸，他觉得我去做这个工作也合适，就把我调了过去。当时的金山报 4、5 个采编人员，采编合一。甚至划版样、校对都是自己负责，一周一期。期间也还评了个编辑还是记者的中级职称。我一直没有入党，因为我没想过要进官场混，也自知没能力在官场混。在金山报时后来要提一个副主编，部领导决定民选，大家无记名投票，我被大家暗算。后任的宣传部长对我说，可惜你不是党员，是党员的话一切都好办。甚至还想有意“培养”我入党。我一笑置之。再后来我觉得我没义务付出那么多去负那个责任，就辞去了副主编的职务。2005 年底我们几个金山报的元老就各奔东西，我就去了我们区里的清水衙门文化局工作，做一个没有实职的主任科员，直到 2008 年底提前退休。

我到金山报工作后，韩寒还在亭林读小学。因为自己没什么社会背景，还有觉得当时县城朱泾的罗星中学教学质量什么的在全县算是比较好的，全县比较好的学生大都在这所学校里读书，里面有几个特色班，就赞助了几千元钱（相当于择校费吧），让韩寒的初中在罗星中学就读。我想等韩寒长大工作后，他的大部分同学将会是我们这个县里方方面面管事的头面人物，有这么一个人际关系基础，那对他的工作和发展会有所帮助。而我自己认识的朋友中，职位最高的也只是处级干部吧，我帮不上韩寒什么，以后只有靠他自己了，当时这么想的。记得韩寒进罗星中学后，摸底考试几门功课平均考了 91 分（满分 100 分），当时他自我感觉非常好，想这下总会名列前茅了，不料只在班级第 50 名左右，倒数前列，不禁感叹那些同学读书成绩真好。他比较牛逼的是作文，当时一篇介绍自己的作文《我》，让当时的语文老师彭老师赞不绝口。有时为炫耀，一节作文课写二篇。一开始韩寒的成绩还算比较均匀发展，但由于和教数学的班主任老师关系老是处不好，影响了他对这门课的兴趣，但中考前恶补一下后他的数学还是考得可以的，反而是语文考得不理想。他的应试作文在那种正式的考试模式中老是得不了高分。在区里的传统作文比赛里也能拿二等奖，因为他的文章不是传统作文比赛喜欢的类型。但他还是比较喜欢看书的，尤其喜欢民国的，钱钟书和梁实秋等人的文章，家里的一本《围城》不知被他翻过多少遍，第一本翻烂后我又买了一本。因为喜欢，所以他后来在第一本书《三重门》里刻意的模仿他的偶像钱钟书《围城》的风格一点也不奇怪。

韩寒中考考了 468 分，有体育长跑比赛第一名的 8 分加分，就是 476 分，松江二中因为他长跑的成绩好，就又降低了几分特招了他。他当时是寄宿在学校的，不是每个星期都回来，那时松江到金山还得要转二次公共汽车。我发现他写《三重门》是在他差不多要写好的时候。我就对他说，要不要我拿去帮你复印一遍，不然弄丢后无法弥补。他同意了，说等写完后。写完后我就拿到对外营业的金山县图书馆复印室去自费复印了一份。我是这个时候才有了看的机会，总的感觉是非常幽默，有点像《围城》的那种笔法，文笔非常老练，而且我猜想，书中那个主人公身上，可能有他自己的影子。书中描写的生活，也从初中延续到了高中。这是一种没有生活的人想象不出来的情景。一般作者的第一本

书，以自己的生活为基础，模仿自己喜欢的偶像风格写很常见。现在看到有人竟怀疑《三重门》是我写的，那我真要谢谢他们的抬举了，我要写得出《三重门》，我早不是现在的我了。这种只要有脑子，而且能正常思维的人都想得明白的。不是一代人，文笔和经历完全不一样，你能写得出那种情景那种意境那种感觉吗？现在一些人觉得他们不可能做得到，所以断定韩寒在他们这个年龄也做不到。一个 17 岁孩子的文章是他父亲代写的并且还能走红中国十多年，是在编神话吗？如果大家都可以随意恶意的假设和推测，我也可以把你臆想得什么都不是。

韩寒在松江二中读书过程中得了那种浑身奇痒的疥疮，后来学校怕感染其他同学让他回家养病一个星期。我是在那天回家看到他写的《求医》才知道他得疥疮的。那天回家时他去外面玩了，但文章在桌上。我看了以后觉得很幽默，笑坏了。因为我事先看见过新民晚报上一则上海萌芽举办新概念作文大赛的消息，所以就建议韩寒将这篇《求医》和另一篇《书店》参加新概念作文比赛。因为他觉得这次比赛没收参赛费，应该是真正的比赛，不是那种常见的近乎骗人的活动，而且既然是新概念，就不是应试作文的那种，就参加了。后来寄出去后一直没有回音，韩寒和我都有点失望，直到那天上午，一位叫胡玮蔚的萌芽编辑将电话打到家里问为什么不去参加前一天的新概念复赛才知道那天新概念作文比赛就要揭晓颁奖了。（很多评委在初赛的时候就留意到了韩寒的两篇文章，觉得特别老练，就和现在大家的怀疑一样，所以他们委托萌芽的胡玮蔚编辑给了我们电话，一方面是爱惜人才，怕因为客观原因错过了比赛，一方面也想当面考验韩寒）。韩寒说没有接到复赛通知（当时我们一家住在 50 多平米的老公房，楼下的邮箱都是没有锁的），后来胡玮蔚去问了评委后再打电话过来说评委同意韩寒中午前赶到上海市区比赛的地方再考一遍，我带了韩寒就急急忙忙的赶到车站那里找了辆黑车去市区，到那边已经接近中午了。接下来就是评委即兴出题现场一个小时写出《杯中窥人》的事。这事居然让阴谋论者认为我是开了后门事先知道了题目写好后让韩寒背的。这真是天地良心了，我们知道韩寒其实入围了是在当天的上午，此前萌芽的编辑我一个都不认识。包括李其纲，我也是根本不知道那个出题老师叫李其纲，出题的老师叫李其纲是我后来在有关新概念作文比赛和韩寒的补考的新闻报道中才知道的。而且我根本写不出这种文章。我的文章根本不是这个风格的。再说如果新概念作文比赛可以舞弊的话，那韩寒真的不可能有这次机会，因为这个比赛很隆重，有很多的教授和著名作家作评委，真的这个比赛要走关系的话，参赛的学生里有这方面能力的家长实在太多了，能得到好处的肯定不会是我们，我相信有这种能力的家庭也不会只住在 50 多平方米的老公房里。我可以这么说，一切能够靠钱靠关系靠舞弊能获得的好处，都不会轮到我们先得到，我们甚至连号都排不上。这种污蔑直接玷污了这个严肃的比赛。韩寒作为新概念作文比赛的参赛者，我只能告诉这些我所知道的情况。

好多人都说我应该回应，其实我觉得这事真的无聊透了。韩寒诚然有很多不足之处，看他不顺眼可以直接批评或者骂他，而且一直以来也不少这么做的人，但用这种全靠自己的臆想和主观判断来污蔑和传播，我觉得就十分下作和无耻了。我对韩寒说，无论你怎么说，他们还会无中生有找各种各样的茬来污蔑你的，因为他们就是看你不顺眼。有一次韩寒比赛翻车了（拉力赛很容易翻车，他参加拉力赛好几十场一共翻车过两次，算是冠军车手里翻车最少的），他回头发现在新闻的留言里最多的都是咒韩寒怎么还没死的，所以他早就知道了在很多人喜欢他的时候，也有很多人不喜欢他，他几乎从来不回应那些对他的辱骂，甚至有些谣言也不回应，但我觉得这次真的太过分了。包括这篇东西，我也只是说说自己，说说当时的一些情况。其实只要是韩寒的真正的读者，就会发现韩寒的文章和书从一开始到现在，其行文风格是有一条清晰一致的成长脉络的，像幽默什么的只是一开始比较刻意，一直在掉书袋，用典故，和现在的很多专家写的一样，那是因为受到了钱钟书的影响。很多典故和生僻的书本或者英语都是他硬记下来为了炫耀而背的，你要写过文章都知道，你随便记住或者摘抄下几个很生僻的东西，想要硬放到文章里是很容易的。韩寒的引用都不算特别自然，算是明显的故作老成。后来更趋于自然和内在。一直到了近几年，他说他写文章要做到不用典而把事情说清楚。我觉得这是他的进步。一个十五岁的孩子到三十岁肯定是在不停的进步的，这也是人生中一个人改变最大的阶段，阴谋论的人不能要求韩寒永远和十五六岁的时候写的文章一个样子，否则就是有假，那样倒是要被人笑死了。我找了一下，发现韩寒学生时代写在各种作业本笔记本和各种不规则纸上的那些文章手稿都在，当时保存这些也不是为了日后打官司或者让人研究，这些不经意间保存的资料，现在看来是多么珍贵。看到这些原始资料，我不得不再一次为我的儿子骄傲，很多人，包括我，十几岁时根本做不到他那样，韩寒虽然有点虚荣，有点故作老成，但是他做到了，他还在不停的努力和进步。我现在还是认为，韩寒的这种文风是骨子里的，不是随便可以模仿的。如果这年头连手稿都不能用来证明什么了，那大家都不要写作了，真是所谓欲加之罪，何患无辞了。韩寒现在决定要出《三重门》的手稿，你们到时还会惊羨一个十七岁的高中没毕业的少年写的字竟然有这么好这么老练，可以自己去对照一下或者发奋一下，能有几个大学生研究生博士生的字能有这么好看。韩寒不是一个特别喜欢应酬和交际的人，别人可能在玩的时间，他在想东西，写作，阅读，练字，练车。他最大的娱乐就是有时候踢一次球或者周末和朋友们打一个《使命在召唤》的真人射击类的电脑游戏，他的名气不算小，但是他平均一个月都没有一个饭局。大家之所以觉得他能做很多的事情，精力很旺盛，是因为他把别人可能用来应酬和娱乐的时间都用在工作上。你们可能不知道吧，如果没有比赛和游戏，他几乎每天晚上的八点开始写作或者看书，一直到早上六点，连续十个小时都在书房里。所以他的博客大多都是凌晨发的。虽然他口头上不承认，一直说他在玩，但这个就好像一个考试很好的学生喜欢说他在家里从来不复习一样。没有想到，他的努力反而成为了

他的过错。有些人甚至拿出了韩寒小学两年级的作文要说明韩寒未来写文章不好，这真的是不厚道的，其实大部分的小学在二年级的时候还在教认字，那个年代很多的学校是三年级才开始写作文的，我不知道现在的小学生是什么样的，我记得韩寒在二年级的时候好像是主动写作文给语文老师批阅的，也就是说这不是老师布置的作业，同年级里的大部分同学都还没有开始写作文。这种推测是有点胡搅蛮缠了。你不能因为刘翔学走路的时候会摔跤就推测他将来肯定跑不快所以有猫腻。况且我觉得韩寒那篇作文写的挺好的。可能他也觉得以 8 岁小孩子的水平来说还不错，就自豪的贴在了博客上，要不然大家也不会看到这些。韩寒写文章很快，修改也不多，这个可以从他的手稿里看出来。韩寒回应麦田的文章，其实写了两个小时，但是 4 点，6 点，8 点，甚至 10 点又修改过，这说明韩寒一夜睡不着，很在意这件事情，他修改的内容是让文章更简洁一些，语气也更缓和一些，包括一些不太礼貌的气话都删除了，可以说在一个作家遭到了污蔑又没有办法自证以后，他已经做的很平和。我知道他很珍惜自己的名誉，所以一定很生气。从半夜两点修改到十点也是韩寒自己写在博客里说的，以表示他很愤怒，气得都有点不会写文章了。他不说反而没有人拿这个来说事，他就是太老实太坦诚了，自己说了，结果吃了亏。被有的博士生拿来有意曲解为韩寒一般写两千字的文章都要花十个小时，所以有代笔，他的考试文章也不可能一小时写出来。有人也质疑说韩寒这几篇文章写的文采不如以前，说明以前是有人代笔的，这几篇文章可都是被迫的回应和申明啊，而且还是他人构陷在先，结果你不去怪那个加害人，反而要百般刁难受害人。这么说的人真的是太没有意思了。

再回到我。我是从看到韩寒写的《求医》等文章后就慢慢不再写东西的，因为我觉得我已经写不过他了，有点不好意思也懒得再写了，这种感受也许一些从事文化工作方面的父亲能够体会。我以前写的也只是一些农村题材的故事和一些应景宣传用的东西，根本没涉猎过中长篇小说。所以说韩寒的长篇小说是我写的很滑稽。和韩寒写的东西一比，反而是我写的实在太小儿科了。另外我也不怕浅薄，还要告诉大家，我根本不会英语，我们那个年代，从小学到四年中学到自学专科，从来就没有英语这门课。（大家好，我是韩寒，插入一下，我的父亲刚才经过回忆纠正了一下，说他们好像在四年中学里有过几节不正规的英语课，但是好像只学会了字母，他本人几乎不认识一个英语单词）我书读得比韩寒少多了，韩寒说的好多典故都是我不知道的，我甚至四大名著都没全看过（不过韩寒也没有看全过，因为他当时读书有一点炫耀的成分，要去读那些同学们都没有读过的书，才好像显得他很有学识），三国演义努力了几次都没看完第一回，我只看我喜欢的一些东西。他的书中文章中对好多事情的分析思考看法观点，让我受到过好多的启发。我是从心里佩服他的，当然从没嫉妒过他。我后来在韩寒的建议下申请了提前退休。现在的生活，就像我前几天在微博上说的，彻底告别了闹钟，一般睡到上午 10 点左右自然醒吧，先在被窝里和马桶上手机上会网，然后洗洗涮涮，弄点吃的，打扫下卫生，中午开始电

脑上网，挂上 QQ 和 MSN 方便有人找和联系事情，然后看看新闻，翻翻微博，有需要在电脑上处理的事处理一下，下午出去办办事，有时去老家看看父母，遛一下金毛和萨摩耶，对了，金毛叫憨憨，萨摩耶叫闹闹，我的微博头像就是闹闹。韩寒博客头像上的金毛是几年前已经去世的木木。有时去看看小孙女。所谓含饴弄孙吧。好多朋友希望我贴小孙女照片，这个，等稍过几天小孙女的图片正式发布后再贴张我和她的合影吧，绝对小美女一个。当然我也没你们想象的老。我自然年龄 56 岁，社会年龄就是你们看上来的年龄大概要小 10 岁，心理年龄也许还要小 10 岁。没韩寒英俊，但五官还算端正。如果有麻友来约，晚上就会去打半夜麻将。回家后再上网，看会新闻，翻会微博，有时看几集美剧，二三点钟后挂起电话洗洗睡去。就这么循环，很快，一年，一年，慢慢老去。

但韩寒还只是一个 30 岁的青年。我希望韩寒能生活在一个正常的人与人之间具有基本信任和交流的社会环境里。你可以不同意他的观点，去和他争论，去批评他，可以看不顺眼直接骂他，但不要去用恶意的揣测去诬陷他，污蔑他。用各种谣言和臆测来扼杀一个韩寒轻而易举，而且我知道很多人想这么做，今天终于有了机会，所有一直不爽韩寒的人终于可以团结起来。不过我觉得韩寒不是那么轻易可以用谣言扼杀的。或者等他被扼杀了，你们就知道想“人造”一个韩寒是一件不可能的事情了。对不起大家，我已经很久不写长得东西了，所以写的很啰嗦，谢谢你们可以耐心的读完。也许在阴谋论的人眼里，我是故意写的这么啰嗦来和韩寒的东西很简洁的风格区分开来的。那么好吧，我找到了两篇 1999 年的时候，就是韩寒写《三重门》的时候我发表在《故事会》和《现代农村》上的小故事，大家可以和韩寒当时的文章对比一下。我知道如果要牵强附会，那么你也能找出我这个文章和韩寒的某个文章一两个用词是一样的，一两个形容词是一样的，一两个转折词是一样的，甚至可以说我发表在《故事会》等报刊上的文章就是韩寒写的甚至这篇文章就是韩寒写的。我们两个人是儿子代老子写，老子代儿子写，我们两个人太闲了。还是要说，欲加之罪，何患无辞。

2012 年 1 月 27 日

Příloha 4: Tabulky uvádějící pozorování empiricky získaná kvantifikací novinového článku a povídky

Tabulky uvádí délky konstruktů x_i měřených v jejich konstituentech, jejich četnosti výskytu z_i a průměrné délky jejich konstituentů y_i měřených v jednotkách bezprostředně nižších. Výplň šedé barvy buněk značí odebraná pozorování s nízkou frekvencí.

Jazykovou úroveň značíme jako i , kde $i = 4$ představuje jazykovou úroveň znak – prvek, $i = 3$ jazykovou úroveň parcelát – znak, $i = 2$ jazykovou úroveň souvětí – parceláta $i = 1$ jazykovou úroveň odstavec – souvětí; i tedy může nabývat hodnot $i = 1, 2, 3, 4$.

Tabulka A Jazyková úroveň 4: znak (měřený v prvcích) – prvek (měřený v průměrném počtu tahů)

x_4	Novinový článek		Povídka	
	z_4	y_4	z_4	y_4
1	463	5,0994	733	4,7763
2	710	3,4127	782	3,4565
3	696	2,5393	731	2,5860
4	340	2,0824	425	2,2047
5	220	1,8718	214	1,9729
6	83	2,0161	101	1,9076
7	43	1,6246	49	1,8192
8	7	1,6964	22	1,8125
9			6	1,3704
10			2	1,8000

Tabulka B Jazyková úroveň 3: parcelát (měřený ve znacích) – znak (měřený v průměrném počtu prvků)

x_3	Novinový článek		Povídka	
	z_3	y_3	z_3	y_3
1			7	3,1429
2	4	3,1250	11	3,0000
3			21	2,5556
4	14	3,3393	32	2,8281
5	6	2,2667	26	2,9538
6	22	2,9545	30	2,8611
7	6	2,7619	41	2,8153
8	14	2,9911	36	2,7014
9	11	2,9697	27	2,7860
10	5	2,4400	26	2,7923
11	10	3,0182	24	2,7348
12	7	2,7738	13	2,4744
13	12	2,8141	13	2,5740
14	14	2,8776	11	2,6688
15	7	2,7429	13	2,9641
16	4	2,7656	4	2,8750
17	4	3,2206	6	2,5882
18	8	2,7153	3	2,8704
19	3	3,0000	3	2,6842
20	3	2,6167	2	2,2500
21	8	2,7262	5	2,6857
22	5	2,5818		
23	2	2,7391	2	2,8261
24	3	2,7222	1	2,8333
25	2	2,5600		
27			1	2,4074
28	2	2,6964		
29	3	2,6207		
30	1	3,0333		
31	1	2,7742		
33	1	3,1818		
34	1	3,2059		
35	2	2,8429		
36	1	2,9444		
38	1	3,0789		
39	1	3,0000		
42	1	3,0952		

47	1	3,1277		
----	---	--------	--	--

Tabulka C Jazyková úroveň 2: souvětí (měřené v parcelátech) – parcelát (měřený v průměrném počtu znaků)

x_2	Novinový článek		Povídka	
	z_2	y_2	z_2	y_2
1	8	27,3750	23	9,9565
2	14	17,2500	34	10,9265
3	7	14,4286	25	8,2933
4	11	11,8636	11	7,3636
5	3	15,6000	12	7,3667
6	1	16,5000	6	7,6944
7	3	11,9048	1	8,5714
8	3	10,3333	2	6,3125
9			2	9,7222
11	1	7,3636	1	8,3636
12	1	10,2500		

Tabulka D Jazyková úroveň 1: odstavec (měřený v souvětích) – souvětí (měřené v průměrném počtu parcelátů)

x_1	Novinový článek		Povídka - Varianta 1		Povídka - Varianta 2	
	z_1	y_1	z_1	y_1	z_1	y_1
1	2	3,0000	9	3,6667	3	2,6667
2	5	4,7000	10	2,3000	2	4,0000
3	4	2,5833	3	3,5556		
4	3	3,2500				
5	2	3,8000	1	3,0000	1	3,0000
6	1	4,8333			1	2,8333
7			2	2,1429	1	2,1429
8			2	3,6250	2	3,6250
9					1	2,6667
10					1	4,2000
11					1	1,7273
15						
17			1	3,9412	1	3,9412
27			1	2,8519	1	2,8519