

UNIVERZITA PALACKÉHO V OLOMOUCI

Filozofická fakulta
Katedra asijských studií

MAGISTERSKÁ DIPLOMOVÁ PRÁCE

Aplikace Menzerath-Altmanova zákona na psanou japonštinu

Application of the Menzerath-Altmann Law to Written Japanese

Olomouc 2016

Denis Birjukov

vedoucí diplomové práce: Mgr. Ondřej Kučera

Kopie zadání diplomové práce

Prohlášení o samostatnosti

Prohlašuji, že jsem diplomovou práci vypracoval samostatně a uvedl veškeré použité prameny a literaturu.

Olomouc, 30. listopadu 2016

Denis Birjukov

Anotace

Práce si klade za cíl ověřit platnost Menzerath-Altmanova zákona na třech japonsky psaných textech: povídce, básni a vysokoškolské seminární práci. Segmentace textů byla provedena na grafickém principu s jazykovými jednotkami vytvořenými výhradně pro účely segmentace japonského textu a zčásti vycházela z pilotního výzkumu z let 2014–2015. V úvodní kapitole je stručně představen obor kvantitativní lingvistiky, Menzerath-Altmanův zákon a charakteristika a specifika japonského písma. V kapitole druhé je popsána užitá metodika a způsob segmentace textů. Ve třetí kapitole jsou uvedeny hypotézy a v kapitole čtvrté samotné výsledky analýz a jejich interpretace. Poslední pátou kapitolu tvoří závěry pro jednotlivé analyzované texty.

Poděkování

Tímto vyjadřuji díky Mgr. Martině Benešové, Ph.D.; Mgr. Vladimíru Matlachovi; Prof. Dr. Haruko Sanadě; Mgr. Ivoně Barešové, Ph.D. a Mgr. Ondřeji Kučerovi za jejich podporu, cenné rady a trpělivost.

Obsah

1. Úvod	11
1.1. Kvantitativní lingvistika	12
1.2. Menzerath-Altmannův zákon	14
1.3. Stručná charakteristika japonského písma	18
1.3.1. Vývoj japonského písma	18
1.3.2. Směr zápisu textu	21
1.3.3. Kombinované užití různých typů písma v japonštině.....	22
1.3.4. Interpunkční znaménka	23
2. Metodika	24
2.1. Analyzovaný text č. 1 – <i>povídka</i>	26
2.2. Analyzovaný text č. 2 – <i>poezie</i> („západního“ stylu)	26
2.3. Analyzovaný text č. 3 – vysokoškolská seminární práce.....	26
2.4. Segmentace výběrových souborů.....	27
2.5. Definice zvolených jazykových jednotek.....	29
2.5.1. Nadodstavce	29
2.5.2. Odstavce	30
2.5.3. Souvětí.....	30
2.5.4. Mezičáří	31
2.5.5. Znaky.....	33
2.5.6. Prvky	37
2.5.7. Tahy	39
2.6. Ostatní poznámky k segmentaci textu.....	39
2.7. Užitý software	41
3. Hypotézy.....	42
3.1. U_1 – odstavce-souvětí	42
3.2. U_2 – souvětí-mezičáří	43
3.3. U_3 – mezičáří-znaky	44
3.4. U_4 – znaky-prvky.....	46
4. Výsledky analýz jednotlivých textů.....	47
4.1. Analyzovaný text č. 1 – <i>povídka</i>	47
4.1.1. Povídka – U_1 – odstavce-souvětí	47
4.1.2. Povídka – U_2 – souvětí-mezičáří	49
4.1.3. Povídka – U_3 – mezičáří-znaky.....	51

4.1.4.	Povídka – U ₄ – znaky-prvky	54
4.2.	Analyzovaný text č. 2 – poezie.....	56
4.2.1.	Poezie – U ₁ – odstavce-souvětí	57
4.2.2.	Poezie – U ₂ – souvětí-mezičáří	60
4.2.3.	Poezie – U ₃ – mezičáří-znaky.....	62
4.2.4.	Poezie – U ₄ – znaky-prvky	65
4.3.	Analyzovaný text č. 3 – seminární práce	67
4.3.1.	Seminární práce – U ₁ – odstavce-souvětí	67
4.3.2.	Seminární práce – U ₂ – souvětí-mezičáří	69
4.3.3.	Seminární práce – U ₃ – mezičáří-znaky	71
4.3.4.	Seminární práce – U ₄ – znaky-prvky.....	74
4.4.	Doplňující experimenty s analyzovaným textem č. 3	76
4.4.1.	Seminární práce – U _{SD} – souvětí-Dlouhá slova.....	77
4.4.2.	Seminární práce – U _{MD} – mezičáří-Dlouhá slova	80
4.4.3.	Seminární práce – U _{DK} – Dlouhá slova-Krátká slova.....	82
4.4.4.	Seminární práce – U _{KZ} – Krátká slova-znaky.....	84
5.	Závěry.....	86
5.1.	Závěr pro analyzovaný text č. 1 – <i>povídku</i>	86
5.2.	Závěr pro analyzovaný text č. 2 – <i>poezii</i>	87
5.3.	Závěr pro analyzovaný text č. 3 – <i>seminární práci</i>	90
5.4.	Závěr pro doplňující experiment se 3. analyzovaným textem – seminární práci.....	92
5.5.	Souhrnný závěr	94
	Resumé v anglickém jazyce.....	96
	Seznam pramenů	97
	Seznam užitého softwaru.....	97
	Seznam literatury.....	98
	Příloha	101
	Analyzovaný text č. 3 – seminární práce (stať)	101

Seznam grafů a tabulek

Graf 1: povídka – U_1 – odstavce-souvětí	47
Graf 2: povídka – U_2 – souvětí-mezičáří	49
Graf 3: povídka – U_3 – mezičáří-znaky	51
Graf 4: povídka – U_4 – znaky-prvky.....	54
Graf 5: poezie – U_1 – odstavce-souvětí	57
Graf 6: poezie – U_2 – souvětí-mezičáří	60
Graf 7: poezie – U_3 – mezičáří-znaky	62
Graf 8: poezie – U_4 – znaky-prvky.....	65
Graf 9: seminární práce – U_1 – odstavce-souvětí	67
Graf 10: seminární práce – U_2 – souvětí-mezičáří	69
Graf 11: seminární práce – U_3 – mezičáří-znaky.....	71
Graf 12: seminární práce – U_4 – znaky-prvky	74
Graf 13: seminární práce – U_{SD} – souvětí-Dlouhá slova (bez bodu x s hodnotou 134)	77
Graf 14: seminární práce – U_{MD} – mezičáří-Dlouhá slova.....	80
Graf 15: seminární práce – U_{DK} – Dlouhá slova-Krátká slova	82
Graf 16: seminární práce – U_{KZ} – Krátká slova-znaky	84
Tabulka 1: Japonská interpunkční znaménka.....	23
Tabulka 2: Jazykové jednotky	27
Tabulka 3: Jazykové úrovně.....	28
Tabulka 4: A, b, R^2 (povídka – U_1 – odstavec-souvětí)	47
Tabulka 5: Hodnoty x, z, y (povídka – U_1 – odstavec-souvětí).....	47
Tabulka 6: A, b, R^2 (povídka – U_2 – souvětí-mezičáří)	49
Tabulka 7: Hodnoty x, z, y (povídka – U_2 –souvětí-mezičáří).....	49
Tabulka 8: A, b, R^2 (povídka – U_3 – mezičáří-znaky).....	51
Tabulka 9: Hodnoty x, z, y (povídka – U_3 – mezičáří-znaky)	52
Tabulka 10: povídka – U_3 – podíl znaků dle počtu prvků.....	53
Tabulka 11: A, b, R^2 (povídka – U_4 – znaky-prvky)	55
Tabulka 12: Hodnoty x, z, y (povídka – U_4 – znaky-prvky)	55
Tabulka 13: A, b, R^2 (poezie – U_1 – odstavce-souvětí)	57
Tabulka 14: Hodnoty x, z, y (poezie – U_1 – odstavce-souvětí).....	57
Tabulka 15: A, b, R^2 (poezie – U_2 – souvětí-mezičáří)	60
Tabulka 16: Hodnoty x, z, y (poezie – U_2 – souvětí-mezičáří).....	60

Tabulka 17: A, b, R^2 (poezie – U3 – mezičárí-znaky)	62
Tabulka 18: Hodnoty x, z, y (poezie – U3 – mezičárí-znaky)	63
Tabulka 19: poezie – U3 – podíl znaků dle počtu prvků	64
Tabulka 20: A, b, R^2 (poezie – U4 – znaky-prvky)	65
Tabulka 21: Hodnoty x, z, y (poezie – U4 – znaky-prvky).....	66
Tabulka 22: A, b, R^2 (seminární práce – U1 – odstavce-souvětí).....	67
Tabulka 23: Hodnoty x, z, y (seminární práce – U1 – odstavce-souvětí)	67
Tabulka 24: A, b, R^2 (seminární práce – U2 – souvětí-mezičárí).....	69
Tabulka 25: Hodnoty x, z, y (seminární práce – U2 – souvětí-mezičárí)	69
Tabulka 26: A, b, R^2 (seminární práce – U3 – mezičárí-znaky)	71
Tabulka 27: Hodnoty x, z, y (seminární práce – U3 – mezičárí-znaky).....	72
Tabulka 28: seminární práce – U3 – podíl znaků dle počtu prvků	73
Tabulka 29: A, b, R^2 (seminární práce – U4 – znaky-prvky).....	75
Tabulka 30: Hodnoty x, z, y (seminární práce – U4 – znaky-prvky)	75
Tabulka 31: A, b, R^2 (seminární práce – USD – souvětí-Dlouhá slova se všemi body x).....	77
Tabulka 32: A, b, R^2 (seminární práce – USD – souvětí-Dlouhá slova bez bodu x s hodnotou 134)	77
Tabulka 33: Hodnoty x, z, y (seminární práce – USD – souvětí-Dlouhá slova).....	78
Tabulka 34: A, b, R^2 (seminární práce – UMD – mezičárí-Dlouhá slova).....	80
Tabulka 35: Hodnoty x, z, y (seminární práce – UMD – mezičárí-Dlouhá slova)	81
Tabulka 36: A, b, R^2 (seminární práce – UDK – Dlouhá slova-Krátká slova)	82
Tabulka 37: Hodnoty x, z, y (seminární práce – UDK – Dlouhá slova-Krátká slova).....	82
Tabulka 38: A, b, R^2 (seminární práce – UKZ – Krátká slova-znaky)	84
Tabulka 39: Hodnoty x, z, y (seminární práce – UKZ – Krátká slova-znaky).....	84

Ediční poznámka

V práci je použita česká transkripce japonštiny s výjimkou bibliografických citací dokumentů původně uvedených v latině. Za transkribovanými jmény a názvy bibliografických citací jsou v závorkách daná jména a názvy pro dohledatelnost uvedeny také v japonských znacích. U japonských místních jmen, u kterých je český přepis již pevně určen pravidly českého pravopisu, se v textu uvádí tento do češtiny přejatý tvar.

Cizojazyčná příjmení žen jsou uvedena v nepřechýleném tvaru.

Pro zápis čínských slov je užito transkripce putonghua.

Pro „Menzerath-Altmanův zákon“ je v této práci rovněž uváděna zkratka „MAZ“.

V příloze k této práci je uveden pouze jeden ze tří analyzovaných textů z důvodu ochrany autorských a jiných práv, a to text č. 3 – *vysokoškolská seminární práce*; analyzovaný text č. 1 – *povídka* a č. 2 – *poezie* jsou v současné době stále ve volném prodeji. Ačkoliv pro zveřejnění seminární práce nebylo třeba zvláštních povolení, byl přesto vyžádán souhlas autorky práce.

1. Úvod

Přestože se o Menzerath-Altmanovu zákonu v současné době poměrně mnoho hovoří, provádějí se hojně analýzy stále většího množství jazyků a uplatňuje se dokonce i v ne zcela lingvistických oborech, jako je např. biosémiotika aj., až do nedávné doby se analýzy pro ověření jeho platnosti do jisté míry vyhýbaly východoasijským jazykům, užívajícím původem čínské znaky. Několik experimentů s těmito jazyky již v relativně nekrátké historii existence Menzerath-Altmanova zákona provedeno bylo, viz např. inspirativní článek Claudie Prün (1994, s. 148–155) implementující originální způsob segmentace čínských znaků dle Jürgena Stalpha (1989, s. 73), avšak publikovaných prací, zabývajících se např. psanou japonštinou či psanou čínštinou, je dodnes ve srovnání s mnoha jinými jazyky naprosté minimum. V letech 2012–2013 pak Tereza Motalová, Lenka Spáčilová aj. (2013) z Univerzity Palackého v Olomouci provedly několik experimentů s různými texty psanými současnou čínštinou a bylo rovněž provedeno i několik jiných experimentů s tímto jazykem, na nichž se ověřovala platnost Menzerath-Altmanova zákona, což lze považovat za zahájení kvantitativně-lingvistických prací s touto tematikou, zaměřených na východoasijské jazyky, na Univerzitě Palackého. Tato práce pak čerpá mimo jiné právě z výsledků daných experimentů a do jisté míry na ně také navazuje – tentokrát však analýzami textů psaných v jazyce japonském.

Tato diplomová práce se zaměřuje na kvantitativně-lingvistickou segmentaci a analýzu textů psaných v japonštině pro ověření platnosti Menzerath-Altmanova zákona právě na psané japonštině, která prozatím nebývá středem pozornosti podobných experimentů. Pro zahájení prvotních experimentů byla zvolena segmentace textu na grafickém principu. Zvolené jazykové jednotky byly: (nadodstavec), odstavec, souvětí, mezičárí, znak, prvek a tah, přičemž pro účely experimentu každá z nich dostala vlastní definici, která ne vždy odpovídala té, jež se užívá v běžné praxi. Vznikly tak čtyři¹ jazykové úrovně: U_1 – odstavce-souvětí, U_2 – souvětí-mezičárí, U_3 – mezičárí-znaky, U_4 – znaky-prvky pro každý ze tří analyzovaných textů a získané výsledky byly podrobeny následné analýze a komparaci.

Práce je rozdělena do pěti kapitol. V úvodní kapitole je stručně představen obor kvantitativní lingvistiky, Menzerath-Altmanův zákon a charakteristika, vývoj a specifika

¹ Resp. pět se započítáním vynechané úrovně U_0 – nadodstavce-odstavce.

japonského písma, na nějž je práce zaměřena. Kapitola druhá popisuje užitou metodiku, způsob segmentace výběrových souborů a především definice zvolených jazykových jednotek. V kapitole třetí jsou představeny důvody pro formulované hypotézy. Čtvrtá kapitola je zaměřena na výsledky analýzy každého z analyzovaných textů na všech jednotlivých jazykových úrovních a popis jejich možných příčin, příp. formulaci nových hypotéz pro budoucí výzkumy. V závěrečné páté kapitole jsou pak předloženy závěry pro každý ze tří analyzovaných textů a jeden doplňující experiment.

1.1. Kvantitativní lingvistika

„Původ kvantitativní lingvistiky je situován do Evropy. Její vývoj byl nejrychlejší v Německu a východní Evropě, odkud pochází většina současných zástupců této disciplíny. Nejznámějším z nich je profesor matematické lingvistiky na Ruhr-University v Bochumi – Gabriel Altmann, ... který je považován za jednoho ze zakladatelů moderní kvantitativní lingvistiky (Motalová 2014, s. 11).“

Už po celé 20. století se užívalo takových metod v lingvistickém bádání, které dnes lze označit jako metody kvantitativní nebo statistické. Samotný „termín ‚kvantitativní lingvistika‘ je dnes už běžně užíván, ale setkáváme se také s termínem ‚lingvistika statistická‘, kterému někteří jazykovědci dávají přednost s ohledem na převažující metodu. ... I když určité náznaky užití kvantitativních metod v jazykovědě najdeme například už ve starověké Indii, dá se mluvit o vážnějších pokusech zavést kvantitativní (statistické) hledisko do jazykovědného výzkumu teprve od přelomu 19. a 20. stol., kdy matematika začala systematicky pronikat do jiných vědních oborů, mezi nimi i do jazykovědy. ... Z lingvistů se jako první zabýval kvantitativními jevy v jazyce americký jazykovědec W. D. Whitney, který v r. 1874 zkoumal frekvenci anglických hlásek ... V roce 1913 vydal ruský matematik A. A. Markov ‚Příklad statistického výzkumu textu Evžena Oněgina...‘ ... Toto dílo sehrálo významnou roli v dalším rozvoji matematiky i jazykovědy a jako první dílo vůbec vzbudilo zájem matematiků o jazykovědné otázky; je samozřejmé, že prohloubilo i zájem jazykovědců o otázky matematické“ (Černý 1996, s. 248–250).

„V českém lingvistickém prostředí má kvantitativní lingvistika (dále KL) poměrně dlouhou a bohatou tradici ... V rámci této tradice je kvantifikace jazykových jevů v naprosté většině případů používána jako doplňující charakteristika při popisu

a klasifikaci jednotek jazykového systému ... Od 70. let 20. století se však pod stejným označením začíná formovat přístup, který kvantifikaci (a matematickou formalizaci) chápe jako nástroj *pro testování hypotéz* týkajících se vlastností jazykového systému ... Jinými slovy, KL tohoto typu již neklade za cíl klasifikaci jazykových jevů ..., ale jejím cílem je modelovat vlastnosti jazykového systému a procesy s ním související tak, aby tyto vlastnosti bylo možné vyjádřit formou empiricky testovatelných hypotéz. Důležité při tom je, že hypotézy jsou zpravidla odvozovány buď z obecných principů, např. Zipfova principu nejmenšího úsilí ..., nebo (v lepším případě) z obecné jazykové teorie ... KL v tomto smyslu je tedy experimentální vědou stejně jako třeba chemie, fyzika či experimentální psychologie. A je to právě důsledné „lpění“ na experimentálním přístupu (za použití běžně používaných statistických metod), jímž se KL tohoto druhu ... liší od většiny ostatních lingvistických směrů“ (Čech 2014, 7).

Za zakladatele moderní kvantitativní lingvistiky lze považovat především George Kingsleyho Zipfa. Byl první, kdo systematicky studoval text a jazyk s využitím statistiky čistě pro vědecké účely. Od té doby se kvantitativní lingvistika samozřejmě významně rozvíjela a stala se plnohodnotnou větví obecné lingvistiky. Rozvinula se dokonce do takové míry, že je obtížné zachovat přehled o mnoha tématech a předmětech bádání, modelech a metodách aplikovaných i vyvíjených a o mnoha výsledcích publikovaných v knihách a časopisech. Od doby Zipfa lze uvažovat pouze o jediném vědci, který inspiroval a posunul kvantitativní lingvistiku v podobné míře, ne-li více, a kdo v tom i nadále pokračuje – o Gabrielu Altmannovi (Grzybek 2007, viii).

Gabriel Altmann – průkopník a „celosvětový hráč“ na poli kvantitativní lingvistiky – se narodil 24. května 1931 ve slovenském městě Poltár. Mezi lety 1953 a 1958 studoval indonéskou lingvistiku a japonskou filologii na Karlově univerzitě v Praze (Grzybek 2007, ix). Po svých doktorských studiích pracoval postupně v Bratislavě, v Kolíně nad Rýnem, v Mannheimu a nejdéle pak na Univerzitě Ruhr v Bochumi až do roku 1998 (*www 1*). Gabriel Altmann navrhoval objevování a formulování zákonů jako „výzkumný program pro kvantitativní lingvistiku již od 60. let 20. století jako způsob, jak dojít k opravdovému vědeckému poznání o jazycích“ (Hřebíček 2007, s. 237–243).

1.2. Menzerath-Altmanův zákon

Počátky formulace Menzerath-Altmanova zákona lze sledovat do první poloviny 20. století: „První krok k objevu ‚Menzerathova zákona‘ byl učiněn roku 1928, kdy Menzerath zjistil, že ‚...čím delší je celek, ve kterém se zvuk objevuje, tím je tento zvuk kratší (zákon kvantity)‘ a že ‚...čím víc je zvuků ve slabice, tím kratší je jejich relativní délka‘... Jinak řečeno, délka artikulace zvuku se v dlouhých slabikách zkracuje. Později při analýze německých slov dospěl Menzerath k následujícímu závěru: ‚Relativní počet zvuků ve slabice klesá se vzrůstajícím počtem slabik ve slově, nebo řečeno jinak: Čím více je ve slově slabik, tím jsou (relativně) kratší‘... Jinými slovy, čím delší je věta, tím kratší obsahuje slabiky“ (www 2, s. 124).

„Byl to prvý podnět k úvahám o vzájemném vztahu jazykových úrovní. Až v 80. letech 20. století navázal Gabriel Altmann na práci P. Menzeratha a zavedl dva termíny: konstrukt a konstituent. Konstrukt je jazyková jednotka na vyšší jazykové úrovni a jeho konstituent je jazyková jednotka na nejbližší úrovni nižší. G. Altmann testoval vztahy mezi těmito dvěma jazykovými jednotkami a dokázal, že mezi nimi existuje korelace“ (Motalová 2014, s. 16).

Tuto korelaci formuloval takto: „Čím delší je v jazyce nějaký konstrukt, tím kratší jsou v průměru jeho konstituenty.“ Je třeba zdůraznit, že tento vztah nelze brát doslova a do písmene a je nutno jej pojímat pouze jako vyjádření tendence; „mluví se tu ... o průměrných hodnotách. ... Altmann ... Menzerathovo pozorování formuloval jako *Menzerathův zákon* ... Tyto okolnosti nás opravňují k tomu, abychom tento zákon nazývali *Menzerathovým-Altmanovým zákonem* ...

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. Délku konstruktů měříme v konstituentech, jejichž počet bude tedy celé číslo. Délka konstituentů ovšem bude různá, protože ji vyjádříme průměrem jednotlivých zjištěných délek ... 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. 53–54).

Tyto dva „předpoklady tvoří jeden celek: obě proporcionality, nepřímá i přímá, jsou řízeny jedním koeficientem. V případě první hypotézy bude záporný, v případě druhé hypotézy kladný. Vyjadřuje míru, při níž platí oba vztahy úměrnosti“ (Hřebíček 2002, s. 54).

Když dnešní lingvisté mluví o „Menzerathovu zákonu“, obvykle Altmannovým vzorcem myslí $y = ax^b e^{cx}$ či matematicky ekvivalentní diferenciální rovnici. Koeficienty a , b , c by měly být závislé na daném korpusu; často se má za to, že případ $c = 0$ vede k základní „nenarušené“ podobě zákona. Pokud se vezme reálný přístup k práci s tímto zákonem jako výchozí bod, zákon by měl být chápán zhruba následovně:

V lingvistickém korpusu vhodného typu (např. lingvisticky homogenním textu) může být křivka obecné podoby $y = Ax^b e^{cx}$ obvykle dostatečně dobře napasována na vztah mezi délkou konstruktů na určité úrovni (měřené v počtu konstituentů konstruktů na hierarchicky hlubší úrovni segmentace) a délkou těchto konstituentů (měřených v počtu konstituentů na nějaké ještě hlubší úrovni)“ (Meyer 2007, s. 449).

Ve většině případů však pro popis výše uvedeného zákona postačuje speciální případ tohoto zákona: $y = Ax^b$. Tento zákon, který vertikálně propojuje různé vrstvy jazyka, byl „prokázán na mnoha jazycích a na různých jazykových úrovních“²: od fonémů po texty, a dokonce byla nalezena určitá shoda mezi délkou konstruktů a počtem jeho významů. Bylo předpokládáno, že tento zákon byl reflexí obecné lidské schopnosti zpracovávání informace (Roukk 2007, s. 605).

„Menzerath-Altmanův zákon operuje s obecnými koncepty jazykových konstruktů a jejich konstituentů. Vzájemný vztah těchto konceptů je definován jako určitý inverzní vztah jejich příslušných proměnných“ (Hřebíček 2007, s. 237).

Hřebíček uvádí zkrácený matematický vzorec odvozený Altmannem pro vyjádření MAZ takto:

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

² Lépe řečeno byl potvrzen na vybraných textech v určitých jazycích. Roukk např. také poukazuje, že se MAZ neprojevil na některých analyzovaných ruskojazyčných textech.

„kde x je velikost konstruktů a y je průměrná velikost jeho konstituentů; A a b jsou koeficienty. S tím, jak vzrůstá x , se průměrná velikost y snižuje a formuje tudíž typickou klesající křivku“ (2007, 237–8).³

„Čím větší je absolutní hodnota záporného parametru b , tím prudčeji klesá křivka znázorňující funkci y v grafu. Parametr A vyjadřuje průměrnou délku konstituentu za předpokladu, že za x dosadíme 1, tj. že daný konstrukt obsahuje jen jeden konstituent. ... Můžeme také říci, že funkce Ax^b vyhovuje Menzerathovu-Altmanovu zákonu, když b nabývá záporné hodnoty“ (Hřebíček 2002, s. 55–58).

Velikost parametru A „řídí vzdálenost neboli posunutí celé příslušné křivky nahoru nebo dolů vzhledem k ose x , čili je posunuta víc nebo méně od této osy“ (Hřebíček 2002, s. 58). Může být chápán jako „počáteční hodnota“ proložené křivky na ose y (Kelih 2010, s. 71). Strmost křivky se řídí absolutní velikostí parametru b a ten „... představuje onu zmíněnou společnou míru proporcionality, která sjednocuje dvě hypotézy o proporcionalitě uvažovaných veličin“ (Hřebíček 2002, s. 55).

Hřebíček dále uvádí, že „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. V textu tedy je realizován řetězec:

... ↔ KONSTRUKT / KONSTITUENT ↔ KONSTRUKT / KONSTITUENT ↔ KONSTRUKT / KONSTITUENT ↔ ...

Jinak řečeno, každá jazyková entita vůči všem vyšším jazykovým úrovním je konstituentem a vůči všem nižším úrovním je konstruktem“, s výjimkou koncových členů této posloupnosti.

Je to „systém, který jako by se pohyboval od jednoho stavu ke druhému, a ten zase jako by byl dalším výchozím stavem, který přejde na další (vyšší nebo nižší, podle směru

³ „Tento vztah byl potvrzen pro jazykové jednotky na všech úrovních od fonémů po věty. Následně byl zákon aplikován rovněž na nejvyšší testovací úroveň, kde jsou konstrukty reprezentovány určitými sémantickými komplexy založenými na lexikálních jednotkách a jejich konstituenty jsou ve formě textových segmentů. ... Bylo potvrzeno, že čím vyšší je frekvence slova (nahlížená jako výskyty slovních jednotek v různých segmentech), tím nižší je průměrná velikost segmentu. Z citlivosti na jednotku slova a z formování charakteristické distribuce velikostí segmentů a jejich funkčního výskytu – oboje definované Menzerath-Altmanovým zákonem – lze vyvodit následující předpoklad: Různé slovní jednotky vyskytující se v daném textu a jejich segmenty se vzájemně specifikují. Spojování slovních jednotek dělá jejich význam specifičtější“ (Hřebíček 2007, 237–238).

pohledu) stupeň jazykové úrovně. Je to jakýsi kruh nebo spirála, která se postupně rozvírá anebo stahuje ke svému středu. V kruhu nebo spirále se nedostáváme zpět do téhož bodu, ocitáme se jinde (na jiné úrovni), ale všechny takové body mají něco společného, co je vyjádřeno zkoumaným zákonem. Představa pohybu je ovšem nepřesná, v jazykové struktuře se ve skutečnosti nepohybujeme od úrovně k úrovni. Nepohybujeme se vůbec ...“ (2002, s. 59–62).

Hřebíček se také domnívá, že jazyk může být fraktální množinou: „... Menzerathův-Altmanův zákon může vytvářet struktury, které v mnoha ohledech jsou vlastní textu. Můžeme tak konstruovat stupňovitou strukturu úrovní, která vyrůstá z předpokladu, že čím větší je jazykový konstrukt, tím menší jsou v průměru jeho konstituenty, a také že přírůstek konstituentu je úměrný jeho velikosti. ... Protože určitá entita je konstruktem vůči entitám na nižších úrovních a konstituentem vůči entitám na vyšších úrovních, lze v rámci textu dospět ke zřetězení konstruktů a konstituentů a jazykové úrovně pojmut jako soustavu založenou na iteracích. Text v této podobě představuje množinu, která je dynamickým systémem s fraktální strukturou“ (2002, s. 65).

Zároveň je „... přesvědčen o tom, že nerozhoduje, v jakých měřících jednotkách je vyjádřena délka věty, protože Menzerathův-Altmanův zákon je pro strukturu textu zákonem univerzálním“ (2002, s. 171).

Důležitá je i poznámka týkající se ověřitelnosti platnosti celého „zákona“, kterou uvádí Altmann. Poukazuje na to, že derivace diferenciální rovnice nepostačuje k tomu, aby se tvrzení, že délka konstituentů je funkcí délky jazykových konstruktů, přidělil status zákona: „Zůstane teoretickou a ne zcela ověřenou hypotézou do té doby, dokud bude záviset na vztahu k jiným jazykovým zákonům, tj. dokud nebude zahrnut do systému zákonů. Takový systém v současnosti neexistuje, my máme pouze podezření, že je to nějakým způsobem propojené s principem nejmenšího úsilí nebo nějakým jiným dosud neznámým rovnovážným principem, který kompenzuje prodlužování na jedné straně zkracováním na straně druhé. Empirické ověření se skládá z testování hypotézy na datech z mnoha jazyků. V těch mohou být koncepty ‚konstruktů‘ a ‚komponentů‘⁴ interpretovány mnoha způsoby, pokud se zamění za více empirické koncepty jako je zvuk, slabika, morfém, slovo, sousloví, složenina, fráze, klauze, věta atd. Je třeba si položit

⁴ konstituentu

otázku, zda koeficienty křivek závisí na daném jazyce, nebo na daných jednotkách, nebo zda rovněž i text ovlivňuje tvar křivek. ... Konečné potvrzení této hypotézy je nemožné z důvodu jeho univerzality, ale jakýkoliv příznivý výsledek může vylepšit jeho stupeň potvrzení“ (Altmann 1980, s. 125, 127). Z tohoto důvodu by tedy možná bylo vhodnější daný „zákon“ nazývat *teorií*.

1.3. Stručná charakteristika japonského písma

Níže následuje stručný popis vývoje, vlastností a specifik písem v současnosti užívaných pro zápis japonského jazyka.

1.3.1. Vývoj japonského písma

Dnešní japonština užívá pro zápis jazyka převážně původem čínské znaky *kandži* (漢字; nyní v jejich japonské podobě) – grafémy přejaté v 1. polovině 1. tisíciletí z Číny – a dvě původem japonské moraické či sylabografické „abecedy“, souhrnně nazývané *kana* (仮名).

Čínské písmo či čínské znaky⁵ lze charakterizovat jako nástroj sloužící k zápisu jazyka čínského, japonského a jiných jazyků (Jindřichová 2014, s. 10).

„V převážné většině případů zapisujeme jedním čínským znakem (grafémem) morfém, tedy nejmenší významovou jednotku (sémém), která je vázána na určitý zvuk“ (Zádrapa 2009, s. 37). Pokud uvažujeme o čínském písmu tak, že znak (základní jednotka čínského písma) označuje jeden morfém, mluvíme o písmu morfemografickém. Sice je někdy čínské písmo označováno jako logografické, protože pojem slovo nebyl v čínštině jednoznačně definován, avšak podíváme-li se na současnou čínštinu a japonštinu, vidíme, že se skládají z velké části z víceslabičných výrazů. Můžeme tak dojít k závěru, že čínština i japonština jsou přinejmenším systémy logo-morfemografické“ (Jindřichová 2014, s. 10).

Obvykle se uvádí, že čínský znak se skládá z tahů, které tvoří vyšší strukturní složky znaků (prvky, radikály, fonetika, determinativy apod.), ze kterých se teprve utváří samotný znak.

⁵ Ve smyslu čínské svým původem.

K utvoření ustálené podoby čínského písma nedošlo jednorázově, ale měnila se po celou dobu existence písma. Čínské písmo se měnilo z původně logografického typu na převážně morfemografické, později se utvářely i znaky ideofonografické obsahující více složek, aj. (Jindřichová 2014, s. 11).

„V Japonsku před přejímáním kulturních prvků z Číny prostřednictvím Koreje neexistoval vlastní způsob zápisu japonštiny. Proto se z Číny začalo v počátečních staletích našeho letopočtu přejímat spolu s dalšími jevy čínské znakové písmo, které se stalo prostředkem k zápisu japonštiny. Avšak čínské znakové písmo se vyvíjelo pro potřeby čínštiny ... a nebylo příliš vhodné pro zápis japonštiny ... Proto také nebylo možné čínské znakové písmo přímo ‚napasovat‘ na systém japonštiny. Kromě toho se zvukový repertoár japonštiny od čínštiny značně liší. Proto se z počátku spolu s písmem přejal i čínský jazyk a čínština se tím stala oficiálně úředním psaným jazykem. Avšak právě kvůli artikulačním odlišnostem japonštiny bylo k čínským výrazům přiřazováno „pojaponštělé“ čtení, které příliš neodpovídalo čínské ani japonské výslovnosti. Japonština také nerozlišovala tóny a místo hlásek, které v japonštině neexistovaly, byly přiřazeny přibližně stejně znějící hlásky japonské. Tímto způsobem získaly v Japonsku čínské výrazy sinojaponské čtení. Avšak protože čistě sinojaponské čtení čínského textu bylo poměrně nesrozumitelné, postupem času se pro sinojaponské výrazy, které měly svůj ekvivalent v japonštině, začalo používat japonské čtení; pro výrazy, které japonský ekvivalent neměly, zůstalo čtení sinojaponské“ (Jindřichová 2014, s. 15).

Na území Japonska se později vyvinulo i písmo vlastní, souhrnně nazývané *kana*. *Kana* se skládá ze slabičných⁶ abeced *hiragana* (ひらがな) a *katakana* (カタカナ). Ty vznikly asi v 9. století z rozsáhlého souboru znaků *kandži*, které se používaly výhradně foneticky pro zapisování japonských slov v případech, kdy nebylo vhodné užívat přejaté čínské znaky (Kraemerová 2000, s. 28).

Až do konce 19. století se pak japonské písmo živelně vyvíjelo bez jakýchkoliv administrativních zásahů (Kraemerová 2000, s. 29). Kromě více méně oficiálních variant znaků se, obdobně jako v Číně, i v Japonsku používaly zjednodušené varianty znaků, tzv. zkratky. „Užívaly se souběžně s oficiálními tvary znaků nejen pro soukromé účely, ale i v tisku“ (Jindřichová 2014, s. 16). Teprve až v roce 1900 proběhla první reforma

⁶ Přesněji moraických.

japonského písma, vyhlášená japonským ministerstvem školství, která mimo jiné vyřadila zkratkovité znaky z užívání a zavedla znakové minimum pro výuku na základních školách (Kraemerová 2000, s. 29). „Avšak zkratky se přesto používaly i nadále, například v soukromé korespondenci“ (Jindřichová 2014, s. 16).

Další reformní kampaň započala po druhé světové válce. Cílem této kampaně bylo omezit počet znaků užívaných v praxi a vyučovaných ve školách. Na základě mnoha kritérií, jako například četnosti grafémů, japonského i sinojaponského čtení atd., bylo vybráno 1850 znaků, které byly v roce 1946 vydány pod názvem *Tójó kandži hjó* 当用漢字表 – Znaky náležitého použití (Kraemerová 2000, s. 29). „Značným přínosem této reformy bylo ale také přijetí zkrácených tvarů znaků, které se od roku 1900 nesměly oficiálně užívat. Po reformě byly tyto znaky nejen znovu přijaty, ale nahradily jejich složitější formy. Na rozdíl od reformy písma v ČLR však neproběhlo cílené zjednodušování písma, nebyly vytvářeny nové zkratky, pouze byly přijaty již existující zjednodušené varianty, z čehož plyne, že byla reforma písma v Japonsku mnohem snadněji přijata a společnost se s ní mohla lépe vypořádat“ (Jindřichová 2014, s. 16).

Po mnoha dalších úpravách vyšel v roce 1981 nový seznam 1945 znaků, tzv. *Džójó kandži hjó* 常用漢字表 – Znaky běžného použití (Kraemerová 2000, s. 30).⁷

K další a zatím poslední úpravě *džójó kandži hjó* došlo v roce 2010.⁸ Nyní se jedná o seznam 2136 znaků navržených Ministerstvem školství, kultury, sportu, vědy a techniky opět pro užití ve školním vzdělávání, běžných tiskovinách a dalších textech (nezahrnuje však znaky užívané většinou pouze převážně v osobních jménech, místních názvech apod., tzv. *Džinmeijó kandži* 人名用漢字). Neznamená to, že jiné znaky se nesmí či nemohou v textu objevit, ale není to příliš běžné v textech ne odborného či netechnického charakteru a tyto tedy nejsou početné.⁹

Hiragana se dnes užívá např. pro gramatické a jiné afixy, partikule a pomocná slova, některá původem převážně japonská slova, fonetický přepis znaků *kandži* či jednodušší zápis neobvyklých či relativně složitých znaků atd., zatímco *katakana* se běžně užívá např. pro zápis přejatých slov (obzvláště nečínského původu) a cizích jmen a názvů, pro

⁷ Tento seznam byl v platnosti v době prvního vydání 1. analyzovaného textu – povídky.

⁸ Tento seznam pak byl v platnosti v době vydání 2. a 3. analyzovaného textu. V obou případech však autoři obou textů velice pravděpodobně nebrali na tento seznam ohled.

⁹ Viz ([www 3](http://www.3)) pro současný seznam *džójó kandži*.

zvýraznění textu či rovněž jako *hiragana* pro jednodušší zápis neobvyklých či relativně složitých znaků apod. Každá z těchto abeced obsahuje 48 základních znaků (včetně dvou zastaralých, které se však stále ojediněle užívají především v názvech), z nichž 25, resp. 26 v moderní *katakaně*, dále mění svoji podobu pomocí značek *dakuten*¹⁰ a *handakuten*, které připomínají obrácené uvozovky ` a malé kolečko °, psané obvykle v pravém horním kvadrantu znaku. Vytvářejí např. znělostní řady jako je *ha* は → *ba* ば → *pa* ぱ apod. 21 znaků z obou abeced (příp. i více v moderní *katakaně* a ojediněle v *hiraganě*) může být dále kombinováno do párů, tj. digrafů, se znaky *ja* (や nebo ヤ), *ju* (ゆ nebo ュ), *jo* (よ nebo ヨ) a řídčeji i jinými a vytvářet tak kombinované móry *jóon*, kde se druhý znak v pořadí zapisuje obvykle ve znatelně menší podobě než obvykle (např. *ki* き + *ja* や = *kja* きゃ či *de* デ + *i* イ = *di* デイ apod.).

1.3.2. Směr zápisu textu

Japonský jazyk umožňuje dva možné způsoby (směry) psaní textů. Prvním je zápis zleva-doprava odshora-dolů (tj. horizontální zápis, kdy se listuje stránkami stejným způsobem jako u textů českojazyčných, tzv. *jokogaki* 横書き), druhým je pak zápis odshora-dolů zprava-doleva (tj. vertikální zápis, kdy listování začíná „od konce“ v porovnání s českojazyčnými texty, tzv. *tategaki* 縦書き).

Japonské písmo bylo po dlouhou dobu formováno pro vertikální zápis, tudíž jeho rozlišující složky leží na vertikálních osách: Nalevo a napravo od osy, ale ve směru odshora-dolů. Lze tedy říci, že pokud je japonské písmo zapisováno horizontálně, ztrácí přirozenou kontinuitu rozlišujících složek (Kraemerová, 2000, s. 34). Dnes se však užívá obou směrů zápisu textu v porovnatelné míře, kdy jeden ze směrů obvykle převažuje v určitém stylu či typu publikace.¹¹

¹⁰ Rovněž zvaných *nigoriten*.

¹¹ Jelikož se také tištěný text v závislosti na směru jeho zápisu skoro nemění (co se týče např. podoby znaků), s výjimkou některých interpunkčních znamének, čísel aj., lze předpokládat, že na projev MAZ nemá směr zápisu znatelný vliv.

1.3.3. Kombinované užití různých typů písma v japonštině

Další zajímavou vlastností japonského systému zápisu písma je kombinované užívání původem čínských znaků *kandži* (nyní v jejich japonské podobě) zároveň se dvěma původně japonskými moraickými abecedami, souhrnně nazývanými *kana*.

Kombinování těchto odlišných typů písma probíhá nejen v rámci např. jediné věty, ale i v rámci jediného slova.¹²

Japonština tedy kombinuje převážně morfemtický a sylabografický systém (Jindřichová 2014, s. 16).

V japonštině se však navíc běžně užívají i arabské číslice a místy i latinka, převážně ve zkratkách. V tomto ohledu se japonské písmo výrazně odlišuje od písma čínského (tedy od standardní čínštiny *putonghua* (普通话) se zjednodušenými znaky či od tradičních znaků užívaných např. v Hongkongu), které obvykle užívá pouze čínský „ekvivalent“ *kandži*, tzv. *hanzi*, arabské číslice a výjimečně latinku. Navíc všechna japonská slova obsahující znaky *kandži* mohou být rovněž zapsána s užitím výhradně *hiragany* nebo *katakany*, a přestože pro kombinované užití všech těchto typů zápisu platí určitá pravidla či zvyklosti, v četných případech záleží hlavně na autorovi textu, který ze způsobů zápisu užije; mimo jiné lze této částečné volnosti využít pro lepší rozlišení relativně velkého množství homonym a pro usnadnění plynulého porozumění psanému textu.¹³

¹² Příklad nejmenším ve smyslu slova lexikografického.

Kromě zcela běžného kombinování *kandži* a jednoho typu *kany* (obvykle *hiragany*) se lze řidčeji setkat i s kombinací *hiragany* a *katakany*, např. v případech zkrácení původně cizího vypůjčeného slova zapisovaného *katakanou* a následným vytvořením slovesa připojením slovesného sufixu zapisovaného *hiraganou* (サボる, ググる apod.), či u jiného typu slov, nebo spíše ojediněle i s kombinací latinky a japonského písma, obvykle v případech, kdy je latinka ve funkci zkratky (X 線, B ライン apod.), kombinací řecké abecedy a japonského písma i u sémanticky nematematických výrazů (プラス α, β-カロテノ) aj.

¹³ Např. slovo *jasašii* 優しい, znamenající (být) *hodný, milý, něžný* (psáno kombinací *kandži* a *hiragany*) může být rovněž zapsáno bez změny jeho významu s užitím výhradně *hiragany* jako やさしい. (V některých případech odlišný zápis může v malé míře ovlivnit dojem daného slova na čtenáře, význam se však, dá se říci, nezmění.) Tato podoba nám však říká pouze o zvukových kvalitách (výslovnosti) daného slova. Toto tudíž může v případě vágního kontextu vést k chybné homofonní záměně za slovo *jasašii* (易しい), znamenající (být) *prostý, jednoduchý*. V naprosté většině případů by chybná interpretace tohoto druhu nemohla nastat, pokud by se užily znaky *kandži*. V případě substantiv nevyužívajících *hiraganové* sufixy je pak „riziko“ podobné nechtěné homofonní záměny (obzvláště při absenci nebo neurčitosti kontextu) vyšší.

1.3.4. Interpunkční znaménka

Japonština využívá v porovnání s češtinou mnoho odlišných interpunkčních znamének. Jelikož interpunkční znaménka tvoří nedílnou součást psaného textu a v daných experimentech byla navíc zahrnuta do definice jazykové jednotky znaku (viz 2.5.5. Znaky), níže jsou uvedena nejběžněji užívaná interpunkční znaménka relevantní pro tři texty analyzované v rámci této práce, jak je uvádí Barešová (2014, s. 83–84).¹⁴

Symbol	Japonský název	Použití
。	<i>kuten</i> 句点, <i>maru</i> 丸	odpovídá české tečce
？	<i>gimonfu</i> 疑問符	odpovídá českému otazníku
！	<i>kantanfu</i> 感嘆符	odpovídá českému vykřičníku
、 ,	<i>tóten</i> 読点, <i>ten</i> 点	odpovídá české čárce
・	<i>nakaguro</i> 中黒, <i>čúten</i> 中点	používá se k oddělení cizích slov přepsaných do <i>katakany</i> nebo při výčtu
「」	<i>kagikakko</i> 鉤括弧	odpovídá našim uvozovkám
『』	<i>nidžúkagikakko</i> 二重鉤括弧	většinou se používá k označení titulů knih, názvů článků a dále jako uvozovky v rámci uvozovek
()	<i>šókakko</i> 小括弧, <i>kakko</i> 括弧, <i>páren</i> パーレン	odpovídá našim jednoduchým závorkám
… ……	<i>santen rídá</i> 三点リーダー, <i>tensen</i> 点線	apoziopeze, výpustka
々	<i>odoridži</i> 踊り字, <i>kurikaeši</i> 繰り返し, <i>noma</i> ノマ	opakování předešlého znaku
っ	<i>sokuon</i> 促音	označuje zdvojený konsonant
ー	<i>čóonpu</i> 長音符, <i>bósen</i> 棒線	označuje délku u slov zapisovaných <i>katakanou</i> ; při psaní do sloupců se píše svisle
ゝ	<i>dakuten</i> 濁点, <i>nigori</i> 濁り	označuje znělost
。 (malé)	<i>handakuten</i> 半濁点, <i>handaku</i> 半濁	označuje řadu PA

Tabulka 1: Japonská interpunkční znaménka

¹⁴ S přidáním alternativní čárky *tóten* (、) pro účely této práce.

2. Metodika

Pro ověření platnosti Menzerath-Altmanova zákona na japonských psaných textech byly pro tento experiment vybrány tři japonsky psané texty odlišných funkčních a literárních stylů.

Je třeba poznamenat, že první z těchto textů – *povídka* – byl již dříve zanalyzován v rámci pilotního projektu IGA_FF_2014_083 a stručné výsledky byly publikovány v (Benešová 2015). Důvody pro opětovné zahrnutí již dříve analyzovaného textu do této práce společně s texty novými jsou následující:

1. Pro účely nového experimentu bylo využito nově vyvinutého softwaru MA Studio. Aby byla zajištěna porovnatelnost a sjednocenost výsledků nových analýz a analýzy předešlé, první analyzovaný text – *povídka* byl zanalyzován a popsán znovu pomocí nového SW.
2. Možnost obsáhlejšího a přesnějšího popisu hypotéz a získaných výsledků a navázání na předešlá zjištění.
3. Jelikož práce, zabývající se MAZ ve spojitosti s japonskými texty prakticky neexistují a analýzy s využitím zde uvedené metodiky byly prozatím provedeny pouze na daných třech textech, jeví se jako vhodné shrnout výsledky v rámci jedné práce.¹⁵

Co se týče volby samotných japonských textů, předně je třeba stanovit kritéria pro to, jaké má mít text vlastnosti, aby mohl být vybrán pro následnou analýzu a testování platnosti MAZ. Některá kritéria byla vybrána dle Hřebíčka:

1. Jde o útvar v přirozeném jazyce. V tomto případě se jedná o psanou japonštinu.
2. Jedná se o text uvažovaný jako útvar souvislý. Tj.: „V obecném vyjadřování lze nazvat textem útvar přerušovaný např. obrázkem, matematickým vzorcem nebo hudební vložkou či pauzou. ... Takový útvar prozatím nebudeme analyzovat, protože chceme znát tajemství souvislosti ..., která jazykové složky spojuje a kterou text představuje. Textem ... nemůže být útvar získaný náhodným

¹⁵ Popis analyzovaného textu č. 1 – *povídky* a popis metodiky se tedy může částečně shodovat s popisem uvedeným v (Benešová 2015) a práce tedy navazuje na původní experiment. Popisy dvou nových analyzovaných textů – *poezie* a *seminární práce* – jsou vytvořeny nově a pouze autorem této diplomové práce.

výběrem nějakých segmentů různého druhu z delšího textu, např. z románu. Mluvíme-li o textu jako o souvislém útvaru, máme na mysli nepřerušenu sekvenci jazykových jednotek na všech různých jazykových úrovních. Textem ... není seznam tvořený sekvencí jednotlivých jednoslovných nebo několikaslovných výrazů bez gramatické souvislosti, které představují nějaké vyčíslení jednotlivých položek ... Souvislý text má zřetelný začátek a konec a uvnitř této souvislosti není ničím (třeba obrázky, grafy, hudebními vložkami, nadměrnými pauzami) přerušen“ (2002, s. 42).

3. Text vybraný pro analýzu musí mít přiměřenou délku. Je třeba „dbát o to aby námi analyzované texty nebyly ani příliš krátké, ani příliš dlouhé. ... Kdybychom vzali příliš dlouhý text, např. nějaký Tolstého román ..., asi bychom měli velké množství jazykového materiálu, ale v něm by se mohly ztratit strukturní závislosti a jemnosti, o něž pravděpodobně jde. U dlouhých textů můžeme předpokládat, že se nějakým způsobem rozpadají na ‚pod-texty‘. Proto za text budeme brát spíše jednotlivou kapitolu románu, je-li rozumně dlouhá, než celý román. Pokud jde o příliš krátké texty tvořené třeba jednou, dvěma, třemi větami, nebo dokonce jedním slovem atd., musíme při jejich analýze uvážit, jakými nástroji budeme konat naše pozorování. V jedné nebo několika větách nesporně jsou skryty zákonitosti platné pro delší text. Naše nástroje však mají statistickou povahu a vyžadují, aby text byl delší, aby se mohly projevit veličiny, které odhalují tendence k realizaci určitého druhu struktury. ... Je to podobné jako v přírodních vědách, kde musí být splněny určité technické požadavky, aby experiment byl proveditelný a aby bylo možné získat měřitelné hodnoty.“ (2002, s. 44)
4. Text musí být mnohvrstevně segmentovaný, tj. skládat se z řady bohatě vrstvených elementů nebo jednotek: „Řekneme-li, že něco má podobu přirozeného jazyka, že se to vyskytuje v přirozeném jazyce, říkáme tím zároveň, že je to složitě segmentované. Takové jsou zřejmě všechny přirozené jazyky. ... Texty *musejí být členitě organizované*“ (2002, s. 45).

Japonské textové útvary vybrané pro ověření platnosti MAZ vyhovují výše uvedeným kritériím. Jedná se o následující texty:

2.1. Analyzovaný text č. 1 – povídka

Prvním z analyzovaných textů, vybraných pro verifikaci platnosti Menzerath-Altmanova zákona aplikovaného na psanou japonštinu, a zároveň pro započítání výzkumů v této oblasti, byla povídka *Šinkon-san* (新婚さん) od současné japonské spisovatelky Banany Jošimoty.¹⁶

Tato povídka vyšla poprvé ve sbírce povídek *Tokage* (とかげ) roku 1993 v nakladatelství Shinchosha Publishing; analyzovaný text pochází z vydání z roku 2000.

Délka textového vzorku je 6204 znaků (v souladu s definicí znaku přibližné v kap. 2.5.5. Znaky), resp. 14 stran tištěného textu v původní knižní podobě.

Tento text byl již dříve analyzován a stručně popsán v (Benešová 2015). Důvody pro jeho opětovné zahrnutí do této diplomové práce jsou uvedeny v kap. 2.

2.2. Analyzovaný text č. 2 – poezie („západního“ stylu)

Dalším analyzovaným textem pro ověření MAZ byla báseň *Havran* (japonsky *Ógarasu* 大鴉) od amerického básníka Edgara Allana Poea v japonském překladu z roku 2013, vydaném v elektronické podobě pro internetový obchod iTunes Store.

Originál básně pochází již z roku 1845 a na dnešní dobu nevšední jazyková stránka byla v japonském překladu do jisté míry zachována – především je patrné častější užívání znaků *kandži* neobsažených ve standardizovaném seznamu znaků *Džójó kandži hjó*. Délka textu je 2721 znaků.

2.3. Analyzovaný text č. 3 – vysokoškolská seminární práce

Posledním z analyzovaných textů je vysokoškolská seminární práce od japonské studentky z oboru literatury v rozsahu 5594 znaků. Jedním z kritérií pro výběr seminární práce mimo jiné bylo, aby práce byla dokončena a řádně odevzdána k (úspěšnému) ohodnocení. Práce rovněž nesměla být nadměrně přerušována např. grafy, matematickými vzorci, tabulkami aj.

¹⁶ Banana Jošimoto (吉本ばなな), narozena 24. července 1964. Viz (Jošimoto 2000).

2.4. Segmentace výběrových souborů

Pro tento experiment byl zvolen grafický princip segmentace vybraných textových souborů. Grafický princip segmentace byl zvolen jednak z důvodu specifického systému japonského písma a rozvržení japonských textů v porovnání s často analyzovanými jazyky evropskými (ve smyslu geografickém), jednak pro navázání na již uskutečněné experimenty s texty čínskými. Dle Claudie Prün navíc vykazují graficky segmentované znaky *kandži* charakteristiky Menzerath-Altmanova zákona (1994, s. 149); Motalová a Matoušková pak došly k podobně pozitivním závěrům při analýze textů čínských.¹⁷

Segmentace textu na jazykové jednotky byla do jisté míry nastavena tak, aby po úvodní instruktáži a se stejnými prostředky dokázal s minimem chyb nasegmentovat japonský text i člověk, který japonsky číst a psát neumí, tedy jinak řečeno, aby mohla být segmentace provedena v co možná největší přípustné míře na grafickém principu.¹⁸

Vybrané texty byly rozsegmentovány do sedmi jazykových jednotek a následně kvantifikovány a převedeny do formátu pro užití v softwaru MA Studio. Zvolené jazykové jednotky jsou:



Tabulka 2: Jazykové jednotky

¹⁷ Viz Matoušková, 2014.

¹⁸ V praxi se toto kritérium neověřovalo, posloužilo však jako vodítko při volbě jazykových jednotek.

Tím byly vytvořeno 5 jazykových úrovní:

U₀ : Nadodstavce měřené v délce odstavců – odstavce měřené v průměrné délce souvětí
↕
U₁ : Odstavce měřené v délce souvětí – souvětí měřená v průměrné délce mezičárí
↕
U₃ : Souvětí měřená v délce mezičárí – mezičárí měřená v průměrné délce znaků
↕
U₄ : Mezičárí měřená v délce znaků – znaky měřené v průměrné délce prvků
↕
U₅ : Znaky měřené v délce prvků – prvky měřené v průměrné délce tahů

Tabulka 3: Jazykové úrovně

„Je nápadné, že všechny úrovně tvoří žebříček, v němž každá úroveň tvoří jeden nadřazený a podřazený stupeň ...“ (Hřebíček 2002, s. 51).

Pro výpočet parametrů A a b , koeficientu determinace R^2 a dalších statistických údajů bylo užito softwaru MA Studio (ver. 2.11.0.0). Grafy s proloženými křivkami použité v této práci pocházejí rovněž z MA Studio.

Interval pro koeficient determinace R^2 je v rozmezí od 0 do 1. „Hodnoty R^2 větší nebo rovny 0,7 mohou prokazovat adekvátní a dobře sedící model v kvantitativní lingvistice“ (Benešová 2011, s. 77).

Jelikož četnost výskytu jednotlivých bodů x (konstruktů) o dané délce se od sebe často výrazně odlišovala (a to až v řádu tisíců výskytů oproti jednotkám výskytů v rámci jediné jazykové úrovně), což by nebylo respektováno v případě provádění výpočtů s prostým průměrem, všem jednotlivým empiricky získaným bodům byla přiřazena určitá váha dle četnosti jejich výskytu. Díky tomu nebude nutné násilně odstraňovat pozorování s extrémně nízkou četností výskytu. V grafech extrahovaných z programu MA Studio je přítomna teplotní mapa, která představuje četnost výskytu jednotlivých bodů x dle odpovídající barvy na škále od červené (nejvyšší četnost) přes žlutou a zelenou až k modré (nejnižší četnost).

Co se týče výpočtů, bylo počítáno s každým jednotlivým výskytem každé jazykové jednotky, tzn. v případě opakování strukturně totožné jednotky (např. stejného znaku) v textu nebyl tento opakující se znak při druhém a dalším výskytu ignorován; bylo s ním počítáno jako s novým znakem se stejnými číselnými hodnotami. Důvod je ten, že (pokud pomineme nesmyslnost násilné redukce autorem záměrně napsaného textu), např. v případě jazykové hladiny U_4 u analyzovaného textu č. 3 – seminární práce, by došlo k nezohlednění celých 81,9 % z celkového reálného počtu prvků obsažených v textu. To stejné s rozdílným podílem ztrátovosti platí pro všechny jazykové hladiny.

Užitý matematický vzorec pro výpočty je $y = Ax^{-b}$.¹⁹

2.5. Definice zvolených jazykových jednotek

Jazykové jednotky, užití pro segmentaci vybraných japonských textů, byly vytvořeny za účelem užití právě v tomto experimentu a nemusí se tedy svým vymezením shodovat s běžně užívanými jednotkami se stejným názvem. Toto platí především pro jednotky *znak* a *prvek*, které se od běžného pojetí stejnojmenných jednotek v japonštině značně odlišují. Níže uvedené definice je tedy nezbytné chápat v kontextu tohoto experimentu (zaměřeného na segmentaci čistě grafickou) a platí pouze v rámci tohoto experimentu; ne všechny vychází ze zažitého úzu. Přestože budou často popisována i běžně zažitá užití daných jednotek, zde uvedené jednotky a definice by se neměly porovnávat ani zaměňovat s jednotkami a definicemi běžně užívanými.

Následuje podrobný popis výše uvedených jazykových jednotek.

2.5.1. Nadodstavce

Nadodstavec je jednotka oddělující větší tématické úseky textu zřetelněji než běžné odstavce. Graficky je nadodstavec reprezentován odsazením textu o celý jeden nebo dva prázdné řádky či sloupce a započítím nového (nad)odstavce. Není však v japonských textech běžný a nemusí být v textu vůbec užíván – mnohé texty neobsahují vůbec žádné

¹⁹ Pro popis vzorce MAZ viz 1.2. *Menzerath-Altmanův zákon*
Pro důvody užití této podoby vzorce viz (*www 4*).

nadodstavce. Není to běžně citovaná jednotka a byla vybrána pouze na základě grafické výraznosti.

Z důvodu velmi nízké frekvence výskytu této jednotky ve vybraných textech byla jazyková úroveň *nadodstavce-odstavce* označena jako úroveň U_0 a z následné analýzy vypuštěna, jelikož by výsledky analýzy této jazykové úrovně s velkou pravděpodobností neměly velkou vypovídací hodnotu. Je však třeba mít existenci této jazykové úrovně na paměti při případné analýze delších textů obsahujících nadodstavce, jelikož se i tato jazyková jednotka/úroveň může v případě některých textů ukázat jako důležitá.

2.5.2. Odstavce

Odstavec začíná vždy na novém řádku (či sloupci v případě vertikálně psaného textu) s odsazením, reprezentovaným obvykle jedním prázdným grafickým polem na začátku řádku či sloupce. Výjimka z tohoto pravidla nastává v případě, kdy nový odstavec začíná místo odsazení o jedno prázdné grafické pole jedním z uvozovacích znamének, tj.

「 nebo 『 (resp. — nebo ⇔ v případě vertikálně psaných textů) či velice výjimečně znaménkem jiným. Nový odstavec započne i bez odsazení či uvozovacího znaménka v tom případě, že předchozí řádek či sloupec byl viditelně ukončen za účelem vytvoření nového odstavce.

Odstavce mají v japonském textu poměrně jasně vymezená užití. Ačkoliv nemají gramatickou či ortografickou funkci, jejich užití je řízeno určitými (ač nezávaznými) pravidly.²⁰

2.5.3. Souvětí

Jednotka souvětí je obvykle ukončena tečkou 。 (*kuten* 句点), otazníkem ? (*gimonfu* 疑問符), nebo vykřičníkem ! (*kantanfu* 感嘆符). V některých případech může být souvětí ukončeno i jiným způsobem²¹ a lze si představit i souvětí, které není ukončeno, přestože bylo užito jednoho z výše uvedených interpunkčních znamének.²² V takových případech je třeba kromě stránky grafické zohledňovat i sémantiku či kontext

²⁰ Viz (www 5) pro více informací o obvyklém užití odstavců v japonských textech.

²¹ Např. v případě přerušené výpovědi.

²² Např. v případě vložené výpovědi, citace, zvýraznění textu apod.

a ukončovací znaménka souvětí pak výjimečně, ale vždy stejně, ignorovat (resp. nakládat s nimi stejně, jako s jazykovou jednotku znak, viz 2.5.5. Znaky).

Je vhodné poznamenat, že označení „souvětí“ zde figuruje pouze jako název jazykové jednotky, tj. nemusí se jednat vyloženě o souvětí o několika klauzích apod.; pro účely tohoto experimentu může jít např. i o větu holou aj. Pro segmentaci jsou tedy ve většině případů podstatná pouze ukončovací znaménka uvedená výše.

2.5.4. Mezičárí

V japonském jazyce je, jako ostatně ve všech jazycích, definice slova relativně problematická. Existuje několik vlivných teorií gramatik (mimo jiné např. tzv. školská gramatika²³ a velice podobná Hašimotova gramatika²⁴, dále Tokiedovy gramatiky²⁵, Jamadova gramatika²⁶ atd.), které se značně rozcházejí ve svých přístupech k pojetí pojmu „slovo“ či jeho ekvivalentu. S ohledem na výše uvedené bylo tudíž namísto poměrně vágně a především sémanticky specifikované jednotky slovo, umístěné mezi jazykové jednotky souvětí a znak, experimentálně užito snadno graficky segmentovatelné jazykové jednotky mezičárí, rovněž umístěné mezi jednotky souvětí a znak.²⁷

Interpunkční čárky 、 nebo , (*tóten* 読点) mají v japonštině mimo jiné pomocnou funkci napomoci čtenáři číst psaný text plynuleji a snáze mu porozumět. Existují samozřejmě některá více méně běžně aplikovaná pravidla pro užívání čárek (obvykle, avšak ne vždy, za určitými spojkami a partikulemi, pro spojení dvou klauzí, u výčtů, pro odlišení časové a příčinné souslednosti apod.), v mnoha případech však může být čárka užita, aniž by její užití bylo obligatorní, tj. její užití není vždy řízeno pevně danými pravidly. Čárka může rozdělovat klauze, vymezovat výčet atd., ale nutnost jejího užití může být vágní a mnohdy je tak na autorovi, kdy a kde užije čárky ve sporných případech.²⁸

²³ *Gakkó bunpó* 学校文法

²⁴ *Hašimoto bunpó* 橋本文法, resp. *Kokugo hójó secu* 国語法要説

²⁵ *Tokieda bunpó* 時枝文法, resp. např. *Gengo katei secu* 言語過程説

²⁶ *Jamada bunpó* 山田文法, resp. např. *Nihon bunpóron* 日本文法論

²⁷ V budoucích fázích výzkumu se lze pokusit definovat slovo přesněji i s pomocí metod využívajících MAZ.

²⁸ Následující věta (Nicú 2004, s. 25) poslouží příkladem neobligatorního užití čárky pro usnadnění čtení:

„あの選手はオリンピックではなばなしい活躍をした。“

Jelikož partikule で je velmi často následována partikulí は (grafém *hiragany* „は“ je vyslovován

Jedno pravidlo však patrně platí vždy: Pokud se v textu objeví čárka, poskytuje čtenáři vodítko, kde má při čtení udělat krátkou pauzu nebo kde alespoň zpomalit a zběžně se podívat na několik následujících znaků, než bude opět pokračovat ve čtení běžnou rychlostí. Lze tedy předpokládat, že japonská čárka má, mimo své další funkce, roli grafického markeru pauzy či zpomalení během čtení, která se uplatňuje vždy.

Tento výzkum je založen na analýze grafické podoby vybraných textů. Kvůli určité vágnosti v užívání japonské čárky ve sporných případech byl tudíž vytvořen segment mezičárí, vymezený na jednom konci začátkem nového souvětí nebo obvykle čárkou, a čárkou nebo znaménky ukončujícími souvětí²⁹ na straně druhé. Tento způsob segmentace dovoluje snadnou identifikaci segmentu mezičárí na čistě grafickém principu. Je to segment velice podobný *parcelátu* (viz Matoušková 2014 aj.), avšak vzhledem k odlišnostem užití čárky v japonštině a v čínštině a pro předejití dalším neurčitostem byl název tohoto segmentu pozměněn.

V případě kulatých závorek v textu byl způsob segmentace následující: Pokud se jednalo o závorku zpřesňující výpověď dalším delším textem (a obvykle zároveň obsahující sloveso, tj. jako právě tento typ textu mezi závorkami v této větě) a ve většině případů tak určitým způsobem narušující plynulost výpovědi, byla i kulatá závorka považována za konec či začátek segmentu mezičárí. Pokud se jednalo o závorku, jejíž obsah plynulost výpovědi výrazným způsobem nenarušoval, neobsahoval sloveso a byl (většinou) „jednoslovný“, tj. tak, jak právě bylo užit slovo *většinou* v závorce v tomto souvětí, nebyly takové závorky považovány za znaménka ukončující či započínající segment mezičárí. V tomto případě se tedy výjimečně muselo přihlídnout i k jiným než čistě grafickým hlediskům textu.

různými způsoby v závislosti na tom, zda je ho užit jakožto partikule, nebo jako samostatné slabiky), může snadno dojít k tomu, že se v této větě přečte špatně. V této větě grafém は není partikulí a měl by se tudíž vyslovovat jiným způsobem než je intuitivní při rychlém čtení textu. Aby se předešlo nejasnostem, stejnou větu lze zapsat (bez změny samotného významu v obou podobách) rovněž s pomocí čárky (1.), nebo s pomocí *kandži* (2.), které svou grafickou podobou usnadní přečíst větu správně bez nutnosti užití čárky:

1. あの選手はオリンピックで、はなばなしい活躍をした。
2. あの選手はオリンピックで華々しい活躍をした。

Je třeba poznamenat, že ani původní podoba věty není chybná.

²⁹ V tomto případě dojde k ukončení mezičárí i souvětí zároveň. Pokud by se v textu vyskytly arabské číslice s matematickými znaménky, která by za normálních okolností ohraničovala mezičárí či souvětí, výjimečně by se považovala za součást čísla a hranice mezičárí či souvětí by netvořila.

Dá se předpokládat, že mezi jazykovými jednotkami souvětí a znaku chybí ještě nějaké jiné jazykové jednotky, byť se nemusí jednat o jednotky definované čistě na grafickém principu. Těmito (v japonštině opět nesnadno jednoznačně definovatelnými) jednotkami by mohla být např. klauze, slovo (Dlouhé slovo, Krátké slovo či jiné slovo, viz např. *www 7*) nebo možná právě navrhované mezičáří.³⁰

(Určitou alternativu představuje možnost segmentace japonského textu např. na tzv. Dlouhá slova a Krátká slova³¹ pomocí segmentačního SW Comainu s morfologickým analyzárem MeCab a slovníkem UniDic a následnou ruční kontrolou a opravou. Tento způsob segmentace by však již odporoval striktně grafickému principu segmentace textu, jelikož se jedná o segmentaci založenou na analýze morfemické. Jako dodatečný experiment však byla v rámci této práce segmentace výše uvedeným způsobem rovněž experimentálně provedena na analyzovaném textu č. 3 – *seminární práci* (výsledky viz 4.4. Doplnující experimenty s analyzovaným textem č. 3)).

2.5.5. Znak

Znak je považován za jednu ze základních grafických jednotek japonského a čínského písma.

Každý znak zabírá jedno stejně velké grafické pole nezávisle na počtu tahů. Grafické pole má tvar čtverce, nebo obdélníku, jehož výška není o moc větší než šířka (Švarný 1967, s. 31).

Všechny japonské znaky (grafémy)³² jsou tedy zapisovány dovnitř stejně velkých (předpřipravených či imaginárních) grafických polí, vždy jeden znak do jednoho pole s výjimkou některých interpunkčních znamének, která mohou být za určitých podmínek „nacpána“ do pole předcházejícího (viz 2.6. Ostatní poznámky k segmentaci textu)³³.

³⁰ Je však třeba zdůraznit, že tento výzkum je zaměřen především na grafickou, tj. ne nutně sémantickou, strukturu textu.

³¹ Japonsky *čótan’i* (長單位), doslova „dlouhé jednotky“, a *tantan’i* (短單位), doslova „krátké jednotky“.

³² V případě *kany* není příliš vhodné za běžných okolností užívat slova „znak“, zde se však jedná o definici jazykové jednotky pro účely analýzy a (nejen) grafémy *kany* budou tedy v této práci shodně nazývány „znaky“.

³³ Další méně významnou výjimku tvoří zvláštní případy, kdy se v textovém procesoru více znaků záměrně zapíše (naformátuje) do velikosti jednoho čtvercového pole, buď z důvodu zkrácení (např. měrné jednotky jako ミリ → り či キロメートル → ー), nebo odlišného formátování (např. チヨ → ちヨ; v tomto případě se však celkový počet zaregistrovaných znaků nezmění) apod.

Opačný případ pak existuje např. v podobě emotikonů, které nezřídka zabírají více teoretických grafických polí v textovém procesoru, avšak do standardního formátu *genkó jóši* (viz dále) by se

Jednotlivá slova nejsou v japonštině oddělena mezerami (s výjimkou např. knih pro děti psaných v *hiraganě* (Kraemerová 2000, s. 33), reklamy aj.). Mezery se nevkládají ani mezi jednotlivé znaky³⁴ a ve výsledku je tak velmi obtížné přesně vymezit hranice japonských (ortografických) slov. Velice podobná situace platí i pro psanou čínštinu.

Segmentace vybraných textů na jednotlivé znaky se striktně přidržovala kritéria čtvercových grafických polí. Vše obsažené uvnitř rámce vymezeného pomyslným čtvercem/obdélníkem (resp. opravdovým čtvercem/obdélníkem dle *genkó jóši*³⁵), a nic navíc, bylo považováno za jeden samostatný znak, s výjimkou čísel a latinky.

Co se týče čísel, jelikož není logické rozdělit delší číslo přesahující jedno čtvercové grafické pole (např. 1868) dle kritéria grafických polí na větší počet „kratších“ čísel (v tomto případě např. na 18 a 68 vždy s jednou číslicí na každé polovině grafického pole, tj. se dvěma číslicemi v jednom poli), bylo s celým číslem zacházeno jako s jedním znakem *kandži*, s počtem prvků odpovídajícím počtu číslic, z nichž je tvořeno.³⁶ S racionálním číslem, které se však ve vybraných textech nevyskytlo, by bylo zacházeno stejně. O určité výjimce by se dalo uvažovat, pokud by byla čísla zapsána takovým způsobem, že jediná číslice by sama zabírala celé grafické pole (japonština umožňuje zápis jak dvou číslic, tak i jedné číslice, graficky o něco větší, do jednoho grafického pole): v takovém případě by teoreticky bylo možné segmentovat každou číslici zvlášť jako jednotlivé znaky. Tento způsob zápisu se však v textech neobjevil a rovněž i v tomto případě se jako vhodnější varianta jeví segmentace prvním uvedeným způsobem.

Ve vybraných textech se vyskytla i slova a zkratky psané latinkou, ačkoliv jejich počet byl minimální. Protože umístění počátečního písmene latinky do čtvercového pole je závislé na předchozích znacích, či jinak řečeno může nastat buď situace, že je písmeno umístěno do levé poloviny čtverce a druhé následně do poloviny pravé (čtverec by tedy obsahoval dvě písmena), stejně tak jako situace, kdy je levá (resp. pravá) polovina čtverce ponechána prázdná a první (resp. poslední) písmeno slova je umístěno až do

zapisovaly jen velmi ojediněle, obtížně a pravděpodobně nejednotně, protože se jedná o relativně nový a stále velice produktivní jazykový fenomén. Japonské emotikony jsou navíc obvykle zapisovány lineárně, tedy bez náklonu o 90° jako je běžné např. v češtině, což umožňuje vytvářet komplikovanější a delší obrazce. (Např. (ˆ◇ˆ) (*'卍) (°Д°;) Σ(□□|||) (ノ□益□)ノ多—— ((ˆ▽ˆ*))ケラケラ aj.)

³⁴ Až na vzácné případy, jako např. mezi příjmením a vlastním jménem osob v některých typech textů.

³⁵ Papír s předtisknutými čtverci pro zápis znaků.

³⁶ Tedy číslo 1868 je segmentováno jako 1 znak o 4 prvcích a 4 tazích dle obvyklého japonského způsobu zápisu číslic.

pravé (resp. levé) poloviny čtverce a výsledná konfigurace by pak závisela na větším množství do značné míry náhodných podmínek, jako např. na předcházejících znacích, číslicích, interpunkčních znaménkách a také na tom, zda je počet písmen ve slově sudý či lichý, což není možné systematicky a objektivně měřit na textu, jehož čtvercová grafická pole jsou pouze imaginární (např. prostý digitalizovaný text v elektronickém formátu *.txt či *.doc), bylo s celým graficky odděleným slovem zapsaným latinkou zacházeno stejně jako s jedním znakem *kandži*, s počtem prvků odpovídajícím počtu písmen v ortografickém slově.³⁷

Interpunkční znaménka, včetně čárek, teček a dalších, byla rovněž považována za znaky. Přestože při grafické analýze mnoha jiných jazyků není obvyklé do samotné analýzy zahrnovat i interpunkční znaménka (naopak, v mnoha případech se ignorují), tato volba je podpořena kritériem grafických polí v japonském jazyce, kde jedno znaménko obvykle zabírá jedno celé pole, dále existencí tzv. „mlčících vět“³⁸ (*činmokubun* 沈黙文) a v neposlední řadě také silnou rolí interpunkčních znamének na modalitu výpovědi.³⁹ Z toho plyne, že interpunkční znaménka (nejen) v japonském jazyce by neměla být zcela ignorována.⁴⁰

Znaky *kany* bývají méně složité než znaky *kandži* (co do počtu prvků a tahů) a frekvence výskytu některých znaků *hiragany* v psaných textech je několikanásobně vyšší z důvodu jejich výrazně menší variability (ve srovnání s *kandži*) a četných

³⁷ Např. zkratkové slovo *OL* je tedy segmentováno jako 1 znak o 2 prvcích a 2 tazích dle obvyklého japonského způsobu zápisu latinky.

³⁸ Viz (Koike 2000, s. 65). Koike uvádí následující dialog jako příklad mlčících vět:

„Otec: 「おまえ、大学へ行くのか。」 (Poslyš, na výšku ale půjdeš, ne?)

Syn: 「.....」

Otec: 「行かないのか。」 (Nebo snad nepůjdeš?)

Syn: 「.....」

Otec: 「黙ってないで、なんとか言いなさい！」 (Co mlčíš? Tak už něco řekni!)

Syn: 「.....」

Otec: 「考えが、まあだ、まとまってないのか！」 (To ses pořád ještě nerozhodnul?)

Syn: 「.....」 “

V případě vynechání těchto mlčících vět by dialog postrádal smysl. Výše uvedené mlčící věty by při segmentaci byly rozděleny následujícím způsobem: 4 znaky (v souladu se 4 zabranými grafickými poli v případě tisku či psaní rukou); 9 prvků; 11 tahů.

³⁹ Např. i české věty „Je to pravda.“, „Je to pravda?“, „Je to pravda!“ a podobné značným způsobem mění svůj význam v závislosti na užitých interpunkčních znaménkách.

⁴⁰ Viz (Barešová 2014, s. 83) pro více informací o japonských interpunkčních znaménkách.

gramatických funkcí.⁴¹ Princip grafického pole však platí jak pro zápis v *kandži*, tak i pro *kanu* a interpunkční znaménka.

Je však třeba podotknout, že ačkoliv se znaky *kana* vyvinuly ze znaků *kandži*, jejich zápis je řízen podobnými pravidly a jsou mnohdy vzájemně nahraditelné, je sporné, zda o nich při segmentaci textu lze uvažovat ve smyslu ekvivalence. Jelikož se jedná o typově odlišná písma a velice zjednodušeně řečeno lze říci, že znaky *kandži* v sobě „zahrnují“ ideje, což pro moraicke (syllabografické) znaky *kana* neplatí, dalo by se o *kandži* uvažovat také jako o nadřazené, nebo snad ještě lépe paralelní jazykové jednotce vůči znakům *kana*. V takovém případě však vyvstává otázka, jakým způsobem nejlépe provést segmentaci textu. Jako příklad problémů, pokud se o těchto dvou písmech nebude uvažovat jako o jednotkách ekvivalentních, lze uvést nestoprocentní jednotnost přepisu *kandži* do *kany* (některé znaky lze buď přepsat více způsoby beze změny významu,⁴² možnost více způsobů přidávání *hiraganových* afixů, byť se může jednat o verzi zápisu méně užívanou,⁴³ nebo možnost převést celý text obsahující *kandži* i *kanu* do v podstatě totožného textu obsahujícího pouze *kanu* atd.). Hlavní překážkou by pak byl jednoduše fakt, že přestože lze v moderní japonštině mít text tvořený pouze znaky *kany*, nelze (pokud neuvažujeme krátké či cíleně a nepřirozeně zkonstruované texty) mít text tvořený pouze znaky *kandži*, a dochází tedy k jejich přirozené kombinaci, což značným způsobem komplikuje chápání *kandži* jako nadřazené jednotky. Navíc by při pouhé konverzi celého textu do znaků *kany* (pokud pomineme jen občasný problém s nejednotností přepisu) mohlo docházet k určité ztrátě informace: *kandži* se neužívají zcela nahodile a násilnou konverzí může dojít ke změně záměru autora textu. V případě uvažování jako o jednotce paralelní by pak bylo třeba vytvořit nový sjednocený postup segmentace, zohledňující všechny výše uvedené a některé další problémy.

Tento jazykový experiment operuje se znaky *kana* jako s jednotkami na ekvivalentní jazykové úrovni s *kandži*, latinkou a dalšími znaky.

⁴¹ Znaky *hiragany*, které jsou obvykle jednodušší (co do počtu prvků a tahů) než znaky *kandži*, se často objevují ve funkci slovesných afixů a v důsledku aglutinační povahy japonského jazyka do značné míry prodlužují podobu sloves (tj. zároveň s tím stoupá počet relativně jednoduchých znaků ve větě). Současně je *hiragana* často užívána pro zápis např. spojky a plní mnoho dalších gramatických funkcí. Dá se předpokládat, že čím delší je věta, tím více tato věta v průměru obsahuje spojku, slovesných afixů apod., v nichž se uplatňují znaky *hiragany*.

⁴² Např. 寂しい jako さびしい i さみしい nebo „obráceně“ ひっこし jako 引っ越し i 引越 apod.

⁴³ Např. やわらかい jako 柔らかい i 柔かい apod.

2.5.6. Prvky

Obecně lze říci, že znaky⁴⁴ jsou tvořeny z prvků, které jsou tvořeny z tahů. Jak však uvádí Motalová (2013) a jiní, „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 ... obecné definice, která tvrdí, že prvek je nižší jednotkou než znak.“

Existuje více možných způsobů rozdělení jednotlivých prvků (částí) znaku; tyto však nejsou zcela jednotné a často vycházejí ze sémantických hledisek, což není příliš vhodné pro čistě grafickou segmentaci a analýzu vybraných textů.

„V dějinách bylo vytvořeno mnoho seznamů prvků *kandži* a všechny z nich vykazují mnoho shodných, ale zároveň i odlišných bodů v závislosti na tom, k jakému účelu byly seznamy vytvořeny. Účelem může být hledání znaků ve znakovém slovníku, učení a zapamatování *kandži* nebo etymologická diskuse o formách znaků. Většina seznamů zahrnuje prvky, které jsou celé složené z jiných prvků ze stejného seznamu, nebo lze nalézt *kandži*, pro která existuje více způsobů redukce na prvky“ (Prün 1994, s. 148), což také zohlednil Stalph (1989, s. 73) při tvorbě svého vlastního seznamu prvků *kandži*.

Současné japonské slovníky většinou uvádějí 214 znakových radikálů (Vochala 1975, s. 136), tj. složek, dle kterých se znaky řadí do slovníků, které se často zaměňují za prvky či determinativy⁴⁵ nebo jiné složky znaku. Existuje však více variant stejných radikálů i prvky, které radikály nejsou.

Pro zajištění alespoň částečné porovnatelnosti výsledků analýzy psané japonštiny a psané čínštiny (nezávisle na tom, jak se tyto dva systémy při bližším zkoumání navzájem odlišují), a zároveň pro otestování této nové metody na japonských textech, bylo při segmentaci znaků na prvky užito stejného kritéria tzv. „ostrovů“, které Motalová aj. užily při své analýze čínských textů.

Dle kritéria ostrovů jsou pouze vzájemně se dotýkající tahy považovány za prvky: Jelikož definice prvků se různí, byla užita segmentační metoda, která rozděluje znaky na prvky na základě dotyků tahů, tedy tzv. „ostrovů“. Na základě této koncepce se prvek

⁴⁴ Zde myšleno znaky *kandži*, resp. čínské *hanzi*. V souvislosti s odlišnými znaky *kana* aj. se běžně, tj. mimo rámec tohoto experimentu, o konceptu prvku neuvažuje.

⁴⁵ Determinativem se nazývá složka znaku mající sémanticky distinktivní funkci (napovídá významový okruh znaku). Neobsahuje ji však každý znak, přestože každý znak obsahuje radikál (Jindřichová 2014, s. 10).

považuje za tzv. „ostrov“, tj. za oddělenou část znaku, která je tvořena jedním tahem nebo skupinou spolu propojených tahů a samozřejmě oddělených od ostatních částí (tj. prvků) znaku. Různé kombinace těchto jednotek vytvářejí další jazykovou jednotku – znak⁴⁶ (Motalová 2013, s. 40).

Jelikož tento způsob segmentace se již na první pohled odlišuje od běžně užívaných způsobů rozdělování znaků na prvky, je lépe ho chápat spíše jako vypůjčení názvu „prvek“ pro daný segment, tj. nejedná se o nahrazení sémanticky a lexikograficky zavedených pojetí prvků (které se však nepříliš hodí pro segmentaci grafickou) v japonských či čínských znacích. Navíc experimenty s čínskými texty s užitím takto definovaných prvků vykazaly vysokou míru platnosti Menzerath-Altmanova zákona. To je další z důvodů pro experimentální zavedení výše uvedeného způsobu segmentace; tímto způsobem zároveň bude možné porovnat, zda segmentace japonského textu, který obsahuje více typů písma než texty čínské, prokáže stejně vysokou platnost MAZ.

Motalová a Spáčilová rovněž poukazují na to, že tento způsob segmentace znaků na prvky bude ovlivněn užitým typem fontu písma (2013, s. 40). Jelikož jedním z nejčastěji používaných fontů v japonštině je font Mincho (明朝体), bylo ho rovněž užito pro analýzu vybraných textů.⁴⁷

Protože zatím neexistuje žádný software, který by byl schopen automaticky vypočítat počet ostrovů ve znacích, byla kvantifikace provedena ručně pro každý jednotlivý znak. Aby se předešlo obrazovým chybám, způsobeným nízkým zobrazovacím rozlišením (kdy se při nižším rozlišení budou jednotlivé části znaků nutně propojovat kvůli fixní velikosti pixelů), byly všechny znaky zvětšeny na velikost 500 v editoru Microsoft Word 2013 při rozlišení monitoru 1600×900 pixelů (131 PPI).

Pokud se způsob segmentace metodou ostrovů ukáže jako nevhodný, může být v pozdějších výzkumech užito i jiného způsobu identifikace znakových prvků např. na základě některého široce užívaného slovníku japonštiny, nebo např. využít originální metodu segmentace navrženou Jürgenem Stalphem.⁴⁸

⁴⁶ Např. znak 一 je dle této definice tvořen 1 prvkem, znak 魔 rovněž 1 prvkem a znak 感 8 prvky.

⁴⁷ Přesněji fontu MS Mincho v Microsoft Wordu 2013.

⁴⁸ Viz (Stalph 1989, s. 73).

2.5.7. Tahy

Tah byl vybrán jako nejmenší grafická jazyková jednotka japonského textu. Každý japonský znak (v případě standardního zápisu) je tvořen určitým počtem a pořadím tahů,⁴⁹ jichž je omezený počet druhů.

Tah je linie (i zakřivená), která je napsána nepřerušovaně, tj. bez nadzvednutí psacího náčiní během psaní znaků rukou (pozastavení bez nadzvednutí se tedy nepovažuje za přerušování tahu), pokud je striktně zachováno běžné pořadí a způsob zápisu tahů.⁵⁰

Běžný počet tahů japonských znaků *kandži* se rovněž v některých případech odlišuje od běžného počtu tahů identicky vypadajících čínských znaků *hanzi*. V tomto experimentu se vycházelo z počtu tahů běžného pro japonštinu.⁵¹

Zde je vhodné poznamenat, že v tomto případě nejde o segmentaci na čistě grafickém principu, ve smyslu možnosti snadné identifikace jednotlivých tahů již na první pohled, jako např. v případě znakových prvků (ostrovů).

2.6. Ostatní poznámky k segmentaci textu

Tečka na konci japonské věty obvykle zabírá celé grafické pole. Výjimkou z tohoto pravidla je případ, kdy tečka připadá na úplný konec řádku či sloupce textu, kde již nezbývá žádné prázdné imaginární či reálné pole a za normálních okolností by většina jiných znaků byla zapsána na začátku následujícího řádku či sloupce. V takovém případě může být tečka zapsána buď za běžně vyhrazenou psací plochu mimo čtvercové grafické pole, nebo může být namísto toho vložena do pole předchozího společně s předchozím znakem. Tečky a čárky ve vybraných textech byly vždy považovány za jeden samostatný znak sestávající z 1 prvku a 1 tahu a zabírající samostatné grafické pole, s výjimkou následujícího případu:

⁴⁹ Některé slovníky se však v počtu či pořadí tahů u malého počtu znaků v popisu mírně rozcházejí.

⁵⁰ Habein (2000) definuje tah následujícím způsobem: „Za jednotlivý tah považujeme jakoukoliv linii napsanou perem či tužkou v nepřerušovaném kontaktu s papírem, nehledě na to, jak se tato linie zaostřuje a ohýbá.“ Samotná klasifikace různých typů tahů není jednotná a různé publikace, zabývající se čínským nebo japonským písmem či kaligrafií, často uvádějí odlišné počty a typy tahů. Pro tento experiment to však není relevantní, jelikož se užívá pouze zde uvedené obecné definice tahu, resp. počtu tahů, uvedeného ve slovníku *Shin Kangorin* (Kamada 2004), a není tedy nezbytné jednotlivé tahy blíže specifikovat.

Např. znak 一 se tedy skládá z 1, znak 及 ze 3 a znak 藍 z 18 tahů.

⁵¹ Dle znakového slovníku *Shin Kangorin* (Kamada 2004).

Jelikož tečka ve funkci ukončení přímé řeči je obvykle při psaní rukou do *genkó jóši* zapisována společně s interpunkčním znaménkem uvozujícím přímou řeč dohromady do jednoho sdíleného grafického pole, tato dvě znaménka byla považována za jeden znak sestávající z 2 prvků. To samé platí pro kombinaci čárky a uvozovacího znaménka, což není v japonštině jev běžný, objevil se však v analyzovaném textu č. 2 – *poezii*.

Lze rovněž uvažovat i o rozdělení samotného čtvercového grafického pole dále na poloviny či dokonce čtvrtiny aj., není však možné takto zacházet pouze s kombinací uvozovacího znaménka a tečky (příp. čárky), ale vynechat přitom všechny ostatní znaky. Různé sektory čtvercového grafického pole⁵² by však místo toho mohly fungovat jako základ pro nové samostatné jazykové (grafické) jednotky v případných budoucích experimentech.

Přerušeni výpovědi elipsou, označené zdvojenou japonskou výpustkou (*santen ridá* 三点リーダー) nebylo považováno za ukončení věty. V tomto případě se jednalo o výjimečné posouzení nejen grafické, ale i sémantické stránky textu s přihlédnutím ke kontextu věty.

Z důvodu kritéria grafických polí byly kombinované móry (*jóon* 拗音), zapisované jako digrafy *kja* (きゃ), *šu* (しゅ), *čo* (ちょ) apod. ve standardní velikosti,⁵³ považovány za dva samostatné znaky.

V analyzovaném textu č. 3 – *seminární práci* byly v původním textu místy přítomny krátké citace z odkazované literatury. Tyto tvořily nedílnou součást celku. Z toho důvodu byly pouze obligatorní odkazy na citovanou literaturu v závorkách při segmentaci pro účely tohoto experimentu ignorovány, jelikož se jednalo o čistě formální vložky (aby text vyhovoval kritériím japonské vysokoškolské práce), do velké míry nezávislé na původním přání a běžné produkci textu autora dané práce.

⁵² Viz (Spahn 1989).

⁵³ O teoretických případech, kdy by se více znaků pohromadě z důvodu neobvyklé velikosti naopak považovalo za jeden znak dle definice aj. viz 2.5.5. Znaky Takové případy se však v žádném z analyzovaných textů neobjevily.

Ze stejných důvodů byly při analýze odstraněny i jiné formální aspekty textů jako např. údaje o studentovi, *furigana*⁵⁴ v závorkách u analyzovaného textu č. 2 – *poezie* apod.

Japonská beletrie se většinou tiskne s vertikálně strukturovaným textem a to platí i pro analyzovaný text č. 1 – *povídku*. Z toho důvodu byl v tomto případě text ponechán beze změny (tj. ve vertikálním směru zápisu) i pro následnou analýzu.⁵⁵

Analyzované texty č. 2 – *poezie* a č. 3 – *seminární práce* byly naopak v původní podobě zapsány horizontálně; opět byl zachován původní směr zápisu textu i pro účely následné analýzy.

2.7. Užitý software

Výpočty pro analýzu dat byly provedeny pomocí softwaru MA Studio⁵⁶ (ver. 2.11.0.0), vyvinutého na Univerzitě Palackého v Olomouci.

Veškerá segmentace textů byla provedena ručně pomocí Microsoft Word 2013 a Microsoft Excel 2013.

Dodatečná experimentální segmentace na Dlouhá slova a Krátká slova byla provedena pomocí softwaru Comainu (ver. 0.71) s morfologickým analyzérem MeCab a slovníkem UniDic a následnou ruční kontrolou a opravou.

⁵⁴ *Furigana* je fonetický přepis (obvykle) znaků *kandži*, příp. i latinky aj., většinou v podobě sylabografických znaků *kany* zapisovaných nad daným znakem či znaky *kandži* aj. (v případě horizontálního zápisu), či vedle *kandži* aj. (v případě zápisu vertikálního), méně obvykle pak např. i v závorce za znaky *kandži* aj. Jelikož lze říci, že *furigana* ve většině případů nevnáší žádnou novou informaci pro čtenáře obeznámené se správným čtením znaků v daném textu a jedná se tedy o jakousi redundantní pomocnou informaci (jinak řečeno, pokud čtenář správné čtení zná, uvidí více méně stejnou informaci dvakrát, asi jako pokud by se v češtině např. nad číslicí 9 napsalo slovně *devět*), a tedy v případě odstranění *furigany* nedojde ke ztrátě výchozích znaků ani významu, byla při analýzách textů *furigana* ignorována, aby nedocházelo k dvojímu započítávání „týchž“ jednotek.

⁵⁵ Ve většině případů by směr zápisu japonského textu neměl ovlivnit proces grafické segmentace. Vzácné případy, kdy k drobným změnám dochází, však existují, jak je patrné např. u pořadí tahů interpunkčních znamének apod. Pokud se však nejedná o systém extrémně citlivý na vstupní podmínky, vzhledem k pouze nepatrnému množství změn způsobených odlišným směrem zápisu textu by k významným rozdílům ve výsledcích analýzy pravděpodobně docházet nemělo.

⁵⁶ Software určený pro kompletní práci s multidisciplinárně exponovaným Menzerath-Altmanovým zákonem.

3. Hypotézy

3.1. U_1 – odstavce-souvětí

Na jazykové úrovni odstavce-souvětí, měřené v průměrné délce mezičárí, očekávám rozdílné výsledky dle toho, o jaký styl textu se jedná. Je to dáno tím, že podoba odstavců se v různých stylech liší a následná segmentace proběhla s určitými změnami. Lze předpokládat, že u povídky a seminární práce se MAZ neprojeví, nebo nebude model příliš vyhovující, což by byl výsledek v souladu s obdobnými experimenty s texty čínskými (Motalová 2013) a s předchozím experimentem s japonskou povídkou.

Segmentace odstavců u těchto dvou textů bude probíhat přesně v souladu s definicí odstavce uvedené v 2.5.2. Odstavce Maximální délka odstavců (konstruktů), počet souvětí (konstituentů) ani průměrná délka mezičárí, tvořících tato souvětí, nejsou graficky omezeny, jako se děje v případě básně; tyto faktory tedy závisely zcela na vůli autorů. Pokud má MAZ obecnou platnost pro všechny přirozené texty nezávisle na omezení maximální velikosti jazykových jednotek, bude logické předpokládat, že na této jazykové úrovni by se měla potvrdit, nebo alespoň by měla být naznačena shoda s matematickým modelem. S ohledem na již získané výsledky analýzy čínských textů však nebude překvapující, pokud se ani u textů japonských (s možnou výjimkou básně) shoda neprokáže. Lze tak předpokládat, že na jazykové úrovni U_1 sice může být v případě povídky tendence pro platnost MAZ nastíněna, ale v nepříliš vysoké míře (zčásti kvůli vybranému literárnímu stylu). Je možné, že MAZ se projeví u ostatních dvou analyzovaných textů v o něco větší míře.

Možný odlišný výsledek analýzy předpokládám v případě básně, a to kvůli mírně odlišnému přístupu k pojetí odstavce. U tohoto textu bude za jeden odstavec považována jedna celá strofa, tedy ne ukončení řádku a započetí nového řádku (s případným odsazením) jako v případě textů ostatních.⁵⁷ Nebude se tedy jednat o segmentaci striktně dle uvedené definice odstavce, ačkoliv se i tento způsob segmentace stále přidrží grafického principu. (Pokud by však segmentace proběhla striktně dle definice, tj. ukončení řádku (verše) by znamenalo ukončení odstavce, pak by

⁵⁷ Dalo by se říci, že na konci verše řádek není ukončen, ale pouze zalomen kvůli prostorovým restrikcím a k ukončení odstavce jako obsahového celku dojde až s ukončením celé strofy.

odstavec byl často ukončen dříve, než by došlo k ukončení souvětí nebo mezičárí a byla by zcela narušena hierarchická kontinuita jazykových jednotek.⁵⁸) Jelikož vizuálně jsou všechny strofy v básni přibližně stejné dlouhé (tedy co do počtu znaků) a MAZ říká, že „čím delší je konstrukt, tím kratší jsou jeho konstituenty“, tj. v daném případě „čím delší je odstavec v počtu souvětí, tím kratší jsou tato souvětí v průměrném počtu mezičárí“, na této jazykové hladině by na rozdíl od ostatních dvou analyzovaných textů MAZ mohl platit. Podstatným důvodem pro tuto hypotézu je právě omezená (grafická) délka samotného odstavce, resp. strofy, což znamená, že čím více bude takto omezený odstavec obsahovat souvětí, tím méně by daná souvětí měla být schopna pojmout mezičárí, aby ve výsledku příliš nepřekračovala omezenou maximální délku odstavce. Tuto hypotézu do jisté míry podporuje přirovnání k pozorování o restriktivním čtvercovém grafickém poli, do kterého se zapisují jednotlivé znaky, kdy se na čínských textech ukazuje, že čím více má znak uvnitř prostoru omezeného grafického pole prvků, tím méně tahů mají v průměru tyto prvky (aby se stále vměstnaly do prostoru omezeného grafického pole).

Pokud se tato hypotéza ukáže jako platná, mohlo by to znamenat, že v případě grafické segmentace textů mají na projev vztahu, který udává MAZ, nezanedbatelný vliv různá grafická omezení textu, a to v tom smyslu, že čím silnější jsou grafická omezení, tím výraznější je platnost MAZ.

3.2. U₂ – souvětí-mezičárí

Na této jazykové úrovni by se vztah popsaný Menzerath-Altmanovým zákonem mohl potvrdit. Pro vybranou povídku a seminární práci na této úrovni platí, že texty byly autory tvořeny bez výraznějších grafických omezení, autoři mohli sami poměrně volně ovlivňovat délku souvětí i mezičárí a přizpůsobit tak text té podobě, která jim vyhovovala. Pokud by zároveň platila domněnka, že na projev vztahu formulovaného MAZ má nezanedbatelný vliv grafické omezení na některé jazykové jednotce, mohl by se MAZ ukázat jako o něco lépe sedící model na seminární práci než na povídce, a to z důvodu,

⁵⁸ Je vhodné zmínit, že tento způsob segmentace vede i k porušení definice jazykové jednotky nadodstavce, jelikož ignoruje prázdný řádek mezi jednotlivými odstavci. V případě budoucí analýzy delších básnických textů bez vyřazení jednotky nadodstavce by tedy mohlo být nutné poupravit i obecnou definici nadodstavce, aby se dala lépe aplikovat i na básnické texty.

že grafické rozvržení seminární práce jako celku přeci jen není zcela volné (přidrží se jistých psaných i nepsaných pravidel či zvyklostí pravděpodobně ve větší míře než povídka), tudíž koeficient determinace R^2 by se mohl více přibližovat hodnotě 1 než v případě povídky.

U básnického textu předpokládám, že by výsledek mohl být podobný jako v případě jazykové úrovně odstavce-souvětí, tedy že se vztah formulovaný MAZ rovněž může potvrdit, ale z poněkud odlišného důvodu. Maximální délka souvětí je totiž do jisté míry ovlivněna omezením maximální délky vyšší jazykové jednotky, tj. odstavce, a nemůže tedy být neomezená. Z toho důvodu by tedy opět mělo platit, že čím více budou souvětí (konstrukty) obsahovat mezičárí (konstituentů), tím kratší budou mezičárí v průměrném počtu znaků. Jelikož však není samotná jednotka souvětí v případě básnického textu tak striktně graficky omezena jako odstavce a počet znaků v mezičárí může být díky „drobnosti“ jednotky znaku variabilnější než počet mezičárí v souvětí, nemusí se tendence MAZ projevit stejně výrazně jako na úrovni U_1 . Pokud však platí hypotéza o vlivu grafických omezení textu na projev MAZ, měl by se v tomto textu MAZ ukázat jako lépe sedící model než u ostatních dvou analyzovaných textů.

3.3. U_3 – mezičárí-znaky

Po provedených experimentech s do značné míry podobnou jednotkou parcelátu při dřívější analýze čínských textů a japonské povídky v předešlém experimentu by nebylo překvapující, pokud by se na této jazykové úrovni platnost vztahu formulovaného MAZ neprokázala, nebo by byla naznačena jen ve velice malé míře.

Důvodů pro předpokládaný negativní výsledek může být hned několik. Mezi hlavní může patřit samotné užití jednotky mezičárí, která je založena na čistě grafickém principu; jelikož jazyková jednotka mezičárí je z hlediska do značné míry zavedených konvencí poměrně umělá a neobvyklá (lze předpokládat, že pravděpodobně jen málokdo vnímá japonské texty na úrovni „od čárky k čárce“, byť je pravda, že čárky v japonském textu mnohdy usnadňují jeho čtení a porozumění), nebylo by překvapující, kdyby se MAZ stabilně neprojevoval přinejmenším na jazykových úrovních U_3 a U_4 .

Dále je to fakt, že skok mezi jazykovou jednotkou mezičárí a jazykovou jednotkou znaku se zdá být poněkud velký. Zavedení další jedné (či více) mezijednotky však

komplikuje absence jasně definovaného či alespoň graficky jasně rozpoznatelného ortografického slova v japonštině. Díky současnému užívání více druhů písma se sice na rozdíl od čínštiny dají některá slova poměrně jednoznačně odlišit i graficky, nelze však hovořit o tom, že by takto bylo možno systematicky a bez odchylek nasegmentovat běžný, natož zcela jakýkoliv, japonský text.

Motalová aj. (2013) pak uvádí i některé další možné důvody neplatnosti MAZ na podobné jazykové úrovni v čínštině, které by mohly v podobné míře platit i pro texty japonské a lze tak předpokládat neplatnost MAZ na dané jazykové úrovni:

Zaprvé je to fakt, že mezičárí je jazyková jednotka s délkou proměnlivou (do značné míry v závislosti na vůli autora), zatímco jeho konstituenty, tj. znaky, jsou jednotky s délkou v počtu prvků fixní (tj. autor nemůže vlastní vůlí měnit počet prvků v užitém znaku, pouze může napsat znak jiný, což však často změní zamýšlený význam sdělení, nebo je to i zcela nemožné). Je pravda, že na rozdíl od čínštiny existuje v japonštině možnost „nahradit“ znaky *kandži* „ekvivalentními“ znaky *kany*, což již autor může vlastní vůlí ovlivňovat, mnohdy však i tato změna může nést určité následky a výsledný text tak nemusí být 100% ekvivalentní textu původnímu, alespoň co se týče výsledného dojmu na čtenáře (viz 2.5.5. Znaky). Nehledě na to, že záměrné nahrazování znaků za jiné jen z toho důvodu, že by to možná mohlo mít vliv na projev MAZ, jistě žádný producent běžného textu ať už vědomě či nevědomě nedělá (ačkoliv je pravda, že i volba, zda užít *kandži* či *kany* někdy bývá do určité míry ovlivňována zavedenými zvyklostmi či funkčním stylem textu). Druhou uvedenou příčinou neplatnosti MAZ na čínských textech byla malá variabilita znaků, co se týče v nich obsažených prvků, což by mohlo narušit vztah nepřímé úměrnosti mezi konstrukty a konstituenty (Motalová a Spáčilová 2013, s. 115).

Pokud se prokáže, že v souladu s hypotézou ani v japonštině MAZ na úrovni mezičárí-znaky neplatí na žádném z analyzovaných textů, bude blíže prozkoumán a porovnán podíl znaků podle počtu prvků na celkovém počtu znaků obsažených v textech pro zjištění toho, zda by toto mohlo být důvodem neplatnosti MAZ také u textů japonských.

Jelikož na této jazykové úrovni již ani v případě básnického textu nemá na délku konstituentů či nižších jednotek vliv omezená délka odstavců, výše uvedené by pravděpodobně mělo ve stejné míře platit tentokrát i pro analyzovaný text č. 2 – *poezii* a výsledky analýzy by tedy mohly být podobné jako v případě ostatních dvou textů.

3.4. U₄ – znaky-prvky

Na nejnižší jazykové úrovni U₄ by se na první pohled možná dalo očekávat, že se vztah vyjádřený MAZ projeví ve značně vysoké míře, podobně jako v případě čínštiny, jelikož japonština také obsahuje původem čínské znaky. Je však velice důležité neopomenout, že japonština obsahuje i znaky *kana*, které jsou navíc v jakémkoliv japonském textu obvykle velice frekventované. Na této úrovni tedy proti sobě stojí dva problémy:

Na jedné straně se všechny japonské znaky (dle zde uvedené definice znaku) zapisují do striktně vymezených čtvercových grafických polí, které limitují množství prvků, jež se mohou dovnitř vměstnat. Právě tento argument by mohl nahrávat tomu, že na jazykové úrovni znaky-prvky by se shoda s matematickým modelem MAZ mohla projevit.⁵⁹

Na straně druhé však stojí argument, že znaky *kandži* a znaky *kana* (a případné další, např. latinka či interpunkční znaménka, která rovněž spadají do uvedené definice znaku) možná nejsou jednotkami ekvivalentními, tj. ideálně by měly tvořit hierarchicky odlišné, příp. paralelní jazykové jednotky. Je také fakt, že zatímco u znaků *kandži* se běžně mluví o konceptech prvků, tahů apod. (nezávisle na jejich definici), u znaků *kana* a dalších se s těmito koncepty obvykle nepočítá. Ačkoliv jsou tedy limitovány stejně jako znaky *kandži* pevně danými grafickými poli (přinejmenším ve smyslu jeden znak *kany* do jednoho pole, resp. určitý fixní počet (obvykle) jiných znaků do jednoho pole), je možné, že v tomto bodě by bylo vhodné s jejich totožným zacházením skončit a dále tyto znaky segmentovat nějakým odlišným způsobem; ekonomizační hledisko, patrně platné pro znaky *kandži*, resp. *hanzi*, uvnitř grafických polí, totiž pro ostatní znaky kvůli jejich odlišnému vývoji nemusí platit.

Ve výsledku je tedy otázkou, který z těchto argumentů bude mít na projev Menzerath-Altmanova zákona větší vliv. Na základě výše uvedeného a výsledků z předešlého experimentu s japonskou povídkou předpokládám, že na této jazykové úrovni bude u japonských textů vztah vyjádřený MAZ potvrzen kladnými hodnotami parametru *b*, ale ve výsledku bude koeficient determinace R^2 nízký a MAZ se neprojeví zdaleka v tak vysoké míře, jako v případě textů čínských.

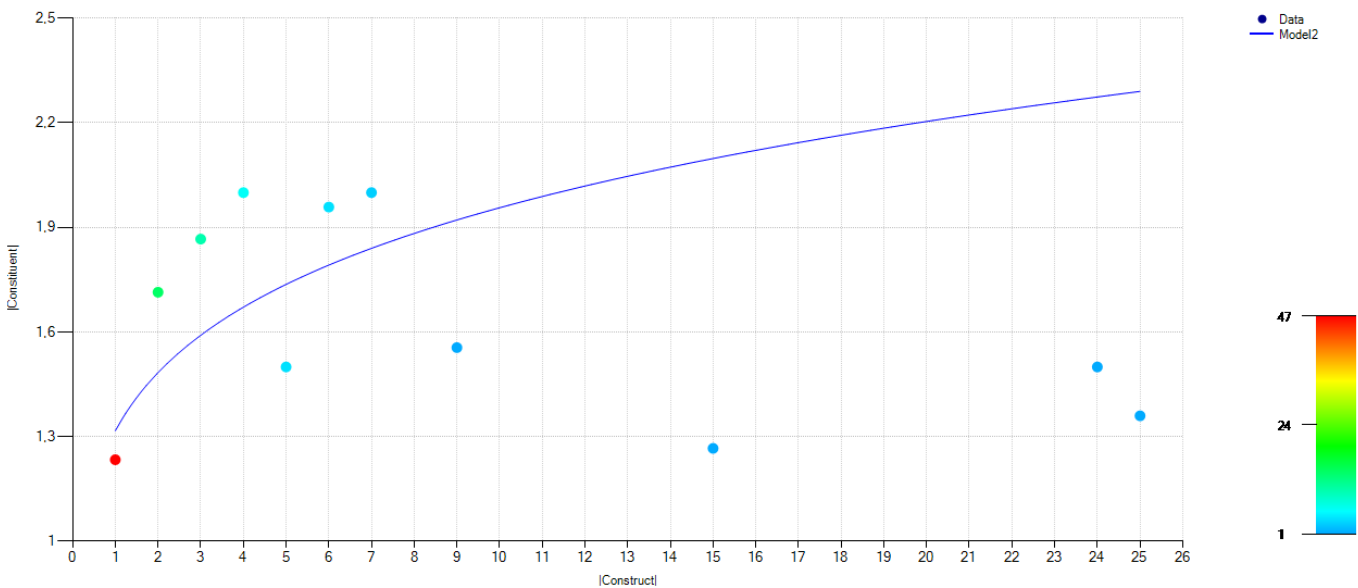
⁵⁹ Viz (Motalová a Spáčilová 2013, s. 107).

4. Výsledky analýz jednotlivých textů

4.1. Analyzovaný text č. 1 – povídka

Níže jsou uvedeny výsledky analýzy 1. analyzovaného textu pro jazykové úrovně U_1 až U_4 .

4.1.1. Povídka – U_1 – odstavce-souvětí



Graf 1: povídka – U_1 – odstavce-souvětí

$$A = 1,3171581582$$

$$b = -0,1718294869$$

$$R^2 = 0,4874410620$$

Tabulka 4: A, b, R^2 (povídka – U_1 – odstavec-souvětí)

x	z	y
1	47	1,23404255319149
2	14	1,71428571428571
3	10	1,86666666666667
4	6	2
5	4	1,5
6	4	1,95833333333333

x	z	y
7	3	2
9	1	1,55555555555556
15	1	1,26666666666667
24	1	1,5
25	1	1,36

Tabulka 5: Hodnoty x, z, y (povídka – U_1 – odstavec-souvětí)

V tabulce 5 jsou uvedeny hodnoty získané kvantifikací analyzovaného textu č. 1 – *povídky* na jazykové úrovni U_1 . Konstrukt x zde představuje délku odstavců, měřených v počtu souvětí. Hodnoty z představují frekvenci výskytu konstruktů o dané délce. Konstituent y pak představuje průměrnou délku souvětí, měřených v počtu bezprostředně nižších jazykových jednotek, tj. v mezičáří.

Délka konstruktů vyjádřených v počtu konstituentů na této jazykové úrovni je od 1 do 25. Průměrná délka konstituentů je v rozmezí $\langle 1,234; 2 \rangle$. Přesné hodnoty parametru A , parametru b a koeficientu determinace R^2 pro jazykovou úroveň U_1 jsou uvedeny v tabulce 4.

Na jazykové úrovni U_1 nelze říci, že s rostoucí délkou odstavců klesá průměrná délka souvětí. Už z tabulky je patrné, že předpoklad MAZ (čím delší konstrukt, tím kratší konstituent) na této jazykové úrovni nebude platit.

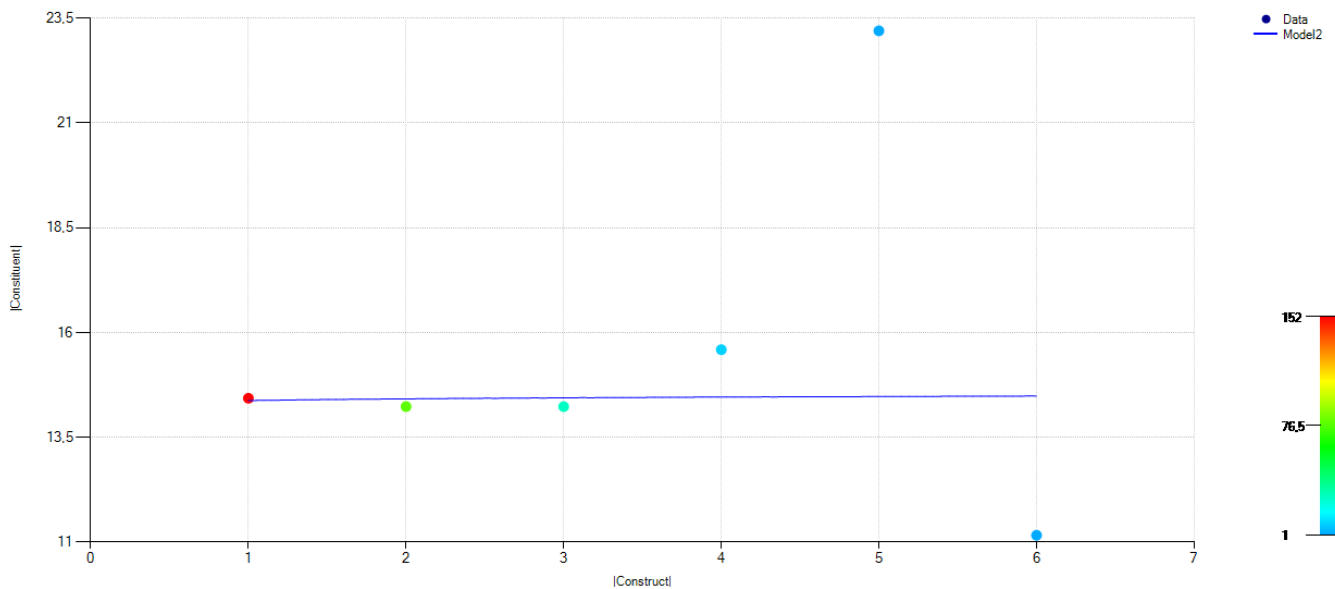
To potvrzuje i vypočtený parametr b , jehož hodnota je záporné číslo. Funkce je tudíž stoupající a konkávní a předpoklad platnosti MAZ není na této úrovni naplněn. Je tedy zřejmé, že MAZ se na této jazykové úrovni neprojevil. Hodnota koeficientu determinace R^2 je přibližně 0,487.

V případě povídky jakožto 1. analyzovaného textu tedy nelze konstatovat, že čím více odstavce obsahuje souvětí, tím méně mezičáří obsahují v průměru daná souvětí.

Jedním z možných důvodů takového výsledku může být fakt, že se v textu vyskytuje poměrně velké množství dialogů a každé ukončení přímé řeči bylo zároveň doprovázeno ukončením odstavce. To vedlo k tomu, že některé odstavce byly v textu často velice krátké, tj. o velikosti jednoho souvětí či dokonce jednoho mezičáří. Takto krátké odstavce se v ostatních výběrových souborech nevyskytovaly.⁶⁰ Dalším možným důvodem může být i zavedení striktně graficky segmentované jednotky mezičáří, která je graficky velice snadno identifikovatelná, ale může ignorovat jiné (např. sémantické) aspekty textu. Výsledek je však v souladu s hypotézou (viz 3.1. U_1 – odstavce-souvětí).

⁶⁰ Po úpravě definice v případě básnického textu.

4.1.2. Povídka – U₂ – souvětí-mezičáří



Graf 2: povídka – U₂ – souvětí-mezičáří

$$A = 14,3790897404$$

$$b = -0,0041691435$$

$$R^2 = 0,0026799784$$

Tabulka 6: A , b , R^2 (povídka – U₂ – souvětí-mezičáří)

x	z	y
1	152	14,4342105263158
2	78	14,2371794871795
3	27	14,2345679012346
4	8	15,59375
5	1	23,2
6	1	11,1666666666667

Tabulka 7: Hodnoty x , z , y (povídka – U₂ – souvětí-mezičáří)

V tabulce 7 jsou uvedeny hodnoty získané kvantifikací 1. analyzovaného textu (povídky) na jazykové úrovni U₂. Konstrukt x zde představuje délku souvětí, měřených v počtu mezičáří. Hodnoty z představují frekvenci výskytu konstruktů o dané délce.

Konstituent *y* pak představuje průměrnou délku mezičárí, měřených v počtu bezprostředně nižších jazykových jednotek, tj. ve znacích.

Délka konstruktů vyjádřených v počtu konstituentů na této jazykové úrovni je od 1 do 6. Průměrná délka konstituentů je v rozmezí $\langle 11,167; 23,2 \rangle$. Přesné hodnoty parametru *A*, parametru *b* a koeficientu determinace R^2 pro jazykovou úroveň U_2 jsou uvedeny v tabulce 6.

Tendence naznačená jazykovou úrovní souvětí-mezičárí není o mnoho lepší. Průměrné délky konstituentů pro tři nejfrekventovanější velikosti konstruktů jsou přibližně stejné a poté dokonce stoupají. Parametr *b* je číslo záporné a předpoklad pro platnost MAZ tedy nebyl splněn; hodnota koeficientu determinace R^2 je pouhých 0,003.

Je však třeba podotknout, že v této tabulce není mnoho bodů *x*, takže statistická relevance nemusí být příliš vysoká. Navíc pátý a šestý bod *x* mají pouze jeden jediný výskyt.

Je tedy zřejmé, že u povídky na jazykové úrovni souvětí-mezičárí MAZ neplatí a model nelze považovat za dobře sedící a nelze tedy říci, že čím více souvětí obsahuje mezičárí, tím méně znaků obsahují v průměru mezičárí.

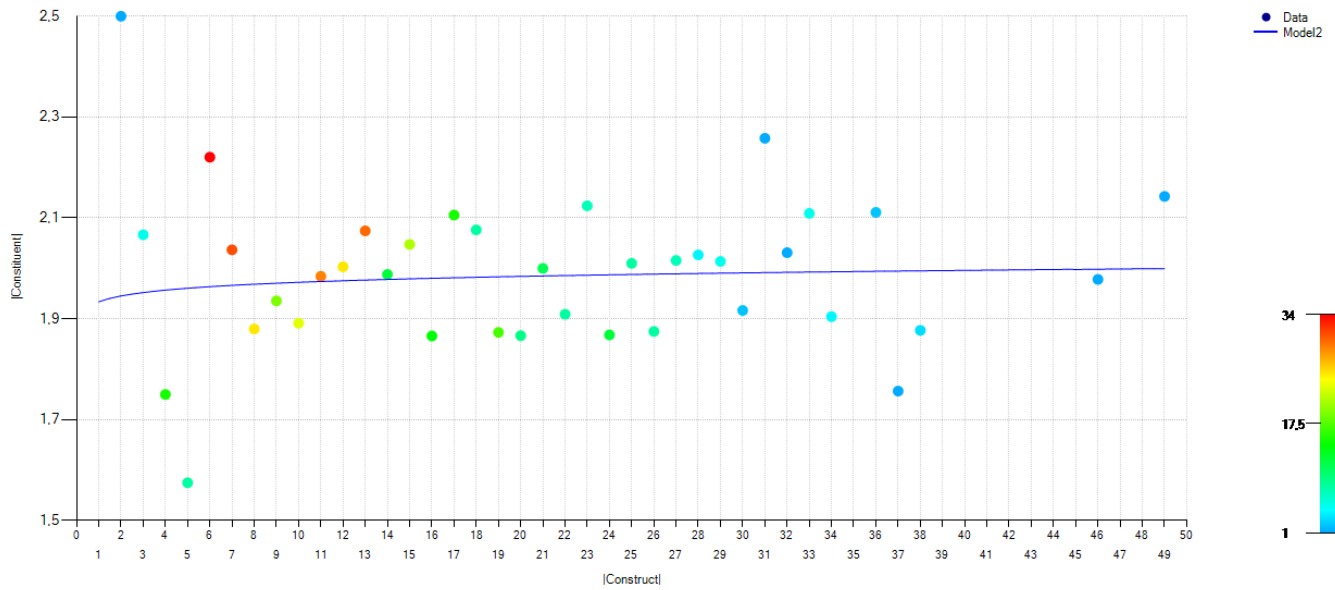
Výsledky na této úrovni tedy neodpovídají hypotéze; původní předpoklad byl, že na jazykové úrovni U_2 by se vztah popsaný Menzerath-Altmanovým zákonem mohl potvrdit.

Projev daného vztahu mohl být narušen užitím graficky definované jednotky mezičárí, která místy ignoruje sémantická či syntaktická hlediska textu. Rovněž užití jednotky znaku jako bezprostředně nižší jazykové jednotky nemusí být ideální: Je možné, že mezi jednotkami mezičárí a znak by bylo vhodné zavést další jednotky, které by figurovaly namísto velice obtížně definovatelné jednotky (ortografického) slova, nebo by zastávaly jeho funkci.⁶¹

Také zahrnutí interpunkčních znamének do jazykové jednotky znaku mohlo ovlivnit výsledky analýzy na této jazykové hladině, jelikož interpunkční znaménka do jisté míry ovlivňují délku mezičárí (tj. konstituentu).

⁶¹ Malý a pouze doplňující experiment s jinými jazykovými jednotkami byl v rámci této práce proveden, viz 4.4. Doplňující experimenty s analyzovaným textem č. 3.).

4.1.3. Povídka – U₃ – mezičárí-znaky



Graf 3: povídka – U₃ – mezičárí-znaky

$$A = 1,9337840531$$

$$b = -0,0085903054$$

$$R^2 = 0,0060178368$$

Tabulka 8: A, b, R² (povídka – U₃ – mezičárí-znaky)

x	z	y
2	1	2,5
3	5	2,06666666666667
4	15	1,75
5	8	1,575
6	34	2,22058823529412
7	31	2,036866359447
8	25	1,88
9	19	1,93567251461988
10	23	1,89130434782609
11	29	1,98432601880878
12	25	2,00333333333333
13	30	2,07435897435897

x	z	y
14	12	1,98809523809524
15	21	2,04761904761905
16	14	1,86607142857143
17	15	2,10588235294118
18	8	2,07638888888889
19	17	1,87306501547988
20	9	1,86666666666667
21	11	2
22	8	1,90909090909091
23	7	2,12422360248447
24	12	1,86805555555556
25	8	2,01

<i>x</i>	<i>z</i>	<i>y</i>
26	8	1,875
27	7	2,01587301587302
28	4	2,02678571428571
29	5	2,01379310344828
30	2	1,91666666666667
31	1	2,25806451612903
32	1	2,03125

<i>x</i>	<i>z</i>	<i>y</i>
33	5	2,10909090909091
34	4	1,90441176470588
36	2	2,11111111111111
37	1	1,75675675675676
38	3	1,87719298245614
46	1	1,97826086956522
49	1	2,14285714285714

Tabulka 9: Hodnoty *x*, *z*, *y* (povídka – U_3 – mezičárí-znaky)

V tabulce 9 jsou uvedeny hodnoty získané kvantifikací 1. analyzovaného textu (povídky) na jazykové úrovni U_3 . Konstrukt *x* zde představuje délku mezičárí, měřených v počtu znaků. Hodnoty *z* představují frekvenci výskytu konstruktů o dané délce. Konstituent *y* pak představuje průměrnou délku znaků, měřených v počtu bezprostředně nižších jazykových jednotek, tj. v prvcích.

Délka konstruktů vyjádřených v počtu konstituentů na této jazykové úrovni je od 2 do 49. Průměrná délka konstituentů je v rozmezí $(1,575; 2,5)$. Přesné hodnoty parametru *A*, parametru *b* a koeficientu determinace R^2 pro jazykovou úroveň U_3 jsou uvedeny v tabulce 8.

Tabulka 9 ukazuje, že průměrné délky znaků sice zpočátku s narůstající délkou mezičárí klesají, ovšem již od mezičárí o délce 6 znaků a dále je tato tendence mnohokrát porušena.

I na jazykové úrovni mezičárí-znaky lze tak pozorovat více méně rovný tvar regresní křivky. Parametr *b* je opět záporné číslo a tvar proložené křivky je tedy mírně rostoucí. Rovněž i koeficient determinace R^2 vykazuje kompatibilitu s teoretickým modelem pouze zhruba 0,006. Všechny body *x* jsou v Graf 3: povídka – U_3 – mezičárí-znaky rozprostřeny poměrně nahodile, a přestože při vizuálním zhodnocení s pomocí teplotní mapy se zdá, že body *x* s větší frekvencí se nacházejí spíše v levém horním kvadrantu, parametr *b* vychází záporně a proložená křivka je tedy mírně stoupající.

Tento negativní výsledek je však v souladu s hypotézou. Možné důvody pro neplatnost MAZ na této jazykové úrovni jsou podrobněji rozvedeny v kap. 3.3. U_3 – mezičárí-znaky Ve stručnosti to může být právě implementace samotné jednotky mezičárí,

kteřá je založena na čistě grafickém principu. Dále pravděpodobně příliš velký skok mezi jazykovou jednotkou mezičárí a jazykovou jednotkou znaku, mezi nimiž mohou chybět ještě nějaké další mezijednotky. Rovněž fakt, že mezičárí je jazyková jednotka s délkou proměnlivou (v závislosti na vůli autora), zatímco jeho konstituenty, tj. znaky, jsou jednotky s délkou v počtu prvků fixní. A nakonec také malá variabilita znaků, co se týče počtu v nich obsažených prvků, což by mohlo narušit vztah nepřímé úměrnosti mezi konstrukty a konstituenty.

Co se týče posledního argumentu, pro ověření toho, zda by mohl platit pro japonštinu v případě povídky, je podíl znaků dle počtu v nich obsažených prvků uveden v následující tabulce.

celkový počet znaků: 6204			celkový počet unikátních znaků: 629		
data pro všechny znaky v textu			data pro unikátní znaky v textu		
počet prvků ve znaku	četnost znaků o daném počtu prvků	podíl na celkovém počtu všech znaků (v %)	počet prvků ve znaku	četnost znaků o daném počtu prvků	podíl na celkovém počtu unikátních znaků (v %)
1	3235	52,144	1	163	25,914
2	1273	20,519	2	123	19,555
3	958	15,442	3	139	22,099
4	344	5,545	4	85	13,514
5	224	3,611	5	47	7,472
6	101	1,628	6	39	6,2
7	25	0,403	7	19	3,021
8	29	0,467	8	10	1,59
9	14	0,226	9	3	0,477
10	0	0	10	0	0
11	0	0	11	0	0
12	1	0,016	12	1	0,159

Tabulka 10: povídka – U₃ – podíl znaků dle počtu prvků

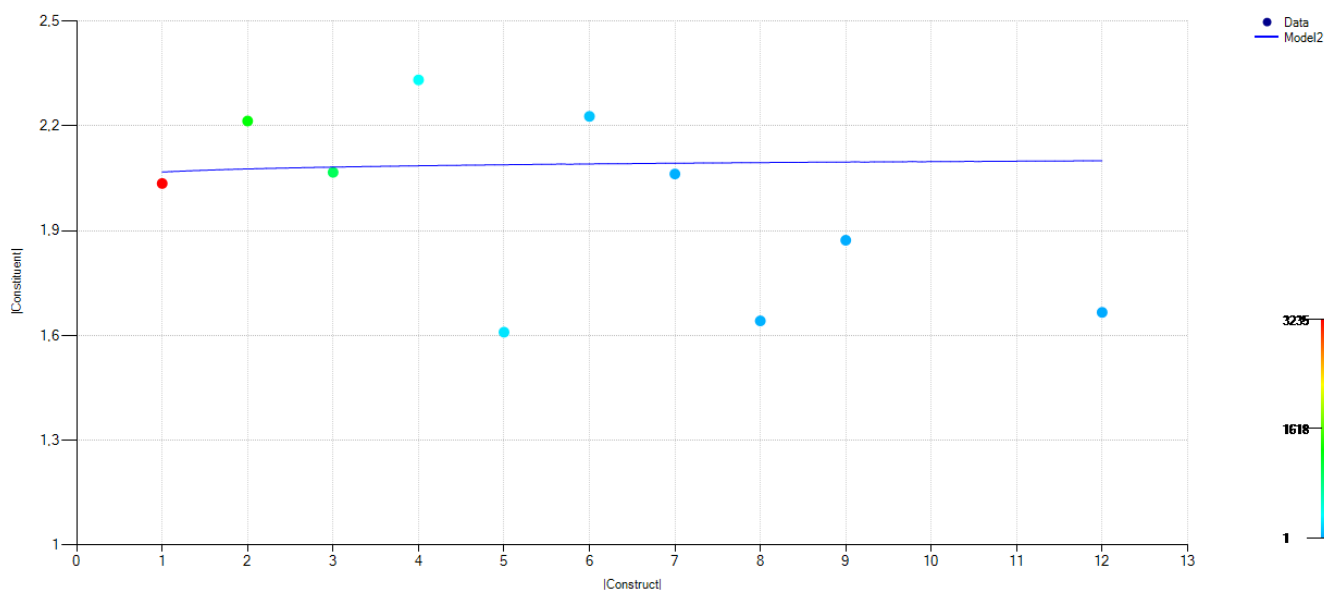
Jak je patrné z procentuálního zastoupení znaků, reálná variabilita znaků dle počtu jejich prvků je velmi nízká. Jednoprvkové znaky tvoří více než celou polovinu všech znaků

obsažených ve vybraném textu (a více než jednu čtvrtinu, pokud se zohlední pouze unikátní znaky bez opakování), přičemž jednoprvkové až tříprvkové znaky pak tvoří drtivou většinu všech znaků v textu – přes 88 % (a více než dvě třetiny, pokud se zohlední pouze unikátní znaky bez opakování). Takto výrazně malá variabilita znaků, co se týče počtu v nich obsažených prvků, by rovněž mohla být jednou z příčin neplatnosti vztahu mezi délkou konstruktů a jejich konstituentů.⁶²

Nakonec by mohlo být důvodem pro negativní výsledek i zahrnutí interpunkčních znamének do jednotky znaku. Ta interpunkční znaménka, která neukončují mezičárí, do jisté míry délku mezičárí zvyšují.

U jazykové úrovně mezičárí-znaky tedy v případě povídky neplatí tvrzení, že čím více mezičárí obsahuje znaků, tím méně prvků obsahují v průměru znaky.

4.1.4. Povídka – U₄ – znaky-prvky



Graf 4: povídka – U₄ – znaky-prvky

⁶² Nejfrekventovanějšími jednoprvkovými až tříprvkovými znaky jsou znaky *kany* a interpunkční znaménka.

$$A = 2,0682800446$$

$$b = -0,0062272080$$

$$R^2 = 0,0028836487$$

Tabulka 11: A, b, R^2 (povídka – U_4 – znaky-prvky)

x	z	y
1	3235	2,0312596719282
2	1273	1,95400516795866
3	958	1,8794916739702
4	344	2,25531914893617
5	224	1,60982142857143
6	101	2,22772277227723
7	25	2,06285714285714
8	29	1,64224137931034
9	14	1,87301587301587
12	1	1,66666666666667

Tabulka 12: Hodnoty x, z, y (povídka – U_4 – znaky-prvky)

V tabulce 12 jsou uvedeny hodnoty získané kvantifikací 1. analyzovaného textu (povídky) na jazykové úrovni U_4 . Konstrukt x zde představuje délku znaků, měřených v počtu prvků. Hodnoty z představují frekvenci výskytu konstruktů o dané délce. Konstituent y pak představuje průměrnou délku prvků, měřených v počtu bezprostředně nižších jazykových jednotek, tj. v tazích.

Délka konstruktů vyjádřených v počtu konstituentů na této jazykové úrovni je od 1 do 12. Průměrná délka konstituentů je v rozmezí $\langle 1,61; 2,255 \rangle$. Přesné hodnoty parametru A , parametru b a koeficientu determinace R^2 pro jazykovou úroveň U_4 jsou uvedeny v tabulce 11.

I na této nejnižší jazykové úrovni jsou výsledky podobné. Z tabulky 12 není patrná žádná závislost délky konstituentů na délce konstruktů. Parametr b je znovu záporné číslo, tvar proložené křivky je skoro rovný (mírně konkávní) a předpoklad pro platnost MAZ tedy nebyl naplněn. Navíc i koeficient determinace R^2 ukazuje pouze 0,003 a hodnoty x jsou rozprostřeny značně nerovnoměrně. Ani na této jazykové úrovni se tedy nenaplnily předpoklady pro platnost Menzerath-Altmanova zákona a v případě

povídky není možné říci, že čím více znaky obsahují prvků, tím méně tahů obsahují v průměru prvky.

Tento výsledek již nebyl natolik nepřekvapující jako některé výsledky na vyšších jazykových hladinách. Velkou otázkou je, z jakého důvodu se ani na této jazykové úrovni vztah formulovaný MAZ nejen neprojevil a parametr b je číslo záporné, zatímco se na podobné jazykové úrovni při analýze čínských textů⁶³ MAZ naopak potvrdil jako nadmíru sedící model s většinou hodnot koeficientu determinace R^2 nad 0,9.

Patrně hlavním problémem v případě japonštiny může být japonské užívání dvou „abeced“, které se vyvinuly z komplexnějších znaků *kandži*, ale které se od nich dnes v mnoha aspektech odlišují. Jak bylo uvedeno v kap. 3.4. U4 – znaky-prvky., v případě japonštiny zde proti sobě stojí dva problémy. Na jednu stranu se všechny japonské znaky zapisují do striktně vymezených čtvercových grafických polí, která omezují množství prvků, jež se mohou dovnitř polí vměstnat. Na straně druhé pak stojí fakt, že znaky *kandži*, znaky *kana* aj. možná nejsou ekvivalentní, tj. mohly by tvořit hierarchicky odlišné, či paralelní jazykové jednotky. V tom případě by bylo vhodné tyto znaky segmentovat nějakým odlišným způsobem; ekonomizační hledisko, pravděpodobně platné pro znaky *kandži*, totiž pro ostatní znaky nemusí platit.

Ze získaných výsledků se zdá, že druhý argument může mít převahu nad prvním a přestože jsou nejen znaky *kandži*, ale i všechny ostatní znaky řízeny a limitovány stejnými pravidly, co se týče široce přijímaného kritéria grafických polí, patrně by neměly být segmentovány stejným způsobem, přestože se nezdá příliš logické ani snadné zacházet s nimi při grafické segmentaci jinak než se znaky *kandži*.⁶⁴ Na závěry je však zatím příliš brzy; je potřeba získat a porovnat výsledky analýz jazykové úrovně U₄ z většího množství japonských textů.

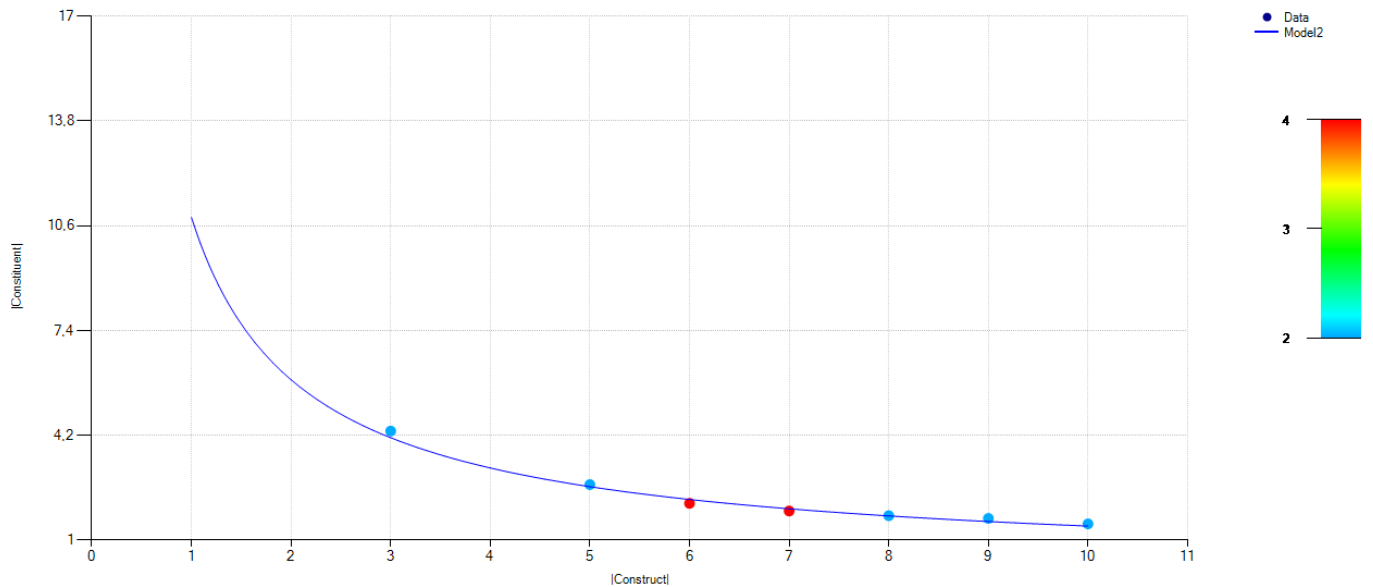
4.2. Analyzovaný text č. 2 – poezie

Níže jsou uvedeny výsledky analýzy 2. analyzovaného textu (poezie) pro jazykové úrovně U₁ až U₄.

⁶³ Viz (Motalová 2013).

⁶⁴ O problémech s odlišnou segmentací *kandži* a *kany* viz 2.5.5. Znaky

4.2.1. Poezie – U₁ – odstavce-souvětí



Graf 5: poezie – U₁ – odstavce-souvětí

$$A = 10,8574257162$$

$$b = 0,8796006386$$

$$R^2 = 0,9790415441$$

Tabulka 13: A, b, R^2 (poezie – U₁ – odstavce-souvětí)

x	z	y
3	2	4,33333333333333
5	2	2,7
6	4	2,125
7	4	1,89285714285714
8	2	1,75
9	2	1,66666666666667
10	2	1,5

Tabulka 14: Hodnoty x, z, y (poezie – U₁ – odstavce-souvětí)

V tabulce 14 jsou uvedeny hodnoty získané kvantifikací 2. analyzovaného textu (poezie) na jazykové úrovni U₁. Konstrukt x zde představuje délku odstavců, měřených v počtu souvětí. Hodnoty z představují frekvenci výskytu konstruktů o dané délce.

Konstituent y pak představuje průměrnou délku souvětí, měřených v počtu bezprostředně nižších jazykových jednotek, tj. v mezičáří.

Délka konstruktů vyjádřených v počtu konstituentů na této jazykové úrovni je od 3 do 10. Průměrná délka konstituentů je v rozmezí $\langle 1,5; 4,333 \rangle$. Přesné hodnoty parametru A , parametru b a koeficientu determinace R^2 pro jazykovou úroveň U_1 jsou uvedeny v tabulce 13.

Z tabulky 14 lze vyčíst stabilně klesající průměrné délky konstituentů s rostoucí délkou konstruktů, a to bez jediného vychýlení. Je pravda, že celkový počet bodů x není příliš velký a bylo by tudíž vhodné provést výpočty i na dalších básnických textech pro zjištění toho, zda lze tento výsledek považovat za vypovídající; na tomto vybraném textu se však potvrdila platnost Menzerath-Altmannova zákona na jazykové úrovni U_1 . Parametr b je číslo kladné a tvar proložené křivky je tedy konvexní a klesající. Koeficient determinace R^2 ukazuje nadměrně vysokou hodnotu 0,979, což značí extrémně vysokou shodu s matematickým modelem.

Ze získaných výsledků lze tedy konstatovat, že na jazykové úrovni U_1 u básnického textu opravdu platí vztah mezi délkou konstruktů a jejich konstituentů, tedy čím více odstavce obsahují souvětí, tím méně mezičáří obsahují v průměru daná souvětí.

Na jazykové úrovni odstavce-souvětí se tak patrně potvrzuje hypotéza (viz 3.1. U_1 – odstavce-souvětí), že u básnického textu může být výsledek analýzy poněkud odlišný od výsledků na stejné jazykové úrovni ostatních analyzovaných textů. Zatímco segmentace odstavců u povídky a seminární práce probíhala přesně v souladu s definicí odstavce (viz 2.5.2. Odstavce), v případě básně byla za jeden odstavec považována celá strofa. Nejednalo se tedy o segmentaci striktně dle uvedené definice odstavce, ačkoliv se i tento způsob segmentace stále přidržuje grafického principu a, především, je s ohledem na styl textu logičtější.

Tento básnický text měl, na rozdíl od ostatních dvou analyzovaných textů, značně omezenou minimální i maximální délku odstavců (konstruktů), a tedy i počet souvětí (konstituentů) a průměrnou délku mezičáří, aby nebyla příliš narušena grafická podoba odstavců (strof). Přestože se nejedná o původem japonský text, ale o překlad starší anglojazyčné básně do japonštiny, překladatel se rozhodl zachovat původní grafický formát básně a nemohl tak příliš ovlivňovat délku odstavců (strof). Navíc jsou všechny

strofy v básni v počtu znaků podobně dlouhé.⁶⁵ A právě toto může být významný důvod platnosti MAZ na této jazykové úrovni. Odstavce jsou graficky ohraničeny svou délkou, což znamená, že čím více bude takto omezený odstavec obsahovat souvětí, tím méně by daná souvětí měla být schopna pojmut mezičárí, aby ve výsledku příliš nepřekračovala omezenou maximální délku odstavce. Platí tedy, že čím delší je konstrukt, tím kratší jsou (resp. musí být) jeho konstituenty, tj. v daném případě „čím delší je odstavec v počtu souvětí, tím kratší jsou tato souvětí v průměrném počtu mezičárí. Jak již bylo uvedeno v kap. 3.1. U₁ – odstavce-souvětí, do jisté míry to lze přirovnat k pozorování o restriktivním čtvercovém grafickém poli, do kterého se zapisují jednotlivé znaky. Při analýze čínských textů⁶⁶ se ukázalo, že čím více má znak (uvnitř omezeného grafického pole) prvků, tím méně tahů mají (resp. musejí mít) v průměru tyto prvky, aby stále byly schopny se vměstnat do prostoru těchto omezených grafických polí, přičemž hodnoty koeficientu determinace R^2 v případě daných čínských textů byly obdobně vysoké, tj. většina nad 0,9.⁶⁷

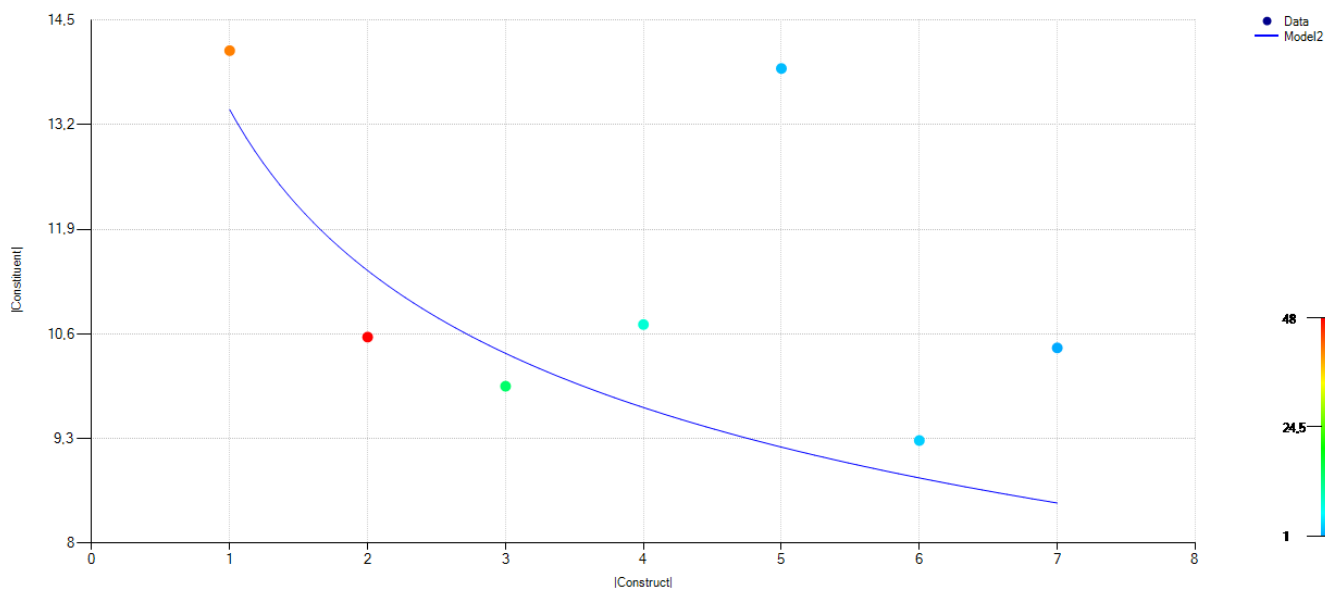
Bylo by vhodné provést analýzy dalších básnických textů (západního stylu i japonského stylu) pro ověření, zda výše uvedené platí univerzálně i na dalších textech a zda v otázce platnosti MAZ opravdu hrají významnou roli restriktivní omezení grafické podoby textu.

⁶⁵ Počet znaků v jednotlivých odstavcích se pohyboval v rozmezí (125; 184), průměr 146, medián 142.

⁶⁶ (Motalová 2013)

⁶⁷ V případě japonských vybraných textů se sice na úrovni U₄ – znaky-prvky při daném způsobu segmentace MAZ neukazuje jako příliš vhodný model (viz příslušné kapitoly U₄ – znaky-prvky), kvůli užívání více typů písma se tato jazyková úroveň značně odlišuje od stejné jazykové úrovně v čínštině. Proto je v tomto případě na místě pro srovnání uvést výsledky analýzy textů čínských namísto textů japonských, byť se v ostatních ohledech jedná o jazyky velice odlišné.

4.2.2. Poezie – U₂ – souvětí-mezičáří



Graf 6: poezie – U₂ – souvětí-mezičáří

$$A = 13,3889029977$$

$$b = 0,2333956687$$

$$R^2 = 0,6614184918$$

Tabulka 15: A , b , R^2 (poezie – U₂ – souvětí-mezičáří)

x	z	y
1	41	14,1219512195122
2	48	10,5625
3	14	9,95238095238095
4	8	10,71875
5	2	13,9
6	3	9,27777777777778
7	1	10,4285714285714

Tabulka 16: Hodnoty x , z , y (poezie – U₂ – souvětí-mezičáří)

V tabulce 16 jsou uvedeny hodnoty získané kvantifikací 2. analyzovaného textu (poezie) na jazykové úrovni U₂. Konstrukt x zde představuje délku souvětí, měřených v počtu mezičáří. Hodnoty z představují frekvenci výskytu konstruktů o dané délce.

Konstituent y pak představuje průměrnou délku mezičárí, měřených v počtu bezprostředně nižších jazykových jednotek, tj. ve znacích.

Délka konstruktů vyjádřených v počtu konstituentů na této jazykové úrovni je od 1 do 7. Průměrná délka konstituentů je v rozmezí (9,277; 14,12). Přesné hodnoty parametru A , parametru b a koeficientu determinace R^2 pro jazykovou úroveň U_2 jsou uvedeny v tabulce 15.

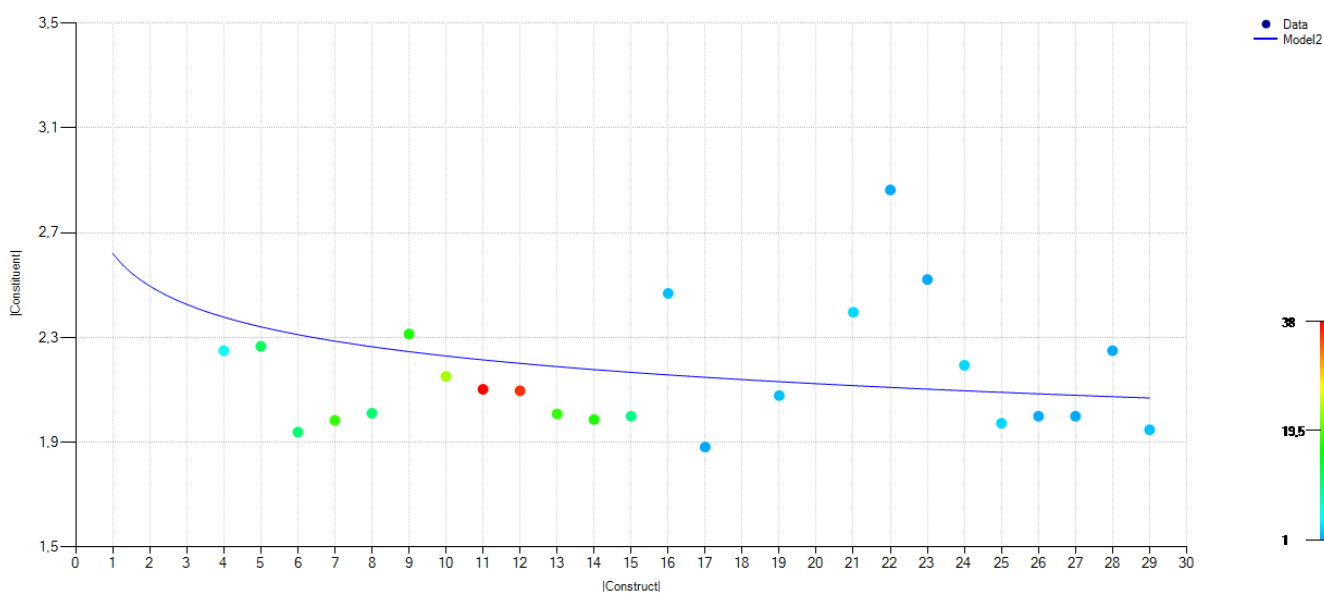
Průměrné délky konstituentů nejfrekventovanějších konstruktů zpočátku dodržují tendenci poklesu se vzrůstající délkou konstruktů, avšak se snižující se frekvencí konstruktů se tato tendence porušuje a opět dochází k nárůstu. Jelikož se však jedná o konstrukty s relativně nízkou četností, ve výsledku tyto odchylky nemají příliš silný vliv a výsledný tvar křivky je konvexní a klesající. Parametr b je kladné číslo, což vyhovuje předpokladu pro platnost MAZ a koeficient determinace R^2 je 0,661.

S takovýmto výsledkem není snadné zcela jednoznačně říci, zda můžeme považovat MAZ jako platný pro jazykovou úroveň souvětí-mezičárí. Domnívám se však, že vzhledem k tvaru proložené křivky, poměrně vysokému koeficientu determinace a nízkému počtu konstruktů, u kterých se tendence klesajících délek konstituentů porušila, lze říci, že u tohoto básnického textu platí tvrzení, že čím více souvětí obsahují mezičárí, tím méně znaků obsahují v průměru mezičárí. Je však žádoucí analýza dalších výběrových souborů napsaných v tomto stylu, než bude možné dané tvrzení zobecnit na většinu japonských básnických textů.

Co se týče důvodu pravděpodobné platnosti Menzerath-Altmanova zákona (pouze u básnického textu) na této jazykové úrovni, lze uvažovat o tom, že maximální délka souvětí je poměrně výrazně ovlivněna omezením maximální délky odstavce jakožto vyšší jazykové jednotky, která sice nevstupuje do výpočtů, avšak stále restriktivně působí na délky či počet souvětí, která tak nemohou být neomezeně dlouhá. Z toho důvodu by tedy opět mělo platit, že čím více budou souvětí (konstrukty) obsahovat mezičárí (konstituentů), tím kratší budou mezičárí v průměrném počtu znaků, aby se příliš nenarušila celková délka odstavce (tj. strofy). Jelikož však není jazyková jednotka souvětí tak výrazně graficky omezena jako jednotka odstavce a zároveň počet znaků obsažených v jazykové jednotce mezičárí může být o něco variabilnější než počet mezičárí (jakožto vyšší a větší jednotky než znaky) uvnitř souvětí, neprojevil se na této jazykové úrovni vztah mezi konstrukty a konstituenty tak výrazně, jako se projevil na vyšší úrovni U_1 .

MAZ se zároveň u básnického textu projevil výrazněji než u ostatních dvou analyzovaných textů, tj. povídky a seminární práce. Zatím je třeba dalších výzkumů v této oblasti, ale lze uvažovat o tom, že grafická omezení textu opravdu mají významný vliv na projev MAZ. V případě budoucích experimentů s dalšími japonskými texty bude vhodné toto pozorování podrobit dalšímu ověření.

4.2.3. Poezie – U₃ – mezičárí-znaky



Graf 7: poezie – U₃ – mezičárí-znaky

$$A = 2,6212939040$$

$$b = 0,0701825858$$

$$R^2 = 0,1017813902$$

Tabulka 17: A, b, R² (poezie – U₃ – mezičárí-znaky)

x	z	y
3	8	3,54166666666667
4	5	2,25
5	12	2,26666666666667
6	11	1,93939393939394
7	18	1,98412698412698
8	11	2,01136363636364
9	17	2,31372549019608

x	z	y
10	23	2,15217391304348
11	38	2,10287081339713
12	36	2,09722222222222
13	18	2,00854700854701
14	17	1,98739495798319
15	10	2
16	2	2,46875

<i>x</i>	<i>z</i>	<i>y</i>
17	1	1,88235294117647
19	2	2,07894736842105
21	3	2,3968253968254
22	1	2,86363636363636
23	1	2,52173913043478
24	3	2,19444444444444

<i>x</i>	<i>z</i>	<i>y</i>
25	3	1,97333333333333
26	1	2
27	1	2
28	1	2,25
29	2	1,94827586206897

Tabulka 18: Hodnoty *x*, *z*, *y* (poezie – U_3 – mezičáří-znaky)

V tabulce 18 jsou uvedeny hodnoty získané kvantifikací 2. analyzovaného textu (poezie) na jazykové úrovni U_3 . Konstrukt *x* zde představuje délku mezičáří, měřených v počtu znaků. Hodnoty *z* představují frekvenci výskytu konstruktů o dané délce. Konstituent *y* pak představuje průměrnou délku znaků, měřených v počtu bezprostředně nižších jazykových jednotek, tj. v prvcích.

Délka konstruktů vyjádřených v počtu konstituentů na této jazykové úrovni je od 3 do 29. Průměrná délka konstituentů je v rozmezí $\langle 1,882; 3,541 \rangle$. Přesné hodnoty parametru *A*, parametru *b* a koeficientu determinace R^2 pro jazykovou úroveň U_3 jsou uvedeny v tabulce 17.

Z tabulky 18 je patrné, že délky konstituentů ne vždy klesají společně s nárůstem délek konstruktů a body *x* jsou rozprostřeny spíše lineárně. Parametr *b* je na této jazykové úrovni kladné číslo, ale hodnota koeficientu determinace R^2 je pouhých 0,102. Na jazykové úrovni U_3 tedy v případě básnického textu sice lze říci, že MAZ dle hodnoty parametru *b* platí, avšak regresní křivka nesedí na daná empirická pozorování.

Jelikož na jazykové úrovni mezičáří-znaky již u tohoto textu neovlivňuje omezená délka odstavců délku konstituentů či ještě nižších jednotek, jak se pravděpodobně dělo na vyšších jazykových úrovních, kde se vztah MAZ ukázal jako platný a hodnota R^2 byla poměrně vysoká, výsledek je podobně negativní jako na téže jazykové úrovni ostatních analyzovaných textů. Možné důvody jsou tedy rovněž totožné jako v případě analýzy povídky a seminární práce a odpovídají hypotéze (viz 3.3. U_3 – mezičáří-znaky): Ve stručnosti to může být právě implementace samotné jednotky mezičáří, která je založena na čistě grafickém principu. Dále pravděpodobně příliš velký skok mezi jazykovou jednotkou mezičáří a jazykovou jednotkou znaku, mezi nimiž mohou chybět ještě další mezijednotky. Rovněž fakt, že mezičáří je jazyková jednotka s délkou

proměnlivou (v závislosti na vůli autora), zatímco jeho konstituenty, tj. znaky, jsou jednotkami s délkou v počtu prvků fixní. A nakonec také malá variabilita znaků, což se týče počtu v nich obsažených prvků, což by mohlo narušit vztah nepřímé úměrnosti mezi konstrukty a konstituenty.

Co se týče posledního argumentu, pro ověření toho, zda by mohl platit pro japonštinu v případě básnického textu, je podíl znaků dle počtu jejich prvků uveden v následující tabulce.

celkový počet znaků: 2721			celkový počet unikátních znaků: 484		
data pro všechny znaky v textu			data pro unikátní znaky v textu		
počet prvků ve znaku	četnost znaků o daném počtu prvků	podíl na celkovém počtu všech znaků (v %)	počet prvků ve znaku	četnost znaků o daném počtu prvků	podíl na celkovém počtu unikátních znaků (v %)
1	1365	50,165	1	122	25,207
2	585	21,499	2	91	18,802
3	375	13,782	3	101	20,868
4	134	4,925	4	59	12,19
5	133	4,888	5	42	8,678
6	65	2,389	6	30	6,198
7	36	1,323	7	22	4,545
8	7	0,257	8	6	1,24
9	16	0,588	9	6	1,24
10	4	0,147	10	4	0,826
11	1	0,037	11	1	0,207

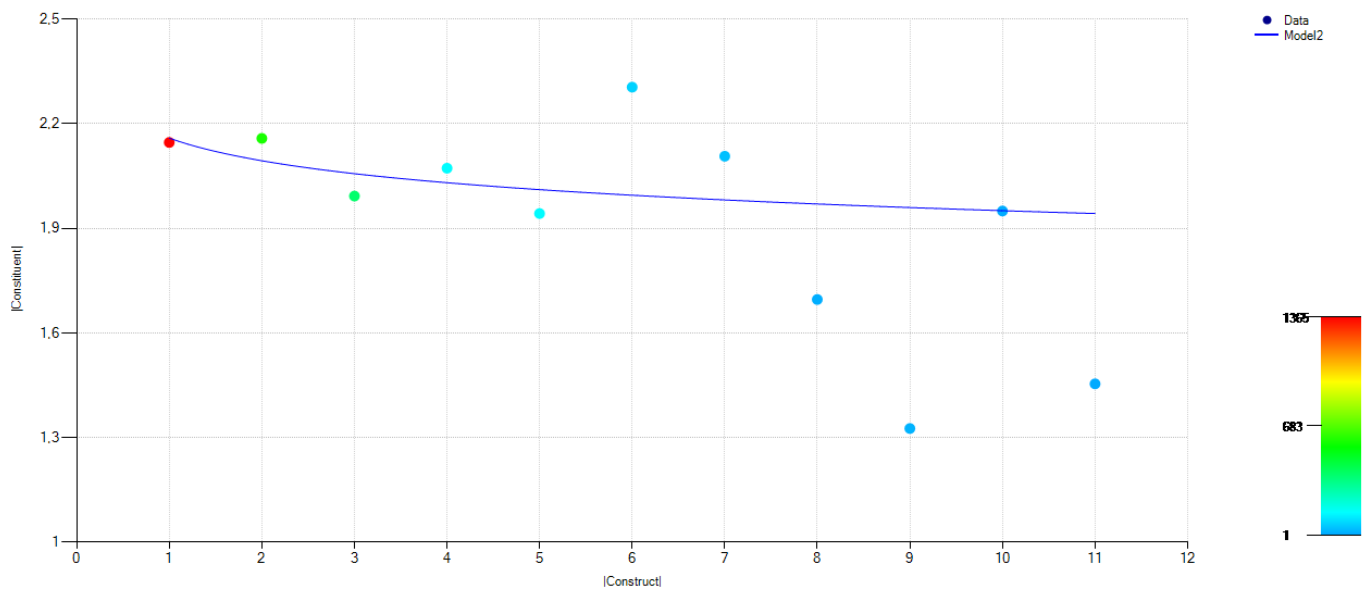
Tabulka 19: poezie – U₃ – podíl znaků dle počtu prvků

Jak je patrné z procentuálního zastoupení znaků, reálná variabilita znaků dle počtu jejich prvků je (opět) velmi nízká. Jednoprvkové znaky tvoří více než celou polovinu všech znaků obsažených ve vybraném textu (a více než jednu čtvrtinu, pokud se zohlední pouze unikátní znaky bez opakování), přičemž jednoprvkové až tříprvkové znaky pak tvoří drtivou většinu všech znaků v textu, přes 85 % (a necelé dvě třetiny, pokud se zohlední pouze unikátní znaky bez opakování). Takto výrazně malá variabilita znaků, co

se týče počtu v nich obsažených prvků, by rovněž mohla být jednou z příčin negativního výsledku na této jazykové úrovni.⁶⁸

Nakonec by mohlo být důvodem pro negativní výsledek i zahrnutí interpunkčních znamének do jazykové jednotky znaku. Ta interpunkční znaménka, která neukončují mezičárí, do jisté míry délku mezičárí zvyšují, což by také mohlo mít vliv na projev vztahů mezi konstrukty a konstituenty.

4.2.4. Poezie – U₄ – znaky-prvky



Graf 8: poezie – U₄ – znaky-prvky

$$A = 2,1603134081$$

$$b = 0,0445206618$$

$$R^2 = 0,2754253404$$

Tabulka 20: A, b, R^2 (poezie – U₄ – znaky-prvky)

x	z	y
1	1365	2,1466275659824
2	585	2,15825491873396

⁶⁸ Nejfrekventovanějšími jednoprvkovými až tříprvkovými znaky jsou i v tomto případě znaky *kany* a interpunkční znaménka.

x	z	y
3	375	1,99287622439893
4	134	2,07276119402985
5	133	1,94285714285714
6	65	2,3051282051282
7	36	2,10714285714286
8	7	1,69642857142857
9	16	1,32638888888889
10	4	1,95
11	1	1,45454545454545

Tabulka 21: Hodnoty x , z , y (poezie – U_4 – znaky-prvky)

V tabulce 21 jsou uvedeny hodnoty získané kvantifikací 2. analyzovaného textu (poezie) na jazykové úrovni U_4 . Konstrukt x zde představuje délku znaků, měřených v počtu prvků. Hodnoty z představují frekvenci výskytu konstruktů o dané délce. Konstituent y pak představuje průměrnou délku prvků, měřených v počtu bezprostředně nižších jazykových jednotek, tj. v tazích.

Délka konstruktů vyjádřených v počtu konstituentů na této jazykové úrovni je od 1 do 11. Průměrná délka konstituentů je v rozmezí $\langle 1,326; 2,305 \rangle$. Přesné hodnoty parametru A , parametru b a koeficientu determinace R^2 pro jazykovou úroveň U_4 jsou uvedeny v tabulce 20.

Na této úrovni je z tabulky 21 patrné, že průměrné délky konstituentů s rostoucími délkami konstruktů střídavě klesají a opět stoupají; proložená regresní křivka má spíše rovný tvar. Parametr b je zde kladné číslo a MAZ lze tedy prohlásit za teoreticky platný. Koeficient determinace R^2 má však jen velmi nízkou hodnotu 0,275, což značí, že celý model dobře neseďí.

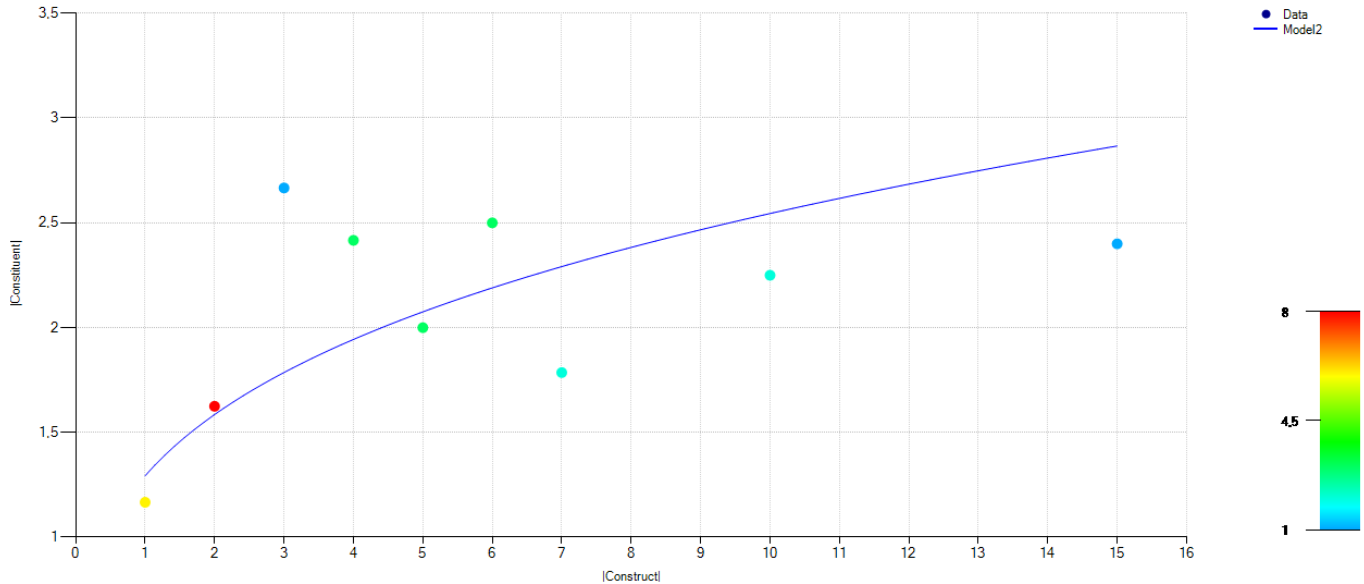
Výsledek pro 2. analyzovaný text je tedy obdobný jako pro 1. a částečně 3. analyzovaný text: Na jazykové úrovni znaky-prvky nelze potvrdit, že čím více znaky obsahují prvků, tím méně tahů obsahují v průměru prvků.

Důvody pro daný výsledek na této jazykové úrovni jsou pravděpodobně totožné s důvody neplatnosti MAZ na stejné jazykové úrovni u analyzovaných textů č. 1 a č. 3; především se jedná o běžné užívání dvou sylabografických abeced a dalších znaků společně se znaky *kandži* v rámci jednoho textu, čímž je japonština do jisté míry unikátní.

4.3. Analyzovaný text č. 3 – seminární práce

Níže jsou uvedeny výsledky analýzy 3. analyzovaného textu (seminární práce) pro jazykové úrovně U_1 až U_4 .

4.3.1. Seminární práce – U_1 – odstavce-souvětí



Graf 9: seminární práce – U_1 – odstavce-souvětí

$$A = 1,2930143335$$

$$b = -0,2939941035$$

$$R^2 = 0,7163675860$$

Tabulka 22: A, b, R^2 (seminární práce – U_1 – odstavce-souvětí)

x	z	y
1	6	1,16666666666667
2	8	1,625
3	1	2,66666666666667
4	3	2,41666666666667
5	3	2
6	3	2,5
7	2	1,78571428571429
10	2	2,25
15	1	2,4

Tabulka 23: Hodnoty x, z, y (seminární práce – U_1 – odstavce-souvětí)

V tabulce 23 jsou uvedeny hodnoty získané kvantifikací 3. analyzovaného textu (seminární práce) na jazykové úrovni U_1 . Konstrukt x zde představuje délku odstavců, měřených v počtu souvětí. Hodnoty z představují frekvenci výskytu konstruktů o dané délce. Konstituent y pak představuje průměrnou délku souvětí, měřených v počtu bezprostředně nižších jazykových jednotek, tj. v mezičáří.

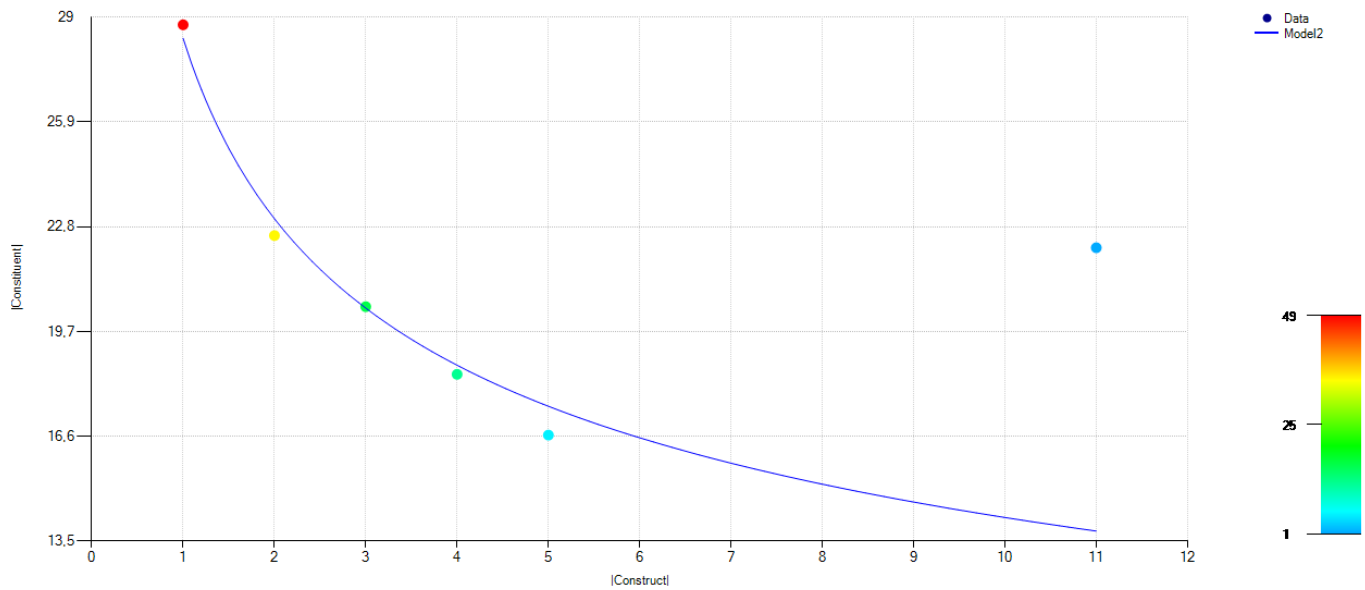
Délka konstruktů vyjádřených v počtu konstituentů na této jazykové úrovni je od 1 do 15. Průměrná délka konstituentů je v rozmezí $\langle 1,167; 2,667 \rangle$. Přesné hodnoty parametru A , parametru b a koeficientu determinace R^2 pro jazykovou úroveň U_1 jsou uvedeny v tabulce 22.

Na této úrovni se vztah zkracování průměrných délek konstituentů se vzrůstající délkou konstruktů neprojevil. Naopak, z tabulky 23 je patrná tendence spíše opačná s občasným poklesem délky konstituentů a následným opětovným růstem. Stejně tak parametr b je číslo záporné a proložená křivka je tudíž rostoucí a konkávní, což odporuje předpokladu platnosti MAZ. Koeficient determinace R^2 má sice zajímavou hodnotu 0,716, avšak za daných okolností je to patrně nepříliš relevantní.

Na úrovni U_1 tedy lze říci, že vztah „čím více odstavce obsahují souvětí, tím méně mezičáří obsahují v průměru daná souvětí“ neplatí.

Přesto však tento výsledek není příliš překvapivý a je v souladu s hypotézou. Jedním z možných důvodů může být zavedení striktně graficky definované jednotky mezičáří, která je sice vizuálně velice snadno identifikovatelná, ale může ignorovat syntaktické a jiné aspekty textu a není tudíž moc přirozená. Dalším možným důvodem by mohl být jednoduše nedostatečný počet různých bodů x , kterých se objevilo pouze 9. To mohlo způsobit, že statistická relevance není dostatečně průkazná a je třeba provést další analýzy textů, obsahujících větší počet bodů x , tj. odstavců.

4.3.2. Seminární práce – U₂ – souvětí-mezičáří



Graf 10: seminární práce – U₂ – souvětí-mezičáří

$$A = 28,3781248539$$

$$b = 0,3006119296$$

$$R^2 = 0,9273694696$$

Tabulka 24: A, b, R² (seminární práce – U₂ – souvětí-mezičáří)

x	z	y
1	49	28,7755102040816
2	35	22,5428571428571
3	16	20,4375
4	12	18,4375
5	5	16,64
11	1	22,1818181818182

Tabulka 25: Hodnoty x, z, y (seminární práce – U₂ – souvětí-mezičáří)

V tabulce 25 jsou uvedeny hodnoty získané kvantifikací 3. analyzovaného textu (seminární práce) na jazykové úrovni U₂. Konstrukt x zde představuje délku souvětí, měřených v počtu mezičáří. Hodnoty z představují frekvenci výskytu konstruktů o dané délce. Konstituent y pak představuje průměrnou délku mezičáří, měřených v počtu bezprostředně nižších jazykových jednotek, tj. ve znacích.

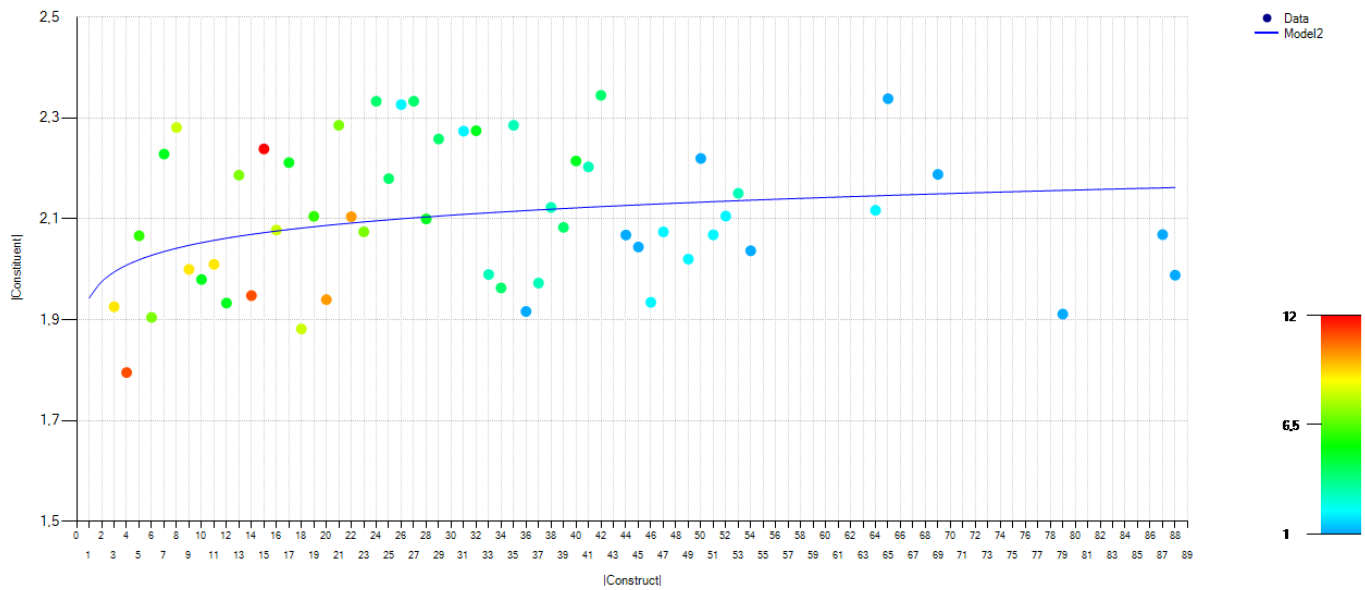
Délka konstruktů vyjádřených v počtu konstituentů na této jazykové úrovni je od 1 do 11. Průměrná délka konstituentů je v rozmezí (16,64; 28,776). Přesné hodnoty parametru A , parametru b a koeficientu determinace R^2 pro jazykovou úroveň U_2 jsou uvedeny v tabulce 24.

Na rozdíl od 1. analyzovaného textu (povídky) tentokrát již výsledek odpovídá hypotéze. Data v tabulce 25 ukazují, že s výjimkou posledního konstruktů s nejdelší délkou v počtu konstituentů, který se objevil pouze jedinkrát, u všech ostatních konstruktů se potvrdil vztah nepřímé úměry, tedy se stoupající délkou konstruktů klesá průměrná délka jejich konstituentů. Výsledkem je kladný parametr b , což je předpoklad platnosti MAZ, a koeficient determinace R^2 s extrémně vysokou hodnotou 0,927.

Na této úrovni platí, že při tvorbě daného textu mohl autor více méně volně ovlivňovat délku souvětí i mezičárí a přizpůsobit ho tak té podobě, která mu vyhovovala. Seminární práce sice obsahovala několik přímých citací z jiné literatury, tyto však nebyly zdaleka tak početné jako např. dialogy v povídce, což mohlo přispět k lepší shodě s modelem MAZ. Na jazykové úrovni souvětí-mezičárí u seminární práce tedy platí tvrzení, že čím více souvětí obsahují mezičárí, tím méně znaků obsahují v průměru mezičárí.⁶⁹

⁶⁹ Zároveň se jedná o styl, který přeci jen nemá zcela volnou grafickou podobu (autor se při tvorbě seminární práce částečně musí přidržovat určitých psaných i nepsaných pravidel či zvyklostí). Výsledná podoba textu tedy, dá se říci, „směřuje“ k určité finální (grafické) podobě (na rozdíl od povídky s veskrze svévolnou možnou podobou), podobně jako tomu bylo u 2. analyzovaného textu. To by opět bylo v souladu s předpokladem, že grafická omezení textu mají vliv na projev MAZ. Zda je však toto platné právě pro jazykovou úroveň U_2 u seminární práce, když (na rozdíl od básnického textu) úroveň U_1 nevykázala žádnou shodu s matematickým modelem MAZ, zůstává zatím nezodpovězenou otázkou – může se jednat o pouhou náhodnou korelaci.

4.3.3. Seminární práce – U₃ – mezičárí-znaky



Graf 11: seminární práce – U₃ – mezičárí-znaky

$$A = 1,9431475310$$

$$b = -0,0238655529$$

$$R^2 = 0,0842810418$$

Tabulka 26: A, b, R² (seminární práce – U₃ – mezičárí-znaky)

x	z	y
3	9	1,92592592592593
4	11	1,79545454545455
5	6	2,06666666666667
6	7	1,9047619047619
7	5	2,22857142857143
8	8	2,28125
9	9	2
10	5	1,98
11	9	2,01010101010101
12	5	1,93333333333333
13	7	2,18681318681319
14	11	1,94805194805195
15	12	2,23888888888889
16	8	2,078125

x	z	y
17	5	2,21176470588235
18	8	1,88194444444444
19	6	2,10526315789474
20	10	1,94
21	7	2,28571428571429
22	10	2,10454545454545
23	7	2,07453416149068
24	4	2,33333333333333
25	4	2,18
26	2	2,32692307692308
27	4	2,33333333333333
28	5	2,1
29	4	2,25862068965517
31	2	2,2741935483871

x	z	y
32	5	2,275
33	3	1,98989898989899
34	4	1,96323529411765
35	3	2,28571428571429
36	1	1,91666666666667
37	3	1,97297297297297
38	3	2,12280701754386
39	4	2,08333333333333
40	5	2,215
41	3	2,20325203252033
42	4	2,3452380952381
44	1	2,06818181818182
45	1	2,04444444444444
46	2	1,93478260869565

x	z	y
47	2	2,07446808510638
49	2	2,02040816326531
50	1	2,22
51	2	2,06862745098039
52	2	2,10576923076923
53	3	2,15094339622642
54	1	2,03703703703704
64	2	2,1171875
65	1	2,33846153846154
69	1	2,18840579710145
79	1	1,91139240506329
87	1	2,06896551724138
88	1	1,98863636363636

Tabulka 27: Hodnoty x, z, y (seminární práce – U₃ – mezičárí-znaky)

V tabulce 27 jsou uvedeny hodnoty získané kvantifikací 3. analyzovaného textu (seminární práce) na jazykové úrovni U₃. Konstrukt *x* zde představuje délku mezičárí, měřených v počtu znaků. Hodnoty *z* představují frekvenci výskytu konstruktů o dané délce. Konstituent *y* pak představuje průměrnou délku znaků, měřených v počtu bezprostředně nižších jazykových jednotek, tj. v prvcích.

Délka konstruktů vyjádřených v počtu konstituentů na této jazykové úrovni je od 3 do 88. Průměrná délka konstituentů je v rozmezí (1,795; 2,345). Přesné hodnoty parametru *A*, parametru *b* a koeficientu determinace R^2 pro jazykovou úroveň U₃ jsou uvedeny v tabulce 26.

Na jazykové úrovni mezičárí-znaky hodnoty průměrných délek konstituentů se vzrůstající délkou konstruktů střídavě klesají a opět stoupají a na žádné delší sekvenci po sobě jdoucích konstruktů nelze pozorovat stabilně klesající délku konstituentů. Parametr *b* je na této úrovni číslo záporné, takže proložená regresní křivka je stoupající a konkávní, což nespĺňuje předpoklad platnosti MAZ. Hodnota koeficientu determinace R^2 rovněž vykazuje naprosto minimální shodu s matematickým modelem, a to 0,084.

Negativní výsledek platnosti MAZ rovněž odpovídá na jazykové úrovni U_3 hypotéze. Možné důvody pro neplatnost MAZ na této jazykové úrovni jsou podrobněji rozvedeny v kap. 3.3. U_3 – mezičáří-znaky Může to být samotné užití jednotky mezičáří, která je založena na čistě grafickém principu. Dále to může být způsobeno pravděpodobně příliš velkým skokem mezi jazykovou jednotkou mezičáří a jazykovou jednotkou znaku, mezi nimiž mohou chybět nějaké další mezijednotky.⁷⁰ Rovněž fakt, že mezičáří je jazyková jednotka s proměnlivou délkou, zatímco znaky, ze kterých se mezičáří skládají, jsou jednotka s délkou fixní (co se týče počtu prvků v každém konkrétním znaku). A nakonec také malá variabilita znaků, co se týče počtu v nich obsažených prvků, což by mohlo narušit vztah nepřímé úměrnosti mezi konstrukty a konstituenty.

Co se týče posledního argumentu, pro ověření toho, zda by mohl platit pro japonštinu v případě seminární práce a podobných akademických textů, je podíl znaků dle počtu jejich prvků uveden v následující tabulce.

celkový počet znaků: 5594 data pro všechny znaky v textu			celkový počet unikátních znaků: 702 data pro unikátní znaky v textu		
počet prvků ve znaku	četnost znaků o daném počtu prvků	podíl na celkovém počtu všech znaků (v %)	počet prvků ve znaku	četnost znaků o daném počtu prvků	podíl na celkovém počtu unikátních znaků (v %)
1	2632	47,05	1	175	24,929
2	1198	21,416	2	144	20,513
3	1048	18,734	3	150	21,368
4	332	5,935	4	93	13,248
5	170	3,039	5	54	7,692
6	98	1,752	6	38	5,413
7	61	1,09	7	24	3,419
8	33	0,59	8	14	1,994
9	19	0,34	9	8	1,14
10	3	0,054	10	2	0,285

Tabulka 28: seminární práce – U_3 – podíl znaků dle počtu prvků

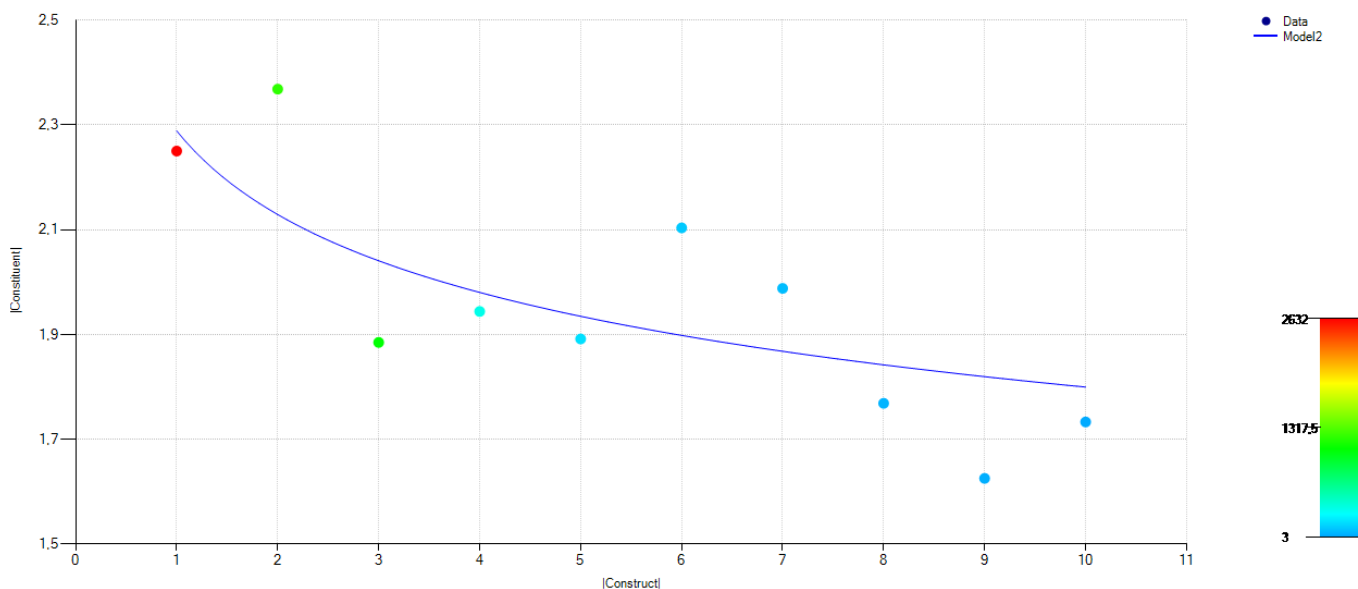
⁷⁰ Doplnující experimentální analýzy s užitím jazykových jednotek Dlouhé slovo a Krátké slovo, postavených mezi, či namísto jednotek mezičáří a znak, však ukázaly, že ani se začleněním těchto (negraficky segmentovaných) jednotek nedošlo k žádnému zlepšení, co se týče platnosti MAZ ve vztahu k tomuto analyzovanému textu; viz kap. 4.4. Doplnující experimenty s analyzovaným textem č. 3

Jak je patrné z procentuálního zastoupení znaků, reálná variabilita znaků dle počtu jejich prvků je i v tomto případě opět velmi nízká. Jednoprvkové znaky tvoří skoro celou polovinu všech znaků obsažených ve vybraném textu (a prakticky jednu čtvrtinu, pokud se zohlední pouze unikátní znaky bez opakování), přičemž jednoprvkové až tříprvkové znaky pak tvoří drtivou většinu všech znaků v textu – přes 87 % (a rovné dvě třetiny, pokud se zohlední pouze unikátní znaky bez opakování). Takto výrazně malá variabilita znaků, co se týče počtu v nich obsažených prvků, by rovněž mohla být jednou z příčin neplatnosti vztahu mezi délkou konstruktů a jejich konstituentů.⁷¹

Nakonec by mohlo být důvodem negativního výsledku i zahrnutí interpunkčních znamének do jednotky znaku, což není příliš běžná praxe. Ta interpunkční znaménka, která neukončují mezičárí, do jisté míry délku mezičárí zvyšují, což by také mohlo mít vliv na projev vztahů mezi konstrukty a konstituenty.

U jazykové úrovně mezičárí-znaky tedy ani v případě seminární práce neplatí tvrzení, že čím více mezičárí obsahuje znaků, tím méně prvků obsahují v průměru znaky.

4.3.4. Seminární práce – U₄ – znaky-prvky



Graf 12: seminární práce – U₄ – znaky-prvky

⁷¹ Nejfrekventovanějšími jednoprvkovými až tříprvkovými znaky jsou opět znaky *kany* a interpunkční znaménka.

$$A = 2,2889145244$$

$$b = 0,1043646604$$

$$R^2 = 0,4851495041$$

Tabulka 29: A , b , R^2 (seminární práce – U_4 – znaky-prvky)

x	z	y
1	2632	2,25037993920973
2	1198	2,36853088480801
3	1048	1,88517811704835
4	332	1,94427710843373
5	170	1,89176470588235
6	98	2,10374149659864
7	61	1,98829039812646
8	33	1,76893939393939
9	19	1,62573099415205
10	3	1,73333333333333

Tabulka 30: Hodnoty x , z , y (seminární práce – U_4 – znaky-prvky)

V tabulce 30 jsou uvedeny hodnoty získané kvantifikací 3. analyzovaného textu (seminární práce) na jazykové úrovni U_4 . Konstrukt x zde představuje délku znaků, měřených v počtu prvků. Hodnoty z představují frekvenci výskytu konstruktů o dané délce. Konstituent y pak představuje průměrnou délku prvků, měřených v počtu bezprostředně nižších jazykových jednotek, tj. v tazích.

Délka konstruktů vyjádřených v počtu konstituentů na této jazykové úrovni je od 1 do 10. Průměrná délka konstituentů je v rozmezí $\langle 1,626; 2,369 \rangle$. Přesné hodnoty parametru A , parametru b a koeficientu determinace R^2 pro jazykovou úroveň U_4 jsou uvedeny v tabulce 29.

Na nejnižší jazykové úrovni U_4 je výsledek analýzy poměrně zajímavý. Průměrné délky konstituentů s narůstající délkou konstruktů střídavě klesají a stoupají a ke stabilnímu poklesu průměrných délek konstituentů dochází až u větších délek konstruktů (s výjimkou velice málo frekventovaných 10prvkových znaků, jež se v textu objevily pouze třikrát). Vypočtený parametr b je ve výsledku číslo kladné a předpoklad platnosti MAZ byl tedy splněn. Co se týče hodnoty koeficientu determinace R^2 , nelze z něj jednoznačně určit, zda se jedná o spíše dobře či alespoň přijatelně, nebo o vyloženě

špatně sedící model: R^2 dosáhl hodnoty 0,485. Jisté shody s matematickým modelem MAZ však dosaženo bylo, což je patrné i po vizuálním zhodnocení grafu.

Přestože současná data naznačují, že tomu tak patrně může být, pro přesnější posouzení, zda na úrovni znaky-prvky v seminární práci platí argument „čím více znaky obsahují prvků, tím méně tahů obsahují v průměru prvky“ je nutné provést další experimenty.

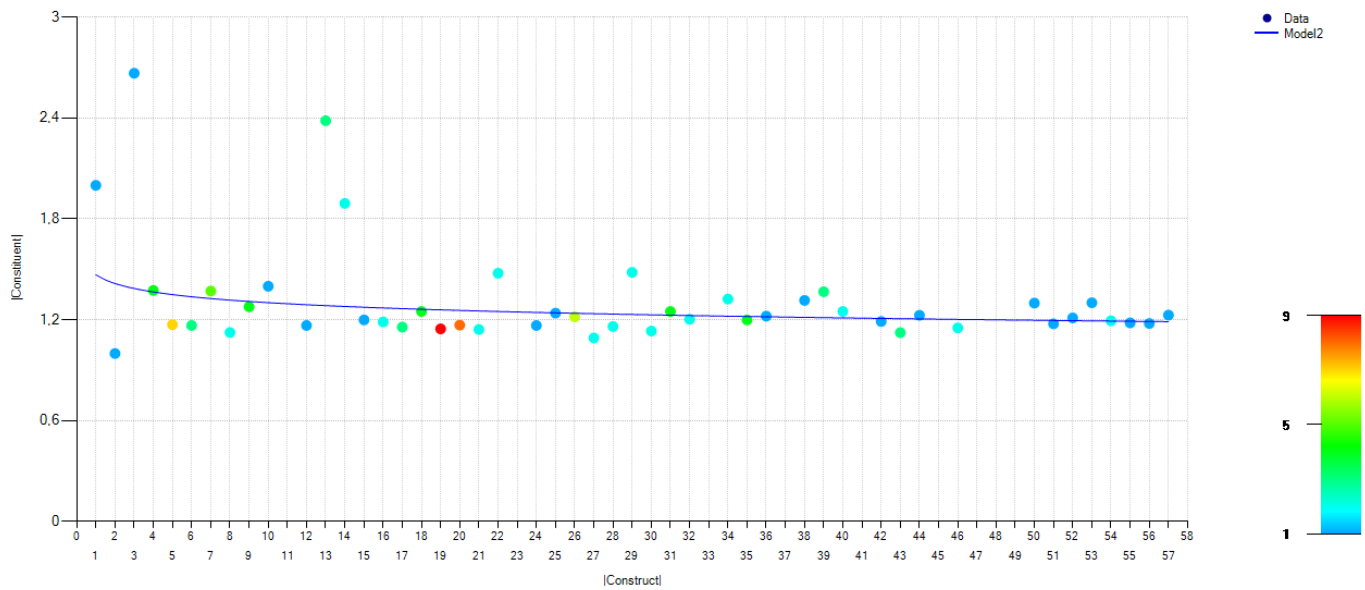
4.4. Doplnující experimenty s analyzovaným textem č. 3

Jako dodatečný experiment byl na 3. analyzovaném textu (seminární práci) otestován odlišný způsob segmentace textu na jazykové jednotky, a to zahrnutí dvou zcela negraficky segmentovaných jednotek slova (tzv. Dlouhého slova a Krátkého slova), založených na morfemické a syntaktické analýze textu a definovaných Národním institutem pro japonský jazyk a lingvistiku (NINJAL). Tyto dvě jednotky nevyhovují kritériu grafické segmentace, proto nebyly zahrnuty do všech výše provedených výpočtů. Nicméně, jak již bylo v této práci několikrát uvedeno, existuje podezření, že mezi zde užívanou jazykovou jednotkou mezičárí a jednotkou znaku je možná příliš náhlý přechod, tzn. je možné, že mezi nimi chybí ještě další mezijednotky a mohly by se tak vytrácet některé vztahy důležité pro projev Menzerath-Altmanova zákona na jazykových úrovních U_2 a U_3 .

Z toho důvodu bylo na analyzovaném textu č. 3 provedeno několik pokusných experimentů s rozdílnými způsoby segmentace a pouze pro úplnost (tj. vzhledem k (patrné) neexistenci jiných publikovaných podobných experimentů provedených na japonských textech) jsou výsledky popsány v rámci této práce. Jelikož původní metodika užitá v této práci nepočítala s těmito novými jednotkami, a zároveň (a to především) tyto nové jednotky odporují základnímu kritériu čistě grafické segmentace textu, bude se v tomto případě jednat pouze o experimentální pokusy bez stanovení hypotéz. Pro podrobnější popis užitého softwaru a uvedených jazykových jednotek viz (www 6) a (www 7).

Níže následují výsledky analýzy pro jednotlivé zkušební jazykové úrovně.

4.4.1. Seminární práce – U_{SD} – souvětí-Dlouhá slova



Graf 13: seminární práce – U_{SD} – souvětí-Dlouhá slova (bez bodu x s hodnotou 134)

$$A = 1,4682855486$$

$$b = 0,0520951469$$

$$R^2 = 0,0687516782$$

$$A = 1,4626715357$$

$$b = 0,0505995813$$

$$R^2 = 0,0681887690$$

Tabulka 31: A, b, R² (seminární práce – U_{SD} – souvětí-Dlouhá slova se všemi body x)

Tabulka 32: A, b, R² (seminární práce – U_{SD} – souvětí-Dlouhá slova bez bodu x s hodnotou 134)

x	z	y
1	1	2
2	1	1
3	1	2,66666666666667
4	4	1,375
5	7	1,17142857142857
6	3	1,16666666666667
7	5	1,37142857142857
8	2	1,125
9	4	1,27777777777778
10	1	1,4
12	1	1,66666666666667
13	3	2,38461538461538
14	2	1,89285714285714
15	1	1,2

x	z	y
16	2	1,1875
17	3	1,15686274509804
18	4	1,25
19	9	1,14619883040936
20	8	1,16875
21	2	1,14285714285714
22	2	1,47727272727273
24	1	1,16666666666667
25	1	1,24
26	6	1,21794871794872
27	2	1,09259259259259
28	2	1,16071428571429
29	2	1,48275862068966
30	2	1,13333333333333

x	z	y
31	4	1,25
32	2	1,203125
34	2	1,32352941176471
35	4	1,2
36	1	1,22222222222222
38	1	1,31578947368421
39	3	1,36752136752137
40	2	1,25
42	1	1,19047619047619
43	3	1,12403100775194

x	z	y
44	1	1,22727272727273
46	2	1,15217391304348
50	1	1,3
51	1	1,17647058823529
52	1	1,21153846153846
53	1	1,30188679245283
54	2	1,19444444444444
55	1	1,18181818181818
56	1	1,17857142857143
57	1	1,2280701754386

Tabulka 33: Hodnoty x, z, y (seminární práce – U_{SD} – souvětí-Dlouhá slova)

V tabulce 33 jsou uvedeny hodnoty získané experimentální kvantifikací 3. analyzovaného textu (seminární práce) na nově vytvořené jazykové úrovni U_{SD}. Konstrukt **x** zde představuje délku souvětí, měřených v počtu Dlouhých slov. Hodnoty **z** představují frekvenci výskytu konstruktů o dané délce. Konstituent **y** pak představuje průměrnou délku Dlouhých slov, měřených v počtu bezprostředně nižších jazykových jednotek, tj. v tomto experimentu v Krátkých slovech.

Posloupnost jazykových jednotek na této jazykové úrovni je tedy následující:

(odstavce) \Rightarrow **souvětí** \Rightarrow **Dlouhá slova** \Rightarrow **Krátká slova** \Rightarrow (znaky) \Rightarrow (prvky) \Rightarrow (tahy)

Jazyková jednotka mezičárí tedy byla na této jazykové úrovni zcela nahrazena jazykovými jednotkami Dlouhé slovo a Krátké slovo.

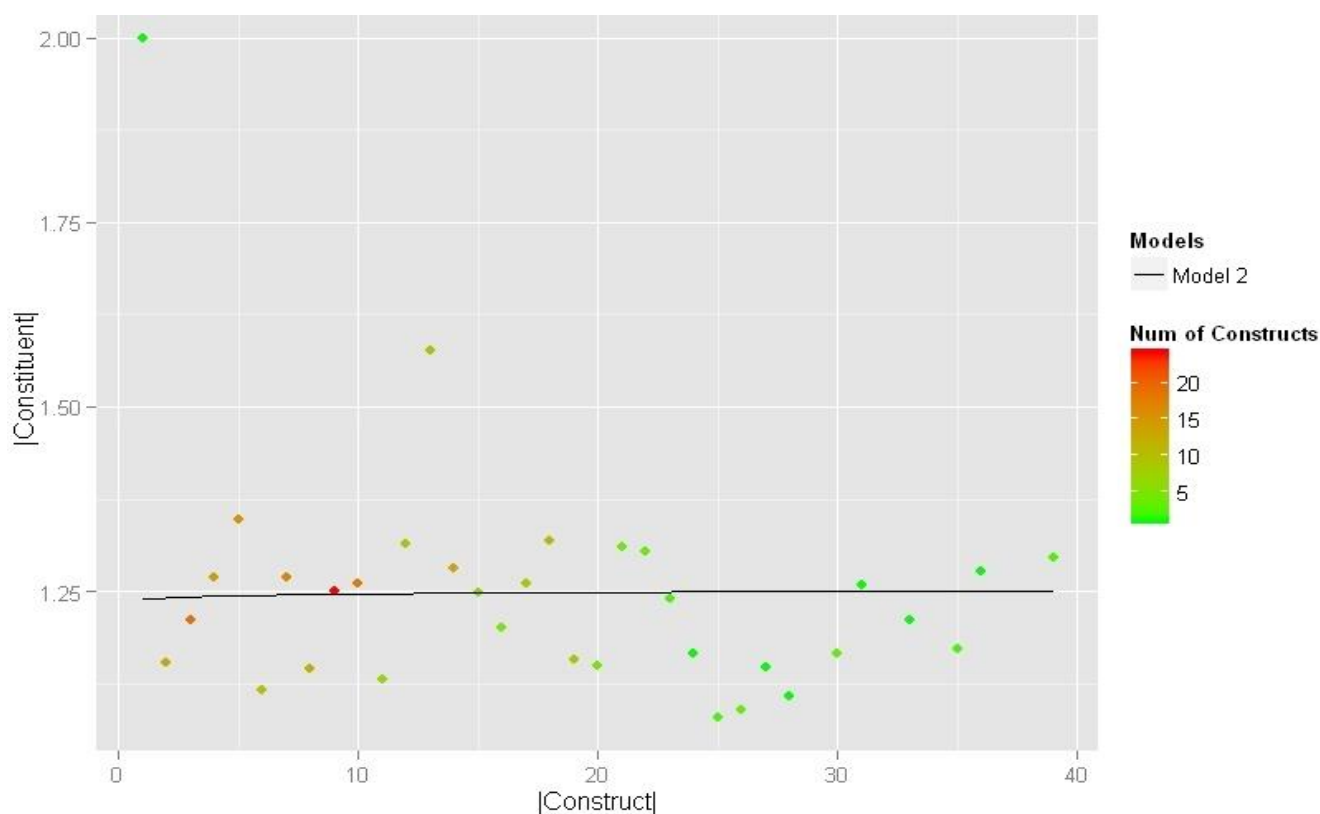
Délka konstruktů vyjádřených v počtu konstituentů na této jazykové úrovni je od 1 do 134. Průměrná délka konstituentů je v rozmezí $\langle 1; 2,667 \rangle$. Přesné hodnoty parametru *A*, parametru *b* a koeficientu determinace R^2 pro jazykovou úroveň U_{SD} jsou uvedeny v tabulce 31. Pro přehlednost byl z grafické vizualizace odstraněn poslední bod *x* o délce 134, jehož zahrnutí by odsunulo všechny zbývající body *x* v grafu do jeho levé poloviny a graf by se stal nepřehledným. Jelikož se konstrukt takové délky v textu objevil pouze jednou, jeho odstranění nemělo prakticky žádný vliv na hodnoty obou parametrů ani koeficientu determinace. Přesto jsou v tabulkách pro úplnost uvedeny hodnoty parametrů a koeficientu determinace s tímto bodem *x* i bez něj.

Se vzrůstající délkou konstruktů průměrná délka konstituentů v nahodilých intervalech klesá a opět stoupá. Z grafu je patrné, že body x jsou rozprostřeny poměrně lineárně, což potvrzuje tvar proložené křivky. Hodnota parametru b je sice kladné číslo a předpoklad platnosti MAZ tak byl splněn, avšak koeficient determinace R^2 ukazuje pouze minimální shodu s matematickým modelem: 0,068.

Při užití tohoto způsobu segmentace se tedy platnost MAZ na 3. analyzovaném textu na jazykové úrovni U_{SD} – souvětí-Dlouhá slova nepotvrdila a není možné prohlásit, že čím více souvětí obsahují Dlouhých slov, tím méně Krátkých slov obsahují v průměru Dlouhá slova.

Nahrazení čistě grafické jednotky mezičárí za odlišným způsobem vytvořené jednotky slov na jazykové úrovni U_{SD} tedy nevedlo k vysoké hodnotě R^2 či ke klesajícímu a konvexnímu tvaru proložené křivky. Naopak, ve srovnání s jazykovou úrovní U_2 (souvětí-mezičárí), na které se ukázala velmi vysoká míra shody s modelem MAZ ($R^2 = 0,927$), došlo k poklesu hodnoty koeficientu determinace na téměř naprostou nulu. Jelikož však došlo k nahrazení jazykové jednotky mezičárí (která byla vyšší než jednotka teoretického slova) rovnou za slovo, došlo k relativně velkému skoku od jednotky souvětí rovnou ke slovu. To je pravděpodobně přechod příliš náhlý a ignorování jednotky mezičárí či např. klauze apod., která by stála mezi souvětím a Dlouhým slovem, mohlo ovlivnit výsledky v neprospěch shody s matematickým modelem MAZ. Daný výsledek tedy nelze považovat za příliš překvapivý.

4.4.2. Seminární práce – U_{MD} – mezičáří-Dlouhá slova



Graf 14: seminární práce – U_{MD} – mezičáří-Dlouhá slova

$$A = 1,2389925784$$

$$b = -0,0026383692$$

$$R^2 = 0,0005933666$$

Tabulka 34: A, b, R² (seminární práce – U_{MD} – mezičáří-Dlouhá slova)

x	z	y
1	1	2
2	13	1,15384615384615
3	19	1,21052631578947
4	14	1,26785714285714
5	15	1,34666666666667
6	10	1,11666666666667
7	17	1,26890756302521
8	12	1,14583333333333
9	24	1,25
10	18	1,26111111111111

x	z	y
11	7	1,12987012987013
12	9	1,31481481481481
13	10	1,57692307692308
14	13	1,28021978021978
15	7	1,24761904761905
16	5	1,2
17	9	1,26143790849673
18	11	1,31818181818182
19	9	1,15789473684211
20	4	1,15

<i>x</i>	<i>z</i>	<i>y</i>
21	4	1,30952380952381
22	3	1,3030303030303
23	2	1,23913043478261
24	1	1,16666666666667
25	2	1,08
26	3	1,08974358974359
27	1	1,14814814814815

<i>x</i>	<i>z</i>	<i>y</i>
28	1	1,10714285714286
30	3	1,16666666666667
31	1	1,25806451612903
33	1	1,21212121212121
35	2	1,17142857142857
36	1	1,27777777777778
39	2	1,29487179487179

Tabulka 35: Hodnoty *x*, *z*, *y* (seminární práce – U_{MD} – mezičáří-Dlouhá slova)

V tabulce 35 jsou uvedeny hodnoty získané experimentální kvantifikací 3. analyzovaného textu (seminární práce) na nově vytvořené jazykové úrovni U_{MD} . Konstrukt *x* zde představuje délku mezičáří, měřených v počtu Dlouhých slov. Hodnoty *z* představují frekvenci výskytu konstruktů o dané délce. Konstituent *y* pak představuje průměrnou délku Dlouhých slov, měřených v počtu bezprostředně nižších jazykových jednotek, tj. v tomto experimentu v Krátkých slovech.

Posloupnost jazykových jednotek na této jazykové úrovni je tedy následující:

(odstavce) \Rightarrow (souvětí) \Rightarrow **mezičáří** \Rightarrow **Dlouhá slova** \Rightarrow **Krátká slova** \Rightarrow (znaky) \Rightarrow (prvky) \Rightarrow (tahy)

Mezi jazykové jednotky mezičáří a znaky tedy byly na této jazykové úrovni vloženy jazykové jednotky Dlouhé slovo a Krátké slovo.

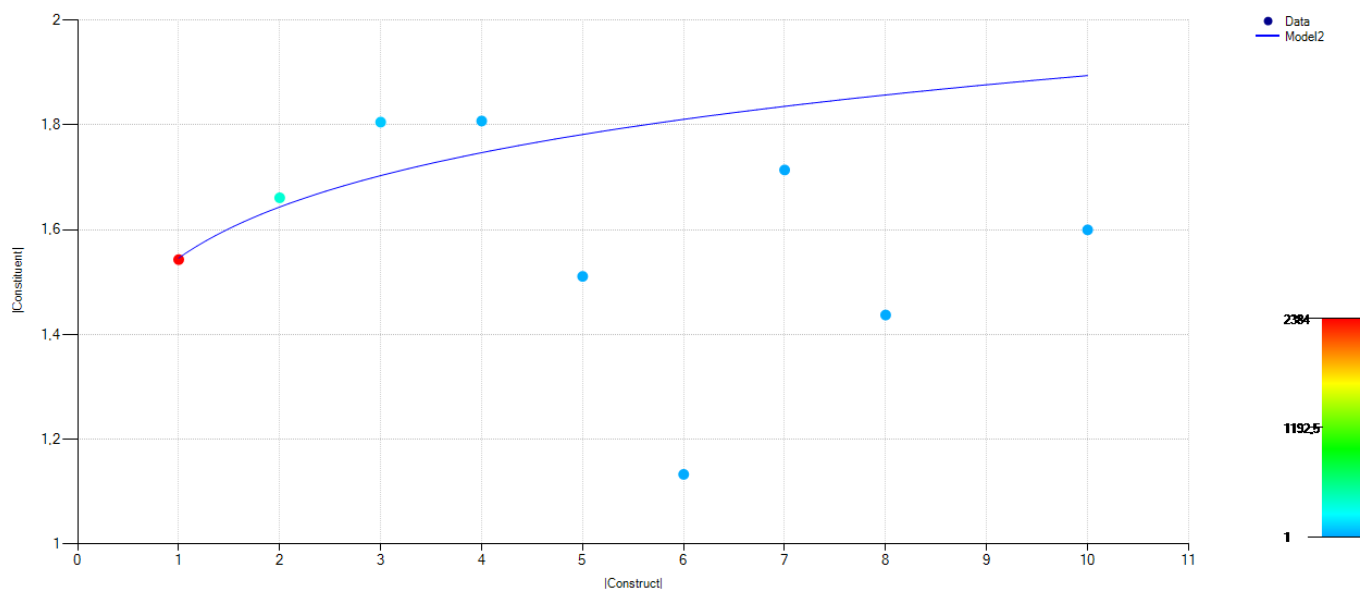
Délka konstruktů vyjádřených v počtu konstituentů na této jazykové úrovni je od 1 do 39. Průměrná délka konstituentů je v rozmezí $\langle 1,08; 2 \rangle$. Přesné hodnoty parametru *A*, parametru *b* a koeficientu determinace R^2 pro jazykovou úroveň U_{MD} jsou uvedeny v tabulce 34.

Se vzrůstající délkou konstruktů průměrná délka konstituentů v nahodilých intervalech klesá a opět stoupá. Z grafu je patrné, že body *x* jsou rozprostřeny poměrně lineárně, což potvrzuje tvar proložené křivky. Hodnota parametru *b* je na této jazykové úrovni záporné číslo, takže předpoklad platnosti MAZ nebyl splněn a koeficient determinace R^2 ukazuje zcela nulovou shodu s matematickým modelem: 0,001.

Při užití tohoto způsobu segmentace se tedy platnost MAZ na 3. analyzovaném textu na jazykové úrovni mezičáří-Dlouhá slova nepotvrdila a není možné prohlásit, že čím více

mezičáří obsahují Dlouhých slov, tím méně Krátkých slov obsahují v průměru Dlouhá slova. Ani zavedení mezijednotek mezi jazykové jednotky mezičáří a znaky tedy nevedlo k větší shodě s matematickým modelem MAZ; jak jazyková úroveň U_{MD} , tak i jazyková úroveň U_3 přinesly obdobně negativní výsledky.

4.4.3. Seminární práce – U_{DK} – Dlouhá slova-Krátká slova



Graf 15: seminární práce – U_{DK} – Dlouhá slova-Krátká slova

$$A = 1,5461186043$$

$$b = -0,0882158756$$

$$R^2 = 0,5622455758$$

Tabulka 36: A, b, R^2 (seminární práce – U_{DK} – Dlouhá slova-Krátká slova)

x	z	y
1	2384	1,543204698
2	360	1,661111111
3	84	1,805555556
4	26	1,807692308
5	9	1,511111111

x	z	y
6	5	1,133333333
7	1	1,714285714
8	4	1,4375
10	1	1,6

Tabulka 37: Hodnoty x, z, y (seminární práce – U_{DK} – Dlouhá slova-Krátká slova)

V tabulce 37 jsou uvedeny hodnoty získané experimentální kvantifikací 3. analyzovaného textu (seminární práce) na nově vytvořené jazykové úrovni U_{DK} . Konstrukt x zde představuje délku Dlouhých slov, měřených v počtu Krátkých slov. Hodnoty z představují frekvenci výskytu konstruktů o dané délce. Konstituent y pak představuje průměrnou délku Krátkých slov, měřených v počtu bezprostředně nižších jazykových jednotek, tj. v tomto experimentu ve znacích.

Posloupnost jazykových jednotek na této jazykové úrovni je tedy následující:
(odstavce) \Rightarrow (souvětí) \Rightarrow (mezičáří) \Rightarrow **Dlouhá slova** \Rightarrow **Krátká slova** \Rightarrow **znaky** \Rightarrow (prvky)
 \Rightarrow (tahy)

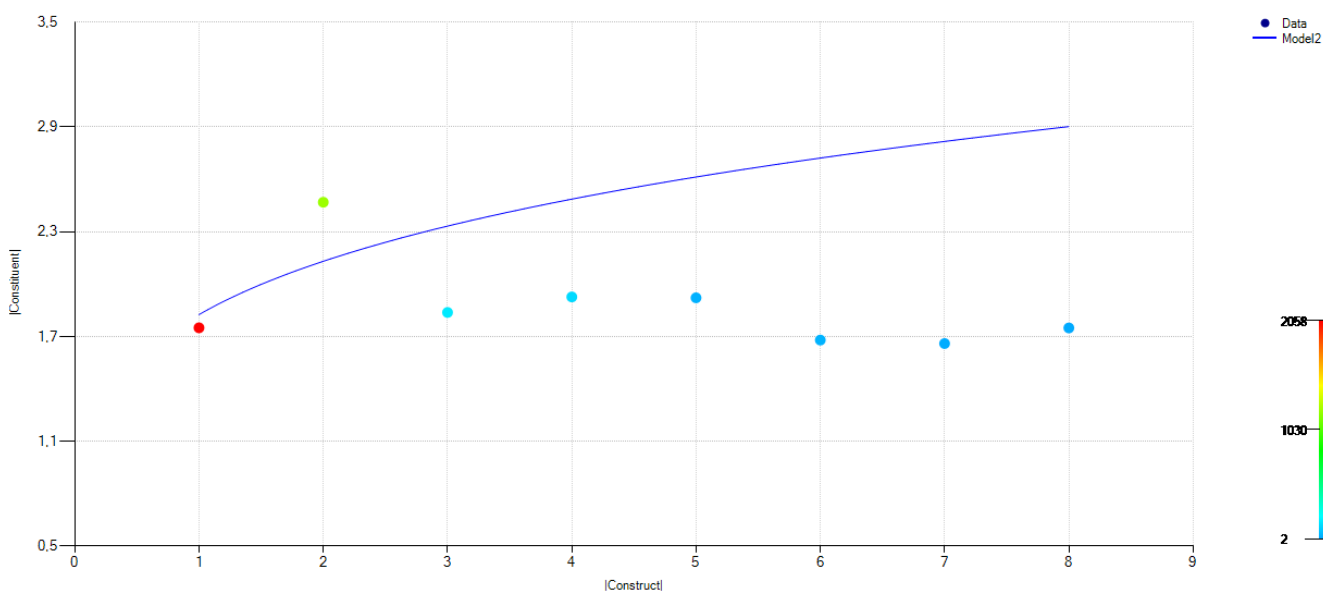
Mezi jazykové jednotky mezičáří a znaky tedy byly na této jazykové úrovni opět vloženy jazykové jednotky Dlouhé slovo a Krátké slovo.

Délka konstruktů vyjádřených v počtu konstituentů na této jazykové úrovni je od 1 do 10. Průměrná délka konstituentů je v rozmezí $\langle 1,133; 1,808 \rangle$. Přesné hodnoty parametru A , parametru b a koeficientu determinace R^2 pro jazykovou úroveň U_{DK} jsou uvedeny v tabulce 36.

Se vzrůstající délkou konstruktů průměrná délka konstituentů zpočátku naopak stoupá. Z grafu je patrné, že body x jsou rozprostřeny poměrně nahodile. Hodnota parametru b je na této jazykové úrovni záporné číslo, takže předpoklad platnosti MAZ nebyl splněn a proložená křivka je mírně stoupající a konkávní. Koeficient determinace R^2 o hodnotě 0,562 by v ideálním případě naznačoval poměrně nadějně sedící model, se záporným parametrem b tomu tak však patrně není.

Při užití tohoto způsobu segmentace se tedy platnost MAZ na 3. analyzovaném textu na jazykové úrovni Dlouhá slova-Krátká slova nepotvrdila a není možné prohlásit, že čím více Dlouhá slova obsahují Krátkých slov, tím méně znaků obsahují v průměru Krátká slova. Ani na této jazykové úrovni tedy zavedení dvou mezijednotek mezi jazykové jednotky mezičáří a znaky nevedlo k platnosti MAZ; jak jazyková úroveň U_{KZ} , tak i jazyková úroveň U_3 přinesly obdobně negativní výsledky.

4.4.4. Seminární práce – U_{KZ} – Krátká slova-znaky



Graf 16: seminární práce – U_{KZ} – Krátká slova-znaky

$$A = 1,8263555055$$

$$b = -0,2227927686$$

$$R^2 = 0,3858377595$$

Tabulka 38: A , b , R^2 (seminární práce – U_{KZ} – Krátká slova-znaky)

x	z	y
1	2058	1,75072886297376
2	1199	2,46955796497081
3	157	1,83864118895966
4	117	1,92735042735043
5	18	1,92222222222222
6	24	1,68055555555556
7	8	1,66071428571429
8	2	1,75

Tabulka 39: Hodnoty x , z , y (seminární práce – U_{KZ} – Krátká slova-znaky)

V tabulce 39 jsou uvedeny hodnoty získané experimentální kvantifikací 3. analyzovaného textu (seminární práce) na nově vytvořené jazykové úrovni U_{KZ}. Konstrukt x zde představuje délku Krátkých slov, měřených v počtu znaků. Hodnoty

z představují frekvenci výskytu konstruktů o dané délce. Konstituent y pak představuje průměrnou délku znaků, měřených v počtu bezprostředně nižších jazykových jednotek, tj. v prvcích.

Posloupnost jazykových jednotek na této jazykové úrovni je tedy následující:

(odstavce) \Rightarrow (souvětí) \Rightarrow (mezičáří) \Rightarrow (Dlouhá slova) \Rightarrow **Krátká slova** \Rightarrow **znaky** \Rightarrow **prvky** \Rightarrow (tahy)

Mezi jazykové jednotky mezičáří a znaky tedy byly na této jazykové úrovni navíc znovu vloženy jazykové jednotky Dlouhé slovo a Krátké slovo.

Délka konstruktů vyjádřených v počtu konstituentů na této jazykové úrovni je od 1 do 8. Průměrná délka konstituentů je v rozmezí $\langle 1,661; 2,47 \rangle$. Přesné hodnoty parametru A , parametru b a koeficientu determinace R^2 pro jazykovou úroveň U_{KZ} jsou uvedeny v tabulce 38.

I na této jazykové úrovni se vzrůstající délkou konstruktů průměrná délka konstituentů v nahodilých intervalech klesá a opět stoupá. Z grafu je patrné, že body x jsou rozprostřeny poměrně lineárně. První dva konstrukty se zdaleka nejvyšší frekvencí výskytu pak určují stoupající a konkávní tvar proložené křivky, což není v souladu s typickým tvarem křivky MAZ; hodnota parametru b je na této jazykové úrovni záporné číslo. Hodnota koeficientu determinace R^2 je 0,386.

Ani při užití tohoto způsobu segmentace se tedy platnost MAZ na 3. analyzovaném textu na jazykové úrovni Krátká slova-znaky nepotvrdila a není možné prohlásit, že čím více Krátká slova obsahují znaků, tím méně prvků obsahují v průměru znaky. Ani zde zavedení mezijednotek mezi jazykové jednotky mezičáří a znaky nevedlo k platnosti MAZ či větší shodě s matematickým modelem; jak jazyková úroveň U_{DK} , tak i jazyková úroveň U_3 přinesly obdobně negativní výsledky.

5. Závěry

Níže následují jednotlivé závěry analýz pro každý ze tří analyzovaných textů, na kterých se v rámci této práce ověřovala platnost Menzerath-Altmanova zákona na psané japonštině. Pro grafickou segmentaci bylo pro všechny texty vytvořeno sedm a z toho užito šest jazykových jednotek (odstavec, souvětí, mezičárí, znak, prvek, tah), resp. čtyři jazykové úrovně, pojmenované U_1 (odstavce-souvětí) až U_4 (prvky-tahy).

5.1. Závěr pro analyzovaný text č. 1 – povídku

Pro ověření platnosti Menzerath-Altmanova zákona na japonském textu v povídkovém stylu při užití grafické segmentace byla vybrána povídka *Šinkon-san* o délce 6204 znaků dle zde zavedené definice znaku.

Na jazykové úrovni U_1 měl vypočítaný parametr b zápornou hodnotu a proložená křivka byla stoupající. Jedním z možných důvodů negativního výsledku může být výskyt poměrně velkého množství dialogů v textu, kdy každé ukončení přímé řeči vedlo k ukončení odstavce, které tak mohly být velice krátké; takto krátké odstavce se v ostatních vybraných textech nevyskytovaly. Dalším možným důvodem může být užití grafické jednotky mezičárí.

Na jazykové úrovni U_2 byla situace velice podobná a hodnota parametru b byla rovněž záporná. Na rozdíl od ostatních jazykových úrovní však tento negativní výsledek neodpovídal hypotéze. Možných důvodů, proč se MAZ na této hladině neukázal jako dobře sedící model, je opět několik. Především je to fakt, že různých bodů x bylo velice málo, a navíc dva poslední body x se vyskytly pouze jednou, takže statistická relevance dat nemusí být příliš vysoká. Projev vztahu mezi konstrukty a konstituenty mohl být narušen užitím graficky definované jednotky mezičárí, která místy ignoruje sémantická či syntaktická hlediska textu, a rovněž užití jednotky znaku jako bezprostředně nižší jazykové jednotky mohlo narušit ideální posloupnost jazykových jednotek v případě, že mezi nimi chybí nějaké další jednotky, které by figurovaly namísto velice obtížně definovatelné jednotky (ortografického) slova, či by zastávaly jeho funkci. Zahrnutí interpunkčních znamének, která do jisté míry ovlivňují délku mezičárí, do jazykové jednotky znaku také mohlo ovlivnit výsledky analýzy na této jazykové úrovni.

I na jazykové úrovni U_3 nedošlo ke změně a parametr b je opět záporný. MAZ tedy ani na této úrovni neplatí. Možnými důvody mohou být implementace jazykové jednotky mezičárí, která je založena na grafickém principu; příliš velký skok mezi jednotkou mezičárí a jednotkou znaku, mezi nimiž mohou chybět jiné mezijednotky; proměnlivost délky mezičárí v závislosti na vůli autora, zatímco jeho konstituenty – znaky – mají fixní délku, co se týče počtu jejich prvků; interference interpunkčních znamének, zahrnutých do jednotky znaku; a nakonec také potvrzená velice malá variabilita znaků, co se týče počtu v nich obsažených prvků.

Na nejnižší jazykové úrovni U_4 se výsledek opakuje a hodnota parametru b je i na této jazykové úrovni záporná. Na tomto výsledku bylo poměrně překvapující, že na rozdíl od čínských textů,⁷² kde se MAZ potvrdil s velmi vysokými hodnotami R^2 (většinou nad 0,9), v případě japonštiny tomu bylo zcela naopak. Patrně hlavním důvodem takového rozdílu může být japonské užívání dvou sylabografických (moraických) abeced *kana* společně se znaky *kandži* a znaky dalšími. Na jednu stranu se japonské znaky zapisují do striktně vymezených grafických polí, která omezují množství prvků, jež se mohou dovnitř polí vměstnat. Na druhou stranu však existuje předpoklad, že znaky *kandži*, znaky *kana* aj. nejsou ekvivalentní, tj. mohly by tvořit hierarchicky odlišné, či paralelní jazykové jednotky a ekonomizační hledisko, patrně platné pro znaky *kandži*, nemusí pro ostatní znaky platit.

Celkový závěr pro 1. analyzovaný text je, že žádná jazyková úroveň neprokazuje platnost MAZ při implementaci užitých způsobů grafické segmentace. Jelikož se však jednalo prozatím o jediný vzorek povídky, je ještě příliš brzy na vyvození závěru, že MAZ neplatí pro všechny japonské texty v povídkovém stylu.

5.2. Závěr pro analyzovaný text č. 2 – poezii

Pro ověření platnosti Menzerath-Altmanova zákona na japonském básnickém textu při užití grafické segmentace byla vybrána báseň *Havran* v japonském překladu o délce 2721 znaků dle zde zavedené definice znaku.

Na jazykové úrovni U_1 byl vypočítaný parametr b kladný, hodnota koeficientu determinace R^2 měla extrémně vysokou hodnotu 0,979 a proložená regresní křivka tedy

⁷² (Motalová 2013)

byla klesající a konvexní s téměř ideálním tvarem. Přestože celkový počet bodů x nebyl vysoký a bylo by vhodné provést další experimenty s básnickými texty v japonštině, na tomto výběrovém souboru se na jazykové úrovni U_1 potvrdila platnost Menzerath-Altmanova zákona. Zároveň se potvrdila hypotéza, že u básnického textu může být výsledek analýzy odlišný od výsledků na stejné jazykové úrovni ostatních analyzovaných textů. Tento text měl, na rozdíl od ostatních dvou analyzovaných textů, na jazykové úrovni U_1 značně omezenou možnou délku konstruktů, a tedy i počet konstituentů a průměrnou délku mezičárí, aby nebyla příliš narušena grafická podoba strof (odstavců), jež byly navíc všechny v počtu znaků podobně dlouhé. Jinak řečeno, odstavce byly graficky ohraničeny svou délkou, tudíž čím bude takový odstavec obsahovat delší souvětí, tím kratší by daná souvětí měla být schopna pojmout mezičárí, aby příliš nepřekračovala omezenou délku odstavce. Platí tedy, že čím delší je konstrukt, tím kratší jsou (resp. musí být) jeho konstituenty. Toto se do jisté míry dá přirovnat k chování znaků v restriktivních čtvercových grafických polích, při jejichž analýze na čínských textech se ukázalo, že čím více má znak (uvnitř omezeného grafického pole) prvků, tím méně tahů mají (resp. musejí mít) v průměru tyto prvky, aby byly schopny se vměstnat do prostoru omezených grafických polí; hodnoty R^2 v případě těchto čínských textů byly většinou vyšší než 0,9. Bylo by proto vhodné provést analýzy dalších básnických textů pro ověření, zda má opravdu na platnost MAZ významný vliv restriktivní omezení grafické podoby textu.

I na jazykové úrovni U_2 je parametr b číslo kladné a koeficient determinace R^2 vykázal poměrně vysokou hodnotu 0,661. Se snižující se frekvencí konstruktů sice dochází k nárůstu průměrných délek konstituentů, avšak jedná se o konstrukty s relativně nízkou četností. S takovýmto výsledkem není snadné vyvozovat jednoznačné závěry, avšak vzhledem ke tvaru křivky i relativně vysoké hodnotě R^2 pravděpodobně lze na této úrovni předběžně prohlásit MAZ za platný. Co se týče důvodů, proč se MAZ na této úrovni projevil pouze u básnického textu, lze uvažovat o tom, že maximální délka souvětí je poměrně výrazně ovlivněna omezením maximální délky odstavce jakožto vyšší jazykové jednotky, která restriktivně působí na délky či počet souvětí. Podobně jako na jazykové úrovni U_1 by tedy mělo platit, že čím více budou souvětí (konstrukty) obsahovat mezičárí (konstituentů), tím kratší budou mezičárí v průměrném počtu znaků, aby se příliš nenarušila celková délka odstavce (strofy). Jelikož však jednotka souvětí není natolik graficky omezená jako jednotka odstavce a počet znaků v mezičárí může být

variabilnější než počet mezičárí uvnitř souvětí, neprojevil se pravděpodobně na této jazykové úrovni vztah mezi délkami konstruktů a konstituentů tak výrazně, jako na vyšší úrovni U_1 . Zatím je třeba dalších výzkumů v této oblasti, lze však opět uvažovat o tom, že grafická omezení textu mají významný vliv na projev MAZ.

Na jazykové úrovni U_3 již výsledky tak pozitivní nejsou. Parametr b je sice kladné číslo a proložená křivka je konvexní a klesající, avšak hodnota koeficientu determinace R^2 je pouhých 0,102. Na jazykové úrovni U_3 tedy sice lze na základě kladného parametru b prohlásit, že vztah mezi délkami konstruktů a konstituentů platí, avšak proložená regresní křivka nesedí na tato pozorování a výsledky tak celkově nelze označit za platné. Na této jazykové úrovni již neovlivňuje omezená délka odstavců délku konstituentů či ještě nižších jednotek a výsledek je podobně negativní jako na těžší jazykové úrovni ostatních analyzovaných textů. Možné důvody jsou rovněž podobné jako u ostatních analyzovaných textů: Může to být např. užití jednotky mezičárí, založené na čistě grafickém principu; příliš velký skok mezi jazykovou jednotkou mezičárí a jazykovou jednotkou znaku, mezi nimiž by mohly chybět další mezijednotky; proměnlivost délky mezičárí, zatímco jeho konstituenty jsou jednotkami s délkou v počtu prvků fixní; zahrnutí interpunkčních znamének do jednotky znaku; a rovněž také prokázaná velmi nízká variabilita znaků, co se týče počtu v nich obsažených prvků.

Na úrovni U_4 je parametr b kladné číslo, avšak koeficient determinace R^2 má hodnotu pouhých 0,275. Na jazykové úrovni znaky-prvky u tohoto textu tedy sice MAZ opět teoreticky platí, avšak vzhledem k hodnotě R^2 zatím nelze zobecněně říci, že čím více znaky obsahují prvky, tím méně tahů obsahují v průměru prvky. Důvody pro platnost MAZ s jen velmi nízkou hodnotou R^2 na jazykové úrovni U_4 jsou pravděpodobně totožné s důvody u analyzovaného textu č. 1: především tedy japonské užívání znaků *kandži* společně se znaky *kana* a dalšími v rámci jednoho textu. Pokud se nejedná o hierarchicky rovné jazykové jednotky, může to být jeden z důvodů pro daný výsledek na této jazykové úrovni.

Celkový závěr pro 2. analyzovaný text je, že na jazykových úrovních, u kterých lze předpokládat výrazná grafická omezení maximální délky textu, tj. především U_1 a U_2 , Menzerath-Altmanův zákon platí a jeho platnost je výrazně podpořena tvarem proložené křivky a poměrně vysokou hodnotou koeficientu determinace R^2 . Na těch úrovních, kde již podobná omezení nejsou, tj. U_3 a U_4 , Menzerath-Altmanův zákon sice

dle kladné hodnoty parametru b rovněž platí, avšak hodnota koeficientu determinace je velice nízká a výsledky tak nelze považovat za platné. Prozatím je tak třeba provést analýzu dalších textů v podobném stylu, než bude možné získané závěry více zobecnit.

5.3. Závěr pro analyzovaný text č. 3 – seminární práci

Pro ověření platnosti Menzerath-Altmanova zákona na japonské vysokoškolské seminární práci při užití grafické segmentace byla vybrána práce z oboru literatury o délce 5594 znaků dle zde zavedené definice znaku.

Na jazykové úrovni U_1 měl sice koeficient determinace R^2 poměrně vysokou hodnotu, avšak parametr b měl hodnotu zápornou a proložená křivka tedy nabyla rostoucího a konkávního tvaru. MAZ tedy není na této jazykové úrovni platný. Přesto však tento výsledek není příliš překvapivý a je v souladu s hypotézou. Možným důvodem pro takový výsledek může být např. zavedení striktně graficky definované jednotky mezičárí, která ignoruje některé syntaktické aspekty textu a není tudíž zcela přirozená. Dalším možným důvodem pak může být nedostačující počet vzniklých bodů x , kterých se objevilo pouze 9 a statistická relevance tak nemusí být dostatečně průkazná; je tedy třeba provést další analýzy textů v podobném stylu.

V případě jazykové úrovně U_2 je výsledek opačný: Menzerath-Altmanův zákon se ukazuje jako platný. U všech konstruktů, kromě jediného, se potvrdil vztah klesajících délek konstituentů s rostoucími délkami konstruktů. Parametr b je kladný a koeficient determinace R^2 vykázal nadmíru vysokou hodnotu 0,927. Na této jazykové úrovni mohl autor daného textu poměrně volně ovlivňovat délku souvětí i mezičárí a přizpůsobit ho tak přirozeně podobě, která mu vyhovovala. Seminární práce obsahovala několik přímých citací z jiné literatury, tyto však nebyly zdaleka tak početné jako např. dialogy v povídce, což mohlo přispět k lepší shodě s matematickým modelem. Zde tedy platí tvrzení, že čím více souvětí obsahují mezičárí, tím méně znaků obsahují v průměru mezičárí.

Jazyková úroveň U_3 měla výsledek opět negativní. Hodnota parametru b je záporná a hodnota koeficientu determinace R^2 je pouhých 0,084. Jedním z možných důvodů pro neplatnost MAZ na této jazykové úrovni může být užití graficky segmentované jednotky mezičárí. Dalším možným důvodem by mohl být příliš náhlý přechod mezi jazykovou

jednotkou mezičárí a jazykovou jednotkou znaku, mezi nimiž mohou chybět nějaké další mezijednotky, nicméně jak je patrné z dodatečného experimentu se 3. analyzovaným textem, ani několik pokusných analýz s přidáním (negraficky segmentovaných) jednotek Dlouhého a Krátkého slova mezi či namísto jazykových jednotek mezičárí a znaku nepřineslo žádné pozitivní výsledky. Rovněž fakt, že mezičárí je (oproti jeho konstituentům – znakům) jednotka s proměnlivou délkou, mohl přispět v neprospěch projevu MAZ. Nakonec i potvrzená velmi malá variabilita znaků, co se týče počtu v nich obsažených prvků, a také zahrnutí interpunkčních znamének do jednotky znaku, mohlo nepříznivě ovlivnit vztah mezi konstrukty a konstituenty.

Výsledky pro jazykovou úroveň U_4 jsou rozporuplné. Parametr b je tentokrát číslo kladné a koeficient determinace R^2 dosáhl hodnoty 0,485, tedy zhruba 50% shody. To není příliš vysoká hodnota a navíc průměrné délky konstituentů s narůstající délkou konstruktů střídavě klesají a stoupají a ke stabilnímu poklesu délek konstituentů dochází až u větších délek konstruktů. Avšak po vizuálním zhodnocení tvaru křivky a vzhledem k tomu, že hodnotu R^2 nelze označit ani za extrémně nízkou, nelze vyvodit jednoznačný závěr, že MAZ, resp. získaná data, na této jazykové úrovni neplatí. Pro posouzení, zda na této jazykové úrovni v seminární práci Menzerath-Altmanův zákon platí, či nikoliv, je tak třeba nejprve provést další experimenty a prozatím považovat výsledky první analýzy za přinejmenším nadějně.

Celkový závěr pro analyzovaný text č. 3 – seminární práci je, že výsledky na všech jazykových úrovních více méně odpovídají hypotézám: Platnost MAZ se potvrdila na úrovni U_2 , nepotvrdila na úrovních U_1 a U_3 a vyšla nejednoznačně na úrovni U_4 . Překvapivé jsou výsledky na jazykové úrovni U_2 , kde na rozdíl od povídky vyšla extrémně vysoká hodnota R^2 , což mohlo být zapříčiněno např. absencí velkého počtu dialogů či větším omezením, kladeným na výslednou podobu textu, nicméně takto velký rozdíl je udivující. Proto je nutné provést další experimenty s texty v podobném stylu, aby se ověřilo, zda se nejedná o pouhou výjimku, a dále pro zjištění, zda hodnota R^2 na úrovni U_4 bude u tohoto typu textů v průměru spíše vyšší či nižší než zdejších 0,5, aby tak bylo možné vyvozovat hodnocení obecnějšího rázu.

5.4. Závěr pro doplňující experimenty se 3. analyzovaným textem – seminární prací

V rámci doplňujícího experimentu byl na 3. analyzovaném textu otestován odlišný způsob segmentace na jazykové jednotky: zahrnutí negraficky segmentovaných jednotek slova (Dlouhého slova a Krátkého slova) definovaných Národním institutem pro japonský jazyk a lingvistiku (NINJAL). Analýza byla provedena především pro částečné ověření podezření, že mezi jazykovou jednotkou mezičárí a jazykovou jednotkou znaku může být příliš náhlý přechod, tj. je možné, že mezi nimi chybí další mezijednotky a mohly by se tak vytrácet některé vztahy důležité pro projev Menzerath-Altmanova zákona především na jazykových úrovních U_2 a U_3 . Z toho důvodu bylo na analyzovaném textu č. 3 – seminární práci – provedeno několik pokusných experimentů s rozdílnými způsoby segmentace, přestože tyto nové jazykové jednotky odporují základnímu kritériu čistě grafické segmentace textu. Jedná se tedy pouze o doplněk k původním stěžejním experimentům.

Posloupnost jazykových jednotek na jazykové úrovni U_{SD} je (odstavce) \Leftrightarrow **souvětí** \Leftrightarrow **Dlouhá slova** \Leftrightarrow **Krátká slova** \Leftrightarrow (znaky) \Leftrightarrow (prvky) \Leftrightarrow (tahy). Jazyková jednotka mezičárí tedy byla na této jazykové úrovni zcela nahrazena jazykovými jednotkami Dlouhé slovo a Krátké slovo. Hodnota parametru b na této úrovni je sice číslo kladné, avšak koeficient determinace R^2 ukázal pouze naprosto minimální shodu s matematickým modelem: 0,068; tvar proložené křivky je skoro rovný. Při užití tohoto způsobu segmentace, kdy jednotka mezičárí byla zcela nahrazena dvěma jednotkami slova, se tedy platnost MAZ na jazykové úrovni U_{SD} nepotvrdila. Ve srovnání s jazykovou úrovní U_2 (souvětí-mezičárí), na které se ukázala velmi vysoká míra shody s matematickým modelem ($R^2 = 0,927$), naopak došlo k poklesu hodnoty koeficientu determinace na téměř naprostou nulu. Jelikož však na této jazykové úrovni byla jazyková jednotka mezičárí (která byla vyšší než jednotka teoretického slova) nahrazena rovnou za slovo, došlo opět k relativně náhlému přechodu od jednotky souvětí rovnou k jednotce slova. Ignorování jednotky mezičárí či např. klauze apod., která by stála mezi souvětím a Dlouhým slovem, mohlo ovlivnit výsledek v neprospěch MAZ, pokud tedy vůbec přistoupíme na možnost analýzy odlišně segmentovaných jazykových jednotek.

Na jazykové úrovni U_{MD} je posloupnost jazykových jednotek následující:

(odstavce) \Rightarrow (souvětí) \Rightarrow **mezičáří** \Rightarrow **Dlouhá slova** \Rightarrow **Krátká slova** \Rightarrow (znaky) \Rightarrow (prvky) \Rightarrow (tahy). Mezi jazykové jednotky mezičáří a znaky tedy byly vloženy jazykové jednotky Dlouhé slovo a Krátké slovo. Na této jazykové úrovni je hodnota parametru b záporné číslo a hodnota koeficientu determinace R^2 je skoro nulová: 0,001. Při užití tohoto způsobu segmentace se tedy platnost MAZ na 3. analyzovaném textu na jazykové úrovni U_{MD} nepotvrdila. Podstatné zde je, že ani zavedení dalších mezijednotek mezi jazykové jednotky mezičáří a znaky nevedlo k pozitivnějším výsledkům a výsledky jsou tedy obdobné jako na jazykové úrovni U_3 .

(odstavce) \Rightarrow (souvětí) \Rightarrow (mezičáří) \Rightarrow **Dlouhá slova** \Rightarrow **Krátká slova** \Rightarrow **znaky** \Rightarrow (prvky) \Rightarrow (tahy) je posloupnost jazykových jednotek na úrovni U_{DK} . Mezi jazykové jednotky mezičáří a znaky byly na této jazykové úrovni opět vloženy jazykové jednotky Dlouhé slovo a Krátké slovo, analyzovala se však odlišná sekvence jazykových jednotek. Výsledek je však obdobný jako u předchozích pokusných úrovní. Přestože koeficient determinace R^2 nabyl nadějně hodnoty 0,562, body x byly v grafu rozprostřeny poměrně nahodile, hodnota parametru b je záporná a proložená křivka je tak mírně stoupající a konkávní. Ani při užití tohoto způsobu segmentace se platnost MAZ nepotvrdila a není možné prohlásit, že čím více Dlouhá slova obsahují Krátkých slov, tím méně znaků obsahují v průměru Krátká slova; výsledek je podobně negativní jako na jazykové úrovni U_{KZ} a U_3 .

Ani u čtvrté a poslední pokusné jazykové úrovně U_{KZ} nelze hovořit o platnosti Menzerath-Altmanova zákona. Posloupnost jazykových jednotek na této úrovni je (odstavce) \Rightarrow (souvětí) \Rightarrow (mezičáří) \Rightarrow (Dlouhá slova) \Rightarrow **Krátká slova** \Rightarrow **znaky** \Rightarrow **prvky** \Rightarrow (tahy), tzn., že mezi jazykové jednotky mezičáří a znaky byly navíc znovu vloženy jazykové jednotky Dlouhé slovo a Krátké slovo, avšak analyzovala se opět odlišná sekvence jazykových jednotek. Parametr b je i na této jazykové úrovni záporný; hodnota koeficientu determinace R^2 je 0,386. Užití tohoto způsobu segmentace se tedy opět nepotvrdila platnost MAZ na 3. analyzovaném textu; jazykové úrovně U_{KZ} i U_3 přinesly obdobně negativní výsledky.

Celkový závěr pro doplňující experimenty se třetím analyzovaným textem – seminární prací – je, že při užití daných způsobů segmentace, kdy bylo namísto graficky segmentované jednotky mezičáří, či společně s ní, užito negraficky segmentovaných jednotek Dlouhého slova a Krátkého slova, aby se ověřilo, zda v původní posloupnosti

jazykových jednotek není příliš náhlý přechod od jednotky mezičárí k jednotce znaku a nechybí tak mezi nimi nějaká další mezijednotka (kterou pravděpodobně může být slovo, jež v japonštině není jednoznačně definováno), se MAZ neukazuje jako platný na žádné z pokusně vytvořených úrovní. Na dvou jazykových úrovních naopak došlo k výraznému poklesu hodnoty koeficientu determinace R^2 a parametr b nabyl záporné hodnoty. Bylo by žádoucí provést experimenty s dalšími typy textů, ale je pravděpodobné, že kombinace striktně graficky a negraficky segmentovaných jednotek v mnoha případech nepovede k pozitivním výsledkům. Při případných budoucích experimentech tedy bude logicky žádoucí implementovat jazykové jednotky Dlouhé slovo a Krátké slovo spíše do jiné než grafické analýzy japonského textu.

5.5. Souhrnný závěr

Celkový závěr pro 1. analyzovaný text – povídku je, že žádná jazyková úroveň neprokazuje platnost MAZ při implementaci užitých způsobů grafické segmentace. U analyzovaného textu č. 2 – poezie se ukázalo, že na jazykových úrovních, u kterých lze předpokládat výrazná grafická omezení maximální délky textu, tj. především U_1 a U_2 , Menzerath-Altmanův zákon platí a jeho platnost je výrazně podpořena tvarem proložené regresní křivky a poměrně vysokou hodnotou koeficientu determinace R^2 . Na těch úrovních, kde již podobná omezení nejsou, tj. U_3 a U_4 , Menzerath-Altmanův zákon sice dle kladné hodnoty parametru b rovněž platí, avšak hodnota koeficientu determinace je velice nízká a výsledky tak celkově nelze prohlásit za platné. Pro analyzovaný text č. 3 – seminární práci platí, že výsledky na všech jazykových úrovních více méně odpovídají hypotézám: Platnost MAZ se potvrdila na úrovni U_2 , nepotvrdila na úrovních U_1 a U_3 a vyšla nejednoznačně na úrovni U_4 . Překvapivé jsou výsledky na jazykové úrovni U_2 , kde na rozdíl od povídky vyšla extrémně vysoká hodnota R^2 , což mohlo být zapříčiněno např. absencí velkého počtu dialogů či větším omezením, kladeným na výslednou podobu textu, nicméně takto velký rozdíl je udivující. Závěr pro doplňující experiment se třetím analyzovaným textem – seminární prací je, že při užití daných způsobů segmentace, kdy bylo namísto graficky segmentované jednotky mezičárí, či společně s ní, užito negraficky segmentovaných jednotek Dlouhého slova a Krátkého slova, aby se ověřilo, zda v původní posloupnosti jazykových jednotek není

příliš náhlý přechod od jednotky mezičáří k jednotce znaku a nechybí tak mezi nimi nějaká další mezijednotka (kterou pravděpodobně může být slovo, jež v japonštině není jednoznačně definováno), MAZ neprokazuje svou platnost na žádné z pokusně vytvořených úrovní. Na dvou jazykových úrovních naopak došlo k výraznému poklesu hodnoty koeficientu determinace R^2 a parametr b nabyl záporné hodnoty. Na úrovni U_{SD} sice parametr b nabyl hodnoty kladné, avšak při hodnotě R^2 nižší než 0,1 nelze výsledek, že MAZ platí, prohlásit za platný.

Jelikož se zatím jednalo o pilotní experimenty s japonskojazyčnými texty a každý z analyzovaných stylů byl zastoupen prozatím pouze jednou, je nutné provést další experimenty, aby bylo možné vyvozovat závěry obecnějšího rázu. Prozatím nelze jednoznačně tvrdit, že MAZ neplatí pro všechny japonské texty v povídkovém stylu. U textů typu poezie by bylo vhodné zjistit, zda se i na jiných takových textech prokáže platnost MAZ a vysoká hodnota koeficientu determinace na úrovních U_1 a U_2 , a naopak velice nízká na úrovních U_3 a U_4 , čímž by se mohl ověřit vliv grafických omezení textu na projev MAZ. V případě textu č. 3 je třeba dalších experimentů se stejným stylem pro ověření, zda mimořádně vysoká hodnota R^2 na úrovni U_2 nebyla pouhá výjimka, a dále pro zjištění, zda hodnota R^2 na úrovni U_4 bude u tohoto typu textů v průměru spíše vyšší či nižší než zdejších přibližně 0,5, aby tak bylo možné jednoznačněji rozhodnout o platnosti získaných výsledků. K dodatečnému experimentu s textem č. 3. je třeba poznamenat, že je pravděpodobné, že kombinace striktně graficky a negraficky segmentovaných jednotek ani u jiných textů nepovede k pozitivním výsledkům. Při případných budoucích experimentech tedy bude logicky žádoucí implementovat jazykové jednotky Dlouhé slovo a Krátké slovo spíše do jiné než grafické analýzy japonského textu.

Resumé v anglickém jazyce

The objective of this thesis was to verify the validity of the Menzerath-Altmann law applied to written texts in the Japanese language. The analysed sample texts were a short story, a poem, and a seminar paper. The methodology used in this analysis was partially based on a previous pilot study project from 2014–2015. The results of the short story analysis matched the results of the previous pilot study. The results of the poem analysis are conflicting; while the Menzerath-Altmann law has been proven valid for half of the language levels, it hasn't been so for the other half. However, these results may signify that the hypothesis that graphical text constrictions have a significant effect on the manifestation of the law might be true. As for the seminar paper, the Menzerath-Altmann law has been clearly manifested on only one of the four language levels. An additional experiment with different language units was conducted too, with negative results. However, as this is one of the very first studies of Japanese texts in relation to the Menzerath-Altmann law, it should be noted that it is necessary to analyse more texts before a generalised conclusion can be drawn.

Seznam pramenů

- www 1: MOTALOVÁ, Tereza. Gabriel Altmann. *Encyklopedie lingvistiky*, ed. Kateřina Prokopová [online]. Olomouc: Univerzita Palackého v Olomouci, 2014 [cit. 2015-08-15]. Dostupné z:
http://oltk.upol.cz/encyklopedie/index.php5/Gabriel_Altmann
- www 2: ALTMANN, Gabriel. Prolegomena to Menzerath's law. *Glottometrica* [online]. 1980, č. 2 [cit. 2015-01-16]. Dostupné z:
<https://googlegroups.com/group/ankara-linguistic-circle/attach/73d2b32da50082f1/PROLEG.pdf?part=2b-2x4adpDai8-4YM>
- www 3: *Džójó kandži hjó (常用漢字表)*. 2010.
http://www.bunka.go.jp/kokugo_nihongo/pdf/jouyoukanjihyou_h22.pdf
- www 4: ANDRES, Ján, Martina Benešová, Martina Chvosteková a Eva Fišerová. Optimization of Parameters in the Menzerath–Altmann Law, II. In: *Acta Universitatis Palackianae Olomucensis. Facultas Rerum Naturalium. Mathematica* [online]. 2014, č. 53, 2, s. 5–28. Dostupné z <http://eudml.org/doc/262150>
- www 5: Bunšó no kakikata, totonoekata: 5, danraku no cukurikata (文章の書き方・ととのえ方: 5, 段落の作り方). *Sanseidō Web Dictionary* [online]. 2014 [cit. 2014-01-04].
<http://www.sanseido.net/main/words/hyakka/howto/05.aspx>
- www 6: Introduction to the BCCWJ. *Center for Corpus Development, NINJAL* [online]. Dostupné z: http://pj.ninjal.ac.jp/corpus_center/bccwj/en/
- www 7: KOZAWA, Šunsuke (小澤俊介), Kijotaka UČIMOTO (内元清貴) a Jasuharu DEN (伝康晴). BCCWJ ni motodzuku čótan'i bunseki cúru Comainu. In: *Gengo šori gakkai dai nidžúukai nendži taikai ronbunšú (言語処理学会第20回年次大会論文集)* [online]. The Association for Natural Language Processing, 2014, s. 582–585. Dostupné z:
http://www.anlp.jp/proceedings/annual_meeting/2014/pdf_dir/P6-2.pdf

Seznam užitého softwaru

1. MATLACH, Vladimír, Martina BENEŠOVÁ a Martina CHVOSTEKOVÁ. Univerzita Palackého v Olomouci. *MA Studio* [software]. 2015. Verze MA Studio 2.11.0.0.
2. Microsoft Corporation. *Microsoft Office 2013* [software]. Dostupné z: <https://products.office.com/en-us/microsoft-office-2013>
3. The UniDic consortium. *Comainu* [software]. Verze Comainu-0.71. Dostupné z: <http://comainu.org>

4. The UniDic consortium. *UniDic* [software]. Verze unidic-mecab 2.1.2. Dostupné z: <https://osdn.jp/projects/unidic/>

Seznam literatury

1. BAREŠOVÁ, Ivona a DYTRTOVÁ, Monika. *Problematika české transkripce japonštiny a pravidla jejího užívání*. Olomouc: Univerzita Palackého, 2014. ISBN 978-80-244-4017-0.
2. BENEŠOVÁ, Martina. *Kvantitativní analýza textu se zvláštním zřetelem k analýze fraktální*. Olomouc, 2011. Disertační práce (Ph.D.). Univerzita Palackého v Olomouci.
3. BENEŠOVÁ, Martina a Denis BIRJUKOV. Application of the Menzerath-Altmann Law to Contemporary Written Japanese in the Short Story Style. In: TUZZI, Arjuna, Martina BENEŠOVÁ a Ján MAČUTEK (eds.). *Recent Contributions to Quantitative Linguistics*. Berlin, München, Boston: DE GRUYTER, 2015, s. 13–26. DOI: 10.1515/9783110420296-003. ISBN 9783110420296. Dostupné z: <http://www.degruyter.com/view/books/9783110420296/9783110420296-003/9783110420296-003.xml>
4. ČECH, Radek, Ioan-Iovitz POPESCU a Gabriel ALTMANN. *Metody kvantitativní analýzy (nejen) básnických textů*. 1. vyd. Olomouc: Univerzita Palackého v Olomouci, 2014. Qfwfq. ISBN 978-80-244-4044-6.
5. ČERNÝ, Jiří. *Dějiny lingvistiky*. 1. vyd. Olomouc: Votobia, 1996, 517 s. ISBN 80-858-8596-4.
6. GRZYBEK, Peter, Reinhard KÖHLER (eds.). *Exact Methods in the Study of Language and Text*. New York: Mouton de Gruyter, 2007, 767 s. ISBN 978-311-0193-541.
7. HABEIN, Yaeko S. *Decoding Kanji: A Practical Approach to Learning Look-Alike Characters*. 1. vyd. Tokio: Kodansha International, 2000. ISBN 978-477-0024-985.
8. HŘEBÍČEK, Luděk. Contextual Word Prominence. In: GRZYBEK, Peter a Reinhard KÖHLER (eds.). *Exact Methods in the Study of Language and Text*. New York: Mouton de Gruyter, 2007, s. 237–243. DOI: 10.1515/9783110894219.237. ISBN 9783110193541.
9. HŘEBÍČEK, Luděk. *Vyprávění o lingvistických experimentech s textem*. Praha: Academia, 2002. 195 s. ISBN 80-200-0973-6.
10. JINDŘICHOVÁ, Lucie. *Srovnání reformy čínského písma v ČLR a Japonsku*. Olomouc, 2014. Bakalářská práce (Bc.). Univerzita Palackého v Olomouci. Filozofická fakulta.

11. JOŠIMOTO, Banana (吉本ばなな). *Tokage (とかけ)*. Tokio: Shinchōsha, 2000. ISBN 978-410-1359-120.
12. KAMADA, Tadaši (鎌田正) a Torataró JONEJAMA (米山寅太郎). *Shin Kangorin Jórei Purasu (新漢語林用例プラス)*. Tokio: Taishūkan Shoten, 2004. ISBN 44-690-3162-3.
13. KELIH, Emmerich. Parameter interpretation of the Menzerath law: evidence from Serbian. *Text and Language: Structures, Functions, Interrelations : Quantitative Perspectives*. Praesens, 2010, 71–79.
14. KOIKE, Seidži. *Úvod do gramatiky moderní japonštiny*. 1. vyd. Brno: Tribun EU, 2010, 237 s. ISBN 978-807-3990-503.
15. KRAEMEROVÁ, Alice. *Úvod do japanologie: jazyk a literatura*. 1. vyd. Olomouc: Univerzita Palackého v Olomouci, Filozofická fakulta, 2000, 138 s. ISBN 80-244-0063-4.
16. MATOUŠKOVÁ, Lenka a Tereza MOTALOVÁ. *An Application of the Menzerath-Altman Law to Contemporary Written Chinese*. 1. vyd. Olomouc: Univerzita Palackého v Olomouci, 2014. Qfwfq. ISBN 978-80-244-4221-1.
17. MEYER, Peter. Two semi-mathematical asides on Menzerath-Altman's law. In: GRZYBEK, Peter a Reinhard KÖHLER (eds.). *Exact Methods in the Study of Language and Text*. New York: Mouton de Gruyter, 2007, s. 449–460. DOI: 10.1515/9783110894219.449. ISBN 9783110894219.
18. MOTALOVÁ, Tereza a Lenka SPÁČILOVÁ. *Aplikace Menzerath-Altmanova zákona na současnou psanou čínštinu*. Olomouc, 2013. Diplomová práce (Mgr.). Univerzita Palackého v Olomouci. Filozofická fakulta.
19. MOTALOVÁ, Tereza, Lenka SPÁČILOVÁ, Martina BENEŠOVÁ a Ondřej KUČERA. An Application of the Menzerath-Altman Law to Contemporary Written Chinese. *Czech and Slovak Linguistic Review*. 2013, 3 (1), 22–53. ISSN 1805-1502.
20. NICÚ, Nobuko (二通信子) a Fudžiko SATÓ (佐藤不二子). *Rjúgakusei no tame no ronritekina bunšó no kakikata (留学生のための論理的な文章の書き方)*. Tokio: 3A Corporation, 2004. ISBN 48-831-9257-1.
21. POE, Edgar Allan. *Ógarasu (大鴉)*. Macho Pubhouse, 2013. ISBN 978-80-89674-94-7. Dostupné z: <https://itunes.apple.com/jp/book/da-ya/id806435887?mt=11>
22. PRÜN, Claudia. Validity of Menzerath-Altman's law: Graphic representation of language, information processing systems and synergetic linguistics. *Journal of*

Quantitative Linguistics. 1994, 1(2), 148–155. DOI: 10.1080/09296179408590009. ISSN 0929-6174.

23. ROUKK, Maria. The Menzerath-Altmann law in translated texts as compared to the original texts. In: GRZYBEK, Peter a Reinhard KÖHLER (eds.). *Exact Methods in the Study of Language and Text*. New York: Mouton de Gruyter, 2007, s. 605–610. DOI: 10.1515/9783110894219.605. ISBN 9783110193541.
24. SATÓ, Sawako (佐藤沙和子). "Šárokku Hómuzu" širízu ni oite sakuša Doiru wa naze Hómuzu no seikaku wo hidžó ni reisei de džišinka no ronriteki ningen ni settei šita no ka (『シャーロック・ホームズ』シリーズにおいて作者ドイルはなぜホームズの性格を非常に冷静で自信家の理論的人間に設定したのか). 2012. Seminární práce. Utsunomiya University.
25. SPAHN, Mark, Wolfgang HADAMITZKY a Kimiko FUJIE-WINTER. *Kan-Ei Jukugo Ribāsu Jiten*. Shohan. Tokio: Hatsubai Kinokuniya Shoten, 1989, xviii, 1669 s. ISBN 48-169-0828-5.
26. STALPH, Jürgen. *Grundlagen einer Grammatik der sinojapanischen Schrift*. Wiesbaden: O. Harrassowitz, 1989. ISBN 3447029447.
27. ŠVARNÝ, O. et al. *Úvod do hovorové čínštiny: Příručka pro vys. šk. 2*. Praha: SPN, 1967.
28. VOCHALA, Jaromír, Miroslav NOVÁK a Vladimír PUCEK. *Úvod do čínského, japonského a korejského písma, I. díl: vznik a vývoj*. 1. vyd. Praha 1: Státní pedagogické nakladatelství, 1975, 134 s.
29. ZÁDRAPA, Lukáš a Michaela PEJČOCHOVÁ. *Čínské písmo*. vyd. 1. Praha: Academia, 2009, 297 s. Orient (Academia), sv. 5. ISBN 978-802-0017-550.

Příloha

Analyzovaný text č. 3 – seminární práce (stat')⁷³

シャーロック・ホームズシリーズは世界で最も有名な探偵小説の 1 つであると言っても過言ではないだろう。いまだなお、シャーロキアンと呼ばれるこの作品の根強いファンが世界中に存在する。また最近ではガイ・リッチー監督により映画化されたり、イギリス BBC 放送により現代版シャーロック・ホームズが製作されたりしており人気は衰えていないと言える。今回は、最終目標としてシャーロック・ホームズの原文と他のヴァリエーション作品や推理小説との比較、考察を行うために、その前の段階としてシャーロック・ホームズの原文の分析を行う必要があると判断し、原文の分析ということに焦点を当てて研究を行った。特に今回は主人公ホームズの特異な性格、つまり彼の相棒ワトソンが表現しているように、文学や哲学、天文学の知識がゼロで、毒物や地質学や化学、通俗文学に関しては深遠な知識を持っており、自信家で時に道徳心が欠けているととれるほどの理論的思考で事件を解決していくという彼の特徴は何が原因なのかを分析していくことで、作者であるコナン・ドイルがどのような思想を持って、ホームズという人物を作り出したのか、またドイルに影響を与えたものは何であったのかを解明し、他の作品との比較前に原文の理解を深めたい。方法として、先行研究の文献をもとに原文を抜粋し、ホームズの性格の分析と比較を行った。

使用テキスト

コナン・ドイル 延原謙訳 新潮文庫

- ・ 『シャーロック・ホームズシリーズ 緋色の研究』 1886-1887 執筆 1953 年延原訳初版 使用テキストは 2011 年版
- ・ 『シャーロック・ホームズの冒険 ボヘミアの醜聞』 1891 年ストランド誌発表 1953 年延原訳初版使用テキストは 2011 年版
- ・ 『シャーロック・ホームズの叡智 ライゲートの大地主』 1893 年ストランド誌発表 1955 年延原訳初版 使用テキストは 2010 年版

舞台は 19 世紀末ヴィクトリア朝のイギリス、ロンドン。第 2 次アフガン戦争下(1878-1880)に軍医としてアフガニスタンに派遣されていたジョン・ワトソン博士は負傷し、祖国イギリスに帰り新たな下宿先を探していた。ワトソンはロンドンで偶然再会した昔の友人の紹介をきっかけに顧問探偵をしている主人公シャーロック・ホームズと下宿の家賃を折半するために共同生活を送ることとなる。(『緋色の研究』)短編集が多いシャーロック・ホームズシリーズは話の主な流れがパターン化している。ある国の王室の人間や家庭教師の女性など、社会的身分が様々な依頼人が奇妙な事件に巻き込まれて状況に困惑し、ホームズの助けを求めて彼の下宿先を訪ねてくる。それらの事件をホームズが助手のワトソンを伴い解決していくという流れである。ストーリー・テラーはワトソンであり、ホームズの優れた事件解決の手法や推理論を身近で見た感想や、ホームズがいかに自信家であるかを、主観を交えながら語っているという設定である。

作者はコナン・ドイル (Sir Arthur Conan Doyle 1859-1930) であり、彼はアイルランド移民家庭の長男としてスコットランドのエディンバラに生まれる。エディンバラ大学医学部を卒業。その後ロンドンで開業するが上手いかず、家計の足しにするためシャーロック・ホームズシリーズの第 1 作目である長編『緋色の研究』を執筆し発表した。世間の反応は芳しくなかった。その後、眼科医に転身するもそれも上手いかず、執筆する時間だけがだいに余ってしまった。そこで彼は当時としては新しい手法である、1 人の主人公を続けて登場させる短編連載物、今日という読み切り連載という手法を編み出し、シャーロック・ホームズシリーズを続けて月刊誌『ストランド・マガジン』に投稿し始める。

⁷³ (Sató 2012)

長編のものであると1章、2章を別々の冊子に掲載することになり、もし読者が1号でも読み落としてしまったらそれ以上は読んでもらえなくなる。また、1回きりの短編のものであると、当たり外れが大きく、安定した収入を望むことができないので彼はこういった形式をとった。その後、1週間に1編のスピードで『ボヘミアの醜聞』、『赤毛連盟』、『花婿失踪事件』、『ボスコム谷の秘密』等を書き上げこれらが爆発的人気を得る事となった。晩年は死者との霊的交信をはかるという心霊主義に傾倒、1930年7月死去した。(情報元：河村幹夫 『コナン・ドイル—ホームズ・SF・心霊主義』 講談社 1991)

探偵はホームズの天職と言っても良い。なぜならば、彼は初めて出会った人物の職業を見抜けるほどの鋭い観察眼と洞察力という探偵に必要な才能を備えているからだ。ドイル本人はホームズにはモデルとなった実在の人物が存在すると発言している。ドイルがエディンバラ大学で医学を学んでいた時代に出会ったベル教授という人物である。ベル教授は「医者としての患者の見立て、また症状だけでなく職業も性格も見抜くべきだ」(河村 1991 p40)と学生に指導していた。ここにドイルが自伝で書き残しているベル教授に関する有名な逸話がある。

ある日、普通の服を着た男が外来患者としてベル教授の前に立った。じっと見ていたベル教授は「君は軍隊勤務をしたな」と尋ねた。

「はい、さようで」

「除隊後、まだ間がないね」 「その通りです」

「スコットランド高地連隊だね」 「はい、そうです」

「任官将校ではなかったね」 「その通りです」

「西インドのバルバドス諸島に駐屯していたね」 「仰る通りです」

「さて、諸君」とベル教授は学生たちに説明する。

「この患者は立派な感じの人だったが帽子をとらなかった。軍隊では脱帽しないからだ。しかし除隊してから長くたっていたら、普通の市民の作法通り帽子をとったはずである。彼の態度には威厳があり、明らかにスコットランド人である。バルバドスについては、この患者の症状は象皮病であるが、これは西インド諸島に特有のものであり、我が英国で生じる病ではない」(河村 1991 p40)

このようにモデルになったベル教授と同様、ホームズもシャーロック・ホームズシリーズの第1巻の『緋色の研究』で初めて出会ったワトソンと握手をただけで彼がアフガニスタンで従軍していた事を言い当てている。またホームズは「自己の前に展開してくるものを、精確かつ組織的に検討することによって、観察力に富む人間がいかにも多くのことを学びうるものであるか」(『緋色の研究』p32)という発言をしている。彼は目の前の物を良く観察、分析した上で、その途中の推理の過程を省いて相手に伝えるため、まるでホームズが手品を人々に見せているような印象を抱かせるのである。実際に、ホームズはどうしてワトソンがアフガニスタンで従軍していた事がわかったのかという解説をしている。

ここに医者タイプで、しかも軍人風の紳士がいる。すると軍人にちがいない。顔は真っ黒だが、黒さが生地でないのは、手首が白いのでわかる。してみると熱帯地がえりなのだ。艱難をなめ病気で悩んだことは、憔悴した顔が雄弁に物語っている。左腕に負傷している。動かしがたがぎこちなくて不自然だ。

わが陸軍の軍医が艱難をなめ腕に負傷までした熱帯地はどこだろう？むろんアフガニスタンだ。一と、これだけの過程を終わるには1秒も要しなかった。それで僕はそれを言ったら、君は驚いたというわけさ。(『緋色の研究』p36)

また、ホームズは彼の推理において、ある事象を客観的に観察し、それを理論的に思考する事が何よりも大切であると位置づけている。そういった彼の考えは作品の中で度々自身の発言で表現されている。例えば、『シャーロック・ホームズの冒険 ボヘミアの醜聞』の中でホームズは、差出人不明の手紙を見て奇妙だと言うワトソンに対して「まだ判断材料が1つも無い。資料もないのに、ああこうだと理論的な説明をつけようとするのは、大

きな間違いだよ。人は事実のほうを知らず知らず曲げがちになる。」(p14)と注意している。また、ホームズは『シャーロック・ホームズの叢書 ライゲートの大地主』の中で、「探偵術では、数多くの事実の中から、果たしてどれとどれが偶然の事柄で、どれが必然の事柄であるかを判別し得る能力が、もっとも重要です。この能力に欠けているときは、精力の浪費となり、注意力は散漫になって集中されません。」(p121)と述べている。ホームズは確固とした理由に裏打ちされない推測や結論への飛躍を嫌う。彼は現場に出ると虫眼鏡を持ち、現場を犬のように駆けずり回り、物的証拠品を集め自分の科学研究に基づいて結論を下す。実際、彼は様々な場所の土の質の違いを調べ上げ記憶しワトソンの靴の裏に付いている土を見分け、彼がどこへ出かけてきたかを当ててみせている。また数多くの葉巻の灰を調べ、現場の遺留品と照合するという事もしている。このホームズの几帳面かつ冷静な性格はホームズのモデルとなったベル教授だけの影響ではなく Doyle 自身の思想や経験が反映されていると考えられる。その根拠として Doyle が 1900 年に Strand Magazine のインタビューを受けた際の発言があげられる。

私が初めて探偵を思いついた時—あれは 1886 年頃のことでしたか—いくつも探偵小説を読む経験を持っていましたが、何というくだらぬもの(これでも穏やかな言い方をしたつもりですよ)かと思いました。なにしろミステリーの解決に当って、作者はいつも偶然に頼っていたのですからね。これはフェアプレイではないと思いました。だって、探偵が成功するためには自分自身の頭に頼るべきで、必ずしも実人生で起こるとは限らない偶発事なんかに頼るべきではないのですからね。(ノウン 1988 p8)

Doyle は、上記の発言の通り、現実ではあり得ない都合のよい偶然に頼り事件を解決する探偵小説を嫌った。だから彼は、ホームズの性格を少々道徳に欠けるとも思えるほど実用主義で理論的思考をする人間として作り上げた。しかし様々な地域の土の種類を調べ記憶していたり、膨大な種類のある葉巻を研究していたりという事は現実的ではないという指摘もあるだろう。しかし Doyle が重点を置いたのは、事件が発生してから探偵が解決に導くまでのプロセスの中で犯人につながる手掛かりを理論的にかつ客観的に収集していくという事であった。また、緻密な証拠集めと科学的分析を行うという事について Doyle が死去(1930)する少し前に、『ホームズ全集』の序文にて、こう述べている。

古い流派の探偵が文学で描かれる時には、解決に到達するに当たって実に安易な方法をとっていたので、これに反発して『緋色の研究』を書いたのである。私は医学における診察術の厳しい訓練を受けた経験を持っていたので、同じ厳密な観察と推理の方法を犯罪解明に応用すれば、もっと科学的なシステムを確立できるのではないかと思った。(ノウン 1988 p9)

これらの 2 つのインタビューから、Doyle がいかに徹底して厳密な推理と論理にこだわっていたかが読み取ることができる。これらの強いこだわりがホームズの特異な性格を作り出したと言えるのではないだろうか。また自分が書く以前の探偵小説に対して「くだらない」や「実に安易な方法をとっていた」と評価していることから、Doyle は自分が主張する推理や理論の手法に自信を持っていたという事が読み取れる。また彼の影響を受けたキャラクターであるホームズの強い自信の誇示はワトソンに「傲慢ちき」とまで評されている。『緋色の研究』の中でのホームズの次の発言を例にあげたい。

「近頃は犯罪も犯人もさっぱりなくなった」彼は不平そうにいった。「われわれの職業に頭脳があるということが、何の役に立つのか？僕は有名になれるだけの頭脳を持っていることを自分で良く知っている。過去現在にわたって、犯罪捜査に関して僕ほどの研究を積み、また僕ほどの天分をもつものは、ひとりだっていやしない。しかもその結果はどうか？腕をふるおうにも、犯罪が 1 つもないのだ。」(河村 1991 p38)

ホームズのこの自信過剰な性格、またそのもととなったドイルの性格は何に影響を受けたのか。それはドイルが子供時代の母親メアリの厳しい教育に起因すると考えられる。ドイルの母親は彼に期待をかけ、酒におぼれていた自分の夫と同じ人生を歩ませないようにと厳しくしつけをしていた。具体的に彼の母親はドイルに「物心ついたころから中世騎士の英雄譚とか紋章のきまりとか親の家系のいわれとかを徹底的にたたきこんだ。」また、「騎士は強くあらねばならない。メアリはアーサーが近所の子供と喧嘩して負けて泣いて帰ると、叱りとばしてまた喧嘩に行かせた。」これによって「ドイルの負けず嫌いの性格、誤りをなかなか認めようとしないで、がんばり通そうとする頑固な態度はすでにこの頃に形成されていた」(河村 1991 p32)という。ドイルの母親の存在がドイルの自尊心を強くしたのであろう。

ホームズの初見でその人の性質を当てるとい推理のスタイルはドイルが大学時代に出会ったベル教授に基づいており、どこまでも徹底的に理論的分析に基づく推理論はドイル以前の探偵小説に反発するという形、つまり反動として冷徹さが表現されている。またホームズの自信過剰ともとれる性格は作者ドイルが幼少時代に受けた母親からの教育がこの作品にまで影響していると言えるのではないか。

物語の登場人物を詳しく分析していくと作者の育った環境や作者の持つ思想や当時の風潮、時代背景を知ることができる。今回は主人公ホームズの推理論と性格のみの分析を行ったが、ホームズにコカインの常習癖があることや他の登場人物が多国籍であることをより深く分析していく事でヴィクトリア朝時代の当時の文脈に基づいてより作品を深く理解する事ができるだろう。今回は、他のヴァリエーションや推理小説との比較前段階における原文の分析であったが、この分析を詳しくおこなう事で他の作品と比べた時に時代の文脈の中での解釈の違い等、深いレベルでの比較ができるのではないか。