

UNIVERZITA PALACKÉHO V OLOMOUCI

Filozofická fakulta

Katedra asijských studií

BAKALÁŘSKÁ DIPLOMOVÁ PRÁCE

Analýza grafické struktury čínských znaků v odborném textu

Analysis of graphical structure of chinese characters in a scientific text

Olomouc 2017, Dobromila Dvořáková

Vedoucí diplomové práce: Mgr. Tereza Slaměňíková Ph.D.

Podklad pro zadání BAKALÁŘSKÉ práce studenta

PŘEDKLÁDÁ:	ADRESA	OSOBNÍ ČÍSLO
DVOŘÁKOVÁ Dobromila	Nálepkovo náměstí 934/17, Ostrava - Poruba	F13595

TÉMA ČESKY:

Analýza grafické struktury čínských znaků v odborném textu

TÉMA ANGLICKY:

Analysis of graphical structure of Chinese characters in a scientific text

VEDOUCÍ PRÁCE:

Mgr. Tereza Slaměniková, Ph.D. - ASH

ZÁSADY PRO VYPRACOVÁNÍ:

Čínské znakové písmo se vyznačuje komplikovanou grafickou strukturou, kterou je možné rozdělit do tří úrovní, tah, prvek a celý znak. Jednotlivé prvky jsou pak ve znaku uspořádány do tří základních grafických struktur, horizontální, vertikální a centrální. Z této kategorizace jsou vyjmuty znaky jednoduché, které netvoří žádnou strukturu a budou v práci vyčleněny do samostatné skupiny. Existují statistiky popisující strukturní složení znaků, které udávají počty tahů, prvků a zastoupení jednotlivých struktur v rámci znakového inventáře moderní čínštiny. Tématem mé bakalářské práce je analýza grafické struktury čínských znaků v odborných textech týkajících se chemie. Cílem této práce je zjistit počet tahů, prvků a četnost výskytů jednotlivých strukturních typů znaků v daném korpusu a výsledky této analýzy pak srovnat s již provedenými statistikami o četnosti výskytu jednotlivých struktur v obecném čínském textu, jež jsou obsaženy ve znakovém inventáři slovníku "Hanzi xinxi zidian - HXZ - Informační slovník čínských znaků". Analyzovaný korpus textu bude obsahovat přibližně 100 000 znaků a skládá se z několika odborných textů na různá témata, které se všechny týkají chemie a obsahují především termíny z chemického názvosloví. V první části práce se budu věnovat přiblížení základních teorií týkajících se čínských znaků, jako jsou například typologie struktur, způsob vzniku a etymologie znaku, zjednodušování atd. V druhé části práce vytvořím soupis znaků z analyzovaného korpusu a za pomoci HXZ v něm budu zkoumat tři stěžejní aspekty, počet tahů, počet prvků a typ struktury. Pak budu prezentovat výsledky analýzy zmiňovaných tří úrovní grafické struktury znaku. V poslední části práce porovnam výsledky svého výzkumu se znakovým inventářem slovníku HXZ a vyvodím závěr.

SEZNAM DOPORUČENÉ LITERATURY:

- Hanzi Xinxi Zidian. Beijing: Kexue Chubanshe, 1988, 1346 s.
SEHNAL, David. "Čínské znakové písmo, jeho povaha a vývoj". In Z myšlenek a představ Žluté země. Brno: Moravské zemské muzeum, 2002, s. 145-160.
SLAMĚNÍKOVÁ, Tereza. Ideogramy v moderní čínštině. Olomouc: Univerzita Palackého 2013a, 164 s.
UHER, David. Hanská grammatologie. Olomouc: Univerzita Palackého 2013, 334 s.
VOCHALA, Jaromír. Chinese Writing System: Minimal Graphic Units. Praha: Univerzita Karlova 1985, 143 s.
ZÁDRAPA, Lukáš a Michaela Pejčochová. Čínské písmo. Praha: Academia, 2009. 297 s.

Podpis studenta: Dvořáková

Datum: 26.4.2017

Podpis vedoucího práce: Juráček

Datum: 26.4.2017

Prohlašuji, že jsem diplomovou práci vypracovala samostatně a uvedla v ní veškeré použité prameny a literaturu.

V Olomouci dne Podpis

Anotace

Příjmení a jméno autora: Dvořáková Dobromila

Název katedry a fakulty: Katedra Asijských studií, Filozofická fakulta

Název diplomové práce: Analýza grafické struktury čínských znaků v odborném textu

Vedoucí diplomové práce: Mgr. Tereza Slaměnková Ph.D

Počet znaků: 91 937

Počet stran: 59

Počet zdrojů: 19

Klíčová slova: čínské znaky, grafická struktura, tah, prvek, strukturní typ

Tématem této bakalářské práce je analýza tříúrovňové grafické struktury čínských znaků v odborných textech týkajících se chemie a fyziky. Cílem této práce je zjistit počet tahů, prvků a četnost výskytů jednotlivých strukturních typů znaků v daném korpusu a výsledky této analýzy pak srovnat s již provedenými statistikami, jež jsou obsaženy ve znakovém inventáři *Informačního slovníku čínských znaků* 汉字信息字典 *Hànzì Xīnxi Zìdiǎn*. Analyzovaný korpus textu čítá 86 745 znaků a skládá se z populárně naučných textů na různá témata z oblasti chemie a fyziky. V první části práce se věnuji přiblížení základních teorií týkajících se čínských znaků, jako jsou typologie struktur, etymologie znaku, zjednodušování atd. V druhé části práce byl vytvořen frekvenční soupis znaků z analyzovaného korpusu a za pomoci slovníku HXZ v něm byly zkoumány tři stěžejní aspekty, počet tahů, počet prvků a strukturní typ. V poslední části práce jsem porovnála výsledky svého výzkumu se znakovým inventářem slovníku HXZ a vyvodila příslušné závěry. V rámci mé analýzy se prokázaly rozdílné hodnoty zastoupení tahů, prvků a struktur mezi souvislým textem a znakovým inventářem. Hodnoty počtu tahů a prvků znakového inventáře jsou obecně vyšší, zatímco hodnoty vycházející ze souvislého textu poukazují na větší grafickou úspornost.

Chtěla bych zde poděkovat vedoucí mé bakalářské práce Mgr. Tereze Slaměnikové Ph.D za velkou ochotu a vstřícný přístup při vedení mé práce a také za mnoho skvělých postřehů a tipů, které mi pomohly vyřešit problémy, jež se během zpracování mé bakalářské práce vyskytly.

OBSAH

Anotace.....	5
Seznam tabulek a grafů.....	9
Seznam zkratk.....	10
Ediční poznámka.....	11
ÚVOD.....	12
1. TEORIE.....	15
1.1. Kvantitativní lingvistika.....	15
1.2. Stručná historie čínského písma.....	16
1.3. Reforma znaků.....	18
1.4. Konstrukční typy čínských znaků.....	19
1.5. Grafické členění čínského znaku.....	21
1.5.1. Tahy.....	22
1.5.2. Prvky.....	24
1.5.3. Strukturní typy.....	25
2. ANALÝZA.....	28
2.1. Korpus.....	28
2.2. „Informační slovník čínských znaků“ 汉字信息字典 hànzi xìnxī zìdiǎn.....	30
2.3. Zpracování analýzy.....	31
2.3.1. Frekvenční tabulka znaků.....	32
2.3.2. Analytická tabulka struktur.....	34
3. VÝSLEDKY ANALÝZY.....	35
3.1. Oborově specifické znaky.....	35
3.2. Tahy.....	38
3.2.1. Ukázka frekvence zastoupení počtu tahů v textu.....	40
3.3. Prvky.....	41
3.3.1. Ukázka frekvence zastoupení počtu prvků v textu.....	43
3.4. Struktury.....	44

3.4.1. Ukázka frekvence zastoupení strukturních typů v textu.....	46
3.5. Shrnutí výsledků.....	47
3.5.1. Srovnání korpusu se znakovým inventářem slovníku.....	49
4. ZÁVĚR.....	54
Resumé.....	56
SEZNAM LITERATURY.....	57
Seznam příloh.....	59

Seznam grafů a tabulek

Graf 1: Kvalitativní zastoupení počtu tahů	38
Graf 2: Kvantitativní zastoupení počtu tahů	39
Graf 3: Počet tahů ve znacích korpusu 1+2	41
Graf 4: Srovnání kvalitativního zastoupení počtu prvků	42
Graf 5: Kvantitativní zastoupení počtu prvků v korpusu 1+2.....	43
Graf 6: Kvantitativní zastoupení struktur.....	45
Graf 7: Srovnání počtu prvků kvalitativní a kvantitativní analýzy	49
Graf 8: Srovnání zastoupení počtu tahů korpus vs. inventář	50
Graf 9: Srovnání zastoupení počtu prvků korpus vs. inventář	51
Graf 10: Srovnání zastoupení strukturních typů korpus vs. inventář	52
Tabulka 1: Ukázka upraveného frekvenčního seznamu znaků.....	33
Tabulka 2: Relativní četnost výskytu strukturních typů v textu	34
Tabulka 3: Vybrané oborově specifické znaky	36
Tabulka 4: Srovnání kvalitativní a kvantitativní analýzy tahů	48

Seznam zkratk

HXZ *Informační slovník čínských znaků* 汉字信息字典 *Hànzì Xìnxī Zìdiǎn*

JS Jednoduchá struktura

HS Horizontální struktura

VS Vertikální struktura

CS Centrální struktura

RS Rámcová struktura

Ediční poznámka

Pro fonetický přepis čínských znaků používám oficiální čínskou transkripci 拼音 *pīnyīn*, kterou od ostatního textu odlišuji kurzívním fontem. Čínské znaky v rámci této práce jsou až na některé výjimky (v kapitole o etymologii) uvedeny ve zjednodušených variantách. Přímé citace jsou vyznačeny kurzívou a uvozovkami, neoficiální překlady uvozovkami bez kurzívy.

ÚVOD

Popularita čínské kultury je v posledních desetiletích na vzestupu v souvislosti s globalizací a ekonomickým rozvojem Číny. Téměř v každé zemi na světě můžeme narazit na čínské restaurace, obchody, výuku tradičních čínských bojových umění nebo na tradiční čínskou medicínu. Zvyšující se vliv Číny na mezinárodním poli se projevuje také ve stále narůstajícím zájmu o studium čínštiny, neboť v něm mnoho lidí spatřuje dobrou příležitost pro pracovní účely. Zatímco angličtina se vlivem globalizace již stala naprostým základem mezinárodní komunikace, čínština se v současnosti pozvolna dostává do popředí světového zájmu.

Znaky jsou neodmyslitelnou součástí čínského jazyka a v průběhu dějin přetrvaly navzdory mnoha pokusům o jejich obměnu, zjednodušení, ba dokonce i o zrušení. Jejich atraktivita spočívá především v množství a rozmanitosti forem, které tento grafický systém skýtá. Čínské písmo je mnohem komplexnější než písmo evropská, neboť v sobě obsahuje jak fonetickou, tak zároveň i významovou složku. Lze tedy na něj pohlížet z několika různých úhlů.

Moje práce se bude zabývat čínskými znaky pouze z hlediska jejich grafické stránky. Každý čínský znak se obecně dělí do tří základních úrovní na tahy, prvky a struktury. Tah je základní, zcela nejmenší složka, kterou musí znak obsahovat. Existuje množství různých typů tahů, které se kombinují a propojují do větších jednotek nazývaných prvky. Prvky v čínských znacích zpravidla již nesou buď fonetickou, nebo významovou informaci, případně její část. Jednotlivé prvky se spolu spojují na základě určitých pravidel do různých strukturních typů, ze kterých je tvořen finální znak.

Obecné statistiky udávají, že některé počty tahů, prvků a strukturní typy jsou v rámci inventáře moderní čínštiny velmi časté, jiné jsou spíše vzácné. Ve své práci se zaměřím na již zmíněné tři hlavní grafické aspekty čínského znaku, které budu analyzovat pomocí základních postupů kvantitativní lingvistiky ze dvou hledisek. Z hlediska kvalitativního, jež vytváří statistiku v rámci jedinečných znaků, které se v textu vyskytly a z hlediska kvantitativního, jež vytváří statistiku ze všech znaků, které se v textu opakovaně vyskytly.

Cílem mé práce bude vyhodnotit statistiku četnosti počtu tahů, prvků a strukturních typů ve znacích populárně naučného textu a porovnat její výsledky s již zmíněnou obecnou statistikou, kterou v této práci reprezentuje *Informační slovník čínských znaků* 汉字信息字典 *Hànzì Xìnxī Zìdiǎn* (zkratka HZX). Současné

grammatologické studie se při popisu vlastností znaků v moderní čínštině obvykle zaměřují zejména na popis znakového inventáře, zatímco já chci v této práci prokázat, že znaky v souvislém textu vykazují poměrně odlišné vlastnosti.

Předpokládám, že v rámci této analýzy se tedy vyskytnou rozdílné výsledky především mezi kvantitativními hodnotami četnosti výskytu jednotlivých aspektů v souvislém textu a kvalitativními hodnotami četnosti výskytu ve znakovém inventáři slovníku. Zároveň se také domnívám, že i přes specifické zaměření mého korpusu se jeho kvantitativní data nebudou nijak výrazně lišit od dat, týkajících se obecného textu, neboť všechny texty napříč stylistickým spektrem podléhají univerzálním gramatickým náležitostem, které jejich variabilitu z velké míry omezují.

Tato práce se skládá ze dvou hlavních částí, teoretické a analytické. V první části, zaměřené na teorii se budu nejprve stručně zabývat pojmy, týkajícími se kvantitativní lingvistiky, vzhledem k tomu, že celá má práce ze zákonitostí kvantitativní lingvistiky vychází. Dále krátce nastíním vývoj čínského písma a stěžejní reformy, které měly na jeho grafickou podobu největší dopad. Také zmíním tradiční kategorizaci čínských znaků, která s grafickým členěním znaků úzce souvisí, a především pak budu věnovat velkou pozornost strukturnímu členění znaků, jehož definování je velmi důležité pro pozdější analytickou část.

V druhé části práce, která se zaměřuje na samotnou analýzu, nejprve popíšu metodologii a výchozí materiály, se kterými jsem pracovala, tj. zvolený korpus a slovník HXZ. V další části pak v rámci jednotlivých kapitol rozpracuji analýzy tahů, prvků a strukturních typů. Nakonec své výsledky porovnam se statistikami, obsaženými ve slovníku HXZ a shrnu poznatky, které jsem ve své práci učinila.

Při zpracování teoretické části mé práce jsem vycházela především z česky psané literatury, vzhledem k tomu, že problematikou čínských znaků se v České republice podrobně zabývalo poměrně velké množství jazykovědců. Co se týče vývoje čínských znaků, nejvíce jsem využila publikace *Úvod do čínského, japonského a korejského písma I. Vznik a vývoj* od Jaromíra Vochaly, *Čínské znakové písmo, jeho povaha a vývoj* od Davida Sehnala a úvodní teoretický výklad *Učebnice čínských znaků I.* od Ondřeje Kučery. Informace týkající se kategorizace znaků jsem pak čerpala především z knih *Hanská grammatologie* od Davida Uhra, *Čínské písmo* od Lukáše Zádrapy a *Analysis of Chinese Characters* od George Wildera. Strukturní uspořádání znaků bylo zpracováno v publikacích *Chinese Writing system: Minimal graphic units* a *Úvod do čínského, japonského a korejského písma II. Praktický kurz* od Jaromíra Vochaly a *Ideogramy v*

moderní čínštině od Terezy Slaměnikové, ze kterých jsem nejčastěji vycházela. Při tvorbě tabulek a grafů v analytické části jsem se opírala zejména o knihy *Quantitative linguistics* od Marie Těšitelové a *Úvod do studia jazyka* od Jiřího Černého, jež mi zároveň posloužily při interpretaci mnou zjištěných dat a jejich aplikaci na obecné principy.

1. TEORIE

1.1. Kvantitativní lingvistika

Kvantitativní lingvistika se zabývá způsobem, jak zjistit častost výskytu jazykových jevů a zkoumá vzájemná pravidla jejich užití. V česky psané literatuře se kvantitativní lingvistikou podrobně zabývala Marie Těšitelová, která tuto jazykovědnou disciplínu definovala jako složku matematické lingvistiky, která zjišťuje kvantitativní data týkající se různých jazykových jevů a interpretuje tyto výsledky tak, aby byly pochopeny mechanismy jejich vzájemného fungování.¹

Tyto mechanismy jsou analyzovány z textových korpusů, neboť zákonitosti panující mezi jednotlivými jazykovými jevy jsou zpravidla rozpoznatelné až v rámci větších textových celků. Vždy je však otázkou, do jaké míry je daný korpus reprezentativním zdrojem. Michal Šulc ve své publikaci *Korpusová lingvistika* zmiňuje, že jazykový korpus je na jednu stranu přizpůsoben vyjadřovacími schopnostmi jednotlivých mluvčích, na druhou stranu však všechny promluvy podléhají stejným gramatickým náležitostem a dochází v nich k pravidelnému opakování jevů, a proto lze dle některých lingvistů považovat korpusy za dostatečně vypovídající.²

Z pohledu kvantitativní lingvistiky rozlišujeme dva základní pohledy na jazykové jednotky, kvalitativní a kvantitativní. Kvalita je určitá vlastnost předmětu, která jej činí jedinečným a nelze ji od dané věci odloučit bez toho, aby pozbyla svého významu. Na jazykové úrovni se tato kvalita spojuje s pojmem, označujícím konkrétní jedinečný objekt,³ jež se v zahraniční literatuře nazývá type „typ“. Pokud však rozdělujeme souvislý text, vznikne určité množství jednotek, ze kterých se skládal. Jednotka textu se nazývá token „token“. Ne všechny typy se v rámci textu vyskytují stejně často, takže ne všechna slova jsou v jazyce stejně užitečná, co se týče četnosti jejich využití.⁴ Frekvence neboli četnost výskytu je tedy další ze stěžejních pojmů kvantitativní lingvistiky, který je důležitý nejen pro statistické účely, ale také zejména při výuce cizích jazyků. Pokud se například žák

¹ TĚŠITELOVÁ, Marie. *Quantitative linguistics*. Praha: Academia, 1992, 256s. ISBN 80-200-0131-X. str. 11.

² ŠULC, Michal. *Korpusová lingvistika: první vstup*. Praha: Karolinum, 2000, 94 s. ISBN 8071848476. str. 14.

³ TĚŠITELOVÁ, Marie. *Quantitative linguistics*. str. 11.

⁴ *ibid.* str. 12.

učí cizímu jazyku, potřebuje z pravidla znát především slova, která se v rámci textu nejčastěji používají. Při tvorbě učebnic a učebních slovníků je tedy důležité klást důraz na aplikaci metod kvantitativní lingvistiky, aby materiály obsahovaly relevantní využitelnou slovní zásobu.⁵

Existují určitá pravidla, kterými se texty obecně řídí bez ohledu na jejich původ, jež byla vypořádána v rámci kvantitativních analýz korpusů z mnoha různých jazyků světa. V rámci těchto pravidel tedy univerzálně platí například, že nejfrekventovanější slova bývají obvykle nejjednodušší a jedná se o zejména gramatické ukazatele, spojky atd. Také je obecně platné, že nejčastější slovo v textu pokrývá přibližně 5 % a spolu s prvními deseti nejfrekventovanějšími slovy tvoří až okolo 20 % výsledného textu.⁶ V rámci mé následné analýzy ve druhé části práce se většina těchto obecně platných závěrů potvrdila také, což dále rozebírám v oddílech týkajících se zpracování korpusu.

1.2. Stručná historie čínského písma

Podle jedné z legend byly čínské znaky vytvořeny před více než 4500 lety, za vlády mýtického čínského Žlutého císaře 皇帝 *Huángdì* ve 27. století před naším letopočtem.⁷ Jeho dvorní historik 仓颉 *Cāng Jié* se údajně nechal inspirovat otisky zvířecích stop, které jej přivedly k myšlence sestavení systému symbolů, které by byly schopny reprezentovat jednotlivá slova stejně výstižně, jako dokáží zvířecí stopy reprezentovat své původce.⁸ První skutečně doložené historické nálezy písma v Číně pocházejí z období vlády dynastie 商 *Shāng* okolo 13. století př. n. l. Tato starodávná forma logografického čínského písma, která se ryla ostrým předmětem do kostí, se označuje jako nápisy na želvích krunýřích a hovězích lopatkách 甲骨文 *jiǎgǔwén*⁹ a sloužila především k zaznamenávání věšteb týkajících se důležitých událostí, ale také běžných každodenních záležitostí jako bylo počasí, nebo zemědělství.

⁵ TĚŠITELOVÁ, Marie. *Quantitative linguistics*. str. 16.

⁶ ČERNÝ, Jiří. Úvod do studia jazyka. 2. vyd. Olomouc: Rubico, 2008, 248 s. ISBN 978-80-7346-093-8. str. 149.

⁷ KUČERA, Ondřej a kol. Učebnice čínských znaků. Olomouc: Univerzita Palackého, 2005, 237 s. ISBN 80-244-1118-0. str. xi.

⁸ SLAMĚNÍKOVÁ, Tereza. Ideogramy v moderní čínštině. Olomouc: Univerzita Palackého v Olomouci, 2013, 164 s. ISBN 978-80-244-3909-9. str. 14.

⁹ ZÁDRAPA, Lukáš a Michaela Pejčochová. Čínské písmo. Praha: Academia, 2009, 297 s. ISBN 978-80-200-1755-0. str. 30.

Ve stejném období se také objevují nápisy na bronzech 金文 *jīn wén*, které sloužily jako prostředek pro zaznamenávání dvorních rituálů a událostí, během nichž byly odlity.¹⁰ Po rozpadu čínského území za vlády Východní Zhou na sedm navzájem nepřátelských států došlo k různým lokálním variacím tohoto písma. Tyto variace se obvykle hromadně označují jako větší písmo 大篆 *dàzhuàn*, které se rozšířilo v období válečných států¹¹ 战国 *zhànguó*, psané na hedvábí či na bambusových destičkách. Poté, co státu 秦 *Qín* povedlo znovu sjednotit říši ve 3. století př. n. l., byla provedena zásadní úprava tohoto písma. Vzniklo takzvané menší písmo 小篆 *xiǎozhuàn*, které nechal kodifikovat první čínský císař 秦始皇帝 *Qín Shǐ huángdì*. V rámci této reformy byla sjednocena forma, kterou se zapisovaly jednotlivé grafické prvky, a také byla ustanovena pevná pozice prvků ve znaku, což omezilo počet regionálních variací znaků¹² a stanovilo tak první kodifikovanou verzi čínského písma.

Za dynastie 汉 *Hàn* (206 př. n. l. – 220 n. l.), bylo vytvořeno takzvané úřednické písmo 隶书 *lìshū*.¹³ Toto písmo se původně používalo ve státní správě pouze mezi úředníky, později se rozšířilo do všech sfér a nevratně poškodilo strukturu tehdejších znaků. Nové znaky byly psány již pravidelnými hranatými a lomenými tahy avšak postrádaly původní ideografickou motivovanost.¹⁴ Současně vznikaly i neoficiální formy úsporného písma, kurzivní písmo 行书 *xíngshū* a trávové písmo 草书 *cǎoshū*, které ještě více oslabovaly významovou motivovanost znaků.¹⁵ Poslední oficiální historickou vývojovou fází čínského písma bylo vzorové písmo 楷书 *kǎishū*, jehož reforma dále upevnila změny v realizaci tahů a prohloubila další schematizaci znaků. Tato forma písma se používá od svého vzniku na konci dynastie Han v takřka nezměněné podobě až do dnešní doby.¹⁶

¹⁰ KUČERA, Ondřej a kol. *Učebnice čínských znaků*. str. xi.

¹¹ SLAMĚNÍKOVÁ, Tereza. *Ideogramy v moderní čínštině*. str. 15.

¹² UHER, David. *Hanská grammatologie*. Olomouc: Univerzita Palackého v Olomouci, 2013, 334 s. ISBN 978-80-244-3866-5. str. 99.

¹³ SEHNAL, David. „Čínské znakové písmo, jeho povaha a vývoj“. In *Z myšlenek a představ Žluté země*. Brno: Moravské zemské muzeum, 2002, s. 145-160. str. 26.

¹⁴ VOCHALA, Jaromír, Miroslav Novák a Vladimír Pucek. *Úvod do čínského, japonského a korejského písma I. Vznik a vývoj*. Praha: Státní pedagogické nakladatelství 1989a, 132 s. str. 22.

¹⁵ WILDER, George D., INGRAM James H.: *Analysis of Chinese Characters*. New York: Dover, 1974, 263 s. ISBN 0486230457. str. V.

¹⁶ SEHNAL, David. „Čínské znakové písmo, jeho povaha a vývoj“. str. 26.

1.3. Reforma znaků

S narůstajícím počtem obyvatel a zvyšující se administrativní zátěží na konci dynastie 清 *Qīng* vyvstala nutnost dalšího zjednodušení stávajících znaků, které byly často komplikované a většina obyvatelstva byla tudíž negramotná. Na přelomu 19. a 20. století se objevily první tendence k přechodu ze znaků na fonetickou formu písma.¹⁷ V tomto období vzniklo několik fonetických transkripcí čínštiny, z nichž nejvýznamnější byla transkripční abeceda Wade-Giles 韦氏拼音 *wéi shì pīnyīn*, která se uchytila mezi zahraničními sinology a dodnes se používá pro přepis znaků na Taiwanu.

V průběhu první poloviny 20. století pak vznikají také další fonetické abecedy, avšak snahy o zavedení latinky místo znaků nikdy nebyly přijaty čínským obyvatelstvem, proto se tyto abecedy neuchytily jako primární formát zápisu čínštiny.¹⁸ Nedlouho po ustanovení Čínské lidové republiky, roku 1954 je založen Čínský výbor pro reformu písma, jehož cílem je zjednodušení znaků, tvorba nové transkripční abecedy a prosazování obecné čínštiny po celé zemi.¹⁹ Tento výbor schválil reformu, která zavedla postupné zjednodušení (zejména snížení počtu tahů) asi 500 znaků a prvků, které byly od roku 1964 používány jako jediná oficiální varianta znaků jak samostatně tak ve znacích, složených z několika komponentů.²⁰ Bylo zároveň ustanoveno 54 komponentů, které jsou ve všech znacích zjednodušeny automaticky.²¹

Zjednodušování znaků se provádělo na základě několika postupů, jako například vypuštění jedné či více složek znaku (號 – 号), nahrazení složité složky jinou složkou (陽 - 阳), nahrazení znaku jeho kursivní variantou (爲 - 为), nahrazení znaku jiným jednodušším znakem (幾 – 几), vytvoření nového jednoduššího znaku (淚 - 泪).²² Zjednodušení znaků bylo přijato kladně a přineslo řadu pozitivních změn, především pak nárůst gramotnosti populace. Došlo však k nenávratnému zásahu do etymologie znaků, jejichž složky někdy zcela přestaly být funkčními ukazateli jak významu, tak výslovnosti. Dalším problémem zjednodušení je také fakt, že na Taiwanu se dodnes používají

¹⁷ SLAMĚNÍKOVÁ, Tereza. *Ideogramy v moderní čínštině*. str. 16.

¹⁸ VOCHALA, Jaromír, Miroslav Novák a Vladimír Pucek. *Úvod do čínského, japonského a korejského písma I. Vznik a vývoj*. str. 48.

¹⁹ SLAMĚNÍKOVÁ, Tereza. *Ideogramy v moderní čínštině*. str. 17.

²⁰ VOCHALA, Jaromír, Miroslav Novák a Vladimír Pucek. *Úvod do čínského, japonského a korejského písma I. Vznik a vývoj*. s. 51.

²¹ ŠVARNÝ, Oldřich. *Úvod do hovorové čínštiny*. Praha: Státní pedagogické nakladatelství, 1967, 458 s. str. 28.

²² ZÁDRAPA, Lukáš a Michaela Pejčochová. *Čínské písmo*. str. 168.

nezjednodušené formy znaků, což ztěžuje komunikaci samotným Číňanům mezi sebou a paradoxně je nutí učit se dvě různé formy jednoho znaku.²³

1.4. Konstrukční typy čínských znaků

Různí jazykovědci dělí čínské znaky do několika kategorií, podle jejich významové či zvukové motivovanosti. Již za dynastie Han vznikla vůbec první a zcela zásadní publikace, týkající se analýzy čínských znaků, o kterou se badatelé opírají prakticky dodnes. Jedná se o slovník *Výklad významu obrysových a rozboru struktury odvozených znaků* 说文解字 *Shuō Wén Jiě Zì*, který sestavil učenec 许慎 *Xǔ Shèn* na základě své perfektní znalosti čínské kanonické literatury.²⁴ Jednalo se o zcela jedinečnou sbírku, čítající okolo 10 000 znaků, které byly seřazeny podle významových kritérií svých determinativů (zde je determinativů ještě 534, později byly zredukovány na 214) a také byly znaky poprvé systematicky kategorizovány do šesti druhů 六书 *liù shū*, jež jsou uznávány i v dnešní době.²⁵

Jak již název tohoto díla napovídá, druhy znaků tak, jak je vymezuje, spadají do dvou větších celků na základě způsobu jejich vzniku. Původní znaky, které odkazují na jednoznačný význam, nazývá obrysové 文 *wén*.²⁶ Mezi ně zařazuje piktogramy 象形 *xiàngxíng* a symboly 指事 *zhǐshì*. Mezi odvozené znaky 字 *zì*, jež vznikly kombinováním obrysových znaků, patří ideogramy 会意 *huìyì* a fonogramy 形声 *xíngshēng*. Následují znaky, jež vznikly změnou v užívání, které se nazývají výpůjčky 假借 *jiǎjiè* a variety 转注 *zhuǎnzhù*.²⁷

Obrysové znaky 文 *wén*:

Piktogramy 象形 *xiàngxíng*, kterých je ve slovníku 364, schematicky vyobrazují objekty, které jsou označeny určitým morfémem. Mezi tyto znaky patří například slunce 日 *rì* a měsíc 月 *yuè*, jejichž podoba byla nejlépe viditelná na nápisech na kostech a dnes

²³ ZÁDRAPA, Lukáš a Michaela Pejčochová. *Čínské písmo*. str. 33.

²⁴ UHER, David. *Hanská grammatologie*. str. 104.

²⁵ WILDER, George D., INGRAM James H.: *Analysis of Chinese Characters*. str. VI.

²⁶ *ibid*

²⁷ KUČERA, Ondřej a kol. *Učebnice čínských znaků*. str. xiv.

již není příliš vypovídající.²⁸ Dále pak znak hora 山 *shān*, který zobrazuje tři horské vrcholky, nebo znak dveře 門 *mén*, jež ve své nezjednodušené formě zobrazuje dvoukřídlové rozklápěcí dveře s panty.²⁹

Symboły 指事 *zhǐshì* jsou nejméně početnou kategorií ze všech s pouhými 125 zástupci. Tyto znaky zobrazují zejména abstraktní jevy, které jsou vyjádřeny symbolickým způsobem. Patří sem například znaky dole 下 *xià* a nahoře 上 *shàng*, které odkazují na to, že je něco pod, respektive nad horizontem, dále pak třeba znak rozbřesk 旦 *dàn*, který symbolizuje slunce nad obzorem.³⁰ Mezi piktogramy a symboly existuje řada sporných bodů, které stěžují jejich klasifikaci. Jde například o znak měsíc 月 *yuè*, z jehož původní piktografické formy se vyvinul ještě druhý znak s významem soumrak 夕 *xī*.³¹

Odvozené znaky 字 zì:

Ideogramy 会意 *huìyì* jsou poměrně početná skupina, zastoupená 1167 znaky, které jsou tvořeny dvěma a více prvky. Složky ideogramu spolu tvoří logicky související spojení, ze kterého vzniká výsledný význam. Sem náleží například znak skladiště 库 *kù*, který je složen z komponentů střecha 广 a vůz 车.³² Dalším příkladem je pak znak křivý 歪 *wāi*, jehož komponenty doslova znamenají ne rovný 正.³³

Fonogramy 形声 *xíngshēng* představují zcela nejrozsáhlejší kategorii znaků se 7697 zástupci, mezi něž spadá většina v současnosti užívaných čínských znaků. Tento typ se obvykle skládá ze dvou složek, determinativu, který nese význam znaku a fonetika, které naznačuje výslovnost znaku.³⁴ Patří zde například znak řeka 江 *jiāng*, jehož determinativem je řeka/voda 水 *shuǐ* a fonetikem je práce 工 *gōng*³⁵, nebo řeka 河 *hé*, který se skládá z determinativu řeka/voda 水 *shuǐ* a fonetika moci 可 *kě*.

²⁸ SEHNAL, David. „Čínské znakové písmo, jeho povaha a vývoj“. str. 20.

²⁹ WILDER, George D., INGRAM James H.: *Analysis of Chinese Characters*. str. VII.

³⁰ ibid

³¹ SEHNAL, David. „Čínské znakové písmo, jeho povaha a vývoj“. str. 21.

³² SLAMĚŇÍKOVÁ, Tereza. *Ideogramy v moderní čínštině*. str. 28.

³³ SEHNAL, David. „Čínské znakové písmo, jeho povaha a vývoj“. str. 22.

³⁴ UHER, David. *Hanská grammatologie*. str. 44.

³⁵ ZÁDRAPA, Lukáš a Michaela Pejčochová. *Čínské písmo*. str. 53.

Výpůjčky 假借 *jiǎjiè* vznikají tak, že je pro určitý význam převzat existující znak pro jiné slovo, které nijak nesouvisí s daným významem, ale vyslovuje se podobně.³⁶ Mezi tyto znaky patří například původní znak zpívat 哥 *gē*, který byl vypůjčen pro význam starší bratr a současný znak zpívat byl obohacen další složkou na 歌 *gē*. Stejně tak je výpůjčkou i znak ostatní 它 *tuō*, který původně znamenal had, avšak pro hada byl později vytvořen jiný znak 蛇 *shé*.³⁷

Variety 转注 *zhuǎnzhù* jsou velmi podobné fonogramům, avšak způsob jejich vzniku se liší. Tyto znaky zřejmě vznikají odvozováním od původního znaku na základě významové podobnosti.³⁸ Tento příklad je ilustrován na znacích stařec 考 *kǎo*, starý 老 *lǎo* a senilní 耄 *mào*, které všechny obsahují stejnou významovou složku a zároveň mají i podobnou výslovnost.

Existují i další kategorie znaků, jako třeba znaky transfigurované 变体字 *biàntǐzì* (např. deska 片, pingpong 乒乓)³⁹ a znaky konvenční 记号字 *jìhàozì* (např. číslovky 7 七 a 10 十), avšak jsou natolik obskurní, že stěžejním dělením stále zůstává původních šest kategorií podle Xǔ Shèna.⁴⁰

1.5. Grafické členění čínského znaku

Znak jakožto základní jednotka čínského textu je možné dále dělit na menší složky, mezi kterými platí různé zákonitosti. Jednotlivé grafémy ve znaku jsou od sebe precizně odděleny, avšak jsou vždy konzistentní a v textu se nevyskytují mezery mezi slovy. Do stejně velkého čtvercového grafického pole⁴¹ se musí vejít znak s 1 tahem i znak se 30 tahy nejlépe tak, aby celé pole stejnoměrně vyplňoval.

Jednotlivé složky znaku by také k sobě měly být těsně přimknuty, aby nepůsobily, že stojí samostatně.⁴² Znaky se tedy rozpadají do tří základních úrovní, z nichž nejmenší jsou tahy 笔画 *bǐhuà*. Ve znacích, které jsou tvořeny více, než jedním tahem se spojují

³⁶ UHER, David. *Hanská grammatologie*. str. 45.

³⁷ SEHNAL, David. „Čínské znakové písmo, jeho povaha a vývoj“. str. 24.

³⁸ ZÁDRAPA, Lukáš a Michaela Pejčochová. *Čínské písmo*. str. 58-59.

³⁹ SEHNAL, David. „Čínské znakové písmo, jeho povaha a vývoj“. str. 24.

⁴⁰ SLAMĚNÍKOVÁ, Tereza. *Ideogramy v moderní čínštině*. str. 32.

⁴¹ UHER, David. *Hanská grammatologie*. str. 39.

⁴² ZÁDRAPA, Lukáš a Michaela Pejčochová. *Čínské písmo*. str. 65.

různé tahy mezi sebou do prvků 部件 *bùjiàn*. Kombinace několika prvků spojených dohromady se pak nazývá celý znak 整字 *zhěngzì*.⁴³ Pokud jsou znaky tvořeny jedním prvkem, nazývají se jednoduché 独体字 *dú tǐ zì*, zatímco znaky složené z více prvků jsou 合体字 *hé tǐ zì*. Prvky ve znacích jsou členěny do strukturních typů 结构 *jiégòu*, které se mezi sebou odlišují v pozici jednotlivých prvků. Obecně se udává, že existují tři až čtyři druhy základních struktur, které lze dále dělit na vyšších stupních dekompozice do podstruktur.⁴⁴

1.5.1. Tahy

Tah je nejmenší nedělitelná grafická jednotka čínského znaku, která je vždy psána nepřerušovaně a najednou. Tahy se spojují do větších celků, prvků a z těch se pak skládá celý znak. Existuje několik základních druhů tahů, které byly převzaty z kaligrafické tradice, avšak jejich počet se mezi různými jazykovědci na domácím i zahraničním poli liší, neboť se neshodují na tom, co je samostatný typ tahu a co je jeho pouhá modifikace.⁴⁵ Čínská Norma pořadí tahů u běžně užívaných znaků současné čínštiny 现代汉语通用字笔顺规范 *Xiàndài Hànyǔ Tōngyòng Zì Bǐshùn Guīfàn* z roku 1997 uvádí pět základních tahů. Jsou to tah vodorovný 横 *héng*, svislý 竖 *shù*, levý 撇 *piě*, bodový 点 *diǎn* a zahnutý 折 *zhé*.⁴⁶

Grammatolog Ondřej Kučera v publikaci *Učebnice čínských znaků* uvádí pět základních tahů jako bodový tah 点 *diǎn*, vodorovný tah 横 *héng*, svislý tah 竖 *shù*, levý tah 撇 *piě* a pravý tah 捺 *nà*. Jeho dělení se tedy s normou neshoduje v klasifikaci jednoho tahu.⁴⁷ Lukáš Zádrapa ve své knize *Čínské písmo* rozděluje základních tahů šest. Jedná se o tah vodorovný 横 *héng*, svislý 竖 *shù*, levý 撇 *piě*, pravý 捺 *nà*, bodový 点 *diǎn* a stoupavý 提 *tí*. Navíc zde uvádí jako možný sedmý základní tah provedení svislého tahu s hákem 钩 *gōu*. Většina grammatologů jej považuje za pouhou modifikaci svislého tahu

⁴³ SLAMĚNÍKOVÁ, Tereza. „Nové trendy v čínské grammatologii“. Dálný východ. 2011, č. 1, s. 13-32. ISSN 1805-1049. str. 19.

⁴⁴ VOCHALA, Jaromír, Miroslav Novák a Vladimír Pucek. *Úvod do čínského, japonského a korejského písma II. Praktický kurz*. str. 3.

⁴⁵ ZÁDRAPA, Lukáš a Michaela Pejčochová. *Čínské písmo*. str. 67.

⁴⁶ SLAMĚNÍKOVÁ, Tereza. *Ideogramy v moderní čínštině* str. 18.

⁴⁷ KUČERA, Ondřej a kol. *Učebnice čínských znaků*. str. xii.

neboť jeho užití především v rukopisu je ve většině případů fakultativní. Existují však případy, kdy jeho výskyt od sebe odlišuje dva zcela různé znaky, jako je tomu například u znaků 于 *yú* a 干 *gàn*.⁴⁸

Jaromír Vochala v knize *Chinese Writing system: Minimal graphic units*, rozlišuje dokonce sedm základních (vodorovný, svislý, levý, pravý, stoupavý, pravý bodový a levý bodový) a čtyři až pět hákových tahů (vodorovný hák, svislý hák, ohnutý svislý hák, pravý hák a stoupavý hák). Také se dále podrobně zabývá rozdělením jejich mnoha odvozenin a hlavně vzájemnými kombinacemi jednotlivých základních i odvozených tahů.⁴⁹

Tahy se mezi sebou mohou kombinovat třemi základními způsoby. Oddělená kombinace je taková, ve které se jednotlivé tahy ve znaku navzájem nedotýkají. Tento typ kombinace můžeme vidět ve znacích 二 *èr* a 川 *chuān*. Dalším typem je kontaktní kombinace, kdy se tahy dotýkají v určitém bodě, někdy se dotýkat nemusejí, avšak nesmí se křížit. Zde patří znaky 刀 *dāo*, 人 *rén* a 口 *kǒu*. Třetím typem je protínající se kombinace, která nastane v situaci, kdy se kříží dva tahy a oba dále pokračují. Tuto kombinaci lze vidět například ve znacích 十 *shí* a 力 *lì*. Tyto základní kombinace se mohou v různých znacích objevovat současně, jako je tomu například ve znaku 女 *nǚ*.⁵⁰

Směr, pořadí a počet tahů jsou také velmi důležité aspekty, na které je třeba se zaměřit. Směr psaní tahů má určitá pevná pravidla, podle nichž se tahy vedou zleva doprava, shora dolů, z levého horního rohu do pravého dolního, z pravého horního rohu do levého dolního a z levého dolního do pravého horního rohu.⁵¹ Pořadí tahů ve znaku je důležité především v psaném projevu, avšak dlouhou dobu žádná striktní pravidla neexistovala. Na základě současné normy se tahy vždy píše shora dolů a zleva doprava. Pokud se ve znaku kříží svislý a vodorovný tah, první se píše tah vodorovný. Když se ve znaku kříží levý a pravý tah, jako první bude napsán tah levý a bodový tah vpravo nahoře je zpravidla psán až na konec.⁵² Z pohledu vztahu jednotlivých tahů mezi sebou, jsou tahy vždy ve vztahu předchozího a následujícího. Z pohledu vztahu tahů a celého znaku je pak první tah ve znaku iniciální a všechny ostatní jsou již následující.⁵³

⁴⁸ ZÁDRAPA, Lukáš a Michaela Pejčochová. *Čínské písmo*. str. 67-69.

⁴⁹ VOCHALA, Jaromír. *Chinese writing system: minimal graphic units*. Praha: Univerzita Karlova, 1986, 143 s. Acta Universitatis Carolinae. Philologica. Monographia LXX/78. str. 30.

⁵⁰ *ibid.* str. 24.

⁵¹ VOCHALA, Jaromír. *Chinese writing system: minimal graphic units*. str. 22.

⁵² ŠVARNÝ, Oldřich. *Úvod do hovorové čínštiny*. str. 33-35.

⁵³ VOCHALA, Jaromír. *Chinese writing system: minimal graphic units*. str. 28.

Počet a pořadí tahů jsou důležité především při vyhledávání znaků, neboť v mnoha slovnících jsou řazeny právě podle něj. Některé slovníky navíc také uvádějí číselné kódy, které značí pořadí jednotlivých tahů nebo jejich typ, jako například slovník HXZ. V rámci 7000 běžně používaných čínských znaků je nejvyšší počet tahů 36, zatímco nejnižší je 1. Nejvíce znaků v rámci základní znakové sady moderní čínštiny je tvořeno 9 tahy, průměrný počet tahů na znak této základní sady se pak uvádí 10,75.⁵⁴

1.5.2. Prvky

Na rozdíl od tahů se prvky podrobně zabývá poněkud méně jazykovědců. Pohled na to, co přesně je prvek je totiž poměrně nejasný, neboť se přibližně definuje jako jednotka větší než tah a menší než znak.⁵⁵ Jaromír Vochala v knize *Úvod do čínského, japonského a korejského písma II. Praktický kurz* také užívá místo prvků termín grafické složky, což jsou malé grafické jednotky, které jsou určitým způsobem uspořádány do znaku.⁵⁶ Jak jsem již zmínila, znaky se tedy dělí podle počtu prvků na jednoduché a složené.

Jednoduché znaky 独体字 *dú tǐ zì* se skládají pouze z jednoho prvku, jako například 山 *shān*, 心 *xīn*, a 肉 *ròu*. V rámci jednoduchých znaků se vyskytují specifické případy, které tvoří znaky, jež obsahují pouze jeden tah, jeden prvek a jsou samostatným znakem 一 *yī* a 乙 *yǐ*.⁵⁷ Složené znaky 合体字 *hé tǐ zì* se skládají ze dvou a více prvků, z nichž nejčastější jsou tříprvkové a dvouprvkové znaky, jako například 很 *hěn*, 有 *yǒu* a 她 *tā*.⁵⁸ Vzhledem k rozvoji technologií po celém světě je otázka přesného definování hranice prvku velmi důležitá hlavně v rámci digitálního kódování znaků v počítačích. Z tohoto důvodu byl v roce 2009 vydán Normativní seznam prvků a jejich názvů v rámci frekventovaných znaků 现代常用字部件及部件名称规范 *Xiàndài Chángyòng Zì*

⁵⁴ SLAMĚNÍKOVÁ, Tereza. „Nové trendy v čínské grammatologii“. s 19.

⁵⁵ SLAMĚNÍKOVÁ, Tereza. *Ideogramy v moderní čínštině*. str. 18.

⁵⁶ VOCHALA, Jaromír, Miroslav Novák a Vladimír Pucek. *Úvod do čínského, japonského a korejského písma II. Praktický kurz*. str. 58.

⁵⁷ SLAMĚNÍKOVÁ, Tereza. „Nové trendy v čínské grammatologii“. str. 19.

⁵⁸ KUČERA, Ondřej a kol. *Učebnice čínských znaků*. str. xiii.

Bùjiàn jí Bùjiàn Míngchēng Guīfàn, který vymezuje 514 minimálních prvků 基础部件 *jīchǔ bùjiàn*, což jsou prvky, které se již dále nedají nijak dělit, a vymezuje i jejich užití.⁵⁹

Ve znaku se prvky mohou rozpadat na několika stupních, z nichž největší zřetel se klade na prvky 1. stupně dekompozice, neboť tento stupeň je důležitý při analýze strukturních typů znaků. V rámci některých komplikovaných znaků však lze dílčí prvky dále dělit například až do čtyř stupňů, jako je tomu třeba u znaku 臀 *dún*. Z pohledu dělení znaku přímo na minimální prvky je tento znak přímo složen ze šesti prvků.⁶⁰

1.5.3. Strukturní typy

Prvky ve znaku mají své pevné uspořádání, které není možné libovolně měnit, neboť by mohlo dojít ke změně významu.⁶¹ Různí jazykovědci dělí obecně strukturní typy na tři až čtyři základní, které se na vyšších stupních dekompozice rozpadají do dílčích podstruktur. Základní tři struktury, na kterých se všichni shodují, jsou následovné:

Horizontální struktura 左右结构 *zuǒyòu jiégòu*, ve které jsou prvky v 1. stupni dekompozice skládány vedle sebe. Do této struktury spadají znaky jako například 你, 欢 a 路.

Vertikální struktura 上下结构 *shàngxià jiégòu*, kde jsou prvky v 1. stupni dekompozice uspořádány nad sebe. Tato struktura se vyskytuje například ve znacích 累, 喜 a 怒.

Centrální struktura 包围结构 *bāowéi jiégòu*, v níž jeden prvek obklopuje jiný prvek. V rámci této struktury mohou vnější prvky být dvojramenné, třiramenné a čtyřramenné. Tato struktura tvoří například znaky 回, 看 a 医.⁶²

⁵⁹ SLAMĚNÍKOVÁ, Tereza. Synchronní model tradiční kategorizace čínského znakového písma: Dizertační práce. [nepublikováno] Olomouc: Univerzita Palackého, 2015, 231 s. str. 31.

⁶⁰ *ibid.*

⁶¹ KUČERA, Ondřej a kol. *Učebnice čínských znaků*. str. xiii.

⁶² VOCHALA, Jaromír, Miroslav Novák a Vladimír Pucek. Úvod do čínského, japonského a korejského písma II. Praktický kurz. Praha: Státní pedagogické nakladatelství 1989b, 250 s. str. 58-59.

Někdy je jako čtvrtý typ odlišována **Rámcová struktura** 交叉结构 *jiāochā jiégòu*, kam spadají znaky jako např. 乘 a 爽, ve kterých jedna složka znaku tvoří rám a ostatní složky jsou okolo tohoto rámu souměrně situovány.⁶³ Ve slovníku HXZ, ze kterého vycházela většina analytické části této práce, se místo této rámcové struktury vyskytla „**Hnízdící struktura**“ 嵌套结构 *qiàntào jiégòu*. Mohlo by se zdát, že tato struktura je formou rámcové struktury, avšak to se neshoduje s jejím obvyklým vymezením, neboť se zde vyskytují například znaky jako 园 a 幽,⁶⁴ které obvykle spadají pod centrální strukturu. Formální prvky této struktury jsou také natolik shodné se strukturou centrální, že užitek z jejího vymezení je nulový a proto jsem ji v pozdější analýze zařadila jako součást centrální struktury.

Některé čínské zdroje vymezují více základních strukturních typů, které se dále dělí do podstruktur. Tyto struktury nemají oficiální české názvy, proto jsou zde přeloženy v uvozovkách. Vedle základních třech typů se totiž rozlišuje ještě marginální „symetrická struktura“ 对称结构 *duìchèn jiégòu*, kam patří znaky, které jsou graficky perfektně souměrné, například 巫 a 噩. Další kategorií je „struktura stejných prvků“ 品字结构 *pǐnzi jiégòu*, do které jsou členěny pouze znaky skládající se z jednoho opakujícího se prvku. Do této struktury také spadá jen velmi málo znaků jako například 品 a 森, proto se obvykle v rámci uznávaných vědeckých prací nevymezuje.⁶⁵

Také oficiální terminologie týkající se struktur má jisté sporné body, neboť se autoři různých publikací navzájem neshodují na správném českém překladu některých názvů. Například v *Učebnici čínských znaků I*. Ondřeje Kučery jsou pojmy horizontální a vertikální struktura užity opačně,⁶⁶ než jak je definují zahraniční publikace, nebo například Jaromír Vochala.⁶⁷ Dle tohoto pohledu jsou do horizontální struktury členěny prvky nad sebou po horizontální ose a do vertikální struktury jsou členěny prvky po vertikální ose vedle sebe. Tento způsob pojmenování používá v knize *Ideogramy v moderní čínštině* také Tereza Slaměniková, avšak pro snadnější porozumění nabízí také

⁶³ SLAMĚNÍKOVÁ, Tereza. „*Nové trendy v čínské grammatologii*“. str. 21.

⁶⁴ Hanzi Xinxì Zìdiǎn 《汉字信息字典》 [Informační slovník čínských znaků]. Peking: Kexue Chubanshe, 1988, 1346 s. ISBN 7-03-000869-3. str. 1071.

⁶⁵ Wenkubaidu. 汉字结构类型图示表 [online]. 2013 [cit. 2017-04-15]. Dostupné z: <https://wenku.baidu.com/view/11ec942d3968011ca30091f8.html>.

⁶⁶ KUČERA, Ondřej a kol. *Učebnice čínských znaků*. str. xiii.

⁶⁷ VOCHALA, Jaromír, Miroslav Novák a Vladimír Pucek. *Úvod do čínského, japonského a korejského písma II. Praktický kurz*. str. 58-59.

alternativní termíny pravo-levá a horní-dolní struktura, které lépe vystihují původní čínské termíny.⁶⁸

V rámci základních grafických struktur lze rozlišit mnoho různých podstruktur, které se mezi sebou dále kombinují v různých variacích. V 1. stupni dekompozice se vícečlenné struktury vyskytují pouze jako horizontální 粥, 街, 班 a vertikální 黄, 冀, 言. Na vyšších stupních dekompozice se struktury kombinují velmi různorodě. Vyskytují se například horizontální struktura obsahující vertikální podstrukturu 活, 酸, 搬, centrální podstrukturu 欧, 烟, nebo vícečetnou složku v centrální podstruktuře 滴, dále také vertikální struktura s horizontální podstrukturou 筑, 盐, 夜 nebo centrální podstrukturou 窗, 甚. Vyskytuje se také centrální struktura s vertikální podstrukturou 圆, horizontální podstrukturou 阔, 逝, centrální podstrukturou 疯 nebo se složenou vnější složkou 疫 a jiné, viz přehled kombinací v knize *Úvod do čínského, japonského a korejského písma II. – Praktický kurz.*⁶⁹

⁶⁸ SLAMĚNÍKOVÁ, Tereza. *Ideogramy v moderní čínštině*. str. 19.

⁶⁹ VOCHALA, Jaromír, Miroslav Novák a Vladimír Pucek. *Úvod do čínského, japonského a korejského písma II. Praktický kurz*. str. 90-109.

2. ANALÝZA

Tato část bakalářské práce se věnuje mému vlastnímu výzkumu, jehož hlavním účelem je zjistit zastoupení počtu tahů, prvků a strukturních typů v populárně naučném čínském textu zabývajícím se problematikou chemie a fyziky z kvalitativního i kvantitativního hlediska. V zadání práce bylo udáno, že ve finálním korpusu by se mělo vyskytovat přibližně 100 000 znaků. O výběru a způsobu kompilace mého korpusu se zmíním později v další samostatné části kapitoly, která mu bude věnována. Po provedení analýzy a vyhodnocení výsledků pak provedu jejich komparaci jak mezi sebou, z kvalitativního a kvantitativního hlediska, tak i se znakovým inventářem slovníku HXZ, který v mé práci reprezentuje obecný čínský textový korpus. Hlavním cílem této práce je tedy zjistit, jestli se znakový inventář a skladba znaků v populárně naučném textu liší od textu obecného, a pokud ano, tak ve kterých aspektech.

2.1. Korpus

V původním zadání mé práce bylo shromáždit vědecké články na dané téma z oblasti fyziky a chemie. Toto téma bylo zvoleno proto, že vzhledem k vysoce specializované slovní zásobě v textech odborného stylu se ve statistikách předpokládá jistá odchylka od obecného čínského textu a také od jiných typů textů, např. beletristického nebo publicistického.

Již při shromažďování však nastaly potíže, které zapříčinily následnou drobnou změnu tématu práce. Těžká přístupnost vědeckého materiálu v čínštině a faktická nemožnost shromážděné články převést do formátu, se kterým by bylo možné v analýze dále pracovat, způsobily mírné odchýlení od původního zadání a namísto vědeckého textu jsem zvolila populárně naučné texty na stejné téma. Populárně naučný text je obsahem velmi podobný textu odbornému, je však formálně přizpůsoben běžnému čtenáři tak, aby jej zaujal a zároveň mu předal důležité informace, proto jsem jej shledala jako vhodnou alternativu, která splní stejný účel.

Výchozím bodem analýzy byl tedy tento korpus, respektive dva různé samostatné korpusy, které jsem později sloučila do jednoho většího celku. Oba korpusy byly složeny převážně z vysokoškolských absolventských prací psaných v čínském jazyce za použití zjednodušených znaků, které byly veřejně přístupné na čínském internetovém portálu

Baidu.com. Server Baidu.com by se dal přirovnat k čínskému hybridu mezi Google.com a Wikipedia.org, kterým se svými možnostmi využití velmi podobá. Poskytuje širokou škálu různých služeb od vyhledávání informací, přes překladač, mapy, úložiště souborů a mnoho dalších.

Při kompilaci korpusu jsem zahrnula pouze věcný obsah výše zmíněných prací, očištěný od záhlaví stran a ostatních formálních prvků nepodstatných pro analýzu. Fyzikální a chemický korpus jsem kompilovala odděleně, neboť jsem předpokládala, že ze statistického hlediska by mohlo toto oddělení poskytnout další zajímavá porovnání dvou různých typů populárně naučných textů. Po odstranění mezer, interpunkce, nečínských značek a symbolů vznikly tedy dva různé korpusy. Chemický korpus (dále označovaný jako korpus 1) čítal celkem 41 821 znaků, z nichž bylo 1653 jedinečných znaků, fyzikální korpus (dále označovaný jako korpus 2) pak čítal 44 924 znaků, z nichž bylo 1533 jedinečných znaků. Sloučený finální korpus tedy obsahoval 86 745 zjednodušených čínských znaků, z nichž bylo jedinečných 2027 znaků. Oba korpusy měly společných 1159 znaků, zatímco v korpusu 1 se výlučně vyskytlo 494 znaků a v korpusu 2 to bylo 374 znaků.

Všechny znaky, které se v textu vyskytly a počet jejich četnosti jsem zjistila s pomocí korpusového manažeru na internetovém portálu cncorpus.org, díky kterému bylo možné vytvořit frekvenční list znaků, který se stal výchozím bodem celé mé analýzy. Tento seznam znaků jsem později přenesla do tabulek v programu MS Excel, kde jsem jej dále upravovala.

V mé práci narážíme na dvě již dříve zmiňovaná hlediska, kvalitativní a kvantitativní, pro která se v korpusové lingvistice v anglickém jazyce používají termíny Type frequency a Token frequency. Kvalitativní pohled (Type frequency) zohledňuje každý výskyt jednoho znaku, jakožto nositele významu v textu pouze jednou, na základě abstraktní ideje, která se k němu váže. Z hlediska kvalitativního obsahoval tedy korpus celkem 2027 jedinečných znaků (tzv. „typů“). Kvantitativní pohled (Token frequency) pak zohledňuje absolutní četnost znaků, tedy kolikrát se v daném textu fyzicky všechny jednotlivé znaky opakovaně vyskytly.⁷⁰ Z kvantitativního hlediska tedy obsahoval celkem 86 745 znaků (tzv. „tokenů“), neboť mnoho ze znaků se v korpusu vyskytlo několikanásobně.

⁷⁰ ŠULC, Michal. *Korpusová lingvistika: první vstup*. str 12-14.

2.2. Informační slovník čínských znaků 汉字信息字典 *hànzì xìnxī zìdiǎn*

Informační slovník čínských znaků (zkr. HXZ) byl po samotném korpusu druhým nejdůležitějším a nepostradatelným nástrojem mé analýzy. Tento slovník byl vydán ve spolupráci několika autorů z Šanghajske univerzity roku 1988. Na svých 1346 stranách popisuje celkem 11254 čínských znaků, z toho 7785 zjednodušených znaků a 3469 nezjednodušených znaků.⁷¹ Jak už samotný název napovídá, jedná se o výkladový slovník čínských znaků, který se znaky zabývá především po jejich grafické stránce. Vzhledem k tomu, že kromě fonetického přepisu, výslovnosti znaku a základní definice neobsahuje žádný další jazyk, je primárně zamýšlen jako pomůcka pro učitele, studenty, jazykovědce a jiné výzkumné pracovníky, kteří se zajímají o čínské znakové písmo, a to zejména o statistické hodnoty týkající se například četnosti výskytu znaků, jejich strukturního členění, nebo počtu tahů.

Z pohledu makrostruktury je slovník rozdělen do tří základních částí, které se dále dělí na menší úseky. První část je zaměřená na seznámení čtenáře se slovníkem. Obsahuje úvod, ve kterém autoři vysvětlují, v čem tkví přínos slovníku a co je jeho primárním cílem. Dále je zde detailně rozepsáno vše ohledně samotné práce se slovníkem, uspořádání hesel a jejich mikrostruktura. Následuje seznam symbolů a značek, které jsou ve slovníku použity, a jejich vysvětlení. Dále se zde nacházejí seznamy znaků, seřazené podle různých kritérií (např. počet tahů, radikály). Také zde nalezneme tabulky znaků zaměřující se na transkripci pinyin, nebo srovnání starých a nových verzí prvků.

Druhou velkou část lze rozdělit na tři menší oddíly. Největší z nich obsahuje znaky, foneticky seřazené od A do Z, podle jejich primární výslovnosti. Z hlediska mikrostruktury jsou hesla uspořádána přehledně, což značně ulehčuje orientaci ve slovníku a rychlost práce s ním. Každý znak má své jedinečné pořadové číslo na základě fonetické posloupnosti všech znaků, jak jsem již zmínila výše. Nezjednodušené znaky jsou sice odděleny do samostatné části seznamu, ale nejsou řazeny znovu od čísla 1, nýbrž navazují na pořadí předešlých zjednodušených znaků. Každé lemma dále obsahuje číselný kód, ve kterém jsou zaneseny informace jako tónová značka, přesný počet tahů, pořadí a typologie prvních pěti tahů. Další částí hesla je výslovnost zapsaná v transkripci pinyin a základní definice významu znaku. Pro mou práci byly však nejdůležitější zejména další tři body každého lemmatu, které popisují determinativ, pod který daný znak

⁷¹ *Hanzi Xinxī Zidian* 《汉字信息字典》 [*Informační slovník čínských znaků*] str. 5.

spadá, strukturní členění jednotlivých prvků ve znaku, pořadí tahů a někdy také poznámku, týkající se možného špatného přečtení nebo nejasností s užitím znaku. Druhým důležitým pododdílem druhé části slovníku jsou statistiky (vycházející z korpusu čítajícího 21 656 578 znaků), týkající se například počtu významů znaku, počtu homofonních znaků, nejfrekventovanějších struktur znaku, nejčastějšího počtu prvků, nebo nejčastějšího počtu tahů. Třetí část slovníku obsahuje přílohy, které obsahují fyzikální jednotky, tabulky týkající se přepisu slabik několika způsoby u nejčastěji užívaných znaků, seznam všech zjednodušených znaků spolu s původními verzemi nebo seznam znaků řazený podle počtu tahů. Tyto části jsou v porovnání s běžným výkladovým slovníkem jedinečné a nemají v dílech podobného formátu obdobu.

Práce se slovníkem je vzhledem k jeho přehlednému členění poměrně jednoduchá za předpokladu, že si čtenář důkladně nastuduje vysvětlivky na začátku knihy. Při analýze jednotlivých hesel jsem si všimla drobných chyb, jako například chybně uvedeného počtu tahů ve znaku, nebo neodpovídající struktury, což je ale dle mého názoru u díla čítajícího 1346 stran téměř nevyhnutelné a sami autoři v úvodu zmiňují, že si jsou chyb vědomi a uvítají případné připomínky či kritiku.

2.3. Zpracování analýzy

Postup práce s korpusem a slovníkem lze rozdělit na několik etap. Nejprve jsem si vytvořila korpus adekvátní pro účel této práce dle kritérií, která jsem již zmínila v samostatné kapitole týkající se korpusu. Po shromáždění korpusu jsem použila internetový program cncorpus.org, který z daného korpusu vytvořil frekvenční list znaků, které se v něm vyskytly. Tento list, v počátku mé práce dva samostatné seznamy, byl seřazen podle abecedy a obsahoval počet absolutních výskytů daného znaku v textu a počet relativní četnosti výskytu daného znaku v textu, což velmi usnadnilo mou práci v pozdější fázi analýzy, neboť nebylo nutné ručně počítat procentuální zastoupení všech daných znaků v textu.

Frekvenční list jsem vložila do první tabulky v programu MS Excel, kde jsem si vytvořila různé kategorie, pro lepší přehlednost některé označila barevně a nastavila filtry obsahu, které byly taktéž nezbytné v pozdější části analýzy. Zejména kategorie počet tahů, počet prvků a struktura, byly zcela zásadní pro mou následnou analýzu a vyhodnocení výsledků. Informace o tazích, prvcích a strukturním členění jsem pak následně

vyhledávala ve slovníku HXZ a zapisovala je po jednom do příslušných kolonek ve sloučené tabulce znaků z obou korpusů, protože většina znaků se shodovala a nebylo by efektivní je vyhledávat dvakrát. Úskalím této části práce byla její časová náročnost, vzhledem k tomu že bylo nutné každý z 2027 znaků, které se v korpusech dohromady vyskytly, vyhledat ručně ve slovníku.

Ve slovníku HXZ jsou strukturní členění znaku rozlišena do 5 typů, kterých jsem se držela ve své práci i já, avšak při závěrečném vyhodnocení výsledků jsem se rozhodla pátou strukturu inkorporovat do struktury centrální vzhledem k její nízké četnosti. Jak již bylo nastíněno v teoretické části popisující strukturní typy, obecně se uvádí struktury pouze 4 (rámcová struktura se dle příruček moderní grammatologie započítává do struktury centrální, vzhledem k její nízké četnosti). Autoři slovníku si vytvořili čtvrtou strukturu, jejíž vymezení s rámcovou strukturou ostatních publikací nekoreluje, což byl další z důvodů k jejímu sloučení s centrální strukturou. Ukázka upravené části tabulky se nachází na další straně a celková tabulka tak, jak jsem ji používala pro analýzu, je pak k dispozici v přílohách na CD.

Po přepsání všech potřebných dat ze slovníku jsem si vytvořila tři různé oddíly tabulek, zabývající se samostatně jednotlivými počty tahů, prvků a struktur z kvalitativního i kvantitativního hlediska v absolutních i relativních číslech pro každý korpus samostatně, sloučený korpus i inventář HXZ. Tyto tabulky vzhledem k jejich obsáhlosti nebudou součástí textu práce s výjimkou ukázky jedné z nich v další části práce. Jsou však také k dispozici k nahlédnutí v přílohách na CD stejně jako všechny ostatní materiály, které byly součástí mé analýzy.

2.3.1. Frekvenční tabulka znaků

Níže uvedená tabulka slouží pro ilustraci toho, jak vypadal analyzovaný materiál v druhé fázi mé analýzy před tím, než jsem začala provádět výpočty zastoupení jednotlivých elementů v korpusu. Zde uvedená část tabulky je upravena pro potřeby ilustrace tak, že obsahuje pouze důležité kategorie, jako jsou znak, počet tahů, počet prvků, struktura a absolutní četnost výskytu. Byly z ní odstraněny některé dílčí pomocné kategorie, které nebyly pro výslednou analýzu relevantní, a zobrazuje pouze náhodně vybranou množinu znaků. Také struktury v níže uvedené ukázce části tabulky jsou přepsány slovně, zatímco v originálu celkové tabulky jsou značeny zástupnými symboly,

které jsem si zvolila dle legendy v prvním listu dokumentu MS Excel, který se nachází spolu s tabulkou v přílohách na CD.

ZNAK	POČET TAHŮ	POČET PRVKŮ	STRUKTURA	ABSOLUTNÍ ČETNOST
电	5	1	jednoduchá	1395
磁	14	4	horizontální	961
一	1	1	jednoduchá	521
理	11	2	horizontální	369
这	7	2	centrální	295
生	5	1	jednoduchá	244
通	10	3	centrální	212
斯	12	3	horizontální	156
体	7	2	horizontální	145
由	5	1	jednoduchá	119
料	10	2	horizontální	100
算	14	3	vertikální	90
外	5	2	horizontální	85
常	11	3	vertikální	70

Tabulka 1: Ukázka upraveného frekvenčního seznamu znaků

2.3.2. Analytická tabulka struktur

Zde uvádím příklad, jak vypadal jeden z finálních produktů mé analýzy, konkrétně tabulka relativní četnosti výskytu strukturních typů v textu z pohledu kvantitativní analýzy. Jako příklad jsem zvolila tuto tabulku, protože byla nejméně obsáhlá v porovnání s ostatními tabulkami počtu tahů, nebo prvků.

STRUKTURY	JS	VS	HS	CS	RS
korpus 1 酸	20,06 %	18,39 %	46,59 %	13,99 %	0,97 %
korpus 2 物理学	23,83 %	17,59 %	42,75 %	14,77 %	1,06 %
sloučený korpus	22,01 %	17,98 %	44,60 %	14,40 %	1,01 %
HXZ	25,91 %	19,35 %	40,09 %	13,31 %	1,35 %

Tabulka 2: Relativní četnost výskytu strukturních typů v textu

V tabulce jsou zohledněny jak oba korpusy zvlášť, tak korpus 1+2 dohromady a také data vyhledaná ve slovníku HXZ. Barevně jsou označeny tři nejvyšší hodnoty výskytů v rámci každého korpusu, aby se v tak velkém množství dat bylo možné lépe orientovat, tmavě růžová barva značí nejvyšší hodnotu, středně růžová druhou v pořadí a světle růžová třetí nejvyšší hodnotu.

Interpretací dat obsažených v této tabulce se bude dále zabývat příslušná kapitola, zaměřená na výsledky analýzy strukturních typů v další části této práce. Výsledky, které jsem vyhodnotila ze všech jednotlivých tabulek, jsem pak mezi sebou porovnála v rámci jednotlivých korpusů i slovníku HXZ a poskytla interpretaci závěrů, které z této komparace vycházejí.

3. VÝSLEDKY ANALÝZY

V této kapitole budu prezentovat postupně jednotlivé výsledky analýzy tahů, prvků a struktur v mém korpusu a provedu srovnání těchto výsledků se statistikami obsaženými ve slovníku HXZ.

3.1. Oborově specifické znaky

V rámci analýzy znaků obsažených v korpusech jsem provedla dílčí výzkum s cílem zjistit, zda specifické znaky z oblasti fyziky a chemie měly nějaký vliv na výsledky závěrečné analýzy. V korpusu 1, který se zabýval chemií, se vyskytlo celkem 1653 jedinečných znaků, z nichž 494 se vyskytlo pouze v tomto korpusu. Korpus 2 zaměřený na fyziku čítal 1533 jedinečných znaků, ze kterých se jen v tomto korpusu výlučně objevilo 374. V každém korpusu zvlášť jsem se pak zaměřila na prvních 200 nejfrekventovanějších znaků celkově a také na prvních 200 nejfrekventovanějších znaků v rámci těch znaků, které se výlučně vyskytly v prvním, či druhém z nich.

V níže uvedené tabulce jsou zobrazeny některé vybrané oborově specifické znaky - pro korpus 1 vlevo a korpus 2 v pravé části tabulky. Vzhledem k tomu, že mým primárním cílem bylo zkoumat pouze grafickou podobu znaků v daných korpusech, znaky v této tabulce byly vybrány z množiny oborově specifických znaků náhodně pro účel ilustrace. Nejsou tudíž seřazeny podle jejich četností v textech a nereprezentují celou množinu oborově specifických znaků, jejichž analýzou by se mohla zabývat samostatná diplomová práce.

Červenou barvou jsou označeny znaky, které se vyskytují v obou dvou korpusech, zeleně pak znaky, které se vyskytly pouze v chemickém korpusu 1, a modře znaky specifické pro fyzikální korpus 2. Jak je z těchto tabulek už na první pohled zřejmé, mnohem více znaků, které byly obsaženy v obou korpusech, se vyskytuje na prvních příčkách v korpusu 2, zatímco korpus 1 vykazuje vysoký počet znaků, které se vyskytly pouze v něm, pocházejících převážně z chemické terminologie.

Korpus 1			Korpus 2		
酸	<i>suān</i>	kyselina	电	<i>diàn</i>	elektřina
机	<i>jī</i>	organický	磁	<i>cí</i>	magnetismus
氨	<i>ān</i>	amoniak	场	<i>chǎng</i>	pole
液	<i>yè</i>	kapalný	力	<i>lì</i>	síla
丙	<i>bǐng</i>	propionát	量	<i>liàng</i>	množství
氧	<i>yǎng</i>	kyslík	铁	<i>tiě</i>	železo
焊	<i>hàn</i>	svářet	度	<i>dù</i>	stupeň
烯	<i>xī</i>	olefin	验	<i>yàn</i>	pokus
酶	<i>méi</i>	enzym	器	<i>qì</i>	zařízení
油	<i>yóu</i>	olej	材	<i>cái</i>	materiál
钢	<i>gāng</i>	ocel	微	<i>wēi</i>	mikro
酮	<i>tóng</i>	keton	机	<i>jī</i>	stroj
钙	<i>gài</i>	vápník	速	<i>sù</i>	rychlost
苯	<i>běn</i>	benzen	螺	<i>luó</i>	spirála
硅	<i>guī</i>	silikon	频	<i>pín</i>	frekvence
醇	<i>chún</i>	alkohol	伏	<i>fú</i>	volt
羧	<i>suō</i>	karboxyl	秤	<i>chèng</i>	škála
醛	<i>quán</i>	aldehyd	矢	<i>shǐ</i>	vektor
氟	<i>fú</i>	fluor	钳	<i>qián</i>	svorka
腈	<i>jīng</i>	nitril	伽	<i>gā</i>	gamma

Tabulka 3: Vybrané oborově specifické znaky

V obou seznamech se mi podařilo identifikovat přibližně 50 (korpus 1) a 40 (korpus 2) oborově specifických znaků, z nichž některé byly velmi četné a vyskytovaly se v obou korpusech (byť v různé míře), zatímco jiné byly poměrně málo frekventované. V rámci prvních 200 nejfrekventovanějších znaků v korpusu 1 se nejčastěji vyskytoval oborově specifický znak 酸 *suān* kyselina, který se svými 1006 výskyty tvořil bezmála 2,5 % textu. Dále se pak hojně objevovaly znaky jako například 氨 *ān* amoniak, 氧 *yǎng* kyslík a 酶 *méi* enzym, které se vyskytovaly zároveň i v korpusu 2, ačkoli pouze marginálně. Již na druhé příčce mezi prvními 200 nejčetnějšími znaky korpusu 2 se objevil znak 电 *diàn*

elektřina s 1395 výskyty, zabírající 3,1 % textu. Následovaly pak znaky 磁 *cí* magnetismus, 力 *lì* síla, nebo 铁 *tiě* železo, které se rovněž vyskytly v rámci korpusu 1, avšak jejich četnost v něm byla menší.

Mezi prvními 200 nejčetnějšími znaky se v korpusu 1 objevilo ve vysoké četnosti několik znaků, které se výlučně vyskytovaly pouze zde. Nejzajímavější z nich byly například 丙 *bǐng* propionát, který se vyskytl již na 62. pozici ve frekvenční tabulce a 烯 *xī* olefin na 83. pozici. Ve stejné kategorii v rámci korpusu 2 se žádné výlučně se vyskytující znaky neobjevily. První takovýto znak v korpusu 2 螺 *luó* spirála byl až na 248. místě ve frekvenční tabulce a dále následoval znak 频 *pín* frekvence, který byl 289. v pořadí.

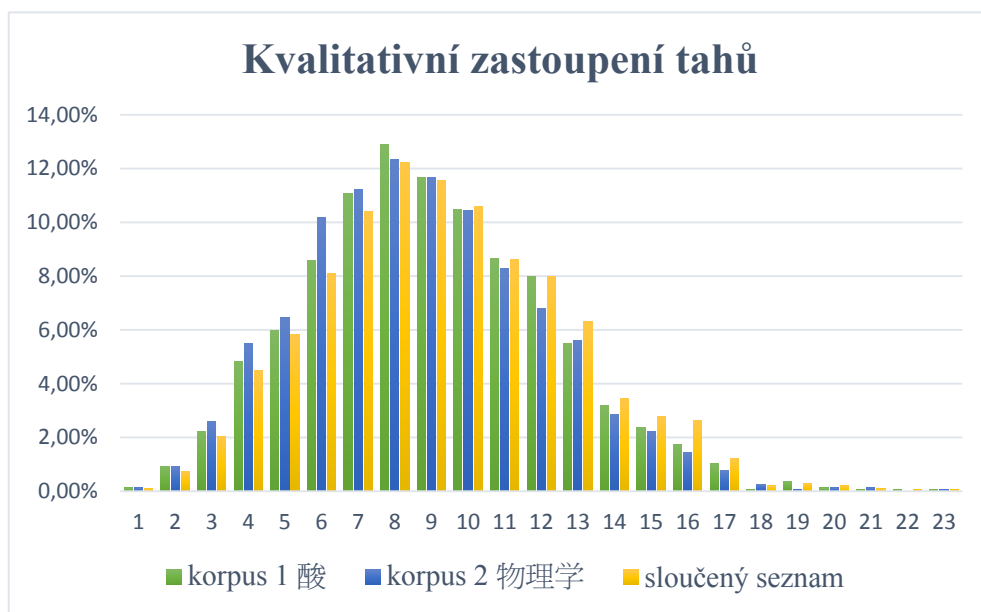
Co se týče oborově specifických znaků korpusu 1, převládala mezi nimi horizontální struktura, v níž byla uspořádána většina znaků s nejčastějšími determinativy oheň, kov a kámen. V celém textu se pak hojně vyskytuje názvosloví kyselin (复合酸 *fùhé suān* komplexní kyselina, 有机酸 *yǒujī suān* organická kyselina, 无机酸 *wújī suān* anorganická kyselina), které obsahuje druhý nejčastější znak v celém textu 酸 *suān* kyselina, jenž je také členěn horizontálně. Další poměrně velkou část pak tvořily především znaky obsažené v názvech prvků plynného skupenství, například 氮 *dàn* dusík, 氰 *qín* kyanid a 氩 *yà* argon, které spadají pod centrální strukturu. V korpusu textu byla také mezi oborově specifickými termíny četná vertikální struktura, která se vyskytuje ve složených výrazech (复合型 *fùhé xíng* sloučenina, 复合酸 *fùhé suān* komplexní kyselina), v nichž znaky 复 *fù* a 合 *hé* oba s významem skládat, jsou tvořeny právě touto strukturou.

Oproti tomu v korpusu 2 lze vidět relativně větší množství jednoduchých znaků, jako například 矢 *shǐ* vektor, 电 *diàn* elektřina (v textu ve slovech 电流 *diànlíu* elektrický proud, 电场 *diànchǎng* elektrické pole, atd.) nebo 力 *lì* síla. Mezi ostatními znaky sice převládá horizontální členění (v textu například 磁场 *cíchǎng* magnetické pole, 磁场强度 *cíchǎng qiángdù* intenzita magnetického pole), ale objevuje se i centrální a vertikální struktura, vzhledem k různosti determinativů, jimiž jsou znaky tvořeny. Strukturní uspořádání oborově specifických znaků korpusu 2 je tedy barvitější, než jak je tomu u korpusu 1, kde zcela převládá horizontální struktura.

3.2. Tahy

Analýza tahů přinesla poměrně zajímavé výsledky vzhledem k tomu, že se projeví významné rozdíly mezi kvalitativní a kvantitativní statistikou. Naopak se prokázalo, že velikost korpusu nemá na průměrný počet tahů ve znaku tak zásadní vliv, jak by se dalo očekávat.

Korpus 1, korpus 2 a sloučený korpus 1+2 s 2027 jedinečnými znaky se převážně shodují v nejčastějším počtu tahů ve znacích mezi 6 a 12 tahy, jak můžeme vidět na grafu kvalitativního zastoupení počtu tahů na další straně. Nejvíce se pak v sloučeném seznamu znaků vyskytují znaky s 8 tahy, které tvoří (12,23 %), za nimi následují znaky s 9 tahy, které tvoří (11,54 %). V další kategorii však vznikl rozdíl mezi dílčími seznamy a sloučeným seznamem, neboť v obou z nich jsou třetí nejčastěji se vyskytující položkou znaky se 7 tahy, zatímco ve sloučeném seznamu jsou to znaky s 10 tahy (10,61 %). Ve kvalitativní statistice znaků tedy můžeme vidět stoupající tendence počtu tahů v závislosti na velikosti zkoumaného vzorku, rozdíly ve výsledcích však nejsou nikterak velké.

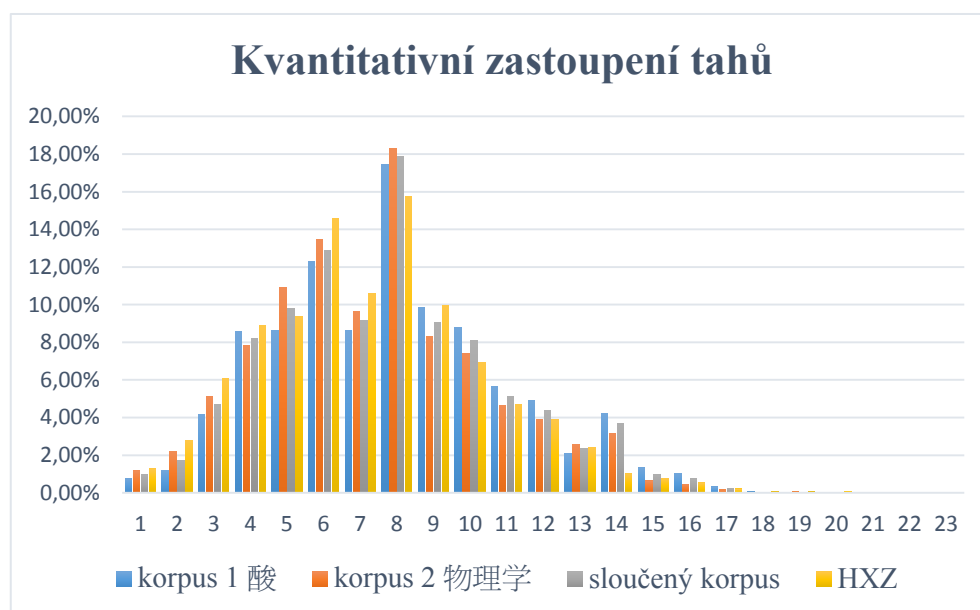


Graf 1: Kvalitativní zastoupení počtu tahů

Je zde důležité podotknout, že jsou tato čísla také posunutá o jeden tah dolů v průměru od obecných statistik, nacházejících se ve slovníku HXZ, který má nejvyšší počet znaků s 9, 10 a 11 tahy. Kromě prvních tří nejčastějších výskytů se nejčastěji vyskytující počty tahů ve slovníku s těmi v mém korpusu více méně překrývají, avšak

mají mírně větší amplitudu od 7 do 14 tahů. Z tohoto výsledku můžeme tedy usuzovat, že zkoumaný vzorek této práce byl dostatečně vypovídající, přestože byl podstatně menší než vzorek slovníku. Myslím si, že drobné rozdíly vznikly proto, že vzhledem k obecnosti korpusu HXZ je větší pravděpodobnost výskytu více obskurních mnoho-tahových znaků, které mohou statistiku ovlivňovat a posouvat ji směrem nahoru.

Mnohem zajímavější je však rozdíl mezi kvalitativním a kvantitativním pohledem na nejrozšířenější počty tahů. Z grafu kvantitativního zastoupení počtu tahů totiž vyplývá, že v reálném textu mají oproti kvalitativní statistice nejvyšší zastoupení znaky se 4 až 10 tahy, na čemž se překvapivě shodují jak všechny mé korpusy, tak také slovník HXZ. Zdaleka nejvíce bylo v textech (stejně jako ve znakovém inventáři) znaků s 8 tahy (17,88 %), pak se však na druhé místo dostávají znaky se 6 tahy (12,88 %) a na třetí místo znaky s 5 tahy (9,79 %). Je tedy zřejmé, že jsou v čínštině obecně nejvíce používány znaky s poměrně málo tahy, protože je snadné je napsat a přečíst, zatímco znaky se 14 a více tahy jsou odsouvány na okraj používání, bez ohledu na typ či účel výchozího textu.



Graf 2: Kvantitativní zastoupení počtu tahů

Znak s nejvyšším počtem tahů v mých korpusech byl 罐 *guàn* s 23 tahy, který se v celém textu vyskytl pouze čtyřikrát. Znak s nejnižším počtem tahů v mém korpusu byl 一 *yī*, tvořený 1 tahem, potažmo 1 prvkem, který se vzhledem ke svému významu číslovky vyskytl v korpusu 765 krát.

Vůbec nejčastěji se vyskytujícím znakem byl osmitahový znak 的 *de*, který je pro jakýkoli text nepostradatelný vzhledem ke své gramatické funkci. Tento znak zaujal první místo v obou korpusech s 3803 výskyty a sám tvořil 4,5 % veškerého textu. Zajímavý je také fakt, že hned druhý nejrozšířenější znak v obou korpusech se však již neshoduje. V chemickém korpusu 1 to byl znak 酸 *suān* kyselina s 1006 výskyty, zatímco ve fyzikálním korpusu 2 to byl 电 *diàn* elektřina s 1395 výskyty. Oba tyto znaky se však v druhém korpusu vyskytly v rámci prvních 200 nejčtetnějších znaků. Můžeme tedy usuzovat, že typ textu sice příliš neovlivní obecnou skladbu znaků, ale je schopen ovlivnit četnost výskytu znaků specifických pro svůj obor.

3.2.1. Ukázka frekvence zastoupení počtu tahů v textu

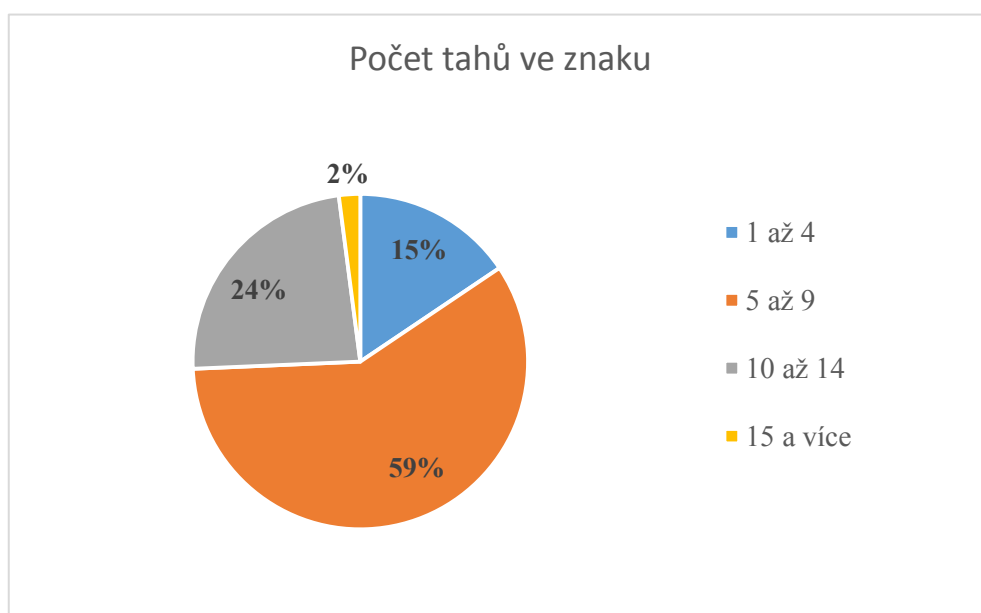
Tento oddíl slouží k názorné ilustraci frekvence zastoupení počtu tahů v textu mého korpusu. Z chemického korpusu 1 jsem vybrala odstavec, který čítá 272 znaků, ve kterém jsem barevně označila počty tahů v intervalech, jež jsem si stanovila. Tyto intervaly slouží ke zlepšení přehlednosti, neboť barevné označení každého počtu tahů jinou barvou by působilo spíše rušivě.

Modrou barvou jsou označeny znaky s 1 až 4 tahy, žlutou jsou označeny znaky s 5 až 9 tahy, u kterých předpokládám nejvyšší četnost výskytu, a červenou barvou jsou označeny znaky s 10 až 14 tahy, jejichž frekvenci předpokládám nižší. Zelenou barvou jsou pak označeny znaky s 15 a více tahy, jichž bylo v textu tak málo, že jsem všechny sloučila do jediné kategorie, aby se ve vzorku vůbec zobrazily.

有机酸的选择实验版前言目前酸化剂种类繁多琳琅满目日本文旨在通过试验文献来评价各有机酸的特点为大家提供一定的参考依据本人才疏学浅文章如有不对之处烦请大家指正酸化剂功能和种类酸化剂具有调控动物肠道微生物平衡增殖有益菌抑制有害菌降低肠道值提高消化道酶的活性提高营养物质消化率减慢胃排空速度提高动物抗应激和免疫力等方面的功能目前市面上的酸化剂分为四类无机型有机型无机有机复合型有机有机复合型可以说各有各的特点复合型酸化剂酸化剂较为人们认可最初酸化剂主要功能被用于降低饲料系酸力但目前的饲料配方设计理念中饲料的系酸力已经比较低因此纯无机的酸化剂逐渐淡出人们

V úvodní kapitole týkající se mé analýzy jsem již zmínila, že korpus mého textu byl zbaven nečínských textových prvků, zejména pak interpunkčních znamének. V této kapitole daná modifikace vhodně slouží k lepší zřetelnosti distribuce jednotlivých struktur v textu, neboť je zcela souvislý a není narušen žádnými cizími prvky.

Jak lze na tomto ukázkovém úryvku pozorovat, zcela nejvyšší četnost mají v textu znaky s 5 až 9 tahy, které v korpusu tvoří 59 %. Znaků s 1 až 4 tahy a znaků s 10 až 14 tahy je podstatně méně, ale stále dohromady tvoří 39 %, což jasně ukazuje na to, že znaky s více než 15 tahy jsou opravdu zcela na okraji zájmu používání. Toto rozložení frekvence počtu tahů ve znacích korpusu dalším způsobem demonstruje také Graf 3 níže.

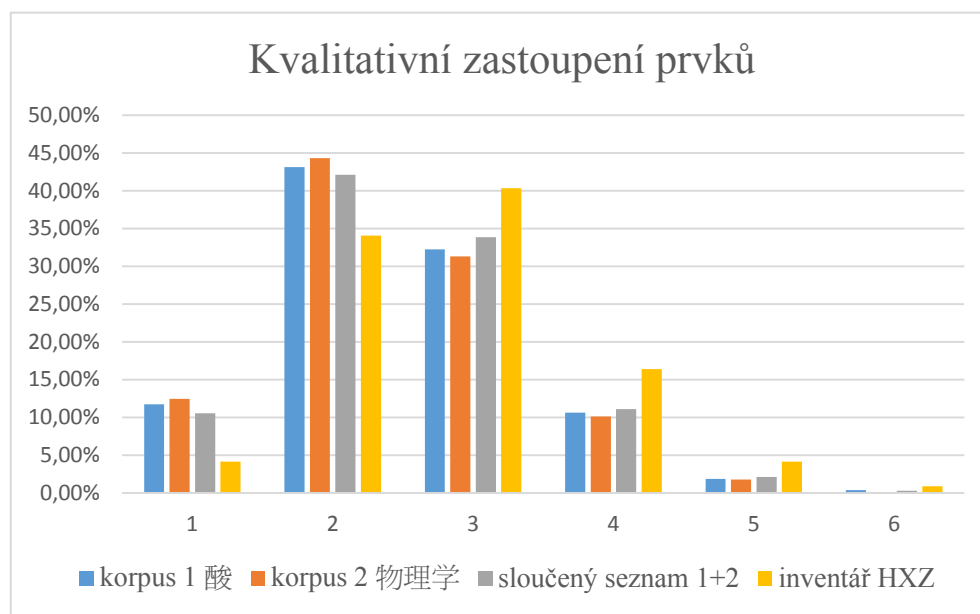


Graf 3: Počet tahů ve znacích korpusu 1+2

3.3.Prvky

Nejvyšší počet prvků ve znaku mého korpusu bylo 6, takovýchto znaků se v celém korpusu však vyskytlo pouze šest, z nichž většina náležela pouze k chemickému korpusu 1. Tento fakt je ovšem zřejmě náhodný, neboť jsem nepřišla na žádnou očividnou spojitost mezi těmito znaky a daným korpusem. Znaků s nejnižším počtem prvků, tedy jedním, byla v seznamu přibližně desetina. Celkem bylo tedy nalezeno 217 jednotlivých znaků, které se s různou mírou četnosti v textu vyskytovaly opakovaně.

Stejně jako v analýze tahů i zde se vyskytl poměrně velký rozdíl mezi kvalitativní a kvantitativní analýzou. Nejprve se však budu věnovat výsledkům analýzy kvalitativní. Z ní vyplývá, že nejvíce se ve frekvenčním seznamu 2027 znaků vyskytují znaky se dvěma prvky (42,08 %) a za nimi pak následují znaky tříprvkové (33,84 %), jak můžeme vidět na grafu kvalitativního zastoupení prvků níže.



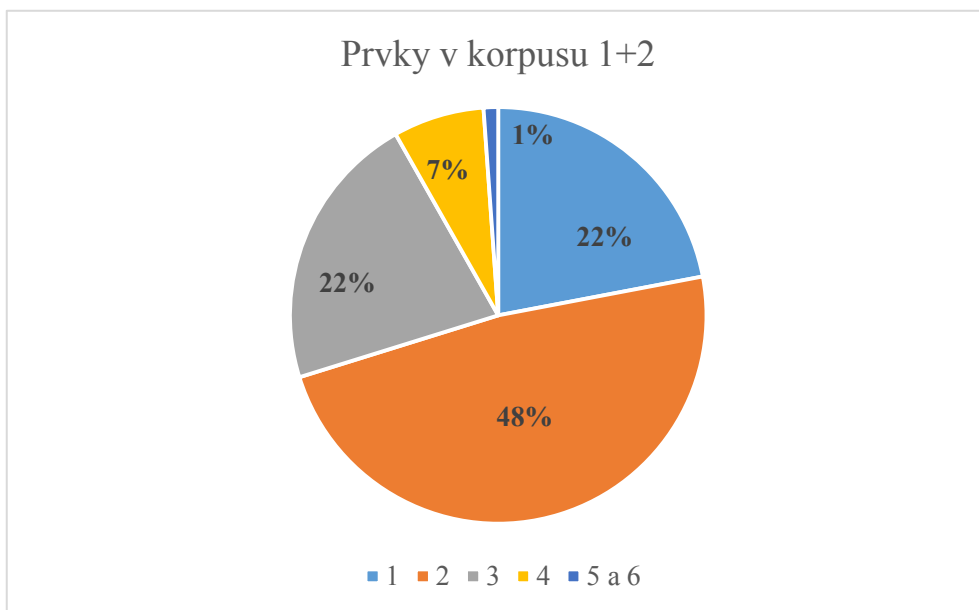
Graf 4: Srovnání kvalitativního zastoupení počtu prvků

Tento výsledek se však odlišuje od inventáře slovníku HXZ, který uvádí nejvyšší počet znaků se třemi prvky (40,32 %) a až pak znaky dvouprvkové (34,04 %). Můžeme tudíž usuzovat, že tento rozdíl lze opět odůvodnit rozdílem ve velikosti zkoumaných vzorků mého korpusu a slovníku HXZ, kdy počet prvků přímo úměrně klesá s počtem znaků ve vzorku.

Skutečný výskyt počtu prvků ve znacích celého korpusu je však od kvalitativní statistiky odlišný. Na základě kvantitativní analýzy bylo zjištěno, že stejně jako ve frekvenčním seznamu velkou většinu reálného textu zaujímají dvouprvkové znaky, které se 48 % tvořily téměř polovinu celého textu. Podobné hodnoty výskytu dvouprvkových znaků vykazuje jak můj korpus, tak korpus HXZ, který jich v textu uvádí 47 %.

Je však viditelný i značný rozdíl oproti jeho frekvenčnímu seznamu, kterému dominují tříprvkové znaky, jak jsem již zmínila výše. Můžeme zde tedy pozorovat projev obecně známých tendencí ohledně nejvíce používaného typu znaků, kdy se v textech ve větší frekvenci využívají graficky méně komplikované znaky. Výskyt jednoprvkových znaků (22,01 %) a tříprvkových znaků (21,60 %) v rámci mého korpusu byl srovnatelný,

pouze s marginálním rozdílem necelých 0,5 %, což lze vidět na grafu kvantitativního zastoupení prvků v obou korpusech z kvantitativního hlediska. Tento výsledek není příliš překvapivý, neboť daná statistika souhlasí také s HXZ, ačkoli rozdíl mezi jednoprvkovými a tříprvkovými znaky ve slovníku je asi 4,5 %, což je dáno jeho obsáhlostí.



Graf 5: Kvantitativní zastoupení počtu prvků v korpusu 1+2

Zároveň také můžeme vidět velký rozestup mezi četností výskytu tříprvkových a čtyřprvkových znaků, které se vyskytují v textu pouze v 7 % případů. Znaky s větším počtem prvků se v textu vyskytují pouze v zanedbatelném množství, což opět jasně ukazuje na snahu o minimalismus a jednoduchost nejvíce frekventovaných znaků.

3.3.1. Ukázka frekvence zastoupení počtu prvků v textu

Tato ukázka slouží k ilustraci frekvence zastoupení počtu prvků v rámci mého korpusu. V tomto oddílu používám pro lepší přehlednost a možnost porovnání stejný úryvek korpusu 1, jako v předchozí kapitole. Také zde používám stejnou škálu barev pro vyznačení jednotlivých počtů prvků.

Modrou barvou jsou označeny jednoprvkové znaky, žlutou barvou jsou označeny znaky dvouprvkové, červenou barvou jsou označeny znaky se třemi prvky a zelenou barvou jsou označeny znaky se čtyřmi a více prvky, neboť frekvence výskytu znaků s pěti a šesti prvky byla dle předchozí analýzy celého textu poněkud zanedbatelná.

有机酸的选择实验版前言目前酸化剂种类繁多琳琅满目本文旨在通过试验文献来评价各有机酸的特点为大家提供一定的参考依据本人才疏学浅文章如有不对之处烦请大家指正酸化剂功能和种类酸化剂具有调控动物肠道微生物平衡增殖有益菌抑制有害菌降低肠道值提高消化道酶的活性提高营养物质消化率减慢胃排空速度提高动物抗应激和免疫力等方面的功能目前市面上的酸化剂分为四类无机型有机型无机有机复合型有机有机复合型可以说各有各的特点复合型酸化剂酸化剂较为人们认可最初酸化剂主要功能被用于降低饲料系酸力但目前的饲料配方设计理念中饲料的系酸力已经比较低因此纯无机的酸化剂逐渐淡出人们

Jak vyplývá z tohoto vzorku textu, nejvyšší četnost mají dvouprvkové znaky, které tvoří téměř polovinu korpusu, zatímco frekvence jednoprvkových a tříprvkových znaků je přibližně stejná, okolo 22 %. Znaky se čtyřmi a více prvky se v textu vyskytují v menší míře, v rámci této množiny se však vyskytl pouze jeden znak s pěti prvky a žádný šestiprvkový znak se zde nevyskytl. Tato ukázka tedy stejně jako v předchozím oddílu opět souhlasí s výsledky kvantitativní analýzy.

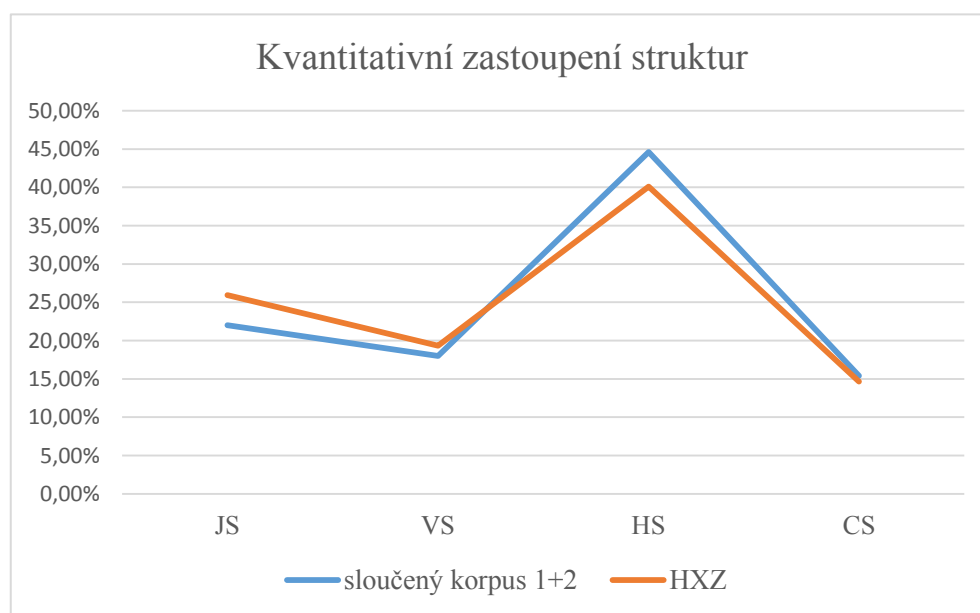
3.4.Struktury

V rámci analýzy struktur bylo v zásadě potvrzeno to, co již vyplývá z analýzy tahů a prvků, vzhledem k nejčastějšímu výskytu dvouprvkových znaků, které se nejčastěji člení horizontálně. Z pohledu kvalitativního se v analýze strukturních typů našly shody i nesrovnalosti mezi jednotlivými dílčími korpusy a slovníkem HXZ. Všechny analýzy uvádí mezi „typy“ jako nejčastěji používanou horizontální strukturu (korpus 1+2 54,12 %, HXZ 64,93 %) a vertikální strukturu (korpus 1+2 23,14 %).

Zajímavý je však rozdíl mezi statistikou sloučeného korpusu 1+2 a slovníkem HXZ. Obě statistiky sice shodně uvádějí, že třetím nejčastějším strukturním typem je centrální struktura (korpus 1+2 11,20 %, HXZ 9,18 %) v mém korpusu však procento výskytu této struktury jen těsně překračuje procento výskytu jednoduché struktury (10,75 %), zatímco u statistiky slovníku HXZ je tento rozdíl více než dvojnásobný (jednoduchá struktura 4,15 %, centrální struktura 9,18 %).

Vzhledem k tomu, že znaky s jednoduchou strukturou, v mém korpusu tak četné, byly různých druhů nelze říci, že by jednoznačně byly charakteristické pouze pro populárně naučný text z oblasti fyziky a chemie. Tento fakt tak tedy dle mého názoru poukazuje spíše na relativně malou velikost korpusu, než na zásadní rozdíl ve skladbě znaků obecného a populárně naučného textu.

Zatímco kvalitativní analýza přináší rozdílné výsledky, analýza kvantitativní zcela potvrzuje, že obecné statistiky o výskytu strukturních typů ve znacích obecného textu platí také pro jiné typy textů bez ohledu na jejich typ, styl, nebo funkci, což můžeme vidět na grafu kvantitativního zastoupení struktur v textu na další straně. Stejně jako v kvalitativní analýze i zde v textu převládají znaky v horizontální struktuře (44,60 %), jejichž výskyt je v textu oproti druhé nejčastější struktuře dvojnásobný.



Graf 6: Kvantitativní zastoupení struktur

Na rozdíl od kvalitativní analýzy však za horizontální strukturou následují znaky jednoduché (22,01 %), které ačkoli netvořily velkou část seznamu, hojně se v reálném

textu opakovaly. Vertikální struktura (17,98 %) se v kvantitativní analýze umístila tedy až za jednoduché znaky spolu s centrální strukturou (14,40 %), což je poměrně zajímavý fakt, vzhledem k tomu, že data kvalitativní analýzy se výrazně lišila, ale data kvantitativní analýzy jsou si velmi podobná. Také bych zde chtěla zmínit, že všechny výsledky výskytu strukturních typů v textu jak ve sloučeném korpusu 1+2 tak i v slovníku HXZ se od sebe liší maximálně v rozmezí 5 %, což ukazuje na vypovídající úroveň této analýzy.

3.4.1. Ukázka frekvence zastoupení strukturních typů v textu

Tento oddíl slouží stejně jako předchozí oddíly tohoto typu v kapitolách 3.2.1 a 3.3.1. k názorné ilustraci frekvence zastoupení jednotlivých strukturních typů v textu mého korpusu. Opět jsem zde vycházela ze stejného textu jako v předchozích dvou kapitolách, aby se daly mezi sebou lépe srovnávat. Horizontální struktura je v textu vyznačena žlutou barvou, vertikální struktura zelenou barvou, jednoduché znaky jsou zvýrazněny červeně a centrální struktura modře.

Po barevném označení všech znaků v ukázkové části textu korpusu 1 se vizualizuje obecný předpoklad o vysoké četnosti horizontální struktury. Díky velkému obsahu žluté plochy lze jasně vidět, že je horizontální členění zcela dominantním strukturním typem, který zabírá více než polovinu textu.

有机酸的选择实验版前言目前酸化剂种类繁多琳琅满日本文旨在通过试验文献来评价各有机酸的特点为大家提供一定的参考依据本人才疏学浅文章如有不对之处烦请大家指正酸化剂功能和种类酸化剂具有调控动物肠道微生物平衡增殖有益菌抑制有害菌降低肠道值提高消化道酶的活性提高营养物质消化率减慢胃排空速度提高动物抗应激和免疫力等方面的功能目前市面上的酸化剂分为四类无机型有机型无机有机复合型有机有机复合型可以说各有各的特点复合型酸化剂酸化剂较为人们认可最初酸化剂主要功能被用于降低饲料系酸力但目前的饲料配方设计理念中饲料的系酸力已经比较低因此纯无机的酸化剂逐渐淡出人们

Tento fakt se dá jednoduše vysvětlit pomocí výsledků analýzy celého korpusu, která prokázala, že znaků s horizontální strukturou je v čínštině celkově nejvíce. Obecně je tato vysoká frekvence horizontální struktury v textu způsobena například vysokou frekvencí výskytu znaku 的 *de*, vzhledem k jeho gramatické funkci.

Je třeba zmínit, že centrální struktura, tak jak jsem si ji definovala zde, do sebe inkorporuje také znaky obecně patřící do rámcové struktury a také znaky patřící do speciálního typu hnízdící struktury uváděné ve slovníku HXZ, vzhledem k tomu, že jejich zastoupení v textu bylo většinou poměrně zanedbatelné.

Zároveň byla v rámci celého textu mezi prvními 200 nejfrekventovanějšími znaky, které samy o sobě zastupovaly většinu textu, největší část členěna horizontálně a následovaly znaky jednoduché, což podporuje správnost výsledků analýzy frekvence strukturních typů. Jak jsem tedy z této názorné ukázky frekvence strukturních typů ve vzorku korpusu 1 zjistila, je možné se domnívat, že v rámci krátkého úseku textu je jeho typ schopen ovlivnit četnost určitých strukturních typů, avšak v textech, které čítají větší množství znaků, se tento rozdíl stírá a jsou potvrzeny obecné statistiky.

3.5. Shrnutí výsledků

V dílčích kapitolách jednotlivých analýz byly některé výsledky poměrně překvapující, a je proto vhodné je zde stručně shrnout. Analýza tahů vykázala relativně významné rozdíly mezi kvalitativní a kvantitativní statistikou, což byl jev, který se v menší míře vyskytl i v analýze prvků a struktur. Na základě shromážděných dat o četnosti výskytu určitých počtů tahů bylo zjištěno, že velikost korpusu nemá na průměrný počet tahů ve znaku tak zásadní vliv, jak by se dalo očekávat.

Bylo zjištěno, že v mém seznamu znaků kvalitativní analýzy je nejvíce znaků o počtu tahů mezi šesti a dvanácti, se zcela nejvyšším výskytem znaků s osmi, devíti a deseti tahy. Kvantitativní analýza zastoupení počtu tahů v rámci skutečného textu však ukazuje, že mají oproti kvalitativní statistice nejvyšší zastoupení znaky se čtyřmi až deseti tahy. To je zcela zásadní poznatek, který ukazuje na to, že ačkoli kvalitativní analýza udává počty tahů ve znacích, které se v korpusu vyskytly, není schopna reálně zobrazit výskyt těchto znaků v textu, a v důsledku toho není pro korpus příliš vypovídající.

Nejvíce se v soupisce znaků i v textu ale vyskytují znaky s osmi tahy, což je poměrně zajímavé a ukazuje to na opravdu vysokou četnost těchto znaků, vzhledem

k tomu, že se shodly kvalitativní i kvantitativní pohledy. Avšak co se týče četnosti v textu, za osmi-tahovými znaky následují šesti-tahové a pěti-tahové znaky, což je významný rozdíl oproti kvalitativní analýze a značí to, jak už jsem zmínila dříve, významnou preferenci užívání znaků s malým počtem tahů.

Všechny tyto poznatky jsou dobře znázorněny na níže uvedeném porovnání tabulky kvalitativní a kvantitativní analýzy počtu tahů, kde tmavě růžová barva znázorňuje nejvyšší četnost výskytu již zmíněných osmi-tahových znaků středně a světle růžová pak druhý a třetí nejvyšší počet výskytů v soupisce i v textu.

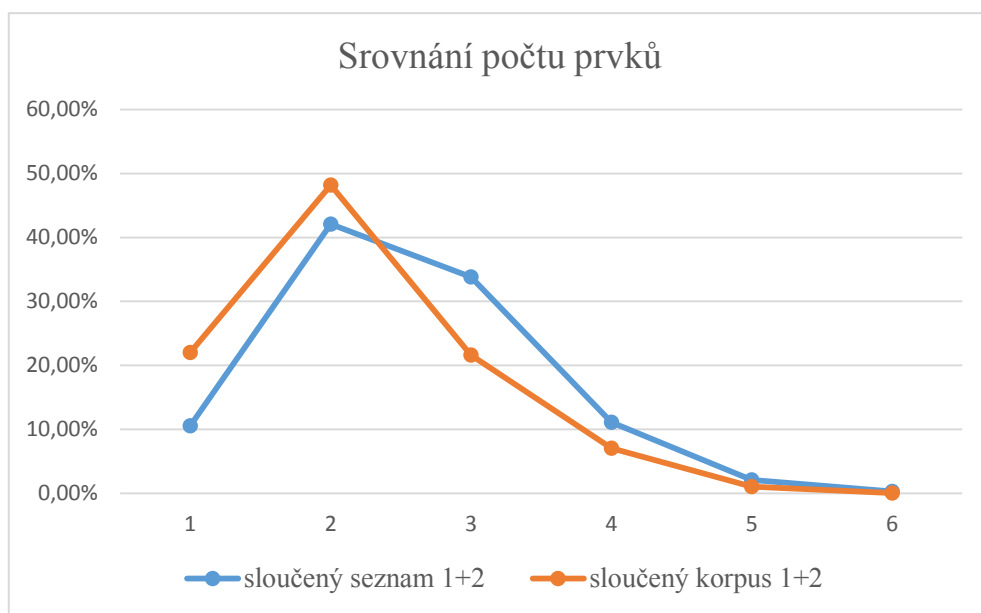
	4	5	6	7	8	9	10	11	12
kvalitativní	4,49 %	5,82 %	8,09 %	10,41 %	12,23 %	11,54 %	10,61 %	8,63 %	7,99 %
kvantitativní	8,20 %	9,79 %	12,88 %	9,15 %	17,88 %	9,04 %	8,08 %	5,13 %	4,38 %

Tabulka 4: Srovnání kvalitativní a kvantitativní analýzy tahů

Stejně jako v analýze tahů i v analýze prvků byla jistá odlišnost ve výsledcích kvalitativní a kvantitativní analýzy, což lze vidět v Grafu 7, který porovnává výsledky těchto dvou pohledů na výskyt počtu prvků. Jak můžeme z grafu vyčíst, jedná se o rozdíl v počtu prvků, kdy kvalitativní analýza uvádí nejvyšší četnost v rámci soupisky znaků u dvouprvkových, tříprvkových a čtyřprvkových znaků.

Oproti tomu kvantitativní analýza udává nejvyšší četnost znaků dvouprvkových, jednoprvkových a tříprvkových přesně v tomto pořadí. Můžeme tedy jasně vidět, že se sice v soupisce vyskytuje poměrně velké množství víceprvkových znaků, ale jejich reálné využití v textu je malé oproti znakům s menším počtem prvků.

Rozdíl ve statistice výskytu strukturních typů vznikl tentokrát pouze mezi kvalitativní a kvantitativní analýzou, avšak je důležité zmínit, že data kvantitativní analýzy se přesně shodují s obecnými statistikami slovníku HXZ, což poměrně důležitý poznatek vzhledem k zaměření této bakalářské práce. Znamená totiž, že statistika výskytu strukturních typů ve znacích obecného textu platí i pro populárně naučný text z velmi specifického odvětví a zásadnějším prvkem, který ovlivňuje kredibilitu analýzy je spíše velikost textu.

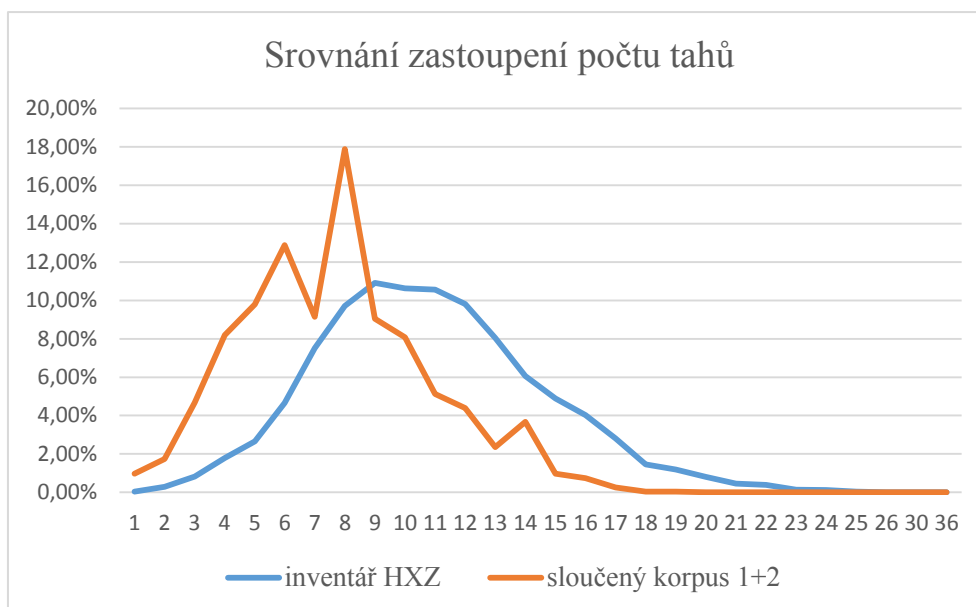


Graf 7: Srovnání počtu prvků kvalitativní a kvantitativní analýzy

Dle kvantitativní analýzy textu dominuje horizontální struktura, tvořící polovinu celého textu, která je zároveň nejčastější struktura i mezi jednotlivými znaky v soupisce. Následují ji však jednoduché znaky, které se v kvalitativní analýze na prvních třech pozicích ve sloučeném korpusu 1+2 a slovníku HXZ ani nevyskytly. Zajímavé však je to, že v rámci kvalitativní analýzy jednotlivých dílčích korpusů jsou jednoduché znaky třetí nejčetnější struktura, avšak po jejich spojení jim již předcházejí znaky s centrální strukturou.

3.5.1. Srovnání korpusu se znakovým inventářem slovníku

V této části budu podrobněji srovnávat výsledky své analýzy korpusu čítajícího 86 745 znaků se znakovým inventářem slovníku HXZ. Ten čítá 7785 znaků, které byly extrahovány z obecného čínského textového korpusu o 21 656 578 znacích. Pro znázornění srovnání výsledků analýz tahů používám níže uvedený graf srovnání kvalitativní analýzy tahů inventáře HXZ a kvantitativní analýzy tahů v korpusu 1+2. Ve znakovém inventáři slovníku se vyskytují znaky obsahující jeden až třicet šest tahů, zatímco v mém korpusu byl nejvyšší počet tahů pouze dvacet tři.



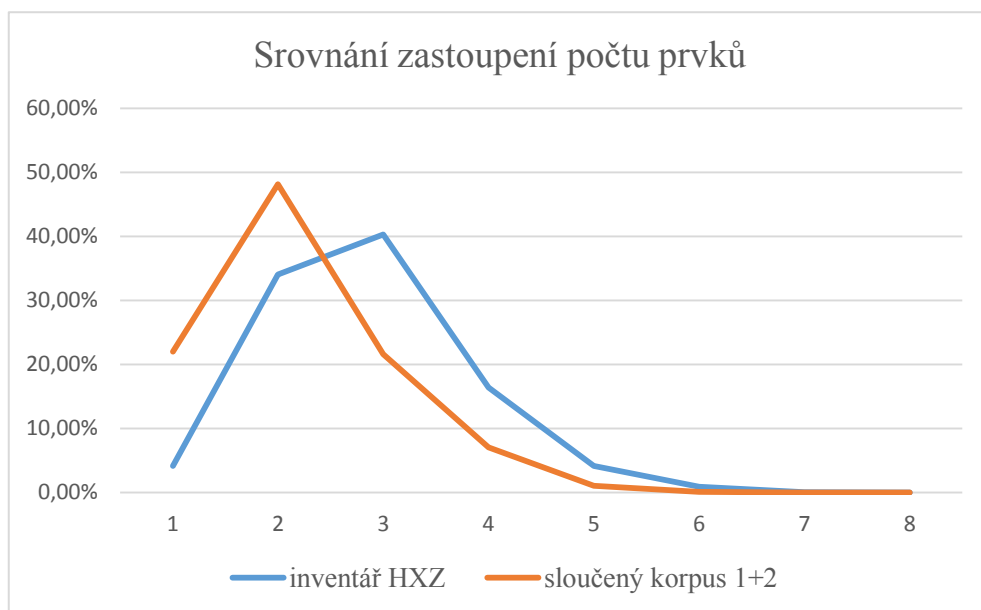
Graf 8: Srovnání zastoupení počtu tahů korpus vs. inventář

Průměrný počet tahů ve znaku tohoto inventáře je 10,94 a zcela nejfrekventovanější jsou devítitahové znaky. V populárně naučném textu je však nejvyšší frekvence znaků s osmi tahy. Nejčetnější výskyt v rámci znakového inventáře slovníku měly znaky s počtem sedmi až třinácti tahů, zatímco analýza korpusu 1+2 uvádí nejvyšší frekvenci užívání znaků se čtyřmi až devíti tahy.

Zatímco křivka frekvence počtu tahů ve znacích z inventáře HXZ postupně stoupá od sedmitahových znaků nahoru, dostává se do svého vrcholu u devítitahových znaků a klesá až u třináctitahových znaků, křivka frekvence počtu tahů v textu takto hladký průběh nemá. Obsahuje řadu výkyvů ve frekvenci užití, jako například u sedmitahových znaků, u kterých je zaznamenán značný propad, oproti šesti a osmitahovým znakům. Po dosažení své nejvyšší hodnoty u osmitahových znaků prudce klesá ke znakům devítitahovým, které jsou poslední s relativně vysokou frekvencí výskytu v textu.

Tyto výsledky poukazují na značný rozdíl v mezi tím, jak graficky komplikované znaky se vyskytují v inventáři slovníku a ve skutečném textu. Zatímco slovník shromažďuje data o mnohotahových znacích, jejichž reálná četnost užití se však blíží nule, princip jazykové ekonomie,⁷² kdy převládá preference častého užívání znaků s nízkým počtem tahů v textu, je zřejmý.

⁷² ČERNÝ, Jiří. Úvod do studia jazyka. str. 149.

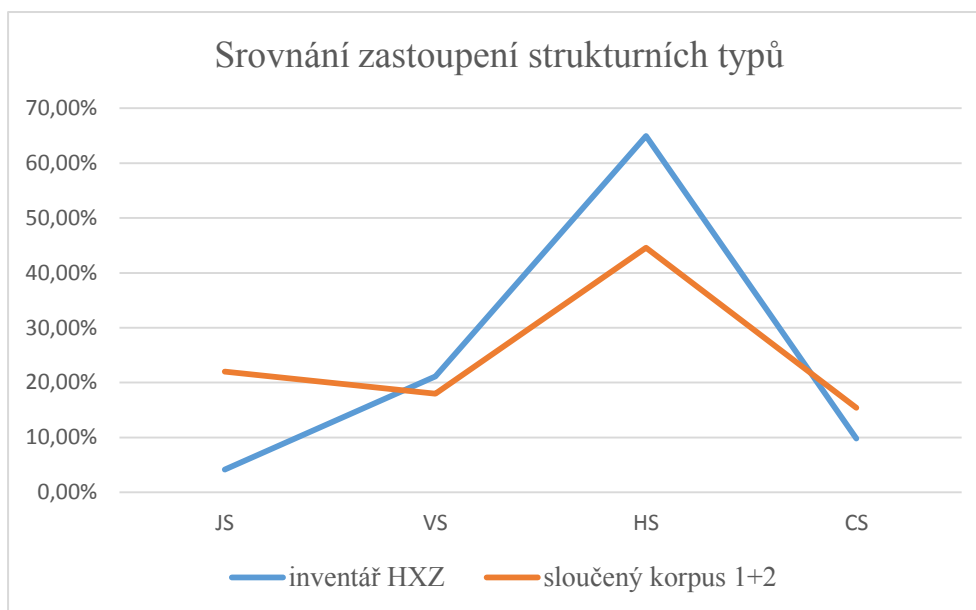


Graf 9: Srovnání zastoupení počtu prvků korpus vs. inventář

Výsledky analýzy počtu prvků v rámci inventáře slovníku HXZ a mého korpusu jsou znázorněny na Grafu 9. Jak z průběhu obou křivek v grafu vyplývá, kvalitativní a kvantitativní analýza se odlišuje ve všech hodnotách. Nejčetnější jsou v inventáři slovníku znaky se třemi prvky, následují je znaky se dvěma prvky a třetí v pořadí jsou znaky se čtyřmi prvky, jak ukazuje modrá křivka. V mém textovém korpusu však měly nejvyšší frekvenci znaky dvouprvkové, za nimi následovaly znaky jednoprvkové a pak až znaky tříprvkové, což je v grafu znázorněno oranžovou křivkou.

Výsledky analýzy počtu tahů a počtu prvků spolu poměrně úzce souvisí, jak se zde prokázalo. Z interpretace výše uvedeného grafu je tedy zřejmé to, co jsem již popsala v předchozí analýze tahů. S narůstajícím počtem tahů se zároveň zvyšuje také počet prvků, což je zachyceno v inventáři slovníku. Vzhledem k tomu, že jsou však v reálném textu upřednostňovány znaky s malým počtem tahů, které se zároveň velmi často opakují, logicky tím klesá i počet prvků, ze kterých se tyto znaky skládají a proto jsou hodnoty kvantitativní analýzy korpusu tolik odlišné od analýzy kvalitativní.

Ze srovnání výskytu strukturních typů v rámci znakového inventáře slovníku HXZ a v rámci mého sloučeného korpusu 1+2 opět vyplývá míra odlišnosti pohledu na vlastnosti znaku v rámci kvalitativního a kvantitativního hlediska. Jak znázorňuje Graf 10 níže, korpus 1+2 i inventář slovníku HXZ uvádějí jako nejčetnější strukturu horizontální, což vzhledem k ostatním výsledkům v této práci i jiným výzkumům podobného zaměření není příliš překvapivé.



Graf 10: Srovnání zastoupení strukturních typů korpus vs. inventář

Kvantitativní pohled na korpus 1+2 však udává jako druhou nejčetnější skupinu jednoduché znaky, což s kvalitativním pohledem na inventář slovníku HXZ nesouhlasí, neboť uvádí jako druhou nejčetnější vertikální strukturu. Rozdíl byl zde v podstatě očekávaný, neboť kvalitativní analýza znaků v inventáři slovníku HXZ nezachycuje jejich skutečné využití v textu. Zatímco se v něm tedy vyskytuje více znaků s vertikální strukturou než znaků jednoduchých, v textu jsou jednoduché znaky mnohem čtější. Tento jev se dá vysvětlit tím, že jednoduchou strukturou jsou řazeny v čínštině piktogramy a symboly, jakožto základní znaky, ze kterých vznikaly znaky odvozené. Obecně v jazycích platí, že je ve frekvenčních seznamech mnohem více slov odvozených, než základních, avšak využití již zmíněných základních znaků je přesto vysoké.⁷³

Znovu tedy výsledky práce poukazují na fakt, že údaje o četnosti strukturních typů znaků ve slovnících, jsou často pouze produktem jejich kvalitativní analýzy. Tyto hodnoty vycházejí z četnosti v rámci znaků v základním inventáři moderní čínštiny, avšak reálná četnost strukturních typů v textech, kterou hodnotí kvantitativní lingvistika je někdy opomíjena.

Má analýza tedy dokládá, že v souvislých textech se znaky z hlediska své struktury chovají poněkud jinak, než v inventářích znaků příslušících daným textům. To potvrzuje obecně platné teorie o strukturním chování prvků v souvislých textech, které se dají

⁷³ ČERNÝ, Jiří. Úvod do studia jazyka. str. 149.

aplikovat na všechny jazyky, čínštinu nevyjímaje.⁷⁴ Četné výskyty číslovek, sloves a jiných syntakticky důležitých prvků v textu vždy zásadně ovlivní četnost výskytu strukturních typů, pod které spadají, avšak jsou obsaženy stejnou měrou v různých typech textů. Porovnání kvalitativního a kvantitativního pohledu na daný korpus je tedy důležité pro pochopení, které znaky jsou v textech skutečně časté.

⁷⁴ ČERNÝ, Jiří. Úvod do studia jazyka. str. 198-199.

4. ZÁVĚR

Hlavním cílem této práce bylo analyzovat grafickou podobu znaků vyskytujících se v populárně naučných textech z oblasti fyziky a chemie. Analýza se zaměřila na tři hlavní grafické aspekty čínského znaku, kterými jsou tah, prvek a strukturní typ. Tyto aspekty analyzovala z pohledu kvalitativního, ten vytváří statistiku z jedinečných znaků „typů“ v textu a kvantitativního, jež vytváří statistiku ze všech znaků, které se v textu opakovaně vyskytly. Tyto dva odlišné pohledy byly v konečné části analýzy mého korpusu, čítajícího 86 745 znaků, vyhodnoceny a porovnány s daty o znacích obsažených v inventáři slovníku HXZ a bylo zjištěno, že se tyto data od sebe zásadním způsobem liší.

Jak jsem zjistila v předešlé části práce, výsledky mé kvalitativní analýzy byly v některých bodech od obecných statistik mírně odlišné, avšak z větší části se nimi shodovaly v závislosti na velikosti zkoumaného korpusu. V souladu s mým původním očekáváním, se největší rozdíly v zastoupení počtu tahů, prvků a struktur v mé práci vyskytly nikoli mezi populárně naučným textem a obecným textem, jak by se na první pohled mohlo zdát, nýbrž mezi kvalitativní a kvantitativní analýzou všech textů bez výjimky.

To dokazuje, že při analýze zastoupení četnosti jakýchkoli aspektů čínského textu je vždy nutné zohlednit oba tyto pohledy. Jak jsem již v této práci zmiňovala, je totiž nezbytné na text pohlížet nejen jako na sadu znaků, které se v něm vyskytly, ale především jako na pomyslnou buňku, která se skládá z různých počtů jednotlivých částí, které mezi sebou mají určité vztahy. Kvalitativní hledisko je tedy důležité, pokud chceme analyzovat strukturní členění čínských znaků jako takových, zatímco kvantitativní hledisko se zabývá členěním znaků a jejich výskytem z pohledu celého textu. V této práci jsem tyto dva pohledy zohlednila tak, aby rozdíl mezi nimi byl pokud možno co nejzřetelnější.

Na základě analýzy provedené v této práci jsem dospěla k několika zjištěním. Přestože existuje mnoho různých žánrů a typů čínských textů, jejich znaková skladba je v zásadě vždy velmi podobná, protože obsahují stále stejné znaky, vycházející ze základní znakové sady. Bez ohledu na svůj typ, musí totiž čínský text vždy obsahovat řadu formálních a gramatických prvků, bez kterých by nebylo možné poskládat smysluplnou větnou konstrukci a četnost těchto znaků pak ovlivňuje dané statistiky.

Dalším zjištěním je, že ačkoli se v čínštině vyskytuje velké množství komplikovaných znaků s mnoha tahy, jejich reálná frekvence užití v textu je malá.

Z výzkumů vyplývá, že čínština stejně jako ostatní jazyky inklinuje k co největší grafické úspornosti často používaných znaků, které však zároveň zaznamenávají adekvátní míru informací. Při zpracování této bakalářské práce byly tedy identifikovány rozdíly v chování znaků v rámci souvislého populárně naučného textu a inventáře moderní čínštiny, čímž byla potvrzena prvotní stanovená hypotéza, že tento rozdíl mezi kvalitativní a kvantitativní rovinou existuje.

Největší odchylky se vyskytly ve statistikách četnosti tahů a prvků ve znacích, kdy inventář HXZ upřednostňoval znaky s větším počtem tahů a prvků, než které byly nejčetnější v souvislém textu. Stejně tak v rámci četnosti strukturních typů uváděl slovník vyšší počet znaků s komplikovanou grafickou strukturou, než jak je tomu ve skutečném textu. Z těchto výsledků je tedy zřejmé, že pohled na znak jako samostatnou jednotku a znak jako součást celistvého textu se značně liší a je důležité brát v potaz jak kvalitativní, tak kvantitativní hledisko.

Přestože se již čínskými znaky zabývala řada jazykovědců a byly stanoveny normy ohledně jejich grafické podoby, stále zůstává mnoho nedořešených mnoho nesrovnalostí, týkajících se grafické struktury znaků a její dekompozice. Věřím, že tato práce může představovat přínos v rámci podrobnějšího výzkumu strukturního členění čínských znaků a mohla by být také dále rozšířena o analýzu dekompozice prvků vyšších úrovní, která by představila ještě podrobnější pohled na tuto často opomíjenou problematiku. Další zajímavou možností by bylo rovněž srovnání výsledků analýzy s daty, vyhodnocenými v obdobných výzkumech, zabývajících se například literárním či publicistickým stylem.

Resumé

This thesis examined the graphical form of Chinese characters included in scientific texts that dealt with chemistry and physics topics. Chinese characters can be divided into three levels of decomposition, strokes, components and structures that form the final character. The objective of this research was to define the percentage of each of these three levels of decomposition and find out, if there are differences in graphical structures between characters contained in scientific text and characters contained in the inventory of characters in modern Chinese. The task was divided into two main parts. In the first theoretical part, I introduced the origin and evolution of Chinese script, briefly mentioned the etymology of Chinese characters and outlined crucial scholarly views on their composition and structure. In the second analytical part I analysed the above mentioned three levels of decomposition of characters in two separate text corpuses composed of scientific texts dealing with chemistry and physics and later on joined the two into a final corpus of 86 745 characters. After gathering all the data I discovered that there are considerable differences in frequency of structure types, components and numbers of strokes between the text corpus and the character inventory of dictionary HXZ. Tendencies of the character inventory to indicate complicated characters as more frequent, than how it is in the real text proved to be existing. This indicates the need to take both qualitative and quantitative linguistics approach into account while analysing text corpuses and not to mistake tokens for types.

Key words:

Chinese characters, graphical structure, stroke, component part, structural form

SEZNAM LITERATURY

Monografie:

ČERNÝ, Jiří. *Úvod do studia jazyka*. 2. vyd. Olomouc: Rubico, 2008, 248 s. ISBN 978-80-7346-093-8.

KUČERA, Ondřej a kol. *Učebnice čínských znaků*. Olomouc: Univerzita Palackého, 2005, 237 s. ISBN 80-244-1118-0.

SLAMĚNÍKOVÁ, Tereza. *Ideogramy v moderní čínštině*. Olomouc: Univerzita Palackého v Olomouci, 2013, 164 s. ISBN 978-80-244-3909-9.

SLAMĚNÍKOVÁ, Tereza. *Synchronní model tradiční kategorizace čínského znakového písma: Dizertační práce*. [nepublikováno] Olomouc: Univerzita Palackého, 2015, 231 s.

ŠULC, Michal. *Korpusová lingvistika: první vstup*. Praha: Karolinum, 2000, 94 s. ISBN 8071848476.

ŠVARNÝ, Oldřich. *Úvod do hovorové čínštiny*. Praha: Státní pedagogické nakladatelství, 1967, 458 s.

TĚŠITELOVÁ, Marie. *Quantitative linguistics*. Praha: Academia, 1992, 256s. ISBN 80-200-0131-X.

UHER, David. *Hanská grammatologie*. Olomouc: Univerzita Palackého v Olomouci, 2013, 334 s. ISBN 978-80-244-3866-5.

VOCHALA, Jaromír. *Chinese writing system: minimal graphic units*. Praha: Univerzita Karlova, 1986, 143 s. Acta Universitatis Carolinae. Philologica. Monographia LXX/78.

VOCHALA, Jaromír, Miroslav Novák a Vladimír Pucek. *Úvod do čínského, japonského a korejského písma I. Vznik a vývoj*. Praha: Státní pedagogické nakladatelství 1989a, 132 s.

VOCHALA, Jaromír, Miroslav Novák a Vladimír Pucek. *Úvod do čínského, japonského a korejského písma II. Praktický kurz*. Praha: Státní pedagogické nakladatelství 1989b, 250 s.

WILDER, George D., INGRAM James H.: *Analysis of Chinese Characters*. New York: Dover, 1974, 263 s. ISBN 0486230457.

ZÁDRAPA, Lukáš a Michaela Pejčochová. *Čínské písmo*. Praha: Academia, 2009, 297 s. ISBN 978-80-200-1755-0.

Články v odborných periodících:

SEHNAL, David. „Čínské znakové písmo, jeho povaha a vývoj“. In *Z myšlenek a představ Žluté země*. Brno: Moravské zemské muzeum, 2002, s. 145-160.

SLAMĚNÍKOVÁ, Tereza. „Nové trendy v čínské grammatologii“. *Dálný východ*. 2011, č. 1, s. 13-32. ISSN 1805-1049.

Slovníky:

Hanzi Xixi Zidian 《汉字信息字典》 [Informační slovník čínských znaků]. Peking: Kexue Chubanshe, 1988, 1346 s. ISBN 7-03-000869-3.

WENLIN INSTITUTE, Inc. 文林 Wenlin Software for Learning Chinese [Software]. Version 4.0.2. Wenlin Institute, Inc. © 1997-2011.

Internetové zdroje:

Dayah, M. *Dynamic Periodic Table* [online]. 1997 [cit. 2017-04-15]. Dostupné z: <http://www.ptable.com>

Wenkubaidu. 汉字结构类型图示表 [online]. 2013 [cit. 2017-04-15]. Dostupné z: <https://wenku.baidu.com/view/11ec942d3968011ca30091f8.html>

Seznam příloh

Příloha č. 1: Frekvenční list korpusu 1 (CD ROM)

Příloha č. 2: Frekvenční list korpusu 2 (CD ROM)

Příloha č. 3: Celková tabulka korpusu 1 a korpusu 2 (CD ROM)

Příloha č. 4: Tabulka analyzovaných jevů (CD ROM)