

Univerzita Karlova

Filozofická fakulta
Ústav asijských studií

Bakalářská práce

Daniel Knapp

**Ortografická variabilita
pomocných podstatných jmen v japonštině**

**Orthographic variability
of auxiliary nouns in contemporary Japanese**

Děkuji všem, kdo mě při psaní této práce podporovali, především pak vedoucí Mgr. Petře Kanasugi Ph.D. za obrovskou trpělivost i cenné podněty, bez kterých by tato práce nikdy nedosáhla stávající podoby.

Prohlašuji, že jsem bakalářskou práci vypracoval samostatně, že jsem řádně citoval všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

Anotace

Tato práce se zabývá ortografickou variabilitou pomocných podstatných jmen. V teoretické části se zabývá zavedením pojmů a postupů, které s problematikou souvisí a jsou využity v následující analytické části, kde se porovnávají způsoby zápisu vybraných pomocných podstatných jmen v jednotlivých subkorpusech korpusu BCCWJ a hledají se významové souvislosti ve větách z těch subkorpusů, na které analýza upozorní.

Klíčová slova

japonský jazyk, pomocná podstatná jména, ortografie, způsob zápisu, variabilita

Annotation

This thesis deals with orthographical variability of Japanese auxiliary nouns. In the theoretical part, there is a definition of used terms and methods related to the following analytical section, where various ways of writing Japanese auxiliary nouns in BCCWJ corpus subcorpora are compared. After analysis identifies subcorpora of interest, sentences containing auxiliary nouns in said subcorpora are checked for similarities in meaning.

Keywords

Japanese language, auxiliary nouns, orthography, way of writing, variability

Obsah

Seznam použitých zkratek.....	7
Úvod.....	8
1. Způsoby zápisu japonštiny.....	10
2. Pomocná podstatná jména.....	11
3. Gramatikalizace.....	14
4. Preskriptivní pravidla	16
5. Korpus	17
6. Metoda analýzy	19
7. Uči (うち、ウチ、内).....	23
8. Šita (した、シタ、下)	29
9. Koto (こと、コト、事)	33
10. Baai (ばあい、バアイ、場合)	37
11. Hodo (ほど、ホド、程).....	41
12. Fú (ふう、フウ、風)	45
13. Cumori (つもり、ツモリ、積り、積もり)	49
14. Jotei (よてい、ヨテイ、予定).....	53
15. Výsledky a diskuse	56
16. Závěr	58
Seznam použité literatury.....	60
Seznam příloh.....	62

Seznam použitých zkratek

BCCWJ: Balanced Corpus of Contemporary Written Japanese

NINJAL: National Institute for Japanese Language and Linguistics

Úvod

Japonština je zajímavá tím, že využívá několik způsobů zápisu, které se vyskytují současně v rámci jedné věty. I když existuje prototypické užití každého z nich, nic nebrání autorům se u jednoho slova rozhodnout v různých situacích pro netypickou možnost zápisu. Toto japonštině propůjčuje stylistický rozměr, kdy je možné rozlišit dvě věty s totožnými slovy, slovosledem i významem, které se liší pouze nuancí nesenou způsobem zápisu.

Ortografie neboli pravopis je souborem dnes autoritativních předpisů k převodu mluveného jazyka do psaného. (Čermák 2009, s. 31) Pro japonštinu existuje seznam základních používaných znaků a jejich čtení. V platnosti jsou také vládní výnosy, které zápis upravují.¹ Přesto si lze všimnout jisté variability používaných způsobů zápisu.

Tato variabilita se již stala předmětem lingvistického výzkumu. Některé práce popisují ortografickou variabilitu jako celek, zkoumali, kterých slov se týká a jaké varianty se objevují, například Ogura (2012). Jiné se soustředí na určitou skupinu slov a tu podrobují podrobné analýze, například Ogura (2013), Kashino a Nakamura (2013), Barešová a Schir (v tisku). Liší se i použité metody. Někteří autoři (Ogura 2012, Ogura 2013, Kashino a Nakamura 2013) analyzovali data získaná z korpusu, jiní (Barešová a Schir, v tisku) se dotazovali rodilých mluvčích.

Tato práce zkoumá ortografickou variabilitu pomocných podstatných jmen pomocí analýzy korpusu. Od ostatních citovaných prací zabývajících se ortografickou variabilitou se odlišuje vymezením předmětu výzkumu - zkoumaným slovním druhem - a tím, že důsledně ověřuje statistickou významnost pozorovaných údajů.

K vyšetření ortografické variability se nabízí logistická regresní analýza, kterou Březina (2018) popisuje slovy: *„technika, která používá vysvětlující proměnné k odhadnutí jejich vlivu na lingvistickou*

¹ Zmíněný seznam, džójó kandži (常用漢字), je v podstatě také vládním výnosem, konkrétně výnosem č. 2 z 30. 11. 2010. Dalším výnosem, který zápis japonštiny upravuje je výnos č. 1 z téhož dne, který je v této práci zmiňován dále.

proměnnou.“ V kontextu této práce je za lingvistickou proměnnou zvolen způsob zápisu daného pomocného podstatného jména. Nabízí se tedy použít vhodné údaje v korpusu současné psané japonštiny BCCWJ jako vysvětlující proměnné, použít popsaný postup a ověřit, zda-li se takto získané údaje odlišují od údajů získaných alternativním postupem, v tomto případě vyčtením z přehledových tabulek.

1. Způsoby zápisu japonštiny

Protože je tato práce zaměřena na variabilitu v použití různých způsobů zápisu pomocných podstatných jmen v japonštině, je nejprve nutné tyto způsoby zápisu a jejich prototypické použití představit. Hlavní způsoby zápisu jsou tři. Jde o zápis pomocí dvou slabičných abeced, *hiragany* a *katakany*, a logografických, původem čínských, znaků *kandži*. Tyto znaky se zpravidla používají pro zápis *kořenů* slov, *hiragana* pro zápis gramatických partikulí a skloňovatelných sufixů apod. a *katakana* pro přepis cizích slov. Tato velice jednoduchá pravidla ovšem vykazují celou řadu výjimek. Existují slova, která se zapisují *hiraganou* celá, ačkoliv pro ně existuje znak a běžně se používá, pro některá slova je *hiragana* výchozím způsobem zápisu a logografický způsob zápisu se nepoužívá, a *katakanou* se zapisují slova, která je třeba v textu zvýraznit. (Nymburská et al. 2007)

Vedle tří hlavních způsobů zápisu se lze v japonském textu dále setkat s arabskými číslicemi a latinkou. Jejich užití však s variabilitou zápisu slov zkoumaných v této práci nesouvisí.

2.Pomocná podstatná jména

Prvním krokem při vyšetření ortografické variability je vymezení zkoumaného vzorku. Jednou z možností, jak k tomuto výběru přistoupit, je vybírat slova podle slovního druhu. Vybraný vzorek by měl být reprezentativní, aby bylo možné pozorované vlastnosti zobecnit ze vzorku zpět na celou kategorii.

Pro tuto práci je zkoumaný vzorek složen z pomocných podstatných jmen. Podstatná jména jsou v japonštině slovním druhem, který je nesklonný a samostatný.² (Novák 1989, s. 27) Definice pomocných podstatných jmen zní: „*Pomocná podstatná jména jsou podstatná jména, která ztratila vlastní význam a jako podstatné jméno fungují jen formálně.*“ (Jošikawa et al. 2003, s. 1) To znamená, že významu se jim dostává díky nějakému rozvíjejícímu členu, který se k nim formálně připojuje jako přívlástek. (Novák 1989)

Aby byl vzorek nějakého celku reprezentativní, musí být proveden správný výběr. Lze-li tento celek rozdělit do několika kategorií, potom nesmí zástupce žádné kategorie být ve výběru opomenut. Pomocná podstatná jména lze rozdělit do čtyř kategorií, (Novák 1989, s. 37) a sice: Pomocná jména upřesňující a rozlišující prostorové a časové vztahy (význam předložkový), pomocná jména vytvářející formální jádro přívlástku (význam zájmenný), pomocná jména vytvářející určení času, způsobu, příčiny apod. (význam příslovečný) a pomocná jména označující nutnost, pravděpodobnost, domněnku apod. (význam modální). (Novák, 1989, s. 37-41) Výběr zkoumaného vzorku Novákovu klasifikaci reflektuje, protože je pro potřeby této práce dostatečně podrobná.

2.1 Předložková pomocná podstatná jména

K těmto pomocným jménům se zpravidla pojí jako přívlástek buď jiná podstatná jména a celá konstrukce potom určuje prostorové nebo časové

² Samostatný v tomto kontextu neznamená schopnost stát ve větě samostatně, ale schopnost být rozvíjen jinými slovy. Nesamostatná slova nemohou být rozvíjena, pouze rozvíjí slova samostatná. (Novák 1989, s. 27)

vztahy, nebo v případě rozvití slovesem vznikají vedlejší věty časové. Pokud nejsou určena přívlastkem³, pak vytvářejí příslovečné určení času nebo místa a mohou být upřesněna jiným příslovcem. (Novák 1989, s. 39) Příklady pomocných jmen patřících do této kategorie jsou například *mae, ato, naka, uči*. (Novák, 1989, s. 38)

2.2 Zájmenná pomocná podstatná jména

Tato pomocná jména mají zájmenný význam v případě rozvití podstatným jménem i slovesem. Zájmenný význam znamená, že se pomocné podstatné jméno chová jako zájmeno. Vznikají tak vedlejší věty podmětné a předmětné. Mezi pomocná jména této kategorie patří například *mono, koto, tokoro*. (Novák, 1989, s. 39-41)

2.3 Adverbiální pomocná podstatná jména

Ve spojení s podstatnými jmény se chovají podobně jako předložková pomocná jména a blíže určují časové vztahy. Ve spojení se slovesem vytvářejí vedlejší věty časové, příčinné a způsobové. Zástupci této kategorie jsou například *toki, tame, jó, mama*. (Novák, 1989, s. 41)

2.4 Modální pomocná podstatná jména

Tato podstatná jména nesou modální význam, který upravuje platnost věty co do nutnosti, pravděpodobnosti nebo domněnky apod.. Na pomocná modální jména se pojí především spona a obvykle nebývají rozvíta jinými podstatnými jmény. Příkladem mohou být například *hazu, cumori, tokoro, jó* (Novák, 1989, s. 41)

2.5 Rozpoznávání pomocných podstatných jmen

Pokud je podstatné jméno v alespoň jednom z jeho užití možné zařadit do jedné z výše uvedených kategorií, pohlížíme na něj jako na pomocné. Přiřazení jména do kategorie probíhá na základě jeho původního

³ Přívlastkovým tvarem ukazovacích zájmen v tomto užití nicméně rozvíty být mohou. (Novák 1989, s. 39)

autosémantického a významových posunů evidentních v existujících konstrukcích.

Pomocná podstatná jména se typicky vyskytují v ustálených spojeních a konstrukcích. V případě, že se podstatné jméno objevuje v konstrukci, která vyžaduje přívlastek, lze na něj nahlížet jako na pomocné, protože ustálená spojení význam slov v nich obsažených posouvají, čímž se snadno naplní definice, že pomocná jména jsou na vlastní význam chudá.

3. Gramatikalizace

Formální definice gramatikalizace zní: „*Proces přechodu zvláště lexému nebo jeho lexikálního významu či funkce k významu či funkci gramatické, někdy spojený s konverzí v jiný slovní druh.*“ (Čermák 2009, s. 243) V souvislosti s pomocnými podstatnými jmény se lze setkat s gramatikalizovanými konstrukcemi, kde došlo k posunu pouze významu. Pomocné jméno zůstává stále podstatným jménem, jeho rozvíjející členy se k němu pojí stejně jako by se pojily k jakémukoliv jinému podstatnému jménu. Pouze význam celé konstrukce je jiný, než by odpovídalo součtu významů jednotlivých slov.

Dalším důvodem, proč uvažovat podstatná jména objevující se v gramatikalizovaných konstrukcích za pomocná, je fakt, že tyto konstrukce jsou vlastně výsledky gramatikalizace. Dnešní pomocná jména byla ve staré japonštině plnovýznamovými slovy, u kterých se postupem času posunul význam, a v některých spojeních se ustálila. (Novák 1989, s. 37) V těchto konstrukcích je nyní hlavním nositelem významu celá konstrukce a nikoliv podstatná jména v nich obsažená.

3.1 Míra gramatikalizace

Pomocná podstatná jména se liší množstvím ustálených spojení, ve kterých se objevují, jsou v tomto ohledu různě „produktivní“. Tuto vlastnost lze také chápat tak, že ukazuje na míru gramatikalizace jednotlivých pomocných podstatných jmen a mohla by mít vliv na výběr způsobu zápisu.

Pro ověření tohoto vlivu byla pro potřeby této práce definována *míra gramatikalizace* jako počet různých konstrukcí uvedených ve slovníku Nihongo bunkei džiten (日本語文型辞典; Sunagawa 1998), které obsahují dané pomocné podstatné jméno. Na základě této míry byla z každé skupiny zmíněné výše (viz kapitola 2) vybrána dvě pomocná jména tak, aby jedno mělo na svoji skupinu tuto míru v rámci své skupiny vysokou, a druhé nízkou.

Spíše než o hodně gramatikalizovaném pomocném podstatném jménu obecně má smysl uvažovat o více či méně gramatikalizovaném pomocném jménu v rámci nějaké skupiny, nebo tuto míru gramatikalizace použít jako skóre k porovnání dvou pomocných podstatných jmen.

Takto formulovaná definice má ale i nedostatky. Není schopna rozlišit, jak moc jsou tyto konstrukce skutečně reprezentativní vzhledem k pomocnému jménu. Teoreticky může existovat podstatné jméno, které se vyskytuje v pouze gramatikalizované formě v rámci jedné konstrukce, která je pro něj ale prototypická a jinde než v ní se neobjevuje, a naproti tomu může existovat takové podstatné jméno, které má míru gramatikalizace vyšší, nicméně všechny konstrukce jsou pro něj okrajové a typicky se používá plnovýznamově. Určit, které z takových podstatných jmen je více gramatikalizované by bylo velmi sporné. Nicméně u podstatných jmen, která byla k analýze zvolena, tento problém pozorován nebyl.

4.Preskriptivní pravidla

Způsob, jakým jsou pomocná podstatná jména zapisována ve veřejných dokumentech, je upraven ve vládní vyhlášce č. 1 z 30. 11. 2010, která nahrazuje vyhlášku č. 1 z 1. 10. 1981. Zatímco vyhláška z roku 2010 je platná dosud, pro tuto práci je relevantní vyhláška z roku 1981,⁴ protože všechny vyšetřované texty byly publikovány nebo napsány do roku 2008.

Žádný z těchto dokumentů se pomocnými podstatnými jmény nezabývá přímo, oba ale obsahují odstavec, ve kterém je uvedeno, že u výrazů podobných těm uvedeným se v užitích podobným těm uvedeným použije k zápisu kana. Následuje seznam slov a vzorových slovních spojení, kde jsou slova ze seznamu podtržena a zapsána hiraganou. Tento seznam obsahující 26 slov je v obou vyhláškách stejný. Objevuje se v něm osm pomocných jmen⁵ (*koto, mono, toki, tokoro, hoka, jue, wake, tóri*), každé pouze v jednom užití. Jak ostatní možná užití těchto slov, tak všechna ostatní pomocná podstatná jména zmíněna nejsou a proto jejich preferovaný způsob zápisu preskriptivně definován není.

⁴ V korpusu jsou texty z let 1971 až 2008. V obou zmíněných vyhláškách jsou sice části relevantní pro zápis pomocných podstatných jmen totožné, šanci ovlivnit způsob zápisu svým vstupem v platnost měla pouze ta z roku 1981.

⁵ V jiných odstavcích se objevují i spojovací výrazy *ところが, ところで, ゆえに*, pomocné sloveso *ようだ* a partikule *ほど*. Jedná se o ukázky gramatikalizace pomocných podstatných jmen spojené se změnou slovního druhu.

5.Korpus

Za účelem zjištění ortografické variability pomocných podstatných jmen v japonštině byla provedena frekvenční analýza jejich výskytů v korpusu. Korpus je soubor digitalizovaného textu, který je uzpůsoben pro snadné prohledávání a mnohdy obsahuje i informace o textech v něm obsažených, například údaje o autorovi, vydavateli, díle, označení slovních druhů, apod.

Zvolený korpus k provedení analýzy je Vyvážený korpus současné psané japonštiny (現代日本語書き言葉均衡コーパス, BCCWJ), který provozuje Národní institut pro japonštinu (国立国語研究所, NINJAL). Rozsah tohoto korpusu je zhruba sto miliónů slov, které jsou získány sběrem úryvků z různých děl s cílem získat reprezentativní vzorek psaného japonského jazyka. Označení vyvážený znamená, že data byla sbírána z různých zdrojů, aby jejich reprezentativnost byla co nejvyšší. Reprezentativní korpusová data jsou taková, která mají stejné vlastnosti, jako má souhrn textů, ze kterých se data vybírala. (Brezina 2018) Sběrem vzorků z co nejširšího počtu různých druhů těchto zdrojů se docílí toho, že v korpusu jsou zastoupeny různé styly psaného jazyka. (Brezina 2018) Do korpusu BCCWJ byla data sbírána z písemností ze tří hlavních zdrojů, podle kterých lze korpus rozdělit na tři subkorpusy.

Jeden subkorpus obsahuje vzorky z písemností, které byly vydány a žádné další požadavky na ně kladeny nebyly. Druhý je složen z knih, které byly vybírány tak, aby se o nich dalo říci, že byly v oběhu, to znamená, že byly opravdu čteny. Tyto knihy byly vybírány z katalogu Tokijské městské knihovny (東京都立図書館). Třetí subkorpus obsahuje ostatní texty, které nelze zařadit ani do jedné z předcházejících kategorií.

Tyto tři subkorpusy lze dále rozdělit na celkem třináct subkorpusů. Knihovnický subkorpus žádné další dělení nemá, subkorpus vydaných písemností lze rozdělit na subkorpusy, jejichž texty byly získány z knih, časopisů a novin a subkorpus pro ostatní texty lze rozdělit na blíže neurčený subkorpus bestsellerů, dále subkorpus s texty získanými

z japonské verze webu Yahoo Answers, zákonů, záznamů vládních jednání, PR informačních bulletinů, učebnic, oficiálních vládních zpráv, veršů a blogů. (Center for Corpus Development, b.r.)

6. Metoda analýzy

Cílem práce je zjistit, jak vypadá ortografická variabilita pomocných podstatných jmen v japonštině a co ji ovlivňuje. Toto bylo provedeno v několika krocích. Nejprve proběhl výběr pomocných jmen k další analýze. Následovalo získání dat z korpusu, ověření jejich relevance a případná filtrace. Poté proběhlo vlastní zpracování pomocí tabulkového procesoru Microsoft Excel a on-line statistických nástrojů Lancaster Stats Tools online. (Brezina 2018) Nakonec byly získané výsledky zhodnoceny a převedeny do předkládané formy.

6.1 Výběr pomocných jmen

Výběr zkoumaných pomocných jmen je naprosto zásadní. Od správně provedeného výběru se odvíjí úspěch celé analýzy. Výběr by měl být reprezentativní, to znamená zahrnout v sobě podstatná jména napříč všemi druhy, a zvolená jména by měla být pro svůj druh typická. Pro analýzu byla zvolena dvě pomocná jména z každé kategorie, které definuje Novák (1989). Pomocná podstatná jména zmíněná v knize Gramatika japonštiny, (Novák 1989) byla rozřazena do kategorií a byla jim určena míra gramatikalizace. Poté bylo v rámci každé kategorie vybráno jedno pomocné jméno, které mělo ve své kategorii tuto míru malou a druhé, které mělo tuto míru ve své kategorii velkou. Pomocná jména, která mohou v závislosti na užití patřit do více kategorií, nebyla pro výběr uvažována. Celkem zkoumaný vzorek tedy obsahuje osm pomocných podstatných jmen, a sice: uči (うち), šita (した), koto (こと), baai (ばあい), hodo (ほど), fuu (ふう), cumori (つもり) a jotei (よてい).

6.2 Získání dat

Dalším krokem je získat potřebná data k vlastní analýze. Pro zamýšlenou frekvenční analýzu je vhodné, když je použitý vzorek dostatečně velký a reprezentativní. Kdyby bylo vstupních dat do analýzy příliš málo, byla by neprůkazná. Popsaný korpus BCCWJ je s těmito cíli sestaven. (Center for Corpus Development, b.r.) Proto je vhodným zdrojem dat.

Alternativní metody, například dotazování rodilých mluvčích, sice umožňují získat informace, které z korpusu vyčíst nelze, například o motivaci konkrétních rozhodnutí, byl by ale takto získaný statistický vzorek řádově menší než ten, který lze získat ze zavedeného korpusu. S větším počtem vyšetřovaných záznamů lze očekávat větší statistickou významnost.

6.3 Filtrace dat z korpusu

Pro každé pomocné jméno byly do analýzy zahrnuty všechny jeho výskyty v korpusu, které splňují následující podmínky.

Použitelné záznamy musí opravdu reprezentovat zkoumané slovo, nesmí se jednat o jiné slovo, které je s ním pouze homonymní. K zaručení této podmínky byly využity atributy v korpusu, které ke každému slovu přidružují údaje o jeho čtení a zápisu. Záznamy psané hiraganou nebo katakanou byly filtrovány podle znaku a záznamy psané znakem byly omezeny pouze na ty, které mají potřebné čtení.

Použitelné záznamy musí být v korpusu označeny jako obecné podstatné jméno. Tímto se odstraní jednak ty případy, kdy je vyhledané slovo použito jako jiný slovní druh, například zájmeno, také se tímto odstraní výskyty ve vlastních jménech.

Použitelné záznamy musí být rozvity přívlastkem. Pomocné podstatné jméno z definice přívlastek mít musí a také se touto podmínkou z analýzy vyloučí případná plnovýznamová užití zkoumaného slova. Tato podmínka je realizována pomocí omezení se na ty záznamy, kterým předchází hiraganou nebo katakanou napsaná jedna z těchto slabik: の, な, い, た, だ, る, う, く, ぐ, す, つ, ぬ, む a ぶ. Slovesa a přídavná jména, která rozvíjejí podstatné jméno, musí být ve slovníkovém tvaru. Jedná se tedy o seznam posledních slabik slovníkových tvarů přídavných jmen a sloves, jejich záporů a minulých časů, i záporných, kvůli na přídavným jménům je zahrnuto な a také je přidána partikule の.

Použitelné záznamy nesmí být začátkem složeného slova. Toto je garantováno webovým rozhraním, pomocí kterého jsou data získávána.

Toto rozhraní umožňuje vyhledávat pouze celá slova. Vyhledá-li se například *こと*, ve výsledcích se *ことがら* neobjeví.

6.4 Vlastní analýza

Cílem vlastní analýzy bylo popsání toho, jak se v korpusu objevují jednotlivá pomocná jména v různých způsobech zápisu. První náhled na data byl realizován jejich sdružením do několika tabulek, popsanych výše v kapitole 6, s cílem zjistit podíly jednotlivých způsobů zápisu nejen v celém korpusu, ale i jednotlivých subkorpusech. Tento krok je nutný k získání představy o tom, který způsob zápisu je dominantní a jestli lze pozorovat rozdíly ve způsobech zápisu mezi zdroji.

Takto lze získat základní přehled, ale získané údaje se musí statisticky ověřit. Pro ověření toho, jestli jsou rozdíly mezi subkorpusemi statisticky významné a jestli lze najít nějaký vztah mezi subkorpusemi a způsoby zápisu, byla použita logistická regresní analýza.

6.4.1 Logistická regresní analýza

Použitá metoda, logistická regresní analýza, je statistický postup, který v kontextu korpusové lingvistiky slouží k ověření toho, jestli je možné takzvanou lingvistickou, neboli vysvětlovanou, proměnnou vysvětlit nějakými vysvětlujícími proměnnými. (Brezina, 2018) V tomto případě je vysvětlovanou proměnnou způsob zápisu a vysvětlující proměnnou příslušnost k subkorpusu. Výstupem tohoto postupu je tabulka, kterou lze interpretovat a určit, které vlivy jsou statisticky významné a jak se podílejí na zjištěném výsledku.

Logistická regresní analýza má sadu předpokladů, které musí být splněny, aby ji bylo možné použít. Tyto předpoklady jsou tři, a sice nezávislost statistického souboru (dat z korpusu), žádná kovariance mezi vysvětlujícími proměnnými a lineární vztah mezi vysvětlujícími a vysvětlovanými proměnnými. (Brezina, 2018)

Všechny záznamy statistického souboru (dat z korpusu) musí na sobě být nezávislé. Toto je problematická podmínka, protože jednotlivá

slova na sobě nezávislá nejsou. Skutečné nezávislosti by bylo dosaženo, kdyby jazyk byl náhodným slepencem slov. (Brezina 2018) Protože ale existují pravidla a osobní preference, je tato podmínka vždy do jisté míry porušena. Budu však předpokládat, že tato závislost je příliš malá na to, aby získaná data zkreslila.

Další podmínky se týkají pouze číselných, takzvaných škálových proměnných. V této práci se s takovými proměnnými nepracuje a tak se ani kovariance, ani linearita ověřovat nemusí.

Protože je podíl zápisu pomocných podstatných jmen v katakaně malý, pro žádné vyšetřované pomocné jméno nepřekročil 1%, a protože vysvětlovaná proměnná, která může nabývat více než dvě hodnoty, celý proces komplikuje, je do následujících vyšetření (logistická regresní analýza a konstrukce grafu pro závislost na čase) katakana uvažována společně s hiraganou do jedné společné kategorie.

K realizaci tohoto postupu jsou využity on-line nástroje Lancaster Stats Tools online. (Brezina 2018)

6.5 Souhrn

Na závěr jsou poznatky z předchozích kroků shrnuty a pro lepší přehled soustředěny do jednoho odstavce.

7.Uči (うち、ウチ、内)

Prvním vyšetřeným pomocným podstatným jménem je *uči*. Jeho hlavním významem je *uvnitř*, ale objevuje se i v konstrukcích s významem posunutým na časové užití, které lze přeložit jako *dokud*. Protože upřesňuje časoprostorové zařazení svého přívlastku, jedná se v Novákové klasifikaci o pomocné podstatné jméno s předložkovým užitím. S mírou gramatikalizace jedenáct bylo zvoleno jako více gramatikalizovaný zástupce své kategorie.

7.1 První náhled

Po stažení dat a jejich filtraci zůstalo 33 199 záznamů, z čehož 30 650 (92,32%) je zapsáno hiraganou, 87 (0,26%) katakanou a zbylých 2 462 (7,42%) znakem. Podrobné informace jsou k dispozici v následujících tabulkách:

Subkorpus	ウチ	うち	内	Souhrn
Vydané texty	20,69%	32,33%	36,31%	32,59%
Vydané texty - noviny	0,00%	3,66%	1,68%	3,49%
Vydané texty - knihy	77,78%	86,78%	92,39%	87,23%
Vydané texty - časopisy	22,22%	9,56%	5,93%	9,28%
Knihy v knihovnách	4,60%	33,39%	34,48%	33,40%
Ostatní texty	74,71%	34,28%	29,20%	34,01%
Ostatní - blogy	76,92%	17,48%	35,33%	18,96%
Ostatní - bestselery	3,08%	12,84%	18,22%	13,12%
Ostatní - vládní jednání	0,00%	9,22%	1,11%	8,65%
Ostatní - informační bulletin	1,54%	5,41%	3,34%	5,25%
Ostatní - učebnice	0,00%	2,22%	1,39%	2,15%
Ostatní - zákoníky	0,00%	6,38%	0,70%	5,98%
Ostatní - vládní zprávy	0,00%	27,89%	3,89%	26,20%
Ostatní - Čiebukuro	18,46%	18,01%	35,33%	19,11%
Ostatní - verše	0,00%	0,56%	0,70%	0,57%

Tabulka 1: Distribuce uči v korpusu (NINJAL 2009)

V *Tabulce 1* jsou zaneseny výskyty slova *uči* v korpusu a zobrazeny jako procenta všech výskytů tohoto slova. Jednotlivé způsoby zápisu jsou rozděleny po sloupcích a ve sloupci *Součet* jsou počty výskytů *uči* v daném subkorpusu sečteny. Zvýrazněny jsou tři hlavní subkorporusy. Za zmínku stojí hodnoty sloupce *Souhrn* těchto tří záznamů. Výskyty *uči* jsou

v těchto třech subkorpusech rozděleny zhruba po třetinách (32,59%, 33,40% a 34,01%).

Při jemnějším dělení těchto subkorporusů získáme 13 různých subkorporusů nižší úrovně, které jsou zastoupeny nezvýrazněnými řádky. Procentuální údaje v těchto řádcích udávají poměrné zastoupení vůči nadřazenému subkorporusu. Například z tabulky lze zjistit, že v subkorporusu *vydaných textů se učí* psané katakanou objevilo v 20,69% všech výskytů, z čehož 77,78% výskytů bylo v *knihách* a 22,22% v *časopisech*. Subkorporus *knih v knihovnách* žádné další dělení nemá. Zde si zvláštní pozornost zaslouží fakt, že většina výskytů zápisu katakanou pochází ze subkorporusu *ostatních textů* (74,71%), konkrétně z *blogů* (76,92%) a *Čiebukura* (japonská verze Yahoo! Answers, 18,46%), což jsou internetové zdroje. Tyto dva subkorporusy do své skupiny také přispěly většinovým podílem zápisů znakem (v součtu 70,65%).

Tento druh tabulky je vhodný pro první náhled na data a utvoření si základního přehledu, pro podrobnější interpretaci ale už vhodný není, protože nezohledňuje velikosti jednotlivých subkorporusů. Mohlo by se zdát, že například *vydané knihy* mají oproti *časopisům* a *novinám* vysoký podíl zápisu *učí* hiraganou (86,78%), ale to je způsobeno tím, že v tomto subkorporusu je výskyt *učí* vysoký celkově (87,23%). Není překvapivé, že když se druhé dva subkorporusy dělí o zbývajících 13,22% výskytů *učí* v hiraganě, budou jejich hodnoty oproti *knihám* menší.

Tabulka 2 ukazuje procentuální zastoupení způsobů zápisu pro každý subkorporus. Zde je už každý řádek nezávislý na ostatních, nezvýrazněné řádky se interpretují stejně jako zvýrazněné. Řádek Celý korpus obsahuje průměrné podíly zápisů pro celý korpus. V této tabulce zaujmou velmi nízké podíly zápisu znakem v *přepisech vládních jednání, zákoníků* a *vládních zpráv*. Společnou vlastností těchto subkorporusů je to, že všechny obsahují texty oficiální a formální.

Další náhled na data je v *Tabulce 3*. V této tabulce jsou relativní četnosti různých zápisů slova *učí* na milion slov. Tyto relativní četnosti se získají vydělením počtu výskytů zkoumaného slova počtem všech slov v

subkorpusu a vynásobením vzniklé hodnoty milionem. (Brezina 2018)
 Toto umožňuje porovnávat jednotlivé subkorpusey přímo mezi sebou, protože je odstraněna závislost na velikosti korpusu. Je pozoruhodné, že se v subkorpusech *Ostatní - zákoníky* a *Ostatní - vládní zprávy* uči vyskytuje mnohem hustěji než v ostatních.⁶

Subkorpus	ウチ	うち	内
Vydané texty	0,17%	91,57%	8,26%
Vydané texty - noviny	0,00%	96,03%	3,97%
Vydané texty - knihy	0,15%	91,10%	8,75%
Vydané texty - časopisy	0,40%	94,32%	5,28%
Knihy v knihovnách	0,04%	92,31%	7,66%
Ostatní texty	0,58%	93,06%	6,37%
Ostatní - blogy	2,34%	85,80%	11,86%
Ostatní - bestselery	0,13%	91,03%	8,84%
Ostatní - vládní jednání	0,00%	99,18%	0,82%
Ostatní - informační bulletiny	0,17%	95,78%	4,05%
Ostatní - učebnice	0,00%	95,88%	4,12%
Ostatní - zákoníky	0,00%	99,26%	0,74%
Ostatní - vládní zprávy	0,00%	99,05%	0,95%
Ostatní - Čiebukuro	0,56%	87,67%	11,77%
Ostatní - verše	0,00%	92,19%	7,81%
Celý korpus	0,26%	92,32%	7,42%

Tabulka 2: Podíly způsobů zápisu v jednotlivých subkorpusech (NINJAL 2009)

⁶ Téměř všechny výskyty *uči* v těchto subkorpusech mají místní význam. Je možné, že jejich větší hustota je zapříčiněna povahou textů, které často obsahují seznamy a výčty.

Subkorpus	ウチ	うち	内	Součet
Vydané texty	0,53	289,33	26,11	315,96
Vydané texty - noviny	0,00	265,01	10,95	275,96
Vydané texty - knihy	0,49	302,21	29,03	331,73
Vydané texty - časopisy	0,90	214,03	11,98	226,91
Knihy v knihovnách	0,13	337,67	28,01	365,82
Ostatní	1,62	262,31	17,95	281,88
Ostatní - blogy	4,94	181,42	25,08	211,44
Ostatní - bestselery	0,54	360,92	35,05	396,50
Ostatní - vládní jednání	0,00	189,91	1,57	191,48
Ostatní - informační bulletiny	0,27	151,45	6,40	158,11
Ostatní - učebnice	0,00	251,91	10,81	262,72
Ostatní - zákoníky	0,00	620,90	4,63	625,53
Ostatní - vládní zprávy	0,00	600,50	5,74	606,24
Ostatní - Čiebukuro	1,17	184,85	24,82	210,84
Ostatní - verše	0,00	264,36	22,40	286,76
Celý korpus	0,83	292,99	23,53	317,35

Tabulka 3: Počty výskytů slova uči na milion slov (NINJAL 2009)

7.2 Vliv subkorpusu

Předcházející tabulky nabídly několik různých náhledů na výskyty pomocného podstatného jména *uči* v korpusu. Tyto náhledy od sebe vždy odlišovaly jednotlivé subkorpusy. Nabízí se tedy otázka, jestli pozorované rozdíly jsou statisticky významné a jestli má rozdělení těchto výskytů do subkorpusů na vyskytující se způsoby zápisu nějaký vliv, a pokud ano, jaký.

K vyšetření těchto vlivů byla zvolena logistická regresní analýza s využitím nástroje *Lancaster Tools Online* (Brezina 2018). Podrobný popis této analýzy však přesahuje rámec této práce a je uveden v *Příloze 1*. Zde jsou namísto toho uvedeny relevantní informace, které lze z této analýzy získat, ve zvláštní tabulce. Jedná se o *Tabulku 4*, která obsahuje pro každý subkorpus podíl zápisu pomocného podstatného jména *uči* kanou, zápisy hiraganou a katakanou jsou tedy pohromadě. Tento údaj je doplněn o tzv. *interval spolehlivosti* na hladině významnosti 95%, který říká, že na základě pozorovaných hodnot skutečný podíl zápisu kany s pravděpodobností 95% bude náležet do tohoto intervalu. Nese informaci o přesnosti, protože procentuální údaj získaný z velkého množství dat bude mít tento interval

užší než procentuální údaj získaný z malého množství dat. Je také užitečný při porovnávání dvou hodnot, pokud se jejich intervaly spolehlivosti nepřekrývají, můžeme s určitostí konstatovat, že jedna je větší než druhá. Tabulka také obsahuje údaje o statistické významnosti jednotlivých řádků, ty, které statisticky významné nejsou, jsou označeny červeně. Toto znamená, že takový řádek buď reprezentuje příliš malý vzorek dat, nebo že je příliš blízký referenčnímu řádku⁷, subkorpusu *Vydané texty – knihy* a není mezi nimi statisticky významný rozdíl.

Z této tabulky lze potvrdit závěry vypořádané z *Tabulky 2*. Dolní meze intervalů spolehlivosti trojice subkorpusů, *Ostatní - vládní jednání*, *Ostatní - zákoníky* a *Ostatní - vládní zprávy* jsou větší než horní mez intervalu spolehlivosti kteréhokoliv jiného subkorpusu, v těchto subkorpusech je tedy větší četnost zápisu *učí* kanou.

Subkorpus	Podíl kany	Interval spolehlivosti	
		dolní mez	horní mez
Vydané texty - noviny	96,03%	93,28%	97,86%
Vydané texty - knihy	91,25%	90,67%	91,81%
Vydané texty - časopisy	94,72%	92,69%	96,29%
Knihy v knihovnách	92,34%	91,05%	93,47%
Ostatní - blogy	88,14%	85,66%	90,27%
Ostatní - bestselery	91,16%	88,82%	93,10%
Ostatní - vládní jednání	99,18%	98,36%	99,65%
Ostatní - informační bulletiny	95,95%	93,72%	97,54%
Ostatní - učebnice	95,89%	92,38%	98,06%
Ostatní - zákoníky	99,26%	98,29%	99,75%
Ostatní - vládní zprávy	99,05%	98,55%	99,41%
Ostatní - Čiebukuro	88,23%	85,76%	90,36%
Ostatní - verše	92,19%	82,93%	97,33%

Tabulka 4: Podíly kany ve vzorku a intervaly spolehlivosti na hladině významnosti 95% (NINJAL 2009, Brezina 2018)

7.3 Souhrn

Na základě provedené analýzy bylo určeno několik faktů, které lze o ortografické variabilitě pomocného podstatného jména *učí* zkonstatovat.

⁷ Určení výchozích hodnot, vůči kterým se ostatní hodnoty porovnávají, je součástí postupu zvolené analýzy. Pro podrobnosti viz *Příloha 1*.

Uči je gramatikalizovaný zástupce pomocných podstatných jmen s předložkovým užitím. Dominantním způsobem zápisu je hiragana a naopak zápis katakanou je pouze okrajový. Podíl zápisu znakem má celokorpusový průměr je 7,42%. Doplněk do 100% připadá na zápis kanou.

Vliv subkorpusů je statisticky významný. Jinými slovy, mezi jednotlivými subkorpusy jsou statisticky významné rozdíly. Byly určeny následující subkorpusy, které jsou nějakým způsobem odlišné: *Ostatní - blogy* a *Ostatní - Čiebukuro*, které jsou internetovými zdroji a obsahují většinu zápisů katakanou z celého korpusu. Subkorpusy *Ostatní - vládní jednání*, *Ostatní - zákoníky* a *Ostatní - vládní zprávy* obsahují oficiální, případně formální texty, a *uči* je v nich psáno hiraganou v drtivé většině výskytů. V subkorpusech *Ostatní - zákoníky* a *Ostatní - vládní zprávy* se slovo *uči* vyskytuje častěji, než je v celém korpusu běžné. Jeho četnost na milion slov je dvounásobkem celokorpusového průměru. Vliv subkorpusů *Ostatní - bestselery* a *Ostatní - verše* není statisticky významný.

8.Šita (した、シタ、下)

Druhým vyšetřeným pomocným podstatným jménem je *šita*, které znamená *pod, dole, spodek*. Jeho znak má vícero čtení, pro analýzu byly vybrány pouze výskyty se čtením *šita*. Jeho význam je čistě prostorový, posun do časové roviny se nevyskytuje. Upřesňuje tedy prostorový vztah a lze ho proto pokládat za pomocné podstatné jméno s předložkovým užitím. Jeho míra gramatikalizace je nula a je zástupcem méně gramatikalizovaného pomocného jména ve své kategorii.

8.1 První náhled

Použitá data obsahují 9270 výskytů pomocného jména *šita*. Z těchto výskytů je 148 (1,60%) zapsáno hiraganou a 9 121 (98,40%) zapsáno znakem. V katakaně výskyt není ani jeden. Následují popisné tabulky 5, 6 a 7, které se interpretují se stejně jako tabulky 1, 2 a 3. Pro podrobnosti viz kapitola 7.1.

Subkorpus	シタ	した	下	Souhrn
Vydané texty	0,00%	31,76%	33,19%	33,16%
Vydané texty - noviny	0,00%	0,00%	1,42%	1,40%
Vydané texty - knihy	0,00%	87,23%	86,06%	86,08%
Vydané texty - časopisy	0,00%	12,77%	12,52%	12,52%
Knihy v knihovnách	0,00%	40,54%	37,47%	37,52%
Ostatní	0,00%	27,70%	29,34%	29,31%
Ostatní - blogy	0,00%	39,02%	22,94%	23,19%
Ostatní - bestselery	0,00%	4,88%	16,33%	16,16%
Ostatní - vládní jednání	0,00%	0,00%	4,67%	4,60%
Ostatní - informační bulletiny	0,00%	2,44%	1,98%	1,99%
Ostatní - učebnice	0,00%	0,00%	3,48%	3,42%
Ostatní - zákoníky	0,00%	0,00%	4,75%	4,67%
Ostatní - vládní zprávy	0,00%	0,00%	17,60%	17,34%
Ostatní - Čiebukuro	0,00%	26,83%	26,49%	26,50%
Ostatní - verše	0,00%	26,83%	1,76%	2,13%

Tabulka 5: Distribuce *šita* v korpusu, absolutní četnost (NINJAL 2009)

Z této tabulky nelze usoudit, že by v rozdělení slova *šita* do subkorpusů byly nějaké výrazné tendence. Lze si všimnout, že některé subkorpusy (*Vydané texty - noviny*, *Ostatní - vládní jednání*, *Ostatní - učebnice*, *Ostatní -*

zákoníky a *Ostatní - vládní zprávy*) mají výskyty slova *šita* pouze jedním způsobem zápisu. Jedná se o subkorpusey obsahující oficiální nebo publicistické texty. *Tabulka 6* toto ilustruje lépe.

Subkorpus	シタ	した	下
Vydané texty	0,00%	1,53%	98,47%
Vydané texty - noviny	0,00%	0,00%	100,00%
Vydané texty - knihy	0,00%	1,55%	98,45%
Vydané texty - časopisy	0,00%	1,56%	98,44%
Knihy v knihovnách	0,00%	1,73%	98,27%
Ostatní	0,00%	1,51%	98,49%
Ostatní - blogy	0,00%	2,54%	97,46%
Ostatní - bestselery	0,00%	0,46%	99,54%
Ostatní - vládní jednání	0,00%	0,00%	100,00%
Ostatní - informační bulletiny	0,00%	1,85%	98,15%
Ostatní - učebnice	0,00%	0,00%	100,00%
Ostatní - zákoníky	0,00%	0,00%	100,00%
Ostatní - vládní zprávy	0,00%	0,00%	100,00%
Ostatní - Čiebukuro	0,00%	1,53%	98,47%
Ostatní - verše	0,00%	18,97%	81,03%
Celý korpus	0,00%	1,60%	98,40%

Tabulka 6: Podíly způsobů zápisu slova *šita* v jednotlivých subkorpusech (NINJAL 2009)

V *Tabulce 6* zaujme subkorpus *Ostatní - verše* s vysokým podílem zápisu *šita* v hiraganě. U části korpusu obsahujícím básně by se nemuselo jednat o nic překvapivého, ale podle sloupce Součet v *Tabulce 5* se jedná o velmi malý subkorpus. Je proto nutné ověřit, jestli tento vysoký podíl je i statisticky významný.

Tabulka 7 je tabulkou relativních četností slova *šita* na milion slov. I v této tabulce je pozoruhodný subkorpus *Ostatní - verše*, jehož relativní četnost slova *šita* je oproti průměru dvounásobná. Je sice malý, ale toto podstatné jméno se v něm vyskytuje poměrně často.

Subkorpus	シタ	した	下	Součet
Vydané texty	0,00	1,37	88,39	89,77
Vydané texty - noviny	0,00	0,00	31,39	31,39
Vydané texty - knihy	0,00	1,44	91,56	93,00
Vydané texty - časopisy	0,00	1,36	85,66	87,01
Knihy v cirkulaci	0,00	1,98	112,78	114,76
Ostatní	0,00	1,02	66,80	67,82
Ostatní - blogy	0,00	1,58	60,64	62,22
Ostatní - bestselery	0,00	0,54	116,92	117,45
Ostatní - vládní jednání	0,00	0,00	24,50	24,50
Ostatní - informační bulletiny	0,00	0,27	14,13	14,40
Ostatní - učebnice	0,00	0,00	100,55	100,55
Ostatní - zákony	0,00	0,00	117,69	117,69
Ostatní - vládní zprávy	0,00	0,00	96,50	96,50
Ostatní - Čiebukuro	0,00	1,07	69,27	70,34
Ostatní - verše	0,00	49,29	210,59	259,88
Celý korpus	0,00	1,41	87,19	88,60

Tabulka 7: Počty výskytů slova šita na milion slov (NINJAL 2009)

8.2 Vliv subkorpusu

Vyhodnocení proběhlo obdobně jako v kapitole s pomocným jménem *učí*. Pro podrobné vysvětlení postupu a způsobu interpretace viz *Příloha 1*.

Výsledkem logistické regrese je *Tabulka 8*. Z osmi⁸ subkorpusů, které popisuje, jich je 6 nemá statisticky významný vliv. Statisticky významný je subkorpus *Vydané texty – knihy*, který byl zvolen jako základní hodnota. To, že ostatní subkorpusy, s výjimkou *veršů*, nemají statisticky významný vliv, znamená, že nelze potvrdit, že se od subkorpusu *vydaných knih* liší. U subkorpusu *Ostatní – verše* odlišnost potvrdit lze, ale protože se jedná o malý vzorek, je interval spolehlivosti velice široký.

⁸ Z hlediska logistické regresní analýzy jsou problematické ty subkorpusy, ve kterých se zkoumané pomocné jméno vyskytuje pouze v jednom způsobu zápisu. Pokud se takový subkorpus objeví, tak v lepším případě Lancaster Stats Tools online (Brezina 2018) pro něj vrátí statisticky nevýznamný řádek s nic neříkajícími hodnotami, v horším případě nástroj selže a model neurčí. V takovém případě je třeba takový subkorpus z regresní analýzy vyloučit a spustit nástroj znovu.

Subkorpus	Podíl kany	Interval spolehlivosti	
		dolní mez	horní mez
Vydané texty - knihy	1,57%	1,09%	2,06%
Vydané texty - časopisy	1,58%	0,42%	4,44%
Knihy v knihovnách	1,75%	0,82%	3,40%
Ostatní - blogy	2,58%	0,98%	5,76%
Ostatní - bestselery	0,46%	0,05%	1,95%
Ostatní – informační bulletiny	1,88%	0,07%	10,65%
Ostatní - Čiebukuro	1,55%	0,53%	3,76%
Ostatní - verše	19,22%	7,07%	38,55%

Tabulka 8: Podíly kany ve vzorku a intervaly spolehlivosti na hladině významnosti 95% (NINJAL 2009, Brezina 2018)

8.3 Shrnutí

Pomocné podstatné jméno *šita* je ve většině případů zapsáno znakem. Malá část zápisů je hiraganou a katakana se neobjevuje vůbec. Je zapisováno podobně napříč všemi zdroji s výjimkou veršů, ty jediné lze statisticky odlišit. Objevuje se v nich častěji zápis hiraganou.

9.Koto (こと、コト、事)

Třetím vyšetřeným pomocným podstatným jménem je *koto*. *Koto* je slovo velmi rozšířené a používá se ve velkém množství gramatikalizovaných konstrukcí. Ve svém základním významu ho lze přeložit jako *věc, záležitost*. Pro jeho zařazení do kategorie je určující jeho nominalizující funkce. V případě rozšíření slovesem vzniká konstrukce, která plní funkci od tohoto slovesa odvozeného podstatného jména. (Novák 1989, ss. 40 - 41) Nominalizační funkci mají pomocná jména zájmenná. (Novák 1989, s. 41) Míra gramatikalizace *koto* je vysoká, 79, což z něj činí významně gramatikalizovaného zástupce své kategorie.

9.1 První náhled

Celkový počet výskytů slova *koto* splňujících podmínky určené v kapitole 6.3 je celkem 715 229, z čehož 681 896 (95,34%) záznamů je zapsáno hiraganou, 636 (0,09%) katakanou a 32 696 (4,57%) znakem. Tabulky 9, 10 a 11 poskytují bližší informace o distribuci jednotlivých způsobů zápisu v korpusu a jeho subkorpusech. Pro podrobnosti viz. kapitola 7.1, tyto tabulky se vyhodnocují stejně jako tabulky 1, 2 a 3.

Subkorpuse	コト	こと	事	Souhrn
Vydané texty	24,37%	35,69%	9,96%	34,51%
Vydané texty - noviny	0,00%	2,13%	0,58%	2,11%
Vydané texty - knihy	82,58%	88,96%	95,39%	89,04%
Vydané texty - časopisy	17,42%	8,91%	4,02%	8,85%
Knihy v knihovnách	8,96%	32,48%	8,00%	31,34%
Ostatní texty	66,67%	31,82%	82,04%	34,15%
Ostatní - blogy	74,76%	18,18%	35,54%	20,18%
Ostatní - bestselery	1,18%	13,41%	1,61%	12,09%
Ostatní - vládní jednání	0,00%	26,87%	0,03%	23,87%
Ostatní - informační bulletiny	0,00%	4,76%	0,44%	4,27%
Ostatní - učebnice	0,00%	2,51%	0,02%	2,24%
Ostatní - zákoníky	0,00%	3,12%	0,00%	2,78%
Ostatní - vládní zprávy	0,00%	10,90%	0,04%	9,69%
Ostatní - Čiebukuro	24,06%	19,97%	62,29%	24,63%
Ostatní - verše	0,00%	0,28%	0,02%	0,25%

Tabulka 9: Distribuce slova *koto* v korpusu (NINJAL 2009)

Koto psané katakanou se v korpusu vyskytuje okrajově. Většina těchto nečetných výskytů je rozdělena mezi internetové zdroje *blogy* a *Čiebukuro*. Do těchto subkorpuseů patří i většina zápisů znakem.

Subkorpus	コト	こと	事
Vydané texty	0,06%	98,62%	1,32%
Vydané texty - noviny	0,00%	99,64%	0,36%
Vydané texty - knihy	0,06%	98,53%	1,41%
Vydané texty - časopisy	0,12%	99,28%	0,60%
Knihy v knihovnách	0,03%	98,81%	1,17%
Ostatní texty	0,17%	88,84%	10,98%
Ostatní - blogy	0,64%	80,02%	19,34%
Ostatní - bestselery	0,02%	98,52%	1,46%
Ostatní - vládní jednání	0,00%	99,98%	0,02%
Ostatní - informační bulletiny	0,00%	98,87%	1,13%
Ostatní - učebnice	0,00%	99,89%	0,11%
Ostatní - zákoníky	0,00%	100,00%	0,00%
Ostatní - vládní zprávy	0,00%	99,95%	0,05%
Ostatní - Čiebukuro	0,17%	72,05%	27,78%
Ostatní - verše	0,00%	99,01%	0,99%
Celý korpus	0,09%	95,34%	4,57%

Tabulka 10: Podíly způsobů zápisu slova *koto* v jednotlivých subkorpusech (NINJAL 2009)

Subkorpus	コト	こと	事	Součet
Vydané texty	4,53	7 107,46	95,08	7 207,07
Vydané texty - noviny	0,00	3 786,76	13,87	3 800,63
Vydané texty - knihy	4,50	7 610,23	109,17	7 723,90
Vydané texty - časopisy	6,10	4 902,62	29,61	4 938,33
Knihy v knihovnách	1,88	7 308,49	86,35	7 396,72
Ostatní texty	10,58	5 416,86	669,57	6 097,01
Ostatní - blogy	31,31	3 895,78	941,46	4 868,56
Ostatní - bestselery	1,34	7 785,33	115,58	7 902,25
Ostatní - vládní jednání	0,00	11 426,10	1,76	11 427,87
Ostatní - informační bulletiny	0,00	2 751,92	31,46	2 783,39
Ostatní - učebnice	0,00	5 896,60	6,49	5 903,09
Ostatní - zákoníky	0,00	6 284,04	0,00	6 284,04
Ostatní - vládní zprávy	0,00	4 847,68	2,25	4 849,93
Ostatní - Čiebukuro	9,97	4 233,90	1 632,36	5 876,22
Ostatní - verše	0,00	2 679,44	26,88	2 706,32
Celý korpus	6,08	6 518,31	312,54	6 836,93

Tabulka 11: Počty výskytů slova *koto* na milion slov (NINJAL 2009)

Z *Tabulky 10* lze soudit, že poměry různých způsobů zápisu pomocného jména *koto* si jsou napříč subkorporusy podobné s výjimkou již zmíněných subkorporusů *blogů* a *Čiebukura*. *Tabulka 11*, která obsahuje relativní četnosti slova *koto* na milion slov v jednotlivých subkorpusech, dále potvrzuje, že tyto internetové zdroje mají největší variabilitu ve způsobech zápisu, což znamená, že se v nich objevuje nejvíce zápisů, které nejsou zapsány hiraganou. Dále data v této tabulce potvrzují, že *koto* je velice frekventované slovo.

9.2 Vliv subkorporusu

Rozdíly v distribuci různých způsobů zápisu mezi subkorporusy byly vyšetřeny stejnou metodou, která je podrobně popsána v *Příloze 1*. Protože je slovo *koto* velice frekventované, bylo nutné data pro logistickou regresi redukovat, protože vyšetření 715 tisíc záznamů najednou je příliš výpočetně náročné. Rozdílem oproti odkazovanému popisu je tedy výběr 30 000 náhodných výskytů, které je možné logistickou regresní analýzou vyšetřit. Její výsledky jsou uvedeny v *Tabulce 12*.

Na modelu v *Tabulce 12* je zajímavé, že ač je jednak statistický významný a je z provedených analýz postaven nad největším množstvím dat, tak více než polovina řádků v něm obsažených statisticky významná není. Při tomto množství dat z toho lze konstatovat, že podíly jednotlivých způsobů zápisu jsou si mezi jednotlivými subkorporusy velmi podobné. Ovšem část subkorporusů, která obsahovala pouze jeden způsob zápisu, byla z analýzy z technických důvodů vyloučena. Podrobnosti viz poznámka pod čarou 8 v kapitole 8.2.

Subkorpus	Podíl Kany	Interval spolehlivosti	
		dolní mez	horní mez
Vydané texty – knihy	98,58%	98,33%	98,81%
Vydané texty – časopisy	99,24%	98,23%	99,73%
Knihy v knihovnách	98,80%	98,19%	99,22%
Ostatní – blogy	81,67%	75,38%	86,64%
Ostatní - bestselery	98,66%	97,46%	99,34%
Ostatní - vládní jednání	99,92%	99,69%	99,99%
Ostatní - informační bulletiny	98,38%	96,25%	99,42%
Ostatní - Čiebukuro	72,03%	63,85%	78,53%

Tabulka 12: Podíly kany ve vzorku slova koto a intervaly spolehlivosti na hladině významnosti 95% (NINJAL 2009, Brezina 2018)

9.3 Shrnutí

Pomocné podstatné jméno *koto* je velice čtné slovo, v relativních četnostech na milion slov převyšuje všechna ostatní pomocná podstatná jména. Je pro něj dominantní zápis hiraganou a zápis katanou je okrajový.

Podle poměru zápisu kanou lze rozdělit subkorpusy na tři skupiny. V subkorpusech obsahující formální a úřední texty je užitá téměř výhradně hiragana. V ostatních tiskovinách se zápis kanou používá v drtivé většině případů, už ho však nelze označit za výhradní. V neformálních a neregulovaných internetových zdrojích se objevují zápisy znakem i katanou. Hiragana je ale i v těchto zdrojích dominantní.

10. Baai (ばあい、バアイ、場合)

Čtvrtým zkoumaným pomocným podstatným jménem je *baai*, které znamená *případ*. Toto slovo běžně tvoří formální jádro přívlastku a lze ho užít k vytvoření vedlejší věty podmíněné, což je vlastnost pomocných jmen zájmených. (Novák 1989, s. 40) S mírou gramatikalizace pět se jedná o málo gramatikalizovaného zástupce své kategorie.

10.1 První náhled

Pomocné podstatné jméno *baai* je v korpusu obsaženo v 94 525 výskytech. Z toho je 515 (0,54%) zapsáno v hiraganě, 1 v katakaně a 94 009 (99,45%) znakem. Jedná se tedy o pomocné podstatné jméno, které má dominantní způsob zápisu znakem. Rozložení těchto výskytů v jednotlivých subkorpusech je popsáno tabulkami 13, 14 a 15. Tyto tabulky jsou srovnatelné s tabulkami 1, 2 a 3 v kapitole 7.1, které jsou podrobně popsány.

Subkorpus	バアイ	ばあい	場合	Souhrn
Vydané texty	0,00%	45,63%	38,32%	38,36%
Vydané texty - noviny	0,00%	0,00%	1,11%	1,11%
Vydané texty - knihy	0,00%	99,15%	92,44%	92,48%
Vydané texty - časopisy	0,00%	0,85%	6,45%	6,41%
Knihy v knihovnách	100,00%	47,96%	19,93%	20,08%
Ostatní texty	0,00%	6,41%	41,75%	41,56%
Ostatní - blogy	0,00%	9,09%	6,64%	6,64%
Ostatní - bestselery	0,00%	27,27%	3,96%	3,98%
Ostatní - vládní jednání	0,00%	0,00%	16,19%	16,18%
Ostatní - informační bulletiny	0,00%	0,00%	13,66%	13,65%
Ostatní - učebnice	0,00%	0,00%	2,15%	2,14%
Ostatní - zákoníky	0,00%	0,00%	13,56%	13,54%
Ostatní - vládní zprávy	0,00%	0,00%	8,49%	8,48%
Ostatní - Čiebukuro	0,00%	63,64%	35,36%	35,38%
Ostatní - verše	0,00%	0,00%	0,01%	0,01%

Tabulka 13: Distribuce slova *baai* v korpusu (NINJAL 2009)

Na základě dat v *Tabulce 13* lze konstatovat, že se zápisy *baai* v hiraganě objevují převážně v knihách. Na subkorpus *Vydané texty – knihy* připadá 99,15% z 45,63% záznamů, to znamená, že 45,24%, téměř polovina všech

výskytů v hiraganě patří do tohoto subkorpusu. Druhá velká část, 47,96% náleží do subkorpusu *Knih v knihovně*.

Subkorpus	バアイ	ばあい	場合
Vydané texty	0,00%	0,65%	99,35%
Vydané texty - noviny	0,00%	0,00%	100,00%
Vydané texty - knihy	0,00%	0,69%	99,31%
Vydané texty - časopisy	0,00%	0,09%	99,91%
Knihy v knihovnách	0,01%	1,30%	98,69%
Ostatní texty	0,00%	0,08%	99,92%
Ostatní - blogy	0,00%	0,12%	99,88%
Ostatní - bestselery	0,00%	0,58%	99,42%
Ostatní - vládní jednání	0,00%	0,00%	100,00%
Ostatní - informační bulletiny	0,00%	0,00%	100,00%
Ostatní - učebnice	0,00%	0,00%	100,00%
Ostatní - zákoníky	0,00%	0,00%	100,00%
Ostatní - vládní zprávy	0,00%	0,00%	100,00%
Ostatní - Čiebukuro	0,00%	0,15%	99,85%
Ostatní - verše	0,00%	0,00%	100,00%
Celý korpus	0,00%	0,54%	99,45%

Tabulka 14: Podíly způsobů zápisu slova *baai* v jednotlivých subkorpusech (NINJAL 2009)

Tabulka 14 dokládá, že ač se výskyty hiraganou koncentrují do dvou subkorpusů zmíněných výše, stále se i v těchto subkorpusech jedná o okrajové užití, kde je nejvyšší zastoupení pouhých 1,30% v subkorpusu *Knihy v knihovně*.

Tabulka 15 obsahuje relativní četnosti výskytů pomocného podstatného jména *baai*. Subkorpus *Ostatní - zákoníky* s četností téměř 5000 výskytů *baai* na milion slov vysoce převyšuje všechny ostatní subkorpuse. Tato vysoká četnost zřejmě souvisí s významem *baai* (případ) a povahou právního textu. Dále stojí za zmínku dvojice internetových zdrojů *blogy* a *Čiebukuro*. Zatímco v případě jiných pomocných jmen často vypadají podobně, zde je pomocné jméno *baai* v *Čiebukuru* obsaženo v relativní četnosti mnohem častěji, než v subkorpusu *blogů*.

Subkorpus	バアイ	ばあい	場合	Součet
Vydané texty	0,00	6,86	1 051,95	1 058,82
Vydané texty - noviny	0,00	0,00	292,75	292,75
Vydané texty - knihy	0,00	8,19	1 170,42	1 178,61
Vydané texty - časopisy	0,00	0,45	525,25	525,70
Knihy v knihovnách	0,03	8,15	618,06	626,25
Ostatní texty	0,00	0,82	979,86	980,68
Ostatní - blogy	0,00	0,30	257,26	257,56
Ostatní - bestselery	0,00	2,41	415,77	418,18
Ostatní - vládní jednání	0,00	0,00	1 245,48	1 245,48
Ostatní - informační bulletiny	0,00	0,00	1 429,42	1 429,42
Ostatní - učebnice	0,00	0,00	910,33	910,33
Ostatní - zákoníky	0,00	0,00	4 931,04	4 931,04
Ostatní - vládní zprávy	0,00	0,00	682,46	682,46
Ostatní - Čiebukuro	0,00	2,05	1 355,97	1 358,02
Ostatní - verše	0,00	0,00	22,40	22,40
Celkový součet	0,01	4,92	898,64	903,57

Tabulka 15: Počty výskytů slova *baai* na milion slov (NINJAL 2009)

10.2 Vliv subkorpusu

I pro toto pomocné jméno byla provedena logistická regresní analýza, která je popsána v Příloze 1. Jejím výstupem je Tabulka 16. Protože se zastoupení kany v jednotlivých subkorpusech výrazně neliší, je v této tabulce mnoho řádků, které nejsou statisticky významné. Znamená to, že podíly výskytů pomocného jména *baai* zapsaných hiraganou v těchto subkorpusech zhruba odpovídají podílu hiraganových zápisů v subkorpusu *Vydané texty – knihy*, který byl zvolen za výchozí pro porovnávání. Subkorpus *Ostatní – verše* obsahuje jen málo výskytů slova *baai* a tyto jsou navíc zapsány pouze znakem. Výsledkem je řádek, který na interpretaci nemá statisticky významný vliv. Data v této tabulce potvrzují, že rozdíly v podílech zápisu kanou popisované v kapitole 10.1 jsou statisticky významné.

Subkorpus	Podíl kany	Interval spolehlivosti	
		dolní mez	horní mez
Vydané texty - noviny	0,00%	0,00%	0,00%
Vydané texty - knihy	0,70%	0,60%	0,81%
Vydané texty - časopisy	0,09%	0,01%	0,31%
Knihy v knihovnách	1,31%	0,94%	1,78%
Ostatní - blogy	0,12%	0,02%	0,34%
Ostatní - bestselery	0,58%	0,24%	1,20%
Ostatní - vládní jednání	0,00%	0,00%	0,00%
Ostatní - informační bulletiny	0,00%	0,00%	0,00%
Ostatní - učebnice	0,00%	0,00%	0,00%
Ostatní - zákoníky	0,00%	0,00%	0,00%
Ostatní - vládní zprávy	0,00%	0,00%	0,00%
Ostatní - Čiebukuro	0,15%	0,08%	0,26%
Ostatní - verše	0,00%	0,00%	100,00%

Tabulka 16: Podíly kany ve vzorku slova *baai* a intervaly spolehlivosti na hladině významnosti 95% (NINJAL 2009, Brezina 2018)

10.3 Shrnutí

Pomocné podstatné jméno *baai* má dominantní způsob zápisu znakem. Katakanaou je v korpusu pouze jeden zápis. Zápis hiraganou je okrajový, a když se už objeví, tak je to v nejčastěji v knize. Nejvyšší podíly zápisu hiraganou mají totiž subkorpora *Vydané texty – knihy* a *Knihy v knihovnách*. Nej hustěji, to znamená s největší relativní četností, se toto slovo objevuje v subkorpusu *Ostatní – zákoníky*. Rozdíly mezi subkorpora, které obsahují více než jeden druh zápisu, jsou statisticky významné.

11.Hodo (ほど、ホド、程)

Páté zkoumané pomocné jméno je *hodo*, které ve svém základním významu znamená *míra*. Přísllovečná pomocná jména určují čas, míru a způsob (Novák 1989, s. 41), proto do této kategorie toto podstatné jméno bezesporu lze. Míru gramatikalizace má šestnáct a ve své skupině patří k těm více gramatikalizovaným jménům.

11.1 První náhled

Pomocné podstatné jméno *hodo* má v korpusu 574 výskytů. Z toho jich je 455 (79,27%) napsáno hiraganou, 7 (1,22%) katakanou a 112 (19,51%) znakem. Jako dominantní se ukazuje zápis hiraganou. Podrobnější informace o těchto výskytech jsou v tabulkách 17, 18 a 19. Pro podrobný popis struktury těchto tabulek viz popis tabulek 1, 2 a 3 v kapitole 7.1.

Subkorpus	ホド	ほど	程	Souhrn
Vydané texty	42,86%	33,41%	12,50%	29,44%
Vydané texty - noviny	0,00%	36,18%	0,00%	32,54%
Vydané texty - knihy	66,67%	48,03%	85,71%	51,48%
Vydané texty - časopisy	33,33%	15,79%	14,29%	15,98%
Knihy v knihovnách	0,00%	19,56%	20,54%	19,51%
Ostatní texty	57,14%	47,03%	66,96%	51,05%
Ostatní - blogy	100,00%	21,96%	41,33%	27,99%
Ostatní - bestselery	0,00%	10,28%	0,00%	7,51%
Ostatní - vládní jednání	0,00%	36,45%	0,00%	26,62%
Ostatní - informační bulletiny	0,00%	11,68%	0,00%	8,53%
Ostatní - učebnice	0,00%	0,93%	0,00%	0,68%
Ostatní - zákoníky	0,00%	0,00%	0,00%	0,00%
Ostatní - vládní zprávy	0,00%	0,47%	0,00%	0,34%
Ostatní - Čiebukuro	0,00%	14,49%	57,33%	25,26%
Ostatní - verše	0,00%	3,74%	1,33%	3,07%

Tabulka 17: Distribuce slova *hodo* v korpusu (NINJAL 2009)

Tabulka 17 obsahuje údaje o rozmístění různých způsobů zápisu slova *hodo* v jednotlivých subkorpusech. Je z ní patrné, že se toto slovo v subkorpusu *zákoníků* vůbec nevyskytuje. Dále nadpoloviční většina všech výskytů

katakanou je v subkorpusu *blogů* a většina zápisů znakem je rozdělena mezi *blogy* a *Čiebukuro*. Tabulka 18 obsahuje jiný pohled na tatáž data.

Subkorpus	ホド	ほど	程
Vydané texty	1,78%	89,94%	8,28%
Vydané texty - noviny	0,00%	100,00%	0,00%
Vydané texty - knihy	2,30%	83,91%	13,79%
Vydané texty - časopisy	3,70%	88,89%	7,41%
Knihy v knihovnách	0,00%	79,46%	20,54%
Ostatní texty	1,37%	73,04%	25,60%
Ostatní - blogy	4,88%	57,32%	37,80%
Ostatní - bestselery	0,00%	100,00%	0,00%
Ostatní - vládní jednání	0,00%	100,00%	0,00%
Ostatní - informační bulletiny	0,00%	100,00%	0,00%
Ostatní - učebnice	0,00%	100,00%	0,00%
Ostatní - zákoníky	0,00%	0,00%	0,00%
Ostatní - vládní zprávy	0,00%	100,00%	0,00%
Ostatní - Čiebukuro	0,00%	41,89%	58,11%
Ostatní - verše	0,00%	88,89%	11,11%
Celý korpus	1,22%	79,27%	19,51%

Tabulka 18: Podíly způsobů zápisu slova *hodo* v jednotlivých subkorpusech (NINJAL 2009)

Z *Tabulky 18* lze vyčíst, že rozdělení podílů různých způsobů zápisu je různorodé. Existuje skupina subkorpusů (*Vydané texty – noviny*, *Ostatní – bestselery*, *Ostatní – vládní jednání*, *Ostatní – informační bulletiny*, *Ostatní – učebnice* a *Ostatní – vládní zprávy*), ve kterých je *hodo* zapsáno výhradně hiraganou. Naproti tomu v subkorpusu *Ostatní – Čiebukuro* je naopak nadpoloviční podíl zápisu znakem. V ostatních subkorpusech (s výjimkou *Ostatní – zákoníky*), je hiragany nadpoloviční většina a alternuje především se znaky.

V následující *Tabulce 19* jsou relativní četnosti jednotlivých způsobů zápisu slova *hodo* v přepočtu na milion slov. Je v ní vidět, že nejvyšších relativních četností dosahují subkorpuse *Vydané texty – noviny* a *Ostatní – básně*.

Subkorpus	ホド	ほど	程	Celkem
Vydané texty	0,09	4,44	0,41	4,94
Vydané texty - noviny	0	40,15	0	40,15
Vydané texty - knihy	0,07	2,57	0,42	3,06
Vydané texty - časopisy	0,23	5,42	0,45	6,10
Knihy v knihovnách	0	2,94	0,76	3,70
Ostatní texty	0,10	5,34	1,87	7,31
Ostatní - blogy	0,40	4,64	3,06	8,10
Ostatní - bestselery	0	5,89	0	5,89
Ostatní - vládní jednání	0	15,29	0	15,29
Ostatní - informační bulletiny	0	6,67	0	6,67
Ostatní - učebnice	0	2,16	0	2,16
Ostatní - zákoníky	0	0	0	0
Ostatní - vládní zprávy	0	0,20	0	0,20
Ostatní - Čiebukuro	0	3,03	4,20	7,23
Ostatní - verše	0	35,85	4,48	40,33
Celý korpus	0,07	4,35	1,07	5,49

Tabulka 19: Počty výskytů slova *hodo* na milion slov (NINJAL 2009)

11.2 Vliv subkorpusu

Za účelem zjištění, zda jsou rozdíly mezi jednotlivými subkorpusy statisticky významné, byla použita logistická regresní analýza. Více podrobností je v *Příloze 1*. Výstup Lancaster Stat Tools online (Brezina 2018) je zpracován a zanesen v *Tabulce 20*.

Ty řádky, které nemají statisticky významný vliv, jsou označeny červeně a významnou část jich tvoří ty subkorpusy, které obsahují pouze jeden způsob zápisu. U statisticky významných řádků je pozoruhodné, že se jejich intervaly spolehlivosti překrývají, protože jsou velmi široké. To znamená, že konstatování o podílech zápisu kanou nejsou tolik průkazná, jako konstatování založená na řádcích, jejichž intervaly spolehlivosti jsou disjunktní, což znamená, že se nepřekrývají. Zde je důvodem je příliš malý vzorek, *hodo* má v porovnání s ostatními zkoumanými pomocnými podstatnými jmény vyskytuje v korpusu málo výskytů.

Subkorpus	Podíl kany	Interval spolehlivost	
		dolní mez	horní mez
Vydané texty - noviny	100%	100%	100%
Vydané texty - knihy	86,21%	77,95%	92,36%
Vydané texty - časopisy	92,59%	63,87%	99,39%
Knihy v knihovnách	79,46%	49,84%	94,05%
Ostatní - blogy	62,17%	29,79%	86,89%
Ostatní - bestselery	100%	0,00%	100%
Ostatní - vládní jednání	100%	0,00%	100%
Ostatní - informační bulletiny	100%	0,00%	100%
Ostatní - učebnice	100%	0,00%	100%
Ostatní - vládní zprávy	100%	0,00%	100%
Ostatní - Čiebukuro	41,82%	15,53%	74,46%
Ostatní - verše	88,89%	42,14%	99,67%

Tabulka 20: Podíly kany ve vzorku slova *hodo* a intervaly spolehlivosti na hladině významnosti 95% (NINJAL 2009, Brezina 2018)

11.3 Shrnutí

Pomocné podstatné jméno *hodo* se v korpusu vyskytuje vzhledem k ostatním pomocným jménům relativně málo, což má vliv na přesnost statistické analýzy. Celkový vzorek je 574 výskytů. Nejčastěji je zapisováno hiraganou. Nejvíce se v relativní četnosti objevuje v novinách a verších. Zápisy znakem se nejvíce objevují na internetu.

12.Fú (ふう、フウ、風)

Šestým zkoumaným pomocným jménem je *fú*. Nejzákladnějším významem tohoto slova je *vítr*, což je plnovýznamové podstatné jméno. Má ale i další významy, z nichž pro tuto práci jsou relevantní *zvyk* a *způsob*. Oba slouží k rozvíjení převážně sloves, jedná se tedy o pomocné jméno příslovečné. S mírou gramatikalizace tři je toto pomocné jméno ve své skupině málo gramatikalizované.

12.1 První náhled

Toto pomocné podstatné jméno se v korpusu objevuje 19 136 krát. Z toho je 17 202 (89,89%) záznamů zapsáno hiraganou a 1 934 (10,11%) znakem. Zápis katakanou se nevykytuje a obvyklým způsobem zápisu tohoto jména je hiragana. Bližší informace o zápisech slova *fú* v subkorpusech jsou v tabulkách 21, 22 a 23. Tyto tabulky odpovídají tabulkám 1, 2 a 3 a jsou podrobně popsány v kapitole 7.1.

Subkorpus	フウ	ふう	風	Souhrn
Vydané texty	0,00%	10,71%	21,61%	11,82%
Vydané texty - noviny	0,00%	0,87%	1,67%	1,02%
Vydané texty - knihy	0,00%	86,98%	86,36%	86,86%
Vydané texty - časopisy	0,00%	12,15%	11,96%	12,12%
Knihy v knihovnách	0,00%	14,31%	27,66%	15,66%
Ostatní texty	0,00%	74,98%	50,72%	72,53%
Ostatní - blogy	0,00%	0,95%	32,93%	3,21%
Ostatní - bestselery	0,00%	3,30%	9,68%	3,75%
Ostatní - vládní jednání	0,00%	93,92%	0,10%	87,29%
Ostatní - informační bulletiny	0,00%	0,06%	0,61%	0,10%
Ostatní - učebnice	0,00%	0,16%	0,00%	0,15%
Ostatní - zákoníky	0,00%	0,00%	0,00%	0,00%
Ostatní - vládní zprávy	0,00%	0,00%	0,10%	0,01%
Ostatní - Čiebukuro	0,00%	1,59%	49,85%	5,00%
Ostatní - verše	0,00%	0,02%	6,73%	0,49%

Tabulka 21: Distribuce slova *fú* v korpusu (NINJAL 2009)

Podle *Tabulky 21* je drtivá většina výskytů slova *fú* v zápisech *vládních jednání*. Zápisy znakem se nejvíce vyskytují v internetových zdrojích, (*Ostatní - blogy* a *Ostatní - Čiebukuro*) a knihách (*Vydané texty - knihy* a *knihy*

v knihovnách). V subkorpusu *Ostatní – zákoníky* se toto podstatné jméno nevyskytuje.

Subkorpus	フウ	ふう	風
Vydané texty	0,00%	81,51%	18,49%
Vydané texty - noviny	0,00%	69,57%	30,43%
Vydané texty - knihy	0,00%	81,62%	18,38%
Vydané texty - časopisy	0,00%	81,75%	18,25%
Knihy v knihovnách	0,00%	82,14%	17,86%
Ostatní texty	0,00%	92,93%	7,07%
Ostatní - blogy	0,00%	27,42%	72,58%
Ostatní - bestselery	0,00%	81,77%	18,23%
Ostatní - vládní jednání	0,00%	99,99%	0,01%
Ostatní - informační bulletiny	0,00%	57,14%	42,86%
Ostatní - učebnice	0,00%	100,00%	0,00%
Ostatní - zákoníky	0,00%	0,00%	0,00%
Ostatní - vládní zprávy	0,00%	0,00%	100,00%
Ostatní - Čiebukuro	0,00%	29,54%	70,46%
Ostatní - verše	0,00%	2,94%	97,06%
Celý korpus	0,00%	89,89%	10,11%

Tabulka 22: Podíly způsobů zápisu slova *fú* v jednotlivých subkorpusech (NINJAL 2009)

Tabulka 22 ukazuje různé poměry způsobů zápisu uvnitř subkorpuseů. Existují subkorpuse, ve kterých se vyskytuje pouze jeden způsob zápisu. Těmi jsou *učebnice* a *vládní jednání*. V těch je jediný záznam znakem, který statisticky tento subkorpus od učebnic neodliší. Jeden jedinný záznam znakem je i ve *vládních zprávách* (viz nízká hodnota podílu zápisu znakem v Tabulce 22.) V internetových zdrojích (*Ostatní – blogy* a *Ostatní – Čiebukuro*) je zápis znakem naopak častější než zápis hiraganou.

Tabulka 23 obsahuje relativní četnosti slova *fú* na milion slov. Je z ní patrné, že v prepisech *vládních jednání* zcela zásadně toto slovo převyšuje obdobné četnosti *fú* ve všech ostatních subkorpusech.⁹

⁹ Většina výskytů je v užití podobném *to iu fú ni*, které je ve vládních jednáních zřejmě oblíbené.

Subkorpus	フウ	ふう	風	Součet
Vydané texty	0,00	53,82	12,21	66,02
Vydané texty - noviny	0,00	11,68	5,11	16,79
Vydané texty - knihy	0,00	56,34	12,69	69,03
Vydané texty - časopisy	0,00	50,63	11,30	61,93
Knihy v knihovnách	0,00	81,20	17,65	98,85
Ostatní texty	0,00	321,97	24,49	346,46
Ostatní - blogy	0,00	12,05	31,90	43,95
Ostatní - bestselery	0,00	113,97	25,42	139,39
Ostatní - vládní jednání	0,00	2374,16	0,20	2374,35
Ostatní - informační bulletiny	0,00	2,13	1,60	3,73
Ostatní - učebnice	0,00	22,70	0,00	22,70
Ostatní - zákoníky	0,00	0,00	0,00	0,00
Ostatní - vládní zprávy	0,00	0,00	0,20	0,20
Ostatní - Čiebukuro	0,00	20,03	47,77	67,80
Ostatní - verše	0,00	8,96	295,72	304,69
Celý korpus	0,00	164,44	18,49	182,92

Tabulka 23: Počty výskytů slova *fú* na milion slov (NINJAL 2009)

12.2 Vliv subkorpusu

Poměry zápisu pomocného jména *fú* v jednotlivých subkorpusech byly zanalyzovány logistickou regresní analýzou, jejíž podrobný popis je uveden v Příloze 1. Výsledky obsahuje Tabulka 24. Ta ukazuje že, mezi subkorpusem *Vydané texty – knihy* a ostatními tiskovinami není statisticky významný rozdíl. V ostatních subkorpusech je zápis znakem o něco častější s výjimkou subkorpusu *Ostatní – verše*, kde je zápis znakem dominantní.

Korpus	Podíl kany	Interval spolehlivosti	
		dolní mez	horní mez
Vydané texty - noviny	69,57%	46,38%	87,05%
Vydané texty - knihy	81,62%	79,87%	83,29%
Vydané texty - časopisy	81,75%	74,41%	87,56%
Knihy v knihovnách	82,14%	77,99%	85,68%
Ostatní - blogy	27,40%	21,00%	35,00%
Ostatní - bestselery	81,77%	75,79%	86,65%
Ostatní - informační bulletiny	57,12%	29,21%	82,05%
Ostatní - Čiebukuro	29,45%	23,40%	36,44%
Ostatní - verše	3,01%	0,40%	9,88%

Tabulka 24: Podíly kany ve vzorku slova *fú* a intervaly spolehlivosti na hladině významnosti 95% (NINJAL 2009, Brezina 2018)

12.3 Shrnutí

Pomocné podstatné jméno *fú* má ve svých zápisech největší podíl hiragany. Katakana se nevyskytuje vůbec a zápisy znakem jsou méně obvyklé. Nicméně v rámci subkorpusů se podíly mezi jednotlivými způsoby zápisu velmi liší. V některých je výskyt výhradně hiraganou, v jiných je většina zápisů znakem. V oficiálních textech se toto pomocné jméno nevyskytuje téměř vůbec. Popsané rozdíly mezi subkorpusy jsou statisticky významné.

13. Cumori (つもり、ツモリ、積り、積もり)

Dalším vyšetřeným pomocným podstatným jménem je *cumori*, které znamená *záměr*. Pojí se se sponou a i význam vyjadřuje vůli mluvčího, jedná se tedy o pomocné jméno modální. Jeho míra gramatikalizace je devět a proto bylo zvoleno jako gramatikalizovaný zástupce modálních jmen.

13.1 První náhled

Cumori se od ostatních pomocných jmen v této studii odlišuje tím, že se u něj objevují dvě varianty zápisu znakem. V korpusu se toto slovo vyskytuje v 11 009 záznamech, z nichž je 10 902 (99,03%) hiraganou a 107 (0,97%) znakem. Varianta 積もり se objevuje 29 krát a 積り 78 krát. Varianta zápisu katakanou se v korpusu nenachází.

Následující tři tabulky 25, 26 a 27 tyto výskyty v korpusu dále upřesňují. Odpovídají tabulkám 1 2 a 3, pro podrobný popis viz kapitola 7.1.

Subkorpus	ツモリ	つもり	積もり	積り	Souhrn
Vydané texty	0,00%	30,46%	41,38%	38,46%	30,55%
Vydané texty - noviny	0,00%	2,71%	0,00%	0,00%	2,68%
Vydané texty – knihy	0,00%	87,11%	100,00%	100,00%	87,27%
Vydané texty - časopisy	0,00%	10,18%	0,00%	0,00%	10,05%
Knihy v knihovnách	0,00%	37,94%	20,69%	48,72%	37,97%
Ostatní texty	0,00%	31,60%	37,93%	12,82%	31,48%
Ostatní - blogy	0,00%	27,63%	72,73%	30,00%	27,78%
Ostatní - bestselery	0,00%	18,78%	18,18%	20,00%	18,78%
Ostatní - vládní jednání	0,00%	16,08%	0,00%	0,00%	15,98%
Ostatní - informační bulletiny	0,00%	0,67%	0,00%	20,00%	0,72%
Ostatní - učebnice	0,00%	0,70%	0,00%	0,00%	0,69%
Ostatní - zákoníky	0,00%	0,00%	0,00%	0,00%	0,00%
Ostatní - vládní zprávy	0,00%	0,15%	0,00%	0,00%	0,14%
Ostatní - Čiebukuro	0,00%	35,76%	9,09%	30,00%	35,66%
Ostatní - verše	0,00%	0,23%	0,00%	0,00%	0,23%

Tabulka 25: Distribuce slova *cumori* v korpusu (NINJAL 2009)

První z popisných tabulek, *Tabulka 25*, ukazuje jak jsou jednotlivé způsoby zápisu rozděleny mezi subkorpusy. Lze si povšimnout, že ač jsou zápisy znakem zcela okrajové, většina se z nich objevuje v knihách. Varianta 積も

り je častější v subkorpusu *vydaných knih* a varianta 積り zase převládá v subkorpusu *knih z knihoven*. V oficiálních textech (*zákoníky, vládní zprávy*) se toto slovo nevyskytuje buď vůbec, nebo velmi okrajově.

Subkorpus	ツモリ	つもり	積もり	積り
Vydané texty	0,00%	98,75%	0,36%	0,89%
Vydané texty - noviny	0,00%	100,00%	0,00%	0,00%
Vydané texty - knihy	0,00%	98,57%	0,41%	1,02%
Vydané texty - časopisy	0,00%	100,00%	0,00%	0,00%
Knihy v knihovnách	0,00%	98,95%	0,14%	0,91%
Ostatní texty	0,00%	99,39%	0,32%	0,29%
Ostatní - blogy	0,00%	98,86%	0,83%	0,31%
Ostatní - bestselery	0,00%	99,39%	0,31%	0,31%
Ostatní - vládní jednání	0,00%	100,00%	0,00%	0,00%
Ostatní - informační bulletiny	0,00%	92,00%	0,00%	8,00%
Ostatní - učebnice	0,00%	100,00%	0,00%	0,00%
Ostatní - zákoníky	0,00%	0,00%	0,00%	0,00%
Ostatní - vládní zprávy	0,00%	100,00%	0,00%	0,00%
Ostatní - Čiebukuro	0,00%	99,68%	0,08%	0,24%
Ostatní - verše	0,00%	100,00%	0,00%	0,00%
Celý korpus	0,00%	99,03%	0,26%	0,71%

Tabulka 26: Podíly způsobů zápisu slova *cumori* v jednotlivých subkorpusech (NINJAL 2009)

Subkorpus	ツモリ	つもり	積もり	積り	Součet
Vydané texty	0,00	96,98	0,35	0,88	98,20
Vydané texty - noviny	0,00	65,70	0,00	0,00	65,70
Vydané texty - knihy	0,00	101,69	0,42	1,05	103,16
Vydané texty - časopisy	0,00	76,39	0,00	0,00	76,39
Knihy v knihovnách	0,00	136,47	0,20	1,25	137,92
Ostatní texty	0,00	86,00	0,27	0,25	86,52
Ostatní - blogy	0,00	94,02	0,79	0,30	95,10
Ostatní - bestselery	0,00	173,10	0,54	0,54	174,17
Ostatní - vládní jednání	0,00	108,58	0,00	0,00	108,58
Ostatní - informační bulletiny	0,00	6,13	0,00	0,53	6,67
Ostatní - učebnice	0,00	25,95	0,00	0,00	25,95
Ostatní - zákoníky	0,00	0,00	0,00	0,00	0,00
Ostatní - vládní zprávy	0,00	1,02	0,00	0,00	1,02
Ostatní - Čiebukuro	0,00	120,37	0,10	0,29	120,76
Ostatní - verše	0,00	35,85	0,00	0,00	35,85
Celý korpus	0,00	104,21	0,28	0,75	105,24

Tabulka 27: Počty výskytů slova *cumori* na milion slov (NINJAL 2009)

To, že je zápis znakem okrajový lze očekávat už z prosté převahy zápisu hiraganou, kterých je 99%. *Tabulky 26 a 27* toto potvrzují, nelze v nich najít žádný zdroj, který by výrazně v užití zápisu znakem vybočoval.

13.2 Vliv subkorpusu

Stále ale mohou existovat statisticky významné rozdíly, i když velmi malé. K jejich objevení byla použita logistická regresní analýza. V *Příloze 1* je podrobný popis této metody i možností interpretace jejích výsledků. Tato analýza umožňuje rozlišit pouze dvě hodnoty u zkoumané proměnné, proto byly sdruženy dohromady zápisy nejen hiragany a katakany, ale i obě varianty znakového zápisu. Výsledky jsou zaneseny do *Tabulky 28*.

Z nich je patrné, že se statisticky významné rozdíly najít podařilo. Není jich mnoho, většina řádků není statisticky významná, a protože jsou rozdíly mezi subkorpusy malé, a často se vyskytuje pouze jeden způsob zápisu v celém pro celý subkorpus, v řádcích, které nejsou statisticky významné, se často objevují nesmyslé hodnoty. Je možné od sebe statisticky významně odlišit pouze tři subkorpusy, a sice: *Vydané texty - knihy, Ostatní - informační bulletiny a Ostatní - Čiebukuro*. Je důležité mít ale na zřeteli, že nebyly v této analýze zohledňovány rozdíly mezi 積もり a 積り.

Subkorpus	Podíl kany	Interval spolehlivosti	
		dolní mez	horní mez
Vydané texty - noviny	100,00%	0,00%	100,00%
Vydané texty - knihy	98,57%	98,10%	98,96%
Vydané texty - časopisy	100,00%	36,17%	100,00%
Knihy v knihovnách	98,95%	97,87%	99,50%
Ostatní - blogy	98,86%	97,18%	99,59%
Ostatní - bestselery	99,39%	97,99%	99,87%
Ostatní - vládní jednání	100,00%	97,76%	100,00%
Ostatní - informační bulletiny	92,00%	70,77%	99,02%
Ostatní - učebnice	100,00%	0,00%	100,00%
Ostatní - zákoníky	0,00%	0,00%	0,00%
Ostatní - vládní zprávy	100,00%	0,00%	100,00%
Ostatní - Čiebukuro	99,68%	98,94%	99,93%
Ostatní - verše	100,00%	0,00%	100,00%

Tabulka 28: Podíly kany ve vzorku slova *cumori* a intervaly spolehlivosti na hladině významnosti 95% (NINJAL 2009, Brezina 2018)

13.3 Shrnutí

Pomocné jméno *cumori* je psáno v drtivé většině případů hiraganou. Zápis katakanou se neobjevuje vůbec a zápis znakem pouze okrajově. Zápisy znakem lze rozdělit na dvě varianty, které se liší připojenými slabikami hiragany. Rozdíly mezi jednotlivými subkorpusy jsou nepatrné, i když v několika případech statisticky významné.

14.Jotei (よてい、ヨテイ、予定)

Posledním vyšetřovaným pomocným jménem je *jotei*. Překládá se jako *určený plán, rozvrh* nebo *harmonogram*. Když se použije ve spojení se sponou, určuje tím, že předmět promluvy je nutný nebo záměrný. Proto se toto pomocné jméno řadí mezi pomocná podstatná jména modální. Míru modality má nula a je tak zástupcem málo gramatikalizovaného modálního pomocného jména.

14.1 První náhled

Pomocné podstatné jméno *jotei* se v korpusu vyskytuje celkem 5 814 krát. Z těchto výskytů jsou 4 (0,07%) zapsány hiraganou, zbylých 5 810 (99,93%) je psáno znakem. Zápis katakanou se v korpusu neobjevuje. Toto pomocné jméno se tedy píše výhradně znakem, 4 zápisy hiraganou jsou zanedbatelné. Stejně tak jako pro ostatní zkoumaná pomocná jména byly i pro *jotei* vytvořeny tabulky, které popisují jeho výskyt v korpusu. Jedná se o tabulky 29(43), 30(44) a 31(45). Odpovídají tabulkám 1, 2 a 3 v kapitole 7.1, ve které jsou podrobně popsány.

Subkorpus	ヨテイ	よてい	予定	Souhrn
Vydané texty	0,00%	25,00%	22,43%	22,43%
Vydané texty - noviny	0,00%	0,00%	14,50%	14,49%
Vydané texty - knihy	0,00%	100,00%	60,32%	60,35%
Vydané texty - časopisy	0,00%	0,00%	25,17%	25,15%
Knihy v knihovnách	0,00%	0,00%	12,32%	12,32%
Ostatní texty	0,00%	75,00%	65,25%	65,26%
Ostatní - blogy	0,00%	66,67%	42,76%	42,78%
Ostatní - bestselery	0,00%	0,00%	2,35%	2,35%
Ostatní - vládní jednání	0,00%	0,00%	6,30%	6,30%
Ostatní - informační bulletiny	0,00%	0,00%	12,56%	12,55%
Ostatní - učebnice	0,00%	0,00%	0,66%	0,66%
Ostatní - zákoníky	0,00%	0,00%	0,87%	0,87%
Ostatní - vládní zprávy	0,00%	0,00%	7,81%	7,80%
Ostatní - Čiebukuro	0,00%	33,33%	26,67%	26,67%
Ostatní - verše	0,00%	0,00%	0,03%	0,03%

Tabulka 29: Distribuce slova *jotei* v korpusu (NINJAL 2009)

Z Tabulky 29 lze soudit, že se většina výskytů tohoto slova objevuje v internetových zdrojích (*Ostatní - blogy* a *Ostatní - Čiebukuro*), za kterými

následují knihy (*Vydané texty – knihy* a *Knihy v knihovnách*). Tabulka 30 nic nového neukazuje a z Tabulky 31 lze vyčíst, že se slovo *jotei* vyskytuje s nejvyšší relativní četností v subkorpusu *blogů* a *časopisů*.

Subkorpus	ヨテイ	よてい	予定
Vydané texty	0,00%	0,08%	99,92%
Vydané texty - noviny	0,00%	0,00%	100,00%
Vydané texty - knihy	0,00%	0,13%	99,87%
Vydané texty - časopisy	0,00%	0,00%	100,00%
Knihy v knihovnách	0,00%	0,00%	100,00%
Ostatní texty	0,00%	0,08%	99,92%
Ostatní - blogy	0,00%	0,12%	99,88%
Ostatní - bestselery	0,00%	0,00%	100,00%
Ostatní - vládní jednání	0,00%	0,00%	100,00%
Ostatní - informační bulletiny	0,00%	0,00%	100,00%
Ostatní - učebnice	0,00%	0,00%	100,00%
Ostatní - zákoníky	0,00%	0,00%	100,00%
Ostatní - vládní zprávy	0,00%	0,00%	100,00%
Ostatní - Čiebukuro	0,00%	0,10%	99,90%
Ostatní - verše	0,00%	0,00%	100,00%
Celý korpus	0,00%	0,07%	99,93%

Tabulka 30: Podíly způsobů zápisu slova *jotei* v jednotlivých subkorpusech (NINJAL 2009)

Subkorpus	ヨテイ	よてい	予定	Součet
Vydané texty	0,00	0,03	38,05	38,08
Vydané texty - noviny	0,00	0,00	137,98	137,98
Vydané texty - knihy	0,00	0,035	27,63	27,66
Vydané texty - časopisy	0,00	0,00	74,13	74,13
Knihy v knihovnách	0,00	0,00	23,62	23,62
Ostatní texty	0,00	0,07	94,63	94,71
Ostatní - blogy	0,00	0,20	160,09	160,28
Ostatní - bestselery	0,00	0,00	23,81	23,81
Ostatní - vládní jednání	0,00	0,00	46,84	46,84
Ostatní - informační bulletiny	0,00	0,00	126,92	126,92
Ostatní - učebnice	0,00	0,00	27,03	27,029
Ostatní - zákoníky	0,00	0,00	30,58	30,58
Ostatní - vládní zprávy	0,00	0,00	60,64	60,64
Ostatní - Čiebukuro	0,00	0,10	98,77	98,87
Ostatní - verše	0,00	0,00	4,48	4,48
Celý korpus	0,00	0,04	55,54	55,58

Tabulka 31: Počty výskytů slova *jotei* na milion slov (NINJAL 2009)

14.2 Vliv subkorpusu

S takto malým zastoupením jiného, než znakového zápisu nelze očekávat žádné statisticky významné rozdíly mezi subkorpusy. Když se ty subkorpusy, které obsahují alespoň jeden zápis hiraganou, vyšetří logistickou regresní analýzou, výsledný model není statisticky významný.¹⁰ V takovém případě nemá smysl vytvářet přehledovou tabulku a interpretovat ji.

14.3 Shrnutí

Pomocné jméno *jotei* se téměř výlučně píše znakem. Podíl zápisu hiraganou je velice malý a katakana se nepoužívá. Rozdíly mezi jednotlivými subkorpusy nejsou statisticky významné.

¹⁰ Viz Příloha 1

15. Výsledky a diskuse

Ukazuje se, že všechna sledovaná pomocná podstatná jména mají určitý dominantní způsob zápisu a ostatní způsoby zápisu jsou pro ně pouze okrajové. Zápis katakanou je okrajový vždy a zkoumaná pomocná podstatná jména se tak zapisují především buď hiraganou, nebo znakem. Toto znázorňuje *Tabulka 32*. Z té je také vidět, že u pomocných jmen, která jsou ve své kategorii více gramatikalizovaná, převažuje zápis hiraganou. Naopak zápis znakem převažuje u těch pomocných jmen, která jsou ve své kategorii gramatikalizovaná méně.

Pomocné jméno	Skupina	Míra gramatikalizace	Dominantní způsob zápisu	Podíl výskytů
<i>uči</i>	předložková	vysoká	hiragana	92,32%
<i>šita</i>	předložková	nízká	znak	98,40%
<i>koto</i>	zájmenná	vysoká	hiragana	95,34%
<i>baai</i>	zájmenná	nízká	znak	99,45%
<i>hodo</i>	přísllovečná	vysoká	hiragana	79,27%
<i>fú</i>	přísllovečná	nízká	hiragana	89,89%
<i>cumori</i>	modální	vysoká	hiragana	99,03%
<i>jotei</i>	modální	nízká	znak	99,93%

Tabulka 32: Přehled pomocných jmen a jejich dominantních způsobů zápisu (NINJAL 2009)

Skupina pomocných podstatných jmen s přísllovečným užitím je zajímavá tím, že její méně gramatikalizované pomocné podstatné jméno *fú* je nejčastěji zapisováno hiraganou, zatímco ostatní méně gramatikalizovaní zástupci svých skupin jsou nejčastěji zapisováni znakem. Dále pomocné podstatné jméno *hodo*, které je více gramatikalizovaným zástupcem výše zmíněné skupiny, má ze všech pomocných podstatných jmen nejnižší podíl zápisu svým dominantním způsobem, v tomto případě hiraganou. Znamená to, že se toto pomocné podstatné jméno v korpusu objevuje zapsáno nedominantním způsobem (v tomto případě znakem) nejčastěji ze všech sledovaných podstatných jmen.

Logistická regresní analýza pro každé zkoumané pomocné podstatné jméno s výjimkou *jotei* ukázala alespoň dva subkorpusy, které

mezi sebou vykazují statisticky významné rozdíly v distribuci různých způsobů zápisu. V této souvislosti byly často zmiňovány subkorpusy *Ostatní – blogy* a *Ostatní – Čiebukuro*. Tyto internetové subkorpusy vykazují větší variabilitu zápisů a zastoupení dominantního způsobu zápisu je v nich tedy nižší, než je celokorpusový průměr. Naproti tomu subkorpusy oficiálních textů, např. *Ostatní – zákoníky* a *Ostatní – vládní zprávy* jsou ve sledovaných způsobech zápisu konzistentní a zastoupení dominantního způsobu zápisu v nich bývá nadprůměrné, často se v nich navíc jiný způsob zápisu ani nevyskytuje. Toto ilustruje *tabulka 33*.

Pomocné jméno	Průměr Podíl kany	Blogy Podíl kany	Čiebukuro Podíl kany	Zákoníky Podíl kany	Vládní zprávy Podíl kany
<i>uči</i>	92,58%	88,14%	88,23%	99,26%	99,05%
<i>šita</i>	1,60%	2,58%	1,55%	0,00%	0,00%
<i>koto</i>	95,43%	81,67%	72,03%	100,00%	99,95%
<i>baai</i>	0,55%	0,12%	0,15%	0,00%	0,00%
<i>hodo</i>	80,49%	62,17%	41,82%	0,00%	100,00%
<i>fú</i>	89,89%	27,40%	29,45%	0,00%	0,00%
<i>cumori</i>	99,03%	98,86%	99,68%	0,00%	100,00%
<i>jotei</i>	0,07%	0,12%	0,10%	0,00%	0,00%

Tabulka 33: Podíl zápisu kanou pomocných podstatných jmen ve vybraných subkorpusech (NINJAL 2009)

V *tabulce 33* je zanesen procentuální podíl zápisu hiraganou a katakanou dohromady. Ta pomocná podstatná jména, která mají dominantní způsob zápisu hiraganou, mají nadprůměrné zastoupení tohoto dominantního způsobu zápisu znázorněno číslem vyšším, než je celokorpusový průměr. Naopak ta pomocná podstatná jména, která se zapisují především znakem, mají nadprůměrné zastoupení svého dominantního způsobu zápisu znázorněno číslem nižším.

16.Závěr

Vyšetřovaná pomocná podstatná jména lze z hlediska způsobu zápisu rozdělit do dvou skupin, a sice na pomocná podstatná jména, u kterých dominuje zápis hiraganou, a na pomocná podstatná jména, u kterých dominuje zápis znakem. Pro potřeby této práce byla definována míra gramatikalizace jako počet různých gramatikalizovaných konstrukcí v gramatickém slovníku Nihongo bunkei džiten (日本語文型辞典; Sunagawa 1998). Ta pomocná podstatná jména, která mají tuto míru vyšší, mají zpravidla dominantní způsob zápisu hiraganou. Pomocná podstatná jména s nižší mírou gramatikalizace se zpravidla zapisují znakem.

Přestože zkoumaná pomocná podstatná jména mají svůj dominantní způsob zápisu, objevují se do jisté míry i rozdíly mezi jednotlivými subkorpusy. Tyto rozdíly jsou největší mezi subkorpusy internetových textů a subkorpusy oficiálních a formálních textů. Mezi různými subkorpusy knižních textů se výrazné rozdíly neobjevují.

Ministerská vyhláška č. 1 z 1. 10. 1981, která obsahuje doporučení kdy psát hiraganou a kdy znakem, se přímo pomocnými podstatnými jmény nezabývá. Je v ní zmíněno pouze pomocné podstatné jméno *koto* a doporučuje se zápis hiraganou. V korpusu dominuje zápis pomocného podstatného jména *koto* hiraganou i u textů starších, než je zmíněná vyhláška.

Korpus BCCWJ je členěn na morfémy a proto není snadné v něm filtrovat na úrovni lexémů. Je tedy možné, že se do vyšetřovaných výsledků dostaly i nějaké záznamy, které by do nich správně patřit neměly. Například v nich může být podstatné jméno v plnovýznamovém užití, kdy se nejedná o pomocné podstatné jméno, nebo se může jednat o část složeného slova. I když proběhlo filtrování s cílem všechny nepatřičné záznamy z analýzy vyloučit, bylo zpracováno tak velké množství záznamů, že podrobná kontrola všech není realizovatelná.

Statistická významnost rozdílů mezi jednotlivými subkorpusy byla ověřena pomocí logistické regresní analýzy. Ta svůj účel splnila a potvrdila pozorované výsledky. Nicméně naproti očekávání byl nástroj, kterým lze vysvětlit závislost dvou lingvistických proměnných, využit pouze jako nástroj pro ověření statistické významnosti pozorovaných rozdílů. Tu lze ověřit mnohem snáze jinými postupy, které jsou díky své jednoduchosti pro tento účel vhodnější, například chí-kvadrát test nebo t-test.

Bylo by zajímavé vyšetřit, jak se způsoby zápisu pomocných podstatných jmen vyvíjely v čase, ale korpus BCCWJ není vhodným zdrojem dat k takovému výzkumu. Obsahuje sice data z období zhruba 30 let, ale tato data nejsou reprezentativní. Jednotlivé subkorpusy totiž na sebe často časově navazují. Pro diachronní analýzu je nutné, aby se na časové ose překrývaly.

Vyšetřením ortografické variability dalšího slovního druhu se postupně zvyšuje porozumění japonštině a principům, na kterých funguje. Potvrdilo se, že míra gramatikalizace má zásadní vliv na způsob zápisu pomocných podstatných jmen. I když neexistuje žádný předpis, který by jejich psaní jednoznačně kodifikoval, u pomocných podstatných jmen, která byla vybrána pro svoji vyšší míru gramatikalizace, bez výjimky dominuje zápis hiraganou, zatímco u těch s nižší mírou gramatikalizace dominuje zápis znakem. Zde se ale jedna výjimka vyskytla. Pomocné podstatné jméno *fú* má dominantní zápis hiraganou, i když je méně gramatikalizovaným zástupcem skupiny pomocných podstatných jmen s příslovečným užitím. Nabízí se otázka, je-li tato výjimka způsobena tím, že je toto slovo něčím výjimečné a pokud ano, čím, nebo jestli se jedná o vlastnost, která je pomocným podstatným jménům s příslovečným užitím společná. Toto může být předmětem dalšího zkoumání.

Seznam použité literatury

ANDĚL, Jiří, 2007. *Statistické metody*. 4., upr. vyd. Praha: Matfyzpress. 299 s. ISBN 978-80-7378-003-6.

BAREŠOVÁ Ivona a SCHIR Martin, v tisku. Kana or Kanji? Preferences in Orthographic Representation of Japanese Auxiliary Verbs.

BREZINA, Vaclav, 2018. *Statistics in corpus linguistics: a practical guide*. 1. vyd. Cambridge: Cambridge University Press. xix, 296 stran. ISBN 978-1-107-12570-4.

Center for Corpus Development, b.r. *Summary Balanced Corpus of Contemporary Written Japanese*. National Institute for Japanese Language [online]. [cit. 24. 6. 2019]. Dostupné z: https://pj.ninjal.ac.jp/corpus_center/bccwj/en/

ČERMÁK, František, 2009. *Jazyk a jazykověda*. Přehled a slovníky. 1. vyd. Praha: Nakladatelství Karolinum. 341 s. ISBN 978-80-246-0154-0.

FAWCETT, Tom, 2006. An introduction to ROC analysis. *Pattern Recognition Letters* [online]. 27(8), 861-874 [cit. 25. 6. 2019]. DOI: <https://doi.org/10.1016/j.patrec.2005.10.010>.

JOŠIKAWA Taketoki et al., 2003. *Keišiki meiši ga kore de wakaru*. Tokio: Hicudži šóbó. 215 s. ISBN 4-89476-191-2.

KASHINO, Wakako a NAKAMURA, Takenori, 2013. Frequency of Katakana Representation for Japanese Non-loan Words as Observed in the BCCWJ Corpus. dostupné z: https://www.ninjal.ac.jp/event/specialists/project-meeting/files/JCLWorkshop_no4_papers/JCLWorkshop_No4_35.pdf

Kenkjúša, 2004. *Šin wa-ei daidžiten (新和英大辞典)*. [disk]. 5. vyd.

NINJAL, 2009. BCCWJ.

NOVÁK, Miroslav, 1989. *Gramatika japonštiny. [Díl] 1.* 2. vyd. Praha: SPN. 212 s.

NYMBURSKÁ Dita et al., 2007. *Japonština:Písmo.* Vyd. 1. Voznice: Leda. 383 s. ISBN 978-80-7335-074-1.

OGURA, Hideki, 2012. *Corpus-Based Survey of the Orthographic Variation in Contemporary Japanese: Analysis of the BCCWJ-Core.* [online]. dostupné z: https://www.ninjal.ac.jp/event/specialists/project-meeting/files/JCLWorkshop_no1_papers/JCLWorkshop2012_42.pdf

OGURA, Hideki, 2013. Orthographic Variation of Word-Final Long Vowels in Japanese Loanwords, dostupné z: https://www.ninjal.ac.jp/event/specialists/project-meeting/files/JCLWorkshop_no4_papers/JCLWorkshop_No4_30.pdf

ŘEHÁKOVÁ, Blanka, 2000. Nebojte se logistické regrese: Introducing Logistic Regression. *Sociologický časopis: Czech Sociological Review* [online]. 36(4), 475-492 [cit. 2019-06-26]. ISSN 2336-128X. Dostupné z: <http://www.jstor.org/stable/41131581>

SUNAGAWA, Juriko, 1998. *Nihongo bunkei džiten.* Tokio: Kurosio Publishers. 693 s. ISBN: 4-87424-154-9

Seznam příloh

Příloha 1: Logistická regresní analýza: Popis a výsledky

Příloha 1: Logistická regresní analýza

Tato příloha obsahuje výstupní tabulky nástroje Lancaster Tools Online (Brezina 2018). Tyto tabulky obsahují výsledky logistické regresní analýzy a svým charakterem přesahují rámec této práce. Nejprve je na příkladu první tabulky uveden popis toho, co tato tabulka obsahuje a následuje vysvětlení postupu, jakým se z této tabulky získaly přehledové tabulky, které jsou uvedeny u každého zkoumaného pomocného podstatného jména. Poté jsou zařazeny tabulky pro ostatní zkoumaná slova, tentokrát bez komentáře.

Popis

Před použitím regresní analýzy je třeba udělat několik rozhodnutí. Pro každou z proměnných, které se v ní vyskytnou, je třeba vybrat jednu hodnotu, která bude prohlášena za základní a vůči ní porovnány všechny ostatní hodnoty dané proměnné. Ve vyšetření vlivu subkorpusů na způsob zápisu vystupují proměnné dvě – příslušnost k subkorpusu a způsob zápisu. Protože se u zvolených pomocných podstatných jmen objevuje zápis katakanou pouze okrajově, je pro potřeby regresní analýzy sloučen zápis hiraganou a katakanou do jediné položky, zápis kanou. Způsob zápisu tedy může nabývat dvou hodnot. Zápis je proveden buď kanou, nebo znakem. Použije-li se za základní hodnotu zápis kanou, ve výsledné tabulce se objeví čísla, která se interpretují jako šance na zápis znakem. Při použití znaku jako základní hodnoty vznikne naopak šance na zápis kanou. Obě možnosti si jsou ekvivalentní. Jako základní hodnota byl pro tuto práci zvolen zápis znakem. Interpretace výsledku „šance na zápis kanou“ je vhodnější pro ta pomocná jména, pro která je zápis kanou dominantním způsobem zápisu a těchto je mezi zkoumanými jmény víc. Je sice možné volit tuto hodnotu jinak pro každé zkoumané pomocné jméno, ale v zájmu snadné porovnatelnosti výsledků je lepší, jsou-li základní hodnoty zvoleny u všech jmen stejně. Druhou proměnnou je příslušnost k subkorpusu, kde byl za základní zvolen subkorpus *Vydané texty - knihy*. Tento subkorpus se ukázal jako velmi dobrý základ pro

porovnávání, protože zápis všech zkoumaných pomocných podstatných jmen v něm odpovídá svým průměrným hodnotám a to jak v četnosti, tak v podílu jednotlivých způsobů zápisu. V následujících tabulkách mu přísluší řádek (*Intercept*).

Výsledkem této analýzy je model v *Tabulce 34*. Tato tabulka je převzata z online nástroje Lancaster Stat Tools online (Brezina 2018) a skládá se ze dvou částí. První částí je popis modelu (*Overall model statistics*). Prvním údajem v tomto popisu je výsledek testu pomocí poměru věrohodností (*Likelihood ratio test*), který má za cíl ověřit, že je tento model statisticky významně lepší, než by byl model bez vysvětlujících proměnných. (Brezina 2018) Pokud by v tomto testu vyšlo, že rozdíl není statisticky významný, znamenalo by to, že v další interpretaci by už nemělo smysl pokračovat, protože spočtené hodnoty mohou s dostatečně velkou¹ pravděpodobností být pouze dílem náhody. V tomto případě výsledkem tohoto testu je, že model s příslušností k subkorpusu statisticky významně lepší je. U výsledného čísla a jemu odpovídající hladině významnosti, zde označené jako p , je pouze důležité, aby nebyla překročena hranice 0,05. Čím větší toto číslo je, tím menší je hladina významnosti. Hladina významnosti označuje nejvyšší povolenou chybu. Pro interpretaci se nehledí na konkrétní hodnotu, které model dosáhl, ale pouze na to, zda byla nebo nebyla překročena daná hranice 5%.

¹ Nejčastěji se jako hraniční hodnota uvádí 0,05 (Brezina 2018, Anděl 2007, s. 9). Zde je tato pravděpodobnost menší než 0,0001.

Overall model statistics: Likelihood ratio test (LL): 612,83 ($p < 0,0001$) -> SIGNIFICANT; C-index: 0,61 -> NOT ACCEPTABLE; Nagelkerke R^2 : 0,04; AIC: 16959,91

Coefficients:

	Estimate (log odds)	Standard Error	Z value (Wald)	p- value	Estimate (odds)	95% CI lower	95% CI upper
(Intercept)	2,344	0,036	64,360	0,000	10,426	9,714	11,206
Vydané texty - časopisy	0,543	0,146	3,725	0,000	1,721	1,306	2,316
Vydané texty - noviny	0,842	0,266	3,166	0,002	2,321	1,430	4,086
Knihy v knihovnách	0,145	0,051	2,852	0,004	1,157	1,047	1,278
Ostatní - blogy	-0,339	0,076	-4,452	0,000	0,713	0,615	0,828
Ostatní - bestselery	-0,011	0,098	-0,111	0,912	0,989	0,818	1,204
Ostatní – vládní jednání	2,453	0,357	6,872	0,000	11,617	6,192	25,552
Ostatní - informační bulletiny	0,822	0,212	3,883	0,000	2,274	1,537	3,536
Ostatní - učebnice	0,804	0,325	2,474	0,013	2,235	1,248	4,522
Ostatní - zákoníky	2,554	0,450	5,670	0,000	12,852	5,932	36,008
Ostatní – vládní zprávy	2,307	0,193	11,930	0,000	10,040	7,015	15,012
Ostatní - Čiebukuro	-0,330	0,076	-4,336	0,000	0,719	0,620	0,836
Ostatní - verše	0,124	0,467	0,265	0,791	1,132	0,500	3,248

Tabulka 34: Logistická regresní analýza slova *učí* (Brezina 2018)

Dalším údajem je *C-index*. Tato hodnota udává klasifikační nebo též prediktivní sílu modelu. (Brezina 2018) Součástí dat v druhé části tabulky je šance, že výskyt *učí* v různých subkorpusech bude psán kanou. *C-index* je pravděpodobnost, že náhodně vybranému záznamu bude na základě této šance přiřazen správný způsob zápisu. (Fawcett, 2006, s. 869)² Jedná se o údaj, který při hledání celkového trendu užití nemá a zde je popsán pro úplnost. Výchozí hodnotou je 0,5 a perfektní model by měl *C-index* roven 1. (Fawcett, 2006, s. 869) Za přijatelnou hodnotu Lancaster Stats Online považuje 0,7 a vyšší. (Brezina 2018)

Následuje *Nagelkerkeův R kvadrát*, který určuje množství variace, kterou model vysvětluje. (Brezina 2018) Uvedená 4% jsou velmi malá

² *C-index* se v citovaném článku označuje jako AUC a je definován v jiném kontextu. Interpretace, že se jedná o pravděpodobnost správné funkce modelu, je však stejná. Informace o tom, že *C-index* a AUC je totožná věc je v poznámkovém aparátu Brezina, 2008

hodnota. Poslední vlastnost, *AIC*, je hodnota, která slouží pro porovnávání dvou různých modelů nad stejnými daty. (Brezina 2018) V tomto výzkumu je použit pouze jeden model, proto je hodnota *AIC* irelevantní.

Po údajích, které popisují vlastnosti modelu, následuje část zvaná koeficienty (*Coefficients*), která popisuje model jako takový. Jako *odds* je označena šance, že náhodně vybraný výskyt v daném subkorpusu je zapsán kanou. Je důležité oddělit pojmy šance a pravděpodobnost. Rozdíl mezi nimi je ten, že šance nějakého jevu je číslo určující, kolikrát je pravděpodobnost tohoto jevu větší, než pravděpodobnost toho, že tento jev nenastane. (Brezina 2018)

První sloupec (*Estimate (log odds)*) obsahuje hodnotu přirozeného logaritmu této šance. Je to technický údaj, který je zde ponechán, protože je součástí výstupu použitého nástroje. Pro interpretaci je lepší odpovídající sloupec *Estimate (odds)*, který obsahuje tuto šanci přímo.

Následuje sloupec se *standardní chybou (Standard Error)*, která určuje, jak přesně spočítaná šance reflektuje skutečnost. Čím menší číslo, tím lépe. (Brezina 2018) *Waldovo* z a s ním svázaná *hladina významnosti (p-value)* určují statistickou významnost každého řádku. Ty řádky, které mají *hladinu významnosti* větší než 0,05, lze z další interpretace vyloučit. Znamená to totiž, že nejsou statisticky významné. Hodnoty v nich pozorované mohou být s příliš velkou pravděpodobností dílem náhody. Neznamená to, že tyto údaje jsou špatné, pouze nelze vyloučit jejich náhodný původ. (Brezina 2018)

Další sloupec *Estimate (odds)* obsahuje šanci, že výskyt patří do kategorie určené řádkem, bude napsán kanou. V prvním řádku, který obsahuje popis základních hodnot vysvětlujících proměnných, (Brezina 2018) se s danou hodnotou počítá jinak, než ve zbytku tabulky. Základní hodnotou je subkorpus *Vydané texty – knihy*. Pro ten je uvedena šance zápisu uči kanou 10,426. Pravděpodobnost se z šance určí vztahem:

Rovnice 1: Výpočet pravděpodobnosti z šance (Řeháková 2000, s. 477)

$$P(A) = \frac{\check{S}(A)}{1 + \check{S}(A)} = \frac{10,426}{11,426} \doteq 0,9125 = 91,25\%$$

Kde $P(A)$ značí pravděpodobnost, že nastane nějaký jev, v tomto případě že náhodně vybraný zápis slova uči v subkorpusu *Vydané texty - knihy* je zapsán kanou. $\hat{S}(A)$ je šance téhož jevu. Tato vypočtená hodnota odpovídá *Tabulce 2*, kde je pro tento subkorpus uveden podíl zápisu hiraganou 91,10% a katakanou 0,15%, což je v součtu právě 91,25%.

Všechny ostatní řádky se neinterpretují takto přímo, protože jsou vyjádřeny vůči tomu prvnímu. To znamená, že například hodnota uvedená pro subkorpus *Ostatní - vládní zprávy* (10,040), říká, že šance na výskyt kanou v tomto subkorpusu je desetkrát větší, než v *Vydané texty - knihy*. Platí tedy, že: $10,426 \cdot 10,040 \doteq 104,68$, což po dosazení do vzorečku výše dá pravděpodobnost výskytu v kaně v tomto subkorpusu 0,9905 (99,05%). Toto opět odpovídá údajům 99,05% v *Tabulce 2*. Protože je uvedená šance relativní vůči prvnímu řádku, je hodnota 1 specifická tím, že značí situaci, kdy k žádné změně šance nedojde, hodnoty menší značí menší šanci a s většími hodnotami zase šance na zápis kanou vzrůstá. Lze tedy shrnout, že čím větší číslo v tabulce je, tím větší podíl kany danému řádku odpovídá.

Poslední dva sloupce udávají interval spolehlivosti na hladině významnosti 95%. V korpusu nejsou zastoupeny všechny texty, jedná se o jejich výběr. Interval spolehlivosti s hladinou významnosti 95% říká, že sledovaná hodnota se bude ve všech textech na základě výskytů ve vzorku vyskytovat s pravděpodobností 95% uvnitř tohoto intervalu. Interval spolehlivosti je uveden pro šanci, což znamená, že, vezme-li se za příklad opět *Ostatní - vládní zprávy*, je interpretací výsledků konstatování, že na základě pozorování je ve všech vládních zprávách podíl zápisu kanou něco mezi 98,55% a 99,41%. Vynásobené šance z dolní hranice intervalu jsou $9,714 \cdot 7,015 \doteq 68,14$, po dosazení do vzorečku v Rovnici 1 se získá pravděpodobnost 98,55% a horní hranice se získá obdobně.

V *Tabulce 4* jsou tedy vlastně zakódovány údaje z *Tabulky 2*, doplněné o další údaje, z nichž nejdůležitější je hladina významnosti a interval spolehlivosti. Tyto údaje lze pro snazší čitelnost zanechat do zvláštní tabulky (*Tabulka 5*).

Overall model statistics: Likelihood ratio test (LL): 43,34 ($p < 0,0001$) -> SIGNIFICANT; C-index: 0,58 -> NOT ACCEPTABLE; Nagelkerke R^2 : 0,03; AIC: 1461,87

Coefficients:

	Estimate (log odds)	Standard Error	Z value (Wald)	p-value	Estimate (odds)	95% CI lower	95% CI upper
(Intercept)	-4,152	0,157	-26,377	0,000	0,016	0,011	0,021
Vydané texty - časopisy	0,006	0,441	0,013	0,989	1,006	0,381	2,211
Knihy v cirkulaci	0,109	0,204	0,534	0,593	1,115	0,750	1,675
Ostatní - blogy	0,504	0,298	1,691	0,091	1,656	0,897	2,912
Ostatní - bestselery	-1,235	0,725	-1,703	0,089	0,291	0,047	0,949
Ostatní - informační bulletiny	0,181	1,022	0,177	0,859	1,199	0,067	5,676
Ostatní - Čiebukuro	-0,014	0,342	-0,042	0,967	0,986	0,480	1,860
Ostatní - verše	2,699	0,370	7,294	0,000	14,870	6,912	29,871

Tabulka 35: Logistická regresní analýza slova *šita* (Brezina 2018)

Overall model statistics: Likelihood ratio test (LL): 3003,58 ($p < 0,0001$) -> SIGNIFICANT; C-index: 0,85 -> EXCELLENT; Nagelkerke R^2 : 0,32; AIC: 7873,71

Coefficients:

	Estimate (log odds)	Standard Error	Z value (Wald)	p-value	Estimate (odds)	95% CI lower	95% CI upper
(Intercept)	4,243	0,088	48,219	0,000	69,618	58,865	83,135
Vydané texty - časopisy	0,631	0,389	1,620	0,105	1,880	0,945	4,451
Knihy v knihovnách	0,169	0,129	1,308	0,191	1,184	0,920	1,526
Ostatní – blogy	-2,749	0,106	-26,009	0,000	0,064	0,052	0,078
Ostatní – bestselery	0,053	0,260	0,204	0,838	1,054	0,653	1,818
Ostatní – vládní jednání	2,857	0,713	4,008	0,000	17,402	5,544	105,388
Ostatní – informační bulletiny	-0,139	0,391	-0,356	0,722	0,870	0,436	2,065
Ostatní – Čiebukuro	-3,308	0,098	-33,615	0,000	0,037	0,030	0,044

Tabulka 36: Logistická regresní analýza slova *koto* (Brezina 2018)

Overall model statistics: Likelihood ratio test (LL): 477,76 ($p < 0,0001$) -> SIGNIFICANT; C-index: 0,73 -> ACCEPTABLE; Nagelkerke R^2 : 0,08; AIC: 5954,66

Coefficients:

	Estimate (log odds)	Standard Error	Z value (Wald)	p-value	Estimate (odds)	95% CI lower	95% CI upper
(Intercept)	-4,962	0,066	-75,482	0,000	0,007	0,006	0,008
Vydané texty - časopisy	-2,096	0,710	-2,950	0,003	0,123	0,020	0,383
Vydané texty - noviny	-16,604	1459,798	-0,011	0,991	0,000	0,000	0,000
Knihy v knihovnách	0,638	0,092	6,955	0,000	1,892	1,581	2,265
Ostatní	-1,804	0,581	-3,103	0,002	0,165	0,041	0,431
Ostatní - blogy	-0,189	0,341	-0,555	0,579	0,828	0,393	1,519
Ostatní - bestselery	-16,604	366,697	-0,045	0,964	0,000	0,000	0,000
Ostatní - vládní jednání	-16,604	399,247	-0,042	0,967	0,000	0,000	0,000
Ostatní - informační bulletiny	-16,604	1007,416	-0,016	0,987	0,000	0,000	0,000
Ostatní - učebnice	-16,604	400,745	-0,041	0,967	0,000	0,000	0,000
Ostatní - zákony	-16,604	506,498	-0,033	0,974	0,000	0,000	0,000
Ostatní - vládní zprávy	-1,531	0,228	-6,715	0,000	0,216	0,134	0,330
Ostatní - verše	-16,604	13073,143	-0,001	0,999	0,000	NA	Inf

Tabulka 37: Logistická regresní analýza slova *baai* (Brezina 2018)

Overall model statistics: Likelihood ratio test (LL): 153,15 ($p < 0,0001$) -> SIGNIFICANT; C-index: 0,83 -> EXCELLENT; Nagelkerke R^2 : 0,37; AIC: 437,46

Coefficients:

	Estimate (log odds)	Standard Error	Z value (Wald)	p-value	Estimate (odds)	95% CI lower	95% CI upper
(Intercept)	1,833	0,311	5,894	0	6,25	3,536	12,095
Vydané texty - časopisy	0,693	0,798	0,869	0,385	2	0,5	13,416
Vydané texty - noviny	17,733	1450,071	0,012	0,99	50298562,06	1,8443E+202	7,2794E+265
Knihy v knihovnách	-0,479	0,389	-1,232	0,218	0,619	0,281	1,307
Ostatní - blogy	-1,335	0,385	-3,463	0,001	0,263	0,12	0,548
Ostatní - bestselery	17,733	2292,763	0,008	0,994	50298562,53	0	Inf
Ostatní - vládní jednání	17,733	1217,652	0,015	0,988	50298562,15	0	3,9852E+208
Ostatní - informační bulletiny	17,733	2150,803	0,008	0,993	50298562,14	0	Inf
Ostatní - učebnice	17,733	7604,235	0,002	0,998	50298562,07	0	NA
Ostatní - vládní zprávy	17,733	10754,013	0,002	0,999	50298562,13	0	NA
Ostatní - Čiebukuro	-2,16	0,39	-5,536	0	0,115	0,052	0,241
Ostatní - verše	0,247	1,105	0,223	0,823	1,28	0,206	24,833

Tabulka 38: Logistická regresní analýza slova *hodo* (Brezina 2018)

Overall model statistics: Likelihood ratio test (LL): 1375,75 ($p < 0,0001$) -> SIGNIFICANT; C-index: 0,7 -> ACCEPTABLE; Nagelkerke R^2 : 0,26; AIC: 6889,5

Coefficients:

	Estimate (log odds)	Standard Error	Z value (Wald)	p- value	Estimate (odds)	95% CI lower	95% CI upper
(Intercept)	1,491	0,058	25,589	0,000	4,440	3,967	4,985
Vydané texty - časopisy	0,009	0,167	0,053	0,958	1,009	0,733	1,412
Vydané texty - noviny	-0,664	0,457	-1,453	0,146	0,515	0,218	1,348
Knihy v knihovnách	0,035	0,075	0,469	0,639	1,036	0,893	1,200
Ostatní - blogy	-2,464	0,121	-20,335	0,000	0,085	0,067	0,108
Ostatní - bestselery	0,010	0,128	0,077	0,939	1,010	0,789	1,302
Ostatní - informační bulletiny	-1,203	0,543	-2,215	0,027	0,300	0,104	0,917
Ostatní - Čiebukuro	-2,360	0,102	-23,236	0,000	0,094	0,077	0,115
Ostatní - verše	-4,987	0,719	-6,936	0,000	0,007	0,001	0,022

Tabulka 39: Logistická regresní analýza slova *fú* (Brezina 2018)

Overall model statistics: Likelihood ratio test (LL): 39,38 ($p < 0,0001$) -> SIGNIFICANT; C-index: 0,63 -> NOT ACCEPTABLE; Nagelkerke R^2 : 0,03; AIC: 1189,17

Coefficients:

	Estimate (log odds)	Standard Error	Z value (Wald)	p- value	Estimate (odds)	95% CI lower	95% CI upper
(Intercept)	4,232	0,155	27,232	0,000	68,881	51,524	94,9
Vydané texty - časopisy	15,334	584,941	0,026	0,979	4563903,437	0,011	4,1689E+102
Vydané texty - noviny	15,334	1133,573	0,014	0,989	4563903,444	0,000	4,2072E+192
Knihy v knihovnách	0,311	0,217	1,432	0,152	1,365	0,890	2,09
Ostatní - blogy	0,228	0,341	0,670	0,503	1,256	0,668	2,575
Ostatní - bestselery	0,854	0,525	1,626	0,104	2,348	0,946	7,831
Ostatní - vládní jednání	15,334	456,894	0,034	0,973	4563903,408	0,847	4,10846E+81
Ostatní - informační bulletiny	-1,790	0,753	-2,376	0,018	0,167	0,047	1,061
Ostatní - učebnice	15,334	2195,154	0,007	0,994	4563903,408	0,000	NA
Ostatní - vládní zprávy	15,334	4809,341	0,003	0,997	4563903,442	0,000	NA
Ostatní - Čiebukuro	1,498	0,524	2,856	0,004	4,471	1,804	14,897
Ostatní - verše	15,334	3802,118	0,004	0,997	4563903,407	0,000	NA

Tabulka 40: Logistická regresní analýza slova *cumori* (Brezina 2018)

Overall model statistics: Likelihood ratio test (LL): 0,04 ($p = 0,979$) -> NOT SIGNIFICANT; C-index: 0,5 -> NOT ACCEPTABLE; Nagelkerke R^2 : 0; AIC: 67,97

Coefficients:

	Estimate (log odds)	Standard Error	Z value (Wald)	p-value	Estimate (odds)	95% CI lower	95% CI upper
(Intercept)	-6,667	1,001	-6,663	0,000	0,001	0,000	0,006
Ostatní - blogy	-0,031	1,225	-0,025	0,980	0,970	0,093	20,887
Ostatní - Čiebukuro	-0,252	1,415	-0,178	0,859	0,777	0,031	19,681

Tabulka 41: Logistická regresní analýza slova *jotei* (Brezina 2018)