UNIVERZITA KARLOVA – FILOZOFICKÁ FAKULTA
ÚSTAV ANGLICKÉHO JAZYKA A DIDAKTIKY

Filologie – Anglický jazyk

Marie Vaňková

ANALÝZA STŘEDNÍ ANGLIČTINY ONLINE: TVORBA A VYUŽITÍ DATABÁZE SPELLINGOVÝCH VARIANT ZALOŽENÉ NA LAEME

ANALYSING EARLY MIDDLE ENGLISH ONLINE: CONSTRUCTION AND USE OF A LAEME BASED SPELLING DATABASE

Teze k disertační práci

Vedoucí práce: Mgr. Ondřej Tichý, PhD.

2021

Obsah

Obsah.....**Chyba! Záložka není definována.**
1. Úvod 3
   1.1. Vymezení cílů 3
   1.2. Struktura práce 3
2. Teoretická východiska 3
3. Tvorba nástroje 5
   3.1. LAEME 5
   3.2. Metodologická východiska 5
   3.3. Struktura databáze a zpracování dat z LAEME 6
   3.4. Zpracování dat 6
   3.5. Dotazy a výpočty 7
      3.5.1. Základní jednotky 7
      3.5.2. Vizualizace dat a statistické údaje 8
   3.6. Experimentální funkce 8
   3.7. Návrh postupu při vyhledávání 8
4. Výsledky 8
   4.1. Segmentace forem 8
   4.2. Uživatelské rozhraní 9
5. Závěry 12
6. Zdroje 14
   6.1. Online Resources 18

1. Úvod

1.1. Vymezení cílů

Disertační práce dokumentuje tvorbu výzkumného nástroje pro analýzu textů ve střední angličtině a vytvořený nástroj zároveň hodnotí v kontextu dosavadního vývoje v příslušné oblasti zkoumání. Nástroj je založen na bohatě anotovaném korpusu a atlasu střední angličtiny *A Linguistic Atla of Early Middle English* (*LAEME*) a zahrnuje databázi variant zápisu a rozhraní uzpůsobené pro prohledávání této databáze. Sestavení databáze spočívalo ve zpracovaní dat z LAEME tak, aby bylo možné srovnávat krátké segmenty (většinou jednotlivá písmeny nebo digrafy), které si vzájemně odpovídají. Struktura databáze rozšiřuje možnosti kvantifikace dat, ale také otevírá prostor pro nové typy dotazů. Při navrhování rozhraní byl kladen důraz na usnadnění přístupu k datům a jejich vizualizaci.

Nástroj byl navržen na základě předchozích zkušeností autorky s LAEME. Hlavním záměrem bylo najít cesty, jak efektivněji využít data, která LAEME nabízí.

1.2. Struktura práce

Teoretická část práce se zabývá problémy v oblasti zkoumání střední angličtiny, zejména vztahem mezi psaným a mluveným jazykem a šířením jazykových změn v čase a prostoru. Shrnuje dosavadní poznatky a metody historické dialektologie, které byly v minulosti použity nebo navrženy k analýze středoanglických textů.

Metodologická část charakterizuje LAEME jakožto hlavní zdroj použitý pro tvorbu nástroje, podrobně popisuje strukturu databáze a proces jejího sestavení. Rovněž vysvětluje uspořádání dat, která lze z databáze získat, a nastiňuje možné postupy při dotazování.

Praktická část krátce popisuje, jak byl vyřešen přístup k datům na úrovni uživatelského rozhraní a následně předvádí využití jednotlivých funkcí nástroje formou krátkých mikroanalýz. Tyto experimentální analýzy potom slouží jako podklad pro závěrečné hodnocení nástroje, včetně možností jeho rozšíření.

2. Teoretická východiska

Projekt výzkumného nástroje se teoreticky opírá o přehled poznatků z oblasti zkoumání střední angličtiny, který propojuje teoretické a metodologické koncepty s popisem konkrétních aplikací a elektronických zdrojů. Tento přehled jednak nabízí hlubší vhled do problematiky zkoumání raně střední angličtiny, která v některých ohledech představuje obzvláště náročný materiál, a jednak zmiňuje konkrétní postupy z oblasti historické dialektologie, které bylo možné při tvorbě nástroje využít.

Z teoretických témat, jimiž se práce zbývá, jsou patrně nejpodstatnější problematika psaného jazyka a jeho vztahu k jazyku mluvenému a šíření jazykových změn. Vychází přitom z děl Josefa Vachka (Vachek & Luelsdorff, 1989) i Anguse McIntoshe (McIntosh et al., 1989), kteří shodně zdůrazňovali určitou míru autonomie psaného jazyka coby systému a stavěli se kriticky k pojímání psaného jazyka jako „pouhého" odrazu jazyka mluveného. Psaný jazyk se dále rozebírá přímo v souvislosti se střední angličtinou, modely psaného jazyka a jejich klasifikací. Pozornost je věnována také současným poznatkům o přepisování textů a tzv. „písařským strategiím" („scribal strategies").

Pokud jde o metodologické postřehy, práce se zaměřuje na obor historické dialektologie, který se podle Williamsona (2004) vyznačuje tím, že se vyhýbá zjednodušování jazykových dat, ale snaží se jim porozumět v jejich složitosti. Přitom se neustále potýká s nedostatkem dat, což je v případě raně střední angličtiny obzvláště palčivý problém.

Mezi konkrétní elektronické nástroje z oblasti historické lingvistiky, které mají výrazné styčné body s nově vyvíjenou databází, patří mimo jiné několik projektů vzniklých přímo pod Centrem Anguse McIntoshe (AMC) v Edinburghu, zejména elektronická verze *A Linguistic Atlas of Late Medieval Eglish* (eLALME), *The Corpus of Narrative Etymologies* (CoNE) a především *From English to Scots* (FITS).

Projekt *From English to Scots* (FITs, Kopaczyk et al., 2018), postavený na *Linguistic Atlas of Older Scots* (LAOS) má jednoznačně největší průnik s nově vyvíjeným nástrojem, jelikož je také postaven na segmentaci forem a srovnávání jednotlivých písmen a digrafů. Databáze FITS navíc obsahuje starší reflexy forem z korpusu LAOS, rekonstrukci zvukové stránky jazyka a popis hláskových změn, které lze v datech identifikovat.

Důležitým konceptem, který byl nakonec využity při tvorbě nového nástroje patří *litterae* („litery") (Benskin, 1997; Laing & Lass, 2013), což je rámec, kteří autoři LAEME používají k analýze psaného jazyka a „významu" jednotlivých písmen. „Litera" (písmeno) je „abstraktní jednotka" (Laing & Lass, 2013: 2.3.1), která může zastupovat určitý zvuk, nebo více zvuků, a zapisuje se pomocí *figur* (viditelných tvarů v rukopisu). Upřednostnění tohoto modelu před známějšími pojmy jako *foném* a *grafém* autoři vysvětlují tím, že model liter odráží způsob, jakým byli písaři vzděláváni, zatímco vymezení *fonémů* vyžaduje analýzy na jazykového systému, které lze u písařů stěží přepokládat.

3. Tvorba nástroje

3.1. LAEME (Laing & Lass, 2013)

Sestavení databáze spočívalo v transformaci dat z LAEME. Základem LAEME je korpus raně středověkých textů vytvořený přímo podle rukopisů (popřípadě jejich snímků). V poměru k jiným historickým korpusům byly texty přepsány velmi detailně a každému slovu (s výjimkou vlastních jmen) byl přiřazen tzv. *lexel* (přibližně odpovídá lemmatu) a *grammel* (gramatický tag). Zvláštností LAEME v porovnání s většinou korpusů je, že rozlišuje rozdělení slov na jednotlivé morfémy.

Jednotlivé texty (korpusové soubory) v LAEME by měly představovat víceméně konzistentní „jazyky" jednotlivých písařů. Obsah jednoho rukopisu tak často bývá rozdělen do více souborů podle toho, kolik písařů se na něm podílelo a případně podle odlišností v systémech zápisu využívaných jedním písařem.

3.2. Metodologická východiska

V návaznosti na teoretický základ byly předem stanoveny následující obecné požadavky na finální podobu nástroje:

1. Nástroj by měl co nejvíce usnadnit identifikaci variant, které jsou v korpusu LAEME vzácné.

2. Nástroj by měl urychlit přístup k potencionálně relevantním datům tak, aby mohla být interpretace založena na co nejrozsáhlejším materiálu.

3. Metoda použitá pro tvorbu nástroje by měla být aplikovatelná na další zdroje dat, takže databázi bude možné rozšířit například o staroanglické varianty zápisu.

4. Zkoumání vzorců patrných na úrovni celé databáze by mělo jít snadno propojit s podrobnou analýzou konkrétních forem v konkrétních textech.

Klíčovým mechanismem, který otevírá nové možnosti zpracování dat, je přitom segmentace forem do tzv. slotů (pozic), tedy explicitní určení toho, která písmena, popřípadě krátké úseky 2-3 písmen si vzájemně odpovídají. Například u dvou variant zájmena 1. osoby jednotného čísla *ic* a *ich* si vzájemně odpovídají *i-i* a *c-ch*.

Vzhledem k rozsahu dat v LAEME, si práce mimo jiné klade za cíl vyvinout částečně automatizovanou metodu zpracování. První návrh této metody byl na základě testu vyhodnocen jako nedostačující, a proto byly nadefinovány požadavky na jeho úpravy.

3.3. Struktura databáze a zpracování dat z LAEME

Data z LAEME, byla pro potřeby projektu převedena do relační databáze a příslušné tabulky byly následně propojeny s tabulkami obsahujícími zpracovaná data. Data z LAEME zahrnují vlastní přepis textů, včetně členění na morfémy, i základní metadata, jako údaje o obsažených textech, lokalizaci a dataci rukopisů.

Jádrem nově vytvořené databáze jsou tři tabulky, z nichž první slouží jako rejstřík slov, druhá jako rejstřík jejich forem a třetí určuje obsazení jednotlivých „pozic" v těchto formách. Jednotlivé tabulky jsou vzájemně provázány. Například slovo FATHER/N obsažené v první tabulce je navázáno na 28 různých forem v druhé tabulce (např. *fader*, *federe*, *uader* atd.). Každá varianta je rozdělena na jednotlivé segmenty (např. *f/a/d/e/r*, *u/a/d/e/r*) a realizace jednotlivých segmentů je zaznamenána ve třetí tabulce. Každý řádek ve třetí tabulce se odkazuje na číslo příslušné formy a pořadí segmentu.

Rejstřík slov a rejstřík forem byly vytvořeny převážně pomocí SQL dotazů a nejnáročnější část zpracování dat spočívala v segmentaci jednotlivých forem. Kromě těchto základních tabulek databáze obsahuje ještě odvozené tabulky, které umožňují rychlejší zpracování dotazů (například statistické údaje o četnosti jednotlivých segmentů), a dále experimentální tabulky navržené pro účely propojení dat z LAEME s dalšími zdroji. Databáze pokrývá lexikální výrazy a omezené množství gramatických slov.

3.4. Zpracování dat

Pro účely projektu byl naprogramován počítačový skript v jazyce Python, který umí načíst skupinu forem určitého slova a vytvářet různé varianty jejich segmentace a „zarovnání". Opírá se přitom o seznam písmen a digrafů a dále o seznam pravděpodobných kombinací písmen (*sety*), která si mohou odpovídat, např. {*f*, *v*, *u*}, {*sh*, *sch*, *sc*} atd. Skript nejprve vygeneruje různé řetězce abstraktních „pozic", na něž by bylo možné jednotlivé formy rozdělit, přičemž rozlišuje mezi samohláskami (V) a souhláskami (C). Například pozice ve výše zmíněné formě *fader* (FATHER/N) lze popsat řetězcem „CVCVC". Poté se skript snaží jednotlivé pozice vyplnit segmenty všech forem daného slova tak, aby se na každé pozici spolu vyskytovaly pouze „povolené" kombinace segmentů. Pokud se to podaří, výsledek může být uložen do databáze.

Skript může v omezené míře pracovat automaticky, většinou však vyžaduje manuální zásahy do seznamu povolených kombinací segmentů nebo výběru správného řetězce pozic. Se zásahy tohoto typu bylo počítáno už v původním návrhu skriptu a z hlediska časové náročnosti jsou výrazně rychlejší než plně manuální segmentace.

Přibližně 33 % dat bylo zpracována automaticky 54 % vyžadovalo manuální zásah do počítačového skriptu (například explicitní předepsání počtu a typu pozic) a 13 % nejproblematičtějších slov (případně některých jejich forem) muselo být zpracováno ručně pomocí jednoduchého webového rozhraní. V některých případech bylo třeba přihlédnout k dalším formám v textu, z něhož problematická varianta pocházela.

3.5. Dotazy a výpočty

Kromě vlastního zpracování dat tvorba nástroje vyžadovala také definici dotazů, které lze do databáze zadávat, uspořádání výstupních dat, možností jejich filtrování i pokročilejších výpočtů a postupů při tvorbě map a síťových grafů.

3.5.1. Základní jednotky

Většina dotazů do databáze pracuje se třemi klíčovými typy dat – *litterae* („litery“), *slots* („sloty“), a *sets* („sety“).

a) Litera

Litera většinou odpovídá jednomu písmenu nebo digrafu, ale pro jednoduchost se jako litery označují i delší segmenty, které v databázi zabírají jeden *slot* (pozici ve slově). Aplikace většinou generuje seznamy liter v určitém kontextu, přičemž u každé litery udává frekvenční data. Seznam liter může představovat například inventář liter použitých v určitém textu, nebo seznam polygrafů obsahujících určité písmeno.

b) Slot

Slot je vymezen jako pozice ve slově, která může být vyplněna různými literami. Každý slot je v databázi jednoznačně určen kombinací číselného identifikátoru slova a pořadím pozice ve slově. Například první pozice ve slově FATHER/N je označena „4194-1“. Jelikož číselné označení nemá pro uživatele aplikace žádný význam, při načtení slotů do rozhraní se vždy uvádí příslušné slovo a jeho slovní druh.

c) Set

Set je množinou liter, které se alespoň jednou vyskytují v témž slotu. Například výše zmíněný první slot ve slově FATHER/N může obsahovat litery {*f*, *u*, *w*, *v*}, a ty tvoří jeden set. Na základě setů lze jednak hledat litery, které mají podobné využití, a jednak jednoduše sdružovat sloty (a tedy i slova), v nichž jsou střídavě používány určité litery. Podle toho, v jakém kontexu jsou sety liter zobrazeny, mohou ukazovat na hláskovou změnu, ale také na mísení forem přejatých z předlohy textu s formami vlastními písaři atd.

3.5.2. Vizualizace dat a statistické údaje

Vedle základních dotazů, které pouze vypisují a filtrují seznamy liter, setů a slotů, aplikace generuje data, která lze zobrazit na mapě, a data pro vizualizaci formou síťových grafů. Základem každé mapy je set liter (s frekvenčními daty) sestavený zvlášť pro každý text. Nejjednodušší jsou mapy, které zobrazují všechny realizace určitého segmentu v konkrétním slově (slotu). U složitějších map lze kombinovat více „příbuzných“ slotů, například všechny sloty, v nichž se může vyskytnout *f* a zároveň *u*.

Síťové grafy jsou generovány na základě seznamu liter a počtu slotů, v nichž se některá dvojice liter vyskytuje současně. Každý uzel odpovídá jedné liteře a síla spojnice mezi uzly je určena počtem slotů, kde se mohou vyskytnou obě propojené litery.

3.6. Experimentální funkce

Vedle základních funkcí, které měly být v rámci projektu dokončeny, nástroj zahrnuje dvě funkce, které zatím nebyly plně zprovozněny, a to odkazy do *Corpus of Narative Etymologies* (CoNE) a možnost zahrnout do databáze data z jiných zdrojů, například staroanglické varianty některých slov. Potřebné tabulky a komponenty uživatelského rozhraní jsou v aktuální verzi nástroje hotové, ale databáze zatím obsahuje jen velmi malý vzorek dat (řadově desítky položek).

3.7. Návrh postupu při vyhledávání

V závěru metodologické kapitoly je nastíněn doporučený přístup k vyhledávání v databázi, tzv. „zooming“, který spočívá ve zkoumání pomyslné sítě dat tvořené texty, jejich systémy zápisu i konkrétními formami slov a v kombinování širší perspektivy se zaměřením na detaily. Tento návrh vychází z předpokladu, že jednou z hlavních přednosti nástroje by měla být možnost relativně rychle dohledat a propojit související data, například neobvyklé užití určitého písmene v opisech téhož textu, regionální rozložení variant určitého slova atd.

4. Výsledky

4.1. Segmentace forem

Chybovost určená z náhodného vzorku automaticky zpracovaných slov byla 3.5 % chybně segmentovaných forem. Vzorek čítal celkem 254 slov a 2136 jejich variant. Vzhledem k rozložení chybných variant mezi větší počet slov se malý počet chyb (většinou jedna) může vyskytovat až u 15 % lexika. Chyba zpravidla neznamená, že jsou špatně zarovnané všechny segmenty ve slově.

Většinu forem z korpusu LAEME bylo možné jednoznačně rozdělit na jednotlivé segmenty, vyskytly se ovšem i formy poměrně náročné ke zpracování, u nichž bylo možné uvažovat o víc než jedné variantě segmentace a zarovnání. Většinu těchto forem bylo možné rozdělit do kategorií podle toho jaká problematická sekvence písmen se v nich vyskytovala, např. *sc-*, *-x-*, *-tht*, *-hg-*, *ih-* na začátku slova atd. Výsledky představují jednotlivé kategorie a uvádějí (arbitrárně zvolený) postup segmentace, jenž byl při tvorbě databáze použit.

Problematické formy zahrnují jednak poměrně rozšířené varianty (někdy spojené s konkrétní hláskovou změnou), např. *sc*, *-er* vs *-re*, a jednak velmi neobvyklé způsoby zápisu, které se objevují pouze v malé skupině textů, někdy pouze v jednom textu. Při zpracování druhé skupiny forem bylo v případě potřeby přihlédnuto k dalším variantám zápisu v příslušných textech.

Nejčastějším obecným problémem při segmentaci byl výskyt ojedinělých variant, které měly oproti obvyklejším zápisům téhož slova o jeden segment víc, např. *fehit* vs. *feit* (FEET/N), *hersump+nes* vs. *hersum+ness* (HÉRSUMNESS). Segmentace v takových případech vyžadovala buď přidání dalšího slotu, který by ovšem u drtivé většiny forem zůstal prázdný, nebo umístění předpokládaných dvou segmentů do téhož slotu. Ve většině případů byla nakonec zvolena druhá z těchto možností, protože z uživatelského hlediska se jevilo jako výhodnější, nevytvářet příliš nepředvídatelné sekvence slotů.

4.2. Uživatelské rozhraní

Jednotlivé funkce dostupné v rozhraní nástroje jsou rozebrány z uživatelského pohledu, jejich popise se tedy charakterem blíží návodu k používání nástroje, ale nezachází do přílišných podrobností, jako jsou podrobné instrukce na které tlačítko klikat apod. Soustředí se spíše na uspořádání a zobrazení dat a vazby mezi jednotlivými funkcemi.

a) Procházení textů

Přehled korpusových souborů (textů) z LAEME se v rozhraní zobrazuje jako tabulka, která uvádí základní údaje o každém souboru (zdrojový rukopis, obsažené texty, lokalizace atd.). Zároveň udává počet jedinečných slotů a celkový počet slotů v daném textu a obsahuje odkazy na popis souboru v Indexu zdrojů v rozhraní LAEME a odkaz na profil daného textu v rozhraní nástroje (viz níže). Zároveň spolu s každým textem zobrazuje odkazy na související texty (v Indexu zdrojů LAEME) tak, jak jsou v LAEME uvedeny (například na texty opsané stejným písařem, další opisy téhož díla apod.).

b) Vyhledávání v databázi

Vstupem pro vyhledávání je vždy jedno písmeno (litera), nebo kombinace písmen. Pomocí takto jednoduše formulovaných dotazů lze vyhledávat litery, sety nebo sloty. U vyhledávání liter se dále rozlišují dvě různé funkce. První funkce vypíše digrafy (polygrafy), které obsahují zadanou literu (pokud například uživatel zadá „h“, dostane seznam digrafů jako *h, hw, gh* atd.). Druhá funkce vypíše všechny litery, které se mohou vyskytnout ve stejném slotu jako zadaná litera (v případě *h* by šlo o *g*, *ȝ*, *c* a další).

Vyhledávání slotů vrací seznam slotů (a tedy slov), v nichž se vyskytuje zadaná litera, nebo kombinace liter. Například dotazem „h,g“ se získá seznam slov jako WEIGHT/N, pro které existují doložené formy *pi**h**te* i *pi**g**te*, LAW/N (*lahe* i *lag+es*) atd. Každý seznam slotů lze uložit do místní paměti pro budoucí použití.

Vyhledávání setů vypíše všechny sety, v nichž se objevuje zadaná litera (kombinace liter), spolu s údaji o frekvenci. V případě „h,g“ bude ve výsledcích vyhledávání samotný set {*h*, *g*} a k tomu větší sety, jichž je {*h*, *g*} podmnožinou, např. {*h*, *g, ȝ, ᵹ*}, {*h*, *gh*, *g*} atd. Ke každému setu lze ihned načíst seznam slotů, v nichž se set vyskytuje.

b) Mapy

Údaje o variantách zápisu v jednotlivých textech se na mapě zobrazují jako výsečové grafy, což umožňuje jednoduše znázornit výskyt různých forem v témž textu i poměr jejich četností. Velikost grafu závisí na celkovém počtu segmentů, na jejichž základě byl graf pro příslušný text vytvořen. Z mapy se lze jednoduše odkazovat na jednotlivé texty. Z vybrané mapy je možné vytvořit sérii map pro jednotlivé časové úseky.

c) Síťové grafy

Síťové grafy vizualizují četnost spoluvýskytu liter v témž slotu. Kromě grafů založených na všech literách a setech v databázi je možné vytvořit graf zobrazující záměnnost liter v určitém textu, nebo graf, který ukazuje korespondence mez literami ve dvou různých textech. V posledním případě se litery z jednotlivých textů zobrazují v různých barvách.

d) Profil textu

Obrazovka *text profile* („profil textu“) zobrazuje inventář liter pro určitý text, sety liter, které se v něm vyskytují (s odkazem na příslušné pozice ve slovech) a vlastní text. Tyto tři prvky jsou propojené, takže lze například klikáním na inventář zvýrazňovat slova, kde se vyskytuje určitá litera apod. Dále je možné vyhledávat a zvýrazňovat přímo v textu. Inventář liter na

červeno-zelené škále ukazuje, které litery jsou v daném textu více nebo méně časté v porovnání s ostatními texty v LAEME. Pro účely srovnávání textů rozhraní umožňuje zobrazit více profilů na téže obrazovce.

Praktické využití výše popsaných funkcí je předvedeno formou krátkých analýz, navržených vždy pro testování určité funkce. Jako předmět analýzy byla volena dobře prozkoumaná témata a známé texty, aby bylo případně možné snáze identifikovat chybné výstupy. Následující přehled uvádí testované funkce nástroje společně s tématy jednotlivých analýz:

a) Vyhledávání setů a filtrování: Vyhledávání se zaměřuje na sety liter, které zároveň obsahují *ch* a *k*, což z nich činí relevantní materiál pro zkoumání dobře známé kontextuální palatalizace [k], tedy jeho změny v [tʃ] (CoNE, VP).

b) Profil textu: Funkce je předvedena na profilu textu #246 (rukopis Cambridge, Trinity College B14.39, písař A), který v LAEME patří k textům s největší mírou variace. Analýza se zaměřuje na identifikaci nezvyklých forem zápisu a jejich rozmístění v textu.

c) Item (slot) lists („seznamy slotů“): Analýza volně navazuje na předchozí. Ukazuje, jak lze pomocí nástroje uložit seznam slov, v nichž se v textu #246 vyskytuje *-st* jako reflex staroanglického *-ht*, a následně porovnat výskyt podobných forem v díle dalších písařů, kteří se podíleli na příslušném rukopisu.

d) Síťový graf pro jednotlivý text: Tato část je pojata jako komentář ke dvěma grafům, z nichž první zobrazuje záměnnost liter v *Ormulu* (#301), což je text známý výjimečnou systematičností zápisu, a druhý záměnnost liter ve výše zmíněním textu #246.

e) Síťový graf – srovnání textů: Jako příklad grafu bylo zvoleno srovnání textů #1100 (Oxford, Jesus College, 29, *The Owl and the Nightingale* a další kratší texty) a #3 (London, British Library, Cotton Caligula A ix, *The Owl and the Nightingale*). Oba texty byly pravděpodobně opsány ze stejné předlohy (Laing, 2004). Graf ukazuje, které litery si v textech vzájemně odpovídají.

f) Mapy: Analýza volně navazuje na vyhledávání setů (a). Ukazuje různé varianty map zobrazujících distribuci *c*, *k* a *ch* a komentuje jejich výpovědní hodnotu.

g) Užití *x*: Poslední analýza ukazuje, jak lze kombinovat jednotlivé funkce nástroje na příkladu výskytů litery *x*.

Nedostatky odhalené během testování se týkají spíše detailů ve výpočtech nebo uživatelské přívětivosti rozhraní. Zcela zásadní chyby v samotném návrhu nalezeny nebyly. Patrně největší slabinou nástroje je, že zatím dokáže vyhledávat pouze v rámci jednotlivých morfémů. Například dotaz na sloty, v nichž je písmeno *g* následováno samohláskou, nenajde výskyty *g* na konci morfémů, následované koncovkou *-e*. Jako nejinovativnější prvek nástroje byly vyhodnoceny síťové grafy, jejichž přínos spočívá především v množnosti zobrazit poměrně velké množství vztahů mezi literami nebo texty v rámci jedné obrazovky.

V závěru hodnocení nástroje byly popsány možnosti jeho budoucího rozšíření. Patří mezi ně například přidání dalších staroanglických variant, či dat z *Linguistic Atlas of Late Medieval English* (LALME) nebo vylepšení statistických funkcí nástroje.

5. Závěry

Závěry práce hodnotí vytvořený nástroj ze tří různých hledisek. Nejprve shrnuje silné a slabé stránky jednotlivých funkcí. Poté hodnotí, do jaké míry se podařilo naplnit principy vytčené v metodologické kapitole, jako snadnější přístup k datům a možnost kombinovat širší a užší perspektivu. Třetím hlediskem je potom návaznost na dříve navržené metody výzkumu popsané v teoretické kapitole.

Testování ukázalo, že současná podoba nástroje je již použitelná k provádění analýz, ale stále by měla být považována za testovací, spíše než finální verzi. U většiny součástí nástroje byly odhaleny drobné nedostatky, zejména v oblasti kvantifikace dat. Dále se ukázalo, že nástroj vyžaduje zvýšenou opatrnost při interpretaci dat, což se nejnápadněji projevilo při testování map. Mapy založené na velkém počtu slotů totiž mohou zobrazovat varianty, které není vhodné srovnávat. Například výskyty určité litery ve zcela odlišných kontextech.

Do nástroje se podařilo začlenit funkce, které přímo reagují na všechny předem stanovené obecné požadavky, ale ne vždy bylo dosaženo přepokládané efektivity. Týká se to především funkcí usnadňujících identifikaci neobvyklých variant. Přístup k datům je většinou velmi rychlý, mimo jiné díky častému využívání odkazů na související data, například odkazů na mapy pro jednotlivé sloty nebo seznam forem určitého slova. Odkazy jsou zároveň prvkem, který výrazně napomohl propojení analýz na úrovni obecnějších vzorců se zkoumáním konkrétních forem v textech.

Pokud jde o návaznost na předchozí výzkum, navzdory odlišné struktuře dat je nástroj kompatibilní s modely systémů psaného jazyka použitými autory LAEME (Laing & Lass, 2013). Dalšími styčnými body s dříve navrženými metodami je možnost pomocí map

zobrazovat odlišné časové roviny a možnost vytvářet seznamy prvků (*item lists*), jejichž formy v různých textech nebo regionech mohou být následně srovnávány.

V závěru práce je krátce představena možná forma vizualizace systémů zápisu, kterou lze vytvořit bez interpretace zvuku. Tento model se inspiruje jednak možnostmi nově vytvořeného nástroje a jednak paralelami mezi psaným a mluveným jazykem.

6. Zdroje

Adams, M. (2015). "Introduction: Evidence and method in the historical study of English". In M. Adams, L. J. Brinton, & R. D. Fulk (eds), *Studies in the History of the English Language VI* (pp. 1–12). DE GRUYTER. https://doi.org/10.1515/9783110345957.1

Aitchison, J. (2002). *Language change-progress or decay*. Cambridge: Cambridge University Press.

Alcorn, R. (2016). TIMESS grant application. Unpublished.

Benskin, M. (1997). "Texts from an English Township in Late Mediaeval Ireland." *Collegium medievale: interdisciplinary journal of medieval research*, no. 1, pp. 91-174.

Black, M. (1999). "AB or Simply A? Reconsidering the Case for a Standard." *Neuphilologische Mitteilungen*, vol. 100, no. 2, pp. 155–174. *JSTOR*, www.jstor.org/stable/43346192. Accessed 31 May 2021.

Brook, G. L. (1972). "A Piece of Evidence for the Study of Middle English Spelling". *Neuphilologische Mitteilungen*, *73*(1/3), 25–28.

Browman, C. P., & Goldstein, L. (1995). "Dynamics and articulatory phonology". *Mind as Motion*, 175–193.

Calle-Martín, J. & Moreno-Olalla, D. (2012) "Body of Evidence*:* of Middle English Annotated Corpora and Dialect Atlases". In L. Wright and R. Dance (eds). *The Use and Development of Middle English. Proceedings of the Sixth International Conference on Middle English*, Cambridge 2008. Bern - Berlin - Bruxelles - Frankfurt am main - New York: Peter Lang AG. 17-34.

Cazal, Y., Parussa, G., Pignatelli, C., & Trachsler, R. (2003). "L'orthographe: Du manuscrit médiéval à la linguistique modern". *Médiévales*, *45*, 99–118. https://doi.org/10.4000/medievales.969

Chomsky, C. (1971). "Invented Spelling in the Open Classroom". *Word*, *27*(1–3), 499–518. https://doi.org/10.1080/00437956.1971.11435643

Christian, D. (1991). "The case for 'Big History'". *Journal of World History*, 2(2), 223-238. http://www.jstor.org/stable/20078501

Corrie, M. (2006). "Middle English: Dialects and Diversity". In L. Mugglestone, (ed.) *The Oxford History of English* [Online] (Updated ed.). Oxford: Oxford University Press. [http://site.ebrary.com/lib/cuni/Doc?id=10694614]

Daneš, F. (1966). "The Relation of Centre and Periphery as a language universal". In *Travaux linguistiques de Prague 2*: p. 9-22.

Denison, D., Bermúdez-Otero, R., McCully, C., & Moore, E. (eds). (2011). *Analysing Older English*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139022170**.**

Diemer, S. (2012a). "Orthographic annotation of Middle English Corpora". *Outposts of Historical Corpus Linguistics: From the Helsinki Corpus to a Proliferation of Resources. [Studies in Variation, Contacts and Change in English 10]*. Retrieved 7 December 2020, from https://www.academia.edu/2067888/Orthographic_annotation_of_Middle_English_Corpora

Diemer, S. (2012b). "Spelling variation in Middle English manuscripts: The case for an integrated corpus approach". In M. Markus, Y. Iyeiri, R. Heuberger, & E. Chamson (eds), *Studies in Corpus Linguistics* (Vol. 50, pp. 31–46). John Benjamins Publishing Company. https://doi.org/10.1075/scl.50.05die

Dossena, M., & Lass, R. (2004). *Methods and Data in English Historical Dialectology*. Peter Lang.

Emerson, R. H. (1997). “English Spelling and Its Relation to Sound”. *American Speech*, *72*(3), 260. https://doi.org/10.2307/455654

Faulkner, M. (2020). “Quantifying the Consistency of ‘Standard’ Old English Spelling”. *Transactions of the Philological Society*, *118*(1), 192–205. https://doi.org/10.1111/1467-968X.12182

Fisiak, J. (1986) *A short grammar of Middle English (6. wyd.)*. Warszawa: Państwowe Wydawnictwo Naukowe.

Heggarty, P., McMahon, A., & McMahon, R. (2005). “From phonetic similarity to dialect classification: A principled approach”. In N. Delbecque, J. van der Auwera, & D. Geeraerts (eds), *Perspectives on Variation*. De Gruyter Mouton. https://doi.org/10.1515/9783110909579.43

Hogg, R. (2006). “English in Britain”. In R. Hogg & D. Denison (eds), *A History of the English Language* (pp. 352–383). Cambridge University Press. https://doi.org/10.1017/CBO9780511791154.008

Hopper, P. (1987). “Emergent Grammar”. *Annual Meeting of the Berkeley Linguistics Society*, *13*, 139. https://doi.org/10.3765/bls.v13i0.1834

Horobin, S. (2010). *Studying the History of Early English*. Palgrave Macmillan.

Horobin, S. & Smith, J. (1999). “A database of Middle English spelling”. *Literary and Linguistic Computing*, *14*(3), 359–374. https://doi.org/10.1093/llc/14.3.359

Hudson, A. (1966), “Tradition and Innovation in Some Middle English Manuscripts”, *The Review of English Studies*, Vol. 17, No. 68, 359-372.

Kestemont, M. (2015). “A computational analysis of the scribal profiles in two of the oldest manuscripts of Hadewijch's letters”. *Scriptorium* 69. 159–175.

Kestemont, M. & Karina van Dalen-Oskam (2009). *“*Predicting the Past: Memory Based Copyist and Author Discrimination in Medieval Epics”. In *Proceedings of the twenty-first Benelux conference on artificial intelligence (BNAIC 2009)*. ResearchGate. Retrieved 9 June 2020, from https://www.researchgate.net/publication/237349804_Predicting_the_Past_Memory_Based_Copyist_and_Author_Discrimination_in_Medieval_Epics

Kohnen, Thomas (2014) *Textbooks in English Language and Linguistics (TELL), Volume 6 : Introduction to the History of English*. Frankfurt am Main, DEU: Peter Lang AG. ProQuest ebrary. Web. 10 February 2016.

Kopaczyk, J., Molineaux, B., Karaiskos, V., Alcorn, R., Los, B., & Maguire, W. (2018). “Towards a grapho-phonologically parsed corpus of medieval Scots: Database design and technical solutions”. *Corpora*, *13*(2), 255–269. https://doi.org/10.3366/cor.2018.0146

Kretzschmar, Jr, W. (2015). “Complex systems and the history of the English language”. In *Language and Complex Systems* (pp. 105-130). Cambridge: Cambridge University Press. doi:10.1017/CBO9781316179017.006 Laing, M. (1988).

Laing, M. (2015). “Some illustration of useful ways to compare and contrast the maps of early Middle English data in LAEME with those of late Middle English data in eLALME”. Unpublished. University of Edinburgh.

Laing, M (2004) “Multidimensionality: Time, Space and Stratigraphy in Historical Dialectology”. in M. Dossena & R. Lass (eds), *Methods and Data in English Historical Dialectology: Linguistic Insights 16*. Peter Lang Publishing Group, Bern, 49-96.

Laing, M. (1999). “Confusion "wrs” Confounded: Litteral Substitution Sets in Early Middle English Writing Systems”. *Neuphilologische Mitteilungen, 100(3*), 251-270. Retrieved May 31, 2021, from http://www.jstor.org/stable/43346203

Laing, M. (1993). *Catalogue of Sources for a Linguistic Atlas of Early Medieval English*, Cambridge: Brewer.

Laing, M. (1988). “Dialectal Analysis and Linguistically Composite Texts in Middle English”. *Speculum*, *63*(1), 83–103. https://doi.org/10.2307/2854323

Laing, M. (1992). “A Linguistic Atlas of Early Middle English: The Value of Texts surviving in more than one Version”. *History of Englishes: New Methods and Interpretations in Historical Linguistics: Topics in English Linguistics 10*, 566–581.

Laing, M. & Lass, R. (2013). *Introduction to the Linguistic Atlas of early Middle English.* [http://www.lel.ed.ac.uk/ihd/laeme2/laeme_intro_ch1.html; http://www.lel.ed.ac.uk/ihd/laeme2/laeme_intro_ch2.html]. Edinburgh: © The University of Edinburgh.

Laing, M., & Lass, R. (2009). “Shape-shifting, sound-change and the genesis of prodigal writing systems”. *English Language and Linguistics*, *13*(01), 1. https://doi.org/10.1017/S1360674308002840

Laing, M., & Lass, R. (2003). “Tales of the 1001 nists: The phonological implications of litteral substitution sets in some thirteenth-century South-West Midland texts”. *English Language and Linguistics*, *7*(2), 257–278. https://doi.org/10.1017/S1360674303001102

Laing, M., & Williamson, K. (2004). “The Archaeology of Medieval Texts”. In C. Kay & J. Smith (eds). *Categorization in the History of English*, p. 85- 145.

Lass, R. (2015). “Interpreting Alphabetic Orthographies”. In P. Honeybone & J. Salmons (eds), *The Oxford Handbook of Historical Phonology.* Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199232819.013.024

Lass, R. (2006). “The end of linear narrative? Reflections on the historiography of English”. In N. Love, (ed.), *Language and History: Integrationist Perspectives* (1st ed.). Routledge. https://doi.org/10.4324/9780203592588

Linell, P. (2019). “The Written Language Bias (WLB) in linguistics 40 years after”. *Language Sciences*, 76. Online [https://www.sciencedirect.com/science/article/pii/S0388000118303875] https://doi.org/10.1016/j.langsci.2019.05.003

McIntosh, A., Samuels, M. L., & Laing, M. (1989). *Middle English dialectology: Essays on some principles and problems*. Aberdeen University Press.

McMahon, A. M. S. (1994). *Understanding Language Change* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139166591

McMahon, A., Foulkes, P., & Tollfree, L. (1994). “Gestural Representation and Lexical Phonology”. *Phonology*, *11*(2), 277–316.

Millward, C. M. & Hayes, M. (2012), *A Biography of the English Language*, Wadsworth: Cengage Learning.

Minkova, D. (2015). “Establishing Phonemic Contrast in Written Sources”. In P. Honeybone & J. Salmons (eds), *The Oxford Handbook of Historical Phonology.* Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199232819.013.024

Minkova, D. (2013). *A Historical Phonology of English*. 441. Edinburgh: Edinburgh University Press.

Minkova, D. (2003). *Alliteration and sound change in early English*. Cambridge: Cambridge University Press.

Minkova, D., & Stockwell, R. P. (eds). (2002). *Studies in the history of the English language: A millennial perspective*. Mouton de Gruyter.

Minkova, D., & Stockwell, R. (1998). "Are Diphthongs Neglected?". *Publication of the American Dialect Society*, *80*(1), 34–49. https://doi.org/10.1215/-80-1-34

Mossé, F. (1968). *A handbook of Middle English (5th printing, corrected and augmented.)*. Baltimore: Johns Hopkins University Press.

Ogura, M. & William S-Y. Wang. (2004). "Dynamic Dialectology and Complex Adaptive System". In M. Dossena & R. Lass (eds), *Methods and Data in English Historical Dialectology: Linguistic Insights 16*. Peter Lang Publishing Group, Bern, 137-170.

Pandey, P. K. (1997). "Optionality, Lexicality and Sound Change". *Journal of Linguistics*, *33*(1), 91–130.

Read, C. (1971). "Pre-School Children's Knowledge of English Phonology". *Harvard Educational Review*, *41 (1)I*, p. 1-34.

Sebba, M. (Ed.). (2007). "Between language and dialect: Orthography in unstandardised and standardising vernaculars". In *Spelling and Society: The Culture and Politics of Orthography around the World* (pp. 102–131). Cambridge University Press. https://doi.org/10.1017/CBO9780511486739.010

Smith, J. (2020). "On Scriptae: Correlating Spelling and Script in Late Middle English". *Revista Canaria de Estudios Ingleses*, *80*, 13–27. https://doi.org/10.25145/j.recaesin.2020.80.02

Smith, J. J. (2007). *Sound change and the history of English*. Oxford University Press.

Smith, J. J., Black, M., & Horobin, S. (2002). "Towards a new history of Middle English spelling". In P. J. Lucas & A. M. Lucas (eds), *Middle English from Tongue to Text: Selected Papers from the Third International Conference on Middle English: Language and Text, Held at Dublin, Ireland, 1-4 July 1999* (No. 4; Issue 4, pp. 9–20). Peter Lang. http://eprints.gla.ac.uk/8961/

Stenroos, M. (2004). "Regional Dialects and Spelling Conventions in Late Middle English". In M. Dossena & R. Lass (eds), *Methods and Data in English Historical Dialectology: Linguistic Insights 16*. Peter Lang Publishing Group, Bern, 257-286.

Teresi, L., & University of Manchester (1998). *A computer-assisted analysis of spellings in two vernacular manuscripts of the transition period: MS Cambridge, Corpus Christi College 302 and MS London, British Library, Cotton Faustina A. ix.* Manchester: University of Manchester.

Browman, C. P., & Goldstein, L. (1992). "Articulatory Phonology: An Overview". *Phonetica*, *49*(3–4), 155–180. https://doi.org/10.1159/000261913

Tolkien, J. R. R. (1929). "Ancrene wisse and Hali meiðhad". *Essays and Studies, 14.*

Upward, C., & Davidson, G. (2011). *The history of English spelling* [Online]. Malden, Mass.: Wiley-Blackwell. [http://site.ebrary.com/lib/cuni/Doc?id=10483263]

Vachek, J. (1978) *A brief survey of the historical development of English* (4. ed.). Praha: Státní pedagogické nakladatelství.

Vachek (1966). "*On the Integration of the Peripheral Elements into the System of language*". *Travaux linguistiques de Prague 2*. p. 23—37.

Vachek, J. (1942). "Písmo a transkripce ve světle strukturálního jazykozpytu". *ČMF*, *28*, p. 403-408.

Vachek, J., & Luelsdorff, P. A. (1989). *Written language revisited*. Amsterdam: John Benjamins.

Vaňková, M. (2021, forthcoming). "Testing a Spelling Database Created from The Linguistic Atlas of Early Middle English".

Vaňková, M. (2016). *Localisation of version D of "The Poema Morale" based on "The Linguistic Atlas of Early Middle English*. Unpublished MA thesis.

Venezky, R. L. (2011). *The Structure of English Orthography*. Walter de Gruyter.

Wiggins, A. (2007). "Middle English Romance and the West Midlands". In Scase, W. (ed.), *Essays in manuscript geography: Vernacular manuscripts of the English West Midlands from the Conquest to the sixteenth century*. Brepols.

Williamson, K. (2004). "On Chronicity and Space(s) in Historical Dialectology". In M. Dossena & R. Lass (eds), *Methods and Data in English Historical Dialectology: Linguistic Insights 16*. Peter Lang Publishing Group, Bern, 97-136.

Wood, M. (1982). "Invented Spelling". *Language Arts*, 59(7), 707-717. Retrieved May 31, 2021, from http://www.jstor.org/stable/41405102

6.1. Online zdroje

A Linguistic Atlas of Older Scots, Phase 1: 1380-1500 [http://www.lel.ed.ac.uk/ihd/laos1/laos1.html] (Edinburgh: © 2008- The University of Edinburgh).

Angus McIntosh Centre for Historical Linguistics [http://www.amc.lel.ed.ac.uk/]

Benskin, M., Laing, M., Karaiskos, V. & K. Williamson. An Electronic Version of A Linguistic Atlas of Late Mediaeval English [http://www.lel.ed.ac.uk/ihd/elalme/elalme.html] (Edinburgh: © 2013- The Authors and The University of Edinburgh).

Bosworth, J. *An Anglo-Saxon Dictionary Online*., Ed. Thomas Northcote Toller and Others. Comp. Sean Christ and Ondřej Tichý. Faculty of Arts, Charles University in Prague, 21 Mar. 2010. Web. 24 Feb. 2015. [http://bosworth.ff.cuni.cz/]

Heggarty, Paul, Aviva Shimelman, Giovanni Abete, Cormac Anderson, Scott Sadowsky, Ludger Paschen, Warren Maguire, Lechoslaw Jocz, María José Aninao, Laura Wägerle, Darja Dërmaku-Appelganz, Ariel Pheula do Couto e Silva, Lewis C. Lawyer, Jan Michalsky, Ana Suelly Arruda Câmara Cabral, Mary Walworth, Ezequiel Koile, Jakob Runge & Hans-Jörg Bibiko. 2019. Sound Comparisons: Exploring Diversity in Phonetics across Language Families. (Available online at https://soundcomparisons.com, Accessed on 2021-05-31.)

Laing, M. (2013) *A Linguistic Atlas of early Middle English, 1150–1325*, Version 3.2 [http://www.lel.ed.ac.uk/ihd/laeme2/laeme2.html]. Edinburgh: © The University of Edinburgh.

Lass, R., Laing, M., Alcorn, R. & K. Williamson (2013). A Corpus of Narrative Etymologies from Proto-Old English to Early Middle English and accompanying Corpus of Changes, Version 1.1 [http://www.lel.ed.ac.uk/ihd/CoNE/CoNE.html]. Edinburgh: © The University of Edinburgh. https://soundcomparisons.com/

Treharne, E., Cambridge, Trinity College, B. 14. 52., in *The Production and Use of English Manuscripts 1060 to 1220*, edited by Orietta Da Rold, Takako Kato, Mary Swan and Elaine Treharne (University of Leicester, 2010)**,** accessed 8 March 2016. [http://www.le.ac.uk/english/em1060to1220/mss/EM.CTC.B.14.52.htm]