

# Standaryzacja tekstu w perspektywie historycznej Analiza zbitek leksykalnych\*

Joanna Kopaczyk

---

Uniwersytet im. Adama Mickiewicza

## 1. Pojęcie standaryzacji: standard języka a standard językowy

Standaryzacja to działanie prowadzące do zawężenia zakresu dostępnych opcji; to ujednolicanie, tworzenie wzorców: z szeregu opcji wybieramy jedną według ustalonych kryteriów i stosujemy ją jako model. Widać to w różnych towarzyszących nam standardach, chociażby w standardzie waluty, standardach miar i wag, standardach jakości, standardach bezpieczeństwa czy w standardach produkcji. Standaryzacja to zatem działanie utylitarne, zwiększające pewność, że w danej sytuacji można się spodziewać podobnych lub wręcz takich samych reakcji lub użycia takich samych narzędzi; że dana czynność będzie przebiegać łatwiej i że można ją będzie replikować i porównywać z innymi w usystematyzowany sposób. Takie standardowe działanie wykształca się powoli, na skutek doświadczeń i eliminowania wariantów nieskutecznych, trudnych w użyciu, mało wydajnych, czy po prostu nieznajdujących wystarczającej liczby propagatorów i użytkowników. W związku z tym, każde działanie standaryzujące jakiś obszar ludzkiej aktywności ma charakter diachroniczny i można je badać w perspektywie historycznej.

W kontekście komunikacji i języka standaryzacja toczy się na dwóch zależnych od siebie i przenikających się płaszczyznach. Z jednej strony można mówić o s t a n d a r y z a c j i j ę z y k a, czyli o wyłanianiu się jako modelu jakiegoś wariantu języka promowanego lub uprzywilejowanego społecznie. Taką właśnie definicję standaryzacji przyjmują James Milroy i Lesley Milroy (1991: 22–23), widząc w tym procesie cechy ideologii promującej pewien zestaw abstrakcyjnych norm. Z drugiej zaś strony, w samym języku obserwujemy s t a n d a r y -

---

\* Projekt badawczy finansowany ze środków na naukę w latach 2009–2011 w ramach grantu Ministerstwa Nauki i Szkolnictwa Wyższego nr N N104 014337.

z a c j ę j ę z y k o w ą, czyli ustalanie się norm użycia i wytwarzanie najczęstszych, najlepiej działających strategii na poszczególnych poziomach struktury języka: w pisowni, wymowie, odmianie, składni czy na płaszczyźnie leksyki, a nawet tekstu. Rozdzielenie tych dwóch aspektów standaryzacji pozwala uniknąć problemu, przed którym stanęli Milroy i Milroy (1991: 22) – jak mówić o standardzie, skoro np. jedynie ortografię można dziś uznać za w pełni ustandaryzowaną? Oddzielając konceptualnie standaryzację języka od standaryzacji językowej, możemy precyzyjniej analizować ten proces<sup>1</sup>.

Skupiając się na s t a n d a r d z i e j ę z y k a, można zauważyć, że dyskusja nad jego definicją toczy się najczęściej w nawiązaniu do polityki językowej, tożsamości kształtowanej przez język oraz wielojęzyczności (np. Joseph 2004; Millar 2005). W polskiej tradycji językoznawczej standard języka jest omawiany w kontekście językoznawstwa normatywnego, czyli preskrytywizmu (Miodek 1983; Bugajski 1993; Ruszkowski 1995) oraz szerzej pojmowanej kultury języka (Kurkowska 1971; Bartmiński, red. 2001; Markowski 2005). Peter Trudgill (1999) stara się zdefiniować standard języka poprzez kryteria negatywne, określając, czym standard n i e jest. Nie jest on więc ani równoznaczny z językiem, ani z akcentem, ani ze stylem czy rejestrem wypowiedzi. Nie jest to także jedynie zbiór preskryptywnych zasad. Rozważania te doprowadzają autora do wniosku, że standard to dialekt o szczególnych właściwościach społecznych, intelektualnych i kulturowych, nienależący do typowego kontinuum dialektalnego z powodu swych ponadregionalnych cech. Z reguły, dana cecha albo jest elementem standardu, albo należy do odmian niestandardowych, bez wariantów pośrednich.

## 2. Etapy standaryzacji

Standard języka wykształca się stopniowo, choć może też zostać ‘zaprojektowany’ sztucznie i wprowadzony niejako odgórnie i bez faz przejściowych<sup>2</sup>. W przypadku większości znanych standardów języka, do ich stopniowego

---

<sup>1</sup> W terminologii anglojęzycznej najczęściej oba rodzaje standaryzacji określa się jako *language standardisation*. W nawiązaniu do zaproponowanego rozróżnienia można by wprowadzić konsekwentną nomenklaturę, rozgraniczającą *language standardisation* (standaryzację języka) od *linguistic standardisation* (standaryzacji językowej).

<sup>2</sup> Por. sytuację językową w Norwegii, gdzie jedna z dwóch standardowych odmian języka, zwana Nynorsk, powstała w ramach projektu bazującego na wybranych cechach odmian lokalnych, wyabstrahowanych przez Ivara Aasena w połowie XIX wieku (Haugen 1976: 35–37, 405–409).

wykształcenia się doprowadziły scenariusze, nakreślone początkowo przez Einara I. Haugena (1966/1972). Ważnym przyczynkiem do wczesnej dyskusji o wykształcaniu się standardu języka była praca Heinza Klossa (1967), wprowadzająca rozróżnienie na języki *Abstand* i *Ausbau*. Teorie te zostały rozszerzone i uzupełnione przez takich badaczy standaryzacji, jak Milroy i Milroy (1985/1991), Leith (1983/1997) lub Millar (2005). Scenariusze te zawierają etapy standaryzacji języka, których wystąpienie jest konieczne, aby standaryzacja była kompletna.

Ogólnie, standaryzacja polega na represji dostępnego zróżnicowania w języku (Milroy i Milroy 1991: 8)<sup>3</sup>. Można ją podzielić na etapy, jednak ich kolejność jest w dużej mierze umowna; często też występują one jednocześnie. Pierwszy etap to *s e l e k c j a* – selekcja dialektu, który ma szansę zostać kodem ponadregionalnym i przejawia ku temu predyspozycje, np. ciesząc się prestiżem jako dialekt stolicy. Selekcja może się odbyć w sposób bardziej lub mniej świadomy, jednakże wybrany wariant zawsze będzie posiadał pewne cechy społeczne, które wyznaczają go do pełnienia roli standardu. Kolejnym niezbędnym etapem jest *a k c e p t a c j a*, czyli swego rodzaju ‘zgoda’ na używanie właśnie tego wyselekcjonowanego dialektu w funkcjach standardu. Towarzyszy temu geograficzne i społeczne rozprzestrzenianie się standardu, co Milroy i Milroy (1991: 27) określają mianem *d y f u z j i*. Standard języka powinien odpowiadać wszystkim możliwym zastosowaniom, czyli charakteryzować się maksymalnym zróżnicowaniem funkcji, często nazywanym *e l a b o r a c j ą* funkcji. Wykształcenie tej cechy to kolejny etap standaryzacji. Następnym w kolejności jest etap minimalnego zróżnicowania formy – standard powinien być jednolity i nie dopuszczać wariantów formalnych w takim stopniu, jak robią to odmiany niestandardowe. Aby pomóc użytkownikom w trzymaniu się standardowych form, w procesie standaryzacji następuje *k o d y f i k a c j a* języka. Kompilowane są słowniki, gramatyki i inne pomoce stabilizujące formę języka w jego standardowym wydaniu<sup>4</sup>. Marian Bugajski (1993: 89–90) zwraca uwagę, że już językoznawcy szkoły praskiej rozdzielili normę od kodyfikacji,

<sup>3</sup> Tłumaczenia wszystkich cytatów przez autorkę. W oryginale: „suppression of optional variability in language” (Milroy, Milroy 1991: 8).

<sup>4</sup> Nie wszystkie standaryzujące się języki lokalne muszą dokończyć ów proces. W renesansowej Szkocji standaryzacja języka szkockiego nie osiągnęła jeszcze ostatniego etapu – kodyfikacji, gdy sąsiedni język angielski zaczął wypierać go z wielu funkcji komunikacyjnych na fali sprzyjających okoliczności społeczno-politycznych, szczególnie po unii personalnej obu królestw w 1604 roku (patrz także McClure 1995; Bugaj 2004; Millar 2005: 89–91).

widząc tę ostatnią w funkcji „stabilizatora oraz regulatora funkcjonowania i rozwoju normy”<sup>5</sup>.

Procesowi standaryzacji języka towarzyszą szczególne warunki historyczno-społeczne, wymagające od danej społeczności sprawnej, jednolitej komunikacji, np. za pomocą dokumentów administracyjnych. Standaryzacji sprzyjają wynalazki, umożliwiające jednolitą komunikację na szerokim obszarze – takim kamieniem milowym w standaryzacji języków europejskich było rozwinięcie druku na masową skalę od połowy XV stulecia.

Te sprzyjające okoliczności zbiegły się ze zmianą myślenia o języku lokalnym, rodzimym, i jego roli w kształtowaniu komunikacji wewnątrz państwa. Peter Burke (2004: 81) zauważył, iż standaryzacja języków europejskich rozpoczęła się w momencie, gdy języki rodzime (*vernaculars*) zaczęły zwiększać swój wachlarz zastosowań i funkcji społecznych, czyli w dobie renesansu. O postępującej standaryzacji piśmiennictwa w odrodzeniowej Polsce tak pisał autor *Kształtowania się normy językowej w drukarniach polskich XVI wieku*:

[W] tej dopiero epoce dzięki sprzyjającym warunkom społeczno-kulturalnym poszerzyły się znacznie kręgi ludzi parających się piórem i powstała w pełnym tego słowa znaczenia literatura polska, [na którą złożyła się; przyp. J. K.] twórczość pisarzy należących do czterech co najmniej pokoleń i pochodzących z różnych stron Rzeczypospolitej. [...] W szczegółach różni autorzy mogli mieć jeszcze nieco inne poglądy na ‘lepszość’ i ‘gorszość’ rozmaitych sposobów mówienia czy pisania, ale już przez to samo, że zwracano uwagę na środki wyrazowe, że w nich przebiegano, dokonywało się z wolna ujednoclenie języka literackiego, usuwanie zeń form rażących, dialektycznych i przestarzałych (Borecki 1974: 5–6).

Ponadto, standaryzujące języki rodzime zaczęły być postrzegane jako element tożsamości, jednoczący wspólnoty narodowe i etniczne, a także działający jako przejaw władzy i siły (Millar 2005: 155–205).

Na tym tle należy przejść do drugiego, wewnątrzjęzykowego działania procesu standaryzacji, który wiąże się przede wszystkim z etapem minimalnego zróżnicowania formy oraz z kodyfikacją języka. Celem dalszej części rozdziału jest zwrócenie uwagi na standard językowy na różnych strukturalnych poziomach języka, szczególnie na poziomie tekstu. W tym kontekście przyjrzymy się zastosowaniu zbitek leksykalnych (*lexical bundles*) – narzędziu korpusowemu, przydatnemu w analizie standaryzacji na poziomie leksyki, składni i tekstu. Będzie również mowa o problemach pojawiających się przy gromadzeniu i analizowaniu korpusowych danych historycznych pod kątem standaryzacji.

---

<sup>5</sup> Należy jednak zaznaczyć, że prążanie widzieli normę w różnych odmianach języka, natomiast kodyfikację odnosili jedynie do języka literackiego (Bugajski 1993: 89).

### 3. Standaryzacja a struktura języka

Zwykło się postrzegać standaryzację językową na dwóch płaszczyznach: języka mówionego i języka pisanego (np. Haas, red. 1982; Milroy, Milroy 1991: 64–70; Bugajski 1993: 92–98). Mówi się tu o różnicach w podlegających standaryzacji poziomach struktury języka w obu jego odmianach. I tak, przy języku mówionym można badać standaryzację wymowy, sprowadzanie wielu wariantów z różnych akcentów do jednego, akceptowanego ponadregionalnie, zapisanego w słownikach wymowy. Można zwrócić uwagę na morfologię, na prawidłową odmianę lub poprawność procesu słowotwórczego. Można też skupić się na składni, choć tu zauważono dość wcześnie (Crystal, Davy 1969: 95–124), że składnia języka mówionego ze względu na swą urywaną strukturę nie podlega tym samym regułom co składnia w języku pisanim. Przechodząc więc do języka pisanego, podobnie, można mówić o standaryzacji ortografii (Bugajski 1993: 185), a także śledzić zmiany fonologiczne, sugerowane przez standaryzujące warianty pisowni (przykładowo, oboczność typu ‘pirwszy’ – ‘pierwszy’ w renesansowych drukach polskich prześledził Borecki 1974). W piśmie widoczna jest bardzo dobrze standaryzacja elementów, takich jak wspomniana wyżej morfologia czy składnia, która w języku pisanim jest poddana większej kontroli ze strony osoby piszącej (wybrane aspekty standaryzacji w składni języka mówionego i pisanego zob. w: Cheshire, Steina red. 1997).

Odpowiedź na pytanie, co rozumiemy pod pojęciem standardu językowego na różnych poziomach struktury, musi być złożona. Przykładowo, w tekstach średnioangielskich widać standaryzację fleksji rzeczownikowej i ustalenie morfemu {S} jako markera liczby mnogiej, jednak jego pisownia nie podlega tak szybkiej standaryzacji i jest przez długi czas zależna od czynników regionalnych (np. Blake, red. 1992; MacIntosh i in., red. 1986). Dlatego też Miriam A. Locher i Jürg Strässler (2008: 3) stwierdzili, że standard języka angielskiego „jest niezależny od wymowy, rejestru i stylu, które to mogły ustandaryzować się każde z osobna<sup>6</sup>. Podobnie, z osobna, aczkolwiek we wzajemnej zależności, standaryzują się wymowa i pisownia, derywacja i leksyka, fleksja i składnia. Takie rozumienie standaryzacji języka odpowiada wprowadzonemu wcześniej rozróżnieniu na standaryzację języka i standaryzację językową. Dzięki takiemu podejściu łatwiej jest oddzielić dyskusję nad standardem języka w kontekście pozajęzykowym (społecznym, historycznym, politycznym) od analizy standaryzacji na poszczególnych poziomach struktury języka.

<sup>6</sup> W oryginale: „Standard English is independent of pronunciation, register and style, which may have been standardised independently” (Locher, Strässler 2008: 3).

Do tego skomplikowanego obrazu należy dodać jeszcze jeden poziom struktury, zahaczający o zróżnicowane funkcje języka. Jest to poziom tekstu, a konkretniej typu tekstu i korelacji tegoż z prefabrykowanymi, standardowymi zwrotami i frazami oczekiwanymi w tym właśnie kontekście przez użytkownika. To, że użytkownik oczekuje pewnych konkretnych rozwiązań oraz to, że w przypadku stosowania innej konstrukcji niż oczekiwane tekst może nie spełnić swojej funkcji, pokazuje kolejną cechę standaryzacji – umowną poprawność. Dobór odpowiednich środków wyrazu i konstrukcji językowych wiąże się ze sprawnością językową, zdefiniowaną przez Andrzeja Markowskiego (2001: 103–104) jako umiejętność „przekazania przez osobę mówiącą (piszącą) w swoich tekstach tego, co się chce przekazać [...] w sposób najodpowiedniejszy w danej sytuacji komunikatywnej”. Jako przykład niech posłuży instrukcja obsługi jakiegoś urządzenia. Spodziewamy się po takim typie tekstu konkretnych konstrukcji – czasowników w trybie rozkazującym, konstrukcji warunkowych, braku narracji w pierwszej osobie. Gdyby takowa się pojawiła, mielibyśmy do czynienia nie z instrukcją obsługi, a np. z opowiadaniem o obsłudze urządzenia.

#### **4. Standaryzacja na poziomie tekstu: formuły, zwroty, zbitki leksykalne**

Na poziomie tekstu spodziewamy się więc języka standaryzowanego, formułicznego, zbudowanego z oczekiwanych elementów. Najlepiej widać to w dyskursie specjalistycznym, gdzie znajomość tych właśnie elementów świadczy o biegłości w danej dyscyplinie, czy to w tworzeniu przepisów kulinarnych, czy w języku dokumentów prawnych. Można tu przytoczyć jedną z definicji języka prawnego, który bywa definiowany jako „szczególny przejaw stylu urzędowego, charakteryzujący się m.in. dyrektywnością, bezosobowością, dążeniem do precyzji i standaryzacją” (Zieliński 2000: 50). Ową biegłość w konstruowaniu tekstu można zdobyć poprzez naukę modelu oraz typowych sformułowań czy wyrażań (np. Goźdz-Roszkowski 2004). Amy J. Devitt (2004: 137–163) zwróciła uwagę, że gatunek tekstu stanowi normę, pewien standard zapewniający podobieństwo pomiędzy komunikatami różnych nadawców i oczekiwaniami odbiorców. Użytkownicy, sprawnie posługujący się owym standardem, stanowią swego rodzaju ‘społeczność dyskursową’ (*discourse community*).

Skąd zatem możemy czerpać wiedzę na temat owych wyrażań, charakterystycznych dla danego typu dyskursu i ustandaryzowanych pod względem formy?

Do momentu możliwości skorzystania z obszernych korpusów, badacz zdany był w większości przypadków na osobiste doświadczenie, intuicję i wybiórczość. W taki właśnie intuicyjny sposób David Mellinkoff (1963) napisał jedno z najważniejszych opracowań dotyczących języka prawnego, niemniej jednak jego listy cech leksykalnych i wyrażeń formułicznych w języku prawnym jawią się dziś jako niepełne i nieco przypadkowe.

Współczesne językoznawstwo korpusowe daje lepsze narzędzia do systematycznych badań nad standaryzacją w języku. Metoda badania standaryzacji na poziomie gramatyki poprzez przeszukiwanie korpusów to efekt prac Douglasa Bibera i jego współpracowników. W pracy nad *Longman Grammar of Contemporary English*, jako jedną z metod zastosowali oni wyszukiwanie najczęściej występujących zbitek lub wiązek leksykalnych (*lexical bundles*) w korpusie współczesnej angielszczyzny (Biber i in. 1999). Jest to typowy przykład tzw. analizy indukcyjnej, *corpus-driven*, czyli będącej wynikiem badania samych danych korpusowych, a nie sprawdzeniem za pomocy korpusu uprzednio postawionej hipotezy (rozdzielenie to omawia m.in. Lewandowska-Tomaszczyk 2005: 133). Metoda ta polega na przepuszczeniu korpusu tekstów przez program komputerowy, który wyszukuje *n*-gramy, czyli przesuwające się o jedno słowo zbitki o zadanej długości. Dla przykładu, w tabeli 1 pokazano, jakie zbitki leksykalne o podanej długości *n*, czyli *n*-gramy, można wyodrębnić w pierwszym zdaniu tego akapitu.

**Tabela 1.** Przykładowe *n*-gramy

3-gramy	4-gramy	5-gramy
Współczesne językoznawstwo korpusowe	Współczesne językoznawstwo korpusowe daje	Współczesne językoznawstwo korpusowe daje lepsze
językoznawstwo korpusowe daje	językoznawstwo korpusowe daje lepsze	językoznawstwo korpusowe daje lepsze narzędzia
korpusowe daje lepsze	korpusowe daje lepsze narzędzia	korpusowe daje lepsze narzędzia do
daje lepsze narzędzia	daje lepsze narzędzia do	daje lepsze narzędzia do systematycznych

Automatyczna ekstrakcja zbitek dostarcza informacji o tym, co w korpusie faktycznie się znajduje. Nawet jeśli zbitka nie jest frazą, i tak zostanie wyciągnięta przez program, a co więcej, może okazać się istotnym elementem tworzenia danego tekstu bez względu na swoją strukturę. Przy odpowiednio obszernym korpusie, zbitki zaczynają się powtarzać. Często wychodzą one nieregularnie poza ramy frazy lub zdania, lub stanowią najczęściej występujące elementy wewnątrz fraz. Niektóre zbitki mają identyczną postać i wiele wystą-

pień – te będą najbardziej charakterystyczne dla danego typu dyskursu czy tekstu. Inne dopuszczają pewne odchylenia od dosłowności lub mają mniej wystąpień, niemniej jednak ich ponawiane użycie przez autora świadczy o wykształcaniu się standardu na poziomie tekstu – jego formy i organizacji – dla określonych funkcji i w określonym rejestrze.

Dla Bibera zbitki leksykalne stanowiły narzędzie pomocnicze w tworzeniu opisu i kodyfikowaniu języka angielskiego w postaci książkowej gramatyki. Badania na korpusach, pozwalające na wyodrębnienie najczęściej używanych sformułowań, mają szanse zrewolucjonizować myślenie o standardzie językowym. To użytkownik języka staje się twórcą standardu poprzez częstotliwość użycia, co stanowi odejście od preskrytywizmu, towarzyszącego standaryzacji przynajmniej od czasów oświecenia. W pewnym sensie, podejście empiryczne do standardu językowego i branie pod uwagę struktur faktycznie używanych stanowi powrót do pierwotnego, dynamicznego określania standardu na podstawie tego co najczęstsze, najchętniej wybierane przez użytkownika. Ważnym punktem metodologii Bibera jest więc powiązanie standardu z frekwencją wystąpień. Im częstsza konstrukcja, tym wyraźniej należy ona do standardu języka. Zbitki leksykalne pomogły w ustalaniu standardu konstrukcji gramatycznych – głównie na poziomie zdania, ale też na poziomie frazy, wewnątrz frazy i na styku fraz. Są one jedną z najbardziej obiektywnych metod, które pokazują, co tak naprawdę dzieje się w tekście i jakie sformułowania podlegają standaryzacji – powtarzane są w tej samej formie, w różnych tekstach tego samego typu.

Oprócz dość oczywistych różnic w treści, typy tekstów różnią się od siebie częstotliwością i długością zbitek leksykalnych. Michael Stubbs i Isabel Barth (2003: 62) w swych badaniach korpusowych nad trzema różnymi typami tekstu – tekstem naukowym, fikcją i beletrystką – stwierdzili, że powtarzalne ciągi słów stanowią „przewidywalną cechę różnych rodzajów tekstu” oraz dostarczają „dowodów na możliwość wyodrębnienia jednostek rutynowego użycia języka”<sup>7</sup>. Ponadto, okazało się, że można rozpoznać dany typ tekstu po długości zbitek, które w nim występują. Teksty naukowe wykazywały większą formułiczność i powtarzalność struktur niż literatura piękna oraz zawierały większą liczbę dłuższych zbitek. Najdłuższe powtarzające się sformułowania występowały w języku religijnym: w tłumaczeniach na język rodzimy tekstów biblijnych (np. w *Biblii Króla Jakuba* z początku XVII wieku, która stała się na kolejne stulecia obowiązującą wersją Biblii na Wyspach Brytyjskich) oraz w tekstach liturgicznych. Dyskurs prawny, na którym skupi się część empiryczna niniejszego rozdziału, również należy do niezwykle formulicznych odmian komunikacji.

---

<sup>7</sup> W oryginale: „predictable characteristic of different text types” oraz „evidence of units of routine language use” (Stubbs, Barth 2003: 62).



## 5. Zbitki leksykalne w korpusach historycznych

Zgodnie z wcześniejszą definicją, standaryzacja to proces, a więc ma ona swój początek i trwanie. Dzięki temu można szukać korzeni pewnych zjawisk towarzyszących standaryzacji w historii zewnątrz- oraz wewnątrzjęzykowej, a także śledzić rozwój trendów w czasie, przyjmując perspektywę diachroniczną. Elektroniczne wersje materiału historycznego, zawarte w korpusach językowych, dają się przeszukiwać automatycznie w śledzeniu powtarzalnych fragmentów, świadczących o standaryzacji. Przy podejściu diachronicznym można podzielić korpus na okresy czasowe i obserwować popularność danej zbitki.

**Tabela 2.** Porównanie metodologii doboru długości zbitki leksykalnych i punktów odcięcia w wybranych badaniach (1998–2010)

Autor (data)	Język: typ tekstów w korpusie	Liczba słów w korpusie	Długość badanych zbitek	Punkt odcięcia
Altenberg (1998)	współczesny angielski: mowa	0,5 mln	1- do 8-gramów	10
Biber i in. (1999)	współczesny angielski: konwersacja, fikcja, prasa, dyskurs akademicki	40 mln	3- do 6-gramów	10/5 per mln <sup>8</sup>
Biber i in. (2003)	współczesny angielski: konwersacja, dyskurs akademicki	7 mln + 5,3 mln	4- i 5-gramy	20 per 1mln
Goźdz-Roszkowski (2006a)	współczesny angielski: dyskurs prawny	0,3 mln	4-gramy	15
Culpeper, Kytö (2010)	wczesnonowoangielski: dialogi	0,5 mln	3-gramy	10
Kopaczyk (w przygotowaniu)	średnioszkocki: teksty prawne	0,6 mln	3- do 8-gramów	10

W tego typu metodologii pojawiają się dwa główne problemy: (1) jak długie mają być *n*-gramy, (2) od ilu wystąpień można mówić o formuliczności, czyli jak wyznaczyć tzw. ‘punkt odcięcia’ (*cut-off point*). Językoznawcy zgadzają się co do tego, że wskaźnikiem standaryzowania się struktur jest częstotliwość ich występowania (Hudson 1998: 141; Biber i in. 1999: 38–40; 2003: 74), jednak postawienie granicy frekwencyjnej między tym, co standardowe, a tym co nie,

<sup>8</sup> Biber i in. (1999: 992–993) przyjmują dwa różne punkty odcięcia: dla 3- i 4-gramów jest to 10 wystąpień na 1 mln słów, dla 5- i 6-gramów – 5 wystąpień na 1 mln słów.

jest bardzo trudne. Dużo zależy od rozmiaru korpusu i jego struktury, i w tym przypadku badacz zmuszony jest dokonywać decyzji intuicyjnych. Z zestawienia metodologii w tabeli 2 widać wyraźnie, że badacze w różny sposób zawężają materiał badawczy, przez co trudno jest przeprowadzić komparatywną analizę ilościową zbadanych zbitek, jednak nie utrudnia to zbytnio porównań jakościowych.

Badania dowiodły też, że długość powtarzalnych kombinacji słownych jest odwrotnie proporcjonalna do częstotliwości ich występowania (DeCock i in. 1998: 71). Krótsze zbitki są częste, natomiast dłuższe – o wiele rzadsze (por. tabela 3). Niemniej oba rodzaje dostarczają ciekawego materiału badawczego. Krótsze zbitki dają nam obraz formułczości na poziomie frazy, np. wyrażenia szeregowe (Kania, Tokarski, red. 1984), takie jak *stated and ordained*, lub wewnątrz frazy, np. *of the said*, natomiast zbitki dłuższe są bardziej osadzone w kontekście i pokazują, jakie funkcje tekstu były stosunkowo często wyrażane w postaci tych samych sformułowań. Jeśli można wykazać, że istnieje diachroniczna tendencja do wytwarzania takiego, a nie innego sposobu wyrażania danego znaczenia lub funkcji, lub też dystrybucja takich zbitek w korpusie nie skupia się w jednym regionie czy jednym dokumencie, wówczas możemy mówić o symptomach procesu standaryzacji językowej na poziomie tekstu w danej kulturze słowa pisanego.

**Tabela 3.** Najczęściej występujące zbitki w korpusie średnioszkockim (0,6 mln słów) (według: Kopaczyk, w przygotowaniu)

Długość zbitek	Najczęstsza zbitka	Liczba wystąpień	Liczba tekstów
3-gramy	<i>of ye said</i> 'wspomnianego wyżej'	2 157	773
4-gramy	<i>ye yhere off god</i> 'roku pańskiego'	699	556
5-gramy	<i>for ye wrangus haldin fra</i> 'za bezprawne zawłaszczenie od'	445	171
6-gramy	<i>chalans for ye wrangus haldin fra</i> 'oskarżenie za bezprawne zawłaszczenie od'	401	161
7-gramy	<i>for ye wrangus haldin fra him off</i> 'za bezprawne zawłaszczenie od niego'	325	156
8-gramy	<i>chalans for ye wrangus haldin fra him off</i> 'oskarżenie za bezprawne zawłaszczenie od niego'	297	146
		Liczba wszystkich tekstów w korpusie	1 818

Z przykładów najczęstszych zbitek danej długości w tabeli 3 widać również, że długość badanej zbitki będzie mieć wpływ na rezultaty czy to pod kątem struktury gramatycznej, czy też funkcji w tekście. Przykłady pochodzą z korpusów języka średnioszkockiego, używanych do badania standaryzacji tekstów administracyjnych i prawnych w szkockich miastach w latach 1380–1560 (szczegóły na temat korpusów zostały zamieszczone w dodatku).

Przy pracy z korpusami średniowiecznymi i renesansowymi pojawia się jeszcze problem natury technicznej – przecież pisownia nie została jeszcze ustandaryzowana, co przyjmujemy za pewnik, analizując korpusy współczesne. W związku z tym należy najpierw sprowadzić analizowane teksty do jednolitej formy zapisu, niejako „ustandaryzować” sztucznie pisownię, a dopiero potem wyciągać z tak przygotowanego korpusu zbitki leksykalne. Dla potrzeb projektów historycznych na Uniwersytecie Lancaster powstał program VARD (*Variant Detector*), służący do regularyzacji pisowni w tekstach<sup>9</sup>. Szkocki materiał korpusowy musiał jednak zostać ujednolicony w dużej mierze manualnie ze względu na niespotykane dużą wariantowość pisowni oraz różne praktyki transkrybowania i tagowania w trzech korpusach, na których opierają się badania.

Uzyskane za pomocą tej metody zbitki leksykalne dają nam możliwość prześledzenia procesu wytwarzania się stałych sformułowań w języku. Jako materiał badawczy najlepiej nadają się tu jednorodne korpusy specjalistyczne, czy to pod względem typu tekstu (np. listy), czy tematyki (np. język administracji), gdyż można wówczas wyeliminować czynniki związane z różnorodnością korpusu, a potencjalnie wpływające na dobór słów, konwencji stylistycznych i formuł.

## 6. Klasyfikacja zbitek leksykalnych pod względem struktury gramatycznej

Krótkie zbitki często stanowią niedokończony element frazy, a długie – wychodzą poza jej granice. Przez to tradycyjna metoda klasyfikacji gramatycznej się wymyka i w przypadku zbitek leksykalnych nie należy mówić o frazach, a raczej

---

<sup>9</sup> VARD został stworzony na potrzeby pracy z korpusem tekstów wczesnonowoangielskich i wykorzystuje listy wcześniej zdefiniowanych wariantów, jak również automatyczne rozpoznawanie wariantów za pomocą funkcji Levenshteina i algorytmu Soundex (Rayson i in. 2008). Obecnie dostępna jest ulepszona wersja VARD2.3 (zob. stronę Alistaira Barona <http://www.comp.lancs.ac.uk/~barona/ward2/>). Trwają również prace nad przystosowaniem VARDa do pracy z innymi językami.

o fragmentach fraz (*phrase fragments*). Można je sklasyfikować, szukając pierwszego elementu, mogącego stanowić człon nadrzędny frazy – wówczas uzyskujemy podział na fragmenty frazy rzeczownikowej, przyimkowej czy czasownikowej, a także np. fragmenty zdań podrzędnych lub pytań, tak jak w przypadku korpusu dialogów (Culpeper, Kytö 2010). Dla ilustracji procesu standaryzacji, widocznego w tekstach szkockich, wybrałam najdłuższe badane zbitki – 8-gramy. Fakt, że tak długie ciągi słów pojawiają się w identycznej postaci wiele razy w różnych tekstach świadczy o wykształcaniu się typowych struktur, przenoszących dane znaczenie. Formuliczość jest cechą charakterystyczną dyskursu prawnego i na powtarzających się zbitkach widać wyraźnie, jakiego rodzaju struktury przyczyniają się do owej formuliczości. Tabela 4 zawiera podsumowanie kategoryzacji gramatycznej dla 8-gramów wyodrębnionych w korpusie szkockim, a tabela 5 prezentuje najczęstsze 8-gramy w danej kategorii gramatycznej.

**Tabela 4.** Kategoryzacja gramatyczna zbitek w korpusie szkockim

8-gramy	Okazy >10	Liczba typów	Indeks typ/okaz	Liczba okazów najczęstszego typu
Fragmenty NP	3095	135	4,36	297
Fragmenty PP	1738	92	5,29	91
Fragmenty VP	1204	50	4,15	131
Fragmenty zdań podrzędnych	1078	44	4,08	126

**Tabela 5.** Najczęstsze 8-gramy w poszczególnych kategoriach gramatycznych

8-gramy	Najczęstsza zbitka	Liczba okazów najczęstszego typu
Fragmenty NP	<i>chalans for ye wrangus haldin fra him off</i> 'oskarżenie za bezprawne zawłaszczenie od niego'	297
Fragmenty PP	<i>til all men be thir present letteris me</i> 'wszystkim ludziom na mocy niniejszych pism ja'	91
Fragmenty VP	<i>kend til all men be thir present letteris</i> 'wiadome wszystkim ludziom na mocy niniejszych pism'	131
Fragmenty zdań podrzędnych	<i>be it kend til all men be thir</i> 'niech będzie wiadome wszystkim ludziom na mocy niniejszych'	126

Ze zgromadzonego materiału widać, że 8-gramy to najczęściej fragmenty fraz rzeczownikowych. Potencjalnie jednak długość badanej zbitki może mieć wpływ na to, jakiego rodzaju kategoria gramatyczna będzie dla niej najczęstsza, dlatego w krótszych zbitkach możemy się spodziewać innej prominentnej kategorii gramatycznej (np. najczęstszy 3-gram w korpusie, z liczbą wystąpień 2157 w 773 tekstach, to *off ye said* 'niniejszego', czyli średnioszkocka fraza przyimkowa). Jednym ze wskaźników zróżnicowania leksykalnego, a tym samym wskaźnikiem formułiczności i standaryzacji, jest indeks typ/okaz, czyli *type-to-token ratio*: im częstsza zbitka danego typu, tym niższy indeks typ/okaz, a tym samym wyższy stopień formułiczności – co zarazem prowadzi do standardowego wyrażania funkcji przenoszonej przez ową zbitkę (zob. tabela 6). Mimo że fragmenty fraz rzeczownikowych są najlichniesze, w tabeli 4 na pierwszy plan pod względem standaryzacji wysuwają się fragmenty fraz czasownikowych i zdań podrzędnych (które niejako stanowią swe dopełnienie, co widać w przykładach z tabeli 5).

## 7. Klasyfikacja zbitek leksykalnych pod względem funkcji

Jednak najciekawszym sposobem klasyfikacji zbitek leksykalnych jest kategoryzacja o charakterze funkcjonalnym. W tym nowym sposobie patrzenia na język poprzez struktury formułiczne, widać wyraźnie, jakie funkcje tekstu specjalistycznego mają tendencję do standaryzacji struktury. Ramy teoretyczne przy kategoryzacji funkcjonalnej są zwykle oparte na podziale zaproponowanym przez Michaela A. K. Hallidaya (1978) na: funkcję referencyjną (*referential function*), funkcję organizacyjną tekstu (*organisational function*) oraz funkcję interakcyjną (*interactional function*), która zawiera w sobie również funkcję nastawienia wobec przekazywanych treści (*stance*) (por. Moon 1998; Biber i in. 2003; 2004; Lewandowska-Tomaszczyk 2005; Goźdz-Roszkowski 2006a; 2006b; Culpeper, Kytö 2010). W przykładach zaczerpniętych z badań nad językiem szkockiej administracji (tabela 6) widać, która funkcja jest najbardziej formułiczna i ustandaryzowana w przekazie. Tabela 6 zawiera też ilość wystąpień najdłuższych badanych zbitok, 8-gramów, w zależności od pełnionej w dyskursie funkcji.

Tabela 6. 8-gramy według kategoryzacji funkcjonalnej

Funkcja referencyjna (4289 okazów)	Typ /okaz	Najczęstsza zbitka (8-gram)	Liczba jej wystąpień
Instrument prawny	<b>2,08</b>	<i>chalans for ye wrangus haldin fra him off</i> 'oskarżenie za bezprawne przywłaszczenie od niego (czegoś)'	<b>297</b>
Czas	4,63	<i>ye yhere off oure lord a thousande four</i> 'roku Pańskiego tysiąc cztery(sta)'	120
Miejsce	4,04	<i>haldin in ye tolbutth off ye samyn be</i> 'zebrany w ratuszu tegoż samego przez'	56
Przewinienie	6,13	<i>for ye wrangus haldin fra him off ane</i> 'za bezprawne przywłaszczenie od niego (czegoś)'	31
Uczestnik	6,98	<i>of ye balzeis off Peblis in that tyme</i> 'mieszczan z Peebles tego czasu'	27
Obiekt	7,46	<i>ye said erd and stan in-to ye handis</i> 'wspomnianą ziemię i kamień na ręce'	14
Funkcja interakcyjna i modalna (2026 okazów)			
Dyrektyw*	<b>2,81</b>	<i>kend til all men be thir present letteris</i> 'wiadome wszystkim ludziom na mocy niniejszych pism'	<b>131</b>
Reprezentatyw	3,72	<i>sutis callit ye curt effermit and absentis demyt</i> 'sprawy wniesione sąd zatwierdził i brak obecności odnotował'	60
Komisyw	5,68	<i>in witnes off ye quhilk thing to thir</i> 'poświadczając niniejszą rzecz przed tym'	38
Inwokacja	4,96	<i>James be ye grace off God king of</i> 'Jakub z łaski boskiej król (dopełniacz)'	27
Deklaratyw	7,65	<i>vp frely deliuerit purely and simply resignit fra</i> 'dobrowolnie dostarczył (i) bez przeszkód zrezygnował z'	23
Funkcja organizacyjna tekstu (816 okazów)			
Definiowanie	5,09	<i>as efter followis that is to say that</i> 'jak następuje co ma się rozumieć jako'	48
Intertekstualność	5,59	<i>and effect as efter followis that is to</i> 'i w efekcie jak następuje co ma (się rozumieć)'	24
Narracja	6,69	<i>day thom logan officer brocht zard and stan</i> 'dnia oficer Tom Logan przyniósł ziemię i kamień'	21

\* Polskie odpowiedniki poszczególnych aktów mowy przyjęto za Lipczukiem (1999).

Na pierwszy plan w ilości zbitek wychodzi funkcja referencyjna, osadzając dyskurs w czasie i przestrzeni, a także w realiach prawnych danego okresu i panujących zwyczajach. Najbardziej prominentną i najbardziej ustandaryzowaną w przekazie kategorią jest odwołanie wprost do instrumentu prawnego, czyli zarzutu, oskarżenia, ustawy czy innego nakazu, w jego postaci tekstowej. Poniekąd łączy się to z dyrektywnym aktem mowy – najbardziej formuliczną kategorią spośród zbitek interakcyjnych i modalnych. Znajduje to uzasadnienie w specyfice języka prawnego, który w warstwie ilokucyjnej najczęściej nakazuje lub zakazuje, odwołując się przy tym do poszczególnych instrumentów prawnych, stworzonych przecież także za pomocą języka.

Wśród kolejnych kategorii wykazujących znaczny stopień standaryzacji, można wymienić odniesienia do czasu i miejsca w funkcji referencyjnej. Teksty administracyjne i prawne, ujęte w korpus szkocki, są tak skonstruowane, aby kolejni użytkownicy mogli się odwołać do ich przekazu – stąd dążenie do jednolitego podawania daty i miejsca posiedzenia sądu grodzkiego czy parlamentu. Część tekstów dotyczy przekazywania gruntów w spadku lub ich sprzedaży, widać więc dbałość o sprecyzowanie, jaka nieruchomość podlega transakcji. W funkcji interakcyjnej zauważalna jest standaryzacja wyrażen związanych z działaniami sądu, rady czy parlamentu, takimi jak: zatwierdzanie, potwierdzanie, ustanawianie, konstatowanie i temu podobnymi. Zostały one ujęte w tabeli 6 jako reprezentatywy.

Pozostałe funkcje wykazują mniejszy stopień standaryzacji, jednak jest on i tak wyraźnie widoczny. W funkcji interakcyjnej zwracają uwagę wszelkiego rodzaju obietnice i zobowiązania, czyli komisyjne akty mowy, a także inwokacje do Boga mające poświadczyć autorytet władzy i prawa. Funkcja organizacyjna tekstu jest stosunkowo najmniej ustandaryzowana, jednakże definiowanie i doprecyzowywanie znaczenia wykazuje formułiczne tendencje, co również zgadza się z charakterystyką języka prawnego.

To, że materiał korpusowy obfituje w tak długie, ośmioelementowe zbitki leksykalne, które powtarzają się w niezmienionej formie tak dużą ilość razy, potwierdza empirycznie formułiczną strukturę języka prawnego w średnio-wiecznej i renesansowej Szkocji. Kolejnym krokiem w badaniach może być prześledzenie trendów diachronicznych w materiale, aby odpowiedzieć na pytanie, które zbitki wyłaniają się stopniowo jako te ustandaryzowane w swej formie. Przy funkcjonalnej analizie danych można i wręcz trzeba powrócić do pozajęzykowych czynników, sprzyjających standaryzacji poszczególnych funkcji tekstowych, a co za tym idzie, danego typu tekstu, osadzając go w kontekście historyczno-kulturowym.

## 8. Podsumowanie

Zatem wracamy do standaryzacji jako procesu wynikającego z potrzeb danego środowiska do ujednolicania i ułatwiania komunikacji. Badanie standaryzacji w tekstach historycznych powinno odbywać się przy użyciu obiektywnych danych korpusowych, tak jak dzieje się to przy pracy z materiałem współczesnym. Korpusy historyczne mają jednak swoją specyfikę. Badacz musi mieć na uwadze ograniczoną możliwość doboru reprezentatywnych tekstów, chociażby z tak prozaicznych przyczyn, jak pożary czy pleśń. Ponadto, pojawia się problem różnego tempa standaryzacji na różnych poziomach języka, chociażby w pisowni, co utrudnia badania za pomocą automatycznej ekstrakcji danych. Niemniej jednak, ważne jest, aby badać standaryzację w sposób obiektywny na dostępnym materiale. Proponowana metoda zbitek językowych wydaje się dostarczać satysfakcjonujących rezultatów, ponieważ odchodzi od patrzenia na język i na tekst w sposób atomistyczny. Cytując Jonathana Culpepera i Merję Kytö (2010: 141), „podejście atomistyczne może przeoczyć sporo z tego co się naprawdę dzieje”<sup>10</sup>. Poszczególne typy tekstów, zwłaszcza te ze sfery języka specjalistycznego, wytworzyły standardowe ramy konstrukcyjne oraz standardowe sposoby na przekazywanie określonych funkcji w dyskursie. Badając korpusy diachroniczne, można więc prześledzić tendencje i preferencje na przestrzeni czasu, a tym samym proces wytwarzania się standardu w tekście.

## Dodatek

**Tabela D1.** Szczegóły struktury korpusu szkockiego

Korpus	Długość korpusu (słowa)	Daty korpusu
ECOS	389,514	c. 1380–1500
HCOS	107,533	1450–1560
Wgt	77,640	1512–1534
Razem	574,687	c. 1380–1560

Objaśnienia: ECOS = Edinburgh Corpus of Older Scots, red. Keith Williamson (2008), HCOS = Helsinki Corpus of Older Scots, red. Anneli Meurman-Solin (1993), Wgt = Wigtown Burgh Court Book, niepublikowany transkrypt.

<sup>10</sup> „An atomistic approach is liable to miss much of what is going on” (Culpeper, Kytö 2010: 141).



## Bibliografia

- Altenberg, B. 1998. „On the phraseology of spoken English: The evidence of recurrent word-combinations”, [w:] A. Cowie (red.), 101–122.
- Baron, A. 2010. *VARD2 Homepage*, <http://www.comp.lancs.ac.uk/~barona/ward2/>
- Bartmiński, J. (red.), 2001. *Współczesny język polski*. Lublin: Wydawnictwo UMCS.
- Bartol, D., Duszak, A., Izdebski, H., Pierrel, J.-M. (red.), 2006. *Langue, Droit, Société. Language, Law, Society*. Nancy: Université de Nancy.
- Bex, T., Watts, R. J. (red.), 1999. *Standard English: The Widening Debate*. London: Routledge.
- Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E. 1999. *The Longman Grammar of Spoken and Written English*, London: Longman.
- Biber, D., Conrad, S., Cortes, V. 2003. „Lexical bundles in speech and writing: An initial taxonomy”, [w:] A. Wilson, P. Rayson, T. McEnery (red.), 71–92.
- Biber, D., Conrad, S., Cortes, V. 2004. „If you look at...”. Lexical bundles in university teaching and textbooks”, *Applied Linguistics* 25 (3), 371–405.
- Blake, N. (red.), 1992. *The Cambridge History of the English Language 2: 1066–1476*, Cambridge: Cambridge University Press.
- Borecki, M. 1974. *Kształtowanie się normy językowej w drukach polskich XVI wieku*, Wrocław–Warszawa: Ossolineum, Wydawnictwo PAN.
- Bugaj, J. 2004. „Middle Scots as an emerging standard and why it did not make it”, *Scottish Language* 23, 19–34.
- Bugajski, M. 1993. *Językoznawstwo normatywne*, Warszawa: PWN.
- Burke, P. 2004. *Languages and Communities in Early Modern Europe*, Cambridge: Cambridge University Press.
- Buttler, D., Kurkowska, H., Satkiewicz, H. (red.), 1971. *Kultura języka polskiego. Zagadnienia poprawności gramatycznej*, Warszawa: PWN.
- Cheshire, J., Stein, D. 1997. *Taming the Vernacular. From Dialect to Written Standard Language*, London–New York: Longman.
- Cowie, A. (red.), 1998. *Phraseology. Theory, Analysis and Applications*, Oxford: Clarendon Press.
- Crystal, D., Davy, D. 1969. *Investigating English Style*, London: Longman.
- Culpeper, J., Kytö, M. 2010. *Early Modern English Dialogues: Spoken Interaction as Writing*, Cambridge: Cambridge University Press.
- DeCock, S., Granger, S., Leech, G., McEnery, T. 1998. „An automated approach to the phrasicon of EFL learners”, [w:] S. Granger (red.), 67–79.
- Devitt, A. 2004. *Writing Genres*, Carbondale: Southern Illinois University Press.
- Gotti, M., Giannoni D. (red.), 2006. *New Trends in Specialized Discourse Analysis*, Frankfurt am Main: Peter Lang.
- Goźdz-Roszkowski, S. 2004. „Multi-word lexical units in legal genres – the pedagogic perspective”, [w:] B. Lewandowska-Tomaszczyk (red.), 401–416.

- Goźdz-Roszkowski, S. 2006a. „Frequent phraseology in contractual instruments. A corpus-based study”, [w:] M. Gotti, D. Giannoni (red.), 147–162.
- Goźdz-Roszkowski, S. 2006b. „Recurrent word combinations in judicial argumentation. A corpus-based study”, [w:] D. Bartol, A. Duszak, H. Izdebski, J.-M. Pierrel (red.), 139–152.
- Granger, S. (red.), 1998. *Learner English on Computer*, London: Longman.
- Haas, W. (red.), 1992. *Standard Languages: Spoken and Written*, Manchester: Manchester University Press.
- Halliday, M. A. K. 1978. *Language as a Social Semiotic. The Social Interpretation of Language and Meaning*, London: Edward Arnold.
- Haugen, E. [1966] 1972. „Dialect, language, nation”, [w:] J. Pride, J. Holmes (red.), 97–111.
- Haugen, E. 1976. *The Scandinavian Languages. An Introduction to Their History*, London: Faber and Faber.
- Hudson, J. 1998. *Perspectives on Fixedness: Applied and Theoretical*, Lund: Lund University Press.
- Joseph, J. E. 2004. *Language and Identity: National, Ethnic, Religious*, Basingstoke: Palgrave Macmillan.
- Kania, S., Tokarski, J. (red.), 1984. *Zarys leksykografii i leksykologii polskiej*, Warszawa: Wydawnictwa Szkolne i Pedagogiczne.
- Kloss, H. 1967. „‘Abstand languages’ and ‘Ausbau languages’”, *Anthropological Linguistics* 9, 29–41.
- Kopaczyk, J. W przygotowaniu. *Standardizing Legal Discourse. The Language of Scottish Burghs (1380–1560)*.
- Kurkowska, H. 1971. „Teoretyczne zagadnienia kultury języka”, [w:] D. Buttler, H. Kurkowska, H. Satkiewicz (red.), 11–79.
- Leith, D. [1982] 1997. *A Social History of English*, London: Routledge.
- Lewandowska-Tomaszczyk, B. 2005. „Analiza języka z zastosowaniem korpusu”, [w:] B. Lewandowska-Tomaszczyk (red.), 133–173.
- Lewandowska-Tomaszczyk, B. (red.), 2004. *Practical Applications in Language and Computers*, Frankfurt am Main: Peter Lang.
- Lewandowska-Tomaszczyk, B. (red.), 2005. *Podstawy językoznawstwa korpusowego*, Łódź: Wydawnictwo UŁ.
- Lewandowska-Tomaszczyk, B. (red.), 2008. *Corpus Linguistics, Computer Tools, and Applications – State of the Art. PALC 2007*, Frankfurt am Main: Peter Lang.
- Lipczuk, R. 1999. „O wielości i wieloznaczności terminów (na przykładzie klasyfikacji aktów mowy)”, *Lingua ac Communitas* 9, 169–176.
- Locher, M. A., Strässler, J. (red.), 2008. *Standards and Norms in the English Language*, Berlin–New York: Mouton de Gruyter.
- Locher, M. A., Strässler, J. 2008. „Introduction: Standards and norms”, [w:] M. A. Locher, J. Strässler (red.), 1–20.
- McClure, J. D. 1995. *Scots and Its Literature*, Amsterdam: Benjamins.
- McIntosh, A., Samuels, M. L., Benskin, M. (red.), 1986. *A Linguistic Atlas of Late Medieval English [LALME] 1–4*, Aberdeen: Aberdeen University Press.

- Markowski, A. 2005. *Kultura języka polskiego. Teoria. Zagadnienia leksykalne*, Warszawa: PWN.
- Mellinkoff, D. 1963. *The Language of the Law*, Boston–Toronto: Little, Brown and Company.
- Millar, R. McColl. 2005. *Language, Nation and Power*, Basingstoke: Palgrave Macmillan.
- Milroy, J., Milroy, L. [1985] 1991. *Authority in Language. Investigating Language Prescription and Standardisation*, London–New York: Routledge.
- Miodek, J. 1983. *Kultura języka w teorii i praktyce*, Wrocław: Wydawnictwo UWr.
- Moon, R. 1998. *Fixed Expressions and Idioms in English*, Oxford: Clarendon Press.
- Pisarek, W. (red.), 2000. *Polszczyzna 2000. Orędzie o stanie języka na przełomie tysiącleci*, Kraków: Ośrodek Badań Prasoznawczych UJ.
- Pride, J., Holmes J. (red.), 1972. *Sociolinguistics: Selected Readings*, Harmondsworth: Penguin.
- Rayson, P., Archer, D., Baron, A., Smith, N. 2008. „Travelling through time with corpus annotation software”, [w:] B. Lewandowska-Tomaszczyk (red.), 29–46.
- Ruszkowski, M. 1995. *Teoretyczne i praktyczne zagadnienia językoznawstwa normatywnego*, Kielce: Wydawnictwo WSP im. J. Kochanowskiego.
- Stubbs, M., Barth, I. 2003. „Using recurrent phrases as text-type discriminators: A quantitative method and some findings”, *Functions of Language* 10 (1), 61–104.
- Trudgill, P. 1999. „Standard English. What it isn't?”, [w:] T. Bex, R. J. Watts (red.), 117–128.
- Wilson, A., Rayson, P. McEnery, T. (red.) 2003. *Corpus linguistics by the Lune. A Festschrift for Geoffrey Leech*, Frankfurt am Main–New York: Peter Lang.
- Zieliński, M. 2000. „Języki prawne i prawnicze”, [w:] W. Pisarek (red.), 50–74.