

RUCH W JĘZYKU – JĘZYK W RUCHU

POD REDAKCJĄ
KAROLINY LISZYK-KUBINY
i MARCINA MACIOŁKA



UNIWERSYTET ŚLĄSKI



WYDAWNICTWO GNOME

K A T O W I C E 2 0 1 2

Recenzent
EWA SZKUDLAREK-ŚMIECHOWICZ

© Copyright 2012 by
Uniwersytet Śląski w Katowicach
Wszelkie prawa zastrzeżone

Redakcja i korekta
ZESPÓŁ

Projekt okładki i szaty graficznej
MAREK FRANCIK

Publikacja sfinansowana ze środków
UNIwersytetu Śląskiego w Katowicach

ISBN 978-83-63268-19-0

Złożono czcionką Times
Printed in UE

GNOME — Wydawnictwa Naukowe i Artystyczne
ul. Drzymały 18/6, 40-059 Katowice, Poland
tel.: 603370713 e-mail: wydawnictwognome@gmail.com

JAK ZAPISAĆ RUCH W JEZYKU, JEZYK W RUCHU – KORPUSY JEZYKÓW FONICZNYCH I MIGOWYCH

W artykule opisana zostanie jedna z najmłodszych gałęzi lingwistyki – lingwistyka korpusowa. Chociaż prac korpusowych można doszukiwać się już w średniowieczu (konkordancje tekstów biblijnych i klasycznych), to za narodziny językoznawstwa korpusowego uznaje się lata pięćdziesiąte XX wieku, kiedy zaczęły pojawiać się pierwsze prace korzystające z możliwości maszyn elektronicznych. Jedną z nich był projekt J.W. Ellisona wykorzystania komputera do porównywania różnych tekstów biblijnych. W 1957 r. opracował on pierwszą komputerową konkordancję biblijną.

Dalszy rozwój był niezwykle dynamiczny. Można mówić nie tyle o ewolucji tej dyscypliny, ile rewolucji. Rewolucje pojawiają się również u samych źródeł lingwistyki korpusowej – wymienia się trzy przełomowe nurty: strukturalizm, generatywizm (lingwistykę formalną) i lingwistykę informatyczną (Świdziński 2006). Wszystkie zmieniły zarówno podejście do języka jako przedmiotu analizy, jak i zaoferowały nowy zestaw narzędzi. Obecnie należałoby też dodać rewolucję numer cztery – Internet. W efekcie dostaliśmy możliwość łatwego tworzenia i publikowania wielomilionowych korpusów tekstów oraz zbiorów programów do ich przetwarzania. Pojawia się więc kwestia sporna, czy traktować lingwistykę korpusową jako osobny paradygmat naukowy, czy raczej jako metodologię w obrębie językoznawstwa.

Korpus w najszerszym rozumieniu jest dowolnym zbiorem tekstów wykorzystywanym do badań – przede wszystkim językoznawczych (Rudolf 2004: 19). Typ danych nie podlega żadnym ograniczeniom – mogą to być teksty odmiany pisanej, jak i mówionej języka. Ze względu na nieprecyzyjność takiej definicji (zgodnie z nią biblioteka może być uznana za korpus) powstawały próby zawężenia znaczenia tego terminu przez wprowadzenie kryteriów, które korpus powinien spełniać.

John Sinclair przedstawił następujące cechy korpusu (Sinclair 1996):

- ilość – korpus powinien zawierać „znaczną” liczbę słów, ze względu na nieustanny rozwój trudno podać konkretną wartość;
- jakość – dane powinny być *autentyczne*, tj. stanowić zbiór przykładów z jak najbardziej naturalnej komunikacji międzyludzkiej;
- prostota – materiał w formacie *plain text* w kodzie ASCII;

- dokumentacja – postulat pełnej dokumentacji danych, która trzymana jest w osobnym miejscu niż korpus.

Inne kryteria przyjęli Tony McEnery i Andrew Wilson (McEnery, Wilson 1996: 29–32):

- reprezentatywność – zebrany materiał powinien stanowić przekrój interesującego obszaru językowego w należytych proporcjach, innymi słowy korpus powinien być różnorodny językowo;
- wielkość – korpus powinien mieć skończoną liczbę słów, po osiągnięciu zaplanowanej wielkości rozbudowywanie zostaje wstrzymane, wyjątek stanowi korpus monitorujący (*monitor corpus*);
- format – powinien być użyty format danych, który może być przetwarzany maszynowo (obecnie domyślnie jest to format elektroniczny i przetwarzanie komputerowe);
- odniesienie standardowe – korpus stanowi standardowe odniesienie dla reprezentowanego języka, można więc na jego podstawie oceniać „standardowość” lub jej brak innych form wypowiedzi.

Mimo upływu czasu część tych założeń pozostaje nadal zasadna. Pozostałe zdezaktualizowały się wraz z postępem technologicznym i zwiększeniem mocy komputerów. Trzeba pamiętać o licznych możliwych odchyleniach od tych prototypowych wyznaczników.

Korpusy można klasyfikować ze względu na różne kryteria podziału. Jednym z ważniejszych typów jest *korpus oportunistyczny*, czyli zbiór dostępnych tekstów bez żadnych dodatkowych zasad doboru materiału. Jest to więc raczej przypadkowy zbiór tekstów. Przeciwnieństwem jest *korpus zrównoważony* powstały z tekstów dobranych w ustalonych wcześniej proporcjach (Rudolf 2004: 19). Inne zestawienia przedstawiono poniżej.

KORPUS OGÓLNY – KORPUS SPECJALISTYCZNY

Korpus ogólny stanowi próbę odwzorowania języka na poziomie podstawowym, powszechnym, z pominięciem specjalistycznych żargonów czy dialektów, jakie mogą w jego obszarze występować. Typowym przykładem korpusów ogólnych są korpusy referencyjne, takie jak np. Brytyjski Korpus Narodowy. Natomiast korpus specjalistyczny stanowi próbę odwzorowania języka używanego przez określoną grupę ludzi w określonych sytuacjach, np. języka prawniczego (Waliński 2005). Korpusy specjalistyczne mają największe zastosowanie w projektach leksykograficznych.

KORPUSY ZAWIERAJĄCE TEKSTY W CAŁOŚCI A KORPUSY „PRÓBKOWANE”

Z reguły umieszcza się teksty w całościowym ich brzmieniu, niektóre korpusy jednak zawierają jedynie ich fragmenty (korpus próbkowany), np. korpus Browna, w którym znajdują się fragmenty tekstów prasowych o długości 2000 słów, reszta artykułu została pominięta.

ZRÓWNOWAŻONE KORPUSY REFERENCYJNE ORAZ „MONITORUJĄCE” KORPUSY JĘZYKOWE

Korpus referencyjny ma stanowić jak najwierniejsze odwzorowanie naturalnych struktur w języku przy zachowaniu ich proporcji. Musi być więc zrównoważony, odmiany języka muszą być uwzględnione z odpowiednią wagą. Najczęściej więc są to korpusy zamknięte o bardzo dużej wielkości – im większy rozmiar, tym mniejsze znaczenie pojedynczych tekstów. Korpus monitorujący ma charakter otwarty, cały czas dodawane są do niego teksty. W ten sposób jest narzędziem monitorującym zmiany w języku; wykorzystuje się go w leksykografii.

KORPUSY JĘZYKA PISANEGO I MÓWIONEGO

Ze względów ekonomicznych bardziej opłacalne jest tworzenie korpusów języka pisanego – nie pojawiają się koszty nagrań i transkrypcji zebranego materiału. Na potrzeby niektórych badań tworzy się jednak korpusy tekstów mówionych.

KORPUSY JEDNO- I WIELOJĘZYCZNE

Zgodnie z nazwą korpus jednojęzyczny zawiera teksty jednego języka, podczas gdy teksty korpusu wielojęzycznego są w dwóch lub więcej językach. Korpusy wielojęzyczne ze względu na sposób łączenia zawartych w nich tekstów można podzielić na równoległe i porównywalne.

RÓWNOLEGŁE I PORÓWNYWALNE KORPUSY JĘZYKOWE

Korpus równoległy to taki, który zawiera teksty oraz ich przekłady na język obcy. Natomiast korpus porównywalny zawiera odrębne teksty rodzime w różnych językach, które zostały dobrane za pomocą ściśle określonych cech (tematyka, styl, medium publikacji, data powstania itd.), dzięki czemu są do siebie pod wieloma względami podobne.

KORPUSY ANOTOWANE I NIEANOTOWANE

Korpus nieanotowany zawiera „surowe” dane językowe, czysty tekst bez dodatkowych komentarzy. Przez anotację, czyli dodanie informacji (np. metadanych, opisu kategorii gramatycznych), uzyskujemy korpus anotowany.

KORPUSY SYNCHRONICZNE I DIACHRONICZNE

Korpus synchroniczny zawiera teksty współczesne, z jednego odcinka czasowego. Korpus diachroniczny to zbiór materiałów z różnych epok umożliwiający analizę zmian historycznych języka.

Lingwistyka korpusowa jest na tyle uniwersalnym narzędziem, że znajduje zastosowanie w wielu dziedzinach, w których prowadzi się badania dotyczące

języka naturalnego i różnych jego aspektów. Wykorzystywana jest w pracach czysto teoretycznych, jak i dotyczących praktycznego wykorzystywania języka. Liczba zastosowań korpusów wydaje się nieograniczona. Z korpusów korzysta się w badaniach leksykograficznych, dialektologicznych, diachronicznych, studiach nad gramatyką i semantyką. Obecne są również w analizie mowy i pracach nad przetwarzaniem języka naturalnego (McEnery, Wilson 1996: 11–27). Działem lingwistyki, który najpełniej korzysta z korpusów, jest lingwistyka migowa.

Ze względu na specyficzny charakter języka migowego, który w przeciwieństwie do języków fonicznych jest przestrzenny, a nie linearny, optymalny sposób rejestracji próbek empirycznych to zapis wideo. Wiąże się to z faktem, iż:

aparatu artykulacyjnego obejmuje wiele części ciała, od rąk i tułowia po głowę i twarz z jej wyrafinowaną mimiką. Artykulacja sprowadza się do tworzenia najrozmaitszych układów dynamicznych i konfiguracji owych części ciała i ich segmentów, także w sposób symultaniczny. Ruchy mają przy tym wiele parametrów, a na kształty znaków nakładają się często kształty elementów niejęzykowych (Świdziński, Mikulska 2003: 32).

Dość oczywisty wniosek sprowadza się do stwierdzenia, że najlepszym materiałem do badań języków migowych jest korpus ich tekstów.

Korpusowe metody badań można uznać za fundament lingwistyki migowej. Głusi oprócz „czystego” języka migowego posługują się różnymi hybrydami i pidginami migowo-fonicznymi. Dopiero porównanie pewnej ilości wypowiedzi pozwala wyciągnąć wnioski natury ogólnej. Niewielka jest również liczba rodzimych użytkowników, którzy znają na tyle dobrze lingwistyczny aparat pojęciowy, by dokonywać opisu swojego języka. Dotychczas *gros* prac na temat gramatyki języka migowego tworzone było przez słyszących badaczy, dla których migowy jest językiem drugim, a więc nie mogą oni polegać na swojej intuicji przy formułowaniu tez. Zawsze zwracano się o pomoc do informatorów, jednakże dopiero znaczna ich liczba przemawia za jakością badania. Ze względu na nieistnienie pisma głuchych należy każdorazowo sięgać po zapis wideo. Stworzenie według określonych zasad korpusu tekstów migowych dostarcza reprezentatywnej ilości materiału badawczego i umożliwia podjęcie wielu analiz, nie tylko gramatycznych.

Założeniem większości korpusów migowych jest rejestrowanie „naturalnego” migania. Nie jest to proste w realizacji. Osoby w obecności kamer zmieniają swój sposób wypowiedzania, starają się wyrażać „poprawniej”, w stylu oficjalnym. Dodatkowo mogą nie używać pewnych znaków, które uważają za regionalizmy albo za znaki „nieodpowiednie”. Kolejnym czynnikiem stresującym jest świadomość, że dane będą opublikowane i dostępne dla wszystkich zainteresowanych. Nie można przedstawić tekstu języka migowego oddzielnie od osoby, która go wykonuje, potrzebne jest zapewnienie sobie praw do wizerunku informatora, co pozwala na publikację nagrań. Twórcy korpusów przyznają, że istnieje duże niebezpieczeństwo „paradoksu obserwatora”.

Przed podobną trudnością stają twórcy korpusów mówionych, którym też zależy na naturalności dialogu. Badacze są zgodni co do tego, że ograniczenia metodologiczne, prawne i etyczne nie pozwalają jednak w pełni spełnić tego postulatu (Grucza 2007: 112). Wymienia się trzy sposoby pozyskiwania materiału jak najbliższego naturalnej komunikacji:

- ukryta obserwacja i nagrywanie; badacz z ukrycia rejestruje rozmowę; metoda, która pozwala uzyskać rzeczywiście naturalny dialog, ze względów prawnych i etycznych jest jednak niedopuszczalna;
- udział i jednoczesna obserwacja; badacz bierze udział w akcie komunikacyjnym, jednocześnie go rejestrując; największy zarzut stawiany tej metodzie dotyczy możliwości wypaczenia danych przez obecność badacza i jego aktywne uczestnictwo w rozmowie;
- elicytacja; metoda zbliżona do powyższej; zakłada obecność badacza w rozmowie i jednoczesne próby sterowania procesem komunikacji za pomocą odpowiednich materiałów stymulacyjnych; jest to najczęściej stosowany sposób przy tworzeniu korpusów migowych; w sesji nagraniowej uczestniczy głuchy moderator (niekoniecznie sam badacz), który prezentuje odpowiednio przygotowane zadania elicytacyjne i na tym właściwie kończy się jego rola.

Czasami stosuje się metodę podobną do ukrytej obserwacji – moderator wychodzi z pomieszczenia, pozostawiając włączone kamery, a uczestnicy są pozostawieni sami sobie bez konkretnego zadania. Są jednak poinformowani, że cały czas ich rozmowa będzie rejestrowana.

Spośród korpusów wyróżniają się zwłaszcza te, które zostały poddane „lematyzacji”, czyli każdej formie wyrazowej przypisana została jej reprezentacja słownikowa (leksem, lemat). Rozbudowuje to możliwości wyszukiwania i znacznie ułatwia tworzenie słowników. W kontekście odkrywania wiedzy lematyzacja to odróżnienie słów o innym znaczeniu od form jednego słowa, które przyjmuje inny zapis w zależności od funkcji w zdaniu.

W przypadku języków fonicznych jest to proces łatwy – wiadomo, jaką formę wybrać za podstawową, jedyną trudnością jest homonimia (np. napis *mamy* w zależności od kontekstu sprowadza się do bezokolicznika MIEĆ albo do formy mianownikowej MAMA). I tak: dla przymiotników formą słownikową jest mianownik liczby pojedynczej rodzaju męskiego; dla rzeczowników – mianownik liczby pojedynczej; czasowników – bezokolicznik itd. Dla języków o odmiennych gramatyce i kategoriach fleksyjnych wybór reprezentanta leksemu będzie odpowiednio inny.

Próba takiego opisu staje się jednak nad wyraz trudna dla języków wizualno-przestrzennych. Jako warunek konieczny do stworzenia korpusu jest jednak niezbędna. W punkcie wyjścia nie ma zbioru leksemów, do których można redukować napotkane formy. „Mig” oddaje się za pomocą glosowania. Tradycyjnie rozumiana glosa to:

wszelkie objaśnienie filologa, kopisty lub czytelnika umieszczone na marginesie bądź między wierszami tekstu. Najprostszą formą glosy było podanie synonimu w tym samym języku, co tekst podstawowy lub odpowiednika w innym języku używanym potocznie (Polański 1999).

W przypadku języków migowych można pokusić się o definicję: „głosowanie – próba opisu znaków języka migowego za pomocą słów języka fonicznego wzbogaconych o dodatkowe symbole (mimika, przestrzeń)”.

Języki wizualno-przestrzenne w najmniejszej jednostce, którą da się wydzielić (migu), przekazują czasami równocześnie wiele różnorodnych informacji gramatycznych i przy transkrypcji w języku fonicznym trzeba wykorzystać kilka leksemów. Sekwencja MYĆ_TWARZ realizowana w jednym znaku powinna być opisana przez hipotetyczny czasownik polski **myciotwarzować* (Świdziński, Mikulska 2003: 36). Potencjalnie może pojawić się unilateralny znak MYĆ_TWARZ_DOKŁADNIE niebędący złożeniem MYĆ_TWARZ + DOKŁADNIE, ale rozpowszechnionym w językach migowych zjawiskiem oddawania sposobu wykonywania czynności przez ruch. Abstrahując od analizy słowotwórczej, w pragmatyczną trudność zamienia się tutaj prosta kwestia, jak to zapisać. Jaką formę przyjąć za hasłową i czy w ogóle takowa istnieje?

Jak zauważył Trevor Johnston (Johnston 2008), przypisując glosę do danego znaku, dokonujemy lematyzacji – pomijamy jednostkowe warianty i niesystemowe modyfikacje, starając się uwydatnić lemat. Glosa, w literaturze angielskojęzycznej *ID-gloss od identifying gloss*, chociaż wyrażona w drugim języku, nie jest tłumaczeniem migu. Postulatem Trevora Johnstona jest odwrócenie kolejności znanej z korpusów języków fonicznych – zapisane uprzednio teksty poddajemy lematyzacji. Głosując korpus języka migowego, od razu przeskakujemy do lematyzacji, bez „pośredniego” stadium transkrypcji. Zaletą tego podejścia jest fakt, że nie tracimy żadnych informacji (szczegółowy opis może być dokonany później w dowolnej chwili), a bardzo szybko dostajemy korpus, który można przetwarzać maszynowo.

Na całym świecie istnieje wiele korpusów migowych. Oprócz dużych projektów realizowanych w Stanach Zjednoczonych i krajach europejskich powstają również nagrania w Chinach, Japonii czy w krajach arabskich. Różnią się one od siebie bardzo znacznie – począwszy od procedur zbierania danych, a skończywszy na anotowaniu. Należy pamiętać, że część z tych korpusów powstaje w konkretnym celu i podlega specjalnej obróbce, dlatego podobieństwa między nimi mogą być zerowe. Nie zostały wypracowane również uniwersalne standardy dotyczące sposobu zapisu nagranych materiałów i wytyczne jego opisywania. Każdy ośrodek naukowy używa swoich narzędzi i procedur. Dopiero od niedawna nagrania tworzone są z myślą o porównywaniu ich między różnymi krajami.

Jednakże sam proces tworzenia korpusu jest zazwyczaj bardzo podobny. Pierwszym krokiem jest znalezienie odpowiednich informatorów. Ustala się szereg kryteriów, które muszą być spełnione. Najważniejszym z nich jest znajo-

mość języka migowego. Paradoksalnie nie każdy głuchy umie migać i nie każdy głuchy z danego kraju miga w tym samym języku. Korpus powinien zawierać tekst migowy jak najbliższy „naturalnemu” migowemu. Jest to założenie, którego nie można spełnić w całości. Mimo to należy dążyć do takiego stanu przez dobór głuchych, którzy nabyli język odpowiednio wcześniej.

Nabór osób odbywa się albo przez kampanię ogłoszeniową (Internet, kluby dla głuchych itp.), albo przez specjalne osoby kontaktowe – głuchych, którzy znają środowisko i cieszą się w nim dobrą opinią. Przy wyborze dba się o zachowanie równych proporcji między liczbą kobiet i mężczyzn. Z reguły wprowadza się też kategorie wiekowe, np. 18–30, 30–45, 45–60 i 60+. Szczególny nacisk kładzie się na grupę głuchych z głuchych rodzin – są to tak zwani głusi dynastyczni, których język migowy odznacza się wyjątkowym bogactwem i niewielkim wpływem języka fonicznego.

Ochotnicy dobrani w pary ze względu na zbliżony wiek i pochodzenie są zapraszani w dogodnych terminach do studia nagraniowego. Przed sesją każda osoba wypełnia kwestionariusz zawierający pytania o poziom znajomości języka, historię edukacji, kontakty z innymi niesłyszącymi itp. Koniecznym warunkiem jest podpisanie zgody na upublicznienie swojego wizerunku. Nad przebiegiem nagrań czuwa moderator. Wybór osoby głuchej na to stanowisko eliminuje możliwość interferencji z językiem fonicznym – podczas całego nagrania w studiu nie pojawia się żadna osoba słyszająca. Zadaniem moderatora jest sprawne przeprowadzenie nagrania – kontrola czasu oraz upewnianie się, że uczestnik zrozumiał polecenie kolejnego zadania. Oprócz tego jego rola powinna być ograniczona do minimum.

W przypadku korpusów migowych bardzo ważne stają się kwestie techniczne, a zwłaszcza sposób rejestracji materiału (Johnston, Schembri 2006). Porównując studia w różnych ośrodkach badawczych, można stwierdzić, że pięć kamer to minimalny zestaw, ustawiony w następujący sposób: 2 kamery skierowane *en face* na informatorów, 2 kamery podsufitowe nagrywające informatorów z góry oraz 1 kamera, która obejmuje wszystkie postaci biorące udział w sesji. Powinien być to sprzęt rejestrujący obraz w jakości HD. Dzięki temu możliwe jest późniejsze zbliżanie np. twarzy w celu precyzyjnego odczytania mimiki. Opcjonalne jest poszerzenie zestawu o kamery stereoskopowe lub zestaw typu *motion capture*. Ważne dla jakości nagrań (a tym samym jakości wyprodukowanego tekstu) są też inne zagadnienia pozajęzykowe – odpowiednie oświetlenie, jednolite tło, właściwe zagospodarowanie przestrzeni.

Najważniejszą częścią tworzenia korpusu pozostaje elicytacja. Chociaż w poszczególnych projektach stosuje się odmiennie zestawy materiałów, to większość z nich się powtarza, często jedynie w lekko zmodyfikowanej formie – są to sprawdzone, dokładnie przetestowane polecenia, które dostarczają materiału badawczego obfitego w różnorodne konstrukcje językowe. Najczęściej procedura wywołująca obejmuje:

- Dyskusję. Temat do wyboru z listy bądź stymulowany np. filmem lub obrazkami. Czasami brak sprecyzowanego tematu, informatorzy mogą rozmawiać swobodnie o każdym zagadnieniu.
- Opowiadanie (*retelling*). Informator opowiada treść obejrzanego filmu bądź historyjki obrazkowej. Stosuje się filmy wykorzystywane również w badaniach języków fonicznych.
- Listę wyrazów. Informatorzy proszeni są o podanie odpowiedników migowych danego słowa.
- Zadanie negocjacyjne. Informatorom przedstawiony jest problem wymagający negocjacji i wspólnego podjęcia decyzji.

Dobór materiałów (filmów, obrazków itd.) do danego typu zadania zależy od celu, jaki ma spełniać korpus oraz od realiów kulturowych. Nie jest wskazane powtarzanie nagrań z tymi samymi osobami, dlatego elicytacja musi być należycie przygotowana ze zminimalizowanym ryzykiem wystąpienia jakichkolwiek problemów. Zebrany materiał jest anotowany w odpowiednich programach komputerowych, a następnie publikowany zgodnie z wcześniejszymi założeniami. Od pewnego czasu standardem staje się udostępnianie materiału w Internecie.

Z dużych zagranicznych korpusów migowych warto wymienić: korpus ECHO, korpus Australijskiego Języka Migowego (AUSLAN), korpus Brytyjskiego Języka Migowego (BSL), korpus Niemieckiego Języka Migowego (DGS), korpus Holenderskiego Języka Migowego (NGT) oraz projekt Dicta Sign. Wszystkie dostępne są w całości bądź częściowo w Internecie wraz ze szczegółową dokumentacją. Od 2010 r. w ramach Pracowni Lingwistyki Migowej przy Wydziale Polonistyki Uniwersytetu Warszawskiego powstaje korpus Polskiego Języka Migowego.

Korpus nie jest projektem zamkniętym. Prawie nigdy nie można uznać, że jest już nieprzydatny i odstawić go na półkę. Korpusu języka migowego te słowa dotyczą w jeszcze większym stopniu. Nieustanny rozwój technologiczny dostarcza coraz lepszych narzędzi, co owocuje bardziej precyzyjnymi opisami języków, w tym języka migowego, który nieustannie jest w ruchu.

BIBLIOGRAFIA

- Crasborn O., 2002: *Language content: sign languages*. Berlin.
- Gruca S., 2007: *O konieczności tworzenia korpusów tekstów specjalistycznych*. W: *W kręgu teorii i praktyki lingwistycznej*. Warszawa.
- Johnston T., 2008: *Creating a Corpus of Auslan within an Australian National Corpus*. Macquarie University.
- Johnston T., Schembri A., 2006: *Issues in the creation of a digital archive of a signed language*. Sydney.
- McEnery T., Wilson A., 1996: *Corpus Linguistics*. Edinburgh.

- Meyer Ch. F., 2002: *English Corpus Linguistics*. Cambridge.
- Polański K., red., 1999: *Encyklopedia językoznawstwa ogólnego*. Wrocław.
- Rudolf M., 2004: *Metody automatycznej analizy korpusów tekstów polskich*. Warszawa.
- Sinclair J.M., 1996: *EAGLES: Preliminary recommendations on Corpus Typology*. <http://www.ilc.cnr.it/EAGLES/corpusstyp/corpusstyp.html>.
- Świdziński M., 2006: *Lingwistyka korpusowa w Polsce – źródła, stan, perspektywy*. W: „LingVaria”, 1/2006, s. 23–34.
- Świdziński M., Mikulska D., 2003: *Reprezentacja linearna tekstu Polskiego Języka Migowego*. W: *Studia nad kompetencją językową i komunikacją niesłyszących*. Warszawa.
- Waliński J., 2005: *Typologia korpusów oraz warsztat informatyczny lingwistyki korpusowej*. W: *Podstawy językoznawstwa korpusowego*. Red. B. Lewandowska-Tomaszczyk. Łódź.