

ARTYKUŁY I ROZPRAWY

MAGDALENA DERWOJEDOWA* | UNIwersYTET WARSZAWSKI

Językoznawstwo komputerowe i inżynieria lingwistyczna po 1989 roku

Słowa kluczowe: językoznawstwo komputerowe, inżynieria lingwistyczna, przetwarzanie języka naturalnego.

Lingwistyka komputerowa (informatyczna) to dziedzina językoznawstwa, która wykorzystuje metody komputerowe (statystyczne lub algorytmiczne) do badania i modelowania języka naturalnego, natomiast inżynieria lingwistyczna jest nakierowana na tworzenie aplikacji (programów), które przetwarzają dane językowe (Mykowiecka 2007: 9). Rozgraniczenia te są w dużym stopniu umowne i opierają się na przeciwstawieniu podstaw teoretycznych zastosowaniom praktycznym; często wymiennie używa się też terminu ogólniejszego – przetwarzanie języka naturalnego: (Natural Language Processing (NLP), por. Przepiórkowski 2008: 3–6; Hajnicz 2011: 1).

Dystans dzielący polską technologię językową od rozwiązań, którymi mogą się cieszyć przetwarzający teksty w bardziej zaawansowanych w tej dziedzinie językach, ciągle istnieje, jednak wiele zaległości i braków zostało nadrobionych: występują zasoby korpusowe i leksykalne, a także zestawy narzędzi lingwistycznych. W wielu wypadkach nie była to prosta adaptacja, dostosowanie do fleksyjnej specyfiki polszczyzny wymagało bowiem niemałej inwencji, niektóre przedsięwzięcia to dzieła całkowicie oryginalne. Zasobami, które udało się uzupełnić są przede wszystkim korpusy polszczyzny (Hebal-Jezierska 2014). Zasobami tego rodzaju są też analizatory morfologiczne, słownik walencyjny czy wordnet, które wraz z innymi zasobami niekorpusowymi omówię w pierwszej części. Druga jest poświęcona narzędziom do przetwarzania języka naturalnego.

Słowosieć (<http://plwordnet.pwr.wroc.pl/wordnet/>) jest polskim wordnetem, czyli siecią wyrazów poklasyfikowanych ze względu na łączące je relacje znaczeniowe (Derwojedowa i in. 2007b; Maziarz i in. 2012). Relacja synonimii grupuje jednostki w tak zwane *synsety* (*sets of synonyms*, zbiory synonimów), natomiast hierarchię znaczeń tworzy hiperonimia. Siatka powiązań jednostek Słowosieci jest jednak bardziej złożona, uwzględnia inne relacje leksykalne oraz powiązania słowotwórcze; dodatkowo podano ekwiwalenty angielskie z wordnetu princetońskiego (Fellbaum 1998). Słowosieć jest rozległym zasobem leksykalnym, pod względem liczby haseł (170 tys.) największym obecnie na świecie. Ze względu na bogate polskie słowotwórstwo do konstruowania sieci zależności leksykalnych posłużono się wzorcami

* derwojed@uw.edu.pl

czeskiemu z wielojęzycznego programu EuroWordNet (por. Derwojedowa, Zawisławska 2007). Niewątpliwym osiągnięciem polskiego wordnetu jest zastosowanie wspomaganie maszynowego i eksperymenty w zakresie automatycznego wykrywania relacji znaczeniowych (Piasecki i in. 2009; Maziarz i in. 2012). Dzięki temu możliwe było szybkie powiększenie słownika, prace te stały się też zaczynem ciekawych eksperymentów w zakresie łączliwości wyrazów – na stronie projektu znajduje się także wyszukiwarka funkcji podobieństwa znaczeniowego podająca, jakie wyrazy występują najczęściej w tych samych kontekstach (Broda i in. 2008; Broda, Piasecki 2011). Słowosiec jest zasobem nieustannie rozwijanym przez zespół kierowany przez Macieja Piaseckiego, co pozwala żywić nadzieję, że nie tylko będzie rozległa, ale zostanie też rygorystycznie zweryfikowana. Drugim polskim wordnetem jest PolNet, wykorzystywany na przykład w systemie wspomaganie w sytuacjach kryzysowych (Vetulani i in. 2010a,b; Bouvry i in. 2012).

Zasobami leksykalnymi i narzędziami do ich opracowania są programy do tworzenia słowników nazw wielosegmentowych, odmieniania takich jednostek oraz ich wyszukiwania w tekście. Są to na przykład słownik nazw własnych, słownik nazw geograficznych i słownik gramatyczny frazeologizmów oraz narzędzia do wykrywania jednostek nazewniczych (*named entities*; Graliński i in. 2010; Marciniak, Rabięga-Wiśniewska 2010; Czerepowicka, Kosek 2011; Marciniak i in. 2011; Savary, Waszczuk 2012; Marcińczuk i in. 2013). Inną grupę stanowią słowniki walencyjne – niewielki (1 tys. jednostek), ale bardzo szczegółowy Marka Świdzińskiego (1994), zaadaptowany na potrzeby przetwarzania maszynowego *Słownik syntaktyczno-generatywny* Kazimierza Polańskiego (Polański (red.) 1980–1992; Grund 2000), słownik CEGLEX wykorzystany następnie w narzędziach GRAMLEX (Vetulani i in. 1998) oraz najnowszy z nich, a zarazem najbardziej rozbudowany, rozwijany od kilku lat Walenty (Przepiórkowski i in. 2014a). Słownik ten powstał na podstawie NKJP (Przepiórkowski i in. (red.) 2012) – zawarte w nim informacje to te wymagania, które można wydobyć, analizując dane korpusowe. Jest on ponadto zapisany w formacie umożliwiającym współpracę z innymi programami, np. wykorzystanie w parserach. Choć z rozstrzygnięciami autorów można by dyskutować (w szczególności co do zestawu notowanych fraz czy interpretacji pewnych zjawisk), to jest to największe i najszczegółowiej opracowane polskie źródło walencyjne, zawierające informacje powierzchniowo- i głębinowskładniowe, do tego dostępne dla ogółu zainteresowanych (Przepiórkowski 2004b; Przepiórkowski i in. 2014a,b; Hajnicz i in. 2015). W wypadku Walentego wykorzystano metody automatyczne do identyfikacji fraz zależnych; szczegółowo procedurę identyfikacji składników zdaniowych potencjalnie pełniących funkcje poszczególnych argumentów opisuje Elżbieta Hajnicz, rozważając też możliwe alternacje, powiązania struktury powierzchniowo- i głębinowskładniowej (mapowanie argumentów i fraz). Konstruuje ona odpowiednie testy oraz omawia przeprowadzone eksperymenty, które poprzedziły powstanie Walentego (Hajnicz 2011). Słownik jest rozbudowywany, uzupełniany i ulepszany.

Inna pod względem założeń i celów, choć również zawierająca informację o wymaganiach składniowych, jest nieduża, bo licząca około 200 haseł, baza RAMKI. Wykorzystuje ona teorię ram interpretacyjnych Ch.J. Fillmore'a (1982). Ten szczupły zasób charakteryzuje się olbrzymią homonimią jednostek, przyjęto bowiem, że o wydzieleniu osobnego hasła decyduje różnica w którymkolwiek z kryteriów: znaczenia, własności składniowych (realizowanych schematów),

własności ramowych (przywoływanych ram) oraz odpowiedniości ról głębokich i funkcji powierzchniowych. Każdej jednostce przyporządkowano ramę, zestaw ról, scenariusz, schemat składniowy oraz zilustrowano ją przykładami z korpusu, na których zostały oznaczone podstawowe funkcje powierzchniowe oraz elementy ramy (Derwojedowa i in. 2010).

Wśród programów do przetwarzania języka naturalnego można wyróżnić parsery składniowe, analizatory morfologiczne, dyzambiguousy i tagery oraz narzędzia bardziej wyspecjalizowane, np. do wydobywania z tekstu jednostek określonego typu. Szczególną pozycję w tej grupie zajmuje monografia M. Świdzińskiego zawierająca wyczerpujący formalny powierzchniowy opis składniowy polszczyzny (Świdziński 1992). Są w niej zdefiniowane w języku Prolog (Colmerauer 1978) reguły analizy struktur składniowych polszczyzny, od wypowiedzenia do konstrukcji najniższego rzędu. Autor sceptycznie odnosił się do możliwości jej implementacji. Tymczasem już po kilku latach podjęto takie próby (Bień 1996; Bień i in. 2000), a w 2004 roku powstał program Świgra (od: Świdzińskiego gramatyka) Marcina Wolińskiego. W latach 2009–2013 gramatyka była przez obu badaczy rozwijana, obecnie M. Woliński prowadzi prace nad znaczącym ulepszeniem parsera. Bezpośrednim wynikiem tych badań jest bank drzew rozbiorów gramatycznych Składnica (Świdziński, Woliński 2009, 2010; Woliński i in. 2011). Gramatyka doczekała się też istotnego rozszerzenia przez Macieja Ogrodniczuka (2006) o składnię konstrukcji liczebnikowych.

Zarówno kompletnością, jak i subtelnością opisu różnią się od gramatyki Świdzińskiego inne opisy formalne, jak choćby gramatyka Tomasza Obrębskiego (2002). Jest ona pierwszym formalnym ujęciem zależnościowym, podstawą działającego programu. I chociaż jej zakres jest ograniczony, to gramatyka ta zawiera kilka ciekawych rozwiązań, np. w zakresie opisu interpunkcji czy uzgodnień. Drzewa zależnościowe zawiera też wspomniana już Składnica. Powstały one w wyniku konwersji pierwotnych drzew składnikowych według reguł zaproponowanych przez Alinę Wróblewską (2012). Innym oryginalnym na gruncie polskim rozwiązaniem jest gramatyka opracowana w formalizmie HPSG, a następnie zaimplementowana przez zespół z IPI PAN. Opis ten nie jest kompletny, ale prezentuje główne cechy formalizmu i (niekiedy dyskusyjnych) rozstrzygnięć gramatyki (Przepiórkowski i in. 2001). Kompletna i wyczerpująca ma być rozwijana obecnie gramatyka wykorzystująca teorię LFG (Lexical Functional Grammar; Patejuk, Przepiórkowski 2015). Jej istotnym *novum* jest uwzględnienie semantyki, dotąd w analizie składniowej w zasadzie pomijanej. Analiza semantyczna jest też celem prac nad gramatyką kategoryalną języka polskiego (Marczewski 2012; Jaworski 2013). W gramatyce została podjęta próba rozwiązania problemów, jakich nastęrcza formalizm w odniesieniu do języka polskiego, a więc swobodnej linearyzacji, odróżnienia składników luźnych i wymaganych oraz bogatej morfologii.

Warunkiem wstępnym analizy składniowej jest analiza morfologiczna. Jest ona również niezbędna do bardziej zaawansowanego wyszukiwania w korpusach. Z wielu analizatorów powstałych na przestrzeni ostatnich 25 lat tylko nieliczne przetrwały próbę czasu i rozwoju, przede wszystkim aktualizacji słownika. Sytuacja analizy fleksyjnej jest przy tym wyjątkowa – została ona bowiem formalnie opisana i usystematyzowana wcześniej (Tokarski 1951, 1961a,b, 1962, 1963a,b,c,d, 1973, 1993; Gruszczyński 1989; Saloni 1992, 2000, 2001; Saloni i in. 2007/2012). Ze *Schematycznego indeksu a tergo polskich form wyrazowych* Jana Tokarskiego (1993) korzystały programy SAM-96

i udoskonalony SAM-98 Krzysztofa Szafrana (1997) oraz Morfeusz M. Wolińskiego (2006, 2014). Analizą *a fronte* posłużyli się twórcy analizatora Amor (Rabiega-Wiśniewska, Rudolf 2002; Rudolf 2004). Uniwersalny mechanizm opracowany przez Gábora Prószyńskiego z oryginalnym polskim słownikiem jest wykorzystywany przez analizator PoMor Roberta Wołosza (2005). Program ten wyróżnia się tym, że korzysta z dwóch podsłowników, z których jeden umożliwia włączenie do analizy form dawnych, regionalnych i przestarzałych. Na potrzeby tłumaczenia maszynowego powstał moduł analizy morfologicznej LEX (Graliński i in. 2013), a z programu do korekty pisowni – analizator Morfologik (Woliński i in. 2012).

Narzędzia te były i są w różnym stopniu używane w przetwarzaniu języka naturalnego, np. SAM znalazł zastosowanie w systemach tłumaczenia maszynowego (Fabian i in. 1999; Graliński 2000, 2001), natomiast Morfeusz jest podstawowym narzędziem analizy w zestawie narzędzi korpusowych i parserów IPI PAN, w tym kolejnych programów do znakowania korpusów – tagerów (Przepiórkowski 2004a, 2008; Przepiórkowski i in. (red.) 2012). Morfeusz umożliwia też użytkownikom samodzielną, dość wygodną analizę poszczególnych słów i całych tekstów, jest także wyposażony w moduł syntezy. Od pewnego czasu trwają prace nad rozszerzeniem słownika programu o możliwość analizowania tekstów dawniejszych (z lat 1600–1750 i 1830–1918; por. Derwojedowa i in. 2014; Gruszczyński i in. 2014).

Blisko tych programów lokują się programy do odmiany wyrażen wielosegmentowych. Praktycznym potrzebom szerszego odbiorcy (np. redaktora, ucznia czy studenta) odpowiadają automaty Zbigniewa Bronka do odmiany grup liczebnikowych, nazwisk i nazw geograficznych (<http://nlp.actaforte.pl/>), tworzące wszystkie poprawne formy odmiany. Dla każdej klasy wyrażen autor skonstruował niewielką gramatykę, wypełniając programami pewną lukę – analizatory zwykle koncentrują się na rzeczownikach pospolitych i wyrażeniach jednosegmentowych. W wypadku automatu składającego i rozkładającego grupy liczebnikowe powstało narzędzie o dużej szczegółowości – prezentowane są też warianty skrajnie rzadkie, zresztą jako takie oznaczone.

Programom do analizy fleksji nie towarzyszą w równej obfitości opisy słowotwórcze. Istnieje co prawda teoretyczny opis Joanny Rabiegi-Wiśniewskiej (2008), jednak powiązania słowotwórcze znajdują miejsce co najwyżej w zasobach leksykalnych, np. w Słowsieci i SGJP3 (Derwojedowa i in. 2007a,b; Broda i in. 2008; Piasecki i in. 2008; Maziarz i in. 2012; Saloni i in. 2015) oraz – w niewielkim zakresie – w Morfeuszu i PoMorze (Wołosz 2005; Woliński 2014). Można więc mieć nadzieję, że i w tej mierze nastąpi znaczący postęp.

Dane otrzymane z analizy morfologicznej mogą być wieloznaczne ze względu na homonię. Zadaniem dyzambiguatora jest wybranie interpretacji poprawnej w danym kontekście. Problem ten jest rozwiązywany przez adaptację do specyfiki polszczyzny rozwiązań dla innych języków. Stosowane metody można podzielić na metody bazujące na regułach opisujących ograniczenia danego języka i na metody statystyczne. Oba te sposoby bywają też łączone. Ograniczenia są wykrywane automatycznie w tekście lub formułowane przez eksperta na podstawie jego wiedzy językowej. Przykładem reguły może być sąsiedztwo prawostronne wyrazu *przeszło* – jako operator adnumeratywny poprzedza ono liczebnik (Grochowski 1984), jako czasownik zajmuje pozycje ze znacznie mniejszymi ograniczeniami.

Metodami statystycznymi posłużył się Łukasz Dębowski w pionierskim na gruncie polskim trigramowym dyzambiguatorze opracowanym na potrzeby Korpusu IPI (Dębowski 2003). Do oznakowania finalnej wersji tego korpusu użyto natomiast regułowego tagera TAKIPI (Piasecki, Godlewski 2006), który wykorzystuje drzewa decyzyjne oraz reguły ujednoznaczniania. W NKJP zastosowano jeszcze inne podejście, korpus został oznakowany za pomocą programu Pantera, czyli adaptowanego przez Szymona Acedańskiego tagera Brilla. W module PELCRA NKJP zastosowano ujednoznacznianie wykorzystujące maksimum entropii, program został opracowany przez Piotra Pęczika (Acedański 2010; Przepiórkowski i in. (red.) 2012). Powstały też programy wykorzystujące inne techniki: warstwowe, pamięciowe czy hybrydowe (Radziszewski, Śniatowski 2011; Radziszewski 2012, 2013; Kobylński 2014). Wyniki działania poszczególnych tagerów wahają się zależnie od tego, jakie czynniki są brane pod uwagę (poprawna kwalifikacja jednostek znanych i nieznanymi, błędne przyporządkowanie między klasami wyrazów, błędne przyporządkowanie form synkretycznych itd.), w podstawowym zakresie są jednak dość zbliżone (Mykowiecka 2007; Kobylński 2014). Mimo takiej mnogości programów samodzielne ujednoznacznienie zebranego korpusu jest zadaniem trudnym technicznie – uruchomienie programów wymaga znacznych umiejętności. Innym problemem są zestawy znaczników (tagsety) nie zawsze odpowiadające przyzwyczajeniom czy potrzebom użytkownika. Program w jakimś stopniu konfigurowalny, przyjaźniejszy użytkownikom, a być może też podający informację o homonimiach niemożliwych do rozwiązania, jest ciągle wyzwaniem.

W ostatnim dwudziestopięcioleciu powstały też interesujące narzędzia o węższej specjalizacji, np. analizator tekstów medycznych (Marciniak, Mykowiecka 2014), systemy tłumaczeniowe (Jassem 2001) czy system wspomagający działania w sytuacjach kryzysowych (Vetulani i in. 2010b). Wśród osiągnięć tego typu warto wymienić program odpowiadający na pytania w języku naturalnym na podstawie polskiej Wikipedii (Przybyła 2015) czy programy do wydobycia informacji z tekstów sumeryjskich (Jaworski 2008) oraz prace w zakresie automatycznej lub wspomaganej komputerowo klasyfikacji tekstów (Borchmann i in. 2015). Znacznie większym przedsięwzięciem jest program Thetos, tłumaczący komunikaty polskie na język migany (system językowo-migowy, SJM). Zawiera on moduł analizy i syntezy morfologicznej, analizuje teksty również semantycznie (Szał, Suszczańska 2002). Innym wyspecjalizowanym programem przewidzianym do szerszego nawet niż Thetos stosowania jest sprawdzający trudność tekstów Jasnopis. Pod uwagę brane są cechy leksykalne (wyrazy trudne), morfologiczne i składniowe, a stopień trudności jest wyliczany z uwzględnieniem różnych metod statystycznych. Ocena trudności jest przedstawiana użytkownikowi na obrazowej skali jako liczba klas, które trzeba ukończyć, żeby dany tekst zrozumieć. Wskazywane są też te cechy tekstu, które utrudniają zrozumienie (Gruszczyński, Ogrodniczuk (red.) 2015). Do badań szczegółowych, choć bardziej nastawionych na zastosowanie w innych aplikacjach niż jako pomoc użytkownikowi, należy automatyczne wykrywanie koreferencji (Ogrodniczuk i in. 2015). W szerokim rozumieniu jest to nie tylko powierzchniowa anafora, ale również inne typy odniesień – metaforyczne czy wyrażone za pomocą relacji leksykalnych – hipo- i hiperonimii czy meronimii). Wymaga to nie tylko uchwycenia językowo specyficznych wykładników nawiązania, czyli

analizy składniowej, ale też rozstrzygania o identyczności referentów, a także odróżnienia koreferencji od innych zjawisk tekstowych, np. współlektensji, czyli przynależności do tego samego pola znaczeniowego.

Uproszczona analiza konstrukcji składniowych, a więc taka, w której zakłada się, że wynikiem nie muszą być wszystkie możliwe interpretacje lub interpretacje te mogą być niedookreślone, to tzw. płytkie parsowanie. Jest ono stosowane na przykład do wyszukiwania słów kluczowych, wydobywania informacji z tekstów, automatycznego odpowiadania na pytania, a także automatycznego wykrywania kolokacji czy schematów (ram) walencyjnych. Gramatyką napisaną z myślą o płytkim parsowaniu polszczyzny jest Spejd, a wśród przykładów zastosowania jest ujednoznacznianie składniowe, rozpoznawanie wyrażen wielosegmentowych, nazw własnych, słów pisanych z dywizem, interpretowanie skrótów i skrótowców (w tym pisowni z dywizem i końcówką fleksyjną), liczb zapisanych cyframi jako mających wartości kategorii gramatycznych, dat. Gramatyka definiuje proste konstrukcje składniowe, takie jak rozmaite grupy nominalne czy zdaniowe (Przepiórkowski 2008).

Na koniec trzeba wspomnieć o badaniach ilościowych cech tekstu na użytek filologii, czyli o stylometrii, oraz opracowanym w zespole filologów i informatyków pakiecie *stylo*, który ułatwia takie badania, można go na przykład wykorzystać do analiz wersologicznych czy afiliacyjnych (Pawłowski i in. 2010; Eder 2011; Eder i in. 2013).

We współczesnych badaniach językoznawczych metody i wyniki językoznawstwa komputerowego są wykorzystywane w innych obszarach badań, takich jak wydobywanie danych z korpusów i ich opracowywanie na potrzeby badań różnego typu, a także w leksykografii.

Bibliografia

- Acedański S. 2010: *A morphosyntactic Brill tagger for inflectional languages*, [w:] H. Loftsson, E. Rögnvaldsson, S. Helgadóttir (red.), *Advances in Natural Language Processing*, 7th International Conference on NLP, IceTAL 2010, Reykjavik, Iceland, August 16–18, Springer, s. 3–14.
- Bień J. 1996: *Komputerowa weryfikacja formalnej gramatyki Świdzińskiego*, „Biuletyn Polskiego Towarzystwa Językoznawczego” LII, s. 147–164.
- Bień J., Szafran K., Woliński M. 2000: *Experimental parsers of Polish*, [w:] 3. Europäische Konferenz „Formale Beschreibung slavischer Sprachen, Leipzig 1999”, „Linguistische Arbeitsberichte” 75, Institut für Linguistik, Universität Leipzig, Leipzig, s. 185–190.
- Borchmann Ł., Graliński F., Jaworski R., Wierchoń P. 2015: *A semi-automatic method for thematic classification of documents in a large text corpus*, [w:] F. Mambriini, M. Passarotti, C. Sporleder (red.), *Proceedings of the Workshop on Corpus-Based Research in the Humanities (CRH)*, Instytut Podstaw Informatyki PAN, Warszawa, s. 13–21.
- Bouvy P., Kłopotek M.A., Leprevost F., Marciniak M., Mykowiecka A., Rybiński H. (red.) 2012: *Security and Intelligent Information Systems: International Joint Conference, SIIS 2011, Warsaw, Poland, June 13–14, 2011, Revised Selected Papers*, Lecture Notes in Computer Science, vol. 7053, Springer (online: <http://www.springer.com>).
- Broda B., Derwojedowa M., Piasecki M., Szpakowicz S. 2008: *Corpus-based semantic relatedness for the construction of Polish Word Net*, [w:] *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, European Language Resources Association (ELRA), Marrakech, Morocco, s. 1800–1807 (online: http://www.lrec-conf.org/proceedings/lrec2008/pdf/459_paper.pdf, dostęp: 20 października 2015).
- Broda B., Piasecki M. 2011: *Parallel, massive processing in SuperMatrix – a general tool for distributional semantic analysis of corpora*, „International Journal of Data Mining, Modelling and Management” 5 (1), s. 1–19 (online: doi: 10.1504/IJMMMM.2013.051924).

- Calzolari N., Choukri K., Declerck T., Loftsson H., Maegaard B., Mariani J., Moreno A., Odijk J., Piperidis S. (red.) 2014: *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014*, ELRA, Reykjavík, Iceland (online: <http://www.lrec-conf.org/proceedings/lrec2014/index.html>, dostęp: 30 października 2015).
- Colmerauer A. 1978: *Metamorphosis grammars*, [w:] L. Bolc (red.), *Natural Language Communications with Computers*, Lecture Notes in Computer Science, vol. 63, Springer, Berlin, s. 133–189.
- Czerepowicka M., Kosek I. 2011: *Problemy opisu związków frazeologicznych w formalizmie „Multifleks” (na przykładzie rodzaju wyrażen frazeologicznych)*, [w:] M. Bańko, D. Kopicńska (red.), *Różne formy, różne treści*, Wydział Polonistyki Uniwersytetu Warszawskiego, Warszawa, s. 117–126.
- Derwojedowa M., Kieraś W., Skowrońska D., Wołosz R. 2014: *Współczesne narzędzia leksykograficzne a analiza tekstów dawniejszych*, „Polonica” XXXIV s. 21–27.
- Derwojedowa M., Linde-Usiekniewicz J., Zawisławska M. 2010: *Rygorystyczna aplikacja metodologii kognitywno-interpretacyjnej*, Elipsa, Warszawa.
- Derwojedowa M., Piasecki M., Szpakowicz S., Zawisławska M. 2007a: *Polish WordNet on a shoestring*, [w:] G. Rehm, A. Witt, L. Lemnitzer (red.), *Proceedings of Biannual Conference of the Society for Computational Linguistics and Language Technology, Tübingen, April 11–13 2007*, Universität Tübingen, Tübingen, s. 169–178.
- Derwojedowa M., Piasecki M., Szpakowicz S., Zawisławska M. 2007b: *Relacje w polskim WordNecie*, Raporty Seria PREPRINTY 1, Instytut Informatyki Stosowanej Politechniki Wrocławskiej, Wrocław.
- Derwojedowa M., Zawisławska M. 2007: *Relacje leksykalne w polskiej i czeskiej bazie WordNet (Lexical relations in Polish and Czech WordNet)*, [w:] Z. Rudnik-Karwatowa (red.), *Z Polskich Studiów Slawistycznych. Prace na XVI Międzynarodowy Kongres Slawistów, Ochryda 9–16 września 2008, XI Językoznawstwo*, Instytut Slawistyki PAN, Warszawa, s. 15–23.
- Dębowski Ł. 2003: *A reconfigurable stochastic tagger for languages with complex tag structure*, [w:] *Proceedings of the Workshop on Morphological Processing of Slavic Languages. 10th Conference of the European Chapter of Association for Computational Linguistics*, Budapest, s. 63–70.
- Eder M. 2011: *Style-markers in authorship attribution. A cross-language study of the authorial fingerprint*, „Studies in Polish Linguistics” 6, s. 99–114.
- Eder M., Kestemont M., Rybicki J. 2013: *Stylometry with R: A suite of tools*, [w:] *Digital Humanities 2013. Conference Abstracts*, University of Nebraska-Lincoln, Lincoln, s. 487–489 (online: <http://dh2013.unl.edu/abstracts/>, dostęp: 20 października 2015).
- Fabian P., Migas A., Suszczyńska N. 1999: *Zastosowania analizy morfologicznej i składniowej w procesie rozpoznawania mowy*, „Speech and Language Technology. Technologia Mowy i Języka”, t. 3, s. 155–165.
- Fellbaum C. 1998: *WordNet. An Electronic Lexical Database*, MIT Press, Cambridge, Mass.
- Fillmore C.J. 1982: *Frame semantics*, [w:] *Linguistics in the Morning Calm*, The Linguistic Society of Korea, Seoul, s. 110–137.
- Graliński F. 2000: *Hasłowanie korpusu polskich tekstów informatycznych (1,2 mln słów) – raport*, „Speech and Language Technology. Technologia Mowy i Języka”, t. 4, cz. 2, s. 147–153.
- Graliński F. 2001: *Komputerowa analiza tekstu polskiego. Zastosowanie w systemie POLENG*, praca magisterska, Wydział Matematyki i Informatyki Uniwersytetu im. Adama Mickiewicza, Poznań.
- Graliński F., Jassem K., Junczys-Dowmunt M. 2013: *PSI-Toolkit: A natural language processing pipeline*, [w:] A. Przepiórkowski, M. Piasecki, K. Jassem, P. Fuglewicz (red.), *Computational Linguistics*, Studies in Computational Intelligence, vol. 458, Springer, Berlin–Heidelberg, s. 27–39 (online: doi: 10.1007/978-3-642-34399-5_2).
- Graliński F., Savary A., Czerepowicka M., Makowiecki F. 2010: *Computational Lexicography of Multi-Word Units: How Efficient Can It Be?*, [w:] E. Laporte, P. Nakov, C. Ramisch, A. Villavicencio (red.), *Proceedings of Multiword Expressions: from Theory to Applications (MWE 2010)*, Workshop at COLING 2010, August 28, Beijing, China.
- Grochowski M. 1984: *Projekt klasyfikacji syntaktycznej polskich leksemów nieodmiennych*, „Polonica” X, s. 73–97.
- Grund D. 2000: *Komputerowa implementacja słownika syntaktyczno-generatywnego czasowników polskich*, „Studia Informatica” 21 (3), s. 243–256.
- Gruszczyński W. 1989: *Fleksja rzeczowników pospolitych we współczesnej polszczyźnie pisanej*, Wydawnictwa Uniwersytetu Warszawskiego, Warszawa.
- Gruszczyński W., Adamiec D., Ogrodniczuk M. 2014: *Elektroniczny korpus tekstów polskich z XVII i XVIII w. (do 1772 r.) – prezentacja projektu badawczego*, „Polonica” XXXIII, s. 309–316.
- Gruszczyński W., Ogrodniczuk M. (red.) 2015: *Jasnopis, czyli mierzenie zrozumiałości polskich tekstów użytkowych*, Wydawnictwo ASPRA-JR, Warszawa.

- Hajnicz E. 2011: *Automatyczne tworzenie semantycznych słowników walencyjnych*, Problemy Współczesnej Nauki. Teoria i Zastosowania. Inżynieria Lingwistyczna, Akademicka Oficyna Wydawnicza EXIT, Warszawa.
- Hajnicz E., Nitoń B., Patejuk A., Przepiórkowski A., Woliński M. 2015: *Internetowy słownik walencyjny języka polskiego oparty na danych korpusowych*, „Prace Filologiczne” LXV, s. 95–110.
- Hebal-Jeziarska M. 2014: *Praktyczny przewodnik po korpusach języków słowiańskich*, Wydział Polonistyki Uniwersytetu Warszawskiego, Warszawa (online: http://www.iszip.uw.edu.pl/files/pdf/praktyczny_przewodnik.pdf, dostęp: 30 maja 2016).
- Jassem K. 2001: *POLENG: System tłumaczenia automatycznego z języka polskiego na język angielski*, [w:] S. Gawiejnowicz (red.), *Wybrane zastosowania współczesnej informatyki*, Poznańskie Towarzystwo Przyjaciół Nauk, Poznań.
- Jaworski W. 2008: *Contents modelling of Neo-Sumerian Ur III economic text corpus*, [w:] D. Scott, H. Uszkoreit (red.), *Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008)*, Coling 2008 Organizing Committee, Manchester, UK, s. 369–376.
- Jaworski W. 2013: *Przetwarzanie języka polskiego za pomocą kategoryjnej gramatyki logicznej* (online: <http://fit2013.mat.umk.pl/slides/FIT-2013-Jaworski.pdf>, dostęp: 15 lutego 2016).
- Kłopotek M.A., Wierchoń S.T., Trojanowski K. (red.) 2006: *Intelligent Information Processing and Web Mining: Proceedings of the International IIS: IIPWM'06 Conference held in Ustroń, Poland, June 19–22, 2006*, Advances in Soft Computing, Springer, Berlin.
- Kobyliński Ł. 2014: *PoliTā: A multitagger for Polish*, [w:] N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, D. Tapias (red.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014, ELRA, Reykjavik, Iceland, s. 2949–2954* (online: <http://www.lrec-conf.org/proceedings/lrec2014/index.html>, dostęp: 30 października 2015).
- Marciniak M., Mykowiecka A. (red.) 2009: *Aspects of natural language processing. Essays dedicated to Leonard Bolc on the Occasion of His 75th Birthday*, Lecture Notes in Computer Science, vol. 5070, Springer, Berlin.
- Marciniak M., Mykowiecka A. 2014: *Terminology extraction from medical texts in Polish*, „Journal of Biomedical Semantics” 5 (online: doi: 10.1186/2041-1480-5-24).
- Marciniak M., Rabiega-Wiśniewska J. 2010: *Elektroniczny słownik fleksyjny nazw Warszawy*, „Prace Filologiczne” LVIII, s. 271–282.
- Marciniak M., Savary A., Sikora P., Woliński M. 2011: *Toposlaw – A Lexicographic Framework for Multi-word Units*, [w:] Z. Vetulani (red.), *Human Language Technology. Challenges for Computer Science and Linguistics. 4th Language and Technology Conference, LTC 2009, Poznań, Poland, November 6–8, 2009, Revised Selected Papers*, Lecture Notes in Artificial Intelligence, vol. 6562, Springer, Berlin, s. 139–150.
- Marcińczuk M., Kocoń J., Janicki M. 2013: *Liner2 – A Customizable Framework for Proper Names Recognition for Polish*, [w:] R. Bembenik, L. Skonieczny, H. Rybiński, M. Kryszkiewicz, M. Niezgodka (red.), *Intelligent Tools for Building a Scientific Information Platform. Advanced Architectures and Solutions*, Springer, Berlin–Heidelberg, s. 231–253, (online: doi: 10.1007/978-3-642-35647-6_17).
- Marczewski P. 2012: *Building a lexicon for a categorial grammar of the Polish language*, praca magisterska, Wydział Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego, Warszawa.
- Maziarz M., Piasecki M., Szpakowicz S. 2012: *Approaching plWordNet 2.0*, [w:] *Proceedings of the 6th Global Wordnet Conference*, Matsue, Japan (online: <http://www.plwordnet.pwr.wroc.pl/lgt/files/publications/paper%2042.pdf>, dostęp: 20 października 2015).
- Mykowiecka A. 2007: *Inżynieria lingwistyczna. Komputerowe przetwarzanie tekstów w języku naturalnym*, Wydawnictwo Polsko-Japońskiej Wyższej Szkoły Technik Komputerowych, Warszawa.
- Orębski T. 2002: *Automatyczna analiza składniowa języka polskiego z wykorzystaniem gramatyki zależnościowej*, rozprawa doktorska, Instytut Podstaw Informatyki PAN, Warszawa.
- Ogrodniczuk M. 2006: *Weryfikacja korpusu wypowiedników polskich (z wykorzystaniem gramatyki formalnej Świdzińskiego)*, praca doktorska, Wydział Neofilologii Uniwersytetu Warszawskiego, Warszawa (online: <http://bc.klf.uw.edu.pl/30/>, dostęp: 20 października 2015).
- Ogrodniczuk M., Głowińska K., Kopeć M., Savary A., Zawisławska M. 2015: *Coreference in Polish. Annotation, Resolution and Evaluation*, Walter De Gruyter, Berlin–Boston.
- Patejuk A., Przepiórkowski A. 2015: *Parallel development of linguistic resources. Towards a structure bank of Polish*, „Prace Filologiczne” LXV, s. 255–270.

- Pawłowski A., Krajewski M., Eder M. 2010: *Time series modelling in the analysis of Homeric verse*, „Eos” XCVII, s. 79–100.
- Piasecki M., Godlewski G. 2006: *Reductionistic, Tree and Rule Based Tagger for Polish*, [w:] M.A. Kłopotek, S.T. Wierchoń, K. Trojanowski (red.), *Intelligent Information Processing and Web Mining. Proceedings of the International IIS: IIPWM'06 Conference held in Ustroń, Poland, June 19–22, 2006*, Advances in Soft Computing, Springer, Berlin, s. 531–540.
- Piasecki M., Szpakowicz S., Broda B. 2009: *Wordnet from the ground up*, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław.
- Piasecki M., Szpakowicz S., Derwojedowa M., Zawisławska M. 2008: *Lexical units as the centrepiece of a wordnet*, [w:] *Conference on Intelligent Information Systems 2008, Challenging Problems in Science: Computer Science*, s. 351–357.
- Polański K. (red.) 1980–1992: *Słownik syntaktyczno-generatywny czasowników polskich*, Zakład Narodowy im. Ossolińskich, Wrocław.
- Przepiórkowski A. 2004a: *Korpus Instytutu Podstaw Informatyki PAN. Wersja wstępna*, Instytut Podstaw Informatyki PAN, Warszawa.
- Przepiórkowski A. 2004b: *Towards the design of a syntactico-semantic lexicon for Polish*, [w:] *Intelligent Information Processing and Web Mining. Proceedings of the International IIS: IIPWM'06 Conference held in Ustroń, Poland, June 19–22, 2006*, „Advances in Soft Computing”, Springer, Berlin, s. 237–246.
- Przepiórkowski A. 2007: *Towards a partial grammar of Polish for valence extraction*, [w:] *Selected contributions from the conference Grammar and Corpora, Sept. 25–27, Liblice*, Academia, Prague, s. 127–133.
- Przepiórkowski A. 2008: *Powierzchniowe przetwarzanie języka polskiego*, Akademicka Oficyna Wydawnicza EXIT, Warszawa.
- Przepiórkowski A., Bańko M., Górski R., Lewandowska-Tomaszczyk B. (red.) 2012: *Narodowy Korpus Języka Polskiego*, Wydawnictwo Naukowe PWN, Warszawa (online: http://nkjp.pl/settings/papers/NKJP_książka.pdf, dostęp: 30 października 2015).
- Przepiórkowski A., Hajnicz E., Patejuk A., Woliński M., Skwarski F., Świdziński M. 2014a: *Walenty: Towards a comprehensive valence dictionary of Polish*, [w:] N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, D. Tapias (red.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014, ELRA, Reykjavík, Iceland*, s. 2785–2792 (online: <http://www.lrec-conf.org/proceedings/lrec2014/index.html>, dostęp: 30 października 2015).
- Przepiórkowski A., Kupś A., Marciniak M., Mykowiecka A. 2001: *Formalny opis języka polskiego. Teoria i implementacja*, Akademicka Oficyna Wydawnicza EXIT, Warszawa.
- Przepiórkowski A., Skwarski F., Hajnicz E., Patejuk A., Świdziński M., Woliński M. 2014b: *Modelowanie własności składniowych czasowników w nowym słowniku walencyjnym języka polskiego*, „Polonica” XXXIII, s. 159–178.
- Przybyła P. 2015: *Odpowiadanie na pytania w języku polskim z użyciem głębokiego rozpoznawania nazw*, rozprawa doktorska, Instytut Podstaw Informatyki PAN, Warszawa.
- Rabiega-Wiśniewska J. 2008: *Formalny opis derywacji w języku polskim. Rzeczowniki i przymiotniki*, Wydawnictwo Wydziału Polonistyki Uniwersytetu Warszawskiego, Warszawa.
- Rabiega-Wiśniewska J., Rudolf M. 2002: *Amor – program automatycznej analizy fleksyjnej tekstu polskiego*, „Biuletyn Polskiego Towarzystwa Językoznawczego” LVIII, s. 175–186.
- Radziszewski A. 2012: *Metody znakowania morfosyntaktycznego i automatycznej płytkiej analizy składniowej języka polskiego*, rozprawa doktorska, Wydział Informatyki i Zarządzania Politechniki Wrocławskiej, Wrocław.
- Radziszewski A. 2013: *A tiered CRF tagger for Polish*, [w:] R. Bembienik, L. Skonieczny, H. Rybiński, M. Kryszkiewicz, M. Niezgodka (red.), *Intelligent tools for building a scientific information platform. Advanced architectures and solutions*, Springer, Berlin, s. 215–230.
- Radziszewski A., Śniatowski T. 2011: *A Memory-Based Tagger for Polish*, 4th Language and Technology Conference, LCT 2009, Poznań, November 6–8, 2009 (online: https://github.com/mariana-scorp/lt-project/blob/master/papers_on_pos-tagging/Radziszewski%20-%202009%20-%20A%20Memory-Based%20Tagger%20for%20Polish.pdf).
- Rudolf M. 2004: *Metody automatycznej analizy korpusu tekstów polskich. Pozyskiwanie, wzbogacanie i przetwarzanie informacji lingwistycznych*, Wydział Polonistyki Uniwersytetu Warszawskiego, Warszawa.
- Saloni Z. 1992: *Rygorystyczny opis polskiej deklinacji przymiotnikowej*, „Zeszyty Naukowe Wydziału Humanistycznego Uniwersytetu Gdańskiego. Prace Językoznawcze” 16, Gdańsk, s. 215–228.
- Saloni Z. 2000: *Wstęp do koniugacji polskiej*, Wydawnictwo Uniwersytetu Warmińsko-Mazurskiego, Olsztyn.
- Saloni Z. 2001: *Czasownik polski*, Wiedza Powszechna, Warszawa.
- Saloni Z., Woliński M., Wołosz R., Gruszczyński W., Skowrońska D. 2007/2012: *Słownik gramatyczny języka polskiego* (dokument elektroniczny), wyd. 1, 2, Warszawa.

- Saloni Z., Woliński M., Wołosz R., Gruszczyński W., Skowrońska D. 2015: *Słownik gramatyczny języka polskiego*, wyd. 3 (online: <http://sgjp.pl/>).
- Savary A., Waszczuk J. 2012: *Narzędzia do anotacji jednostek nazewnucznych*, [w:] A. Przepiórkowski, M. Bańko, R. Górski, B. Lewandowska-Tomaszczyk (red.), *Narodowy Korpus Języka Polskiego*, Wydawnictwo Naukowe PWN, Warszawa (online: http://nkjp.pl/settings/papers/NKJP_książka.pdf, dostęp: 30 października 2015), s. 225–252.
- Szafran K. 1997: *Automatyczna analiza fleksyjna tekstu polskiego (na podstawie „Schematycznego indeksu a tergo” Jana Tokarskiego)*, rozprawa doktorska, Wydział Polonistyki Uniwersytetu Warszawskiego, Warszawa.
- Szmal P., Suszczańska N. 2002: *Tłumaczenie tekstów na język migowy w systemie TGT-1: zasady i realizacja*, „Speech and Language Technology. Technologia Mowy i Języka”, t. 6, s. 113–124.
- Świdziński M. 1992: *Gramatyka formalna języka polskiego*, Wydawnictwa Uniwersytetu Warszawskiego, Warszawa.
- Świdziński M. 1994: *Syntactic Dictionary of Polish Verbs*, Uniwersytet Warszawski, Universiteit van Amsterdam, Warszawa–Amsterdam.
- Świdziński M., Woliński M. 2009: *A new formal definition of Polish nominal phrases*, [w:] M. Marciniak, A. Mykowiecka (red.), *Aspects of natural language processing. Essays dedicated to Leonard Bolc on the occasion of his 75th Birthday*, Lecture Notes in Computer Science, vol. 5070, Springer, Berlin, s. 143–162.
- Świdziński M., Woliński M. 2010: *Towards a bank of constituent parse trees for Polish*, [w:] P. Sojka, A. Horák, I. Kopeček, K. Pala (red.), *Text, Speech and Dialogue. 13th International Conference, TSD 2010*, Brno, Czech Republic, Lecture Notes in Artificial Intelligence, vol. 6231, Springer, Heidelberg, s. 197–204.
- Tokarski J. 1951: *Czasowniki polskie. Formy, typy, wyjątki. Słownik*, Wydawnictwo S. Arcta, Warszawa.
- Tokarski J. 1961a: *Fleksja polska, jej opis w świetle możliwości mechanizacji w urządzeniu przekładowym. Deklinacja rzeczownikowa i inne nieprzymiotnikowe*, „Poradnik Językowy”, nr 4, s. 156–163, nr 8, s. 343–352.
- Tokarski J. 1961b: *Fleksja polska, jej opis w świetle możliwości mechanizacji w urządzeniu przekładowym. Założenia ogólne. Deklinacja przymiotnikowa*, „Poradnik Językowy”, nr 3, s. 97–111.
- Tokarski J. 1962: *Fleksja polska, jej opis w świetle możliwości mechanizacji w urządzeniu przekładowym. Koniugacja*, „Poradnik Językowy”, nr 4, s. 145–158.
- Tokarski J. 1963a: *Fleksja polska, jej opis w świetle możliwości mechanizacji w urządzeniu przekładowym. Rozpoznawanie form deklinacji przymiotnikowej*, „Poradnik Językowy”, nr 1, s. 2–21.
- Tokarski J. 1963b: *Fleksja polska, jej opis w świetle możliwości mechanizacji w urządzeniu przekładowym. Rozpoznawanie form deklinacji rzeczownikowej*, „Poradnik Językowy”, nr 2, s. 55–77.
- Tokarski J. 1963c: *Fleksja polska, jej opis w świetle możliwości mechanizacji w urządzeniu przekładowym. Rozpoznawanie form osobowych w koniugacji*, „Poradnik Językowy”, nr 3–4, s. 111–130.
- Tokarski J. 1963d: *Fleksja polska, jej opis w świetle możliwości mechanizacji w urządzeniu przekładowym. Rozpoznawanie koniugacyjnych form nieosobowych*, „Poradnik Językowy”, nr 5–6, s. 173–184.
- Tokarski J. 1973: *Fleksja polska*, Państwowe Wydawnictwo Naukowe, Warszawa.
- Tokarski J. 1993: *Schematyczny indeks a tergo polskich form wyrazowych*, red. i oprac. Z. Saloni, Wydawnictwo Naukowe PWN, Warszawa.
- Vetulani Z. (red.) 2011: *Human Language Technology. Challenges for Computer Science and Linguistics: 4th Language and Technology Conference, LTC 2009, Poznań, November 6–8, 2009, Revised Selected Papers*, Lecture Notes in Artificial Intelligence, vol. 6562, Springer, Berlin.
- Vetulani Z., Kubis M., Obrębski T. 2010a: *PolNet – Polish WordNet: Data and tools*, [w:] N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odiijk, S. Piperidis, M. Rosner, D. Tapias (red.), *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC’10)*, European Language Resources Association (ELRA), Valletta, Malta, s. 3793–3797 (online: <http://www.lrec-conf.org/proceedings/lrec2010/>, dostęp: 20 października 2015).
- Vetulani Z., Marcinak J., Obrębski T., Vetulani G., Dąbrowski A., Kubis M., Osiński J., Walkowska J., Kubacki P., Witalewski K. 2010b: *Zasoby językowe i technologie przetwarzania tekstu. POLINT-112-SMS jako przykład aplikacji z zakresu bezpieczeństwa publicznego*, Wydawnictwo Uniwersytetu im. Adama Mickiewicza, Poznań.
- Vetulani Z., Martinek J., Obrębski T., Vetulani G. 2008: *Dictionary based methods and tools for language engineering*, Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza, Poznań.
- Woliński M. 2004: *Komputerowa weryfikacja gramatyki Świdzińskiego*, rozprawa doktorska, Instytut Podstaw Informatyki PAN (online: <http://www.ipipan.waw.pl/wolinski/publ/mw-phd.pdf>, dostęp: 5 listopada 2015).
- Woliński M. 2006: *Morfheus – a practical tool for the morphological analysis of Polish*, [w:] M.A. Kłopotek, S.T. Wierchoń, K. Trojanowski (red.), *Intelligent Information Processing and Web Mining. Proceedings of the International IIS: IIPWM’06 Conference held in Ustroń, Poland, June 19–22, 2006*, Advances in Soft Computing, Springer, Berlin, s. 503–512.

- Woliński M. 2014: *Morfeusz reloaded*, [w:] N. Calzolari i in. (red.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014, ELRA, Reykjavík, Iceland* (online: <http://www.lrec-conf.org/proceedings/lrec2014/index.html>, dostęp: 30 października 2015), s. 1106–1111.
- Woliński M., Głowińska K., Świdziński M. 2011: *A preliminary version of Składnica – a treebank of Polish*, [w:] Z. Vetulani (red.), *Human Language Technology. Challenges for Computer Science and Linguistics: 4th Language and Technology Conference, LTC 2009, Poznań, Poland, November 6–8, 2009, Revised Selected Papers*, Lecture Notes in Artificial Intelligence, vol. 6562, Springer, Berlin, s. 299–303.
- Woliński M., Miłkowski M., Ogrodniczuk M., Przepiórkowski A., Szalkiewicz Ł. 2012: *PoliMorf: a (not so) new open morphological dictionary for Polish*, [w:] N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, S. Piperidis (red.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation, LREC 2012, ELRA, Istanbul, Turkey*, s. 860–864 (online: <http://www.lrec-conf.org/proceedings/lrec2012/index.html>, dostęp: 20 października 2015).
- Wołosz R. 2005: *Efektywna metoda analizy i syntezy morfologicznej w języku polskim*, Akademicka Oficyna Wydawnicza EXIT, Warszawa.
- Wróblewska A. 2012: *Polish dependency bank*, „Linguistic Issues in Language Technology” 7 (1) (online: <http://elanguage.net/journals/index.php/lilt/article/view/2684>).

Summary

Computational linguistics and language engineering after 1989

Keywords: computational linguistics, language engineering, natural language processing.

The paper presents research on NLP in Poland in the last 25 years. In this review tools and resources for syntactical and morphological research with the exception of corpora and dictionaries have been presented. Formal grammars, parsers, taggers, morphological analysers, valence dictionaries, wordnets, multi-word dictionaries and other lexical resources of this kind are briefly discussed.