

Maurycy Graszewicz  
Uniwersytet Wrocławski

## **Korpus wypowiedzi polskich polityków (KWPP)**

Czasem jest tak, że to, co się liczy, nie da się policzyć,  
a to, co daje się policzyć — nie liczy się.

Albert Einstein

### **Wstęp**

Niniejszy artykuł stanowi wstępny opis przygotowanego już i aktualnie opracowywanego korpusu wypowiedzi polskich polityków (KWPP), projektu realizowanego przez autora od 2011 roku. O ile autorowi wiadomo, nie istnieje tak duży korpus wypowiedzi polskich polityków. Jedynym, podobnym co do zawartości, nie zaś co do wielkości, jest korpus wypowiedzi polityków z lat 2004–2009 [Graszewicz 2011]. Tworzenie korpusów wypowiedzi polityków ma zaś w światowej nauce niemłodą już tradycję i historię [por. np. Adda-Decker *et al.* 2008; Faaß, Heid 2012; Guerini, Strapparava, Stock 2008; Barbaresi 2012; Osenova, Simov 2012]. Po przetworzeniu i przygotowaniu autor zamierza udostępnić publicznie wszystkie te elementy korpusu, które będą możliwe do upowszechnienia.

### **Struktura korpusu**

W trakcie tworzenia KWPP przygotowano łącznie 41 korpusów dla poszczególnych partii politycznych zasiadających w Sejmie RP w latach 2006–2013. Taki wybór partii był podyktowany głównie tym, że pozostałe pozasejmowe ugrupowania mają niewielki wpływ na dominujący polityczny dyskurs i polityczne tematy.

Każdy pojedynczy korpus liczył około 100 000 wyrazów. W tabeli 1 zaznaczone pola oznaczają, że w danym roku przedstawiciele określonej partii zasiadali w Sejmie, czyli że z wypowiedzi polityków tej partii przygotowano korpus wchodzący w skład KWPP.

Tabela 1. Udział poszczególnych partii w składzie KWPP w kolejnych latach

| Rok  | LPR* | PiS | PO | PSL | SRP | SLD | PJN | RP | SP |
|------|------|-----|----|-----|-----|-----|-----|----|----|
| 2006 |      |     |    |     |     |     |     |    |    |
| 2007 |      |     |    |     |     |     |     |    |    |
| 2008 |      |     |    |     |     |     |     |    |    |
| 2009 |      |     |    |     |     |     |     |    |    |
| 2010 |      |     |    |     |     |     |     |    |    |
| 2011 |      |     |    |     |     |     |     |    |    |
| 2012 |      |     |    |     |     |     |     |    |    |
| 2013 |      |     |    |     |     |     |     |    |    |

\* Skróty: LPR — Liga Polskich Rodzin, PiS — Prawo i Sprawiedliwość, PJN — Polska Jest Najważniejsza, PO — Platforma Obywatelska, PSL — Polskie Stronnictwo Ludowe, SRP — Samoobrona RP, SLD — Sojusz Lewicy Demokratycznej, RP — Ruch Palikota (od 2013 roku Twój Ruch), SP — Solidarna Polska.

Podobnie jak w mniejszym korpusie z lat 2004–2009 przyjęto, że najbardziej interesujące w kontekście systemu politycznego będą wypowiedzi jak najbardziej zbliżone do języka naturalnego, do którego badacze mają marginalny dostęp. Każdy korpus poszczególnych partii w danym roku został wyselekcjonowany i zgromadzony w taki sposób, aby 40% stanowiły wywiady, 40% teksty z blogów prowadzonych przez polityków, a pozostałe 20% wypowiedzi sejmowe. Jeśli polityk (co zdarzało się stosunkowo często) nie prowadził bloga, tę część wypowiedzi uzupełniano wypowiedziami najbardziej zbliżonymi, a więc wywiadami. W kategorii wywiady znajdują się wypowiedzi prasowe, radiowe i telewizyjne (w formie transkrypcji lub stenogramów) [por. Graszewicz 2011, s. 179–180]. Wypowiedzi w mass mediach stanowią większość zawartości korpusów także z tego powodu, że politycy, chcąc dotrzeć do odbiorcy, potencjalnego wyborcy używają mass mediów w procesie komunikacji politycznej. Ma to swoje co najmniej dwojakie konsekwencje. Po pierwsze, politycy coraz rzadziej komunikują się bezpośrednio z wyborcami; po drugie zaś media, których funkcją jest samoobserwacja społeczeństwa i zwrotne sterowanie samoobserwacją, używają rozróżnień, dzielą system polityczny na szersze opcje [por. Graszewicz 2011, s. 172]. „Analiza polskiej sceny politycznej zaistniałej po 1990 prowadzi do wniosku, że istotną rolę w procesie kreowania obrazu rzeczywistości społecznej odgrywają media. Politycy, chcąc dotrzeć do masowego odbiorcy

z czytelnym komunikatem, muszą skorzystać z ich pośrednictwa, dostosować język przekazu do stawianych przez nie wymogów” [Biniewicz 2009, s. 305].

Dobór wypowiedzi był zaś całkowicie przypadkowy. Tabela 2 ukazuje udział poszczególnych rodzajów wypowiedzi w kolejnych latach.

Tabela 2. Udział poszczególnych rodzajów wypowiedzi w kolejnych latach

| Rok  | Razem (2006–2013) |        |         |         |
|------|-------------------|--------|---------|---------|
|      | blog              | sejm   | wywiad  | Σ       |
| 2006 | 59179             | 174022 | 370027  | 603228  |
| 2007 | 93564             | 123365 | 388731  | 605660  |
| 2008 | 40916             | 70058  | 301202  | 412176  |
| 2009 | 77998             | 63633  | 277074  | 418705  |
| 2010 | 96805             | 56002  | 255044  | 407851  |
| 2011 | 73128             | 68673  | 277440  | 419241  |
| 2012 | 108279            | 130451 | 361484  | 600214  |
| 2013 | 110248            | 124541 | 372076  | 606865  |
| Σ    | 660117            | 810745 | 2603078 | 4073940 |

Całościowy KWPP składający się z poszczególnych korpusów liczy 4 073 940 leksemów. Jest to około 6224 stron znormalizowanego maszynopisu.

Kolejne tabele zawierają dane dotyczące poszczególnych rodzajów wypowiedzi w korpusach poszczególnych partii politycznych w kolejnych latach.

Tabela 3. Udział poszczególnych rodzajów wypowiedzi w poszczególnych korpusach Ligi Polskich Rodzin

| Rok  | LPR   |       |        |        |
|------|-------|-------|--------|--------|
|      | blog  | sejm  | wywiad | Σ      |
| 2006 | 13360 | 33700 | 53460  | 100520 |
| 2007 | 40127 | 9182  | 49325  | 98634  |
| 2008 | 0     | 0     | 0      | 0      |
| 2009 | 0     | 0     | 0      | 0      |
| 2010 | 0     | 0     | 0      | 0      |
| 2011 | 0     | 0     | 0      | 0      |
| 2012 | 0     | 0     | 0      | 0      |
| 2013 | 0     | 0     | 0      | 0      |
| Σ    | 53487 | 42882 | 102785 | 199154 |

Tabela 4. Udział poszczególnych rodzajów wypowiedzi w poszczególnych korpusach Prawa i Sprawiedliwości

| Rok  | PiS    |        |        |        |
|------|--------|--------|--------|--------|
|      | blog   | sejm   | wywiad | Σ      |
| 2006 | 0      | 9735   | 91673  | 101408 |
| 2007 | 1619   | 11933  | 88634  | 102186 |
| 2008 | 0      | 20279  | 83109  | 103388 |
| 2009 | 10714  | 15030  | 79463  | 105207 |
| 2010 | 30499  | 13440  | 59394  | 103333 |
| 2011 | 31265  | 7831   | 64158  | 103254 |
| 2012 | 12736  | 12241  | 75023  | 100000 |
| 2013 | 23954  | 15190  | 61059  | 100203 |
| Σ    | 110787 | 105679 | 602513 | 818979 |

Tabela 5. Udział poszczególnych rodzajów wypowiedzi w poszczególnych korpusach Polska Jest Najważniejsza

| Rok  | PJN   |      |        |       |
|------|-------|------|--------|-------|
|      | blog  | sejm | wywiad | Σ     |
| 2006 | 0     | 0    | 0      | 0     |
| 2007 | 0     | 0    | 0      | 0     |
| 2008 | 0     | 0    | 0      | 0     |
| 2009 | 0     | 0    | 0      | 0     |
| 2010 | 0     | 0    | 0      | 0     |
| 2011 | 12500 | 5000 | 32500  | 50000 |
| 2012 | 0     | 0    | 0      | 0     |
| 2013 | 0     | 0    | 0      | 0     |
| Σ    | 12500 | 5000 | 32500  | 50000 |

Tabela 6. Udział poszczególnych rodzajów wypowiedzi w poszczególnych korpusach Platformy Obywatelskiej

| Rok  | PO    |       |        |        |
|------|-------|-------|--------|--------|
|      | blog  | sejm  | wywiad | Σ      |
| 2006 | 0     | 28262 | 71694  | 99956  |
| 2007 | 12499 | 5741  | 84182  | 102422 |
| 2008 | 7887  | 7437  | 89428  | 104752 |
| 2009 | 32438 | 7687  | 65577  | 105702 |
| 2010 | 10123 | 8310  | 82093  | 100526 |
| 2011 | 7650  | 20334 | 79684  | 107668 |

|      |       |        |        |        |
|------|-------|--------|--------|--------|
| 2012 | 0     | 14957  | 85043  | 100000 |
| 2013 | 0     | 24307  | 75693  | 100000 |
| Σ    | 70597 | 117035 | 633394 | 821026 |

Tabela 7. Udział poszczególnych rodzajów wypowiedzi w poszczególnych korpusach Polskiego Stronnictwa Ludowego

| Rok  | PSL    |        |        |        |
|------|--------|--------|--------|--------|
|      | blog   | sejm   | wywiad | Σ      |
| 2006 | 16000  | 38794  | 45206  | 100000 |
| 2007 | 11451  | 46190  | 42963  | 100604 |
| 2008 | 15697  | 21597  | 65214  | 102508 |
| 2009 | 12681  | 22082  | 71852  | 106615 |
| 2010 | 36140  | 13985  | 53187  | 103312 |
| 2011 | 25525  | 23460  | 57646  | 106631 |
| 2012 | 27313  | 18139  | 54654  | 100106 |
| 2013 | 18466  | 14379  | 68079  | 100924 |
| Σ    | 163273 | 198626 | 458801 | 820700 |

Tabela 8. Udział poszczególnych rodzajów wypowiedzi w poszczególnych korpusach Samoobrony RP

| Rok  | Samoobrona |       |        |        |
|------|------------|-------|--------|--------|
|      | blog       | sejm  | wywiad | Σ      |
| 2006 | 20144      | 24112 | 56315  | 100571 |
| 2007 | 10100      | 30077 | 60623  | 100800 |
| 2008 | 0          | 0     | 0      | 0      |
| 2009 | 0          | 0     | 0      | 0      |
| 2010 | 0          | 0     | 0      | 0      |
| 2011 | 0          | 0     | 0      | 0      |
| 2012 | 0          | 0     | 0      | 0      |
| 2013 | 0          | 0     | 0      | 0      |
| Σ    | 30244      | 54189 | 116938 | 201371 |

Tabela 9. Udział poszczególnych rodzajów wypowiedzi w poszczególnych korpusach Sojuszu Lewicy Demokratycznej

| Rok  | SLD  |       |        |        |
|------|------|-------|--------|--------|
|      | blog | sejm  | wywiad | Σ      |
| 2006 | 9675 | 39419 | 51679  | 100773 |

|      |        |        |        |        |
|------|--------|--------|--------|--------|
| 2007 | 17768  | 20242  | 63004  | 101014 |
| 2008 | 17332  | 20745  | 63451  | 101528 |
| 2009 | 22165  | 18834  | 60182  | 101181 |
| 2010 | 20043  | 20267  | 60370  | 100680 |
| 2011 | 8688   | 17048  | 75952  | 101688 |
| 2012 | 28604  | 20040  | 51358  | 100002 |
| 2013 | 17278  | 15207  | 67975  | 100460 |
| Σ    | 141553 | 171802 | 493971 | 807326 |

Tabela 10. Udział poszczególnych rodzajów wypowiedzi w poszczególnych korpusach Ruchu Palikota

| Rok  | RP    |       |        |        |
|------|-------|-------|--------|--------|
|      | blog  | sejm  | wywiad | Σ      |
| 2006 | 0     | 0     | 0      | 0      |
| 2007 | 0     | 0     | 0      | 0      |
| 2008 | 0     | 0     | 0      | 0      |
| 2009 | 0     | 0     | 0      | 0      |
| 2010 | 0     | 0     | 0      | 0      |
| 2011 | 0     | 0     | 0      | 0      |
| 2012 | 15002 | 20040 | 65064  | 100106 |
| 2013 | 14789 | 24567 | 60644  | 100000 |
| Σ    | 29791 | 44607 | 125708 | 200106 |

Tabela 11. Udział poszczególnych rodzajów wypowiedzi w poszczególnych korpusach Solidarnej Polski

| Rok  | SP    |       |        |        |
|------|-------|-------|--------|--------|
|      | blog  | sejm  | wywiad | Σ      |
| 2006 | 0     | 0     | 0      | 0      |
| 2007 | 0     | 0     | 0      | 0      |
| 2008 | 0     | 0     | 0      | 0      |
| 2009 | 0     | 0     | 0      | 0      |
| 2010 | 0     | 0     | 0      | 0      |
| 2011 | 0     | 0     | 0      | 0      |
| 2012 | 24624 | 45034 | 30342  | 100000 |
| 2013 | 35761 | 30891 | 38626  | 105278 |
| Σ    | 60385 | 75925 | 68968  | 205278 |

Wszystkie korpusy wypowiedzi zostały przed analizą zlematyzowane za pomocą lematyzatora morfosyntaktycznego języka polskiego [Piasecki 2007, s. 151–167; Broda, Piasecki, Radziszewski 2008].

## Opis statystyczny i ilościowy populacji wypowiedzi w korpusie wypowiedzi polskich polityków (KWPP)

W tabeli 12 znajdują się informacje dotyczące liczby wystąpień (*tokens*) i unikatowych słów (*types*) dla kolejnych korpusów poszczególnych partii politycznych.

Tabela 12. Opis statystyczny i ilościowy populacji badanych korpusów dla poszczególnych partii w latach 2006–2013

| Partia* | Rok  | Wielkość pliku | Słowa w tekście | Słowa użyte do listy wyrazów | Hasło (wyraz słownikowy) | Stosunek słów do haseł (wyrazów słownikowych) TTR |
|---------|------|----------------|-----------------|------------------------------|--------------------------|---------------------------------------------------|
| PSL     | 2006 | 1 328 034      | 101 261         | 100 398                      | 7 192                    | 7,16                                              |
| PSL     | 2007 | 1 346 012      | 101 781         | 100 886                      | 7 392                    | 7,33                                              |
| PSL     | 2008 | 1 374 588      | 104 303         | 103 024                      | 8 322                    | 8,08                                              |
| PSL     | 2009 | 1 442 652      | 109 348         | 107 613                      | 8 551                    | 7,95                                              |
| PSL     | 2010 | 1 437 858      | 107 353         | 105 728                      | 9 281                    | 8,78                                              |
| PSL     | 2011 | 1 443 404      | 109 064         | 106 831                      | 8 731                    | 8,17                                              |
| PSL     | 2012 | 2 053 838      | 101 307         | 99 989                       | 7 683                    | 7,68                                              |
| PSL     | 2013 | 618 789        | 94 672          | 93 872                       | 6 685                    | 7,12                                              |
| SRP     | 2006 | 1 302 846      | 102 138         | 101 089                      | 8 217                    | 8,13                                              |
| SRP     | 2007 | 1 332 636      | 102 361         | 101 538                      | 7 692                    | 7,58                                              |
| LPR     | 2006 | 1 356 838      | 103 091         | 101 932                      | 8 717                    | 8,55                                              |
| LPR     | 2007 | 2 085 590      | 103 091         | 101 932                      | 8 717                    | 8,55                                              |
| PiS     | 2006 | 1 337 660      | 102 857         | 102 408                      | 6 461                    | 6,31                                              |
| PiS     | 2007 | 1 345 618      | 103 609         | 103 088                      | 7 461                    | 7,24                                              |
| PiS     | 2008 | 1 372 640      | 104 824         | 104 287                      | 7 481                    | 7,17                                              |
| PiS     | 2009 | 1 377 160      | 105 514         | 104 891                      | 8 115                    | 7,74                                              |
| PiS     | 2010 | 1 366 206      | 104 940         | 103 960                      | 9 152                    | 8,80                                              |
| PiS     | 2011 | 1 365 372      | 104 613         | 103 788                      | 8 864                    | 8,54                                              |
| PiS     | 2012 | 1 317 456      | 101 137         | 100 046                      | 7 966                    | 7,96                                              |
| PiS     | 2013 | 657 987        | 101 271         | 100 125                      | 8 005                    | 8,0                                               |
| PO      | 2006 | 2 031 203      | 101 235         | 100 781                      | 6 466                    | 6,42                                              |

|     |      |           |         |         |       |       |
|-----|------|-----------|---------|---------|-------|-------|
| PO  | 2007 | 2 404 586 | 118 425 | 118 117 | 7 403 | 6,27  |
| PO  | 2008 | 1 401 574 | 106 356 | 105 634 | 8 052 | 7,62  |
| PO  | 2009 | 2 109 259 | 101 709 | 100 570 | 8 826 | 8,78  |
| PO  | 2010 | 2 024 481 | 101 735 | 101 210 | 7 478 | 7,39  |
| PO  | 2011 | 2 213 636 | 110 891 | 109 535 | 7 711 | 7,04  |
| PO  | 2012 | 1 315 012 | 101 408 | 100 966 | 7 222 | 7,15  |
| PO  | 2013 | 654 713   | 101 488 | 101 045 | 6 939 | 6,87  |
| PJN | 2011 | 647 126   | 50 710  | 50 312  | 5 644 | 11,22 |
| SLD | 2006 | 1 342 030 | 101 883 | 101 065 | 7 750 | 7,67  |
| SLD | 2007 | 1 338 470 | 101 904 | 101 332 | 7 195 | 7,10  |
| SLD | 2008 | 1 353 222 | 102 878 | 102 145 | 8 086 | 7,92  |
| SLD | 2009 | 1 446 414 | 108 965 | 108 138 | 8 816 | 8,15  |
| SLD | 2010 | 1 304 958 | 98 397  | 97 522  | 8 675 | 8,90  |
| SLD | 2011 | 1 345 572 | 102 803 | 101 927 | 8 154 | 8,00  |
| SLD | 2012 | 670 668   | 101 494 | 100 325 | 9 094 | 9,06  |
| SLD | 2013 | 665 181   | 101 847 | 100 644 | 8 438 | 8,38  |
| RP  | 2012 | 657 440   | 101 391 | 100 597 | 8 548 | 8,50  |
| RP  | 2013 | 657 043   | 101 303 | 100 665 | 8 382 | 8,33  |
| SP  | 2012 | 676 169   | 101 102 | 99 973  | 7 929 | 7,93  |
| SP  | 2013 | 748 050   | 111 630 | 110 263 | 8 611 | 7,81  |

\* Legenda: LPR— Liga Polskich Rodzin, PiS — Prawo i Sprawiedliwość, PJN — Polska Jest Najważniejsza, PO — Platforma Obywatelska, PSL — Polskie Stronnictwo Ludowe, SRP — Samoobrona RP, SLD — Sojusz Lewicy Demokratycznej, RP — Ruch Palikota (od 2013 roku Twój Ruch), SP — Solidarna Polska.

Tabela 13 zawiera opis statystyczny i ilościowy populacji KWPP, na który składa się 41 poszczególnych korpusów partii politycznych.

Tabela 13. Opis statystyczny i ilościowy populacji KWPP

| Wielkość pliku | Słowa w tekście | Słowa użyte do listy wyrazów | Hasło (wyraz słownikowy) | Stosunek słów do haseł (wyrazów słownikowych) | Usunięte liczby |
|----------------|-----------------|------------------------------|--------------------------|-----------------------------------------------|-----------------|
| 27 531 386     | 4 103 614       | 4 068 330                    | 48 894                   | 1,20                                          | 35 284          |

## Reprezentatywność korpusu

Kwestia reprezentatywności korpusu jako próby badawczej, szczególnie w przypadku wypowiedzi polityków, jest sprawą problematyczną [Graszewicz 2011, s. 18–21].

Trzeba przypomnieć w tym miejscu, że konstrukcja korpusu, a więc udział w nim poszczególnych „gatunków” wypowiedzi, została ustalona arbitralnie podług problemów i celów badawczych, jakie za pomocą korpusu można rozstrzygnąć. Z punktu widzenia pojęcia reprezentatywności ideałem byłby losowy dobór tekstów, trzeba jednak przypomnieć, że „nie istnieje żaden jednorodny makrotekst jako populacja generalna, wobec której badane teksty można byłoby traktować jako próby z niej wylosowane” [Sambor 1998, s. 54]. Trudno byłoby zatem z jednej strony wyobrazić sobie zamknięty korpus wszystkich wypowiedzi polityków poszczególnych partii politycznych, z drugiej zaś prawie niemożliwe byłoby jego stworzenie zarówno ze względu na rozległość materiału, jak i dostępność tekstu czy zgodę właścicieli praw autorskich. „Intuicja podpowiada nam bowiem (a praktyka potwierdza), że starannie przygotowany korpus jest zawsze reprezentatywny w pewnym stopniu, a błąd popełnia jedynie ten, kto dopuszcza się nadinterpretacji” [Pawłowski 1999, s. 88].

## Bibliografia

- Adda-Decker M., Barras C., Adda G., Paroubek P., Boula de Mareuil Ph., Habert B., 2008, *Annotation and analysis of overlapping speech in political interviews*, [w:] Proc. 6th Internat. Language Resources and Evaluation Conf. (LREC'08), Marakech.
- Anusiewicz J., 1994, *Lingwistyka kulturowa. Zarys problematyki*, Wrocław.
- Anusiewicz J., 1999, *Językowy obraz świata w poglądach niektórych językoznawców niemieckich XX wieku*, [w:] *Językowy obraz świata*, red. J. Bartmiński, Lublin, s. 162–289.
- Barbaresi A., 2012, *German Political Speeches — Corpus and Visualization*. DGfS-CL poster session, Mar, Frankfurt.
- Bartmiński J., 2006, *Językowe podstawy obrazu świata*, Lublin.
- Bartmiński J. (red.), 2004, *Językowy obraz świata*, Lublin.
- Bartmiński J. (red.), 2006, *Język — wartości — polityka. Zmiany rozumienia nazw wartości w okresie transformacji ustrojowej w Polsce: raport z badań empirycznych*, Lublin.
- Berger P.L., Luckmann T., 1983, *Społeczne tworzenie rzeczywistości*, Warszawa.
- Biniewicz J., 2009, *Strategie konwersacyjne w komunikacji publicznej (na przykładzie programów w formule talk-show)*, [w:] *Teorie komunikacji i mediów*, t. 1, red. M. Graszewicz, J. Jastrzębski, Wrocław.
- Broda B., Piasecki M., Radziszewski A., 2008, *Towards a Set of General Purpose Morphosyntactic Tools for Polish*, [w:] *Intelligent Information Systems XVI. Proceedings of the International IIS'08 Conference held in Zakopane*, June 2008, Academic Publishing House EXIT, red. M.A. Kłopotek, A. Przepiórkowski, S.T. Wierchoń, K. Trojanowski, Warszawa, s. 441–450.
- Faaß G., Heid U., 2012, *Deutsche politische Kommunikation der Gegenwart als linguistisch annotiertes Korpus*, Poster at the DGfS-CL Poster-Session, Frankfurt.
- Graszewicz M., 2011, *Polski system polityczny. Semantyki i struktury komunikacji politycznej*, Wrocław.
- Guerini M., Giampiccolo D., Moretti G., Sprugnoli R., Strapparava C., 2013, *The New Release of CORPS: A Corpus of Political Speeches Annotated with Audience Reactions*, [w:] *Multimodal Communication in Political Speech. Shaping Minds and Social Action*, Berlin-Heidelberg, s. 86–98.
- Guerini M., Strapparava C., Stock O., 2012, *Audience reactions for information extraction about persuasive language in political communication*, [w:] *Multimodal Information Extraction*, red. M. Maybury, MIT Press, s. 289–304.

- Guerini M., Strapparava C., Stock O., 2008, *CORPS: A Corpus of Tagged Political Speeches for Persuasive Communication Processing*, „Journal of Information Technology & Politics”, 5(1), s. 19–32.
- Hammerl R., Sambor J., 1990, *Statystyka dla językoznawców*, Warszawa.
- Hammerl R., Sambor J., 1993, *O statystycznych prawach językowych*, Warszawa.
- Osenova P., Simov K., 2012, *The Political Speech Corpus of Bulgarian*, Proceedings of LREC, Istanbul, s. 1744–1747,
- Pawłowski A., 1999, *Metodologiczne podstawy wykorzystywania słowników frekwencyjnych w badaniu językowego obrazu świata*, [w:] *Przeszłość w językowym obrazie świata*, red. A. Pajdzińska, P. Krzyżanowski, Lublin.
- Pawłowski A., 2001, *Metody kwantytatywne w sekwencyjnej analizie tekstu*, Warszawa.
- Piasecki M., 2007, *Polish Tagger TaKIPi: Rule Based Construction and Optimisation*, „Task Quarterly”, 11, s. 151–167.
- Samber J., 1988, *Lingwistyka kwantytatywna. Stan badań i perspektywy rozwoju*, „Biuletyn Polskiego Towarzystwa Językoznawczego”, 41, s. 47–67.

## Corpus of statements by Polish politicians (CSPP)

### Summary

The article describes the corpus of statements by Polish politicians (CSPP), which is currently under preparation. The CSPP comprises text corpora of political parties present in the Polish parliament in 2006–2013. First, the author places the corpus in question in the context of other national corpora of politicians' statements. In addition, he describes the exact size and composition of the corpus, and provides a statistical and quantitative analysis of the statements included in the CSPP.