

Joanna Wołoszyn   
Instytut Podstaw Informatyki PAN, Warszawa  
[joanna.woloszyn@ipipan.waw.pl](mailto:joanna.woloszyn@ipipan.waw.pl)

Marcin Woliński   
Instytut Podstaw Informatyki PAN, Warszawa  
[wolinski@ipipan.waw.pl](mailto:wolinski@ipipan.waw.pl)

Witold Kieraś   
Instytut Podstaw Informatyki PAN, Warszawa  
[witold.kieras@ipipan.waw.pl](mailto:witold.kieras@ipipan.waw.pl)

## SIEĆ POWIĄZAŃ DERYWACYJNYCH NA MATERIALE SŁOWNIKA GRAMATYCZNEGO JĘZYKA POLSKIEGO. PROPOZYCJA KLASYFIKACJI<sup>1</sup>

**Słowa kluczowe:** słowotwórstwo, sieć derywacyjna, słownik gramatyczny  
**Keywords:** derivation, derivational network, grammatical dictionary

### 1. Wstęp

Niniejszy artykuł przedstawia propozycję wprowadzenia do *Słownika gramatycznego języka polskiego* (SGJP) sieci powiązań derywacyjnych. Na początku chcielibyśmy zaznaczyć, że nie jest naszą ambicją stworzenie jakiejś nowej teorii czy kategoryzacji słowotwórczej, a jedynie dostosowanie już istniejących koncepcji do potrzeb SGJP i jego pochodnych, jak np. analizatora fleksyjnego Morfeusz. W prezentowanym tekście skupimy się przede wszystkim na wybranych kwestiach związanych z podziałem derywatów na kategorie, a na końcu krótko podsumujemy przeprowadzony

---

<sup>1</sup> Wkład autorów w powstanie tekstu jest następujący: J. Wołoszyn – 60%, W. Kieraś – 20%, M. Woliński – 20%.

przez nas eksperyment polegający na połączeniu leksemów z wybranej z SGJP listy 2500 najczęściej używanych wyrazów z ich derywatami (wybieranymi z pełnego *Słownika*) oraz kategoryzacji wyrazów pochodnych. Działanie to można traktować jako pilotaż rozszerzenia sieci derywatów SGJP.

## 2. Motywacje

*Słownik gramatyczny języka polskiego* od początku swojego istnienia był projektem nastawionym na rygorystyczny opis polskiej fleksji. Oprócz zjawisk czysto fleksyjnych w pewnych wąsko zakrojonych obszarach systematycznie notował najbardziej regularne zjawiska słowotwórcze. Mianowicie za pomocą tzw. odsyłaczy, czyli powiązań między hasłami, połączono derywowane od czasowników imiesłowy i odsłowniki oraz regularnie tworzone od przymiotników przysłówki i rzeczowniki zakończone na *-ość*. W IV wydaniu słownik wzbogacono o obszerną listę nazwisk i nazw miejscowości, które również systematycznie połączono: nazwiska parami w wariantach męskim i żeńskim, nazwy miejscowości zaś z derywowanymi od nich przymiotnikami i rzeczownikowymi nazwami mieszkańców (również w wariantach męskim i żeńskim). W innych obszarach autorzy notowali powiązania słowotwórcze bardzo rzadko i doraźnie, bez wyraźnego planu rozszerzenia sieci powiązań derywacyjnych. Jedną z przyczyn tego stanu rzeczy był sposób, w jaki SGJP powstał. Celem było ujęcie deklinacji i koniugacji w postaci wzorów odmiany: list zakończeń, które po przyłączeniu do wspólnego rdzenia stworzą pełną listę form fleksyjnych danego paradygmatu. Ujęte w SGJP regularne derywaty (z pewnymi wyjątkami) są tworzone z wykorzystaniem zawartych w słowniku wzorów fleksyjnych. Nieco szersze uwzględnienie powiązań derywacyjnych między hasłami słownika wymagałoby rozbudowania listy wzorów odmiany w taki sposób, by uwzględniały również alternacje zachodzące w derywatach, a nie tylko w formach fleksyjnych danego leksemu. Alternatywne wprowadzenie takich połączeń bez istotnej modyfikacji systemu wzorów odmiany wymagałoby w ogromnej mierze pracy ręcznej.

Wydaje się jednak, że szersze uwzględnienie powiązań derywacyjnych między hasłami w SGJP byłoby korzystne i interesujące zarówno z punktu widzenia jednostkowego użytkownika słownika, jak i użytkowników zautomatyzowanych, czyli po prostu programów komputerowych wykorzystywanych do przetwarzania tekstów polskich. Naturalnym przykładem takiego programu jest *Morfeusz 2* (Kieraś, Woliński 2017), który dzięki wykorzystaniu danych SGJP zwraca wszystkie możliwe interpretacje fleksyjne danego słowa tekstowego. Łatwo można sobie wyobrazić, że *Morfeusz* (lub inny program) mógłby zinterpretowanym fleksyjnie wyrazom przypisywać ich bezpośrednie podstawy derywacyjne oraz typy ich powiązań z podstawą, co z kolei pozwoliłoby np. na ilościowe badania słowotwórcze na obszernych korpusach polszczyzny.

Stworzenie takiej bazy powiązań derywacyjnych na podstawie danych SGJP wydaje się dziś prostsze niż do tej pory dzięki możliwościom, jakie dają duże modele językowe stanowiące istotny element obserwowanego obecnie kolejnego przełomu w rozwoju sztucznej inteligencji (dla języka polskiego np. HerBERT: Mroczkowski et al. 2021). Za pomocą odpowiednio przygotowanych danych treningowych można taki model skłonić do wskazania prawdopodobnych połączeń między hasłami i opatrzenia ich odpowiednią etykietą odpowiadającą typowi derywacji (względnie: do wygenerowania potencjalnych derywatów różnego typu na podstawie hasła wejściowego). Kluczowe jest zatem przygotowanie koncepcji oznaczania powiązań derywacyjnych i odpowiedniej próbki danych treningowych. Techniczna strona automatycznego tworzenia sieci powiązań jest przedmiotem osobnych prac, w niniejszym artykule skupimy się na propozycji wyboru typów produktywnych powiązań słowotwórczych na bazie wyekstrahowanego ze słownika zbioru haseł oraz stworzeniu propozycji zbioru treningowego, który posłuży do dalszych prac nad rozszerzeniem sieci derywacyjnej.

Należy zaznaczyć, że pomysł stworzenia tego typu bazy powiązań słowotwórczych nie jest nowy i podobne bazy powstają również dla innych języków. Przykładem może być stworzony niedawno zasób dla języka rosyjskiego (Kyjánek et al. 2022), włączony do bazy DeriNet. Projekt DeriNet<sup>2</sup> skupia różne sieci derywacyjne dla kilkunastu różnych języków, zbudowane różnymi metodami automatycznymi i półautomatycznymi działającymi z różnym stopniem dokładności, ale sprowadzone do takiego samego formatu i udostępnione we wspólnej wyszukiwarce. Wyczerpujący przegląd innych tego typu baz derywacyjnych można znaleźć w raporcie technicznym (Kyjánek 2018). W zasobach DeriNetu znajduje się również sieć derywacyjna dla polszczyzny (Lango et al. 2021). Została jednak uzyskana metodą wyłącznie automatyczną na podstawie listy form hasłowych słownika Morfeusza (a zatem pewnej zmodyfikowanej wersji listy haseł SGJP), nie zawiera więc żadnej klasyfikacji typów połączeń derywacyjnych.

Proponowaną w niniejszym artykule sieć derywacyjną odróżniają od innych tego typu zasobów dwie cechy. Po pierwsze, byłaby ona zbudowana z wykorzystaniem obszernej bazy fleksyjnej polszczyzny – taką bazą jest SGJP. Sam pomysł przyjęcia takiego zasobu jako podstawy prac jest bardzo naturalny, ale na drodze takiemu rozwiązaniu w wypadku innych języków stawały niekiedy przeszkody licencyjne. Przy SGJP nie jest to jednak problemem, ponieważ jego dane są ogólnie dostępne na liberalnej licencji typu *open source* i są rozpowszechniane w postaci źródłowego słownika Morfeusza. Po drugie zaś – proponujemy klasyfikację, która nie tylko łączy leksemy powiązane ze sobą słowotwórczo, ale również przypisuje tym połączeniom etykiety klasyfikujące typ takiego powiązania<sup>3</sup>, co ułatwi jego

2 <https://ufal.mff.cuni.cz/derinet>.

3 Warto zauważyć, że ten element koncepcji odróżnia tworzony przez nas zasób od *Słownika gniazd słowotwórczych współczesnego języka ogólnopolskiego* (Vogelgesang 2001; Jadacka et al. 2001; Skar-

przeszukiwanie. Ze względu na naturę zjawisk słowotwórczych klasyfikacja tego typu nie jest jednorodna, ma ona jednak charakter głównie pomocniczy i praktyczny – jej celem jest wsparcie użytkownika SGJP (zarówno indywidualnego czytelnika słownika, jak i programu wykonującego jakieś konkretne zadanie na podstawie jego danych). Punktem wyjścia jest zaś zaczątek takiej sieci derywacyjnej istniejący już w SGJP.

### 3. Informacja derywacyjna w SGJP

Przed wszystkim należy zaznaczyć, że informacje słowotwórcze są w SGJP odnotowywane niesystematycznie, zabrakło także jasnej i pełnej ich kategoryzacji. Jak wskazują sami autorzy,

informacje słowotwórcze sprowadzają się do podania leksemów pozostających z danym w najregularniejszych stosunkach derywacyjnych. Są to dla przymiotników: przymiotniki w stopniu wyższym, regularne przysłówki odprzymiotnikowe, regularne nazwy cech na *ość*, a dla czasowników – imiesłowy przymiotnikowe, odsłownik (rzeczownikowa nazwa czynności) oraz czasownik przeciwnego aspektu tworzący z danym parę aspektową (Saloni 2020: 22).

Pośród wielu innych odsyłaczy przypisywanych leksemom pojawiły się wprawdzie takie związane ze słowotwórstwem (np. wariant żeński, wariant męski, nazwa cechy, zdrobnienie, zgrubienie, wariant odmężowski, wariant odojcowski oraz nazwa mieszkańca), ale na pewno nie nazywają one wszystkich relacji derywacyjnych, których można się dopatrzeć. Co więcej, o ile nazwa cechy zakończona na *-ość* była regularnie podawana przy prawie wszystkich przymiotnikach, o tyle pozostałe uwzględniane w słowniku kategorie słowotwórcze takie jak zdrobnienia czy warianty odmężowskie były oznaczane dość przypadkowo, prawdopodobnie raczej przy okazji uzupełniania innych informacji. Dobrą ilustracją tego, jak do tej pory traktowano informacje słowotwórcze w SGJP, jest fakt, że np. wyrazowi *chałupa* przypisano zdrobnienie *chałupię*, ale nie powiązano go z *chałupką*; jako zdrobnienie *gołębia* opisano wyrażenie *gołębisio*, zabrakło natomiast połączenia z *gołąbką* (przepisano informacje o zdrobnieniach z SJPD, który stanowił pierwotną siatkę haseł dla SGJP, bez łączenia w pary wyrażen niemających odpowiednich adnotacji w SJPD). Zauważamy więc potrzebę z jednej strony usystematyzowania występujących już w SGJP kategorii słowotwórczych, z drugiej zaś – uzupełnienia ich o pomijane do tej pory grupy derywatów.

---

żyński et al. 2004; Skarżyński 2004), który przypisuje wyrazom podstawowym wszystkie wyrazy pochodne, ale jednocześnie nie kategoryzuje derywatów.

#### 4. Materiał badawczy i metody

Aby przygotować naszą propozycję wprowadzenia informacji słowotwórczych do *Słownika*, potrzebowaliśmy odpowiedniego zbioru danych dotyczących derywacji; zdecydowaliśmy się przyjąć listę 2500 najczęściej występujących leksemów z SGJP jako punkt wyjścia do przygotowania takiego zbioru. Pierwszy etap polegał na potraktowaniu obecnych na niej wyrażen jako potencjalnych podstaw słowotwórczych i utworzeniu od nich wszystkich możliwych derywatów. Wyrazów pochodnych poszukiwaliśmy w pierwszym rzędzie w obrębie pełnego *Słownika* (bez ograniczania się do jakiejś konkretnej listy frekwencyjnej) oraz we własnej intuicji językowej, a następnie porównywaliśmy wyniki z danymi uzyskanymi przez Mateusza Langę, Zdenka Žabokrtský'ego i Magdę Ševčíkovą (2021), aby uniknąć możliwych przeoczeń<sup>4</sup>. Powstały w wyniku takiego podejścia zbiór postanowiliśmy (ze względu na charakter informacji w SGJP<sup>5</sup>) usystematyzować przy pomocy jak najbardziej gramatycznych kryteriów.

#### 5. Klasyfikacja derywatów

Wybór klasyfikacji słowotwórczej podyktowany był dwoma czynnikami: powinna ona być w jak najmniejszym stopniu oparta na semantyce oraz być zrozumiała i użyteczna dla typowego użytkownika słownika. Ze względu na te założenia najbardziej użyteczna wydaje się klasyfikacja zaprezentowana w pracy Renaty Grzegorzczukowej (1979); jest to podział co najmniej kojarzony przez większość społeczeństwa, przy najmniej częściowo przyswajany w trakcie edukacji szkolnej i chociaż w zakresie derywatów rzeczownikowych rzeczywiście opiera się na semantyce, to wydaje się, że w przypadku tej akurat grupy wyrazów nie da się tego uniknąć.

Derywaty proponujemy więc dzielić na:

- (1) nazwy czynności (np. *działalność*),
- (2) nazwy cech (*czerni*),
- (3) nazwy wykonawców czynności (*czytelnik*),
- (4) nazwy narzędzi (*czernidło*),

---

4 Warto w tym miejscu wspomnieć, dlaczego mimo istnienia tego *de facto* słowotwórczego słownika proponujemy wprowadzenie informacji dotyczącej derywacji do SGJP: powodowane jest to po pierwsze tym, że zależy nam nie tylko na zauważeniu powiązania między wyrażeniami, ale też na sklasyfikowaniu wiążącej je relacji, a po drugie tym, że grafy Lango, Žabokrtský'ego i Ševčíkovéj (2021) często pokazują połączenia ze słowotwórczego punktu widzenia bezsensowne, np. między słowami *armia* i *Warmia*.

5 Próba oparcia naszych danych na wykładnikach semantycznych spowodowałaby konieczność znaczącego zmodyfikowania istniejącej już siatki haseł motywowanych morfologicznie.

- (5) nazwy obiektów czynności (*czytanka*),
- (6) nazwy rezultatów czynności (*napis*),
- (7) nazwy nosicieli cech (*czarnuch*),
- (8) nazwy miejsc (*czytelnia*),
- (9) nazwy zbiorów (*premierostwo* ‘premier z żoną’),
- (10) nazwy mieszkańców (*Ukrainiec*),
- (11) warianty odmězowskie (*doktorowa*),
- (12) warianty odojcowskie (*doktorówna*)<sup>6</sup>,
- (13) warianty żeńskie (*członkini*),
- (14) warianty męskie (*gwiazdor*),
- (15) zdrobnienia (*córeczka*),
- (16) zgrubienia (*człowieczysko*),
- (17) ekspresywizmy (*córuchna*),
- (18) nazwy istot niedorosłych (*kocię*),
- (19) przymiotniki odrzeczownikowe (*czasowy*),
- (20) przymiotniki odczasownikowe (*czytelny*),
- (21) czasowniki odrzeczownikowe (*datować*),
- (22) czasowniki odprzymiotnikowe (*czerwienić*),
- (23) cechy osłabione (*czarniawy*)<sup>7</sup>.

Jak widać podział Grzegorzyczkowej został minimalnie zmodyfikowany: obiekty i rezultaty czynności (stanowiące u niej jedną kategorię) zostały na potrzeby klasyfikacji rozdzielone, a nazwy pokrewieństwa zostały rozbite na warianty odmězowskie i odojcowskie, ponieważ po pierwsze mają wyraźne wykładniki formalne pozwalające łatwo je odróżnić, a po drugie takie właśnie etykiety były do tej pory używane w SGJP.

## 6. Derywaty odrzucone

Chcielibyśmy, aby proponowane przez nas uzupełnienie SGJP o informacje słowotwórcze obejmowało wyłącznie produktywne typy słowotwórcze, dlatego po dokonaniu wyboru klasyfikacji z pierwotnej listy derywatów należało odrzucić te, które

6 Można by się zastanawiać, czy rejestrowanie wariantów odmězowskich i odojcowskich nie ma obecnie waloru jedynie historycznego, zdecydowaliśmy się jednak (przynajmniej na tym etapie) uwzględnić je z kilku powodów: stosowne odsyłacze istniały już w SGJP, w naszej próbie znalazły się wyrażenia, które dało się tak zakwalifikować, a ponadto wydaje się, że formy tego typu wciąż funkcjonują w języku potocznym, zwykle z odcieniem żartobliwym, ironicznym czy negatywnym.

7 W przypadku typów derywatów, które bywały już oznaczane w SGJP (np. wariant żeński), korzystamy z etykiet ustalonych przez autorów *Słownika*.

są jednostkowymi, indywidualnymi przypadkami, niemieszczącymi się w przyjętych kategoriach, jak np. *doktorant* czy *ramiączko*.

Warto podkreślić, że na tym etapie proponujemy zająć się wyłącznie formami opartymi na jednym temacie słowotwórczym, co wyklucza z naszych badań derywaty pochodzące od wyrażen syntaktycznych (np. *poddasze*, *bezdrożę*, *narożnik*) oraz złożenia (np. *rurociąg*, *czarnoziem*).

Istotną grupą, z której zdecydowaliśmy się (przynajmniej na tym etapie) zrezygnować, są derywaty czysto prefiksalne (głównie czasowniki). Było to spowodowane po pierwsze tym, że prefiksy nie zmieniają wzoru odmiany wyrażenia (co oznacza, że z punktu widzenia gramatyki są mniej interesujące niż sufiksy), a po drugie tym, że zmiany znaczenia wprowadzane przez prefiksy są na tyle różnorodne, że nie dają się ująć w wystarczająco ogólne klasy. Przykłady takiego rozdrobnienia przy próbie semantycznego podejścia do derywacji prefiksальной możemy znaleźć w pracy Małgorzaty Kity (1998), gdzie poszczególnym prefiksom przypisano od dwóch do dwunastu znaczeń, np. derywaty z prefiksem *nad(e)-* są charakteryzowane jako: 1. opisujące zbliżanie się do celu, 2. wskazujące na akcję przekraczającą spodziewany rozmiar, 3. zrobienie czegoś dodatkowego, co nie jest nadmiarem takiej samej czynności zasadniczej (Kita 1998: 64).

Grzegorzyczkowa zaznacza, że

od każdego czasownika tworzy się równoległe duża liczba derywatów prefiksalnych o różnych znaczeniach, co daje w rezultacie obraz skomplikowany, nadający się raczej do opisu słownikowego, aniżeli gramatycznego (Grzegorzyczkowa 1979: 78).

Potwierdzają to wyniki naszych obserwacji: przykładowo od leksemu *budować* możemy wprowadzić wyrażenia *zabudować*, *przebudować*, *wybudować*, *rozbudować* i *dobudować*, które choć semantycznie odmienne, gramatycznie zachowują się zupełnie identycznie. Ich forma wskazuje zaś na powiązanie z podstawą słowotwórczą na tyle wyraźnie, że choć można było powiązać te czasowniki bardzo ogólną i niewiele mówiącą etykietą „czasownik prefiksalny”, wydawało się to równie nieużyteczne jak zagłębianie się w drobiazgowo rozróżnienia semantyczne, więc zdecydowaliśmy się w naszej propozycji zupełnie pominąć derywaty czysto prefiksalne.

## 7. Klasyfikacja derywatów – rozstrzygnięcia szczegółowe

Zasadnicza większość derywatów utworzonych od 2500 najczęstszych leksemów z SGJP mieści się bez kontrowersji w jednej z wyznaczonych grup, niemniej jednak przy klasyfikacji pewnych form pojawiły się wątpliwości dotyczące granic między niektórymi grupami oraz kierunków motywacji słowotwórczej. Nie będziemy się więc skupiać na przypadkach prostych i niewymagających dyskusji, przybliżymy natomiast te kwestie, które wymagały rozstrzygnięcia.



## 7.1. Zdrobnienia, zgrubienia, ekspresywizmy

Jednym z największych problemów, z którymi przyszło nam się zmierzyć podczas klasyfikowania wyrażeń pochodnych słowotwórczo, było ustalenie granic pomiędzy zdrobnieniami, zgrubieniami i ekspresywizmami, zwłaszcza że etykiety „zdrobnienie” i „zgrubienie” funkcjonują już w SGJP. Warto wspomnieć, że do tej pory w *Słowniku* oznaczono zaledwie 25 zgrubień; zdrobnień jest zdecydowanie więcej, bo prawie 1200, niemniej jednak większość z nich stanowią imiona, więc konieczne było zarówno przyjrzenie się temu, jakie zdrobnienia odnotowano do tej pory i czy zrobiono to właściwie, jak i zastanowienie się nad tym, o jakie wyrażenia należałoby tę listę uzupełnić.

Grzegorzyczkowa widzi w przypadku zdrobnień, zgrubień i ekspresywizmów nie trzy, a dwie kategorie: jedną są zdrobnienia, drugą tworzą zgrubienia i ekspresywizmy (Grzegorzyczkowa 1979: 53–56). W pierwszej grupie formant wnosi znaczenie małości przedmiotu nazwanego podstawą, ale najczęściej na to znaczenie nakładają się również elementy ekspresywne: *piesek* może być i ‘małym psem’, i ‘psem opisywanym z (pozytywnym) uczuciem’, i ‘młodym psem’ albo połączeniem wszystkich tych znaczeń, a zatem semantyka nie jest tu szczególnie pomocna, bo gdyby chcieć się w nią zagłębiać, trzeba by było te same derywaty umieszczać w dwóch różnych kategoriach w zależności od znaczenia. Przy odróżnianiu zdrobnień od ekspresywizmów lepszym wyznacznikiem tych drugich są formanty używane do ich tworzenia – są to wyłącznie cztery sufiksy: *-ik / -ek* dla rodzaju męskiego, *-ka* dla rodzaju żeńskiego oraz *-ko* dla rodzaju nijakiego; sporadycznie występują jeszcze sufiksy komponowane typu *-uszek* (*wianuszek*) i *-iczka / -yczka* (*twarzyczka*) (Grzegorzyczkowa 1979: 53), *-iszek* (*braciszek*), *-aszek* (*kijaszek*), *-uszko* (*jabłuszko*) i *-uszka* (*kauczuszka*) (Masojeć 2015: 100). Przyjęcie takiego rozwiązania zrodziło jednak problem związany ze zdrobnieniami odnotowanymi już w SGJP, ponieważ większość zdrobnień imion nie odpowiada takim wymaganiom: o ile *Agatkę* jako zdrobnienie *Agaty* można zaakceptować, o tyle *Agunia*, *Agusia*, *Aga* i *Jaga* wpisują się już raczej w ekspresywizmy. W razie przyjęcia naszej propozycji konieczne byłoby więc nie tylko wprowadzenie powiązań pomiędzy zdrobnieniami i ich podstawami w przypadku rzeczowników pospolitych, ale też wprowadzenie poprawek na liście istniejących już w SGJP połączeń.

Zgrubienia i ekspresywizmy często uznawane są za jedną kategorię: na przykład w książce Grzegorzyczkowej (1979: 55–56) jest to jedna grupa określana etykietą „zgrubienia i ekspresywizmy”, z kolei w pracy Kity (1998: 10) ujęte zostały jako „nazwy ekspresywne” (dzielone potem na „dodatnie”, „ujemne” – w których zdają się mieścić zgrubienia – i „różne”). W innych klasyfikacjach zgrubienia i ekspresywizmy bywają rozdzielane na dwie kategorie: „nazwy zgrubiałe” i „nazwy ekspresywne” – w książce Ireny Masojeć (2015: 100–108), „nazwy augmentatywne” i „nazwy ekspresywne” w pracy Grzegorzyczkowej i Jadwigi Puzyniny (1984b: 427–429), niemniej jednak na-



wet w pracach, w których teoretycznie stanowią one dwie odrębne grupy, charakteryzuje się je częściowo za pomocą tych samych formantów: np. sufiks *-idło* opisywany jest jako w zasadzie sygnalizujący negatywne emocje nadawcy (*wierszydło*), ale mogący również mieć znaczenie augmentatywności (*ptaszzydło*), w obu grupach pojawiają się formy z sufiksami *-sko* (np. *gmaszysko*, *babsko*), *-al* (np. *nochal*, *synal*) czy derywaty z ucięciami (np. *ogór*, *ciacho*) (ibid.: 427–429). Trudno więc powiedzieć, gdzie według autorek miałaby przebiegać granica pomiędzy nazwami augmentatywnymi a nazwami ekspresywnymi; wydaje się, że albo był to podział motywowany wyłącznie semantyką, albo wręcz podział pozorny, wymagający kontekstowego rozstrzygnięcia, czy ważniejszy jest element ‘duży rozmiar’ czy ‘stosunek nadawcy’.

Zgrubienia niewątpliwie zawierają poza znaczeniem dużego rozmiaru przedmiotu także elementy ekspresywne, oznaczające zwykle negatywne (bądź kontekstowo żartobliwe czy serdeczne) ustosunkowanie nadawcy do desygnatu podstawy lub sytuacji czy odbiorcy (ibid.: 427), można się więc zastanawiać, czy nie uznać ich za podgrupę nazw ekspresywnych. Ze względu na to, że jest to jednak kategoria dość mocno utrwalona w świadomości odbiorców, w dodatku z już wprowadzoną do SGJP etykietą, proponujemy podobne rozwiązanie jak w przypadku zdrobnień: zakładamy, że zgrubieniem jest każdy derywat utworzony za pomocą sufiksów *-sko*, *-isko*, *-al*, *-uch* oraz derywaty z ucięciem. Mówimy tu jednak wyłącznie o sytuacjach wymagających rozstrzygnięcia, czy dany derywat mieści się w grupie zgrubień czy ekspresywizmów; oczywiście jest, że podane tu sufiksy używane są również do tworzenia innych typów derywatów.

Przyrostek *-idło* (bardzo często również zaliczany do sufiksów typowych dla zgrubień) scharakteryzowany został jako ukazujący przede wszystkim negatywny stosunek nadawcy do przedmiotu nazywanego podstawą (ibid.: 427), a przykłady jego występowania w SGJP (można wymienić choćby *piśmidło*, *filmidło*, *wierszydło*, *melodramidło*) również świadczą o tym, że jest to jego znaczenie podstawowe. Odnotowano zaledwie dwie formy wskazujące na duży rozmiar obiektu: *wąsidło* i *ptaszzydło*, a że i w ich znaczeniu widać wyraźnie obecność komponentu emocjonalnego, utworzone za pomocą sufiksu *-idło* derywaty proponujemy zaliczać do ekspresywizmów, nie do zgrubień<sup>8</sup>.

Stosunkowo nowym wykładnikiem ekspresywności jest przyrostek *-ol* (służący do wyrażania silnych negatywnych emocji, niechęci i pogardy), któremu towarzyszy ucięcie części podstawy, np. *Anglik* → *Angol* (Masojć 2015: 106).

8 Tu także należałoby podkreślić, że sufiks *-idło* może być wykorzystywany do tworzenia innych typów derywatów (nie tylko ekspresywizmów). Nie postulujemy automatycznego kategoryzowania każdego wyrazu pochodnego z tym sufiksem jako ekspresywizmu, a jedynie stwierdzamy, że w grupie wyrażen wymagających rozstrzygnięcia, czy są zgrubieniami czy ekspresywizmami, te utworzone poprzez dodanie sufiksu *-idło* uznajemy za ekspresywizmy.

Ponadto odsyłacz „ekspresywizm” przypisaliśmy w naszej propozycji również wyrażeniom z sufiksem *-ina*, wskazującym na lekceważenie czy politowanie nadawcy wobec tak określanego przedmiotu (*babina*, *głowina*, *koszulina*). Bardzo liczną podgrupę ekspresywizmów stanowią spieszczenia (które przez użytkowników języka często zaliczane są do zdrobnień), tworzone za pomocą np. przyrostków: *-uś* (*tatuś*), *-iś* (*koniś*), *-ula* (*matula*), *-cia* (*mamcia*), *-unio* (*dziadunio*), a także poprzez „pewien typ derywacji wstecznej (dezintegralnej), polegającej na skróceniu tematu i wprowadzeniu na jego miejsce elementu *ch* i *ś*, np. *Stach*, *Stas*” (Grzegorzycowa 1979: 56). To właśnie do tej podgrupy należałoby według nas przenieść znaczną większość form, które do tej pory były oznaczone w SGJP jako zdrobnienia imion (np. *Anula*, *Anusia*, *Aneczka*), zwłaszcza że Grzegorzycowa zaznacza, iż spieszczenia tworzone są przede wszystkim od nazw własnych; rzeczowniki pospolite, od których daje się je wyprowadzić, to głównie nazwy członków rodziny (*mamusia*, *babunia*, *córuchna*) oraz nazwy pewnych zjawisk mogących wywołać postawę pieszczośliwą (*nózia*, *brzusio*) (ibid.).

## 7.2. Przymiotniki odprzymiotnikowe

W przyjętej klasyfikacji derywatów wprowadziliśmy jeszcze jedną zmianę dotyczącą przymiotników odprzymiotnikowych. Grzegorzycowa dzieli je na oznaczające wzmocnienie i osłabienie cechy, przy czym zaznacza, że wśród tych pierwszych istnieje duża grupa, w której sufiks wnosi nie tyle znaczenie intensyfikacji, ile informację o nadawcy, jego stosunku emocjonalnym czy stylu mówienia (ibid.: 69–70). *Głupiutki* to przecież nie ‘bardzo głupi’, a raczej ‘taki, którego głupotę nadawca traktuje pobłażliwie, z politowaniem’; *maluchny* może i jest ‘bardzo mały’, ale i tu na pierwszy plan wybijają się emocje nadawcy bądź charakterystyczny dla niego sposób mówienia. Co więcej, *nagutki* nie może być ‘bardzo nagi’, bo nagość nie jest cechą stopniowalną, a zatem formant jest tutaj wyłącznie wyrazem uczuć mówiącego. Dodawanie sufiksów ekspresywnych do przymiotników niestopniowalnych wnosi znaczenie wyłącznie ekspresywne, bez intensyfikacji cechy (Kallas 1984: 507), a zatem przymiotników takich jak *calutki* czy *pewniutki* nie możemy opatrzyć etykietą „cecha wzmocniona”. Warto podkreślić, że o ile Grzegorzycowa uznaje, że derywaty oznaczające intensywność cechy tworzone są za pomocą prefiksów i sufiksów (Grzegorzycowa 1979: 69), o tyle Krystyna Kallas twierdzi, że „w funkcji intensyfikującej występują tylko formanty prefiksalne” (Kallas 1984: 506), takie jak: *nad-*, *hiper-*, *prze-*, *arcy-*, *ekstra-*, *super-*. Formy utworzone za pomocą sufiksów (typu *-utki*, *-uchny*, *-uśki* itp.) Kallas zalicza do derywatów ekspresywnych; wskazuje przy tym, że łączą one wysoki stopień intensywności z nacechowaniem ekspresywnym (ibid.: 506–507).

Początkowo planowaliśmy dzielić przymiotniki odprzymiotnikowe na trzy grupy: cechy wzmocnione, cechy osłabione oraz ekspresywizmy, jednak ze względu na

to, że – zwłaszcza w przypadku derywatów z formantem *-utki* – musielibyśmy dokonywać rozróżnień między intensyfikacją cechy a ekspresywizmem na podstawie semantyki, czego w SGJP zawsze starano się unikać, postanowiliśmy przychylić się do rozwiązania zaproponowanego przez Kallas: cechy osłabione proponujemy pozostawić bez zmian (głównie sufiks *-awy*, np. *białawy*, *grubawy*), etykietę „ekspresywizm” przypisać formom utworzonym za pomocą sufiksów, zaś za wyznacznik cech wzmocnionych uznać prefiksy, więc (jako że na tym etapie wyłączyliśmy derywaty prefiksalne z naszej propozycji zakresu danych do wprowadzenia do słownika) ostatecznie bylibyśmy skłonni dodać tylko dwa pierwsze odsyłacze i wstrzymać się z wprowadzeniem intensyfikacji cech do czasu, kiedy znajdziemy odpowiednie, bazujące na gramatycznych kryteriach podejście do derywacji prefiksalnej.

### 7.3. Wzajemna motywacja słowotwórcza

Ostatnią kwestią, którą należało rozstrzygnąć przy klasyfikowaniu derywatów, było ustalenie, co zrobić z przypadkami tzw. motywacji wzajemnej. W standardowych przypadkach przy ustalaniu kierunku motywacji słowotwórczej kierujemy się dwoma kryteriami – formalnym i semantycznym: w formie wyrazu pochodnego zawarta jest forma podstawy, a jego znaczenie daje się opisać z użyciem podstawy, np. w parze wyrażeń *lekarz* i *leczyć* derywatem jest *lekarz*, formalnie wzbogacony o sufiks *-arz* i objaśniany jako ‘ten, kto leczy’ (nie powiedzielibyśmy przecież, że *leczyć* to ‘wykonywać pracę lekarza’) (Grzegorzczkowska 1979: 15; Nagórko 2007: 184–185).

Problemy z ustaleniem kierunku motywacji słowotwórczej rodzą się w sytuacji, kiedy pojawia się konflikt między kryterium formalnym a kryterium znaczeniowym (forma i znaczenie wskazują na przeciwne kierunki motywacji), bądź w wypadku neutralizacji jednego z tych kryteriów (forma lub znaczenie obu wyrażeń są równie złożone) (Nagórko 2007: 185). To pierwsze zjawisko zaobserwować możemy na przykładzie pary *fryzjer* – *fryzjerstwo* (w zasadzie całej serii nazw – powiedzmy – zajęć oraz ich wykonawców): forma *fryzjerstwo* jest bogatsza o sufiks *-stwo*, natomiast wyrażenie *fryzjer* ma zawarty w znaczeniu dodatkowy komponent osoby, wykonawcy czynności. Drugi przypadek to zestawienia typu  *optymizm* – *optymista*. Obie formy są tak samo złożone (można by się tu dopatrywać jedynie wymianę sufiksu *-izm* na *-ista*), a także występują w swoich parafrazach: *optymizm* to ‘cecha optymisty’, zaś *optymista* to ‘ktoś, kogo charakteryzuje optymizm’. Z semantycznego punktu widzenia takie błędne koło definicyjne byłoby nie do przyjęcia, jednak dobrze oddaje ono sposób, w jaki użytkownicy języka postrzegają relacje między wyrazami (ibid.). Warto podkreślić, że również pary z pierwszej grupy dają się definiować za pomocą drugiego członu zestawu: *fryzjerstwo* jest ‘zajęciem fryzjera’, *fryzjer* to ‘ktoś zajmujący się fryzjerstwem’. Tam jednak, gdzie bogatsza jest forma, uznajemy kryterium formalne za równie ważne jak semantyczne, więc przy zbliżonym stopniu

skomplikowania znaczenia proponujemy wskazywać kierunek motywacji od formy mniej złożonej do formy bogatszej (czyli *fryzjer* → *fryzjerstwo* czy *kolarz* → *kolarstwo*) i traktować taki derywat jak każdą inną nazwę cechy (por. Grzegorzczukowa 1979: 36; Grzegorzczukowa, Puzynina 1984a: 386).

W kwestii par takich jak  *optymizm* –  *optymista* czy  *polityka* –  *polityk* uznajemy, że rzeczywiście mamy tu do czynienia z motywacją wzajemną, a w zasadzie z adaptowanymi na gruncie języka polskiego internacjonalizmami (a zatem wyrażeniami zapożyczonymi jednocześnie, nie zaś wynikami polskiego słowotwórstwa), których „morfologiczna interpretacja – choć uprawniona – pozostaje wtórna” (Nagórko 2007: 186). Na potrzeby klasyfikacji zakładamy kierunek od cechy abstrakcyjnej do nosiciela tejże (czyli  *optymizm* →  *optymista*,  *polityka* →  *polityk*), zaś na poziomie prezentacji danych w słowniku w takich przypadkach oba elementy z pary można by ukazywać jako derywaty od drugiego jej członu: przykładowo pod hasłem  *optymista* umieścić odsyłacz „nosiciel cechy” do  *optymizmu* i odwrotnie – pod hasłem  *optymizm* dostawić odsyłacz „nazwa cechy” do  *optymisty*.

## 8. Podsumowanie eksperymentu

W tabeli poniżej przedstawiamy wyniki klasyfikacji derywatów utworzonych od 2500 najczęściej używanych leksemów z SGJP. Warto zauważyć olbrzymią dysproporcję między poszczególnymi kategoriami: z jednej strony otrzymaliśmy 1053 przymiotniki odrzeczownikowe, z drugiej zaś zaledwie jeden wariant męski ( *gwiazdor*<sup>9</sup>). Zdecydowaną przewagę przymiotników odrzeczownikowych tłumaczy fakt, że rzeczowniki stanowią prawie połowę listy wyrażen przyjętych przez nas za potencjalne podstawy słowotwórcze (jest ich 1229 na 2500 słów) i od prawie każdego z nich da się utworzyć przymiotnik<sup>10</sup>.

Do opracowania wybrano jednostki najczęstsze, żeby uchwycić derywację serijną i produktywną, a także przykłady zjawisk nietypowych. Gdybyśmy bowiem wyszli od próbki losowej z SGJP, to dostalibyśmy sporo leksemów rzadkich, które albo w ogóle nie wchodzą w relacje derywacyjne, albo robią to bardzo przewidywalnie. Jednocześnie trzeba mieć świadomość, że opracowany materiał nie jest dobrą próbą statystyczną dla częstości typów odsyłaczy. Przykładowo warianty od-mężzowski/odojcowski istnieją dla mnóstwa nazwisk, których w naszej próbie po prostu nie ma.

9 Nie jest to może tak trafny przykład wariantu męskiego jak np.  *pielęgniarz*, niemniej jednak był on jedynym występującym w naszej próbie.

10 Niekiedy od jednej podstawy rzeczownikowej daje się utworzyć dwa różne przymiotniki, np.  *ciało* →  *cielisty*,  *ciało* →  *cielesny*.

Tabela 1. Podsumowanie wyników eksperymentu

Typ derywatu	Liczba derywatów	Części mowy	Przykłady
Cecha osłabiona	29	adj→adj	<i>brudnawy</i>
Czasownik odprzymiotnikowy	75	adj→v	<i>szarzyć</i>
Czasownik odrzeczownikowy	197	subst→v	<i>skórować</i>
Ekspresywizm	205	subst→subst, adj→adj	<i>wioszczyzna, szareńki</i>
Nazwa cechy	59	subst→subst, adj→subst	<i>autorstwo, zdrowie</i>
Nazwa czynności	92	v→subst	<i>groźba</i>
Nazwa istoty niedorosłej	4	subst→subst	<i>kocię</i>
Nazwa miejsca	43	v→subst, subst→subst	<i>palarnia, piwiarnia</i>
Nazwa mieszkańca	38	subst→subst	<i>Francuz</i>
Nazwa narzędzia	26	v→subst, subst→subst, adj→subst	<i>otwieracz, papierośnica, czernidło</i>
Nazwa zbiorowości	30	subst→subst, adj→subst	<i>klientela, młodzież</i>
Nosiciel cechy	147	subst→subst, adj→subst, v→subst	<i>brodacz, średniak, plaksa</i>
Obiekt czynności	21	v→subst	<i>gra</i>
Przymiotnik odczasownikowy	37	v→adj	<i>przewidywalny</i>
Przymiotnik odrzeczownikowy	1053	subst→adj	<i>życiowy</i>
Rezultat czynności	12	v→subst	<i>twór</i>
Wariant męski	1	subst→subst	<i>gwiazdor</i>
Wariant odmężowski	12	subst→subst	<i>prezesowa</i>
Wariant odojcowski	8	subst→subst	<i>marszałkówna</i>
Wariant żeński	90	subst→subst	<i>przeciwniczka</i>
Wykonawca czynności	130	v→subst, subst→subst	<i>organizator, rybak</i>
Zdrobnienie	479	subst→subst	<i>żarcik</i>
Zgrubienie	75	subst→subst	<i>aktorzyśko</i>

## 9. Zakończenie

Opracowana sieć derywatów dla najczęstszych jednostek SGJP pokazuje, że zaproponowana klasyfikacja typów relacji derywacyjnych nie ma istotnych pominięć. Przygotowane przez nas dane warto byłoby włączyć do SGJP – dotyczą one jednostek częstych, mogą więc być użyteczne dla użytkowników słownika.

Jak wspomniano wcześniej, drugim celem przedstawionej tu pracy było przygotowanie danych treningowych dla algorytmu, który wykrywałby relacje derywacyjne. Ze względu na dużą dysproporcję częstości między typami derywatów konieczne będzie ręczne rozszerzenie danych w taki sposób, aby uwzględnić typy derywatów o niskiej częstości w pierwszej próbie. Powinno to pozwolić na uzyskanie modelu komputerowego radzącego sobie z rozpoznawaniem wszystkich zaproponowanych typów. Model taki zamierzamy wykorzystać do przygotowania większej sieci derywatów, które, po sprawdzeniu przez człowieka, mogłyby być włączone do słownika.

Jednocześnie warto zauważyć, że nawet takie rozszerzenie słownika nie wyczerpie tematu uzupełniania SGJP o dane słotwórcze, ponieważ – jak wskazywaliśmy już wcześniej – pomijamy w niniejszym tekście kwestię derywacji prefiksальной oraz derywaty pochodzące od więcej niż jednej podstawy, tak więc oba te zagadnienia można uznać za zadania na przyszłość domagające się szczegółowych badań.

## Literatura

- GRZEGORCZYKOWA R., 1979, *Zarys słotwórstwa polskiego. Słotwórstwo opisowe*, wyd. 3, Warszawa.
- GRZEGORCZYKOWA R., PUZYNNINA J., 1984a, *Problemy ogólne słotwórstwa*, [w:] R. Grzegorzczkowa, R. Laskowski, H. Wróbel (red.), *Gramatyka współczesnego języka polskiego. Morfologia*, Warszawa, s. 361–388.
- GRZEGORCZYKOWA R., PUZYNNINA J., 1984b, *Rzeczownik*, [w:] R. Grzegorzczkowa, R. Laskowski, H. Wróbel (red.), *Gramatyka współczesnego języka polskiego. Morfologia*, Warszawa, s. 389–468.
- JADACKA H. et al., 2001, *Słownik gniazd słotwórczych współczesnego języka ogólnopolskiego*, t. II: *Gniazda odrzeczownikowe*, Kraków.
- KALLAS K., 1984, *Przymiotnik*, [w:] R. Grzegorzczkowa, R. Laskowski, H. Wróbel (red.), *Gramatyka współczesnego języka polskiego. Morfologia*, Warszawa, s. 469–523.
- KIERAŚ W., WOLIŃSKI M., 2017, *Morfeusz 2 – analizator i generator fleksyjny dla języka polskiego*, „Język Polski” XCVII, s. 75–83, <https://doi.org/10.31286/JP.97.1.7>.
- KITA M., 1998, *Wybieram gramatykę! Gramatyka języka polskiego w praktyce (dla cudzoziemców zaawansowanych)*, t. 2, Katowice.
- KYJÁNEK L., 2018, *Morphological Resources of Derivational Word-Formation Relations. Technical Report TR-2018-61*, Faculty of Mathematics and Physics, Charles University, Prague.
- KYJÁNEK L., LYASHEVSKAYA O., NEDOLUZHKO A., VODOLAZSKY D., ŽABOKRTSKÝ Z., 2022, *Constructing a Lexical Resource of Russian Derivational Morphology*, [w:] N. Calzolari,



- F. Béchet, P. Blache, K. Choukri, Ch. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, J. Odijk, S. Piperidis (red.), *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, Marseille, s. 2788–2797.
- LANGO M., ŻABOKRTSKÝ Z., ŠEVČÍKOVÁ M., 2021, *Semi-automatic Construction of Word-formation Networks*, „Language Resources & Evaluation” 55, s. 3–32, <https://doi.org/10.1007/s10579-019-09484-2>.
- MASOJĆ I., 2015, *Gramatyka współczesnego języka polskiego. Słowotwórstwo rzeczowników*, Wilno.
- MROCKOWSKI R., Rybak P., Wróblewska A., Gawlik I., 2021, *HerBERT: Efficiently Pre-trained Transformer-based Language Model for Polish*, [w:] *Proceedings of the 8th Workshop on Balto-Slavic Natural Language Processing*, Kiyv, s. 1–10, <https://doi.org/10.48550/arXiv.2105.01735>.
- NAGÓRKO A., 2007, *Zarys gramatyki polskiej*, Warszawa.
- SALONI Z., 2020, *Podstawy teoretyczne Słownika gramatycznego języka polskiego*, wyd. 4, Warszawa.
- SKARŻYŃSKI M. et al., 2004, *Słownik gniazd słowotwórczych współczesnego języka ogólnopolskiego*, t. III: *Gniazda odczasownikowe*, Kraków.
- SKARŻYŃSKI M., 2004, *Słownik gniazd słowotwórczych współczesnego języka ogólnopolskiego*, t. IV: *Gniazda motywowane przez liczebniki, przysłówki, zaimki, przyimki, modulatory, onomatopoeje, wykrzykniki*, Kraków.
- SGJP: M. Woliński, Z. Saloni, R. Wołosz, W. Gruszczyński, D. Skowrońska, Z. Bronk, *Słownik gramatyczny języka polskiego*, wyd. 4, Warszawa 2020, [on-line:] <http://sgjp.pl>.
- VOGELGESANG T., 2001, *Słownik gniazd słowotwórczych współczesnego języka ogólnopolskiego*, t. I: *Gniazda odprzymiotnikowe*, red. H. Jadacka, Kraków.

## **Derivational Network on the Basis of *Słownik gramatyczny języka polskiego*:**

### **Suggestion for Classification**

#### **Abstract**

The article presents a suggestion of adding a derivational network to the *Słownik gramatyczny języka polskiego* (SGJP, Grammatical Dictionary of Polish). To date, SGJP regularly noted selected word-building relations (e.g., names of features created by adding the ending *-ość* to adjectives), several (such as diminutives) were marked inconsistently and many were not introduced there at all. To fill this gap, the authors adjusted the classification of derivatives developed by Renata Grzegorzczkova to the needs of SGJP (with slight modifications discussed in the first part of the text). Next, an experiment was conducted where lexemes from a list of 2500 most common SGJP words were connected with their derivatives and the latter were then classified according to the assumed rules.