

JERZY GASZEWSKI  
Instytut Anglistyki  
Uniwersytet Łódzki, Łódź  
e-mail: jgaszewski@op.pl  
ORCID 0000-0002-9313-6454

## WALENCJA W EUROPIE (ŚRODKOWEJ) – AREALNE SCHEMATY DYSTRYBUCYJNE<sup>1</sup>

---

### VALENCY IN (CENTRAL) EUROPE – AREAL PATTERNS OF DISTRIBUTION

**ABSTRACT:** The text explores the question of similarities in valency position marking among Central European languages. The issue is approached empirically – with an extensive database of valency patterns. To avoid bias in favour of Central European patterns, other languages are also included in the comparison. The descriptive statistical method used here is affinity analysis. The article deals in detail with three groups of corresponding valency markers in the case of which one can claim Central European areal patterns. The semantics behind the observed patterns is also explored, pointing to otherwise unnoticed features that seem to be reflected in the grammatical marking.

**KEYWORDS:** European languages, cases, prepositions, semantic microroles, affinity analysis

**SŁOWA KLUCZOWE:** języki europejskie, przypadki, przyimki, mikrorole semantyczne, analiza koszykowa

---

<sup>1</sup> Niniejszy tekst powstał w ramach projektu finansowanego przez Narodowe Centrum Nauki (nr wniosku 2016/20/S/HS2/00285). Za wnikliwą lekturę wcześniejszej wersji dziękuję Joannie Błaszczak (Uniwersytet Wrocławski). Za pomoc w odkrywaniu zastosowanych tu metod analizy statystycznej ogromnie dziękuję Adamowi Idczakowi (Uniwersytet Łódzki).

## Uwagi wstępne

W pracach porównawczych dotyczących języków Europy Środkowej postuluje się podobieństwa w oznaczaniu pozycji walencyjnych jako cechę łączącą języki regionu (np. Balázs 1983, s. 103–4; Pusztay 1996, s. 20–21; Newerkla 2000, s. 11; Bláha 2015, s. 156–157). Zdarzają się i głosy zaprzeczające istnieniu takich arealnych prawidłowości (np. Pilarský 2001, s. 118). Brak zgodności między źródłami to jeden powód, dla którego warto po raz kolejny spojrzeć na ten temat. Ponadto, przywołane prace w większości opierają twierdzenia nt. walencji na wyliczeniu krótkich list przykładów. Takie zestawienia należy więc traktować raczej jako wstępne obserwacje, a nie ilustrację dokładnej analizy. Ogólnie rzecz biorąc, można spodziewać się, że sposoby formalnego oznaczania walencji wykazują podobieństwa arealne.<sup>2</sup> Wymagania rekcyjne czasowników mają w pewnym stopniu charakter idiomatyczny, a idiomy oraz ogólnie frazeologia i leksyka należą do elementów systemu językowego łatwo ulegających wpływom w sytuacjach kontaktu (Aikhenvald 2006, s. 5; Hickey 2017, s. 6). Badania Sergeya Saya (2014, 2018) wykazują pozytywne korelacje między podobnym użyciem wykładników pozycji walencyjnych a bliskością języków w sensie zarówno arealnym, jak i genetycznym. Jeśli chodzi o przynależność konkretnych czasowników do klas walencyjnych, efekt podobieństwa daje się zauważyć między językami w ramach podrodzin, np. językami słowiańskimi, oraz w dość szerokim promieniu bliskości geograficznej (nawet do 4 tys. km), nie ma go natomiast przy dalszym pokrewieństwie, np. między językami różnych gałęzi rodziny indoeuropejskiej (Say 2018, s. 599–601).

W niniejszym szkicu przedstawiam metodę kompleksowego i empirycznego spojrzenia na ten problem badawczy oraz podaję trzy konkretne przykłady potwierdzonych podobieństw w walencji w językach środkowoeuropejskich. Cel badania można rozwinąć w trzy pytania:

- jakie konkretnie formy gramatyczne odpowiadają sobie jako wykładniki argumentów czasowników?
- jakie języki są związane tymi podobieństwami? Czy możemy doszukać się charakterystycznych schematów środkowoeuropejskich?
- czy zaobserwowane podobieństwa mają jakiś klucz semantyczny?

W badaniach arealnych można rozróżnić za Lylem Campbellem (2017, s. 24) dwie perspektywy. Pierwsza zestawia obserwowane cechy strukturalne języków i poszukuje w ten sposób podobieństw arealnych. Druga natomiast poszukuje

<sup>2</sup> W pracy posługuję się określeniem *arealny* odpowiadającym ang. *areal (linguistics)*, w starszych źródłach polskojęzycznych występuje (raczej rzadko używane w pracach badawczych) określenie (*klasyfikacja*) *geograficzna* (np. EJO 1993; Majewicz 1989).

historycznych uwarunkowań, które dowodzą, że dane podobieństwa należy wywodzić z kontaktów między językami, tzn. faktycznie mają one charakter arealny. Nietrudno zgodzić się z L. Campbellem, który wyżej ceni drugie podejście. Taki podział można jednak postrzegać jako rozróżnienie kolejnych etapów badania – najpierw należy stwierdzić układ podobieństw między badanymi językami (co niekiedy jest banalne, ale w innych przypadkach – samo w sobie stanowi wyzwanie), a dopiero potem można analizować tło historyczne. Ze względu na niezbyt zaawansowany stan dotychczasowych badań, w niniejszym tekście pozostaję na pierwszym etapie.

### Kwestie metodologiczne

Europa Środkowa z pewnością nie jest regionem o jasno zdefiniowanych granicach. Timothy Garton Ash podaje, że geograf austriacki Karl Sinnhuber porównał w latach 50. kilkanaście definicji Europy Środkowej, by stwierdzić ogromne rozbieżności między nimi. Jediną częścią kontynentu europejskiego, który nigdy nie był zaliczany do Europy Środkowej okazał się Półwysep Iberyjski, natomiast obszar zawsze zaliczany do regionu to jedynie Austria, Czechy i Morawy (za Janušką 2020, s. 56). Badane przez K. Sinnhubera definicje opierały się zapewne na kryteriach politycznych i kulturowych. W pracach językoznawczych wariacje nt. Europy Środkowej nie są może aż tak efektowne, ale z pewnością nie ma ustalonego kanonicznego zakresu regionu.

Najważniejszym postulowanym zgrupowaniem języków regionu jest tzw. liga dunajska (niem. *Donaubund*), na którą składają się języki dawnej monarchii habsburskiej tj. czeski, niemiecki<sup>3</sup> (czy wręcz austriacka odmiana niemieckiego), słowacki oraz węgierski. Językami peryferyjnymi ligi są na północy język polski, na południu – chorwacki oraz słoweński (Kurzová 1996, s. 58; Bláha 2015, s. 19; Januška 2020, s. 69). Inną istotną koncepcją jest liga europejska (ang. *Standard Average European*), której rdzeń tworzą germańskie i romańskie języki kontynentalnej części Europy Zachodniej (Haspelmath 2001, s. 1505; Heine, Kuteva 2006, s. 7–9). Języki środkowoeuropejskie (poza niemieckim) są peryferyjnymi członkami takiej grupy. Gyula Décsy (1973) postuluje obok ligi dunajskiej także ligę „rokytnicką” obejmującą języki dawnej Rzeczypospolitej: białoruski, litewski, polski oraz ukraiński, nie są mi jednak znane żadne późniejsze prace, które odnosiłyby się do tak określonego regionu<sup>4</sup>.

<sup>3</sup> Gyula Décsy (1973) nie wlicza niemieckiego do ligi dunajskiej, mimo jego oczywistego wpływu na pozostałe języki.

<sup>4</sup> Pomijam odniesienia polegające na referowaniu propozycji G. Décsy’ego (np. Balázs 1983; Majewicz 1989; Pusztay 1996).

Ostatecznie kwestią zasadniczą jest faktyczne występowanie jakiegoś podobieństwa strukturalnego między sąsiadującymi językami. Opisując zasięg takiej cechy wyznacza się obszar, z którym jest związana. Odnosząc się do wcześniejszych prac językoznawczych, możemy stwierdzić, na ile tak określony region pasuje do dotychczasowego wyobrażenia Europy Środkowej.

Oddzielną sprawą jest wybór języków, które ma uwzględniać analiza porównawcza. Wiarygodne wnioski dotyczące arealnego charakteru badanych struktur da poszerzenie grupy badanych języków. Tylko w ten sposób można stwierdzić, czy zaobserwowane cechy wspólne odróżniają języki regionu od języków w nieco dalszym sąsiedztwie (Haspelmath 2001). Staje się zatem jasne, że grupa języków objętych analizą musi obejmować zarówno języki zawsze zaliczane do środkowoeuropejskich, peryferyjne języki regionu, a także języki w bliższym i dalszym sąsiedztwie. W projekcie, którego efektem jest ten artykuł, wybrałem w sumie 16 języków reprezentujących te różne kategorie. W niniejszym szkicu opieram się na danych z 9 języków: czeskiego, niemieckiego, węgierskiego (liga dunajska); chorwackiego, polskiego (peryferia ligi dunajskiej); litewskiego, rumuńskiego (bliskie sąsiedztwo); hiszpańskiego i szwedzkiego (dalsze sąsiedztwo).

Przejdźmy do struktur, które należy ująć w analizie. Zauważmy po pierwsze, że najwięcej czasowników, prawdopodobnie w każdym języku (por. Bickel i in. 2014, s. 490), reprezentuje dwie klasy walencyjne: jedno- i dwuargumentową, w których wykładnikami argumentów są tylko standardowe formy podmiotu oraz dopełnienia bliższego w danym języku, np. polski mianownik i biernik. Obserwujemy tu paralelizm w językach całej Europy – praktycznie wszystkie reprezentują typ nominatywno-akuzatywny, który zresztą jest najczęstszy na całym świecie (Nichols 1992, s. 190). Biorąc pod uwagę zasięg tej cechy, badanie jej tutaj jest bezproduktywne.<sup>5</sup> Dlatego też w badaniu skupiam się na rzadszych, mniej regularnych<sup>6</sup> wykładnikach pozycji walencyjnych – to wśród nich możemy zaobserwować niebanalne podobieństwa o charakterze arealnym. Tym samym znika jednak możliwość prostego typologicznego charakteryzowania walencji. Poszczególne czasowniki mają własne wymagania rekcyjne, których nie można z góry uogólnić, zwłaszcza w perspektywie porównawczej.

Konieczne jest zatem oparcie analizy o pojedyncze predykaty i sankcjonowane przez nie pozycje argumentowe. W poszczególnych językach predykatom odpowiadają konkretne czasowniki, a pozycje walencyjne są wyrażane przez

<sup>5</sup> Przynależność konkretnych leksemów do klasy czasowników przechodnich ma charakter raczej genetyczny. Podobieństwa są tu silne przy blisko spokrewnionych językach, przy dalszym pokrewieństwie są słabsze, ale zauważalne (Say 2018, s. 601).

<sup>6</sup> Właśnie takie wykładniki pozycji walencyjnych można określać jako „idiomatyczne” (p. pierwszy akapit tekstu).

wykładniki gramatyczne (przypadki i przyimki). Takie szczegółowe ujęcie ról semantycznych stosowano już w badaniach typologicznych nad klasami walencyjnymi (Malchukov, Comrie 2015; Hartmann i in. 2014;<sup>7</sup> Bickel i in. 2014; Say 2014). Iren Hartmann i in. (2014) proponują dla ról semantycznych kojarzonych z pojedynczymi predykatami określenie „mikrorola semantyczna” (ang. *semantic microrole*), które będą stosował poniżej. Oto kilka przykładowych mikroról spośród 181 analizowanych przez I. Hartmann i in. (2014, s. 467): *helper* ‘pomagający’, *helped* ‘otrzymujący pomoc’ (oba przy predykatie o znaczeniu ‘pomagać’), *freezing person* ‘marznący’ (argument predykatu ‘marznąć’), *hitter* ‘uderzający’, *hittee* ‘uderzany’ (argumenty predykatu ‘uderzać’). Są to bardzo szczegółowe koncepty semantyczne związane z konkretnymi predykatami. W każdym języku obserwujemy „koekspresję mikroról” (ang. *microrole coexpression*), tzn. użycie jednej formy gramatycznej dla oznaczania wielu mikroról. Taka analiza bazuje na oddolnym uogólnianiu użycia form gramatycznych i nie narzuca gotowych ról semantycznych typu agens, pacjens, adresat, itp. Relacja form gramatycznych do mikroról nie jest identyczna we wszystkich językach, z pewnością nie jest jednak również przypadkowa (Hartmann i in. 2014, s. 464), co czyni takie relacje i kryjące się pośród nich regularności interesującym przedmiotem badań.

Dane podlegające analizie można obrazowo przedstawić jako wielką tabelę wyznaczoną przez analizowane języki i mikrorole, której komórki wypełniają odpowiednie wykładniki gramatyczne. W tabeli 1. podaję niewielką część danych z mojego projektu w takim formacie<sup>8</sup>. Warto zwrócić uwagę, że nierzadkim zjawiskiem jest możliwość użycia kilku wykładników dla oznaczenia jednej mikroroli, por. np. rumuńskie formy w pierwszym wierszu tabeli.

<sup>7</sup> Pierwsze dwie prace należą do efektów projektu badawczego «Leipzig Valency Classes Project», przeprowadzonego w latach 2009–2013 w Instytucie Maxa Plancka w Lipsku.

<sup>8</sup> W bazie danych projektu nazwy predykatów i mikroról pochodzą od czasowników angielskich. Nazwa mikroroli (np. CURE-3) to nazwa predykatu (np. CURE ‘leczyć’) poszerzona o numer pozycji walencyjnej. W tym tekście stosuję ten format nazewnictwa. Wykładniki gramatyczne podaję, stosując skróty łacińskich nazw przypadków.

**Tabela 1.** Zestawienie wykładników z badanych języków dla przykładowych 7 mikroról.

	chorwacki	czeski	hiszpański	litewski	niemiecki	polski	rumuński	szwedzki	węgierski
BELONG-2	Dat.	Dat.	a + NP	Dat.	Dat.	do + Gen.	Dat.; de + Acc.		All.
PAY-3	za + Acc.	za + Acc.	por + NP	už + Acc.	für + Acc.	za + Acc.	pentru + Acc.	för + NP	Caus.
WAIT-2		na + Acc.	por + NP	Gen.	auf + Acc.; Gen.	Gen.; na + Acc.		för + NP; på + NP	Subl.
REACT-2	na + Acc.	na + Acc.	a + NP	į + Acc.	auf + Acc.	na + Acc.	la + Acc.	på + NP	Subl.
SEARCH-2	za + Inst.			Gen.	nach + Dat.	Gen.	de + Acc.	efter + NP	Nom. + után
CURE-3	od + Gen.	od + Gen.; z + Gen.	de + NP	nuo + Gen.	von + Dat.	z + Gen.	de + Acc.	från + NP	Elat.; Inst.
CURE-4	Inst.	Inst.; pomo- cí + Gen.	con + NP	Inst.; su + Inst.	mit + Dat.	Inst.; za po- mocą + Gen.	cu + Acc.; prin + Acc.	med + NP	Inst.

Puste komórki w tabeli oznaczają usunięte z danych standardowe formy dopełnień bliższych.

Przejdźmy do kwestii „podobieństw w walencji”. Za bazowe podobieństwo strukturalne między językami przyjmuję dowolne paralelne użycie form gramatycznych. Porównać należy więc nie tylko co najmniej dwa języki, ale i minimum dwie struktury czy konteksty gramatyczne. Dla struktur walencyjnych takimi kontekstami są naturalnie różne czasowniki. Poniżej podaję przykłady takich minimalnych podobieństw. W formacie tabeli 1 każdy taki zbiór danych to fragment o wymiarach 2 na 2 komórki, w którym obserwujemy paralelizm form gramatycznych.

- (1) polski ~ węgierski  
 należeć + do + Gen. ~ tartozik + All.  
 prowadzić + do + Gen. ~ vezet + All.

- (2) niemiecki ~ polski  
 zahlen für + Acc. ~ płacić za + Acc.  
 halten für + Acc. ~ uważać za + Acc.
- (3) litewski ~ polski  
 laukti + Gen. ~ czekać + na + Acc.  
 saugotis + Gen. ~ uważać + na + Acc.
- (4) litewski ~ polski  
 sureaguoti + į + Acc. ~ zareagować + na + Acc.  
 žiūrėti + į + Acc. ~ patrzeć + na + Acc.
- (5) czeski ~ polski  
 patří + Dat. ~ należeć + do + Gen.  
 volat + Dat. ~ dzwonić + do + Gen.

Zwróćmy uwagę, że relacje genetyczne są na tym etapie ignorowane, zestawiamy tylko występowanie wykładników gramatycznych w poszczególnych językach. Jest to zgodne z praktyką typologii językoznawczej, a także z charakterem badania, które nie jest historyczno-porównawcze. Natomiast przy ocenie i interpretacji stwierdzonych podobieństw kwestie genetyczne jak najbardziej powinny być brane pod uwagę. We wcześniejszym tekście (Gaszewski 2020, s. 36–7) proponuję metodę korekty danych dla par spokrewnionych języków w badaniach arealnych. Polega ona na usunięciu z danych struktur, w których zarówno czasownik, jak i wykładnik gramatyczny w obu językach pochodzą od wspólnej formy w prajęzyku, przy czym kwestię ewentualnego pominięcia rozważa się oddzielnie dla każdej struktury. Wyniki przedstawione w niniejszym szkicu nie uwzględniają takiej korekty. Wykrywając paralele w użyciu wykładników pozycji walencyjnych, nie biorę również pod uwagę szerszej perspektywy semantycznej. Paralele możemy zaobserwować między formami, które ogólnie są ekwiwalentami, jak w (1), lub nie (pozostałe przykłady powyżej).

Podobieństwo strukturalne według zaproponowanej koncepcji daje się mierzyć, a kluczowa jest liczność paralelnych struktur. Oczywiście podobieństwa oparte na dwóch strukturach są bardzo słabe, celem jest więc znalezienie możliwie najliczniejszych paraleli w użyciu form gramatycznych. W ten sposób rozwiązuje się problem „dziwnych” podobieństw, jak w (2–5). Przyimki niemiecki i polski podane w (2) odpowiadają sobie w wielu kontekstach rekcyjnych, jednak za podstawowy odpowiednik *für* + Acc. należy uznać raczej polskie *dla* + Gen. Z kolei, polskie *na* + Acc. odpowiada w (3) i (4) różnym formom litewskim. Mając po 2 przykłady, obie ekwiwalencje musimy przyjmować za równie „podobne”. Jednak w całym zbiorze danych jako odpowiednik polskiej formy wybija się litewskie *į* + Acc. Zaznaczmy przy tym, że w kontekstach nierekcyjnych, jako przyimek przestrzenny, forma ta odpowiada raczej polskiemu *do* + Gen.,



natomiast etymologicznie – *w* (+ Acc.). Najbardziej „podejrzane” ekwiwalenty zawiera przykład (5). Owszem, czeskiemu celownikowi niekiedy odpowiada polskie *do* + Gen., ale podstawowym odpowiednikiem jest zdecydowanie polski celownik, także etymologicznie. Odzwierciedla to znacznie większa liczba mikroról dla tej drugiej pary wykładników (bez przykładu powyżej).

Pozostaje wciąż pytanie, jakie mikrorole ma objąć analiza. Zbadanie wszystkich jest niemożliwe przede wszystkim z powodów praktycznych. Poza tym, nie można zakładać, że istnieje jakiś jeden uniwersalny zestaw znaczeń wyrażanych przez czasowniki i zbiór sankcjonowanych przez nie mikroról. Trzeba zatem dokonać wyboru i tak czynią autorzy wcześniejszych prac (Comrie i in. 2015, s. 15; Haspelmath 2015, s. 134; Say 2014, s. 121). Dokładniejsza analiza pokazuje, że podawane kryteria wyboru nie są do końca jasne, a procesu selekcji żadnego z podanych źródeł nie da się odtworzyć. Próbę uściślenia kryteriów z wcześniejszych prac opisuję dokładnie w oddzielnym tekście (Gaszewski w przygotowaniu a). Metoda selekcji opiera się na obiektywnym kryterium frekwencji korpusowej oraz adekwatnym dla badania kryterium występowania z danym czasownikiem form gramatycznych innych niż standardowe wykładniki podmiotu i dopełnienia bliższego. Kryteria zastosowałem do trzech spośród wybranych języków – niemieckiego, polskiego i węgierskiego<sup>9</sup> i na tej podstawie skompilowana została lista 151 predykatów oraz związanych z nimi mikroról semantycznych.

Lista predykatów stanowi punkt wyjścia do zbierania danych. Spośród różnych możliwych źródeł informacji o walencji zdecydowałem się na najbardziej pierwotne – rodzimych użytkowników badanych języków. Taki wybór niweluje różnice w jakości, szczegółowości oraz dostępności istniejących opisów poszczególnych języków. Zaletą pracy z informantami jest również możliwość doprecyzowania informacji, jeśli pierwotnie zebrane dane pozostawiają jakieś wątpliwości interpretacyjne. Z drugiej strony, informantów nie należy pytać o struktury walencyjne jako takie, dlatego też otrzymywali oni kwestionariusz składający się ze zdań reprezentujących dane struktury walencyjne oraz opisu kluczowych dla analizy części zdań. Kwestionariusz był udostępniany przez utworzoną na potrzeby projektu aplikację internetową (Gaszewski w przygotowaniu b). Ze względów praktycznych zdania wyjściowe w kwestionariuszu sformułowane były po angielsku<sup>10</sup>.

<sup>9</sup> Korzystałem z następujących korpusów: niemieckiego DeReKo (IDS 2012, 2013), polskiego NKJP (Przepiórkowski i in. 2012) oraz węgierskiego MNSz (Oravec i in. 2014; Sass 2015).

<sup>10</sup> Znajomość języka angielskiego była zatem jednym z kryteriów, które musieli spełnić informanci.



**Tabela 2.** Zbieranie danych: przykład czeskiego zdania z predykatem CURE ‘leczyć’.

Doctors had to cure the infection with antibiotics.			
Lékaři museli vyléčit tuto infekci pomocí antibiotik.			
część zdania wejściowego	część zdania informanta	forma podstawowa	wykładnik gramatyczny mikroroli
had to cure	<b>museli vyléčit</b>	<b>vyléčit</b>	(predykat)
the infection	<b>tuto infekci</b>	<b>infekce</b>	Acc.
with antibiotics	<b>pomocí antibiotik</b>	<b>antibiotikum</b>	pomocí + Gen.

czcionka zwykła: elementy wyjściowe

pogrubienie: elementy podawane przez informanta

kursywa: opis gramatyczny materiału od informanta

Na podstawie zdań wprowadzonych przez informantów i zidentyfikowanych w nich form gramatycznych wytwarzane jest zestawienie informacji o sposobie wyrażania mikroról w poszczególnych językach. Niewielką porcję danych podaje w tabeli 3, taki format danych jest pierwotny w stosunku do zestawienia języków i mikroról (tabela 1). W danych czeskich i rumuńskich zaobserwować można wariację wykładników oznaczających tę samą mikrorolę (por. uwagi nt. tabeli 1). W danych rumuńskich możemy zauważyć także synonimiczne czasowniki podane przez informantów jako różne realizacje tego samego predykatu.

**Tabela 3.** Zestawienie informacji o pozycjach walencyjnych (przykładowy fragment danych).

język	predykat	mikrorola	czasownik	wykładnik gramatyczny
Czech	CURE	CURE-3	vyléčit	z + Gen.
Czech	CURE	CURE-3	vyléčit	Acc.
Czech	CURE	CURE-4	vyléčit	pomocí + Gen.
Czech	CURE	CURE-4	vyléčit	Inst.
Romanian	CURE	CURE-3	a trata	de + Acc.
Romanian	CURE	CURE-3	a vindeca	de + Acc.
Romanian	CURE	CURE-4	a trata	cu + Acc.
Romanian	CURE	CURE-4	a vindeca	cu + Acc.
Romanian	CURE	CURE-4	a vindeca	prin + Acc.
Swedish	CURE	CURE-3	att bota	från + NP
Swedish	CURE	CURE-4	att bota	med + NP

## Analiza koszykowa

Zebrany zasób danych wymaga opisu statystycznego. W kolejnych częściach tekstu przedstawiam wyniki uzyskane z wykorzystaniem analizy koszykowej<sup>11</sup> (Agrawal i in. 1993), tu przedstawię pokrótce działanie samej metody. Polega ona na wyszukiwaniu połączeń elementów w ramach całego zbioru danych. W obecnym badaniu „połączenie” oznacza wyrażanie przez wykładniki z różnych języków tej samej mikroroli. Wyniki zostały opracowane przy pomocy algorytmu Apriori dostępnego w ramach pakietu *arules* (Hahsler i in. 2005) w środowisku programowania R (R Core Team 2020).

Podstawowy sposób przedstawiania powiązań elementów w analizie koszykowej to reguły asocjacyjne, mające formę warunku (jeśli występuje  $x$ , to występuje  $y$ ). Prostsza metodą jest przedstawianie zestawów współwystępujących elementów (zbiór  $\{x,y\}$ ). W badaniu korzystałem zasadniczo z drugiej formy, ponieważ lepiej odzwierciedla ona rzeczywistość języka niż reguły. Wyszukane wykładniki są używane w tym samym znaczeniu, ale nie ma bezpośredniego warunkowania użycia formy z jednego języka przez formę z innego. Każdy język jest przecież używany jako samodzielny system. Praktyczną zaletą wyboru jest mniejsza liczba wyników do opracowania. Przykładowe reguły i odpowiadający im zestaw podaję w (6).

- (6)  $\text{czes. na} + \text{Acc.} > \text{niem. auf} + \text{Acc.}$   
 $\text{niem. auf} + \text{Acc.} > \text{czes. na} + \text{Acc.}$   
 $\{\text{czes. na} + \text{Acc.}, \text{niem. auf} + \text{Acc.}\}$

Podstawowe cechy każdego zestawu form gramatycznych podawane przez algorytm to jego rozmiar (ile zawiera elementów, ang. *length*, dosł. ‘długość’) oraz licznosc (jak często cały zestaw występuje w zbiorze danych, tzn. ile mikroról się z nim wiąże, ang. *count*). Wariantem licznosci podawanym w wynikach jest wsparcie (ang. *support*) – licznosc wyrażona jako ułamek liczby wszystkich mikroról. Kolejną miarą jest ufność (dokładność) reguły (ang. *confidence*), iloraz frekwencji całego zestawu oraz frekwencji poprzednika reguły. Wsparcie i ufność są wykorzystywane przez algorytm jako parametry – po podaniu wartości progowych w wynikach nie pojawiają się zestawy o niższych wartościach którekolwiek z parametrów. Jest to jeden z zasadniczych elementów kontroli ilości uzyskanych wyników oraz wydobywania najlepszych zestawów.

<sup>11</sup> Nazwa pochodzi od koszyków zakupów, bo pierwotnym zastosowaniem metody było poszukiwanie regularnych powiązań w sprzedaży towarów.

**Tabela 4.** Przykładowe zestawy wykładników wyszukane przez algorytm.

wykładniki	rozmiar	wsparcie	ufność	liczność
chorw. za + Acc., czes. za + Acc., hiszp. por + NP, lit. už + Acc., niem. für + Acc., pol. za + Acc., rum. pentru + Acc., szw. för + NP, węg. Caus.	9	0,025	0,956	6
hiszp. por + NP, lit. už + Acc., pol. za + Acc., rum. pentru + Acc., szw. för + NP, węg. Caus.	6	0,038	0,936	9
czes. za + Acc., lit. už + Acc., niem. für + Acc., pol. za + Acc., rum. pentru + Acc., węg. Caus.	6	0,034	0,963	8
chorw. za + Acc., hiszp. por + NP, lit. už + Acc., niem. für + Acc., pol. za + Acc., węg. Caus.	6	0,025	0,911	6
hiszp. por + NP, rum. pentru + Acc., szw. för + NP, węg. Caus.	4	0,055	0,967	13
czes. za + Acc., hiszp. por + NP, niem. für + Acc., rum. pentru + Acc.	4	0,046	0,941	11
chorw. za + Acc., czes. za + Acc., pol. za + Acc., węg. Caus.	4	0,025	0,881	6
lit. į + Acc., niem. auf + Acc., pol. na + Acc., rum. la + Acc., szw. på + NP, węg. Subl.	6	0,021	1	5
chorw. na + Acc., hiszp. a + NP, lit. į + Acc., pol. na + Acc., węg. Subl.	5	0,021	0,886	5
chorw. na + Acc., czes. na + Acc., niem. auf + Acc., pol. na + Acc., węg. Subl.	5	0,021	0,821	5
czes. na + Acc., hiszp. a + NP, rum. la + Acc.	3	0,021	0,704	5

Jak pokazuje tabela 4, w wynikach otrzymujemy zestawy o różnych rozmiarach. Jeśli w wynikach mamy duży zestaw np. 9-elementowy (elementy ze wszystkich badanych języków), program podaje również jego podzbiory. Taka jest zależność między pierwszym zestawem w tabeli 4 a sześcioma kolejnymi. Podzbiory mają licznosc taką samą lub większą niż cały zbiór. Przechodząc od rozmiaru 9 (wszystkie języki w badaniu) do rozmiaru 2 (minimalne porównanie) otrzymujemy bardzo wiele zestawów będących podzbiorymi większych zestawów (stanowią one zdecydowaną większość wyników). Są to różne kombinacje form występujących w pełnym zbiorze. Będę odtąd używał określenia grupa (wykładników)

w takim sensie – mając na myśli nie tylko jeden konkretny zestaw form, ale i jego liczne podzbiory. Niekiedy pełną grupę wykładników trzeba odtworzyć z zestawów o mniejszym rozmiarze. Ostatnie cztery zestawy podane w Tabeli 4 składają się na jedną spójną grupę form ze wszystkich 9 języków, choć największy z nich ma rozmiar tylko 6. W wynikach nie ma większych zestawów z tej grupy (mają za niskie wartości parametrów).

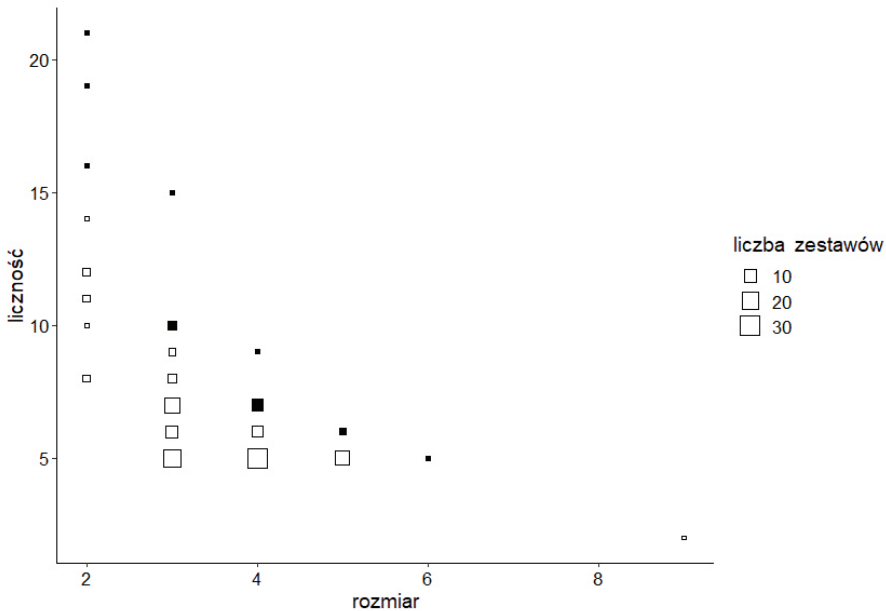
Na potrzeby niniejszego badania odrzuciłem mikrorole, które były realizowane tylko jako standardowe podmioty lub dopełnienia bliższe we wszystkich dziełach badanych językach (p. wyżej). Pozostało 238 mikroról związanych z wybranymi predykatami. Ze względu na duże zróżnicowanie zbioru danych zostały ustalone niskie wartości progowe parametrów: wsparcie – na poziomie 0,02 (co odpowiada minimalnej licznosci 5), a ufność – na poziomie 0,5. Na podstawie zestawów występujących w wynikach zostały określone grupy często odpowiadających sobie wykładników. Najsilniejsze grupy podaje w tabeli 5.

**Tabela 5.** Grupy ekwiwalentnych wykładników pozycji walencyjnych w badanych językach.

maksymalny zestaw grupy			zestawy grupy w wynikach		
wykładniki	rozmiar	liczność	max rozmiar	min liczność	max liczność
chorw. za + Acc., czes. za + Acc., hiszp. por + NP, lit. už + Acc., niem. für + Acc., pol. za + Acc., rum. pentru + Acc., szw. för + NP, węg. Caus.	9	6	9	6	24
chorw. Dat., czes. Dat., hiszp. Dat., lit. Dat., niem. Dat., pol. Dat., rum. Dat., szw. till + NP, węg. Dat.	9	8	9	8	41
chorw. od + Gen., czes. od + Gen., hiszp. de + NP, lit. iš + Gen., niem. von + Dat., pol. od + Gen., rum. de la + Acc., szw. från + NP, węg. Abl.	9	7	9	7	19
chorw. Inst., czes. Inst., hiszp. con + NP, lit. Inst., niem. mit + Dat., pol. Inst., rum. cu + Acc., szw. med + NP, węg. Inst.	9	8	9	8	27

maksymalny zestaw grupy		zestawy grupy w wynikach			
chorw. s + Inst., czes. s + Inst., hiszp. con + NP, lit. su + Inst., niem. mit + Dat., pol. z + Inst., rum. cu + Acc., szw. med + NP, węg. Inst.	9	5	9	5	27
chorw. iz + Gen., czes. z + Gen., hiszp. de + NP, lit. iš + Gen., niem. von + Dat., pol. z + Gen., rum. din + Acc., szw. från + NP, węg. Elat.	9	4	8	5	19
chorw. na + Acc., czes. na + Acc., hiszp. a + NP, lit. į + Acc., niem. auf + Acc., pol. na + Acc., rum. la + Acc., szw. på + NP, węg. Subl.	9	2	6	5	21
chorw. o + Loc., chorw. za + Acc., czes. o + Loc., hiszp. sobre + NP, lit. apie + Acc., niem. über + Acc. pol. o + Loc., rum. de + Acc., rum. despre + Acc., szw. om + NP, węg. Del.	11	3	8	5	12
lit. Gen., pol. Gen., węg. Subl.	3	6	3	6	17

Większość z grup podanych w tabeli 5 wiąże się z bardzo wieloma zestawami, np. pierwsza grupa (zestaw maksymalny to pierwszy zestaw z tabeli 4) ma w wynikach niemal 500 kombinacji elementów. Analizowanie takiej ilości danych jest mozolne, a niekoniecznie potrzebne. Najważniejsze jest określenie, które formy (z których języków) pojawiają się w podzbiorach o najwyższej liczności i na tym opiera się dalsza procedura. W ramach jednej grupy dla każdej wartości rozmiaru brane pod uwagę są tylko zestawy o liczności wyższej niż zestawy o większym rozmiarze (p. poniższy wykres). W ten sposób liczba analizowanych zestawów bardzo się ogranicza, jednak związanych z nimi jest możliwie dużo mikroról. Tak wyłuskane mikrorole możemy uznać za semantyczny „rdzeń” danej grupy wykładników.



Wykres 1. Zależność rozmiaru i liczności zestawów w ramach jednej z grup: zaczerpnięte kwadraty oznaczają zestawy brane pod uwagę w analizie

## Wyniki: grupa „na”

Przedstawiając wyniki skupię się na kilku wybranych grupach wykładników pozycji walencyjnych. Jedna z nich zawiera polskie *na* + Acc. oraz jego (mniej lub bardziej regularne) odpowiedniki w pozostałych badanych językach. Z braku oczywistego klucza nazewnictwa grupy będę określać według występujących w nich polskich form.<sup>12</sup> Grupę „na”, podaną już w tabeli 5, przedstawia (7).

- (7) chorw. *na* + Acc.  
 cz. *na* + Acc.  
 hiszp. *a* + NP  
 lit. *į* + Acc.  
 niem. *auf* + Acc.  
 pol. *na* + Acc.  
 rum. *la* + Acc.  
 szw. *på* + NP  
 węg. Subl.

<sup>12</sup> Jest to tylko przyjęta konwencja nazewnictwa, która nie oznacza w żaden sposób uprzywilejowanego statusu polskich form.

Pełny 9-elementowy zestaw ma licznosc zaledwie 2. Przy tak niskiej wartosci nie mozna uznać, że grupa realnie łączy wszystkie języki, nawet jeśli w każdym z nich algorytm odnalazł pasującą formę gramatyczną. Interesują nas zestawy o najwyższej licznosci (p. wyżej). Istotne jest po pierwsze, które formy w nich występują, a po drugie, jakie mikrorole są z nimi związane. Wykres na końcu poprzedniej części tekstu dotyczy właśnie grupy „na”.

Łatwo zauważyć, że niektóre wykładniki powtarzają się niemal w każdym wierszu tabeli 6, podającej najliczniejsze zestawy z grupy „na” o rozmiarach 2, 3 oraz 4. Są to formy czeska, polska i węgierska. Rzadziej pojawiają się wykładniki litewski i niemiecki, pozostałe języki są reprezentowane zaledwie w jednym lub dwóch zestawach. Jest więc jasne, że grupa „na” najsilniej wiąże formy gramatyczne z języków środkowoeuropejskich. Co ciekawe, najmocniejszy związek jest między polskim a węgierskim, a więc językami niespokrewnionymi. Blisko spokrewniona para czesko-polska ma niższe wartości licznosci i ufności, i to bez jakiegokolwiek korekty dotyczącej struktur potencjalnie odziedziczonych z prajęzyka (p. wyżej).

**Tabela 6.** Zestawy grupy „na” o najwyższej licznosci.

wykładniki	rozmiar	licznosc
pol. na + Acc., węg. Subl.	2	21
cz. na + Acc., węg. Subl.	2	19
cz. na + Acc., pol. na + Acc.	2	16
cz. na + Acc., pol. na + Acc., węg. Subl.	3	15
cz. na + Acc., niem. auf + Acc., węg. Subl.	3	10
lit. į + Acc., pol. na + Acc., węg. Subl.	3	10
niem. auf + Acc., pol. na + Acc., węg. Subl.	3	10
cz. na + Acc., niem. auf + Acc., pol. na + Acc., węg. Subl.	4	9
hiszp. a + NP, lit. į + Acc., pol. na + Acc., węg. Subl.	4	7
niem. auf + Acc., pol. na + Acc., szw. på + NP, węg. Subl.	4	7
lit. į + Acc., pol. na + Acc., rum. la + Acc., węg. Subl.	4	7
chorw. na + Acc., lit. į + Acc., pol. na + Acc., węg. Subl.	4	7
chorw. na + Acc., cz. na + Acc., pol. na + Acc., węg. Subl.	4	7
cz. na + Acc., pol. na + Acc., rum. la + Acc., węg. Subl.	4	7

Każdy z zestawów podanych powyżej wiąże się z konkretnymi mikrorolami, które zebrałem w tabeli 7. Tworzą one semantyczny „rdzeń” grupy. Tabela 7 naczynnie pokazuje arealne zależności w grupie „na”.



**Tabela 7.** Mikrorole związane z zestawami grupy „na” o najwyższej liczności.

	węgierski	polski	czeski	niemiecki	litweski	chorwacki	rumuński	szwedzki	hiszpański
ANSWER-2	x	x	x	x	x	x	x	x	x
REACT-2	x	x	x	x	x	x	x	x	x
GLANCE-2	x	x	x	x	x	x	x	x	
PAY-ATTENTION-2	x	x		x	x	x	x	x	x
INVITE-3	x	x	x		x	x	x		x
LOOK-2	x	x	x	x	x		x	x	
EXCHANGE-3	x	x	x		x		x		x
RELY-2	x	x	x	x		x		x	
COUNT-ON-2	x	x	x	x		x			
DIVIDE-3	x	x	x		x	x			
TRANSLATE-4	x	x			x	x			x
WAIT-2	x	x	x	x				x	
WATCH-OUT-2	x	x	x	x			x		
GROW-INTO-2	x	x			x				x
PREPARE-PREEMPT-2	x	x	x	x					
COLLECT-3	x	x	x						
EARN-4	x	x	x						
LACK-3	x	x	x						
PREPARE-2	x		x	x					
SERVE-4	x		x				x		
VOTE-2	x	x						x	
ALLOW-3	x	x							
LEAVE-TO-3	x		x						
NEED-4		x	x						
PUNISH-2	x	x							
REMEMBER-F-2	x		x						

Znak x w komórce oznacza użycie wykładnika z grupy „na”, niekiedy jest to jeden z kilku używanych wykładników. Wiersze i kolumny zostały posortowane malejąco według liczby niepustych komórek.

W semantyce mikroról można dostrzec kilka prawidłowości. Wykładniki tej grupy używane są z czasownikami wyrażającymi reakcje (ANSWER ‘odpowiadać’, REACT ‘reagować’) do oznaczenia elementu wyzwalającego reakcję i tylko w tym kontekście wszystkie badane języki zgodnie używają form z (7). Znajdujemy kilka predykatów percepcji (GLANCE ‘spojrzeć’, LOOK ‘patrzeć’, PAY-ATTENTION ‘zwracać uwagę’), przy których wykładniki z grupy „na” oznaczają rolę bodźca zmysłowego. Występują predykaty procesów umysłowych (COUNT-ON ‘liczyć na coś’, RELY ‘polegać’, REMEMBER ‘pamiętać’, WAIT ‘czekać’), w tym kontekście formy z (7) oznaczają treść mentalną danego procesu. Ponadto, mamy predykaty oznaczające procesy zmiany (DIVIDE ‘dzielić’, GROW-INTO ‘rozwinąć się w coś’, EXCHANGE ‘wymienić’, TRANSLATE ‘przekładać’), przy których wykładniki z grupy „na” oznaczają stan (rzecz) po zmianie. Wreszcie, w użyciu z niektórymi predykatami formy z (7) oznaczają rolę celu (COLLECT ‘zbierać’, EARN ‘zarabiać’, NEED ‘potrzebować’, PREPARE ‘przygotowywać się’). Do określania tych ostatnich mikroról z wykładników grupy „na” korzystają w zasadzie tylko języki środkowoeuropejskie, i nie w każdym przypadku wszystkie.

Trudno zaproponować cechę semantyczną, która pozwalałaby na uogólnienie obejmujące wszystkie powyższe mikrorole. Możemy się jednak pokusić o dwie niepełne i częściowo pokrywające się generalizacje. Po pierwsze, bez specjalnych kontrowersji można rozszerzyć rolę semantyczną bodźca (ang. *stimulus*) i objąć nią nie tylko argumenty czasowników percepcji (jak LOOK), ale też reakcji (jak ANSWER) i procesów umysłowych (jak WAIT). Druga propozycja dotyczy argumentów większości powyższych predykatów umysłowych (jak WAIT), predykatów zmiany (jak EXCHANGE) oraz argumentów roli celu (jak PREPARE). Wspólną ich cechą jest perspektywa czasowa – wszystkie te mikrorole określają uczestników czy elementy sytuacji, które mają (lub tylko mogą) dopiero się pojawić. Taki charakter niektórych użyc polskiego *na* + Acc. zauważyłem już wcześniej i nazywam taką rolę semantyczną „rzeczą antycypowaną” (Gaszewski 2016, s. 86–89).

Słabą stroną pierwszej propozycji jest fakt, że wiele predykatów, które mają argumenty o podobnej semantyce, nie łączy się z formami z (7), por. pol. czasowniki *widzieć*, *oglądać*, *myśleć*, *zastanawiać się*, które obywają się bez użycia *na* + Acc. Podobnie jest z drugą generalizacją – wskazana cecha nie przesądza o użyciu form z grupy „na”. Wydaje się więc, że obie propozycje wskazują cechy sprzyjające użyciu wykładników z grupy „na”, jednak faktyczne występowanie tych form zależy też od innych czynników. Ogólnie rzecz biorąc, semantyka grupy „na” jest niejednorodna, za użycie wykładników odpowiadają różne motywacje semantyczne. Tym bardziej wymowne jest podobne wyrażanie takiego urozmaiconego zestawu znaczeń w kilku sąsiadujących ze sobą językach.

## Wyniki: grupa „z” i grupa „od”

Przejdźmy do kolejnego przykładu. Wśród grup wykładników, które wykazuje Tabela 5 zwróćmy uwagę na następujące.

- (8) chorw. iz + Gen.  
 czes. z + Gen.  
 hiszp. de + NP  
 lit. iš + Gen.  
 niem. von + Dat.  
 pol. z + Gen.  
 rum. din + Acc.  
 szw. från + NP  
 węg. Elat.
- (9) chorw. od + Gen.  
 czes. od + Gen.  
 hiszp. de + NP  
 lit. iš + Gen.  
 niem. von + Dat.  
 pol. od + Gen.  
 rum. de la + Acc.  
 szw. från + NP  
 węg. Abl.

Jak widać, w części języków formy gramatyczne powtarzają się w obu grupach. Z punktu widzenia np. hiszpańskiego czy litewskiego, grupy te są po prostu wariantami jednej kategorii. Jednak w części języków używane są różne wykładniki – są to języki środkowoeuropejskie: chorwacki, czeski, polski i węgierski, oraz rumuński. O ile w grupie „na” region wskazywany przez dane rozciągał się na północ (litewski), to region rozróżnienia grup „od” i „z” rozciąga się na południe (chorwacki, rumuński). Ponieważ badanie nie obejmowało innych języków bałkańskich, nie da się tu określić południowej granicy tej cechy.

W wynikach oprócz licznych podzbiorów pełnych zestawów (8) i (9) są także zestawy komplikujące prosty podział języków nakreślony powyżej. Mają one mniejsze rozmiary oraz stosunkowo niską licznosc. (10) i (11) podsumowują te serie zestawów, w obu wraz z częścią wykładników grupy „z” występuje jakiś inny wykładnik.

- (10) chorw. od + Gen.  
 czes. z + Gen.  
 hiszp. de + NP  
 pol. z + Gen.
- (11) czes. z + Gen.  
 hiszp. de + NP

niem. aus + Dat.

pol. z + Gen.

Jak widać, zakłócenia pojawiają się w danych chorwackich – forma z grupy „od” łączy się także z elementami grupy „z”. W danych niemieckich pojawia się nieobecny dotąd przyimek *aus* + Dat., w połączeniu z formami grupy „z”. Okazuje się, że język niemiecki jednak wykorzystuje tu dwa różne wykładniki, podobnie jak czeski czy polski, z tą różnicą, że *von* + Dat. dominuje na tyle mocno, że wyszukuje się jako główna forma niemiecka w obu grupach.

**Tabela 8.** Mikrorole związane z zestawami grup „z” i grupy „od” o najwyższej liczności.

	czeski	polski	węgierski	rumuński	chorwacki	niemiecki	szwedzki	litewski	hiszpański
CHANGE-3	z	z	z	z	z	o	w	w	w
GROW-INTO-3	z	z	z	z	z	o	w	w	w
TRANSFORM-3	z	z	z	z	z	o	w	w	w
DEVELOP-FROM-2	z	z	z	z	z	z	w	w	w
TRANSLATE-3	z	z		z		z	w	w	w
RESULT-FROM-2	z	z	z	z		z	w		w
COMPRISE-2	z	z	z	z	o	z		w	w
FIND-OUT-3	o,z	o,z	o,z	o,z	o,z	o	w	w	w
BENEFIT-2	z	z	z		o	o,z		w	w
CURE-3	o,z	z	z		o	o	w		w
GIVE-UP-2	z	z			o,z				w
EXCHANGE-4	z			o			w	w	w
ACCEPT-3	o	o	o	o	o	o	w	w	w
BUY-3	o	o	o	o	o	o	w	w	w
LEARN-3	o	o	o	o	o	o	w	w	w
NEED-3	o	o		o	o	o	w	w	w
RECEIVE-3	o	o	o	o	o	o	w	w	w
RENT-FROM-3	o	o	o	o	o	o	w	w	w
TAKE-3	o	o	o	o	o	o	w	w	w
DEMAND-3	o	o	o	o	o	o	w	w	
HEAR-ABOUT-3	o	o	o	o	o	o	w	w	
STEAL-2			o	o	o	o	w	w	

o – użycie formy z grupy „od”

w – użycie jednej wspólnej formy dla obu grup

z – użycie formy z grupy „z”

Dane chorw. i niem. podano kursywą. Użycie innych form gramatycznych nie jest zaznaczane.

Zebrawszy najliczniejsze zestawy obu grup, podaję w tabeli 8 ich zestawienie z mikrorolami. Cztery języki najbardziej z lewej strony tabeli bardzo wyraźnie rozdzielają dwie grupy mikroról, trzy z prawej konsekwentnie używają jednego wykładnika. Rozkład form chorwackich i niemieckich nie wpisuje się w ogólne prawidłowości, przy czym nie ma również specjalnego podobieństwa między tymi dwoma językami – każdy z nich wyróżnia się na swój sposób.

Jeśli chodzi o semantykę, w grupie „z” występują predykaty oznaczające procesy zmian (CHANGE ‘zmienić’, GROW-INTO ‘rozwinąć się w coś’, DEVELOP-FROM ‘rozwinąć się z czegoś’, TRANSFORM ‘zamienić się’, TRANSLATE ‘przekładać’), a formy z (8) oznaczają stan (rzecz) przed zmianą. Ponadto, występuje mikrorola oznaczająca przyczynę (RESULT-FROM ‘wynikać’). W grupie „od” odnajdujemy predykaty fizycznego transferu (ACCEPT ‘przyjąć’, RECEIVE ‘otrzymać’, STEAL ‘ukraść’, TAKE ‘wziąć’), w tym potencjalnego transferu (DEMAND ‘żądać’, NEED ‘potrzebować’) oraz transakcji (BUY ‘kupować’, RENT-FROM ‘wynajmować od kogoś’) – wykładniki z (9) mają tu zawsze rolę źródła (ang. *source*). Poza tym, znajdujemy predykaty oznaczające przekaz informacji (HEAR-ABOUT ‘słyszeć o czymś’, LEARN ‘uczyć się’), przy których formy grupy „od” oznaczają źródło informacji.

Znaczenia wyliczone powyżej stanowią dość spójną grupę. Najwłaściwszym uogólnieniem jest szerokie rozumienie roli semantycznej źródła – właściwie wszystkie mikrorole w tabeli 8 można uznać za jej podtypy, z uwzględnieniem rozwinień metaforycznych. Pasuje to dobrze do języków korzystających z jednego wykładnika oraz do podstawowego przestrzennego znaczenia form używanych w obu grupach. Jednak rozdział wykładników w językach środkowoeuropejskich również ma podstawy semantyczne. Rolę źródła/stanu przed zmianą w grupie „z” możemy rozszerzyć na mikrorole takich predykatów jak CURE ‘leczyć’, GIVE-UP ‘rezygnować’ czy RESULT-FROM ‘wynikać’. Takie mikrorole są odróżniane od innych typów roli źródła (w procesach transferu fizycznego oraz transferu wiedzy), przy których używane są formy grupy „od”.

### **Wyniki: grupa dopełniaczy**

Ostatnia grupa różni się od poprzednich małym rozmiarem. W efekcie wszystkie związane z nią zestawy wyszukane przez algorytm mieszczą się w tabeli 9. Mamy formy z zaledwie trzech języków, przy czym węgierski sublativus wiąże się tylko z mniej licznym zestawem. Usunięcie tego wykładnika zwiększa licznosc zestawu niemal trzykrotnie, węgierski nie wyszukuje się też na poziomie rozmiaru 2 (odpowiednie podzbiory mają za niską ufność). Istotne w analizowanej grupie wykładników jest więc paralelne użycie litewskiego i polskiego dopełniacza, cecha łącząca w naszym badaniu tylko te dwa języki.

**Tabela 9.** Zestawy reprezentujące grupę dopełniaczy.

wykładniki	rozmiar	liczność
lit. Gen., pol. Gen., węg. Subl.	3	6
lit. Gen., pol. Gen.	2	17

Przyjrzyjmy się mikrorolom związanym z grupą dopełniaczy i ich semantyce. Oba przypadki, litewski i polski, są używane z predykatami oznaczającymi pragnienie, żądanie, itp. (DEMAND ‘żądać’, DESIRE ‘pragnąć’, INSIST ‘nalegać, domagać się’, EXPECT ‘spodziewać się’, COUNT-ON ‘liczyć (na coś)’, WISH ‘życzyć’) i oznaczają pożądaną rzecz (stan). Do tego spójnego zestawu mikroról możemy dołączyć elementy brakujące (a więc niejako „pożądane”) w sytuacjach określanych predykatami NEED ‘potrzebować’, SEARCH ‘szukać’, oraz LACK ‘brakować’, WAIT ‘czekać, oczekiwać’. Nieco przewrotnie obie formy są również używane z predykatami oznaczającymi negatywne emocje (FEAR ‘bać się’, HATE ‘nienawidzić’, WATCH-OUT ‘strzec się’) dla oznaczenia elementu wyzwalającego te odczucia. Wreszcie, mamy blisko związane predykaty oznaczające przekaz wiedzy (LEARN ‘uczyć się’, TEACH ‘uczyć’), przy których dopełniacze litewski i polski oznaczają przekazywaną wiedzę. Można rozszerzyć semantykę przekazu informacji na predykat LISTEN ‘słuchać’. Ostatnia mikrorola w wynikach wiąże się z predykatem ASK-Q ‘pytać (się)’, który również odnosi się do przekazu wiedzy. Jednak tym razem obie formy oznaczają wyraźnie inny element sytuacji – adresata pytania, czyli (potencjalne) źródło wiedzy.

Grupa dopełniaczy jest przykładem ewidentnego podobieństwa litewsko-polskiego, które ma charakter arealny. Oczywiście jest kontakt obu języków, przy czym blisko spokrewniony z polskim czeski nie pasuje do tej grupy wykładników. Podobieństwo wiąże jednak język polski w inny sposób niż w poprzednich grupach, który teraz odróżnia się od języków ligi dunajskiej. Możliwe, że zaobserwowana grupa dopełniaczy odzwierciedla podobieństwa o szerszym zakresie w Europie Wschodniej, jednak jaki to zakres nie można stwierdzić na podstawie obecnego zestawu danych.

### Uwagi końcowe

Przedstawiłem trzy przykłady grup wykładników gramatycznych (przypadków i przyimków) używanych do oznaczania pozycji walencyjnych. We wszystkich możemy zaobserwować różnice między językami, które świadczą o arealnych podobieństwach w użyciu badanych form gramatycznych. W pierwszym przypadku (grupa „na”) języki wyraźnie różniły się liczebnością zestawów, w których były reprezentowane. Języki środkowoeuropejskie: czeski, polski i węgierski były

najsilniej związane, w mniejszym stopniu litewski i niemiecki. W drugim przypadku (grupy „z” i „od”) część języków korzystała z jednej formy gramatycznej dla całego szeregu mikroról, część dzieliła mikrorole między dwa różne wykładniki. Ta druga grupa to języki środkowoeuropejskie: czeski, polski, węgierski wraz z rumuńskim. Rozróżnienie dwóch wykładników w języku chorwackim i niemieckim jest mniej spójne z poprzednimi czterema językami. W ostatnim przypadku (grupa dopełniaczy) podobieństwo zachodzi między językami litewskim i polskim (oraz marginalnie węgierskim), zastosowany algorytm nie podaje w ogóle wyników z pozostałych języków.

Przytoczone przykłady potwierdzają występowanie arealnych schematów w użyciu wykładników pozycji walencyjnych, a w szczególności – schematów charakterystycznych dla języków Europy Środkowej. Warto jednak zwrócić uwagę, że każda z analizowanych grup wykładników miała własny wzorzec rozmieszczenia geograficznego. Dwa pierwsze przypadki pasują do koncepcji ligi dunajskiej, przy czym w obu język polski jawi się jako jeden z języków o charakterze centralnym, niemiecki natomiast jest tu językiem marginalnym. Ponadto, grupa języków raz rozciąga się na północ (litewski), raz na południe (rumuński, chorwacki). Ostatniej analizowanej grupy wykładników nie sposób wiązać z ligą dunajską, może być spójna z szerszym podobieństwem języków w Europie Wschodniej (np. ligą rokytnicką) – stwierdzenie tego wymaga porównania kolejnych języków.

Warto podkreślić dużą szczegółowość prawidłowości dystrybucyjnych związanych z formalnym oznaczaniem walencji. Poszczególne grupy podobnie używanych wykładników wykazują inne schematy podobieństwa między językami. Wybrane tu przykłady były pozytywne, stwierdzały podobieństwo (części) języków środkowoeuropejskich, jednak nie każda grupa wykładników daje taki obraz. Obserwacje te korelują ze współczesnymi trendami w językoznawstwie arealnym, gdzie nacisk kładzie się na konkretne cechy i ich rozmieszczenie geograficzne, a nie na identyfikację lig językowych.

Wreszcie, zwróćmy uwagę, że cechy semantyczne wyłaniające się z zaproponowanych uogólnień mikroról okazują się niekiedy dość szczegółowe i niezbyt oczywiste z punktu widzenia dotychczasowych badań nad językami środkowoeuropejskimi oraz walencją. Takie zjawiska obserwujemy we wszystkich trzech omówionych przypadkach.



## Skróty

- Abl. – ablativus  
Acc. – biernik  
All. – allativus  
Caus. – causalis  
Dat. – celownik  
Del. – delativus  
Elat. – elativus  
Gen. – dopełniacz  
Inst. – narzędnik  
Subl. – sublativus

## Bibliografia

- Agrawal, R., Imieliński, T., Swami, A. (1993). Mining association rules between sets of items in large databases. *Proceedings of the 1993 ACM SIGMOD international conference on Management of data - SIGMOD '93*, 207. DOI: 10.1145/170035.170072.
- Aikhenvald, A.Y. (2006). *Grammars in Contact: A Cross-Linguistic Perspective*. W: A.Y. Aikhenvald, R.M. W. Dixon, (red.), *Grammars in Contact. A Cross-Linguistic Typology* (1–66). Oxford: Oxford University Press.
- Balázs, J. (1983). Areális nyelvészeti kutatások története, módszerei és főbb eredményei. W: J. Balázs (red.), *Areális nyelvészeti tanulmányok* (7–112). Budapest: Tankönyvkiadó/Franklin.
- Bickel, B., Zakharko, T., Bierkandt, L., Witzlack-Makarevich, A. (2014). Semantic role clustering: an empirical assessments of semantic role types in non-default case assignment. *Studies in Language*, 38 (3), 485–511.
- Bláha, O. (2015). *Jazyky střední Evropy*. Olomouc: Univerzita Palackého v Olomouci.
- Campbell, L. (2017). Why is it so Hard to Define a Linguistic Area? W: R. Hickey (red.): *The Cambridge Handbook of Areal Linguistics* (19–39). Cambridge: Cambridge University Press.
- Comrie, B., Hartmann, I., Haspelmath, M., Malchukov, A., Wichmann, S. (2015). Introduction. W: A. Malchukov, B. Comrie (red.), *Valency Classes in the World's Languages* (tom I, 3–26). Berlin, Boston: de Gruyter Mouton.
- Décsy, G. (1973). *Die linguistische Struktur Europas*. Wiesbaden: Harrassowitz.
- EJO – Polański, K. (red.) (1993). *Encyklopedia językoznawstwa ogólnego*. Wrocław, Warszawa, Kraków: Ossolineum.

- Gaszewski, J. (2016). Czy przyimki podlegające rekcji nie wyrażają w ogóle znaczenia?, *Język Polski* 96 (1), 80–95.
- Gaszewski, J. (2020). Does Verb Valency Pattern Areally in Central Europe? A First Look. W: L. Szucsich, A. Kim, Agnes, U. Yazhinova (red.), *Areal Convergence in Eastern Central European Languages and Beyond* (13–53). Wien, Berlin: Peter Lang.
- Gaszewski, J. (w przygotowaniu a.) How to obtain a good sample of verbs? Sampling of open lexical classes for comparative studies. Artykuł na bazie referatu *How to get a sample of verbs for a comparative study?* 52. doroczne spotkanie Societas Linguistica Europaea, sierpień 2019. Lipsk.
- Gaszewski, J. (w przygotowaniu b.) How to collect cross-linguistic data on valency online. Artykuł na bazie referatu *An online questionnaire for collecting valency data*. Warsztaty naukowe „Workshop on questionnaires for linguistic description and typology”, listopad 2017. Paryż.
- Hahsler, M., Grün, B., Hornik, K. (2005). arules – A Computational Environment for Mining Association Rules and Frequent Item Sets, *Journal of Statistical Software*, 14 (15).
- Hartmann, I., Haspelmath, M., Cysouw, M. (2014). Identifying semantic role clusters and alignment types via microrole coexpression tendencies. *Studies in Language*, 38 (3), 463–484.
- Haspelmath, M. (2001). The European linguistic area: Standard Average European. W: M. Haspelmath, E. König, W. Oesterreicher, W. Raible (red.) *Language Typology and Language Universals* (tom II, 1492–1510). Berlin, New York: de Gruyter.
- Haspelmath, M. (2015). Transitivity prominence. W: A. Malchukov, B. Comrie (red.) *Valency Classes in the World's Languages* (tom I, 131–47). Berlin, Boston: de Gruyter Mouton.
- Heine, B., Kuteva, T. (2006). *The Changing Languages of Europe*. Oxford, New York: Oxford University Press.
- Hickey, R. Areas, Areal Features and Areality. W: R. Hickey (red.): *The Cambridge Handbook of Areal Linguistics* (1–15). Cambridge: Cambridge University Press.
- IDS 2012 – *Deutsches Referenzkorpus / Archiv der Korpora geschriebener Gegenwartssprache 2012–I* (wersja z 29.02.2012). Mannheim: Leibniz-Institut für Deutsche Sprache. <http://www.ids-mannheim.de/kl/projekte/korpora/archiv.html>.
- IDS 2013 – *Korpusbasierte Wortgrundformenliste DEREW0*. Mannheim: Leibniz-Institut für Deutsche Sprache. <http://www.ids-mannheim.de/derew0>.
- Januška, J. (2020). Central European Languages as a Complex Research Issue: Summarising and Broadening the Research Foci. W: L. Szucsich, A. Kim, U. Yazhinova (red.), *Areal Convergence in Eastern Central European Languages and Beyond* (55–93). Berlin: Peter Lang.
- Kurzová, H. (1996). Mitteleuropa als Sprachareal. *Acta Universitatis Carolinae – Philologica* 5, *Germanistica Pragensia* XIII, 57–73.
- Majewicz, A.F. (1989). *Języki świata i ich klasyfikowanie*. Warszawa: Państwowe Wydawnictwo Naukowe.

- Malchukov, A., Comrie, B. (2015). *Valency Classes in the World's Languages* (2 tomy). Berlin, Boston: de Gruyter Mouton.
- Newerkla, S.M. (2000). Language Affinity in Central Europe – Some Thoughts on the Interrelations of German, Czech, Slovak and Magyar, *Opera slavica – Slavistické rozhledy*, X/4, 1–16.
- Nichols, J. (1992). *Linguistic Diversity in Space and Time*. Chicago: University of Chicago Press.
- Oravecz, C., Váradi, T., Sass, B. (2014). The Hungarian Gigaword Corpus. W: N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, S. Piperidis (red.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)* (1719–1723). Reykjavik: ELRA.
- Pilarský, J. (2001). *Donausprachbund. Das arealistische Profil einer Sprachlandschaft*. Rozprawa habilitacyjna. Uniwersytet w Debreczynie.
- Przepiórkowski, A., Bańko, M., Górski, R. L., Lewandowska-Tomaszczyk, B. (red.) (2012). *Narodowy Korpus Języka Polskiego*. Warszawa: PWN.
- Pusztay, J. (1996). Közép-Európa nyelvi képe. W: G. Székely, E. Jónás, (red.) *Nyelvek és nyelvoktatás a Kárpát-medencében 1* (11–28). Nyíregyháza: Bessenyei György Könyvkiadó.
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Wiedeń. Strona projektu: <https://www.R-project.org/>.
- Sass, B. (2015). 28 millió szintaktikailag elemzett mondat és 500000 igei szerkezet. W: A. Tanács, V. Varga, V. Vincze (red.), *MSZNY 2015, XI. Magyar Számítógépes Nyelvészeti Konferencia* (303–308). Szeged: SZTE.
- Say, S. (2014). Bivalent verb classes in the languages of Europe: A quantitative typological study. *Language Dynamics and Change*, 4 (1), 116–166.
- Say, S. (2018) Markirovanie aktantov dvuhmestnyh predikatov: predvaritel'nye itogi tipologičeskogo issledovanija. W: S. Say (red.), Valentnostnye klassy dvuhmestnyx predikatov v raznostrukturnyx jazykax (557–616). Sankt-Peterburg: Rossijskaja Akademiya Nauk, Institut Lingvističeskix Issledovanij.

---

**ABSTRAKT:** Tekst porusza kwestię podobieństw w oznaczaniu pozycji walencyjnych pomiędzy językami Europy Środkowej. Temat jest badany empirycznie, na podstawie skompilowanej dużej bazy danych schematów walencyjnych, z użyciem metod analizy koszykowej. Badanie uwzględnia także języki spoza badanego regionu, aby uniknąć fałszywego rozpoznania każdego znalezionej podobieństwa jako cechy środkowoeuropejskiej. Artykuł omawia szczegółowo trzy grupy wykładników walencji, w których można dopatrzeć się podobieństw środkowoeuropejskich. Badanie semantyki stwierdzonych regularności formalnych sugeruje, że użycie form gramatycznych do oznaczania walencji jest warunkowane przez niezauważane w dotychczasowych badaniach cechy znaczeniowe.