

Binarne opozycje *безпека/небезпека* i *bezpieczeństwo/niebezpieczeństwo* w języku ukraińskim i polskim: korpusowa analiza kontrastywna

Binary oppositions *безпека/небезпека* and *bezpieczeństwo/niebezpieczeństwo* in Ukrainian and Polish: a corpus-based contrastive analysis

Mariia ONYSHCHUK

Uniwersytet Państwowy im. Mychajła Drahomanowa (Kijów, Ukraina)/Mykhailo Drahomanov State University (Kyiv, Ukraine)

E-mail: m.i.onyshchuk@udu.edu.ua, 

Abstract: The aim of the article is to examine the binary oppositions *security-insecurity* using statistical data from Polish Web2019 (plTenTen 19) and Ukrainian Web 2020 and 2014 (ukTenTen20), monolingual corpora, to identify commonalities and specific features in their realization and distribution in a real contextual environment. We hypothesize that binary oppositions can be statistically represented in different ways. Various properties of Sketch Engine toolkit are used for this purpose (Thesaurus, Word Sketch Difference) interpreting the statistical findings and their implications for language use by discussing the frequency, score and syntactic relations of the binary constituents across monolingual corpora plTenTen 19 and ukTenTen20. The study revealed the difference in the statistical representation of the binary oppositions based on syntactic relations, while outlining similarities in the lexical, particularly their synonymic range. The frequency of use of grammatical or syntactic constructions is motivated by their productivity/non-productivity in language reproduction. A contrastive study of the binary oppositions *security-insecurity* confirmed the suggestion that the frequency of lexemes is directly related to the ease/difficulty of their perception, the choice of synonymous words in different contexts being determined by mechanisms of categorization and reflection of connections between elements of the real world, by their relevance to the Ukrainian and Polish community. Moreover, the given corpus-based inquiry into the lexico-semantic and grammatical aspects of web corpora upholds the idea of providing new insights into the languages being compared – a snapshot that may go unnoticed in monolingual corpus research.

Keywords: binary opposition, collocation, frequency, thesaurus, contrastive analysis, web corpus.

Wstęp

Celem artykułu jest zaproponowanie technik opartych na korpusach do badania binarnych opozycji w ujęciu międzyjęzykowym, takich jak binarne opozycje *безпека/небезпека* i *bezpieczeństwo/niebezpieczeństwo* w języku ukraińskim i polskim. W trakcie badania zadano następujące pytania badawcze: 1) jak reprezentowane są opozycje binarne w językach ukraińskim i polskim; 2) jak różnią się opozycje binarne pod względem realizacji i dystrybucji na materiale korpusów internetowych;

oraz 3) jakie są podobieństwa i różnice w kontrastujących językach i jak one występują. Zakładamy, że opozycje binarne można przedstawić w różny sposób pod względem statystycznym, co zrealizujemy za pomocą narzędzia *Sketch Engine*.

Techniki wychwytują wzorce współwystępowania reprezentantów opozycji binarnych w zdaniach, tj. ich użycie w konstrukcjach nieciągłych, jak również w konstrukcjach ciągłych lub w ramach kontrastywnych. Podstawowym założeniem jest to, że wszystkie języki posiadają pewną liczbę silnie opozycyjnych leksykalnych par semantycznych wyrażających opozycję binarną.

Badanie prezentuje nowe spojrzenie na semantyczno-dystrybucyjną weryfikację binarnych opozycji *безпека/небезпека* i *безпеczeńство/небезпеczeńство* w języku ukraińskim i polskim. Przedstawiono metodologię i wyniki wyodrębniania pojęć z obszaru konceptualnego analizowanych opozycji. Do badania wykorzystano elektroniczny podkorpus internetowy języka polskiego *Web Polish Web 2019* (plTenTen19) oraz ukraińskiego *Ukrainian Web 2020 i 2014* (ukTenTen20). Za pomocą silnika *Sketch Engine*, narzędzia do analizy tekstu online, które analizuje większe partie tekstu, podjęto działania zmierzające do zidentyfikowania typowych i różnych cech w kontrastujących językach. Oba korpusy stanowią reprezentatywny, ustrukturyzowany zbiór tekstów różnych gatunków w analizowanych językach wraz z oprogramowaniem: wyszukiwaniem słów, form gramatycznych oraz ich kolokacji, wykonywaniem znaczników częściowych słów, a także przetwarzaniem wyników wyszukiwania, sortowaniem, tworzeniem zrównoważonych próbek i uzyskiwaniem różnych informacji statystycznych.

W wyniku wykorzystania korpusów do badania pola semantycznego opozycji binarnych *безпека/небезпека* i *безпеczeńство/небезпеczeńство*, wraz z analizą jednostek leksykalnych, uzyskano informacje na temat frekwencji form wyrazowych, leksemów, kategorii gramatycznych, wariantów połączeń leksemów w zdaniu. Przeanalizowano szczegółowo komponenty, które wyróżniają stabilne wyrażenia z podanym lematem *безпека/небезпека* i *безпеczeńство/небезпеczeńство* w korpusie w porównywanych językach. Korzystając z systemu *Word Sketch* ustawiono lematy do analizy z automatycznego korpusu oferowanego przez program.

W pierwszej części artykułu omówiono teoretyczne podstawy lingwistycznej deskrypcji zjawiska opozycji binarnych, a także badań korpusowych w dziedzinie automatycznej ekstrakcji leksykalnej oraz przedstawiono główne cechy elektronicznego podkorpusu internetowego języka ukraińskiego i polskiego.

Druga część artykułu przedstawia metodologię i wyniki ekstrakcji pojęć ze sfery konceptualnej *безпека/небезпека* i *безпеczeńство/небезпеczeńство*. Do aktualnego badania wykorzystano wyżej wymienione elektroniczne podkorpusy internetowe języka ukraińskiego i polskiego.

W wyniku przeprowadzonego badania uzyskano listę leksemów, które pojawiały się stosunkowo często w podkorpusie i rzadko w korpusie referencyjnym. Ta różnica we względnej częstotliwości była głównym kryterium identyfikacji potencjalnych opozycji leksykalnych, które znajdują się w semantycznym i konceptualnym centrum binarnej opozycji. Weryfikacja zakresu semantycznego analizowanego pojęcia na pod-

stawie danych leksykograficznych wykazała skuteczność zastosowanych metod, ponieważ celem niniejszego artykułu jest próba przedstawienia przydatności elektronicznego korpusu języka ukraińskiego i polskiego do badania i analizy leksyki na materiale korpusów internetowych.

Obecnie wielu językoznawców kompleksowo rozważa opozycje binarne w oparciu o dzieła literackie, a także dyskurs polityczny i publiczny. Przeprowadzając analizę leksykalno-semantyczną języka, możemy znaleźć wiele typów słów fachowych, słów wejściowych, słów terminologicznych, słów dialektalnych i neologizmów, słów przestarzałych, jednostek frazeologicznych, antonimów, synonimów, homonimów, a także przykładów literatury ustnej (S.G. Iskakbekovna/ B.S. Rezhopovich 2023: 160–161). Binarne pary opierają się na różnych kryteriach, takich jak przestrzeń, czas, jakość, kolor, smak, gęstość substancji, kategorie wielkości i ilości, wyrażenia oceniające, zjawiska naturalne, kategorie wiekowe i społeczne, genealogiczne, znaczenia społeczne, ogólny tryb obrazu świata, działania.

Termin *binarne przeciwieństwa* pochodzi ze strukturalizmu, zgodnie z jakim według Ferdinanda De Saussure język jest również uważany za system opozycji binarnych. Twórca nowoczesnej lingwistyki i strukturalizmu opracował koncepcję języka jako systemu opozycji binarnych (A. Putri/ P. Sarwoto 2016: 25). Opozycja binarna jest akceptowana jako model poznania świata (M. Danesi 2009: 12; A. Dundes 1997: 39; C. Lévi-Strauss 1974: 25). Ogólnie rzecz biorąc, opozycje binarne są jednostkami leksykalnymi, które definiują dobro i zło oraz przeciwne strony ludzkich zachowań i cech w relacjach społecznych¹.

Opozycja binarna jest uważana za integralną cechę myślenia dyskursywnego, przeciwstawiającą sobie różne formy intuicji (A. Utaker 1974: 73). Założmy, że postrzeganie binarnych par opozycji *безпека/небезпека* i *bezpieczeństwo/niebezpieczeństwo* ujawnia kontrastujące światopoglądy, ilustruje rolę binarnych par opozycyjnych w konstruowaniu językowego obrazu świata, obrazując ich socjolingwistyczne implikacje we współczesnych językach polskim i ukraińskim. Badanie semantyki na podstawie korpusów językowych jest możliwe dzięki istnieniu różnych rodzajów znaczników semantycznych. Znaczenie leksemów jest określane na podstawie kontekstu ich użycia, poprzez badanie typowych kombinacji słów i kolokacji, tj. konkor-dancji (R.S. Ngula 2017: 220; C. Paradis 2016: 138).

Opozycje binarne można przedstawić w różny sposób pod względem statystycznym. Analiza leksykograficzna oparta na korpusach językowych pomaga analizować

¹ J. Smith stwierdza, że „opozycja binarna to system, za pomocą którego w języku i myśli dwa teoretyczne przeciwieństwa są ściśle zdefiniowane i zestawione ze sobą” (J. Smith 1996: 40). Koncepcja pochodzi od szwajcarskiego lingwisty Ferdinanda de Saussure’a, dla którego język jest zbiorem znaków, przy czym każdy znak zawiera znaczące i znaczone. W ten sam sposób, aby zrozumieć „dobry”, musimy zrozumieć „zły”; analogicznie podchodzimy do deskrypcji innych binarnych opozycji, ponieważ dla „pięknego” istnieje „brzydki”, dla „hałaśliwego” jest „cichy”. Poststrukturaliści wyrazili przekonanie, że „rzeczy nie mogą być rozumiane w izolacji – muszą być postrzegane w kontekście większych struktur, których są częścią” (A.K. Barry 2002: 39).

kontekstowe użycie niektórych słów, zwłaszcza słów synonimicznych lub antonimicznych, ich częstotliwość kolokacji z innymi słowami, regularność ich użycia w różnych stylach, i jasno określić ich semantykę (M. Baker 1995: 223). Binarne opozycje umożliwiają wyjaśnienie i zrozumienie chaotycznych i abstrakcyjnych pojęć w łatwiejszy i skuteczny sposób, zestawiając je ze sobą.

Zakładamy, że we wszystkich językach istnieje pewna liczba silnie przeciwstawnych leksykalnych par semantycznych, których znaczenia są kluczowe dla ludzkiej egzystencji. Istnieje również duża liczba innych par, które są często używane jako antonimy w dyskursie, ale z różnych powodów nie są tak mocno powiązane jak pary kanoniczne. Wymiary znaczeniowe par kanonicznych zostały wykorzystane jako punkt wyjścia dla badań typologicznych w celu zidentyfikowania użycia słów w różnych kombinacjach i w różnych ramach semantycznych w różnych językach i kulturach.

1. Metodologia badania

W badaniu korzystano z automatycznego korpusu języka ukraińskiego *Ukrainian Web 2020 i 2014* (ukrTenTen20), który zawiera 2 592 516 436 słów; dla języka polskiego korzystano z automatycznego korpusu języka polskiego *Polish Web 2019* (plTenTen19), który zawiera 4 253 636 443 słów. Badanie składa się z 2 etapów. W pierwszym etapie za pomocą narzędzia *Sketch Engine* została przebadana reprezentacja oraz dystrybucja binarnych opozycji *безпека/небезпека* i *безпеczeńство/небезпеczeńство* w wymienionych korpusach internetowych; szczególnie skupiamy się na analizie za pomocą narzędzi *Thesaurus* i *Word Sketch Difference*. Drugi etap badania zawierał kontrastywną analizę reprezentacji binarnych opozycji *безпека/небезпека* i *безпеczeńство/небезпеczeńство* w porównywanych korpusach języków ukraińskiego i polskiego.

Główną metodą zastosowaną w badaniu było przeanalizowanie procesu konceptualizacji przez pryzmat binarnych opozycji za pomocą narzędzi korpusowych² *Sketch Engine*.

Podczas analizy skupiono się na tym, w jaki sposób binarne pary opozycyjne *безпека/небезпека* i *безпеczeńство/небезпеczeńство* konstruowały różne ramy poznawcze i światopoglądy w korpusie internetowym. Porównano i skontrastowano te

² <https://www.sketchengine.eu/> to oprogramowanie do zarządzania korpusami i analizy tekstu opracowane przez *Lexical Computing Limited* od roku 2003. Jego celem jest umożliwienie osobom badającym zachowania językowe przeszukiwania dużych zbiorów tekstów zgodnie z zapytaniami językowymi. Narzędzie *Sketch Engine* ma szereg funkcji: „Thesaurus” do budowania tezaury; „Clustering” do grupowania pozycji tezaury w klastry/grupy leksykalne i semantyczne; „Sketch differentiation” do identyfikowania podobieństw i różnic w parach słów; „Collocations” – automatyczne wykrywanie kolokacji; „Word sketch” – wykrywanie kolokacji (kolokacje ograniczone przez model składniowy). Wszystkie z nich wykrywają relacje semantyczne między terminami na różne sposoby i śledzą je w danym korpusie. Wyświetlają również listę wskazującą częstotliwość występowania każdej kolokacji w korpusie, wartość relacji między słowem kluczowym a kolokacją.

pary z ich współczesnymi odpowiednikami w języku i kulturze polskiej oraz ukraińskiej. Przeprowadzając analizę językową, mamy za cel ujawnić rolę opozycji binarnych w kształtowaniu językowej reprezentacji świata w porównywanych korpusach internetowych oraz ich możliwe implikacje dla współczesnego społeczeństwa polskiego i ukraińskiego.

W proponowanym badaniu wykorzystano techniki językoznawstwa korpusowego³, przeprowadzając statystyczną analizę częstotliwości użycia jednostek językowych, która może również służyć jako wskaźnik procesów percepcji, reprodukcji i transmisji danych językowych (J. Aarts 1991: 99). Tak więc częstotliwość leksemów jest bezpośrednio związana z łatwością/trudnością ich percepcji, a wybór słów synonimicznych w różnych kontekstach jest związany z mechanizmami kategoryzacji i odzwierciedlania powiązań między elementami świata rzeczywistego. Częstotliwość użycia konstrukcji gramatycznych lub składniowych jest określona przez ich produktywność/nieproduktywność w reprodukcji językowej.

Z jednej strony, metodologia analizy słów kluczowych (*Key Words in Context* – KWIC) pomaga ujawnić tendencję ich użycia, zapewniając w ten sposób przestrzeń do refleksji nad społecznymi, politycznymi i emocjonalnymi aspektami obu społeczności. Z drugiej strony, najczęstsze kolokacje pokażą specyfikę binarnych opozycji w analizowanym korpusie internetowym w rzeczywistym środowisku kontekstowym. Narzędzia te pomagają zanotować ich częstotliwość i dystrybucję w tekstach internetowych. Opcja tezaurusu zapewnia narzędzie do przeglądania synonimów i podobnych słów, ułatwiając w ten sposób nowy wgląd w relacje leksykalno-semantyczne.

Analiza korpusu daje nowy wgląd w porównywane języki – ujęcie, które może być niezauważone w badaniach korpusów jednojęzycznych oraz oferuje nowe spojrzenie na semantyczno-dystrybucyjną weryfikację pojęcia *безпека/небезпека* i *bezpieczeństwo/niebezpieczeństwo*. Część systemu o nazwie *WordSketch* jest odpowiedzialna za wyszukiwanie kombinacji słów. System przetwarza słowa związane z wyszukiwanym słowem i innymi słowami w jego otoczeniu oraz dostarcza informacji o gramatycznych oraz funkcjonalnych odmianach danego wyrazu. Wyniki analizy są zorganizowane w kategorie zwane relacjami gramatycznymi. O tym, które słowa są uwzględniane w analizie, decydują reguły zawarte lub ujęte w systemie, wykrywającym potencjalne kolokacje. Reguły te zależą od języka, w którym napisane są analizowane teksty.

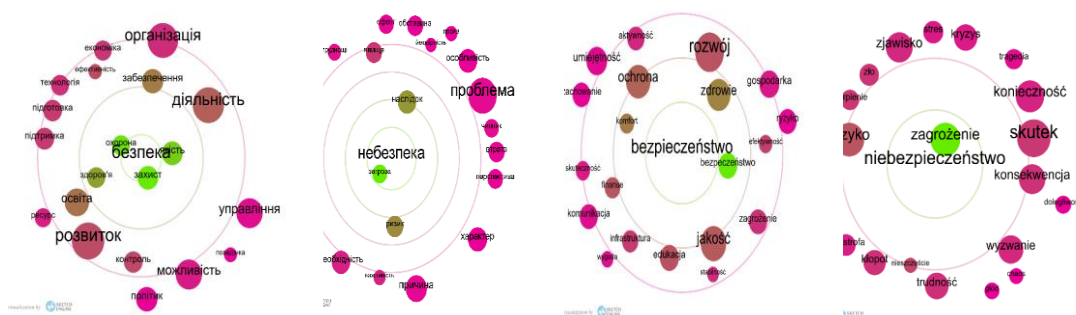
2. Omówienie wyników

Za pomocą funkcji *Thesaurus*⁴ wyodrębniono słowa kluczowe *безпека* – *bezpieczeństwo* w ukraińskim korpusie tekstowym *Web 2020 i 2014* (ukrTenTen20) i polskim

³ Metoda analizy korpusowej jest stosowana od czasu pojawienia się korpusów językowych z adnotacjami semantycznymi. Wciąż nie ma wystarczającej liczby badań w tym obszarze, ale jest to obiecujący kierunek, ponieważ znaczenie jednostek językowych wynika z ich zgodności i częstotliwości użycia w określonych kontekstach językowych. Takie badania przyczyniają się również do wdrażania najnowszych podejść do leksykografii i pedagogiki językowej.

⁴ Opcja tezaurusu zapewnia narzędzie do przeglądania synonimów i podobnych słów, ułatwiając w ten sposób nowy wgląd w relacje leksykalno-semantyczne.

korpusie tekstowym *Polish Web 2019* (plTenTen19). Udało się pokazać jednostki, które posiadają podobny rozkład do danego słowa. Podobieństwo dystrybucyjne słów jest obliczane statystycznie na podstawie miary asocjacji *logDice* (znormalizowana miara Dice) i szablonów leksykalno-syntaktycznych. Rys. 1 przedstawia listę pojęć do lematów *безпека* i *безпеczeńство* w porównywanych korpusach. Tezaurus dystrybucyjny jest budowany na podstawie kolokacji⁵. Najczęstsze kolokacje pokażą specyfikę binarnych opozycji w analizowanym korpusie internetowym w rzeczywistym środowisku kontekstowym. Narzędzie to pomaga odnotować ich częstotliwość i dystrybucję w tekstach internetowych.



Rys. 1. Wizualizacja pojęć 'безпека – bezpieczeńство' i 'небезпека – небезпеченство' w ukraińskim korpusie tekstowym *Web 2020* i *2014* (ukrTenTen20) i polskim korpusie tekstowym *Polish Web 2019* (plTenTen19).

Podczas porównywania leksyki, która znajduje się w obszarze lematu *безпека* – *bezpieczeńство*, zidentyfikowano spójność pojęć za pomocą analizy wyników punktacji, którą przeprowadziliśmy z wykorzystaniem narzędzia *Thesaurus*: zdrowie (0.306 w Ukr, 0.371 w Pol), edukacja (0.296 w Ukr, 0.43 w Pol), ochrona (0.32 w Ukr, 0.351 w Pol), jakość (0.316 w Ukr, 0.345 w Pol), rozwój (0.288 w Ukr, 0.341 w Pol), efektywność (0.288 Ukr i 0.329 w Pol). Jednak, znacznie statystycznie różni się semantyczny zasięg pojęcia *безпека* – *bezpieczeńство* w ukraińskim (np. *діяльність*, *контроль*) i polskim (np. *finanse*, *infrastruktura*) (ryc. 2).

⁵ Wspólne kolokacje można zidentyfikować, jeśli dwa słowa w korpusie mają wiele wspólnych kolokacji, zostaną one uwzględnione w teaurusie dla słowa kluczowego.

	Thesaurus result <i>безпека</i>	Similarity score	Frequency
1	захист	0.322	1060444
2	охорона	0.32	760372
3	якість	0.316	1084563
4	здоров'я	0.306	788175
5	забезпечення	0.298	1134901
6	освіта	0.296	1513729
7	діяльність	0.292	2609475
8	розвиток	0.288	2800055
9	ефективність	0.288	396892
10	контроль	0.29	847733

	Thesaurus result <i>bezpieczeństwo</i>	Similarity score	Frequency
1	zdrowie	0.371	1081606
2	komfort	0.366	233135
3	ochrona	0.351	1227013
4	jakość	0.345	1384298
5	finanse	0.343	295225
6	rozwój	0.341	1902168
7	edukacja	0.34	575867
8	efektywność	0.329	160270
9	infrastruktura	0.33	337666
10	stabilność	0.32	80307

Ryc. 2. Częstotliwość stosowania lematu 'безпека' w ukraińskim korpusie tekstowym Web 2020 i 2014 (ukrTenTen20) i 'niebezpieczeństwo' w polskim korpusie tekstowym Web 2019 (plTenTen19) z wykorzystaniem narzędzia 'Thesaurus'.

Podczas porównywania leksyki, która znajduje się w obszarze lematu *безпека* – *niebezpieczeństwo*, zidentyfikowano spójność pojęć za pomocą analizy wyników punktacji, którą przeprowadziliśmy z wykorzystaniem narzędzia *Thesaurus*: zagrożenie (0.424 w Ukr, 0.468 w Pol), ryzyko (0.336 w Ukr, 0.316 w Pol), trudność (0.228 w Ukr, 0.278 w Pol). Wyniki różnią się w obszarze pojęciowym takiego lematu jak naслідок, który w języku polskim kojarzy się z lematami konsekwencja, skutek (0.348 w Ukr, 0.281 w Pol); lemat *zło* nie jest reprezentowany w analizowanym korpusie języka ukraińskiego; lemat *zjawisko* (0.263 w Ukr – 4 miejsce, 0.271 w Pol – 10 miejsce na liście wyników) jest inaczej reprezentowany w porównywanych językach, co, naszym zdaniem, świadczy o różnych sposobach konceptualizacji świata przez użytkowników, co wyraża socjologiczne preferencje społeczeństwa w konkretnym okresie (ryc. 3).

	Thesaurus result <i>безпека</i>	Similarity score	Frequency
1	загроза	0.424	235773
2	наслідок	0.348	553485
3	ризик	0.336	400125
4	явище	0.263	357449
5	важливість	0.244	89511
6	необхідність	0.24	537534
7	фактор	0.229	387339
8	труднощі	0.228	144085
9	чинник	0.223	223003
10	вплив	0.222	896821

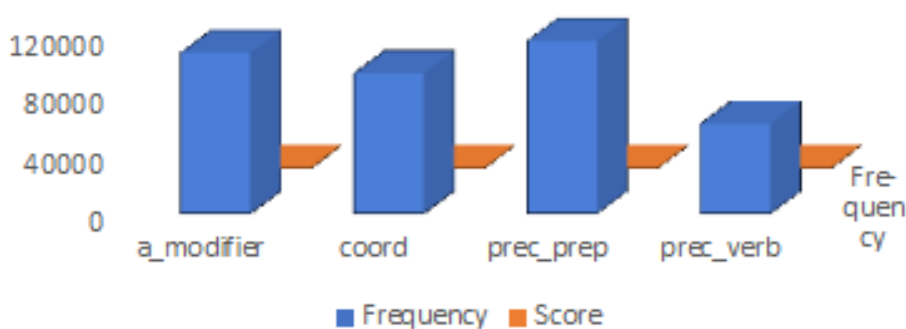
	Thesaurus result <i>niebezpieczeństwo</i>	Similarity score	Frequency
1	zagrozenie	0.468	492608
2	ryzyko	0.316	497710
3	nieszczęście	0.301	73272
4	skutek	0.288	642287
5	cierpienie	0.287	156933
6	zło	0.285	185428
7	konsekwencja	0.281	380466
8	kłopot	0.28	256716
9	trudność	0.278	293089
10	zjawisko	0.271	391273

Ryc. 3. Częstotliwość stosowania lematu 'безпека' w ukraińskim korpusie tekstowym Web 2020 i 2014 (ukrTenTen20) i 'niebezpieczeństwo' w polskim korpusie tekstowym Web 2019 (plTenTen19) z wykorzystaniem narzędzia 'Thesaurus'.

Korzystając z menedżera korpusów systemu *Word Sketch*⁶, wyszukano dane ze wskaźnikami statystycznymi reprezentacji leksykalnej pojęcia *безпека/безпека*

⁶ Domyślnie *Word Sketch* sortuje kombinacje słów według częstotliwości ich użycia, dzięki czemu możliwe jest zauważenie kombinacji częstych, a Indeks Sorensena, test i wzajemna

i *bezpieczeństwo/niebezpieczeństwo* w podkorpusach porównywanych języków. Ważne jest, że w wyniku wykorzystania korpusów do badania pola semantycznego pojęcia *безнека* i *безпе́чення* wraz z analizą jednostek leksykalnych, uzyskano informacje na temat częstotliwości występowania form słownych, leksemów, kategorii gramatycznych, wariantów połączeń leksykalnych w zdaniu oraz kolokacje ze stabilnych wyrażen terminologicznych. Przeanalizujemy bardziej szczegółowo komponenty, które wyróżniają stabilne wyrażenia z danym lematem *безнека* – *безпе́чення* w korpusie porównywanych języków. Korzystając z systemu *Word Sketch*⁷, ustawiamy lematy do analizy z automatycznego korpusu oferowanego przez program. W trakcie językowej charakterystyki użycia lematu *безнека* wykorzystano analizę statystyczną do zbadania częstotliwości jego użycia w wyrażeniach stabilnych.



Ryc. 4. Częstotliwość stosowania lematu 'безнека' w wyrażeniach stabilnych w ukraińskim korpusie tekstowym Web 2020 i 2014 (ukrTenTen20).

informacja są używane do określenia, jak typowa (lub jak silna) jest kolokacja. Wysoka wartość wyniku oznacza, że kolokacja jest silna, a jeśli wartość jest niska, kolokacja jest odpowiednio słaba. Jednak głównym celem punktacji jest sortowanie kolokacji według ich typowości lub siły.

⁷ *Word Sketch* jest odpowiedzialny za wyszukiwanie kombinacji słów. System przetwarza słowa związane z wyszukiwanym słowem i innymi słowami w jego otoczeniu oraz dostarcza informacji o gramatycznych i funkcjonalnych odmianach danego wyrazu. Wyniki analizy są zorganizowane w kategorie zwane relacjami gramatycznymi. O tym, które słowa są uwzględniane w analizie, decydują reguły zawarte lub ujęte w systemie, wykrywającym potencjalne kolokacje. Reguły te zależą od języka, w którym napisane są analizowane teksty.

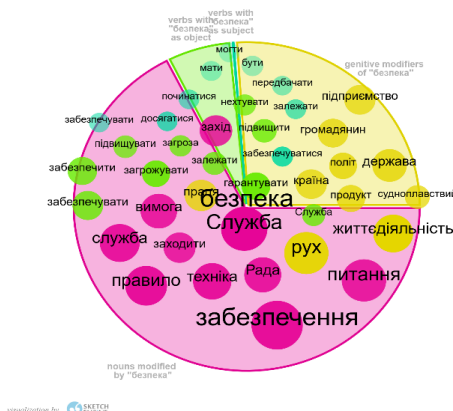


Рис. 5. Типы połączeń dla lematu 'безпека' w ukraińskim korpusie tekstowym Web 2020 i 2014 (ukr-TenTen20).

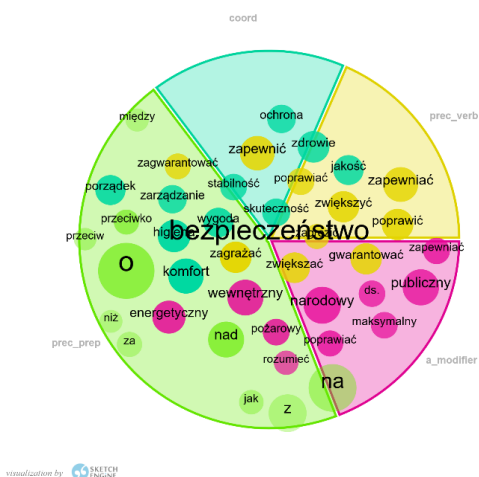


Рис. 6. Типы połączeń dla lematu 'bezpieczeństwo' w korpusie Polish Web 2019 (plTenTen19).

Wyniki analizy gramatycznych relacji w wyrazach stabilnych z lematem *безпека* (na-rzędzie *Word Sketch*) w korpusie tekstowym *Ukrainian Web 2020 i 2014* (ukr-TenTen20).

1. Najczęstsze kolokacje z lematem *безпека* w tekstach analizowanego korpusu reprezentują przymiotniki występujące w stabilnych wyrażeniach w funkcji modyfikatora słowa kluczowego: *служба, забезпечення, правило, техніка, рада, питання, вимога, захід, рівень, система, ремінь, гарантія, подушка*, natomiast najniższą frekwencją kolokacji z analizowanym lematem charakteryzują się rzeczowniki *аналіз, мета, міністерство, ступінь, норматив, регламент, місячник, критерій*.

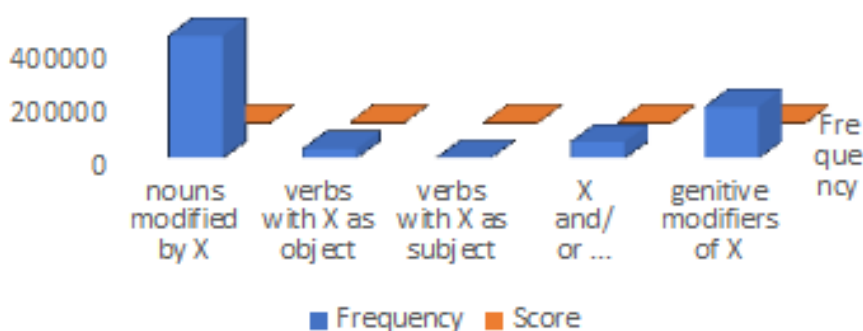
2. Drugą statystycznie reprezentowaną pozycję zajmuje grupa kolokacji z relacjami koordynacji lematu *безпека* w badanym ukraińskim korpusie tekstowym *Web 2020*

i 2014 (ukrTenTen20). Analiza semantyczno-dystrybucyjna kolokacji z tym lematem wykazała przewagę kolokacji z takimi rzeczownikami kluczowymi w funkcji dopełniacza modyfikatora *життєдїяльнїсть, рух, держава, полїт, праця, продукт, громадянин, країна, підприємство, пасажир, перевезення, експлуатація, продукція, пацієнт*.

3. Do trzeciej grupy należą czasowniki *гарантувати, загрожувати, забезпечувати, залежати, підвищити*. Jednocześnie niską frekwencją charakteryzują się kolokacje lematu bezpieczeństwo z czasownikami *полїнувати, зберегти, порушити, підтвердити, управляти*.

Zauważono, że dużo kolokacji używa się w wyrażeniach przyimkowych i frazach czasownika z lematem w różnych połączeniach syntaktycznych.

W trakcie językowej charakterystyki użycia lematu *безпе́czeńство* wykorzystano analizę statystyczną do zbadania częstości jego użycia w wyrażeniach stabilnych.



Ryc. 7. Typy połączeń lematu 'bezpieczeństwo' w wyrażeniach stabilnych w korpusie Polish Web 2019 (plTenTen19).

Wyniki analizy gramatycznych relacji w wyrazach stabilnych z lematem *bezpieczeństwo* (narzędzie *Word Sketch*) w korpusie tekstowym *Polish Web 2019* (plTenTen19).

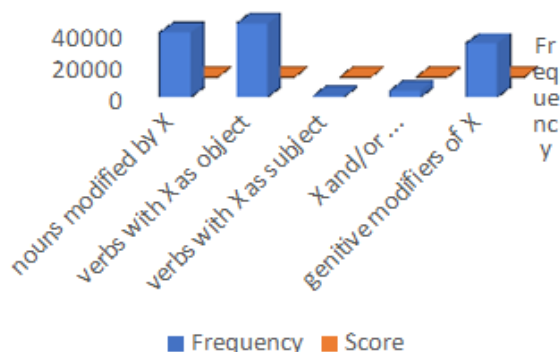
1. Najczęstsze kolokacje z lematem *bezpieczeństwo* w tekstach analizowanego korpusu reprezentują przymiotniki *energetyczne, wewnętrzne, pożarowe, narodowe, publiczne, zdrowotny, pełne, państwowe, własne, finansowe, wysokie*, a także czasowniki *poprawiać, zapewniać*. Jednocześnie niską frekwencją charakteryzują się kolokacje lematu *bezpieczeństwo* z przymiotnikami *surowcowe, informacyjne, pozamilitarne*.

2. Druga grupa z lematem *bezpieczeństwo* obejmuje kolokacje z aktywnymi przyimkami *nad, o, przeciwko, naprzeciw, jak, za, z, niż, między, pomiędzy*, co pozwala znacznie rozszerzyć zakres używania lematu.

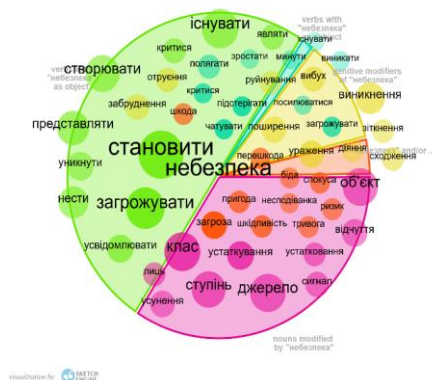
3. Trzecią statystycznie reprezentowaną pozycję zajmuje grupa kolokacji z relacjami koordynacji lematu *bezpieczeństwo* w badanym korpusie. Analiza semantyczno-dystrybucyjna kolokacji z tym lematem wykazała przewagę kolokacji z takimi rzeczownikami kluczowymi jak *komfort, higiena, wygoda, porządek, skuteczność, jakość*,

zdrowie, zarządzanie, stabilność, ochrona, niezawodność, obronność, spokój, współpraca, natomiast najniższą frekwencją kolokacji z analizowanym lematem charakteryzują się rzeczowniki *rzetelność, oświata, technologia*. Najmniejsza grupa składa się z kolokacji lematu *bezpieczeństwo*, które znajdują się w relacjach predykatywnych.

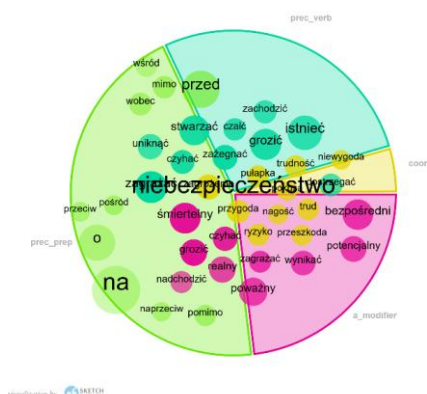
W trakcie językowej charakterystyki użycia lematu *небезпека* wykorzystano analizę statystyczną do zbadania częstości jego użycia w wyrażeniach stabilnych.



Ryc. 8. Częstość stosowania lematu 'небезпека' w wyrażeniach stabilnych w ukraińskim korpusie tekstowym Web 2020 i 2014 (ukrTenTen20).



Ryc. 9. Типы połączeń dla lematu 'небезпека' w ukraińskim korpusie tekstowym Web 2020 i 2014 (ukrTenTen20).



Wyniki analizy gramatycznych relacji w wyrazach stabilnych z lematem *небезпека* (narzędzie *Word Sketch*) w korpusie tekstowym *Ukrainian Web 2020 i 2014* (ukrTenTen20).

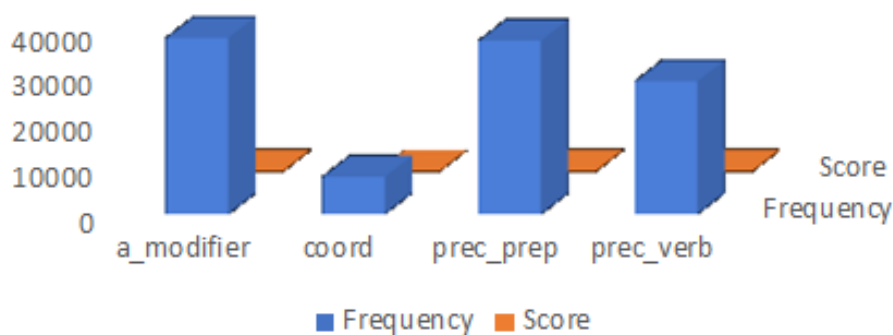
2. Druga grupa z lematem *небезпека* obejmuje kolokacje z grupą czasowników z dopełnieniem jako przedmiotem/obiektom, na przykład *загрожувати, становити, усвідомлювати, представляти, існувати, створювати, нести, являти, критися, уникнути, чатувати, відчутти, виникати, відчувати, піддавати, підстерігати*. Zakres elementów tej grupy wyraża dość szeroki zakres palety lematu *небезпека* w ukraińskim korpusie tekstowym *Web 2020 i 2014* (ukrTenTen20). Najmniejszą grupę stanowią kolokacje z czasownikiem w roli podmiotu, na przykład *підстерігати, чатувати, загрожувати, критися, полягати, посилюватися, зростати, минути, виникати, існувати, чекати, могли*.

3. Trzecią statystycznie reprezentowaną pozycję w badanym korpusie zajmuje grupa kolokacji z modyfikatorami dopełniacza. Analiza semantyczno-dystrybucyjna kolokacji z lematem *небезпека* wykazała przewagę kolokacji z rzeczownikami *ураження, виникнення, отруєння, вибух, забруднення, сходження, поширення, зіткнення, руйнування, діяння, інфікування*. Natomiast najniższą frekwencją kolokacji z analizowanym lematem charakteryzują się rzeczowniki *поневолення, настання, викид, переродження, вплив, речовина, сепаратизм*.

Applied Linguistics Papers: www.alp.uw.edu.pl

небезпека wykorzystano analizę statystyczną, aby rozważyć częstotliwość jego użycia w wyrażeniach stabilnych.

W trakcie językowej charakterystyki użycia lematu *niebezpieczeństwo* wykorzystano analizę statystyczną do rozważenia częstotliwości jego użycia w wyrażeniach stabilnych.



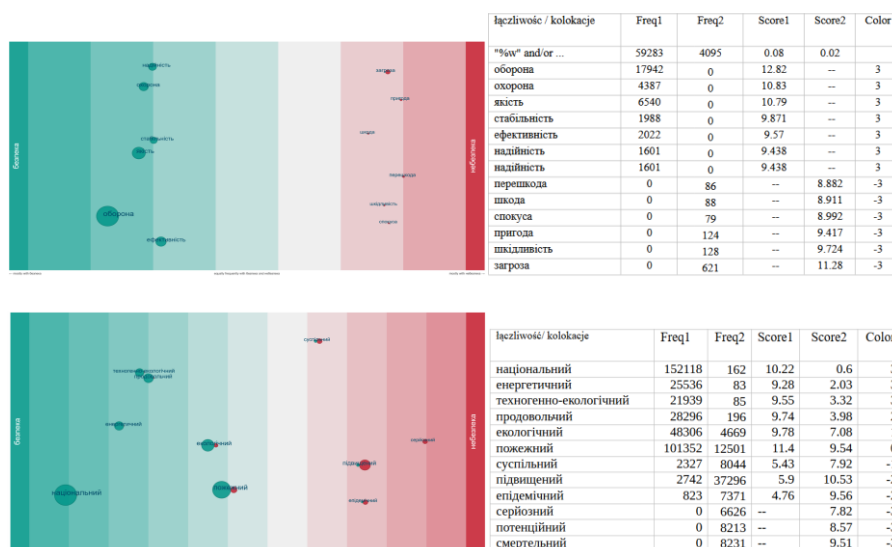
Ryc. 11. Typy połączeń dla lematu 'niebezpieczeństwo' w korpusie Polish Web 2019 (plTenTen19).

Wyniki analizy gramatycznych relacji w wyrazach stabilnych z lematem *niebezpieczeństwo* (narzędzie *Word Sketch*) w korpusie tekstowym Polish Web 2019 (plTenTen19).

1. Najczęstsze kolokacje z lematem *niebezpieczeństwo* w tekstach analizowanego korpusu reprezentują przymiotniki *śmiertelne, bezpośrednie, realne, poważne, potencjalne, groźne, pożarowe, straszliwe*, a także czasowniki *stanowić, grozić, zagrażać, wynikać, nadchodzić, stwarzać*. Jednocześnie niską frekwencją charakteryzują się kolokacje lematu *niebezpieczeństwo* z przymiotnikami *niewyobrażalne, prawdopodobne, lawinowe, bolszewickie, jawne*, a także czasownikami *wyeliminować, wywołać, wzrastać*.
2. Drugą statystycznie reprezentowaną pozycję zajmuje grupa kolokacji z relacjami przyimków używanych z lematem *niebezpieczeństwo* w badanym korpusie z najwyższą frekwencją kolokacji: *przed, pomimo, mimo, wobec, o, na*; najmniejszą frekwencją kolokacji z *wokół i odnośnie*.
3. Trzecią statystycznie reprezentowaną pozycję zajmuje grupa kolokacji z relacjami czasowników w prepozycji. Analiza semantyczno-dystrybucyjna kolokacji z lematem *niebezpieczeństwo* wykazała przewagę kolokacji z takimi czasownikami jak *zagrażać, zażegnać, czyhać, stwarzać, grozić, czaić, istnieć, zachodzić, unikać*, natomiast najniższą frekwencją kolokacji z analizowanym lematem charakteryzują się czasowniki *zignorować, niwelować, narastać, wyolbrzymiać, ominąć*.

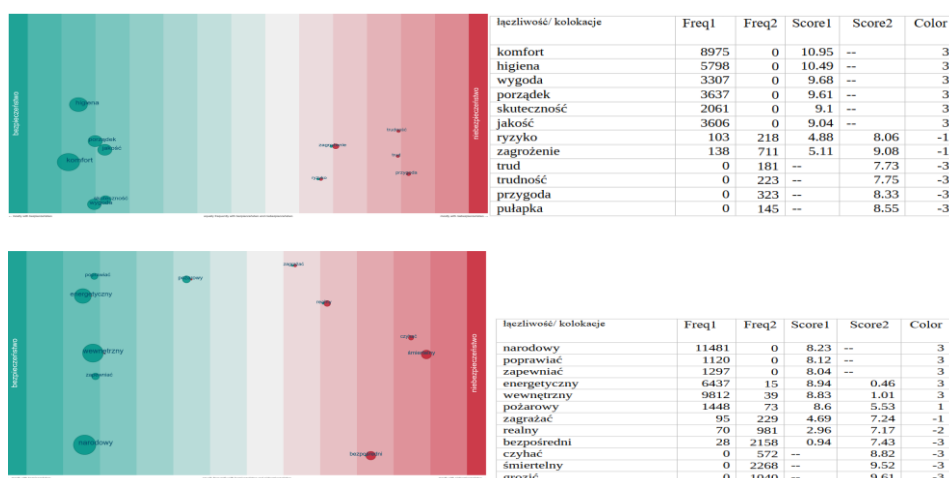
Word Sketch Difference:⁸ porównanie binarnych opozycji *безпека/небезпека* i *безпеczeństwo/небезпеczeństwo*

Korzystając z funkcji wyszukiwania, podjęto próbę porównania kolokacji dwóch różnych lematów w ramach binarnych opozycji. Wybierając opcję wyszukiwania zaawansowanego i określając część mowy jako rzeczownik, uzyskano dane statystyczne, które pomogą zidentyfikować różnicę między cechami semantycznymi i dystrybucyjnymi porównywanych lematów. W korpusie języka ukraińskiego zidentyfikowaliśmy typowe pojęcia dla lematu *безпека*: *оборона, якість, надійність, стабільність, охорона, ефективність*; *небезпека*: *перешкода, пригода, шкода, спокуса*. W polskim korpusie wytypowaliśmy takie pojęcia dla analizowanych binarnych opozycji *bezpieczeństwo*: *obrona, jakość, niezawodność, stabilność, komfort, porządek, jakoś, higiena, wygoda, skuteczność*; *niebezpieczeństwo*: *zagrożenie, ryzyko, trud, trudność, przeszkoda, przygoda, krzywda, pokusa*.



Ryc. 12. Word Sketch: różnica binarnej opozycji 'безпека/небезпека' oraz powiązanych pojęć w korpusie Ukrainian Web 2020 i 2014 (ukrTenTen20) – Relacja koordynacji i Relacja Modifier of_.

⁸ Word Sketch Difference służy do porównywania kolokacji. Dostępne są trzy opcje: 1) lemma, co porównuje użycie wóh różnych lematów poprzez ich kolokacje; 2) formy wyrazów, co porównuje użycie dwóch różnych form słownych tego samego lematu poprzez ich kolokacje; 3) podkorpus, co porównuje użycie tego samego lematu w dwóch różnych podkorpach tego samego korpusu poprzez ich kolokacje.



Ryc. 13. Word Sketch: różnica binarnej opozycji 'bezpieczeństwo/niebezpieczeństwo' i powiązanych pojęć w korpusie Polish Web 2019 (plTenTen19).

Wyniki kontrastywnego badania świadczą o tendencjach użycia wymienionych wyrazów, zapewniając w ten sposób przestrzeń do refleksji nad społecznymi, politycznymi i emocjonalnymi aspektami obu społeczności.

3. Wnioski

Badanie kontrastywne binarnych opozycji *безпека/небезпека* i *bezpieczeństwo/niebezpieczeństwo* potwierdziło sugestie, że częstotliwość leksemów jest bezpośrednio związana z łatwością/trudnością ich percepcji, a wybór słów synonimicznych w różnych kontekstach jest związany z mechanizmami kategoryzacji i odzwierciedlania powiązań między elementami świata rzeczywistego. Częstotliwość użycia konstrukcji gramatycznych lub składniowych jest określona przez ich produktywność/nieproduktywność w reprodukcji językowej.

Zastosowanie narzędzia *Sketch Engine* okazało się skuteczne dla przeprowadzenia badania nad binarnymi opozycjami *безпека/небезпека* i *bezpieczeństwo/niebezpieczeństwo* w języku ukraińskim i polskim w ujęciu kontrastywnym. Za pomocą funkcji wyszukiwarki *Sketch Engine* zidentyfikowaliśmy wspólne i różne charakterystyki semantyczne oraz dystrybucyjne binarnych opozycji *безпека/небезпека* i *bezpieczeństwo/niebezpieczeństwo* w korpusach internetowych języków ukraińskiego i polskiego. Dane statystyczne istotnie pomogły nam zidentyfikować różnicę między cechami semantycznymi i dystrybucyjnymi porównywanych lematów. Zauważyliśmy specyfikę gramatycznych relacji, w których znajdują się opozycje dla porównywanych języków ukraińskiego i polskiego.

Program *Sketch Engine* jest dobrym narzędziem do rozwiązywania istotnych problemów lingwistycznych, głównie dzięki zastosowaniu metod hybrydowych (kombi-

nacji metod statystycznych i lingwistycznych), pomimo różnic w miarach dokładności i kompletności przez różnicę w syntaktycznym markowaniu stabilnych wyrazów w porównywanych korpusach. Podczas korzystania z tego samego korpusu *Sketch Engine* może przypisywać różne dane częstotliwości (i odpowiednio różne dane rankingowe) do identycznych lematów, co jest motywowane różnym systemowym podejściem do doboru markerów dla deskrypcji relacji gramatycznych. Ustalenie przyczyn tej osobliwości może być obszarem dalszych badań.

Bibliografia

- Aarts, J. (1991), *Intuition-based and observation-based grammars*, (w:) K. Aijmer/ B. Altenberg (red.), *English Corpus Linguistics*. London: Longman, 44–62.
- Baker, M. (1995), *Corpora in Translation Studies: An Overview and Some Suggestions for Future Research*, (w:) „Target” 7(2), 223–243.
- Barry, A.K. (2002), *Linguistic Perspectives on Language and Education*. Bloomsbury Publishing.
- Danesi, M. (2009), *Opposition theory and the interconnectedness of language, culture, and cognition*, (w:) „Sign Systems Studies” 37(1/2), 11–42.
- Dundes, A. (1997), *Binary Opposition in Myth: The Propp/Lévi-Strauss Debate in Retrospect*, (w:) „Western Folklore. Western States Folklore Society” 56 (1), 39–50.
- Iskakbekovna, S.G./ B.S. Rezhepovich (2023), *Binary Oppositions as a Method of Conceptualization: A Case of Codex Cumanicus*, (w:) „International Journal of Society, Culture & Language” 11(3), 160–161.
- Lévi-Strauss, C. (1974). *Structural anthropology*. Basic Books.
- Ngula, R.S. (2017), *Corpus Methods in Language Studies*, (w:) D.D. Kuupole (red.) *Perspectives on Conducting and Reporting Research in the Humanities*, Cape Coast: University of Cape Coast Press, 205–223.
- Paradis, C. (2016), *Corpus methods for the investigation of antonyms across languages*, (w:) *The Lexical Typology of Semantic Shifts*, Boston: De Gruyter Mouton, 131–156.
- Putri, A./ P. Sarwoto (2016), *Saussurian Binary Opposition as the Narrative Structure of Williams’ Summer and Smoke*, (w:) „Journal of Language and Literature” (16), 82–98.
- Smith, J. (1996), *Linguistic Evolution*, (w:) „An Historical Study of English”. London: Routledge, 39–52.
- Utker, A. (1974), *On the binary opposition*, (w:) „Linguistics” 12 (134), 73–93.