



Elektroniczny tezaurus czasowników ukraińskich: problemy i metodyka układania

Olena Siruk
Laboratorium Lingwistyki Komputerowej
Narodowy Uniwersytet Kijowski imienia Tarasa
Szewczenki
Kijów, Ukraina
olebosi@gmail.com



Treść

- Tezaurusy Komputerowe (TK): po co i jak?
 - Stosowanie TK w lingwistyce
 - Terminy kluczowe
- Cel badania
- Etapy badania
 - Etap przygotowawczy
 - Teoretyczne problemy układania TK
 - Etap analizy
 - Relacje leksykalno-semantyczne w TK
 - Czasownik w tezaurysie: jaki on jest?
 - Etap syntezy
 - Synoptyczny schemat TK czasowników mówienia
 - Fragment schematu synoptycznego w TK
 - Strefowa struktura artykułu TK
 - Przykład artykułu TK
 - Mikroartykuł czasownika w TK języka ukraińskiego
- Przewagi komputerowego przedstawienia danych lingwistycznych
- Osiągnięcia tworzenia TK



Tezaurusy Komputerowe (TK): po co i jak?

- ❑ Rozwój leksykografii ukraińskiej
- ❑ Potrzeby użytkowników w systematyzowaniu informacji
- ❑ Rozwój technologii komputerowych

Projektowanie ogólnych i specjalistycznych
tezaurusów



Opracowanie **formalizowanej metodyki**
układania tezaurusów





Stosowanie TK w lingwistyce

- 1) wieloplanowany system poradniczy;
- 2) instrumentarium badań;
- 3) sposób przedstawienia, utrwalenia jednostek i relacje w leksykalno-semantycznej strukturze podjęzyków;
- 4) środek do studiowania języków;
- 5) środek wymiaru znaczenia bliskości tekstów;
- 6) automatyczna morfologiczna analiza tekstu;
- 7) leksykon dla gramatyk formalnych.





Cel badania

- opracowanie formalizowanej metodyki układania komputerowego tezaury czasowników języka ukraińskiego
- aprobacja metodyki na materiale czasowników leksykalno-semantycznego pola (LSP) mówienia



Terminy kluczowe

- 1) **Tezaurus**, albo **słownik ideograficzny** – ogólny słownik niealfabetycznego typu, w którym eksplicytnie są odzwierciedlone systemowe leksykalno-semantyczne relacje między jednostkami.
- 2) **Tezaurus komputerowy (TK)** – przedstawiony za pomocą komputera *słownik ideograficzny*.
- 3) **Koncept** - pojęcie, podstawowe znaczenie, z którym są związane warianty leksykalno-semantyczne.
- 4) **Ideografia komputerowa (IK)** – pododdział leksykografii komputerowej, poświęcony badaniu tezaurusów komputerowych, opracowaniu metodyki ich układania i stosowania.
- 5) **Formalizacja** - opis i przedstawienie informacji językowej w postaci, przydatnej do opracowywania komputerowego.





Etapy badania

- 1) Etap przygotowawczy ▶
- 2) Etap analizy ▶
- 3) Etap syntezy ▶
- 4) Etap eksperymentalny ▶



Etap przygotowawczy

- 1) Klasyfikacja TK
- 2) Systematyzacja terminologii IK
- 3) Zbiór teoretycznych i technologicznych danych, dotyczących zasad metodyki układania TK




Teoretyczne problemy układania TK

1. TK: jednostki, relacje, metody analizy, struktura.
2. Osobliwości ideograficznego opisu czasowników.
3. Układanie TK jako procedura formalizowana.
4. Sposoby tworzenia konceptów.
5. Dobór nazw do konceptów.





Etap analizy

1. **Dobór metod do analizy jednostek** (słowotwórcza analiza leksemów LSP mówienia, syntaktyczna analiza zdania lustracyjnego, syntaktyczna analiza tłumaczenia, komponentowa analiza tłumaczenia, stopienewa identyfikacja leksyki).
2. **Ustalanie stosunków między jednostkami** (hyponymia, synonimia, antonimia, derywatywne i inne typy relacji). 
3. **Algorytmizacja** procesów analizy materiału.





Relacje leksykalno-semantyczne w TK

- 1) **hierarchiczne międzyczasownikowe** relacje, albo hyponymia, w szczególności hyponymia derywatywna, przedstawione hyponimem, hyperonimem i czasownikami rodzajów działania (ДРД).
- 2) **międzyczasownikowe** relacje jednego poziomu, czyli **synonimia** (przedstawiona absolutnymi synonimami, w szczególności fonetycznymi wariantami czasowników, stylistycznymi i derywatywnymi synonimami) i **antonimia** (przedstawiona antonimami).
- 3) Relacje czasownika z innymi częściami mowy, które opierają się na derywacji czasownikowej i walencyjnym potencjale czasownika.





Czasownik w teaurusie: jaki on jest?




Ponieważ w czasownikach przeważa semantyka **sygnifikatywna**, one określają **relacje między realiami**. Dla nich charakterystyczne są:

- ❑ **wewnętrzne, sygnifikatywne** uwarunkowanie wyboru konceptów;
- ❑ **indukcyjne** podejście do uporządkowywania leksemów;
- ❑ celowość stosowania **metod komponentowej analizy tłumaczenia i stopenowej identyfikacji znaczenia** dla analizy czasowników oraz elementów **słotwórczej analizy** leksemów i **syntaktycznej analizy tłumaczeń słownikowych** ;
- ❑ relacje, polegające na **typach słotwórczych (ДРД) i rekcji** (podstawa dla relacji między czasownikiem a innymi częściami mowy);
- ❑ **obecność** umiarkowano szerokiej sieci hyponimii, synonimii, antonimii;
- ❑ **nieobecność** stosunku typu «część-całe», taksonomii;
- ❑ niska częstotliwość w teaurusach terminologicznych, nawet innojęzycznych.





Etap syntezy

- 1) Kształtowanie **makrostruktury** TK czasowników (budowa synoptycznego schematu tezaurusa, projektowania wejść do słownika). 
- 2) Okreslenie **mikrostruktury** tezaurusa (opracowanie słownikowego formatu artykułu, oświetlania struktury relacji międzyczasownikowych). 
- 3) **Integracja** czasownikowej części TK w komputerowy tezaurus języka ukraińskiego (na podstawie relacji czasownika z innymi częściami mowy). 





Synoptyczny schemat TK czasowników mówienia

- Діяльність мовленнєва – 1 рівень
 - Висловлення думки / почуття (висловлювати) – 2 рівень
 - Обмін думками (розмовляти) – 2 рівень
 - Особливості вимови (вимовляти) – 2 рівень
 - * багато, беззмістовно, про неістотне – 3 рівень
 - * басом – 3 рівень
 - * включаючи свої слова в чиєсь мовлення – 3 рівень
 - * грубо – 3 рівень
 - * для записування іншою особою – 3 рівень
 - * довго, захоплюючись розмовою – 3 рівень
 - * дотепно – 3 рівень
 - * дуже голосно – 3 рівень
 - * * з негативним наслідком – 4 рівень
 - * * з позитивним наслідком – 4 рівень
 - * * один раз – 4 рівень
 - * * раз по раз – 4 рівень
 - * * постійно – 4 рівень [...]
 - * чітко – 3 рівень [...]
 - Повідомлення інформації (повідомляти) – 2 рівень [...]
- Здатність, спроможність, уміння – 1 рівень [...]



Fragment synoptyki schematu w TK

діяти

Find Add Semapole Sort Edit

- діяти ГФ(34591 _51)||f|Робити, здійснювати що-небудь, виявляти якусь діяльність.||
 - діяльність мовленнєва __ (1100006 _1)C1
 - особливості вимови __ (1100009 _1)C1
 - висловлення думки/почуття __ (1100003 _1)C1
 - обмін думками __ (1100007 _1)C1
 - повідомлення інформації __ (1100002 _1)C1
 - діяти ГФ(34591 _51)1C||f|Робити, здійснювати що-небудь, виявляти якусь діяльність.||
 - буття, існування __ (1100005 _1)C1
 - говорити ГФ(28838 _4)C1|||бути доказом чого-небудь, указувати на щось
 - діяти ГФ(34591 _51)1C||f|Робити, здійснювати що-небудь, виявляти якусь діяльність.||
 - здатність, спроможність, уміння __ (1100004 _1)C1
 - говорити ГФ(28838 _5)C1|||володіти мовою
 - говорити ГФ(28838 _1)C1||f|мати здатність висловлювати думки, почуття
 - діяти ГФ(34591 _51)1C||f|Робити, здійснювати що-небудь, виявляти якусь діяльність.||
 - вплив __ (1100010 _1)C1
 - впливати ГФ(24220 _1)C1|||діяти певним чином на кого-, що-небудь
 - вплив словесний __ (1100011 _1)C1
 - діяти ГФ(34591 _51)1C||f|Робити, здійснювати що-небудь, виявляти якусь діяльність.||



Strefowa struktura artykułu TK

- strefa **stosunków hierarchicznych**;
- strefa **rodzajów działania**:
 - początku działania (інгресивне);
 - końca działania (фінітивне);
 - długotrwałego działania ograniczającego (делімітативне);
 - delimitatywnego działania intensywnego;
 - długotrwałego działania skończonego (пердуративне);
 - działania nagromadzonego (кумулятивне);
 - działania jednokrotnego (семельфактивне);
 - działania zwrotnego (зворотне);
 - długotrwałego działania periodycznego (періодичне);
 - działania narastającego (аугментативне);
 - działania nasycalnego (сатуративне)
 - wygórowanego działania dokonanego;
- strefa **synonimii**;
- strefa **antonimii**;
- strefa **dodatkowej semy** (zawiera reprezentanty pola, w których sema mówienia jest dyferencjalna).



Przykład artykułu TK

вимовляти	
1.	передавати голосом звуки, слова чужої мови, говорити певним чином
Гіперонім	особливості вимови
Гіпонім 35	базікати [*] , басити [*] , бевкати [*] , бубніти [*] , випалювати [*] , витискувати [*] , вишкірятися [*] , відрубувати [*] , вставляти [*] , гаркавити [*] , грубіянити [*] , джеркотати [*] , диктувати [*] , жартувати [*] , жартувати [*] , забалакуватися [*] , заговорюватися [*] , закидати [*] , карбувати [*] , кричати [*] , мугикати [*] , мурмотіти [*] , натякати [*] , повторювати [*] , рокотати [*] , рубати [*] , сичати [*] , скрипіти [*] , співати [*] , стогнати [*] , тягнути [*] , хрипіти [*] , шепелявити [*] , шепотіти [*] , шипіти [*]
Синонім 1	проговорювати [*]
Фінітив 1	промовити [*]
Пердуратив 1	проговорити [*]
Сенельфактив 1	вимовити [*]
Обернена дія 1	вимовлятися [*]
СемантГрупа 4	наголошувати [*] , примовляти [*] , виговорюватися [*] , спотикатися [*]
2.	говорити що-небудь, промовляти
Гіперонім	виражати
Гіпонім 2	вимовити [*] , мовити [*]
3.	дорікати
Гіперонім	докоряти
4.	<i>розм.</i> Благати, випрохувати. Доновляючись про що-небудь, домагатися якоїсь вигоди, одержання чогось і т. ін.
Гіперонім	випрохувати



Mikroartykuł czasownika w TK języka ukraińskiego

базікати	
1.	говорити багато, безмістовно, про щось неістотне, не варте уваги
Гперонім	вимовляти
Синонім 9	нести [•] , верзти [•] , варнякати [•] , просторікувати [•] , ляпати [•] , торочити [•] , плескати [•] , молоти [•] , патякати [•]
Кумулятив 1	набазікати [•]
Надзавершена дія 1	добазікатися [•]
Діяч 1	базіка [•]
Процес 1	базікання [•]
Атрибут 1	балакучий [•]
2.	<i>розн.</i> піддавати розголосові
Гперонім	розголошувати

$$6 + \begin{array}{l} \square \text{ Strefa } \mathbf{substanty} \mathbf{w} \mathbf{6} \mathbf{w} \\ \square \text{ Strefa } \mathbf{atrybuty} \mathbf{w} \mathbf{6} \mathbf{w} \\ \square \text{ Strefa } \mathbf{adwerbaty} \mathbf{w} \mathbf{6} \mathbf{w} \end{array} = 9 \text{ stref}$$



Etap eksperymentalny

- aprobacja opracowanej metodyki na czasownikach, będących reprezentantami innych LSP;
- stosowania TK jako instrumentarium badania tekstów artystycznych (L. Kostenko i W. Stusa).



Przewagi komputerowego przedstawienia danych lingwistycznych

- Rozszerzenie objętości informacji lingwistycznej w teaurusie czasowników przez oboczne układanie ideograficznego, dwujęzykowego, synonimicznego, antonimicznego słowników;
- Możliwość zjednoczenia w jednym systemie trzech hipostaz (semantyczny model leksyki; wieloplanowany system informacyjny, edukacyjny i poradniczy; instrumentarium badań leksykalno-semantycznych);
- Możliwość integracji produktu do sieci oprogramowania lingwistycznego.



Osiągnięcia tworzenia TK

- ❑ Opracowanie formalizowanej metodyki konstruowania tezaurusów komputerowych jak ogólnych, tak i specjalistycznych.
- ❑ Tworzenie słownika specjalistycznego ideografii komputerowej i tezaurusa ukraińskich czasowników LSP mówienia.
- ❑ Modelowanie leksykalno-semantycznego systemu języka, gdzie każdy składnik struktury opisywany jest jako jednostka systemowa.
- ❑ Możliwość wykorzystywania tezaurusa w eksperymentach i badaniach lingwistycznych (użycia tezaurusa czasowników jako komputerowe instrumentarium badań tekstów artystycznych).
- ❑ Możliwość integracji TK do automatycznych systemów obróbki tekstów; użycia jego jako systemu informacyjnego, w celach dydaktycznych.



Laboratorium Lingwistyki Komputerowej (p. 130)
Instytut Filologii
Narodowy Uniwersytet Kijowski imienia Tarasa
Szewczenki
olebosi@gmail.com

- <http://linguist.univ.kiev.ua>;
- <http://mova.info>
- tel.: (+38-044)-239-33-54

DZIĘKUJĘ ZA UWAGĘ!