

ROBERT BOROCH

FORMALNA ANALIZA KONCEPTUALNA – REPREZENTACJA WIEDZY – PRZEKŁAD

WSTĘP

Do przekładu maszynowego wykorzystałem programy firmy Techland z roku 2005 oraz Kompas z roku 2003 wspomagających przekład, a także, dla porównania, wersji programów z roku 2009 (Techland) oraz 2010 (Kompas).

W celu lepszej orientacji wprowadziłem następujące oznaczenia:

1. Techland₂₀₀₅ dla: Techland, *English Translator* 1.0 z roku 2005;
2. Kompas₂₀₀₃ dla: Kompas, *Tłumacz polsko-angielskiego* 2.0 z roku 2003;
3. Techland₂₀₀₉ dla: Techland, *English Translator XT2* z roku 2009.
4. Kompas₂₀₁₀ dla: Kompas, *Tłumacz i słownik języka angielskiego wersja 7.05* z roku 2010.

W tabeli 1 wyjaśnione zostały najważniejsze terminy, którymi posługuję się w artykule.

Tabela 1. Terminologia przyjęta w artykule

Skrót	Nazwa w języku angielskim	Nazwa w języku polskim	Wyjaśnienie
FCA	<i>Formal Concept Analysis</i>	Formalna Analiza Konceptualna	Metoda analizy (przetwarzania) danych: matematycznych i językowych.
KR ¹ / Reprezentacja wiedzy KR	<i>Knowledge Representation</i>	Reprezentacja wiedzy	Zbiór informacji powiązany relacjami semantycznymi, które są tworzone w konkretnym celu ² .
<p>Problem reprezentacji wiedzy jest bardzo złożony i nie istnieje w tej chwili jedna definicja reprezentacji wiedzy, która byłaby akceptowana bez zastrzeżeń. Najbardziej adekwatną definicją reprezentacji wiedzy jest definicja zaproponowana przez Johna F. Sowę w roku 2000, która odnosi się do sposobu przetwarzania wiedzy o konkretnej rzeczywistości, wraz z metodami wnioskowania (interferencji) oraz procesem zarządzania informacją³. Należy tu rozróżnić: 1. reprezentację wiedzy o implementacji w języku naturalnym i 2. reprezentację wiedzy o implementacji w języku sztucznym.</p> <p>Reprezentacja wiedzy KR nie jest tym samym co korpus tekstowy na gruncie lingwistyki korpusowej (ang. <i>Corpus Linguistics</i>)⁴. Korpus jest to: 1. „zbiór pisanych lub mówionych tekstów” („a collection of written or spoken texts”)⁵ i 2. zbiór pewnej wiedzy lub materiałów „dowodowych”, w szczególności zbiór utrwalonych wypowiedzi służących za podstawę analizy opisowej języka („a collection or body of knowledge or evidence; especially: a collection of recorded utterances used as a basis for the descriptive analysis of a language”)⁶. Przykładem korpusu tekstowego może być <i>The Rosetta Project</i>⁷, <i>The British National Corpus</i> (BNC) i <i>Google Scholar Corpora</i>⁸.</p> <p>W polskiej literaturze przedmiotu funkcjonuje potoczne rozumienie terminu <i>korpus tekstowy</i>. Na przykład Sambor Grucza podaje: „[Korpus tekstowy] to każdy zbiór pisanych lub mówionych tekstów sformułowanych w określonym języku (językach), które zostały zebrane i wybrane według określonych kryteriów”⁹. Uję-</p>			

¹ W moich dotychczasowych artykułach *reprezentację wiedzy* oznaczałem symbolem RW, co zmieniam na angielski skrót KR – *Knowledge Representation*.

² Zob. J.F. Sowa, *Knowledge Representation: Logical, Philosophical, and Computational Foundations*, Pacific Grove: Brooks Cole Publishing Co. 2000.

³ Tamże.

⁴ Zob. N.S. Dasha, *Corpus Linguistics. An Introduction*, New Delhi 2008; R.H. Baayen, *Analyzing Linguistic Data. A Practical Introduction to Statistics Using*, Cambridge 2008; J. Newman, *Corpus Linguistics and Variation in English: Theory and Description*, „World English” 3(2012), nr 4, s. 557-559; C. Gabrielatos, T. McEnery, P.J. Diggle, P. Baker, *The Peaks and Troughs of Corpus-Based Contextual Analysis*, „International Journal of Corpus Linguistics” 17(2012), nr 2, s. 151-175.

⁵ *Oxford Advanced Learner's Dictionary*, red. J. Turnbull, Oxford 2010, s. 339. Tłumaczenie R.B.

⁶ *Webster Dictionary*, <http://www.merriam-webster.com/dictionary/corpus> (tłumaczenie i podkreślenia R.B.).

⁷ <http://rosettaproject.org/>

⁸ Zob. V. Brezina, *Use of Google Scholar in corpus-driven EAP research*, „Journal of English for Academic Purposes” 11(2012), s. 319-331.

⁹ *O konieczności tworzenia korpusów tekstów specjalistycznych*, w: *W kręgu teorii i praktyki lingwistycznej*, red. S. Grucza i in., Warszawa 2007, s. 108 (za: M. Łukasik, *Narzędzia*

cie korpusu tekstowego proponowane przez Gruczę jest akceptowalne jako rozumienie potoczne. Biorąc jednak pod uwagę wymogi formalne, propozycja tego badacza musi być uzupełniona, ponieważ termin *zbiór* sugeruje już jakieś uporządkowanie obiektów pod pewnym względem, według jakiejś kategorii. Grucza wprowadza dodatkowy parametr porządkujący (przyporządkowujący), ponieważ w każdym zbiorze tekstów wyróżnia podzbiór tekstów o własnościach szczególnych, co może prowadzić do paradoksu, który jest znany w literaturze przedmiotu pod nazwą *zbioru zbiorów*. Inne rozumienie korpusu tekstowego reprezentują Tony (Anthony) McEnery i Andrew Wilson¹⁰ albo Susan Hunston¹¹. Cele językoznawstwa korpusowego obecnie są inne¹². Wysiłki badaczy koncentrują się na opracowaniu interfejsu, w którym możliwe jest symulowanie naturalnej (nie sztucznej) komunikacji. Oznacza to, że dany tekst został napisany lub wypowiedziany dla celów komunikacyjnych, a nie – przygotowany na potrzeby korpusu.

Warto zwrócić także uwagę na inne cechy korpusu: „korpus jest wielkim zbiorem reprezentatywnych fragmentów („próbek” – *samples*) różnych rodzajów tekstów dotyczących najróżniejszego użycia języka. Innymi słowy: korpus reprezentuje potencjalnie nieograniczoną liczbę tekstów. Korpus charakteryzuje się tym, że: a) nadaje się do współpracy ze środowiskiem komputerowym, b) umożliwia sprawne przetwarzanie i wykorzystanie zgromadzonych danych, c) jest reprezen-

lingwistyki korpusowej w warsztacie terminologa, terminografa i tłumacza tekstów specjalistycznych (cz. I), w: *Debiuty naukowe. Wiedza – korpus – słownik*, red. M. Łukasik, Warszawa 2007, s. 25).

¹⁰ *Corpus Linguistics*, Edynburg 1996 (za: D a s h, *Corpus Linguistics...*, Kindle location: 3872-3873 oraz 3795-3796).

¹¹ *Corpora in Applied Linguistics*, Cambridge 2002 (za: D a s h, *Corpus Linguistics...*, Kindle location: 3872-3873 oraz 3795-3796).

¹² S.T. Gries, *Quantitative Corpus Linguistics with R: A Practical Introduction*, New York–London 2009; P. Baker, *Sociolinguistics and Corpus Linguistics*, Edinburgh 2010; *Corpus Linguistics. An International Handbook*, t. I, red. A. Lüdeling, M. Kytö, Berlin–New York 2008; *Corpus Linguistics. An International Handbook*, t. II, red. A. Lüdeling, M. Kytö, Berlin–New York 2009; *Corpus Linguistics Around the World*, red. A. Wilson, D. Archer, P. Rayson, Amsterdam, New York 2006; *Corpus Linguistics Beyond the World. Corpus Research from Phrase to Discourse*, red. E. Fitzpatrick, Amsterdam, New York 2007; S. Hoffman, *Grammaticalization and English Complex Prepositions. A Corpus-Based Study*, London–New York 2005; T. McEnery, A. Wilson, *Corpus Linguistics. An Introduction*, Edinburgh 2001; *The Dynamics of Linguistic Variation. Corpus Evidence on English Past and Present*, red. T. Nevalainen, I. Taavitsainen, P. Pahta, M. Korhonen, Amsterdam–Philadelphia 2008; *The Routledge Handbook of Corpus Linguistics*, red. A. O’Keeffe, M. McCarthy, London–New York 2010.

¹³ Parafraza oryginalnego tekstu – R.B. Przy przekładzie parafrazowanego fragmentu nie udało się zachować zamysłu autorów oryginału, którzy omówili poszczególne cechy korpusu jako rozwinięcie liter tworzących nazwę *korpus* (ang. *corpus*). Oryg.: „[...] a corpus contains a large collection of representative samples obtained from texts covering wide varieties of language use from numerous domains of interaction. Therefore, a corpus is Capable Of Representing Potentially Unlimited Selections of text. Taking all these factors into careful consideration, we can explain the term, from the features it denotes, in the following way: C: Compatible to computer O: Operational in research and application R: Representative of the source language P: Processeable by both man and machine U: Unlimited in the amount of language data S: Systematic in formation and text representation” (D a s h, *Corpus Linguistics...*, Kindle location: 1097-1104).

	<p>tatywnym świadectwem danego języka, którego dotyczy, d) może być wykorzystywany zarówno przez człowieka, jak i przez maszyny, e) ma nieograniczoną pojemność przechowanych danych językowych, f) zapewnia systematyczność w porządkowaniu i przechowywaniu tekstów¹³.</p> <p>Korpus tekstowy należy do innej kategorii ontologicznej i epistemologicznej niż reprezentacja wiedzy KR. Na potrzeby niniejszego artykułu można założyć wstępnie, że korpus stanowi reprezentację wiedzy KR wtedy, gdy zostanie oparty konkretną sygnaturą w konkretnym celu. Zmiana celu powoduje zmianę sygnatury oraz elementów oznaczonych przez ową sygnaturę.</p> <p>W wypadku reprezentacji wiedzy KR zakładam za Sową, że jakaś baza wiedzy jest indeksowana w konkretnym celu, co wyznacza przyporządkowanie elementów; zindeksowanym elementom zostaje następnie przypisana jakaś sygnatura porządkująca; to właśnie rozumiem pod pojęciem reprezentacji wiedzy KR.</p> <p>Reprezentacja wiedzy KR jest tworzona w konkretnym celu, przez co ma charakter szczególnie, a nie ogólny. Cechą konstytutywną korpusu tekstowego może być jakaś konkretna liczba reprezentacji wiedzy KR, jeżeli możliwe jest zakończenie procesu relacyjnego, co przypomina przeszukiwanie bazy danych pod kątem konkretnej informacji. Innymi słowy: jeden korpus tekstowy może być podstawą wielu reprezentacji wiedzy KR. Problem ten wymaga szerszego oglądu teoretycznego.</p>		
SD/ Depozyt semantyczny SD	<i>Semantic Deposit</i>	Depozyt semantyczny	<p>Jakakolwiek wartość znacząca. SD charakteryzuje się tym, że:</p> <ol style="list-style-type: none"> 1. postrzegający człowiek wie, że to, co postrzega, ma jakieś znaczenie, 2. znaczenie to nie jest jednak sformułowane w postaci definicji realnej lub nominalnej. <p>Innymi słowy: X coś znaczy, lecz nie wiem, co znaczy. Pozwala to na uruchomienie procedury „poszukiwania znaczenia”.</p>
Q	<i>Model Q</i>	Model Q Q model	<p>Termin zaproponowany przez Umberto Eco: „model (który będę nazywał Modelem Q) jest złożony z licznych węzłów (<i>nodes</i>), między którymi istnieją różnego typu połączenia asocjacyjne”¹⁴.</p>
Q _s	<i>Model Q_s</i>	Model Q _s Q _s model	Semantyczny model statyczny, formalny.
Q _D	<i>Model Q_D</i>	Model Q _D Q _D model	Semantyczny model dynamiczny, konceptualny.
jednostka	<p>Pojęcie <i>jednostki</i> jest w artykule rozumiane jako część większej całości, jakiegoś modelu lub systemu. W artykule <i>jednostka</i> jest używana w następujących kontekstach:</p> <ol style="list-style-type: none"> 1. jednostka języka (jednostka leksykalna – jednostka systemu językowego), 2. jednostka reprezentacji wiedzy KR, jednostka robocza reprezentacji wiedzy KR, 3. jednostka hybrydowa modelu Q, czyli jednostka o własnościach formalno-konceptualnych. 		

W tabeli 2 znajduje się wyjaśnienie oznaczeń i symboli przyjętych w artykule.

¹⁴ Oryg.: „[...] model (which will from now be called Model Q) is based on a mass of nodes interconnected by various types of associative links” (U. Eco, *A Theory of Semiotics*, Indiana 1976, s. 122). Tłumaczenie R.B.

Tabela 2. Oznaczenia przyjęte w artykule

Termin	Sposób zapisu	Przykład
depozyt semantyczny SD	brak specjalnego oznaczenia	depozyt semantyczny SD garnek
reprezentacja wiedzy KR	„”	reprezentacja wiedzy KR „fizyka”
wartość	□	wartość [grobowiec]
artefakt	brak specjalnego oznaczenia	artefakt piramida
dana	□	dana [piramida]
termin	kursywa	termin <i>światło</i>
koncept	kursywa	koncept <i>światło</i>
jednostka	kursywa	jednostka <i>światło</i>
obiekt	□	obiekt [A, B, C]
atrybut	□	atrybut [grobowiec]
sygnatura		
1. Zastosowanie ogólne	□	sygnatura [polski]
2. Zastosowanie szczególne, gdzie X oznacza indeks (ogólny), zaś n oznacza sygnaturę (szczególny).	X_n	Na przykład: Język _{polski} Kod językowy _{angielski}

Powyższe oznaczenia stosuję także w innych swoich pracach (wyjątek – patrz przypis 1).

Problem ekwiwalencji przekładowej został w niniejszym artykule pominięty, ponieważ skłaniam się ku opinii Jerzego Pieńkosa, że „[...] wszelki przekład jest zawsze ekwiwalencją”¹⁵. Wszelkie odnoszenie się do ekwiwalencji przekładowej – dynamicznej czy formalnej (ang. *dynamic or formal equivalence*) – spowoduje powrót do sporu o dokładność przekładu, mimo że powszechnie wiadomo, iż nie wypracowano w tej kwestii jednoznacznego stanowiska¹⁶.

Metodologia FCA w mojej propozycji koncentruje się na sposobie zarządzania danymi w reprezentacji wiedzy KR oraz potencjalnej możliwości ulepszenia algorytmu przekładowego. Pojęcie algorytmu przekładowego także wymagałoby wyjaśnienia, na które nie mogę sobie w ramach i tak już obszernego artykułu pozwolić, dlatego odsyłam zainteresowanego Czytelnika do lektury pracy Thomasa H. Cormena, Charlesa E. Leisersona, RONALDA L. RIVESTA i CLIFFORDA STEINA *Wprowadzenie do algorytmów*¹⁷. W niniejszym artykule przyjmuję potoczne rozumienie algorytmu, czyli: „[...] algorytm jest pewną ściśle określoną

¹⁵ *Podstawy przekładoznawstwa. Od teorii do praktyki*, Kraków 2003, s. 170.

¹⁶ Tamże, s. 172.

¹⁷ Przeł. K. Diks, A. Malinowski, D. Roszkowska, D. Rytter, W. Rytter, Warszawa 2012.

procedurą obliczeniową, która dla właściwych danych wejściowych «produkuje» żądane dane wyjściowe zwane wynikiem działania algorytmu»¹⁸.

2. OPIS METODY – FCA

Przedmiotem rozważań w niniejszym artykule są potencjalne możliwości wykorzystania metody FCA: 1. do opisu mechanizmu tworzenia się znaczenia 2. w przekładzie maszynowym. W tym celu wprowadzone zostało rozróżnienie na koncept, pojęcie oraz termin. Proces tworzenia się konceptów, pojęć oraz terminów został przedstawiony na przykładzie jednostki *piramida*, zaś metoda definiowania hybrydowego została wyjaśniona na przykładzie jednostki *światło*.

Przez jednostkę należy w artykule rozumieć część składową obiektu, na przykład:

A : {[piramida] [Cheopsa]},

B : {[piramida] [Chefrena]},

C : {[piramida] [Mykerinosa]}.

Jednostki A, B, C *piramida* tworzą zbiór tego samego typu, co w metodologii FCA nazywa się obiektem. Jednostki A, B, C {[Cheopsa], [Chefrena], [Mykerinosa]} to atrybuty obiektu [piramida].Więcej na temat terminologii FCA piszę w punkcie 2.2.

2.1. FCA – UJĘCIE HISTORYCZNE

FCA jest jedną z metod analizy danych, którą w roku 1982 na gruncie matematyki stosowanej opracował Rudolf Wille. Obecnie metoda FCA jest wykorzystywana w badaniach nad¹⁹:

1. reprezentacją wiedzy KR,
2. zarządzaniem informacją,
3. implementacją wyników zarządzania informacją w konkretnej przestrzeni komunikacyjnej *vel* roboczej²⁰.

¹⁸ Tamże, s. 4.

¹⁹ Zob. R. Wille, *Restructuring Lattice Theory: an Approach Based on Hierarchies of Concepts*, w: *Ordered Sets*, red. I. Rival, Dordrecht, Boston 1982, s. 445-470.

²⁰ Inne możliwości wykorzystania FCA – zob. U. Priss, *Formal Concept Analysis in Information Science*, “Annual Review of Information Science and Technology” 40(2006), s. 521–543; www.upriss.org.uk/top/research.html; taż, *Linguistic Applications of Formal Concept Analysis*, w: *Formal Concept Analysis, Foundations and Applications*, red. B. Ganter,

Za pomocą metod FCA możliwe jest przetwarzanie dużej ilości nie tylko danych matematycznych, lecz także danych językowych²¹. Uważam jednak, że istnieje tu poważne niebezpieczeństwo, że dane językowe mogą zostać odebrane od opisu, co oznaczałoby zerwanie z semantyczną referencją danych językowych i przekształcenie danych językowych w dane matematyczne. W rezultacie wyjściowe dane językowe stałyby się bezużyteczne. Dlatego danym językowym musi zostać nadane jakieś znaczenie: zarówno zewnętrzne, jak i wewnętrzne.

Uważam, że użytecznym narzędziem FCA, które może częściowo rozwiązać opisany wyżej problem, jest semantyczna analiza kontekstowa. Analiza taka, przynajmniej teoretycznie, powinna umożliwić wyznaczenie relacji pomiędzy poszczególnymi jednostkami w konkretnej reprezentacji wiedzy KR. Analiza kontekstowa powinna być, w moim przekonaniu, prowadzona w dwóch etapach. Etap pierwszy to analiza frekwencyjna, która pozwala wyróżnić obiekty, odrzucając te, które nie są istotne ze względu na niskie użycie. Etap drugi to właściwa analiza formalnych konceptów.

2.2. TERMINOLOGIA FCA

Przyjmuję za Utą Priss²², że:

1. elementy jednego typu nazywa się obiektami,
2. elementy innego typu – takie, które można w jakikolwiek sposób połączyć z obiektem – nazywa się atrybutami,
3. zamknięty zbiór obiektów tworzy formalny obiekt wtedy i tylko wtedy, gdy nie jest możliwe dodanie do niego nowych elementów lub kiedy proces dodawania nowych elementów został zakończony,

G. Stumme, R. Wille, Berlin–Heidelberg: Springer Verlag. 2005, s. 149-160, LNAI 3626; <http://upriss.org.uk/papers/fcaic03.pdf>; t a ż, *Formal Concept Analysis as a Tool for Linguistic Data Exploration*, w: *Conceptual Structures in Practice*, red. P. Hitzler, H. Scharfe, Chapman & Hall/CRC Studies in Informatics Series, 2009, s. 177-198; R. B o r o c h, *Formalna Analiza Konceptualna (FCA) w badaniach kulturoznawczych – model analizy obiektu N: prowincjonalizm*, w: *Prowincjonalizm w kulturze europejskiej*, red. B. Płonka-Syroka, K. Marchel, Wrocław 2010, s. 21-46 (badania kulturoznawcze); t e n ż e, *Formalna Analiza Konceptualna (FCA) w badaniach nad postaciami dramatycznymi – jednostka N: sluga na przykładzie Fieraponta w Trzech siostrach i Firsza w Wiśniowym sadzie Antoniego Czechowa*, w: *Czechow współcześnie*, red. R. Strzelecki, G. Guźlak, Bydgoszcz 2010, s. 125-152 (badania literaturoznawcze).

²¹ Zob. P r i s s, *Formal Concept Analysis in Information Science*; t a ż, *Linguistic Applications...*; t a ż, *Formal Concept Analysis as a Tool...*

²² *Formal Concept Analysis in Information Science*.

4. wszystkie zidentyfikowane atrybuty przysługujące obiektom nazywa się formalnymi atrybutami,
5. układ formalnych obiektów oraz formalnych atrybutów wraz z relacjami tworzy formalny kontekst,
6. formalny obiekt i formalny atrybut tworzą formalny koncept wtedy i tylko wtedy, gdy są zamknięte,
7. para formalnych obiektów formalnego konceptu tworzy ekstensję konceptu (jest to zakres nazwy),
8. para formalnych obiektów formalnego atrybutu tworzy intensję konceptu, czyli sposób, w jaki przedmiot się nam przedstawia.

2.3. TEORIA RELACJI

Rozumienie relacji w artykule przyjmuję za teorią relacji w następującej postaci²³:

1. 'Rxy' – dla relacji pomiędzy dwoma elementami,
2. 'Sxyz' – dla relacji między trzema elementami,
3. 'Tx₁,x₂,x₃...x_n' – dla relacji *n*-elementowych²⁴.

Symbole R, S, T oznaczają dowolne relacje, zaś zmienne x, y, z oznaczają zmienne indywidualne, które są nazwami jednostkowymi²⁵.

Zakładam także, że relacje pomiędzy obiektami w reprezentacji wiedzy KR nie są stałe. Zmienność relacji uzależniona jest od zewnętrznych warunków wyjściowych procesu relacyjnego. Przez proces relacyjny rozumiem taki proces, który pozwala na semantyczne przyporządkowanie obiektów i ich atrybutów w takiej formie, która pozwala wyróżnić koncept. Więcej na temat konceptu piszę w punkcie 6.

2.4. PROBLEM ZNACZENIA OBIEKTU W UJĘCIU FCA

Jestem zdania, że w propozycji FCA znaczenie powinno być traktowane jako algorytmiczny proces relacyjny. Przez algorytmiczny proces relacyjny rozumiem taki proces, który wyznacza klasy obiektów oraz przyporządkowuje obiektom atrybuty. Przyporządkowanie uzależnione jest od klasyfikacji obiektu do klasy artefaktów.

²³ L. Gumański, *Wprowadzenie w logikę współczesną*, Warszawa 1990, s. 184.

²⁴ W swoich pracach na oznaczenie relacji w znaczeniu ogólnym stosuję symbol \mathbb{R} .

²⁵ Gumański, *Wprowadzenie...*, s. 184.

Można wyróżnić dwie klasy artefaktów: 1. klasę artefaktów materialnych²⁶ oraz, w moim przekonaniu, 2. klasę artefaktów abstrakcyjnych, chociaż udział ten jest dyskusyjny. Przedwstępnie uznaję, że artefaktem materialnym jest obiekt w jakiejś rzeczywistości, zaś artefaktem abstrakcyjnym jest to, co jest desygnowane²⁷.

Artefakt ma konkretny depozyt semantyczny SD o pewnej wartości. Wartość ta możliwa jest do „odbicia” w danym języku, co wynika z faktu, że artefakt jest realizacją pewnej przyczyny. Przyczyna determinuje sposób zaprojektowania artefaktu, przez co należy rozumieć, że do powstania artefaktu został użyty jakiś algorytmiczny typ myślenia o czymś. Oznacza to, że sam przyczynowy sposób zaprojektowania artefaktu zawiera w sobie implikaty informujące o przeznaczeniu artefaktu albo o sposobie jego użycia²⁸. Artefakty mają „wewnętrzną logikę”, wyjaśniającą przyczynę ich takiej, a nie innej konstrukcji. Jeżeli artefakt nie ma wewnętrznej logiki, to albo artefakt jest skonstruowany niepoprawnie, albo człowiek nie jest w stanie dostrzec wewnętrznej logiki artefaktu, co jednak nie oznacza jej braku²⁹.

3. PROCESY RELACYJNE

W celu lepszego zobrazowania mechanizmu procesów relacyjnych w reprezentacji wiedzy KR posłużę się przykładem. Dana [piramidy w Gizie] jest daną językową o właściwościach artefaktu abstrakcyjnego. Dana [piramidy w Gizie] ma określoną wartość semantyczną, która jest człowiekowi nieznaną. Dopiero po ustanowieniu relacji (proces relacyjny) w reprezentacji wiedzy KR o sygnaturze [piramidy w Gizie] możliwe jest powiązanie danych A, B, C:

[piramida Cheopsa] – A,
 [piramida Chefrena] – B,
 [piramida Mykerinosa] – C.

$$f_1 : A \rightarrow B \rightarrow C$$

Dane A, B, C tworzą obiekt w konkretnej reprezentacji wiedzy KR o sygnaturze [piramidy w Gizie], ponieważ uznaję, czysto hipotetycznie, że proces

²⁶ Zob. P. G a r b a c z, *Logika i artefakty*, Lublin 2006.

²⁷ W artykule został pominięty problem kategoryzacji artefaktów abstrakcyjnych.

²⁸ Zob. G a r b a c z, *Logika i artefakty*.

²⁹ Zob. tamże.

dodawania obiektów jednego typu został w tej konkretnej reprezentacji wiedzy KR „piramidy w Gizie” zakończony.

Istnieje jednak możliwość powiązania danych A, B, C z elementami innego typu:

- [piramida w Lipie] (Polska) – D,
- [piramida w Burdzach] (Polska) – E,
- [piramida w Łaziskach] (Polska) – F.

$$f_2 : D \rightarrow E \rightarrow F$$

Dane D, E, F są atrybutami obiektu o strukturze [A, B, C].

Dziedziny funkcji f_1 i f_2 , mimo że należą do różnych reprezentacji wiedzy KR, mają potencjał umożliwiający skuteczny proces relacyjny. Przez skuteczny proces relacyjny należy rozumieć proces zakończony ustanowieniem konkretnych relacji przyporządkowania, co tworzy w dziedzinach funkcji f_1 i f_2 koncept.

Jeżeli w dziedzinie funkcji f_1 obiekt [piramida] ma znaczenie o sygnaturze [grobowiec], zaś w dziedzinie funkcji f_2 znaczenie o sygnaturze [skarbiec], to do dziedziny funkcji f [piramida, (grobowiec, skarbiec)] jest wprowadzony nowy atrybut w postaci [skarbiec]. Atrybuty [grobowiec] i [skarbiec] mają elementy wspólne z obiektem [piramida] na poziomie, na przykład, identyfikacji ikonicznej. Oznacza to, że pierwotne znaczenie obiektu [piramida] zostało oderwane od rzeczywistego przedmiotu.

Można przypuszczać, że antyczny Egipcjanin w obiekcie [piramida] dostrzega zupełnie inne wartości semantyczne niż współczesny Polak. Dzieje się tak dlatego, że uległy zmianie relacje w reprezentacji wiedzy KR „piramida”. Przekonuje mnie to tylko do stwierdzenia, że konkretny artefakt może być użyty w inny sposób niż jego przeznaczenie, wynikające z jego zaprojektowania. Problem ten może być związany z typem myślenia o artefakcie. Inne użycie artefaktu jest zawsze związane z możliwością realizacji takiego użycia, dlatego, na przykład, hełm (ochronne nakrycie głowy), jeżeli jest wykonany z materiału ognioodpornego, może być użyty w celu zagotowania wody, co jest niezgodne z jego przeznaczeniem, ale jest możliwe³⁰. W tym wypadku tworzą się konkretne, być może tymczasowe, relacje z depozytem semantycznym SD obiektu [garnek] oraz z depozytem semantycznym SD obiektu

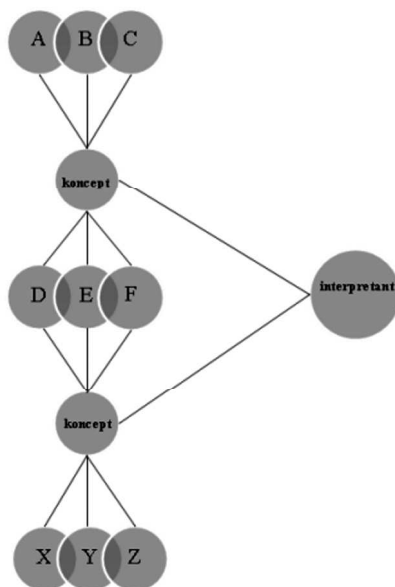
³⁰ Problem ten jest ważny z punktu widzenia tworzenia się znaczenia, czyli znaczenia będącego wewnątrz struktury zdania (artefaktu abstrakcyjnego) albo znaczenia jako przekonania mówiącego. W tym wypadku mówiący musi mieć w swoim umyśle koncepty *garnka* i *gotowania*, bez czego nie będzie możliwe dostrzeżenie w hełmie innych możliwości użycia.

[hełm]. Relacja taka nie mogłaby jednak mieć miejsca, gdyby depozyt semantyczny SD obiektu [garnek] był człowiekowi nieznanym – w tym wypadku [hełm] staje się atrybutem obiektu [garnek].

Konkretna wartość semantyczna jest wynikiem zakończonych procesów relacyjnych w określonej reprezentacji wiedzy KR, czyli procesów umożliwiających łączliwość różnych depozytów semantycznych SD danych, co umożliwia wyłonienie obiektów oraz atrybutów.

Bez procesów relacyjnych dane nie będą mogły być „zawieszane” w reprezentacji wiedzy KR. Wtedy dane są tylko pozbawionym znaczenia zbiorem, który jest lub istnieje³¹, lecz nic z tego nie wynika.

Powróćmy jednak do przykładu piramid. Konceptualne myślenie o artefakcie piramida przedstawia rysunek 1, na którym zostały zaznaczone obszary wspólne, gdzie konkretne obiekty przyjmują pewne wartości, konstituując tym samym jednostkę roboczą, tj. koncept *piramida*.



Rysunek 1. Depozyt semantyczny SD konceptu *piramida*

³¹ Odróżniam tu bycie od istnienia. Przez byty, które są, rozumiem takie byty, które można postrzegać, zaś przez byty, które istnieją, rozumiem byty, które można doświadczać (zob. R. B o r o c h, *Na marginesie semiotyki kultury: problemy tzw. nauk kulturoznawczych*, w: *Znaki czy nie znaki*, red. M. Guławska-Gawkowska, G. Zeldowicz, Warszawa 2013, s. 21-39).

Wielkimi literami A, B, C zaznaczyłem zbiory, w których argumenty funkcji f_1 przyjmują określone wartości. Wielkimi literami D, E, F zaznaczyłem zbiory, w których przyjmują wartości argumenty funkcji f_2 . Zbiory A, B, C, D, E, F... X, Y, Z stanowią dziedzinę funkcji f_1 i f_2 , tworząc depozyt semantyczny SD konceptu *piramida*.

Posłużmy się jeszcze innym przykładem, bardziej związanym z omawianą w niniejszym artykule problematyką.

„Wielki słownik angielsko-polski i polsko-angielski PWN-Oxford” jest zbiorem danych o określonym depozycie semantycznym SD³², które to dane w procesie tworzenia relacji przekształcają się w konkretną semantyczną wartość o ściśle określonym znaczeniu.

W reprezentacji wiedzy KR „słownik PWN-Oxford” mamy do czynienia z jednostkami języka, które uznaję za artefakty abstrakcyjne, mające reprezentację graficzną formułowaną w określonym kodzie. Kod jako artefakt abstrakcyjny sam z siebie informuje o swoim przeznaczeniu oraz o sposobie łączenia jednostek bazowych, które taki kod tworzą. Znaczenie nie jest tu przypisane do danej jednostki kodu, lecz do sposobu jej użycia, który wynika z relacji, w jakie owa jednostka wchodzi z innymi jednostkami tego kodu, a także ze sposobu, w jaki jednostkami kodu posługuje się użytkownik.

W wypadku przekładu konieczne jest odtworzenie semantycznych ekwiwalentów procesów relacyjnych języka wyjściowego w języku docelowym. Nie zawsze jest to jednak możliwe. Brak możliwości odtworzenia procesów relacyjnych ma miejsce na przykład w wypadku hieroglifów, kiedy konieczne jest „przerzucenie” procesu relacyjnego w inną reprezentację wiedzy KR – „egiptologia” czy „archeologia”. Zwróćmy uwagę, że nieumiejętność odczytania znaczeń hieroglifów nie powoduje wymazania zawartego w nich znaczenia³³. Na przykład ostrakony kananejskie są trudne do odczytania ze względu na zachowanie się zbyt małej próbki tekstu, a także ze względu na jakość zapisu. Nawet gdyby jednak nie można było odczytać jakiegoś ostraka-

³² Posłużę się dodatkowo jeszcze innym przykładem: człowiek w artefakcie postrzega przedmiot o określonym kształcie i przeznaczeniu, wynikającym z istoty jego zaprojektowania. Wyobraźmy sobie, iż mamy do czynienia z pewnym zbiorem artefaktów. Człowiek postrzega ów zbiór jako zbiór danych o określonym depozycie semantycznym SD. Następnie, jeżeli będzie to możliwe, ustala relacje semantyczne, w jakich pozostają określone SD w reprezentacji wiedzy KR. Na tej podstawie może stwierdzić, iż postrzegany zbiór artefaktów to, na przykład, galera, mimo iż żaden pojedynczy artefakt sam z siebie takiej informacji człowiekowi nie przekazuje. Jednakże przed przystąpieniem do rekonstrukcji galery człowiek musi mieć koncept *galera*, co ułatwia mu odtwarzanie procesów relacyjnych.

³³ Zob. P. Wilson, *Hieroglyphs. A Very Short Introduction*, Oxford 2003.

konu kananejskiego, to i tak można by stwierdzić, że w tym konkretnym wypadku mamy do czynienia: a) z pismem, b) z pismem kananejskim, ale najprawdopodobniej c) z językiem hebrajskim.

W wypadku kodu językowego o sygnaturze [angielski] procesy rekonstruowania znaczenia w dowolnym obszarze modelowym Q są powszechne, ponieważ rzeczywistość nieustannie wytwarza nowe depozyty semantyczne SD, które są automatycznie wciągane przez procesy relacyjne w siatkę semantyczną określonych jednostek – na przykład *clotted cream/masło*. Jednostka *clotted cream* postrzegana jest przez pryzmat maszynowej podmiany jednostek języka o sygnaturze [polski] jako *gęsta śmietana*, mimo że człowiek postrzegający *clotted cream* widzi najprawdopodobniej polskie *masło*, nie zaś *śmietanę w grudkach, skrzepniętą śmietanę* czy w ogóle *śmietanę*.

Procesy relacyjne w reprezentacji wiedzy KR o sygnaturze [polski] i sygnaturze [angielski] pozwalają na zespolenie *clotted cream* z *gęstą śmietaną* na podstawie podobieństwa wyrabiania tych produktów, a nie na podstawie definicji językowych. W ten sposób powstaje hybrydowy koncept składający się z obiektów [*clotted cream*] i [*gęsta śmietana*].

W koncepcie *clotted cream* znajduje się także termin *scone*, który w reprezentacji wiedzy KR o sygnaturze [polski] jest rodzajem pieczywa albo babeczką. Zatem zdania: *Jan jadł babeczkę z gęstą śmietaną* lub *Jan jadł pieczywo z gęstą śmietaną* nie są tożsame ze zdaniem *John had a scone with a clotted cream*. Okazuje się, że być może formalne procesy relacyjne są niewystarczające w procesie tworzenia znaczenia, bowiem bez empirycznego pozyskania dodatkowej wiedzy znaczenie zdania *John had a scone with a clotted cream* jest w języku polskim niezrozumiałe. Konieczne jest tu odwołanie się do konceptu związanego z wytwarzaniem tych artefaktów, co tłumacz-człowiek czyni automatycznie.

W tłumaczeniu maszynowym zdania *John had a scone with a clotted cream* wygląda to następująco³⁴:

Techland₂₀₀₅

Zdanie wyjściowe:

John had a scone with a clotted cream.

Przekład:

John miał pieczywo z zamienioną się w grudce śmietanką.

Przekład odwrotny I:

³⁴ Tłumaczenie dwukierunkowe z języka angielskiego na język polski oraz wynik tłumaczenia w języku polskim na język angielski.

John was bread with exchanged in the lump the cream.

Kompas₂₀₀₃

Zdanie wyjściowe:

John had a scone with a clotted cream.

Przekład:

John miał rodzaj pieczywa ze śmietaną w grudkach.

Przekład odwrotny I:

John had a scone with the clotted cream.

Techland₂₀₀₉

Zdanie wyjściowe:

John had a scone with a clotted cream.

Przekład:

John miał pieczywo z zamienioną się w grudce śmietanką.

Przekład odwrotny I:

He John was bread with exchanged in the lump the cream.

Kompas₂₀₁₀

Zdanie wyjściowe:

John had a scone with a clotted cream.

Przekład:

John miał babeczkę z gęstą śmietaną.

Przekład odwrotny I:

John had a bun with the clotted cream.

Procedura sprawdzająca (przekład odwrotny) wygląda następująco – na przykładzie Techland₂₀₀₅:

Zdanie wyjściowe I:

John had a scone with a clotted cream.

Wynik przekładu:

John miał pieczywo z zamienioną się w grudce śmietanką.

Przekład odwrotny I:

John was bread with exchanged in the lump the cream.

Zdanie wyjściowe II:

John was bread with exchanged in the lump the cream.

Wynik przekładu:

John był chlebem z wymieniony hurtem śmietankę.

Przekład odwrotny II:

John was bread from exchanged the cream wholesale.

Dopiero przy trzeciej próbie otrzymujemy niezmienny wynik przekładu:

Zdanie wyjściowe III:

John was bread from exchanged the cream wholesale.

Wynik przekładu:

John był chlebem z wymieniony hurt śmietanki.

Widzimy, że dopiero w trzeciej próbie przekładu maszynowego algorytm tłumaczący Techland₂₀₀₅ utrzymuje stabilną, chociaż niedorzeczną, implementację.

Powyższe próby przekładu maszynowego są nonsensami semantycznymi. Zwróćmy uwagę, że kontekstowa analiza semantyczna Techland₂₀₀₅ nie ma nic wspólnego z właściwą analizą kontekstową. Działa tu inny mechanizm, który można nazwać „procesem podmieniania kodu”: a) polskiego na b) angielski i vv., według zasady, że wartość semantyczna obiektu p jest tożsama/synonimiczna z wartością semantyczną obiektu q.

Wnioskuje, że algorytm tłumaczący wymaga weryfikacji przede wszystkim od strony potencjalnych możliwości relacyjnych w konkretnych reprezentacjach wiedzy KR języka wyjściowego i reprezentacji wiedzy KR języka docelowego. Konieczne jest, w mojej ocenie, skierowanie wysiłków badawczych w stronę skonstruowania sensownego algorytmu tłumaczeniowego – takiego, który będzie wielopoziomowo analizował relacje obiektów i ich atrybutów. Być może pomocne tu będą gramatyki formułowane na podstawie *Head-driven Phrase Structure Grammar* (HPSG) oraz *Definite Clause Grammars* (DCG)³⁵.

4. MODEL Q (ECO)

Termin *model Q* został zaproponowany przez Eco: „(...) [model Q zakłada, że] [spójność] systemu może być podtrzymywana przez nowe informacje, zaś dalsze dane mogą być wywnioskowane z danych niepełnych/fragmentarycznych³⁶”.

Właściwości modelu Q rozszerzam o część statyczną modelu (Q_S model) oraz część dynamiczną modelu (Q_D model). Przyjmuję, że Q_S jest modelem statycznym i formalnym, zaś Q_D jest modelem dynamicznym i konceptualnym.

³⁵ Zob. A. Przepiórkowski, A. Kupść, M. Marciniak, A. Mykowiecka, *Formalny opis języka polskiego. Teoria i implementacja*, Warszawa 2002.

³⁶ Oryg.: „(...) the model Q supposes that system can be nourished by fresh information and that further data can be inferred from incomplete data” (Eco, *A Theory...*, s. 124) – przeł. R.B.

Model Q_S jest modelem statycznym, co należy rozumieć tak, że jest to model zamknięty w konkretnych przedziałach wartości semantycznych. Zamknięcie pozwala na identyfikację i opis wszystkich obiektów będących w konkretnym przedziale semantycznym, przez co model Q_S staje się modelem formalnym; możliwe jest bowiem sformułowanie aksjomatyki takiego modelu³⁷.

Model Q_D jest modelem niezamkniętym lub częściowo zamkniętym (domkniętym). Jeżeli model Q_D jest modelem niezamkniętym, to jest niepoznawalny, ponieważ nie można zidentyfikować wszystkich obiektów, które wchodzi w jego skład. Wszystkie ustalenia sformułowane w ramach tego modelu mają wówczas charakter tymczasowy, ponieważ nie jest możliwe określenie stałej minimalnej wartości uporządkowania (CONST). Jeżeli model Q_D jest modelem częściowo zamkniętym (czyli domkniętym), to, przyjmując zamknięcie przedziału jako granicę, możemy zidentyfikować niektóre obiekty wchodzące w zakres domkniętego przedziału oraz w sposób ukierunkowany śledzić ewentualne relacje powstające pomiędzy obiektami³⁸. Będziemy wtedy mówić, że mamy do czynienia z przedziałem jednostronnie domkniętym.

Oba modele: formalny (Q_S) i konceptualny (Q_D) są komplementarne i razem tworzą model Q , stanowiąc wynik implementacji procesów relacyjnych w konkretnej reprezentacji wiedzy KR.

5. POŁĄCZENIE MODELI Q_S I Q_D

FCA koncentruje się na połączeniu systemów formalnych i konceptualnych w jednej komplementarnej implementacji, która jest wynikiem procesów relacyjnych w reprezentacji wiedzy KR. Dlatego FCA jako metoda analizy danych używa w procesie tworzenia relacji w „obszarze roboczym” Q reguł systemów formalnych i systemów konceptualnych jednocześnie. W wyniku takiego połączenia wartości depozytów semantycznych SD danych językowych powstaje jednostka o hybrydowej strukturze znaczeniowej. Znaczenie hybrydowe w ramach systemów formalnych opiera się na desygnacyjnych, denotacyjnych i konotacyjnych teoriach znaczeń, zaś w ramach systemów konceptualnych na asocjacyjnych i intencjonalnych teoriach znaczeń. Oba procesy tworzenia się znaczenia w ujęciu FCA są (lub powinny być) jednocześnie i komplementarne,

³⁷ Zob. L. Bertalanffy, *Ogólna teoria systemów. Podstawy, rozwój, zastosowania*, przeł. E. Woydyłło-Woźniak, Warszawa 1984.

³⁸ Zob. tamże.

co prowadzi do powstania jednostki o wspomnianym już hybrydowym znaczeniu. Jednostka hybrydowa ma większe możliwości łączenia się z innymi depozytami semantycznymi SD danych, przez co tworzy nowe, często pozornie odległe od siebie znaczenia. Proces tworzenia się znaczenia takiej jednostki został przedstawiony w niniejszym artykule na przykładzie jednostki *światło*³⁹. Zanim jednak przejdę do opisu i wyjaśnienia tego procesu, przedstawię teoretyczne zagadnienia związane z procesem tworzenia się znaczenia: różnicę między konceptem a terminem oraz sam mechanizm procesów relacyjnych.

6. KONCEPT VS. TERMIN

Wille wprowadził do analizy danych językowych wspomniane już pojęcie konceptu, charakteryzując koncept jako hierarchiczną strukturę semantyczną. Koncepty w propozycji FCA są podstawowymi jednostkami myślenia, które powstają w „obszarze roboczym” dynamiczno-abstrakcyjnego modelu Q_D ⁴⁰. Ów „obszar roboczy” jest wirtualną przestrzenią komunikacyjną, w której zachodzą procesy relacyjne pomiędzy poszczególnymi depozytami semantycznymi SD danych tworzących obiekty. Wszystkie relacje, jako wynik owych procesów, są symultaniczne i algorytmiczne, co umożliwia ich powtarzanie i odtwarzanie. Owa algorytmiczność nie gwarantuje jednak identyczności (synonimiczności) implementacji. Pozostawię ten problem jako mniej ważny.

Modelowi Q_S jako jednemu z „obszarów roboczych” Q przypisałem następujące własności systemowe: 1. własności konkretnej reprezentacji wiedzy KR,

³⁹ Dotychczasowe próby definiowania jednostki *światło* uważam za nieskuteczne, ponieważ odwołują się one do formalnego systemu języka naturalnego (zob. A. D o b a c z e w s k i, *Światło. Próba analizy semantycznej*, w: *Beiträge der Europäischen Slavistischen Linguistik (POLYSLAV)*, red. K. Böttger, M. Giger, B. Wiemer, Bd. 2, München 1999, s. 93-99; t e n ż e, *Próba opisu znaczeń leksemów jasny/jasno/jasne*, „Prace Językoznawcze UWM” 1(1999), s. 17-26; t e n ż e, *O czasownikach świecić i świecić się*, „Acta Universitatis Nicolai Copernici. Filologia Polska” 55(2001), s. 53-60; t e n ż e, *Czy widzieć jest semantycznie proste?*, w: *Studies on the Syntax and Semantics of Slavonic Languages. Papers in Honour of Andrzej Bogusławski on the Occasion of his 70th Birthday*, red. V. S. Chrakovskij, M. Grochowski, G. Hentschel, Oldenburg 2001, s. 121-132; A. D y s z a k, *Językowy obraz światła w Księdze Rodzaju a współczesne znaczenie rzeczownika „światło”*, w: *Język. Teoria. Dydaktyka. Materiały 21 konferencji językoznawczej zorganizowanej w Trzciny k. Jasła w dniach 27-29 maja 1998 roku*, red. B. Greszczuk, Rzeszów 1999, s. 197-205; t e n ż e, *Językowe wyrażenia zjawisk emisji światła*, Bydgoszcz 1999).

⁴⁰ Zob. R. Wille, *Formal Concept Analysis as Mathematical Theory of Concepts and Concept Hierarchies*, w: *Formal Concept Analysis. Foundations and Applications*, red. B. Ganter, G. Stumme, R. Wille, Berlin, Heidelberg 2005, s. 1-33.

w której znajdują się określone artefakty. Oznacza to, że reprezentacja wiedzy KR jest zbiorem wszystkich danych, zaś obszar Q_S to „obszar roboczy” wykorzystujący jedynie niektóre depozyty semantyczne SD danych w procesie relacyjnym – te, które są aktualnie indeksowane. Kolejną własnością obszaru Q_S to jego 2. algorytmiczność jako systemu obliczeniowego, w którym są implementowane relacje semantyczne pomiędzy konkretnymi już obiektami. Oznacza to, że możliwe jest powtarzanie lub odtwarzanie procesów relacyjnych.

Na przykład: człowiek wie, iż dana [światło] ma jakąś wartość semantyczną, czyli że coś znaczy, ale będzie w stanie dowiedzieć się, jaka jest to wartość, dopiero wtedy, gdy będzie: widział/rozumiał/postrzegał relacje między konkretnymi obiektami w ramach konkretnej reprezentacji wiedzy KR.

W początkowym etapie postrzegania obiektu człowiek nie jest w stanie stwierdzić, jakie są to relacje, lub jakie mogą być, ponieważ postrzega jedynie obiekt, o którym wie, iż ma jakiś depozyt semantyczny SD o konkretnej wartości. To, co pozwala człowiekowi wnioskować, iż określony artefakt jest taki lub taki, to rozmyta semantycznie kategoria konceptu, który jest jednostką dynamiczno-abstrakcyjnego modelu Q_D . Dlatego Wille przyjmuje, iż koncept jest jednostką pierwotną oraz elementarną w procesie myślenia o obiekcie, o którym człowiek wnioskuje za pomocą konkretnej wartości semantycznej, tj. terminu.

Jednakże rodzą się tu dodatkowe pytania, które dotyczą:

1. właściwości dynamiczno-abstrakcyjnego systemu reprezentacji wiedzy. Czy ów system ma zastosowanie tylko i wyłącznie do jednostek *L*-języka⁴¹ czy jednostek języka naturalnego i czy jest częścią modelu Q_S ,

2. właściwości komunikacji kulturowej i roli w takiej komunikacji semantycznych depozytów SD w konkretnej reprezentacji wiedzy KR,

3. procesów myślenia, które uwzględniają: a) modus pozyskiwania informacji, b) modus gromadzenia informacji oraz c) modus interpretowania informacji.

Problemy te, mimo iż są bardzo ważne w teoretycznych refleksjach nad FCA, nie będą w tym artykule dyskutowane. Ważniejsze z punktu widzenia przetwarzania danych językowych jest określenie właściwości takich jednostek, jak *koncept* i *termin*, oraz tego, co w wyniku zakończonych procesów relacyjnych może powstać.

Zakładam za Willem, iż roboczą jednostką systemów konceptualnych i formalnych jest koncept, który w wyniku procesów relacyjnych zostaje prze-

⁴¹ Czyli metajęzyka jakiegoś modelu (zob. S.K. Langer, *On Cassirer's Theory of Language and Myth*, w: *The Philosophy of Ernst Cassirer. The Library of Living Philosophers*, red. P. A. Schilpp, Chicago 1949, s. 381-400).

kształcony w termin o konkretnych właściwościach semantycznych, czyli definitywnie niesprzecznych.

W obszarze Q_S operacyjną ramą procesów relacyjnych (jakkolwiek bliżej jeszcze nieokreślonych) są terminy jako jednostki „widoczne” w systemie w postaci pozostałości kodu (reprezentacja graficzna i/lub fonetyczna), a także jako jednostki o prostych definicjach. Terminy w tym rozumieniu są jednostkami myślenia „wcześniejszymi” niż koncepty, ponieważ należąc do statyczno-abstrakcyjnego systemu, reprezentują koherentną wiedzę na temat konkretnej rzeczywistości.

W modelu Q koncept jest kategorią przynależną semantycznej siatce terminów, które eksplikują relacje pomiędzy sądem (czyli tym, co pomyślane) a konkretnym przedmiotem postrzeganym. Przedmiotem postrzegania może być także sam sąd o przedmiocie, dlatego w wypadku postrzegania samego sądu postrzeganie dotyczy kodu oraz języka. Człowiek myśli o pewnym artefakcie zawsze w kategorii terminu. Jednak to, co człowiek postrzega, nie jest tym, co postrzega faktycznie, ponieważ postrzeganie zachodzi tu przez definicję terminu, a nie przez właściwości przedmiotu, do którego termin się odnosi. Nie mogę mówić o *świecie*, jeżeli nie rozumiem terminów: *widzieć/wiedzieć/wierzyć*, ponieważ nie będzie wówczas możliwe sformułowanie zdania *Widzieć/wiedzieć/wierzyć, że coś tam/tu jest*.

Tabela 3 przedstawia różnice pomiędzy konceptem a terminem.

Tabela 3. *Koncept vs. termin* jako jednostki leksykalne języka naturalnego⁴²

Koncept	Termin
brak jednej reprezentacji fonetycznej	ma jedną typową reprezentację fonetyczną
brak jednej reprezentacji graficznej	ma jedną reprezentację graficzną
jednostka należąca do systemu abstrakcyjno-dynamicznego	jednostka należąca do systemu abstrakcyjno-statycznego
ma ograniczony wpływ na proces komunikacji, z powodu braku precyzji	ma duży wpływ na proces komunikacji, ze względu na precyzję
procesy relacyjne w reprezentacji wiedzy oparte są na prostych terminach – terminach wyjściowych <i>vel</i> jednostkach semantycznie prostych	procesy relacyjne w reprezentacji wiedzy oparte są na złożonych terminach
nie można formułować twierdzeń z wykorzystaniem konceptów	relacje między różnymi terminami w konkretnej reprezentacji wiedzy są twierdzeniami
subiektywna interpretacja relacji depozytów semantycznych	uniwersalna interpretacja relacji depozytów semantycznych

⁴² R. Boroach, *Theoretical Proposal of Hybrid Analysis of Lexical Unit: Light*, w: *Proceedings in Advanced Research in Scientific Areas*. The 1st Virtual International Conference. EDIS – Publishing Institution of the University of Zilina, 2012, s. 1258. Zawartość tabeli 3 przeł. R.B.

7. PROCESY RELACYJNE W REPREZENTACJI WIEDZY KR – ZAŁOŻENIA TEORETYCZNE

Procesy relacyjne w reprezentacji wiedzy KR zachodzą na wielu płaszczyznach, tworząc siatkę znaczeniową, na której opiera się znaczenie. Tabela 4 przedstawia taki proces w ujęciu semiotycznym, opracowany przez Görana Sonessona.

Tabela 4. Proces postrzegania przedmiotu w ujęciu semiotycznym⁴³

Postrzeganie	Znaki	System znaków	Skonkretyzowany znak
Epizodyczne	Mimetyczne	Mityczne	Teoretyczne
	Pantomimiczne Gestyczne Imitowanie artefaktów	Językowe	
Proces	Ikonizacja i indeksacja	Symbolizacja	Ikonizacja i indeksacja

Tabela 4 przedstawia proces postrzegania artefaktu i stopniowego przekształcania się wyniku postrzegania z systemu epizodycznego w metateoretyczny (metajęzykowy), w którym następuje konkretyzowanie znaku jako terminu systemu formalnego. Procesy przedstawione w tabeli 4 są jednocześnie, dlatego nie można w tym wypadku mówić o następstwie chronologicznym. W wypadku artefaktów abstrakcyjnych możemy mieć do czynienia od razu z systemem znaków w postaci znaków języka naturalnego lub znaków *L*-języka.

Formułę tabeli 4 sprowadźmy do zapisu:

[percepcja → (znak → system znaków → skonkretyzowany znak)] = znaczenie

Konieczne jest tu określenie podstawy procesów relacyjnych, ponieważ musi istnieć jakiś impuls, który uruchamia całą skomplikowaną operację. Taką podstawą procesów relacyjnych zarówno w systemach konceptualnych, jak i formalnych, może być, na przykład, zdarzenie w rozumieniu Leona Koja.

Logiczne podstawy zdarzenia zostały opracowane przez Koja w pracy *Zdarzeniowa koncepcja znaku*, w której badacz ten podał następującą definicję zdarzenia: „W czasie *t* dla *v* to, że *p*, jest znakiem tego, że *q*, ze względu na okoliczności *s* i cele *r*”⁴⁴.

Koj zapisuje swoją formułę w postaci:

⁴³ G. Sonesson, *Semiosis and the Elusive Final Interpretant of Understanding*, „Semiotica” 179(2010), s. 145-258. Zawartość tabeli 4 przeł. R.B.

⁴⁴ Warszawa 1998, s. 13.

Z (t, v, p, q, s, r)

Tak rozumiane zdarzenie, jak już wspomniałem, może stanowić podstawę procesów relacyjnych w systemach konceptualnych, ponieważ określa okoliczności, zachodzące w procesie tworzenia się relacji, które wyznaczają konkretne konteksty interpretacyjne w postaci zmiennych zdaniowych p, q, s, r. Zmienne te określają interpretanta, który inicjuje proces jednoczesnego indeksowania systemów formalnych i konceptualnych.

Znaczenie nie powstaje tu w wyniku prostej semantycznej operacji odczytania wartości depozytu semantycznego SD jednostek leksykalnych przez konkretnego człowieka, lecz w wyniku „zdarzenia”, które jest dla człowieka kontekstem intensjonalnym, możliwym do opisanie w postaci funkcji f , gdzie argumenty p i q stanowią wartości ekstensjonalne jakiegoś modelu, zaś s (okoliczności) i r (cele) są wartościami intensjonalnymi, gdzie wartością odniesienia jest komunikacyjność zdania. Zatem zdarzenie musi w swojej implementacji prowadzić do zrozumiałego pod względem komunikacyjnym wypowiedzenia, które jest zgodne z intencją mówiącego ze względu na s (okoliczności) i r (cele).

W tym miejscu muszę zaznaczyć, iż argumenty p, q, s i r nie wykluczają się nawzajem ze względu na przynależność formalną czy konceptualną, ale są symultaniczne i w całości składają się na sytuację znakową, czyli zdarzenie. Zdarzenie jest tu czynnikiem (konceptualnym lub formalnym), który rozpoczyna procesy relacyjne w każdej reprezentacji wiedzy KR.

Taki stan rzeczy jest, oczywiście, niemożliwy do osiągnięcia w tłumaczeniu maszynowym, które opiera się na 1. „zamianie” jednostek leksykalnych oraz 2. prostych algorytmicznych regułach gramatycznych, takich, jak nadawanie wartości przypadka (na przykład derywacja końcówek deklinacji rzeczownikowej)⁴⁵.

Zupełnie inaczej sprawa wygląda w wypadku systemu formalnego, który wymaga stworzenia modelu analizy oraz przełożenia języka naturalnego na metajęzyk w celu weryfikacji logicznej. W wyniku tej operacji powstanie funkcja f , której argumentami będzie zbiór modeli (intensjonalność), zaś wartościami odniesienia będą kategorie prawdy i fałszu (ekstensjonalność) w odniesieniu do badanego modelu.

⁴⁵ Zob. Z. Saloni, M. Świdziński, *Składnia współczesnego języka polskiego*, Warszawa 1985; Przepiórkowski, Kupść, Marciniak, Mykowiecka, *Formalny opis języka polskiego...*

Wartość odniesienia: prawda i fałsz w systemach formalnych jest tożsama z komunikacyjnością zdania w systemach konceptualnych.

W systemach formalnych zdarzeniem będzie prawdopodobieństwo wystąpienia danej w konkretnej sekwencji tekstowej, przy czym tekst jest tu rozumiany jako uschematyzowany ciąg znaków⁴⁶. Systemy formalne na mocy stosowanych praw językowych interesują się możliwościami wystąpienia konkretnej danej w pewnej sekwencji, zaś pojawienie się spodziewanej danej jest tu rozumiane jako zdarzenie. Teoretyczne podstawy tak rozumianego zdarzenia znajdują się w prawach, na przykład: 1. łańcuchy Markowa, 2. prawa Zipfa czy 3. prawo Menzeratha⁴⁷.

Problem ten pozostawię w tym artykule bez komentarza, ponieważ nie jest moim zamierzeniem referowanie teorii QL (*Quantitative Linguistics* – lingwistyka kwantytatywna).

Odnieśmy to do problematyki związanej z przekładem intralingwarym. Pewne jest, iż przekładalność jednostek leksykalnych (artefaktów abstrakcyjnych) jest ograniczona, jeżeli w procesie przekładu będziemy brać pod uwagę jedynie pierwotne konteksty interpretacyjne, którymi dla tłumacza są zbiory słowników – leksykony.

Ważną uwagą w tym miejscu będzie to, że słowniki zawsze są opracowywane poza czasem faktycznego postrzegania przedmiotu i mają inną wartość odniesienia w postaci definicji słownikowej, dlatego leksykon jest zawsze „opóźniony” w procesie tworzenia się konceptualnego znaczenia. Przypomnę, że leksykon jest jedynie zbiorem danych, w którym to zbiorze nie zostały wyznaczone żadne relacje semantyczne, lecz jedynie przyporządkowanie, na przykład alfabetyczne, ułatwiające odszukiwanie danych.

Dane należące do konkretnego modelu poznajemy za pomocą logiki takiego modelu, odwołując się do formalnych relacji metajęzykowych w konkretnych reprezentacjach wiedzy KR, takich jak, na przykład, „fizyka” czy „biologia”, które są utrwalone. Tak rozumiana dane staje się argumentem o konkretnej wartości depozytu semantycznego SD. Tabela 5 przedstawia powyższy problem bardziej obrazowo.

⁴⁶ Zob. P. Bouissac, *Circus Performance as Text*, „Poetica” 5(1976), s. 101-118; W. Koch, *Le texte normal, le théâtre et le film* (1973), w: tenże, *Das Textem. Gesammelte Aufsätze zur Semematik des Texts*, Hildesheim–New York 1976, s. 98-126.

⁴⁷ Zob. M. Dillon, *The Quantitative Analysis of Language: Preliminary Considerations*, „Computer Studies In Humanities and Verbal Behavior” 3(1970), s. 191-207; A. Pawłowski, *Metody kwantytatywne w sekwencyjnej analizie tekstu*, Warszawa 2001.

Tabela 5. Słownikowe znaczenia jednostki *light* – część A: rok 1911, część B: rok 2003

Część A	Część B
<i>Oxford Dictionary of Current English</i> ⁴⁸	<i>Cassell's Dictionary of Slang</i> ⁴⁹
1) the natural agent that stimulates the sense of sight 2) medium or condition of space in which sight is possible (opp. darkness) 3) appearance of brightness	1) credit; thus <i>strike a light</i> , to open a line of credit; get a light, to obtain credit 2) insanity, craziness; thus <i>have a light</i> , to be crazy 3) intoxicated, esp. by drugs 4) short of money 5) stupid, weak; <i>light in the head</i> , <i>light in the pocket</i> 6) to enlighten someone with general or specific knowledge

Tabela 5 przedstawia słownikowe znaczenia jednostki *light* w języku angielskim. W części A tabeli 5 znaczenia te opierają się na utrwalonych reprezentacjach wiedzy KR „fizyka”. Część B tabeli 5 jest bardziej skomplikowana. Nie można tu wprowadzić określenia ogólnego, na przykład KR „slang”, ponieważ takie określenie odnosiłoby się do bliżej nieznanych wartości semantycznych. Dlatego przyjmuję, że w wypadku znaczeń części B tabeli 5 należy mówić nie o jednej, ale o wielu reprezentacjach wiedzy KR, które są w trakcie utrwalania. Oznaczam to kolejno jako: KR₁ dla frazy (znaczenia) 1, KR₂ dla frazy 2, KR₃ dla frazy 3 itd.

W części A tabeli 5 znajdują się znaczenia jednostki *light* rejestrowane w roku 1911. Użycie materiału z roku 1911 jest tu celowe, ponieważ pokazuje ówczesny sposób rozumienia pojęcia *light*.

W części B znajdują się znaczenia jednostki *light* (KR₁-KR₆) rejestrowane w roku 2003. W obu częściach A i B widoczne są zasadnicze różnice. Na przykład w części A we frazach 1 i 2 jednostka *light* jest w relacji z jednostką *sight*: R *light*, *sight*. Symbolicznie oznaczmy tę relację RLS. Relacja RLS znajduje implementację w postaci frazy 3 w części A, co jest konceptualnym połączeniem wartości semantycznej depozytów SD jednostek *light* i *sight*. Owo połączenie jest spowodowane tym, że nie można odnosić się do wartości SD *light* bez uwzględnienia wartości depozytu semantycznego SD *sight*. Rok 1911 stanowi jednostronne domknięcie (ograniczenie) przedziału o konkretnych wartościach depozytów semantycznych SD.

⁴⁸ *Oxford Dictionary of Current English*, 1911, red. H.W. Fowler, F.G. Fowler, na podstawie *The Oxford Dictionary* (1900).

⁴⁹ *Cassell's Dictionary of Slang*, red. J. Green, London 2003.

W części B tabeli 5 pojawiają się wartości dodatkowe w postaci fraz (znaczeń) 1-6. Widoczne różnice znaczeń 1-3 w części A i znaczeń 1-6 w części B są wynikiem zakończonych procesów relacyjnych, lecz w innych reprezentacjach wiedzy. Reprezentacja wiedzy KR „fizyka” jest reprezentacją utrwaloną i stabilną semantycznie. Stabilność gwarantuje tu inna reprezentacja wiedzy KR „optyka”, co stanowi wartość odniesienia. Reprezentacja wiedzy KR₁-KR₆ w części B jest reprezentacją konceptualną i odnosi się do konkretnego zdarzenia jako znaku w rozumieniu formuły Kojima. W wyniku konkretnego i jednostkowego zdarzenia komunikacyjnego zostały tu uruchomione procesy relacyjne.

Zakończony proces relacyjny pozostawił widoczną tożsamość jednostek *light* (część A) i *light* (część B), ale tylko w reprezentacji: a) graficznej w lekcji [light] i b) fonetycznej [lait], co jest wyłącznie pozostałością użytego kodu. Argument wskazujący na różnice tych jednostek w sensie definicji jednostki leksykalnej (a nie jednostki systemu Q_S lub Q_D) uznaję za istotny, lecz z konieczności na tym poziomie analizy muszę go pominąć. Uwzględnienie różnicy znaczenia jednostek leksykalnych unieruchomiłoby dalsze rozważania, do chwili rozwiązania problemu z algorytmem tłumaczeniowym. To z kolei zmusiłoby do uznania za słuszny argument, że każdy algorytm tłumaczeniowy już w chwili skonstruowania jest nieaktualny pod każdym względem, ponieważ nie rejestruje zmian kolokacyjnych, które pojawiają się podczas komunikacyjnego używania języka.

Kontrargumentacja wskazująca na to, że w części A i w części B tabeli 5 mamy do czynienia z różnymi jednostkami leksykalnymi, jest słuszna. Należy jednak zauważyć, że w artykule odwracam niejako kolejność: przechodzę od myślenia o obiekcie do implementacji owego myślenia w języku naturalnym.

Zwróćmy także uwagę na fakt, że algorytm tłumaczeniowy nie może rozpoznać jakichkolwiek kontekstów metasemantycznych. Do tej pory nie został opracowany skuteczny sposób maszynowego identyfikowania ani kontekstu z poziomu meta, ani zwykłego kontekstu, który wskazywałby na rodzaj lub charakter tekstu, co niewątpliwie stanowiłoby jakąś wartość odniesienia. Można, oczywiście, podnosić postulaty lingwistyki kwantytatywnej czy lingwistyki modelowej i próbować, w celu wyznaczenia kontekstu, poddawać teksty analizie sekwencyjnej⁵⁰. Metody analizy sekwencyjnej są jednak nieskuteczne, ponieważ analizie musi być poddana odpowiednia pod względem długości próba tekstu, aby można było wskazywać na regularność konkretnych sekwencji. Innym pro-

⁵⁰ Pawłowski, *Metody kwantytatywne...*

blemem jest konieczność indukcyjnego interpretowania wyniku takiej kwantyfikacji.

W moim przekonaniu konieczne jest stworzenie bazy operacyjnej, która pozwalałaby jednoznacznie rozstrzygnąć, że: 1. w kontekście K_1 jednostka J_1 znaczy Z oraz 2. w kontekstach $K_1, K_2 \dots K_N$ jednostka J_1 znaczy $Z, Z_1 \dots Z_N$. Wtedy będzie można stwierdzić, czy jednostka J_1 w kontekstach $K_1, K_2 \dots K_N$ jest jedną jednostką leksykalną, czy też nie oraz czy jednostka J_1 jest może jedynie przejawem ekwiwalentnego (równoważnego) myślenia o konkretnym obiekcie, różna jest natomiast implementacja owego myślenia.

Powiedzmy otwarcie: maszynowe mechanizmy sprawdzające są nieskuteczne, ponieważ algorytm nie jest w stanie właściwie identyfikować ani kontekstów, ani właściwych jednostek. Przykład: tłumacz-człowiek frazę z części B tabeli 5: *light in the pocket* odniesie do znaczenia, które reprezentowane jest przez frazę: *(być) splukany*. Zostanie tu dokonana interpretacja wartości semantycznej depozytu SD *light in the pocket* oraz wyznaczenie ekwiwalentu o tej samej wartości semantycznej SD w języku polskim, o implementacji *splukany*. Jednakże wadliwy algorytm przekładu maszynowego nie jest w stanie skutecznie przeprowadzić tej operacji, czego dowody przedstawiam w tabeli 6. Można by więc zaryzykować stwierdzenie, że ekwiwalencja odnosi się – przynajmniej na obecnym etapie stanu badań – jedynie do tłumacza-człowieka i nie ma żadnych algorytmów, które człowieka mogłyby skutecznie zastąpić.

Tabela 6. Nonsensy semantyczne w przekładzie maszynowym *light in the pocket*.

		Techland₂₀₀₅	
	<i>light in the pocket</i>	świeć w mojej kieszeni	
I traveled freely with a light in my pocket.		Podróżowałem swobodnie z światłem w mojej kieszeni.	
I felt light-headed and optimistic about me travelling freely with a light in my pocket.		Czułem się lekko zmierzany i optymistyczny o mnie podróżowanie swobodnie z światłem w mojej kieszeni.	
		Techland₂₀₀₉	
	<i>light in the pocket</i>	światło w kieszeni	
I traveled freely with a light in my pocket.		Podróżowałem swobodnie ze światłem w mojej kieszeni.	
I felt light-headed and optimistic about me travelling freely with a light in my pocket.		Lekcja [zmierzany] od iść w jakimś kierunku – ang. headed.	Czułem lekko zmierzany i optymistyczny o mnie podróżowanie swobodnie ze światłem w mojej kieszeni.

	Kompas₂₀₀₃
<i>light in the pocket</i>	lekki w kieszeni
I traveled freely with a light in my pocket.	Podróżowałem swobodnie ze światłem w mojej kieszeni.
I felt light-headed and optimistic about me travelling freely with a light in my pocket.	Poczułem się bezmyślny i optymistyczny o mnie podróżującym swobodnie ze światłem w mojej kieszeni.
	Kompas₂₀₁₀
<i>light in the pocket</i>	lekki w kieszeni
I traveled freely with a light in my pocket.	Podróżowałem swobodnie ze światłem w mojej kieszeni.
I felt light-headed and optimistic about me travelling freely with a light in my pocket.	Poczułem się bezmyślny i optymistyczny o mnie podróżującym swobodnie ze światłem w mojej kieszeni.

Tabela 6 przedstawia nonsensy semantyczne w przekładzie maszynowym. Jak widać, algorytm tłumaczeniowy nie rozpoznaje ani części zdania, ani części mowy.

Ponadto mimo deklarowanej przez twórców algorytmu analizy semantycznej nie jest możliwe dobranie właściwych znaczeń, czego przykładem jest użycie lekcji [zmierzany] w znaczeniu „iść w jakimś kierunku” w złożeniu w lekcji [light-headed] albo tłumaczenie sformułowania w lekcji [light-headed] jako bezmyślny.

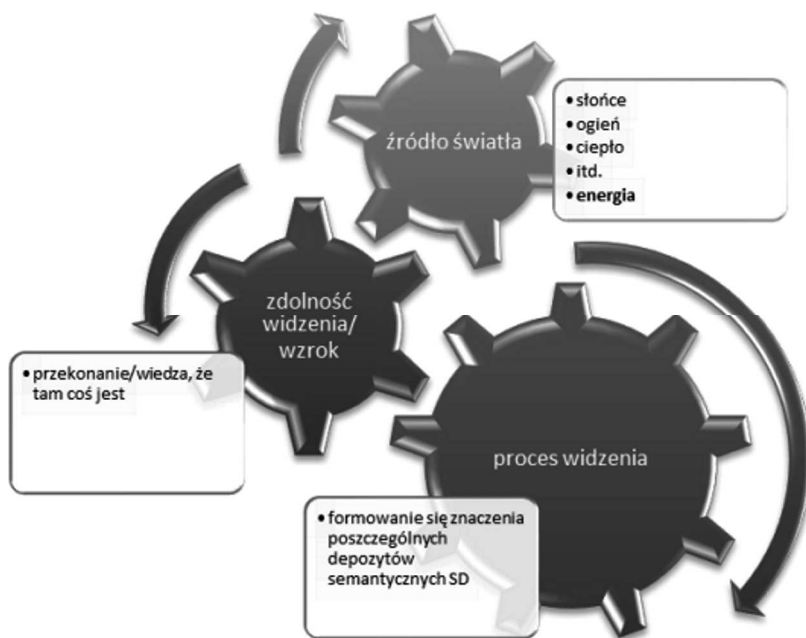
Po dokonaniu przekładu na język docelowy, w tym wypadku na język polski, można podnieść argument odnośnie do tożsamości jednostek leksykalnych, ale jedynie w ramach tego samego języka, czyli w tym wypadku języka polskiego. Istotne mogą tu być pytania dotyczące zakresu jednostek, na przykład, czy jednostka w lekcji [bezmyślny] jest zakresowo tożsama z jednostką w lekcji [lekkomyślny]. Jednostka w lekcji [lekkomyślny] znaczy tyle, co *niezdający sobie sprawy z konsekwencji podejmowanych działań*, zaś jednostka w lekcji [bezmyślny] znaczy tyle, co *niezastanawiający się nad konsekwencjami swoich działań, jednakże owe konsekwencje (prawdopodobnie) znajdujący*.

Część A i część B tabeli 5 pokazują różne reprezentacje wiedzy KR, w których powstały znaczenia jednostki *light*. Reprezentacje te można (i należy) poszerzać na przykład o: KR „source of light” (źródło światła), KR „natural light” (naturalne światło), KR „colours” (barwy/kolory), KR „process of seeing” (proces widzenia), KR „weight” (waga) itd.⁵¹ Wielość reprezentacji wiedzy KR nie jest jednak problemem tak długo, dopóki możliwe jest wyznaczenie przyporządkowania oraz utrzymanie spójności reprezentacji wiedzy KR na poziomie meta.

⁵¹ Zob. *Oxford Advanced Learner's Dictionary*.

Zwróćmy także uwagę, że w części B tabeli 5 jednostka *light* reprezentuje depozyt semantyczny SD, który jest ekstensjonalny i intensjonalny. Ekstensjonalność tego depozytu semantycznego SD przejawia się w tym, iż możliwe jest rozszerzenie wartości semantycznych o inne, pozornie sprzeczne ze znaczeniami z części A, wartości. Jego intensjonalność przejawia się w tym, iż procesy relacyjne odnoszą się do procesów/sposobów myślenia o cechach dystynktywnych w siatce znaczeń jednostki *light*. Wartościami odniesienia będą tu: 1. prawda i fałsz (cechy systemu formalnego) w metajęzykowej reprezentacji wiedzy KR, na przykład KR „fizyka”, oraz 2. komunikacyjność (cecha systemu konceptualnego) w jakiegokolwiek reprezentacji wiedzy KR, która pozwala na zakończenie procesów relacyjnych i implementowanie znaczeń z części B tabeli 5.

Komplementarność systemów formalnych i konceptualnych przedstawiają rysunek 2 i rysunek 3, gdzie pokazana jest „współpraca” różnych reprezentacji wiedzy KR. Więcej na powyższy temat piszę w artykule *Theoretical Proposal of Hybrid Analysis of Lexical Unit: Light*⁵².



Rysunek 2. Konceptualizacja jednostki *light* – model I⁵³

⁵² B o r o c h, *Theoretical Proposal...*

⁵³ Tamże, s. 1259. Przeł. R.B.

Powyższy rysunek przedstawia przykładowy proces konceptualizacji (myślenia o obiekcie) jednostki *światło*, który jest wynikiem „współpracy” takich elementów jak: 1. *źródło światła*, 2. *widzenie* i 3. *przekonanie człowieka, że coś tam jest*. Innymi słowy, mamy tu do czynienia z konceptualizacją jednostki *light*, która to konceptualizacja jest amalgamatem elementów 1-3⁵⁴.

Sekcja (ang. *division*): źródło światła może (teoretycznie) stanowić wartość odniesienia dla reprezentacji wiedzy KR „fizyka”, co automatycznie kieruje nasze myślenie w stronę poziomu meta. Mówiąc prościej: Newtonowska teoria światła (*Newton's Theory of Light*) nie mogła być sformułowana bez odniesienia do falowej teorii światła Kartezjusza (*Descartes' Wave Theory of Light*)⁵⁵, zaś teoria Kartezjusza – bez sądów o świetle sformułowanych przez Arystotelesa, Keplera czy Galileusza, co z kolei zaowocowało pracami Olausu Romera (1664-1710) odnośnie do prędkości światła⁵⁶. Reprezentacja wiedzy KR „fizyka” i KR „źródło światła” jest modelem formalnym (Q_S), zaś reprezentacja wiedzy KR „proces widzenia” i KR „zdolność widzenia” jest modelem konceptualnym (Q_D).

Model I (rysunek 2) stanowi proponowaną wartość odniesienia dla jednostki *light* i może być uznany za przykład metaoperacji semantycznych, które wynikają z procesu postrzegania przedmiotu i przypisywania mu pewnych cech znaczących⁵⁷.

Procesy meta zachodzą pomiędzy jednostkami określonymi przeze mnie jako hybrydowe, o dominancie konceptu lub terminu. Na rysunku 3 takie jednostki zostały oznaczone symbolem H (od *hybrydowy*) z sygnaturą dominandy terminu (H_T) lub konceptu (H_C).

Procesy relacyjne, które są widoczne na rysunku 3 jako implementacja konceptualna jednostki *światło*, prowadzą do wykształcenia terminu *światło* T₃, dla którego najistotniejsze są depozyty semantyczne SD oznaczone jako H_{T2} oraz H_{C4}. Formalnie zapisują to następująco:

$$f: T_3 \rightarrow (H_{T2}, H_{C4}) \text{ wtedy i tylko wtedy, gdy } (H_{T2}, H_{C4}) \in T_3$$

⁵⁴ Tamże.

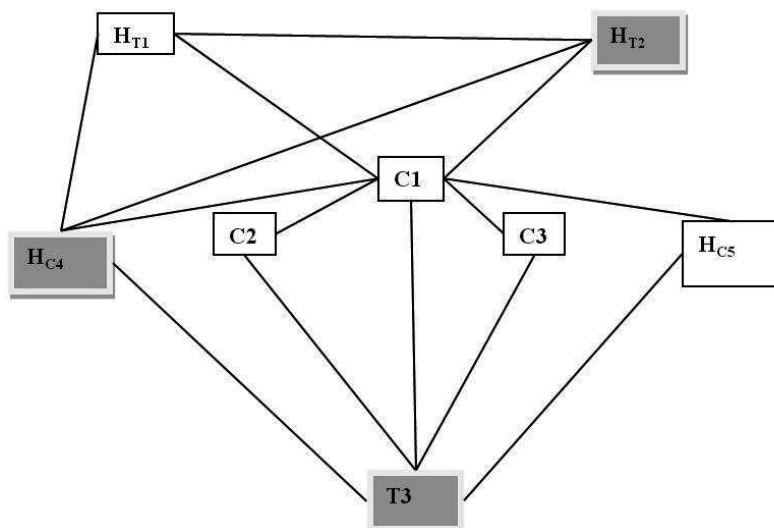
⁵⁵ A.I. S a b r a, *Theories of Light. From Descartes to Newton*, New York 1981.

⁵⁶ P.T. O o n, R. S u b r a m a n i a m, *The Nature of Light: I. A Historical Survey up to the Pre-Planck Era and Implications for Teaching*, „Physics Education” 44(2009), nr 4, s. 384-391.

⁵⁷ Człowiek, postrzegając określony przedmiot, formułuje na jego temat pewne sądy, używając terminów jako jednostek zrozumiałych i koherentnych definicyjnie. Jednakże do semantycznych właściwości terminów w procesie łączenia dodane są inne wartości semantyczne, jako konsekwencja możliwych procesów relacyjnych w konkretnej reprezentacji wiedzy KR z innymi depozytami semantycznymi SD, co uznają za operację z poziomu meta.

Zakładam, iż f jest funkcją, której dziedziną jest ściśle określona reprezentacja wiedzy KR, w której argumenty funkcji f przyjmują określone wartości.

Rysunek 3 przedstawia implementację konceptu *światło* w postaci siatki.



Rysunek 3. Implementacja konceptu *światło*

H_{T1} – jednostka hybrydowa: termin *słońce*: ciepło \vdash światło \vdash widzieć \vdash energia

H_{T2} – jednostka hybrydowa: termin *ogień*: słońce \vdash ciepło \vdash światło \vdash widzieć \vdash energia

T_3 – termin *światło*

C_1 – koncept postrzegania *światła*; formowanie się terminu *światło*

C_2 – koncept *widzieć/być przekonany*

C_3 – koncept *wiedzieć/być przekonany, że coś jest*

H_{C4} – jednostka hybrydowa: koncept *widzieć artefakt przy użyciu zmysłu*; doświadczenie empiryczne

H_{C5} – jednostka hybrydowa: koncept *wiedzieć, że artefakt jest*; dedukcja i indukcja

\vdash – kierunek komplikacji semantycznej

Semantyczne relacje między poszczególnymi elementami wyglądają następująco:

$f: [(H_{T1}, H_{T2}) (H_{C4}, H_{C5})] \rightarrow T_3$ gdzie $(C_1, C_2, C_3) \notin f: [(H_{T1}, H_{T2}) (H_{C4}, H_{C5})]$

„Obszar roboczy” Q_S (w tym wypadku) to wartość depozytu semantycznego SD konceptów (C_1, C_2, C_3) . Koncepty te są pierwotne w procesie relacyjnym reprezentacji wiedzy KR „światło” i stanowią podstawę przekształceń semantycznych w innych reprezentacjach wiedzy KR, jeżeli możliwe jest ustanowienie takich relacji.

Jednostka *światło* występująca w części definicji deskryptywnych innych jednostek jest pozostałością zachodzących relacji w postaci reprezentacji graficznej, która jest wspólna. Tu jednak znaczenie jest już inne: *light* = *dietetyczny* → *light coke* = *dietetyczna cola*.

Przebieg oraz zakończenie procesów relacyjnych przedstawia tabela 7.

Tabela 7. Procesy relacyjne w reprezentacji wiedzy „światło” – kształtowanie się znaczenia

Krok 1: Koncept	Krok 2: Procesy relacyjne	Krok 3: Termin
<ul style="list-style-type: none"> • percepcja przedmiotu • zdolność widzenia/ przekonanie, że tam/gdzieś coś jest <p style="text-align: center;">↓</p> <p style="text-align: center;">koncept <i>światło</i></p>	<ul style="list-style-type: none"> • znaki • system znaków • znak skonkretyzowany 	<ul style="list-style-type: none"> • <i>światło</i>: reprezentacja fonetyczna • <i>światło</i>: reprezentacja graficzna <p style="text-align: center;">↓</p> <p style="text-align: center;">termin <i>światło</i></p>

8. PRZEDMIOT PRZEKŁADU FCA – UWAGI

Zakładam, iż przedmiotem przekładu są synonimiczne relacje depozytów semantycznych SD w języku A oraz ich odpowiedniki w języku B w konkretnych reprezentacjach wiedzy KR. Celem przekładu nie jest określenie denotacji terminów, lecz odszukanie w reprezentacji wiedzy KR w języku docelowym relacji, które są synonimicznym odpowiednikiem relacji w reprezentacji wiedzy KR w języku wyjściowym.

W tłumaczeniu maszynowym proces ten (teoretycznie) jest algorytmiczny, jednakże ów algorytm nie przewiduje, by konkretne reprezentacje wiedzy KR wchodziły między sobą w jakiegokolwiek złożone relacje semantyczne, które analizuje FCA, używając jednocześnie formalnego i konceptualnego modelu opisu, co pozwala na implementację hybrydowego znaczenia konkretnej jednostki wchodzącej w skład obiektu.

Być może wykorzystanie metod kwantytatywnych w sekwencyjnych analizach tekstu stworzy taką możliwość. Jest to nadal wątpliwe, mimo iż Thomas Bernhard Seiler⁵⁸ rozwinął kognitywne podstawy formalnej logiki mechanizmów konstytuujących koncepty, a także zwrócił uwagę na modus odczytania ich depozytów semantycznych SD w tzw. inteligentnym zachowaniu.

⁵⁸ W pracach: *Concept Development and the Development of Word Meaning*, Berlin, New York 1983 oraz *Kognitive Strukturiertheit. Theorien. Analysen. Befunde*, Stuttgart–Berlin–Köln–Mainz 1973.

Aktualne pozostają jednak problemy, o których pisałem wcześniej – konieczność analizy próby tekstowej o odpowiedniej długości oraz brak opierania się w metodach kwantytatywnych na wartościach semantycznych jednostek.

9. PODSUMOWANIE

Metoda FCA umożliwia stworzenie takiego leksykonu danych językowych, który pozwala na automatyczne łączenie się danych poprzez konkretną relację w reprezentacji wiedzy KR bez potrzeby odwoływania się do tradycyjnych definicji słownikowych czy encyklopedycznych, które to definicje mają formalną strukturę utrwaloną w metajęzyku⁵⁹. Owa struktura jest implementacją myślenia o obiekcie: *światło*.

Potencjał, który ma FCA w zarządzaniu informacją, powinien być wspomagany nie przez semantyczną analizę składnikową poszczególnych depozytów semantycznych SD, lecz przez procesy relacyjne, które owe depozyty semantyczne SD łączą, ponieważ to właśnie w tych relacjach powstaje znaczenie, które swoim zakresem obejmuje nie jedną jednostkę leksykalną, lecz cały zespół jednostek, które tworzą obiekty i atrybuty obiektów.

Procesy relacyjne mają charakter pozajęzykowy oraz czasowy, jednakże są one na tyle trwałe, że możliwe jest ich analizowanie w celu adaptacji przez SI (sztuczna inteligencja) do symulacji doświadczenia empirycznego, co może ulepszyć przekład maszynowy. Wymaga to jednak zmiany algorytmu tłumaczącego, na co zwróciłem już uwagę przy okazji tłumaczenia dwukierunkowego translatorów Techland i Kompas.

Uważam, iż w tym celu konieczne jest: 1. stworzenie bazy leksykonów w postaci modeli matematycznych, wykorzystując lingwistykę kwantytatywną, która „definiuje język jako wielowarstwową i wielowymiarową strukturę, złożoną z dyskretnych jednostek połączonych ze sobą siecią relacji”⁶⁰, 2. zarządzanie relacjami depozytów semantycznych SD danych w konkretnej reprezentacji wiedzy KR, wykorzystując w tym celu metodologię FCA oraz 3. konstruowanie hybrydowych definicji depozytów semantycznych SD jednostek opartych na systemach formalnych i konceptualnych, które uwzględniają znaczenie owych relacji w konkretnej reprezentacji wiedzy KR.

⁵⁹ Zob. P r i s s, *Linguistic Applications...*

⁶⁰ P a w ł o w s k i, *Metody kwantytatywne...*, s. 6.

BIBLIOGRAFIA

- Baayen R.H.: *Analyzing Linguistic Data. A Practical Introduction to Statistics Using*, Cambridge 2008.
- Baker P.: *Sociolinguistics and Corpus Linguistics*, Edinburgh 2010.
- Bertalanffy L.: *Ogólna teoria systemów. Podstawy, rozwój, zastosowania*, przeł. E. Woydyłło-Woźniak, Warszawa 1984.
- Boroch R.: Formalna Analiza Konceptualna (FCA) w badaniach kulturoznawczych – model analizy obiektu N: prowincjonalizm, w: *Prowincjonalizm w kulturze europejskiej*, red. B. Płonka-Syroka, K. Marchel, Wrocław 2010, s. 21-46.
- Boroch R.: Formalna Analiza Konceptualna (FCA) w badaniach nad postaciami dramatycznymi – jednostka N: sługa na przykładzie Fieraponta w *Trzech siostrach* i Firsy w *Wiśniowym sadzie* Antoniego Czechowa, w: *Czechow współcześnie*, red. R. Strzelecki, G. Guźlak, Bydgoszcz 2010, s. 125-152.
- Boroch R.: Theoretical Proposal of Hybrid Analysis of Lexical Unit: Light, w: *Proceedings in Advanced Research in Scientific Areas. The 1st Virtual International Conference. EDIS – Publishing Institution of the University of Zilina 2012*, s. 1257-1262.
- Boroch R.: Na marginesie semiotyki kultury: problemy tzw. nauk kulturoznawczych, w: *Znaki czy nie znaki*, red. M. Guławska-Gawkowska, G. Zeldowicz, Warszawa 2013, s. 21-39.
- Bouissac P.: Circus Performance as Text, „*Poetica*” 5(1976), s. 101-118.
- Brezina V.: Use of Google Scholar in corpus-driven EAP research, „*Journal of English for Academic Purposes*” 11(2012), s. 319-331.
- Cassell's Dictionary of Slang, red. J. Green, London 2003.
- Cormen T.H., Leiserson Ch.E., Rivest R.L., Stein C.: *Wprowadzenie do algorytmów*, przeł. K. Diks, A. Malinowski, D. Roszkowska, D. Rytter, W. Rytter, Warszawa 2012.
- Corpus Linguistics. An International Handbook, t. I, red. A. Lüdeling, M. Kytö, Berlin–New York 2008.
- Corpus Linguistics. An International Handbook, t. II, red. A. Lüdeling, M. Kytö, Berlin–New York 2009.
- Corpus Linguistics Around the World, red. A. Wilson, D. Archer, P. Rayson, Amsterdam–New York 2006.
- Corpus Linguistics Beyond the World. Corpus Research from Phrase to Discourse, red. E. Fitzpatrick, Amsterdam, New York 2007.
- Dash N.S.: *Corpus Linguistics. An Introduction*, New Delhi 2008.
- Dillon M.: The Quantitative Analysis of Language: Preliminary Considerations, „*Computer Studies In Humanities and Verbal Behavior*” 3(1970), s. 191-207.
- Dobaczewski A.: Światło. Próba analizy semantycznej, w: *Beiträge der Europäischen Slavistischen Linguistik (POLYSLAV)*, red. K. Böttger, M. Giger, B. Wiemer, Bd. 2, München 1999, s. 93-99.
- Dobaczewski A.: Próba opisu znaczeń leksemów jasny/jasno/jasne, „*Prace Językoznawcze UWM*” 1(1999), s. 17-26.
- Dobaczewski A.: O czasownikach świecić i świecić się, „*Acta Universitatis Nicolai Copernici. Filologia Polska*” 55(2001), s. 53-60.
- Dobaczewski A.: Czy widzieć jest semantycznie proste? w: *Studies on the Syntax and Semantics of Slavonic Languages. Papers in Honour of Andrzej Bogusławski on the Occasion of his 70th Birthday*, red. V. S. Chrakovskij, M. Grochowski, G. Hentschel, Oldenburg 2001, s. 121-132.

- Dyszak A.: Językowy obraz światła w Księdze Rodzaju a współczesne znaczenie rzeczownika „światło”, w: *Język. Teoria. Dydaktyka. Materiały 21 konferencji językoznawczej zorganizowanej w Trzcinicy k. Jasła w dniach 27-29 maja 1998 roku*, red. B. Greszczuk, Rzeszów 1999, s. 197-205.
- Dyszak A.: Językowe wyrażenia zjawisk emisji światła, Bydgoszcz 1999.
- Eco U.: *A Theory of Semiotics*, Indiana 1976.
- Gabrielatos C., McEnery T., Diggie P.J., Baker P.: The Peaks and Troughs of Corpus-Based Contextual Analysis, „*International Journal of Corpus Linguistics*” 17(2012), nr 2, s. 151-175.
- Garbacz P.: *Logika i artefakty*, Lublin 2006.
- Gries S.T.: *Quantitative Corpus Linguistics with R: A Practical Introduction*, New York – London 2009.
- Grucza S.: O konieczności tworzenia korpusów tekstów specjalistycznych, w: *W kręgu teorii i praktyki lingwistycznej*, red. S. Grucza i in., Warszawa 2007, s. 103-122.
- Gumański L.: *Wprowadzenie w logikę współczesną*, Warszawa 1990.
- Hoffman S.: *Grammaticalization and English Complex Prepositions. A Corpus-Based Study*, London–New York 2005.
- Hunston S.: *Corpora in Applied Linguistics*, Cambridge 2002.
- Koch W.: *Le texte normal, le théâtre et le film (1973)*, w: tenże, *Das Textem. Gesammelte Aufsätze zur Semematik des Texts*, Hildesheim–New York 1976, s. 98-126.
- Koj L.: *Zdarzeniowa koncepcja znaku*, Warszawa 1998.
- Langer S.K.: On Cassirer’s Theory of Language and Myth, w: *The Philosophy of Ernst Cassirer. The Library of Living Philosophers*, red. P. A. Schilpp, Chicago 1949, s. 381-400.
- Łukasik M.: Narzędzia lingwistyki korpusowej w warsztacie terminologa, terminografa i tłumacza tekstów specjalistycznych (cz. I), w: *Debiuty naukowe. Wiedza – korpus – słownik*, red. M. Łukasik, Warszawa 2007, s. 23-47.
- McEnery T., Wilson A.: *Corpus Linguistics*, Edinburgh 1996.
- McEnery T., Wilson A.: *Corpus Linguistics. An Introduction*, Edinburgh 2001.
- Newman J.: Corpus Linguistics and Variation in English: Theory and Description, „*World English*” 3(2012), nr 4, s. 557-559.
- Oon P.T., Subramaniam R.: The Nature of Light: I. A Historical Survey up to the Pre-Planck Era and Implications for Teaching, „*Physics Education*” 44(2009), nr 4, s. 384-391.
- Oxford Advanced Learner’s Dictionary*, red. J. Turnbull, Oxford 2010.
- Oxford Dictionary of Current English*, 1911, red. H.W. Fowler, F.G. Fowler, na podstawie *The Oxford Dictionary (1900)*.
- Pańkowski A.: *Metody kwantytatywne w sekwencyjnej analizie tekstu*, Warszawa 2001.
- Pieńkos J.: *Podstawy przekładoznawstwa. Od teorii do praktyki*, Kraków 2003.
- Priss U.: Formal Concept Analysis as a Tool for Linguistic Data Exploration, w: *Conceptual Structures in Practice*, red. P. Hitzler, H. Scharfe, Chapman & Hall/CRC Studies in Informatics Series 2009, s. 177-198.
- Priss U.: Formal Concept Analysis in Information Science, w: *Annual Review of Information Science and Technology*, red. B. Cronin, 40(2006), s. 521-543 (www.upriss.org.uk/top/research.html).
- Priss U.: Linguistic Applications of Formal Concept Analysis, w: *Formal Concept Analysis. Foundations and Applications*, red. B. Ganter, G. Stumme, R. Wille, Berlin–Heidelberg 2005, s. 149-160.
- Przepiórkowski A., Kupś A., Marciniak M., Mykowiecka A.: *Formalny opis języka polskiego. Teoria i implementacja*, Warszawa 2002.

- Sabra A.I.: *Theories of Light. From Descartes to Newton*, New York 1981.
- Saloni Z., Świdziński M.: *Składnia współczesnego języka polskiego*, Warszawa 1985.
- Seiler T.B.: *Concept Development and the Development of Word Meaning*, Berlin–New York 1983.
- Seiler T.B.: *Kognitive Strukturiertheit. Theorien. Analysen. Befunde*, Stuttgart–Berlin–Köln–Mainz 1973.
- Sonesson G.: *Semiosis and the Elusive Final Interpretant of Understanding*, „*Semiotica*” 179(2010), s. 145-258.
- Sowa J.F.: *Knowledge Representation: Logical, Philosophical, and Computational Foundations*, Brooks Cole Publishing Co. Pacific Grove 2000.
- The Dynamics of Linguistic Variation. Corpus Evidence on English Past and Present*, red. T. Nevalainen, I. Taavitsainen, P. Pahta, M. Korhonen, Amsterdam–Philadelphia 2008.
- The Routledge Handbook of Corpus Linguistics*, red. A. O’Keeffe, M. McCarthy, London – New York 2010.
- Webster Dictionary, <http://www.merriam-webster.com/dictionary/>
- Wille R.: *Restructuring Lattice Theory: an Approach Based on Hierarchies of Concepts*, w: *Ordered Sets*, red. I. Rival, Dordrecht–Boston 1982, s. 445-470.
- Wille R.: *Formal Concept Analysis as Mathematical Theory of Concepts and Concept Hierarchies*, w: *Formal Concept Analysis. Foundations and Applications*, red. B. Ganter, G. Stumme, R. Wille, Berlin–Heidelberg 2005, s. 1-33.
- Wilson P.: *Hieroglyphs. A Very Short Introduction*. Oxford 2003.

FORMAL CONCEPT ANALYSIS
– KNOWLEDGE REPRESENTATION – TRANSLATIONS

Summary

The article presents the possibilities of using Formal Concept Analysis, FCA, in intralingual and machine translations. Let me draw your attention to the necessity of joining formal and conceptual systems into one system based on relations of the particular semantic deposits SD in the specific knowledge representation KR with defining linguistic units. I called this defining a *hybrid defining*. The hybrid defining of linguistic units enables to examine their cognitive (hybrid) meaning, which is implemented in a form of a coherent semantic map of meanings. An exemplifying material, by means of which a process of meaning creation is shown, is a *light unit*.

Słowa kluczowe: Formalna Analiza Konceptualna, depozyt semantyczny, reprezentacja wiedzy, definiowanie, znaczenie, przykład.

Key words: Formal Concept Analysis, semantic deposit, knowledge representation, defining, meaning, translations.