

MAGDALENA ZAWISŁAWSKA

## O PROBLEMACH Z KOREFERENCJĄ

### 1. WSTĘP

Przedmiotem niniejszego artykułu jest omówienie źródeł problemów, które występują podczas określania koreferencji w tekście. Analizy takie były prowadzone na potrzeby projektu *CORE – Komputerowe metody identyfikacji nawiązań w tekstach polskich*<sup>1</sup>, którego celem było stworzenie nowatorskich metod i narzędzi informatycznych służących do automatycznego wykrywania koreferencji dla języka polskiego. Jak się okazało, wyznaczenie fraz koreferencyjnych, nie było sprawą prostą nawet dla człowieka, ponieważ po pierwsze, wymagało to od odbiorcy stwierdzenia, czy dane wyrażenie odnosi się rzeczywiście do tego samego obiektu, po drugie – czy faktycznie mamy tu do czynienia z identycznością, a nie np. tylko podobieństwem. Dodatkowym utrudnieniem był fakt, że koreferencja nie znalazła się dotychczas w centrum zainteresowania językoznawców. Zwykle wzmiankuje się o niej przy okazji omawiania spójności semantycznej dyskursu (koherencji). Koreferencja jest ogólnie definiowana jako zjawisko polegające na odnoszeniu się różnych wyrażeń do tego samego referenta, co oznacza, że nie da mówić się o tej kwestii bez odwołania się do pojęcia referencji. Pozostałe, istotne dla koreferencji pojęcia, o których piszę poniżej, to anafora oraz indentyczność.

### 2. REFERENCJA A KOREFERENCJA

Referencja jest przedmiotem zainteresowania przede wszystkim semantyki logicznej, ale także językoznawstwa i w zależności od dziedziny jest różnie rozumiana. W klasycznej semantyce logicznej przyjęło się za Fregem od-

---

<sup>1</sup> Grant kierowany przez dra Macieja Ogrodniczuka, sfinansowany przez Narodowe Centrum Nauki (umowa nr 6505/B/T02/2011/40).

różnianie dwóch aspektów znaczenia jednostek języka naturalnego. Frege określił je jako *Bedeutung* oraz *Sinn*, w pracach innych badaczy są one nazywane odpowiednio: u Milla – *denotation* i *connotation*, u Russela – *denotation* i *meaning*, u Carnapa – *extension* i *intension*, u Blacka – *reference* i *sens*. W tym ujęciu referencja to odniesienie do bytów pozajęzykowych (referentów), natomiast sens to odniesienie wewnątrzjęzykowe – do innych znaków systemu językowego. Według E. Padučевой (1992: 12) takie rozumienie jest jednak błędne. Zwraca ona uwagę, że referencja nie jest elementem znaczenia wyrazów, ale realizuje się w wypowiedzi – jest zatem nie własnością leksemów, ale wyłącznie użyć reprezentujących ich formy w tekście, por.:

Referencja to, ogólnie mówiąc, odniesienie indywidualnych i za każdym razem nowych obiektów i sytuacji. Dlatego referencja dotyczy nie wyrazów i wyrażeń języka, lecz tylko ich użyć w tekście – dotyczy wypowiedzi i jej składników. [...] Nonsensem byłoby więc pytanie „Jaka jest referencja wyrazu *człowiek?*”, jak też każdego innego wyrazu w słowniku czy połączenia wyrazowego (np. *młody człowiek*, *człowiek wysokiego wzrostu*), czy zdania zbudowanego zgodnie z regułami gramatyki. Zdanie *Wszedł człowiek* samo w sobie nie ma referencji – nie jest odniesione do żadnej konkretnej sytuacji, do żadnego obiektu (Padučeva 1992: 12).

Dlatego też, jak zauważa badaczka, każdy syntaktycznie związanym składnikiem zdania ma znacznie, jednak referencja przysługuje tylko niektórym składnikom wypowiedzi. Należy zatem odróżnić referencję jako właściwość aktu mowy (wypowiedzi) od denotacji (ekstensji) leksemu, która byłaby zbiorem wszystkich potencjalnych referentów. Wynika z tego, że wyraz słownikowy (leksem) nie ma referencji, ma jedynie denotację – potencję referencji, która może (lecz nie musi) zrealizować się w tekście.

Z denotacją leksemów wiąże się zjawisko, które Padučeva określa jako *niejednoznaczność denotacyjna*. Polega to na tym, że jeden wyraz może mieć warianty znaczeniowe i w zależności od kontekstu może realizować inny typ referencji, np.:

- (1) Przeczytał całego Hemingwaya.
- (2) Hemingway urodził się w 1899 r.
- (3) Mieszkał w Berlinie.
- (4) Berlin naciska na bankructwo Grecji.
- (5) Zbiła nowy talerz.
- (6) Zupa była tak dobra, że zjadł cały talerz.

W przykładzie (1) mowa o twórczości pisarza, natomiast w (2) o samym pisarzu. Przykład (3) odnosi się do miasta, (4) zaś – do rządu niemieckiego, który ma siedzibę w Berlinie. W zdaniu (5) mowa o naczyniu, a w (6) – o jego zawartości. Wariantywność znaczeniowa tego typu jest bardzo systemowa i często odnotowywana w słownikach, jednak bardzo utrudnia anotację wyrażeń koreferencyjnych w korpusie.

Przesunięcie referencji na płaszczyznę wypowiedzi i traktowanie jej rozłącznie ze znaczeniem leksemu jest zasadniczą kwestią przy budowaniu korpusu wyrażeń koreferencyjnych. Nie jest bowiem tak, że każdą rekurencję wyrazu w tekście można potraktować jako koreferencję, np.:

(7) Każdy szanujący się poseł ma **asystenta**. **Asystentami** są z reguły ludzie młodzi, ale nie brakuje również szczerze zaangażowanych emerytów. Poglądy polityczne **asystenta** powinny być zbieżne z linią szefa. Pracują jako wolontariusze tak jak Marek Hajbos, **asystent** Zyty Gilowskiej. Poseł Adam Bielan (rzecznik PiS) na przykład płaci **asystentom** za wysyłanie korespondencji. Obecny minister sprawiedliwości Grzegorz Kurczuk zaczynał partyjną działalność jako **asystent** Izabelli Sierakowskiej. W ministrowie poszedł też były **asystent** Józefa Oleksego Lech Nikolski. Posłowie nie poprzestają na jednym **asystencie**.

W tekście (7) nadawca osiem razy użył wyrazu *asystent*. Jednak nie ma on tej samej referencji. Mamy tutaj odniesienie do klasy (*Każdy szanujący się poseł ma asystenta; Poglądy polityczne asystenta powinny być zbieżne z linią szefa; Posłowie nie poprzestają na jednym asystencie*), do części tej klasy (*Asystentami są z reguły ludzie młodzi, ale nie brakuje również szczerze zaangażowanych emerytów; asystenci Adama Bielana*), do konkretnych ludzi (*Marek Hajbos, asystent Zyty Gilowskiej, asystent Józefa Oleksego Lech Nikolski*). Włączenie tych wszystkich wyrażeń do jednego klastra koreferencyjnego byłoby więc błędem.

Kolejna kwestia, niezwykle istotna przy badaniu koreferencji, dotyczy pytania, jakim składnikom wypowiedzi przysługuje referencja. W ujęciu klasycznym referencję mogą mieć wyłącznie frazy nominalne, które nazywają obiekty. Tak też widzi to Z. Topolińska (1984: 302), por.:

Jako nazwom przedmiotów materialnych grupom imiennym spotykanym przez nas w tekstach przysługuje określona charakterystyka referencyjna, tj. odniesienie do przedmiotu, który nazywają.

Nie jest do końca jasne stanowisko Padučovej, która w jednym miejscu swojej pracy pisze (1992: 113), że referencja nie przysługuje predykatom (i NP użytym w funkcji predykatywnej), ponieważ nie nazywają obiektów, tylko ich właściwości, w innym zaś mówi o dwóch typach referencji: przedmiotowej i nieprzedmiotowej, por.:

W tradycji logicznej idącej od Fregego przyjęte jest jednolite traktowanie referencji termów przedmiotowych [...] i propozycji (ang. *proposition*). Propozycje – podobnie jak termy – mają referenta (dla propozycji referentem jest jej prawdziwość, znaczenie prawdziwościowe). Natomiast w pracach o profilu pragmatycznym [...] referencja jest rozumiana przede wszystkim jako własność termów przedmiotowych. Dla lingwistyki bardziej naturalne jest jednakowe traktowanie referencji termów przedmiotowych i propozycji: przypisanie wypowiedzi językowej do rzeczywistości dokonuje się nie tylko poprzez referencję termów przedmiotowych, lecz również poprzez referencję kom-

ponentów o znaczeniu propozycjonalnym, które odnoszą się do faktów, zdarzeń, sytuacji (Padučeva 1992: 15).

Według Padučovej referencja jest właściwością całego zdania użytego w wypowiedzi, składników propozycjonalnych, grup imiennych. Środkami referencji nieprzedmiotowej (czyli odniesienia do wydarzeń, sytuacji) są w rozumieniu autorki nie wyrazy, ale kategorie gramatyczne: czas i tryb.

Odmienne rozstrzygnięcie tej kwestii znajdujemy u H. Vatera (2009: 104), który wyróżnia cztery typy referencji:

- 1) referencję sytuacyjną, która jest nadrzędnym typem referencji, ponieważ zdania odnoszą się do jakiego wydarzenia, stanu, akcji;
- 2) referencję czasową, która dotyczy relacji czasowych między sytuacjami;
- 3) referencję miejsca, która może odnosić się do usytuowania jakiegoś przedmiotu w przestrzeni lub kierunku ruchu;
- 4) referencję przedmiotową.

Jak widać, w zależności od tego, jaką definicję referencji się przyjmie, wpłynie to zasadniczo na wyznaczanie łańcuchów koreferencyjnych w tekście. Z reguły w pracach z zakresu NLP referencja jest rozumiana klasycznie – wyłącznie jako referencja przedmiotowa. Wydaje się jednak, że jest to dobre rozwiązanie jedynie na początek prac nad zagadnieniem koreferencji w tekście. Aby w pełni opisać to zjawisko, należałoby uwzględnić wszystkie typy referencji wymienione przez Vatera, np.:

- (8) Zapadał mrok. Bardzo ich to zaniepokoiło.
- (9) Mieszkał w samym centrum. Było tam dość głośno.
- (10) Spotkali się w zeszłym roku. Wtedy właśnie się w sobie zakochali.

W przykładzie (8) zaimek *to* odnosi się do całego wypowiedzenia *Zapadał mrok*. Z punktu widzenia klasycznej teorii referencji nie można połączyć w jeden klaster wyrażen koreferencyjnych tego zdania i zaimka *to*. Analogicznie nie będą uznane za koreferencyjne frazy przysłówkowe *w zeszłym roku* i *wtedy* oraz *w samym centrum* i *tam*. Jednak jeśli uznamy za Vaterem, że referencja nie ogranicza się jedynie do referencji przedmiotowej, trzeba będzie uwzględniać w anotacji również takie przykłady.

Nawet jeśli będziemy definiować referencję wąsko – wyłącznie jako referencję przedmiotową, należy uwzględnić różne jej typy. Zasadniczo w pracach z zakresu semantyki logicznej referencję rozumiano wyłącznie jako odniesienie do konkretnego, wyróżnionego obiektu. Jednak w językoznawstwie za referencję przedmiotową uważa się również odniesienie do zbioru obiektów czy obiektu, którego nadawca nie chce lub nie może zidentyfikować.

E. Padučeva (1992: 118–126) wyróżnia trzy typy referencyjnych grup imiennych:

- 1) grupy imienne określone (z obiektem jednostkowym lub zbiorem obiektów), np. *Ernest Hemingway urodził się w 1899 r., Wszyscy moi studenci zaliczyli kolokwium,*

2) grupy imienne słabo określone, np. *Mam ci coś do powiedzenia, Niektórzy referenci nie dojechali na konferencję,*

3) grupy imienne nieokreślone dla mówiącego, np. *Ktoś zjadł mój jogurt, Jacys ludzie włamali się do jego mieszkania.*

Padučeva postuluje jeszcze, aby wyróżnić czwartą grupę – ze zneutralizowaną opozycją określoności, czyli grupy imienne właściwie nieokreślone. Jest to bardzo istotne dla języków bezrodzajnikowych, jakim jest język polski: np. zdanie *Zatrzymał mnie policjant* jest dwuznaczne – z braku rodzajnika nie wiadomo, czy był to jakiś nieokreślony policjant, czy może konkretny. W języku rodzajnikowym wątpliwość tę rozstrzygnąłby użyty przed grupą imienną rodzajnik.

Padučeva wyróżnia także klasę grup imiennych, które nie mają referencji, czyli w rozumieniu autorki nie oznaczają żadnych wyróżnionych obiektów:

1) grupy imienne egzystencjalne, które odnoszą się do klas obiektów, ale nie wyróżniają żadnego z nich,

a) dystrybutywne grupy imienne oznaczające „uczestników rozdzielenych w pewnym zbiorze zdarzeń jednego typu” (Padučeva 1992: 127), np. *Czasami ktoś z nas go odwiedza, Do każdego wychowanka przyjechali jego krewni.*

b) niekonkretne grupy imienne, występujące w kontekście stłumionej asercji (tzn. z czasownikami *może, chce, powinien, należy*, z formami trybu rozkazującego, w pytaniach, negacji, z czasownikami performatywnymi, itp.), np. *Jan chce się ożenić z jakąkolwiek cudzoziemką.*

c) ogólnogzystencjalne grupy imienne, odnoszące się do obiektów w sposób ogólny, bez wyróżniania konkretnego okazu, np. *Niektórzy ludzie mają alergię na gluten.*

2) grupy imienne uniwersalne, odnoszące się do całej, abstrakcyjnej klasy obiektów, np. *Kto rano wstaje, temu Pan Bóg daje.*

3) grupy imienne atrybutywne, które odnoszą się do jakiegoś jednostkowego bytu, ale nadawca nie ma na myśli jakiegoś konkretnego obiektu, np. *Najsilniejszy człowiek na świecie nie podniósłby 500 kg, Ten, kto wygra, otrzyma nagrodę.*

4) grupy imienne oznaczające rodzaj lub gatunek, np. *On postąpił jak mężczyzna, Jaguary wymierają.*

Z kolei Z. Topolińska (1984: 303–324) w swoim podziale grup imiennych z jednostkowym referentem wyróżnia:

1) deskrypcje określone językowo zupełnie (o jednoznacznej referencji), np. *stolica Polski za Jagiellonów, autor Pana Tadeusza;*

2) deskrypcje określone językowo niezupełnie (których sama formalizacja językowa nie zapewnia jednoznacznej referencji lub kiedy referencja zmienia się wraz z sytuacją mówienia);

a) deskrypcje niezupełne wyznaczające jednoznacznie w określonej sytuacji, np. *Swędzi mnie ręka,*

b) deskrypcje niezupełne skorelowane z gestem jednoznacznego odniesienia, np. *Daj mi ten nóż*,

3) grupy imienne w funkcji wyrażen argumentowych nieidentyfikujących, np. *Coś mi wpadło do oka*.

Grupy nominalne z referentem zbiorowym Topolińska dzieli z kolei na:

1) konstytuujące zbiorowość podaną eksplicytnie (dystrybutywnie lub kolektywnie),

2) różnicujące element zbiorowości podane eksplicytnie (wszystkie elementy są wyróżnione lub tylko część albo jeden z elementów jest wyróżniony).

Zarówno Padučeva, jak Topolińska zwracają uwagę na fakt, że grupy nominalne użyte predykatywnie nie mają referencji, ponieważ nie oznaczają obiektu, ale jego właściwości, np. *Jan jest lekarzem*. Padučeva pisze także, że należy zwrócić uwagę na użycia grup nominalnych, które przypominają użycia predykatywne, ale *de facto* są argumentem przy predykcji identyfikacji, np. *Miasto, które widzimy, to Warszawa*. W tym wypadku wyrażenie *Warszawa* nie jest użyte predykatywnie i ma referencję. Z kolei Topolińska (1988: 324) uważa, że w zdaniach typu: *Na apel zgłosił się dziennikarz, Telefon przyjęła kobieta*, grupy nominalne *dziennikarz* oraz *kobieta* są wyrażeniami predykatywnymi, wtórnie wprowadzonymi na pozycję argumentów. Topolińska twierdzi, że zdania te można eksplikować jako: *Ten, kto się zgłosił na apel, był dziennikarzem, Ta, która przyjęła telefon, była kobietą*.

Jednak należy zauważyć, że grupy nominalne użyte predykatywnie również mogą tworzyć w tekście łańcuchy przypominające klastry koreferencyjne, np. *Mama Jana była architektem i on też chce nim zostać, chociaż wie, że ten zawód tej bardzo wymagający – {architektem, nim, ten zawód}*. Zatem przyjęcie stanowiska Padučovej i Topolińskiej sprawi, że w anotowanych tekstach nie zostaną uwzględnione ciągi tego typu, choć w zasadzie nie różnią się formalnie od „prawdziwych” łańcuchów koreferencyjnych. Czy zatem należy uznać, że definicja koreferencji nie pokrywa się z pojęciem referencji? Byłoby to dość niewygodne rozwiązanie, bo wymagałoby zasadniczego przededefiniowania koreferencji i ponownego rozstrzygnięcia, jaka jest jej relacja do referencji. Czy może potraktować ciągi wyrażen powiązanych z grupą nominalną w funkcji predykatu jako zupełnie odmienne zjawisko? To rozwiązanie można naturalnie przyjąć na początku pracy nad korpusem koreferencji, jednak w końcu trzeba będzie coś zrobić z tymi „pseudokoreferencyjnymi” ciągami w anotowanych tekstach. Zaliczenie tych dwóch zjawisk do różnych kategorii jest też dość problematyczne, ponieważ nie ma formalnych różnic, które by pozwoliły w jednoznaczny sposób oddzielić koreferencję „prawdziwą” od „pozornej”, w której antecedens jest grupą nominalną użytą predykatywnie.

Stawianie tak ostrej granicy między grupami nominalnymi jako nazwami argumentów a tymi samymi grupami użytymi w funkcji predykatów nie ma

właściwie uzasadnienia, jeśli przyjmie się szerszą definicję referencji (czyli także miejsca, czasu i sytuacji). Tym bardziej, że granice między argumentem a predykatem są rozmyte, co widać wyraźnie na przykładzie rzeczowników odczasowników, por.:

(11) Nie mam żadnego pytania (Padučeva 1992: 116).

(12) Bitwa zaczęła się o świcie. Trwała cały dzień i zginęło w niej wielu osadników.

(13) W biegu wzięło udział stu zawodników. Nikomu jednak nie udało się go ukończyć.

W przykładzie (11) wyrażenie *żadnego pytania* ma według Padučovej referencję, choć leksem *pytanie* nie nazywa przecież żadnego obiektu, tylko czynność (*pytanie* 'to, że ktoś pyta'). To, że jest użyte w pozycji argumentu, nie musi oznaczać automatycznie, że ma referencję – skoro Topolińska uważa, że są możliwe sytuacje, kiedy wyrażenie predykatywne jest użyte wtórnie jako argument (por. *Nie mam żadnego pytania – O nic nie chcę cię zapytać*). Analogicznie nie powinno się traktować jako referencyjnych grup nominalnych *bitwa* i *bieg* w (12) i (13), choć jak widać mogą być antecedensami ciągów anaforycznych.

W moim przekonaniu rozróżnienie grup nominalnych z referencją i bez referencji jest w dużej mierze arbitralne. Jeśli przywołamy cytowany wcześniej przykład Topolińskiej: *Jan chce być architektem*, to widać, że wystarczy drobna zmiana formalna, aby nagle analogiczna pod względem treści grupa nominalna zyskała referencję: *Jan chce wykonywać zawód architekta*. Oba wyrażenia: *być architektem* i *wykonywać zawód architekta*, mają identyczny sens. Dlaczego zatem pozbawiać referencji tę pierwszą grupę?

Argumenty semantyczne przedstawione przez Padučewą i Topolińską nie są zbyt przekonujące. Autorki odwołują się do semantyki wyrażenia, choć właściwie bardziej uwzględniają ich cechy gramatyczne (klasę części mowy) i/lub czysto składniowe (funkcję w zdaniu). Nie rozstrzygają one, co zrobić z rzeczownikowymi nazwami czynności – Padučeva najwyraźniej nie widzi różnicy między gerundiami a typowymi rzeczownikami, por. przykład (11), Topolińska zaś tę kwestię całkowicie pomija. Z kolei propozycja tej autorki, aby za użycia predykatywne uważać zdania typu *Telefon odebrała kobieta* wydaje się całkowicie nieuzasadniona. Po pierwsze interpretacja tej wypowiedzi zależy od kontekstu (np. *Telefon odebrała kobieta. Mężczyzna, który jej towarzyszył stanął zaś koło budki* – w tej sytuacji mamy jak najbardziej konkretną referencję), po drugie w zasadzie większość deskryptywnych grup nominalnych da się wyeksplikować w sposób proponowany przez Topolińską, np. *Autor „Pana Tadeusza” przyjechał do Paryża. Ten, który był autorem „Pana Tadeusza”, przyjechał do Paryża; Widział najdłuższą rzekę świata. To, co widział, było najdłuższą rzeką świata*.

Rozstrzygnięcie problemów dotyczących referencji tak naprawdę w dużej mierze zależy od odpowiedzi na pytanie, jak zdefiniujemy świat, do któ-

rego odnoszą się wyrażenia języka naturalnego. W klasycznym rozumieniu referencja dotyczy świata rzeczywistego (tak u Lyonsa czy Searle'a). Oczywiście według takiej definicji wyrażenia typu: *Pegaz, krasnoludek czy smok*, nie będą miały referencji, ponieważ nie odnoszą się do bytów realnych. Padučeva zauważa, że tak ograniczone rozumienie referencji nie ma zastosowania w językoznawstwie, ponieważ dla lingwisty nie ma znaczenia, czy w tekście mówi się o obiekcie istniejącym w świecie materialnym, czy w świecie mitologii starożytnej Grecji. Jako przykład, w którym zderzają się światy możliwe ze światem rzeczywistym, Padučeva (1992: 14) przytacza następujący tekst: „Napoleon w powieści Tołstoja nie jest podobny do Napoleona w książce prof. Tarłégo. [...] Dyskusja na temat, który z tych Napoleonów jest bardziej podobny do historycznego prototypu, nie ma sensu”. W tym przykładzie nadawca wprowadza referentów z trzech światów: świata książki Tołstoja, świata książki Tarłégo oraz historycznego Napoleona, który istniał w świecie realnym.

K. Kunz (2010: 31) zwraca uwagę, że w nowszej literaturze zakłada się, iż referenci wyrażen językowych wcale nie muszą istnieć jako obiekty w rzeczywistości. Przecież, jak podkreśla autorka, możemy mówić o faktach z odległej przeszłości (np. *Platon napisał „Dialogi”*) lub takich, które dopiero zaistnieją (*Za rok wypłacę mu zysk z polisy*), abstrakcyjnych ideach i pojęciach (*demokracja, biel, mądrość*), bytach wyobrażonych lub hipotetycznych (*centaur, Harry Potter, czarna dziura*). Kunz pisze o świecie dyskursu tak:

During speech production, recipients establish a global mental representation on the basis of the linguistic unity of the text – a mental textual world. [...] The mental textual world is considered to be the cognitive input and output of the text processing. [...] It is widely accepted that construing a textual world is a dynamic cognitive process of text representation. It involves multiple composite steps of building, specifying, updating and shifting knowledge. Text recipients start processing a text with the very first textual elements they encounter and integrate new mental concepts and relations into the textual world on the basis of the linear progression of the linguistic structure (Kunz 2010: 31–32).

Świat tekstu rządzi się własną logiką. Może mieć związek ze światem rzeczywistym, ale równie często nie ma z nim nic wspólnego (częste w literaturze czy poezji). U podstaw konstruowanego przez odbiorcę świata tekstu leży tzw. *koherencja testu*, czyli semantyczna spójność tekstu wynikająca z różnego typu związków pojęciowych lub kognitywnych, które zostały wykreowane w tekście. Według de Beaugrandé'a i Dresslera (za: Vater 2009) te związki pojęciowe można opisać jako: „właściwość czegoś”, „miejsce czegoś” i „koreferujący z czymś”. Wynika z tego, że związki koreferencyjne w tekście występują przede wszystkim na jego poziomie treściowym, choć oczywiście bywają też wyznaczane przez formalne środki językowe.

R. Langacker również uważa, że wyrażenia języka naturalnego odnoszą się do świata mentalnego, por.:

Potrąfimy mówić o wszystkim, co daje się wyobrazić. Tylko ograniczona część naszego dyskursu poświęcona jest rzeczywistym sytuacjom, należącym do naszego świata (pomimo ich uprzywilejowanego statusu). A nawet kiedy istotnie mówimy o rzeczywistych sytuacjach, nasze opisy są selektywne i schematyczne. Językowe znaczenia nie tylko odzwierciedlają opisywane sytuacje, lecz powstają w ramach interaktywnego procesu konstruowania i portretowania tych sytuacji na potrzeby komunikacji i ekspresji. W miarę rozwijania się dyskursu interlokutorzy współpracują w budowaniu, uszczegóławianiu i modyfikowaniu struktur pojęciowych, które w najlepszym razie sprowadzają się do bardzo cząstkowych reprezentacji tego, co jest omawiane. To właśnie te pojęciowe reprezentacje sytuacji, a nie ich rzeczywista natura, tworzą bezpośrednią podstawę dla znaczeń wyrazów (Langacker 2009: 350).

Langacker zwraca uwagę, że pojęcie odniesienia (czyli referencji) bywa używane w literaturze przedmiotu w sposób bardzo niekonsekwentny. Według niego klasyczna interpretacja referencji (jako odniesienia do świata realnego) nie jest odpowiednia dla ogólnej charakterystyki grup nominalnych, ponieważ desygnaty wielu grup nominalnych są abstrakcyjne lub ich istnienie obiektywne jest problematyczne (np. *domniemana nieistotność moralnych skrupułów*). Zatem dla językoznawcy najważniejszą kwestią powinno być odniesienie na poziomie dyskursu. Langacker (2009: 353) podważa w ogóle istnienie grup nominalnych niereferencyjnych. Przede wszystkim nie ogranicza on referencji grupy nominalnej do referencji przedmiotowej. To, że jakiś wyraz jest rzeczownikiem, nie wynika z faktu, że nazywa przedmiot, ale jest wytworem dwóch zjawisk poznawczych: grupowania i reifikacji. Tak pojęta „rzecz” może według autora wyłonić się na każdym poziomie organizacji pojęciowej. Jako przykład Langacker (2009: 147) podaje wyraz *przepis*: przepis może być gdzieś zapisany, ale jako taki nie istnieje w przestrzeni jako realny obiekt. Jest jednak „rzeczą”, ponieważ etapy przygotowywania jakiejś potrawy konceptualizujemy jako grupę i reifikujemy je jako jednolity proces, którego efektem jest przyrządzenie dania. Taką samą „rzeczą” jak *talerz* jest *stos talerzy*, *komisja* (złożona przecież z kilku członków) czy *chwila* (ciągły okres, którego składowe – punkty w czasie – są zgrupowane w na podstawie bliskości czasowej). W tym ujęciu nie ma zatem istotnej różnicy między referencją wyrazów typu *drzewo*, *bitwa* czy *bieganie*.

### 3. KOREFERENCJA A ANAFORA

Anafora w odróżnieniu od referencji jest rozumiana jako system wewnętrzny. Topolińska pisze, że „przez anaforę rozumiemy nawiązanie do spraw i przedmiotów już w tekście wspomnianych i w związku z tym po-

przez tekst znanych odbiorcy” (Topolińska 1988: 327). Nieco inaczej definiuje anaforę Padučeva, por.:

Związek anaforyczny jest to relacja powstająca pomiędzy wyrazami w zdaniu lub tekście wtedy, gdy znaczenie jednego wyrazu lub wyrażenia zawiera odesłanie do innego, przy czym nie występuje między nimi związek syntaktyczny. Pierwszy człon relacji anaforycznej nazywa się antecedensem, drugi – anaforem (lub wyrażeniem anaforycznym) (Padučeva 1992: 187).

Padučeva chyba nie do końca odróżnia anaforę i koreferencję, ponieważ pisze najpierw, że treścią związku anaforycznego może być koreferencyjność grup imiennych (którą autorka definiuje jako relację między takimi grupami imiennymi, które oznaczają ten sam obiekt, tzn. mają ten sam referent). Dalej jednak zauważa, że koreferencyjność jest możliwa między NP, które nie oznaczają konkretnego obiektu (czyli nie mają referencji), np. *Każdy człowiek chce, żeby go szanowano*. Według Padučovej w tym przykładzie zachodzi zjawisko (które określa jako *koasygnację*) polegające na tym, że jedna z grup ma referenta w jednym ze światów możliwych, druga zaś jest rozumiana jako odniesiona do samego referenta, por.:

Jeśli zaimek ma ograniczoną referencję, to propozycja, stanowiąca jego bezpośredni kontekst, jest pozbawiona wartości prawdy [...]. Relacja między zaimkiem i antecedensem [...] stanowi odmianę koreferencyjności, która [...] nazywana jest koasygnacją. W ogóle tezę, że wszystkie koreferencyjne grupy imienne są referencyjne, (tzn. mają referenta w świecie mówiącego), należy odrzucić [...] (Padučeva 1992: 134).

Według Padučovej koreferencji nie da się sprowadzić do tożsamości referentów ze względu na nieokreśloność denotacyjną charakterystyczną dla wielu leksemów. Autorka uważa, że w zdaniu: *Hemingwaya czyta się najlepiej, gdy opisuje on polowanie*, zaimek i jego antecedens są koreferencyjne, mimo że nie oznaczają tego samego obiektu (wyrażenie *Hemingwaya* oznacza twórczość pisarza, zaimek zaś odnosi się już do osoby). Z tych rozważań wynika, że Padučeva utożsamia w niektórych fragmentach swojej pracy koreferencję z anaforą.

Badaczka podaje ponadto ograniczenia dotyczące koreferencyjności:

- 1) referencyjne grupy imienne nie są koreferencyjne, jeżeli odnoszą się do różnych obiektów;
- 2) predykatywne NP nie mogą być koreferencyjne, ponieważ znajdują się poza referencją;
- 3) nie mogą być koreferencyjne dwie NP o statusie denotacyjnym różnego typu, np. predykatywna i termowa (Padučeva 1992: 188).

Na nieprzekładalność systemu koreferencji i anafory zwraca uwagę także Topolińska. Jej spostrzeżenia w zasadzie zgadzają się z tezami Padučovej, jednak Topolińska nie utożsamia koreferencji z anaforą. Pisze ona bowiem,

że możliwa jest anafora między grupami nominalnymi, które nie mają referencji, co więcej anaforyzowane mogą być wszystkie treści w zdaniu, np. *Chcę być architektem i nim zostanę, Zapadał mrok. Powiększyło to nasze szanse* (Topolińska 1988: 328).

Tezy Padučevy i Topolińskiej są dość problematyczne z punktu widzenia prac nad korpusem koreferencji. W tekstach bowiem wielokrotnie spotyka się ciągi wyrażenń anaforycznych, których antecedensami są grupy imienne zaliczone przez autorki do klasy wyrażenń niereferencyjnych, np.:

(14) Często tłumy ludzi przechodzą obok jakichś zjawisk i nie dostrzegają sedna sprawy, a dopiero gdy ktoś zbierze i opracuje rządzące nimi prawa, wszyscy dziwią się, że to jest takie proste i oczywiste, że nawet dziecko to zrozumie, a przez tyle lat nikt na to nie wpadł.

W powyższym przykładzie mamy wyłącznie grupy nominalne, które Padučeva uznaje za niereferencyjne. Wyrażenia *tłumy ludzi* i *jakichś zjawisk* znalazłyby się w klasie egzystencjalnej, frazy *ktoś, nikt* w grupie atrybutywnej, a grupy nominalne *rządzące prawa, wszyscy* i *dziecko* w klasie uniwersalnej. Jeśli nie mają one referencji, to nie można uznać za koreferencyjne ciągów: {*tłumy ludzi, nie dostrzegają*}, {*jakichś zjawisk, nimi*}, {*rządzące prawa, to, to, to*}, {*ktoś, opracuje*}. Również grupy nominalne użyte predykatywnie bywają wykorzystywane jako antecedensy w ciągach wyrażenń anaforycznych. Zatem przyjęcie tez Padučevy zasadniczo ograniczyłyby możliwości anotacji korpusu koreferencji. Tym bardziej, że z niektórymi założeniami autorki trudno się zgodzić. Przykład podany przez Padučevą: *Hemingwaya czyta się najlepiej, gdy opisuje on polowanie* jest ewidentną elipsą: [*Książki*] *Hemingwaya czyta się najlepiej, gdy opisuje on polowanie*. Dość jasne jest, że zaimiek *on* odnosi się do pisarza, a nadrzędnik pierwszego wyrażenia został w zdaniu opuszczony.

R. Langacker (2009) z kolei zakłada, że jeśli grupa nominalna może być w tekście antecedensem zaimków, które się niej odnoszą, to znaczy, że ma ona wirtualny punkt odniesienia w świecie dyskursu, czyli w tym ujęciu – ma referencję. Podejście to wydaje się słuszne, zwłaszcza z punktu widzenia prac nad przetwarzaniem języka naturalnego, ponieważ już samo istnienie łańcucha wyrażenń anaforycznych w tekście staje się oznaką, że grupa nominalna jest referencyjna na poziomie świata tekstu i nie potrzeba wprowadzać sztucznego i problematycznego (nawet dla człowieka) odróżnienia grup nominalnych referencyjnych i niereferencyjnych, co proponowały Topolińska i Padučeva.

## 4. KOREFERENCJA A IDENTYCZNOŚĆ

Problem wyznaczenia koreferencji wiąże się z dwoma zasadniczymi problemami: po pierwsze z samym pojęciem *identyczności*, po drugie z faktem, że leksemy łączą także różnego typu związki semantyczne, które utrudniają odbiorcy podjęcie decyzji, czy ma do czynienia z tym samym referentem, czy jednak już innym. Według Anny Wierzbickiej (2010: 61) identyczność jest uniwersalną, podstawową i nierozkładalną jednostką semantyczną. Inaczej widzą to Fauconnier i Turner (2002), którzy uważają identyczność za najbardziej istotną relacją podstawową (*vital relation*). Według autorów nie oznacza to wcale, że jest to relacja nieskomplikowana, ponieważ wymaga ona od odbiorcy wielkiego wkładu wyobraźni, aby mógł on dokonać analizy i syntezy licznych przestrzeni mentalnych – na przykład musi on połączyć przestrzenie mentalne niemowlęcia, dziecka, nastolatka i dorosłego relacją identyczności, mimo oczywistych różnic między nimi, a dodatkowo włączyć jeszcze do procesu integracji inne relacje, typu zmiana, czas i przyczyna-skutek. Z kolei różne związki semantyczne między leksemami mogą spowodować, że odbiorca błędnie uzna za ciąg koreferencyjny zestaw wyrażzeń, które są jedynie połączone jakąś relacją leksykalną, typu synonimia, holo-/meronimia, hipo-/hiperonimia. Takie niejasności doprowadziły do tego, że M. Recasens wprowadziła dodatkowo pojęcie *near-identity*. Twierdzi ona, że „coreference relations between DEs depend on criteria of identity largely determined by the linguistic and pragmatic context” (Recasens et al. 2011; 1142). Autorka pisze dalej, że identyczność jest rozmyta i stopniowalna. Recasens (2010: 151) wyróżnia następujące typy *near-identity*:

**A. Name metonymy**

- a. Role
- b. Location
- c. Organization
- d. Information realization
- e. Representation
- f. Other

**B. Meronymy**

- a. Part\_Whole
- b. Stuff\_Object
- c. Set\_Set

**C. Class**

- a. More specific
- b. More general

**D. Spatio-temporal function**

- a. Place
- b. Time

c. Numerical function

d. Role function

Wyróżnienie pojęcia *near-identity* wynika w moim przekonaniu z tego, że autorka myli poziomy: językowy, pojęciowy oraz realny. Nie podlega dyskusji, że leksemy mają jakieś wspólne elementy semantyczne i że istnieją między nimi różnego typu związki. Są one opisywane w takich projektach, jak WordNet czy FrameNet. Należy jednak pamiętać, że referencja (i co za tym idzie koreferencja) nie jest własnością systemu językowego, ale wypowiedzi, czyli realizuje się dopiero na płaszczyźnie pragmatycznej. Wyrażenia koreferencyjne nie muszą być równokształtne, synonimiczne czy połączone jakimikolwiek związkami semantycznymi, np.:

(15) Jedynym moim pożywieniem w ostatnich trzech dniach były węglowodany – powiedział trzykrotny złoty medalista. Sportowcy niemieccy są ogromnie zaskoczeni wiadomością o pozytywnym wyniku testu antydopingowego Muehlegga.

(16) Kruk krukowi oka nie wykole.

W przykładzie (15) wyrażenia *Muehlegga* i *trzykrotny złoty medalista* są koreferencyjne, choć nie łączą ich żadne związki semantyczne. Z kolei formy tego samego leksemu: *kruk* i *krukowi* w zdaniu (16) nie są koreferencyjne, ponieważ odnoszą się do dwóch różnych referentów, które po prostu należą do tego samego gatunku.

Przykłady *near-identity* podawane przez Recasens w większości wypadków dają się dość łatwo wyjaśnić różnymi zjawiskami zachodzącymi na poziomie gramatycznym, na poziomie semantycznym czy na poziomie pojęciowym. Poniższy przykład Recasens uznałaby w myśl swojej teorii za typ *near-identity* *ROLE*, por.:

(17) Są akcentami dużej wystawy plastycznej, prezentującej twórczość Jana Stępkowskiego, artyści ze Strzegowa. Właściwie artyści przedstawiać nie trzeba; to jeden z najwybitniejszych na północnym Mazowszu rzeźbiarzy, a jednocześnie znakomity malarz. Jako rzeźbiarz – sięga po różne techniki i materiały, nawet tak trudne do obróbki i kształtowania, jak bazalt i granit.

Według Recasens wyrażenia: *jeden z najwybitniejszych na północnym Mazowszu rzeźbiarzy, znakomity malarz, jako rzeźbiarz* powinny być połączone z antecedentem *Jan Stępkowski* relacją *near-identity*. Problem polega na tym, że te anafory są użyte w funkcji predykatów – opisują po prostu właściwości artysty, zatem nie tworzą żadnej identyczności czy *near-identity* z wyrażeniem *Jan Stępkowski*. Wystarczy zamienić te frazy nominalne na czasowniki, aby było to wyraźnie widoczne: *Właściwie artyści przedstawiać nie trzeba; rzeźbi i jednocześnie maluje; Kiedy rzeźbi – sięga po różne techniki i materiały, nawet tak trudne do obróbki i kształtowania, jak bazalt i granit.*

Większość przykładów podawanych przez Recasens dotyczy sytuacji, kiedy między leksemami (czyli jednostkami systemu) zachodzi jakaś relacja semantyczna, np.:

(18) Od czasu okupacji... – Ale tu je masz z powrotem, w metryce, i musisz ich używać w urzędowych papierach – powiedział oschle dyrektor i podsunął mi nowy blankiet do wypełnienia. – Kiedy to jest stara metryka, którą mi odtworzono zaraz po wojnie.

Leksemy *wojna* i *okupacja* są oczywiście związane jakąś relacją semantyczną (w WordNecie określa się ją jako *entailment*), jednak nie oznacza to, że można je automatycznie uznać za przykład *near-identity*. Z tekstu nie wynika, żeby nadawca traktował te wyrażenia jako koreferencyjne. W kreacji świata tekstu nadawca może wykorzystywać własności semantyczne leksemów, ale nie musi. Wyrażenia koreferencyjne nie zawsze są powiązane jakimikolwiek relacjami semantycznymi, por.: *Widziałas dziś naszą księgową? Co ta debilka dziś na siebie założyła!* Leksemów *księgową* oraz *debilka* nie łączy w systemie żadna relacja leksykalna, a przecież są w tym tekście koreferencyjne.

Często niewłaściwa identyfikacja wyrażeń koreferencyjnych wynika z faktu, że odbiorca nie dysponuje taką samą wiedzą o świecie, jaką ma nadawca, por.:

(19) Setnik Lipkowski, rzekomy Mellechowicz, jest synem groźnego Tataru Tuhaj-beja, który wyrządził wiele szkód Polsce. Azja podstępem próbuje porwać Basię.

Jeśli odbiorca nie czytał powieści Henryka Sienkiewicza „Pan Wołodyjowski”, nie będzie wiedział, że *setnik Lipkowski, rzekomy Mellechowicz* oraz *Azja* odnoszą się do tego samego referenta – jednego z bohaterów literackich tej książki.

Zdarzają się oczywiście przykłady bardziej problematyczne, które Recasens również uważa za *near-identity*, np.:

(20) W miejscu dawnej jezdni ryją buldożery. Bez problemu można dojechać ul. Bandurskiego, a że nawierzchnia Retkińskiej była znana jako jedna z najbardziej dziurawych w mieście, nikt nawet specjalnie nie skarży się na utrudnienia w ruchu. Nowa Retkińska będzie miała i sygnalizację u zbiegu z ul. Krzemieniecką, i chodnik (spory odcinek ulicy był go całkowicie pozbawiony).

W tym przykładzie mamy wyrażenia *Retkińska* oraz *nowa Retkińska*, które pozornie odnoszą się do tego samego bytu w świecie rzeczywistym – czyli konkretnej ulicy. Jednak tak naprawdę wyrażenie *nowa Retkińska* odnosi się do bytu wirtualnego – ulicy w przyszłości, po zakończeniu remontu. Zatem nie są to te same referenty w świecie dyskursu. Nadawca chce, aby odbiorca myślał o nich jako o odrębnych obiektach, co sygnalizuje przez dodanie do nazwy ulicy przymiotnika *nowa*.

## 5. PODSUMOWANIE

Podsumowując powyższe rozważania, wydaje się konieczne uporządkowanie pewnych terminów: referencja i koreferencja dotyczą odnoszenia się wyrażen do zjawisk wykreowanych w świecie dyskursu, anafora zaś jest systemem wewnątrztekstowym, który pozwala w tekście wyznaczyć między innymi relacje koreferencji. W moim przekonaniu nie ma potrzeby wprowadzania dodatkowego terminu *near-identity*, który niczego nie wyjaśnia, ale raczej zaburza strukturę anotowanych tekstów, ponieważ miesza ze sobą odmiennie poziomy języka – systemu i mowy. Niejasność dotycząca identyczności referentów bierze się najczęściej z zaburzenia komunikacji pomiędzy nadawcą i odbiorcą na różnych poziomów podczas dekodowaniu tekstu: na poziomie pojęciowym (np. odbiorca i nadawca mogą dysponować różną wiedzą na temat danego obiektu), na poziomie semantycznym (np. odbiorca nie zna słowa, nadawca używa niewłaściwego słowa), na poziomie składniowym (np. elipsa, skrót myślowy). W przypadku określania koreferencji decydujący jest poziom dyskursu – to on wyznacza co jest, a co nie jest koreferencyjne. Mogą być tam wykorzystane własności semantyczne leksemów, świat tekstu może się w jakimś stopniu pokrywać ze światem rzeczywistym, ale nie jest to obligatoryjne. Nadawca równie dobrze jest w stanie wykreować wirtualne typy referentów (np. trzech Napoleonów czy dwie ulice Retkińskie) i w tym przypadku nie ma znaczenia fakt, że w świecie rzeczywistym istniał tylko jeden Napoleon, czy że jest tylko jedna ulica Retkińska. Istotne jest to, że w świecie dyskursu są to różni referenci i jedynie ten poziom powinien być brany pod uwagę podczas anotacji wyrażen koreferencyjnych w korpusie.

## BIBLIOGRAFIA

- Fauconnier G., Turner M. 2002. *The Way We Think. Conceptual Blending and the Mind's Hidden Complexities*, New York: Basic Books.
- Kunz K. 2010. *Variation in English and German Nominal Coreference*, Frankfurt am Main, Berlin, Bern, Bruxelles, New York, Oxford, Wien: Peter Lang.
- Langacker R. 2009. *Gramatyka kognitywna. Wprowadzenie*, Kraków: Universitas.
- Pađučeva E.V. 1992. *Wypowiedź i jej odniesienie do rzeczywistości. (Referencyjne aspekty znaczenia zaimków)*, Warszawa: PWN.
- Recasens M., Hovy E., Marti M. 2010. *A Typology of Near-Identity Relations for Coreference (NIDENT)*. N. Calzolari and K. Choukri and B. Maegaard and J. Mariani and J. Odijk and S. Piperidis and M. Rosner and D. Tapias (ed.), 7. *International Conference on Language Resources and Evaluation, LREC 2010. European Language Resources Association (ELRA)*, Valletta, Malta.
- Recasens M., Hovy E., Marti M. 2011. *Identity, non-identity, and near-identity: Addressing the complexity of coreference*. *Lingua* 121(6), 1138–1152.

- Topolińska Z. 1984. *Składnia grupy imiennej*. Zuzanna Topolińska (ed.), *Gramatyka współczesnego języka polskiego. Składnia*, Warszawa: PWN.
- Van Hoek K. 2009. *Pronouns and point of view: cognitive principles of coreference* [w:] *The new psychology of language: cognitive and functional approaches to language structure*, vol. 2, Tomasello M. (ed.), Tylor & Francis e-library.
- Vater H. 2009. *Wstęp do lingwistyki tekstu. Struktura i rozumienie tekstów*, tłum. Błachut E., Gołębiowski A., Wrocław: Atut.
- Wierzbicka A. 2010. *Semantyka. Jednostki elementarne i uniwersalne*, Lublin: Wydawnictwo UMCS.