

Ten sam czy taki sam?
O problemach z identycznością
z perspektywy odbiorcy

Magdalena Zawisławska (UW)
Maciej Ogrodniczuk (IPI PAN)

Koreferencja

Zjawisko językowe polegające na tym, że zestaw jakichś wyrażen w tekście odnosi się do tego samego przedmiotu (referenta), czyli zachodzi identyczność referencji.

Wykładnikami koreferencji w tekście mogą być:

- imiona własne,
- deskrypcje określone (peryfrazy),
- zaimki anaforyczne,
- wykładniki kategorii osoby wyrażone w formie czasownikowej.

Projekt CORE

Komputerowe metody identyfikacji nawiązań w tekstach polskich

Projekt jest finansowany przez Narodowe Centrum Nauki
(nr kontraktu 6505/B/To2/2011/40).

Ramy czasowe: kwiecień 2011 – kwiecień 2014.

Główny cel projektu:

Opracowanie technik, narzędzi oraz zasobów
do automatycznej identyfikacji nawiązań w języku polskim
o jakości porównywalnej z uzyskiwaną dla innych języków.

Wykonawcy

- **Maciej Ogrodniczuk – kierownik projektu,**
- Barbara Dunin-Kępcicz – nadzór merytoryczny i doradztwo w zadaniach lingwistycznych,
- Adam Przepiórkowski – ekspertyza lingwistyczna i informatyczna, współpraca projektowa z NKJP,
- Agata Savary – ekspertyza dotycząca anotacji korpusu, jednostek nazewniczych i wielowyrazowych oraz narzędzi do anotacji koreferencji,
- Łukasz Dębowski – udział w tworzeniu i rozbudowie narzędzi statystycznych,
- Mateusz Kopeć – główny informatyk, autor środowiska anotacyjnego,
- Katarzyna Głowińska – ekspertyza lingwistyczna
- Magdalena Zawisławska – ekspertyza lingwistyczna i semantyczna, organizacja pracy anotacyjnej, superanotacja korpusu nawiązań,
- Piotr Batko, Maria Głabska, Anna Grzeszak, Emilia Kubicka, Barbara Masny, Paulina Rosalska, Sebastian Żurowski – anotacja koreferencji.

Narzędzie do anotacji tekstów

 [Panie Ministrze]

głowa

Panie

klaster

set_14

link

empty

komentarz

Zastosuj

Cofnij zmiany

 Przeglądarka linków

Tekst

- [CBA] -> [służbach wojskowych]
- [ABW] -> [służbach wojskowych]
- [WSI] -> [służbach wojskowych]
- [jakieś zakazy] -> [występowania w mediach]
- [ujawniania swoich poglądów] -> [jakieś zakazy]

 Przeglądarka klastrow

☐ Rozwiń wszystko

Tekst

- MarszalekBogdanBorusewicz [5]
 - [MarszalekBogdanBorusewicz]
 - [poproszę]
 - [Panie Marszałku]
 - [MarszalekBogdanBorusewicz]
 - [Proszę]
- SenatorCzeslawRyszka [8]
- PelnomocnikRzadudosprawOpracowaniaProgramuZwalczaniaNaduzycwIn:
 - [pana ministra]
 - Panie Ministrze
 - [Panie Ministrze]
 - [PelnomocnikRzadudosprawOpracowaniaProgramuZwalczaniaNaduzycwIn:]
 - [powiem]

MMAX2 dla IPAN wersja 1.1 Tekst: A.mmax

Tekst

Ustawienia

Wygląd

Przeglądarki

Dodatki

[MarszalekBogdanBorusewicz] : [Pan senator [Ryszka]] . I potem [poproszę] [pana [ministra]] o [odповідź] .

[SenatorCzeslawRyszka] : [Dziękuję] , [Panie [Marszałku]] . [Panie [Ministrze]] , [mam] [pytanie] o [nabór [funkcjonariuszy do [CBA]]] . Jak ma przebiegać na [przykład] to [przepytanie za [pomocą [wariografu]]] ? Jakie będą [im] [zadawane pytania] ? Tak z [ciekawości] [pytam] . Skąd w ogóle będą pochodzić [[[urzędnicy] i [funkcjonariusze]] [tego urzędu]] ? Czy to będą [pracownicy [[ABW] , [WSI]] , tacy , którzy , że tak [powiem] , zdali tam [egzamin z [ucziwości]]] ? [Drugie pytanie] to [pytanie o [apolityczność [funkcjonariuszy]]] . [Ja] rozumiem , że [oni] nie mogą należeć do [partii politycznych] . Ale czy będą [ich] obowiązywać jeszcze [jakieś zakazy] , na [przykład] [[występowania w [mediach]] , [ujawniania [swoich poglądów]]] itd . ? [Dziękuję] bardzo .

[MarszalekBogdanBorusewicz] : [Proszę] bardzo , [Panie [Ministrze]] .

[PelnomocnikRzadudosprawOpracowaniaProgramuZwalczaniaNaduzycwInstitucjachPublicznychMariuszKaminski] : Moż najpierw [sprawa legislacyjna] , związana z [WSI] w [ustawie] . Oczywiście , przygotowując [projekt [ustawy]] , musieli [śmy] odnosić się do [obowiązującego aktualnie stanu prawnego] , stąd też [WSI] musiało być wpisane . Nie mieli [śmy] do [końca] [pewności] , jak szybko [parlament] będzie pracował nad [tymi ustawami] , która z [tych ustaw] będzie pierwsza . Al są to tak naprawdę [zmiany legislacyjne w [przepisach przejściowych]] , tak więc to jest [kwestia [prac legislacyjnych nad [ustawami o [służbach wojskowych]]]] , gdzie [te korekty] powinny być dokonane . A [my] nie mogli śmy podejść do [tego] inaczej , nie mogli [śmy] przyjąć [założenia] , że [parlament] na pewno podejmie [decyzję o [reformie [służb wojskowych]]] , projektując [tę ustawę] , odnosić się do [przyszłego i niepewnego stanu prawnego] . Jeżeli chodzi o [pytanie [pana senatora [Piesiewicz]]] , to [powiem] tak : oczywiście , [Panie [Senatorze]] , że [kodeks [postępowania karnego]] , jak już mówili [śmy] na [posiedzeniu [komisji]] , będzie [biblią [funkcjonariuszy [Centralnego Biura Antykorupcyjnego]]] i [wszystkie [rygory] i [zasady] z [tym] związane] muszą być i będą respektowane .

Przy tych słowach **Niełkowski(1)** odsunął gwałtownym ruchem **Władysława(2)** i poszedł do apteki. **Turkowski(2)** chciał biec za **nim(1)** i wypoliczkować **go(1)** ... widok jakiejś staruszki w aptece powstrzymał **krewkiego ucznia(2)**. A w końcu pierwszy gniew przeminął. Rozsądek wziął górę. Bo i cóż by **mu(2)** przyszło z oskarżenia **Niełkowskiego(1)**? Czyżby **mu(2)** uwierzyli? A gdyby **ten nicpoń(1)** istotnie uciekł się do piekielnej broni i rzucił podejrzenie na **niego(2)**? O ... wówczas by przepadł ... zginął chyba ... Wuj by **mu(2)** nie przebaczył.

(1)[**Niełkowski, nim, go, Niełkowskiego, ten nicpoń**]

(2)[**Władysława, Turkowski, krewkiego ucznia, mu, mu, niego, mu**]

Problemy z identyfikacją koreferencji na poziomie językowym

The Procurator's cheek twitched and he said quietly:

‘Bring in the accused.’

At once two legionaries escorted a man of about twenty-seven from the courtyard, under the arcade and up to the balcony, where they placed him before the Procurator's chair. The man was dressed in a shabby, torn blue chiton. His head was covered with a white bandage fastened round his forehead, his hands tied behind his back. There was a large bruise under the man's left eye and a scab of dried blood in one corner of his mouth. The prisoner stared at the Procurator with anxious curiosity.

The Master and Margarita by Mikhail Bulgakov, Translated from the Russian by Michael Glenny, Printed in Great Britain by Collins Clear-Type Press London and Glasgow © 1967

Procuratorowi skurcz wykrzywił policzek. Powiedział cicho:

– Wprowadźcie oskarżonego.

Natychmiast dwóch legionistów wprowadziło między kolumny z ogrodowego placu dwudziestosiedmioletniego człowieka i przywiodło go przed tron procuratora. Człowiek ów odziany był w stary, rozdarty, błękitny chiton. Na głowie miał biały zawój przewiązany wokół rzemykiem, ręce związane mu z tyłu. Pod jego lewym okiem widniał wielki siniak, w kącie ust miał zdartą skórę i zaschlą krew. Patrzył na procuratora z lękliwą ciekawością.

Michał Bułhakow, *Mistrz i Małgorzata*, tłum. Irena Lewandowska, Witold Dąbrowski, Czytelnik, Warszawa 1988.

Właściwości gramatyczne polszczyzny, które wpływają na koreferencję

1. Swobodny szyk zdania
2. Tzw. *podmiot domyślny* i wyrażanie koreferencji jedynie za pomocą końcówki osobowej czasownika
3. Brak systemu rodzajników

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 ministry poszedł też były **asystent** Józefa Oleksego Lech Nikolski. Posłowie nie poprzestają na jednym **asystencie**.

Koreferencja a inne zjawiska semantyczne

Wydawałoby się na mocy reguł językowych, że jeśli N jest nazwą, to dowolne dwa użycia nazwy N mają to samo odniesienie przedmiotowe. Jednak bardzo często użycie w tekście tych samych nazw nie musi wcale oznaczać, że odnoszą się one do tego samego obiektu. Badacze wyróżniają zjawisko określane jako *współklasyfikacja* (ang. *co-classification*) – wyrażenia są powiązane ze sobą na mocy faktu, że odnoszą się do różnych referentów, choć należących do tej samej klasy (Kunz 2009: 32), por.:

Ten chłopiec cały czas rozmawia podczas lekcji, a **tamten chłopiec** jest grzeczny.

– Pożycz mi **długopis**. – A co się stało z **twoim długopisem**?

Odbiorca może mieć wątpliwości, czy dane wyrażenia są koreferencyjne, w sytuacji, gdy jednostki leksykalne łączy jakaś relacja semantyczna. Badacze uważają, że takie wyrażenia są połączone na zasadzie relacji współekstensji (*co-extension*), co oznacza, że dwa lub więcej wyrażeń odnoszą się do czegoś, co znajduje się w tym samym, ogólnie rozumianym polu pojęciowym. Chodzi tu o różnego typu relacje semantyczne między wyrazami, np. mero/holonimie, klasę/egzemplarz, hipo/hiperonimia, antonimie itp. (Kunz 2009: 32-34). Relacje tego typu mogą utrudniać ustalanie, czy mamy do czynienia z identycznością referentów, czy mówimy o różnych obiektach połączonych jedynie jakimś typem podobieństwa.

[...] **mity** są niezastąpionym narzędziem dla psychologa, usiłującego prześledzić wzorce ludzkich zachowań. Wysiłki archeologów, religioznawców, antropologów doprowadziły z jednej strony do porzucenia eurocentrycznego spojrzenia na **mitologię** i odkrycia ...

... – 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**.

Teraz **pożar** zwałił się na budy kupieckie na Velia Carinae, pochłonął je jednym łykiem, łapczywie i prędko, po czym uderzył zaraz szeroką ścianą **ognia** na skupisko suburskie. Wielopiętrowe domy stawały w **płomieniach** jedne za drugimi. Z ogarniętych **ogniem** insul ludzie nie mieli czasu uciekać. [...] Zresztą i tym, co znaleźli się na dole, w ciasnych, krętych uliczkach pełnych szalejącego **ognia** i duszącego **dymu** niełatwo było uciekać. Krzycząc i nawołując się rozpaczliwie ludzie biegali tam i z powrotem, nie znajdując dla siebie wyjścia z **morza ognia**. Wpół oszaleli, w płonących tunikach, pędzili na oślep przed siebie, wpadali w **płomienie** i ginęli. Podcięte **ogniem** insule poczęły się walić. Osuwały się z łoskotem, zawalając ciasne uliczki. **Snopy iskier** buchały w niebo.

Koreferencja a pojęcie identyczności

Identyczność jest zwykle definiowana jako tożsamość, którą można wyrazić za pomocą określnika *ten sam*, uznanego przez Annę Wierzbicką za uniwersalną i elementarną jednostkę semantyczną (Wierzbicka 2010: 61)

Według Fauconniera i Turnera identyczność jest podstawową istotną relacją (ang. *vital relation*). Podkreślają oni dużą złożoność procesów percepcyjnych i pojęciowych, które pozwalają człowiekowi określić identyczność. Identyczność nie musi wynikać z percepcyjnego podobieństwa między obiektami (np. niemowlę i ta sama osoba jako dorosły człowiek) (za: Libura 2010: 98).

Recasens uważa identyczność za zjawisko stopniowalne. W jej ujęciu istnieją ewidentne przykłady pełnej identyczności, braku identyczności oraz przypadki pośrednie, które określa jako prawie identyczność (ang. *near identity*) (Recasens i in. 2011: 1142-1143).

Twój ojciec był największy – powiedziała jakaś nieznana starsza pani, potrząsając ręką Aleksandra – syna Gassmana. – Będzie mi *go* brakować jako *aktora*, ale zwłaszcza będę tęsknił za *ojcem* – powiedział.

Nie widziała „*Przeminęło z wiatrem*”, ale *je* czytała.

Staliśmy, patrząc na dwa portrety *Królowej Elżbiety*. Na tym z lewej była *ona* ubrana jako cesarzowa Indii. Na tym z prawej *królowa* nosiła elegancką, błękitną suknię.

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).

Różne światy – różne koreferencje

Woda – lód – para wodna – H₂O

Pomidor – owoc

Palma – drzewo

Kalafior – kwiat

Kijanka – żaba

Gąsienica – motyl

Różnice w zasobie wiedzy pozajęzykowej między nadawcą i odbiorcą

Wówczas od przywozu samochodu nabytego na szczególnych warunkach nie płaci się **podatku od towarów i usług** [...]

3) Deklaracje. Deklaracje podatkowe w zakresie podatkowe dokonywanych nabyć za okresy miesięczne **podatku VAT**.

Winicjuszowi, który miał ochotę rozpytać Ursusa o ojczysty kraj **Ligii**, słowa te sprawiły pewną przyjemność, albowiem rozmowa z człowiekiem wolnym, jakkolwiek prostym, mniejszą przynosiła ujmę jego rzymskiej i patrycjuszowskiej godności niż rozmowa z niewolnikiem, w którym ni prawo, ni obyczaj nie uznawały ludzkiej istoty. – Toś ty nie Aulusów? – spytał. – Nie, panie. Ja służę **Kallinie**, jako służyłem jej matce, ale po dobrej woli.

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

Podsumowanie

Problemy z identyfikacją wyrażeń koreferencyjnych wynikają z zaburzeń na dwóch płaszczyznach:

1. Płaszczyzna językowa:

- 1.1. Koreferencja jest związana z właściwościami systemu gramatycznego danego języka
- 1.2. Inne semantyczne zależności pomiędzy wyrażeniami (ang. *co-classification* oraz *co-extension*) są wyrażane za pomocą tych samych wykładników co koreferencja

2. Płaszczyzna pozajęzykowa: koreferencja wyznacza relację identyczności pomiędzy referentami na poziomie pojęciowym/doświadczeniowym.

- 2.1. Brak jasnej definicji identyczności i duża złożoność tego procesu na poziomie neurologicznym, percepcyjnym i mentalnym.
- 2.2. Różnice w zasobie wiedzy nadawcy i odbiorcy.

Literatura

Gilles Fauconnier, Mark Turner, *The Way We think. Conceptual Blendings and the Mind's Hidden Complexities*, New York: Basic Books 2002.

Kerstin Anna Kunz, *Variation in English and German Nominal Coreference*, Peter Lang, 2009.

Libura Agnieszka, *Teoria przestrzeni mentalnych i integracji pojęciowej. Struktura modelu i jego funkcjonalność*, Wydawnictwo Uniwersytetu Wrocławskiego, Wrocław 2010.

Marta Recasens, Eduard Hovy, M. Antonia Marti, *A Typology of Near-Identity Relations for Coreference (NIDENT)*. [w:] Nicoletta Calzolari and Khalid Choukri and Bente Maegaard and Joseph Mariani and Jan Odijk and Stelios Piperidis and Mike Rosner and Daniel Tapias (red.) *Materiały 7. International Conference on Language Resources and Evaluation, LREC 2010*. European Language Resources Association (ELRA), Valletta, Malta 2010.

Marta Recasens, Eduard Hovy, M. Antonia Marti, *Identity, non-identity, and near-identity: Addressing the complexity of coreference*. *Lingua* 121(6), 1138–1152 (2011).

Zuzanna Topolińska. *Składnia grupy imiennej* [w:] *Gramatyka współczesnego języka polskiego. Składnia*. Państwowe Wydawnictwo Naukowe, Warszawa 1984.

Anna Wierzbicka, *Semantyka. Jednostki elementarne i uniwersalne*, Wydawnictwo UMCS, Lublin 2010.

DZIĘKUJĘ ZA UWAGĘ