

Negacja w języku polskim,

Implementacja n-słów w środowisku Separ

Dariusz Leniowski

13 września 2011

Streszczenie

Artykuł ten poświęcony jest zjawiskom występującym przy słowach wprowadzających negację (n-słowach). Owocem projektu jest gramatyka, która odtwarza najważniejsze reguły uzgadniania i semantyki. Jako reprezentatywne wybrano słowa: *nikt*, *nic* oraz *zaden*. Całość została zaimplementowana w środowisku Separ.

1 Wprowadzenie

W języku polskim słowa zanegowane wprowadzają dodatkowe ograniczenia na całą frazę. Najważniejszym z nich jest wymuszenie negacji czasownika, niezależnie czy słowo zanegowane znajduje się w kontekście podmiotu czy dopełnienia: ¹

- 1a. Nikt *nie* spał.
- 1b. (Ja) *nie* napisałem **niczego**.
- 1c. (On) *nie* poszedł do **żadnej** szkoły.

W pewnych przypadkach wymuszenie negacji przenosi się na wyższy poziom w hierarchii, np.

- 2a. Marek chciał *nie* iść **nigdzie**.
- 2b. Marek *nie* chciał iść **nigdzie**.
- 2c. Jackowi przeszkadza *nie*obgadywanie **nikogo**.
- 2d. Jackowi *nie* przeszkadza obgadywanie **nikogo**.

Zdarza się jednak, że takie przeniesienie jest niepoprawne:

- 3a. Jackowi przeszkadza, że *nie* obgadują **nikogo**.
- 3b. *Jackowi nie przeszkadza, że obgadują **nikogo**.
- 4a. Jacek pisał książkę *nie* używając **niczego**.
- 4b. *Jacek nie pisał książki używając **niczego**.

Warto zwrócić uwagę różnicę w odmianie wyrazu *książka* w dwóch ostatnich prezentowanych zdaniach. Zjawisko to jest zwane dopełniaczem negacji i powoduje zamianę biernika na dopełniacz w przypadku zanegowanych czasowników.

Dodatkowo zaznaczę, że w języku polskim niektóre słowa wprowadzające negację występują jeszcze w innych znaczeniach, np.

¹Specjalnie wyróżniono słowa *wymuszone* oraz *zanegowane*. Zdania niepoprawne oznaczone są gwiazdką.

- 5a. O **niczym** *nie* rozmawiałem.
- 5b. Rozmawiałem o **niczym**. [Rozmawiałem o czymś, ale nie było to nic istotnego.]
- 6a. **Donikąd** *nie* poszedłem.
- 6b. Poszedłem **donikąd**. [Gdzieś poszedłem, ale nie trafiłem w żadne szczególne miejsce.]
- 7a. Czekanie w korku *nie* było **żadną** przyjemnością.
- 7b. Czekanie w korku to **żadna** przyjemność. [Czekanie w korku to wątpliwa przyjemność.]
8. W tym zdaniu używam słowa **nikt**.

Kluczowe jest, że zdania 5a, 6a i 7a mają znaczenie „nie istnieje”, natomiast 5b, 6b oraz 7b posiadają wydźwięk bliski „istnieje, ale jest to nieważne, wątpliwe, etc.” Naturalnie, w zdaniu 8 słowo *nikt* jest cytowane i nie podlega standardowym regułom.

Więcej informacji na temat negacji w języku polskim, wraz z bardziej wyczerpującą analizą, można znaleźć w [2]; z pracy tej zostały zaczerpnięte niektóre przykłady.

2 Syntaktyka

Aby wymusić zgodność pomiędzy odpowiednimi frazami, każdy wierzchołek w drzewie wyprowadzenia został wzbogacony o dwie właściwości, których możliwe wartości wraz z wyjaśnieniami zostały podane poniżej.

- ZanegowanieGóra
 - Tak — fraza wymaga od wierzchołka-ojca akceptacji negacji.
 - Nie — fraza nie wymaga od wierzchołka-ojca akceptacji negacji.
- ZanegowanieDół
 - Niezanegowane — fraza akceptuje tylko niezanegowane wierzchołki-dzieci.
 - NiePropaguj — fraza akceptuje dowolne wierzchołki-dzieci, nie zmienia to wartości właściwości ZanegowanieGóra.
 - Propaguj — fraza akceptuje dowolne wierzchołki-dzieci, ale wierzchołki wymagające akceptacji negacji powodują, że fraza wynikowa również będzie wymagać akceptacji negacji.

Zakładamy, że pojedyncze słowa posiadają odpowiednie adnotacje w leksykonie, np.

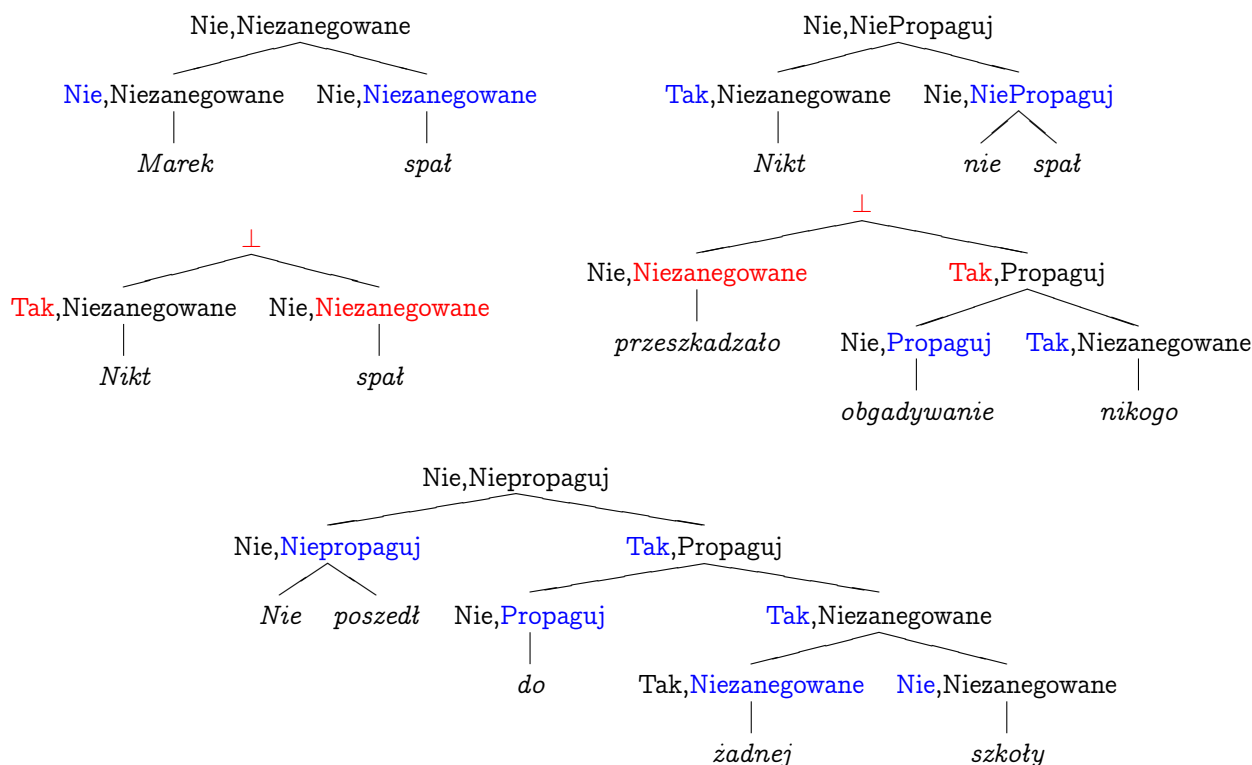
	ZanegowanieGóra	ZanegowanieDół
marek	Nie	Niezanegowane
nic	Tak	Niezanegowane
żaden	Tak	Niezanegowane
spał	Nie	Niezanegowane
nie spał	Nie	NiePropaguj
do	Nie	Propaguj
obgadywanie	Nie	Propaguj
nieobgadywanie	Nie	NiePropaguj
używając	Nie	Niezanegowane
nie używając	Nie	NiePropaguj

Następnie wystarczy je kombinować ze sobą według poniższych reguł (brak dopasowania oznacza niepoprawne wyprowadzenie):

fraza dominująca	fraza zdominowana	fraza wynikowa
α , Niezanegowane	Nie	α , Niezanegowane
α , NiePropaguj	β	α , NiePropaguj
Nie, Propaguj	β	β , Propaguj
Tak, Propaguj	β	Tak, Propaguj

Reguły w dwóch ostatnich wierszach tabeli pokrywają również przypadek wielokrotnych dopełnień dla słów, które propagują negację — wystarczy jedna zanegowana fraza, aby całość przejęła cechę.

W poniższych przykładach błędy dopasowania zaznaczono na **czzerwono**, a poprawne wyliczenia na **niebiesko**.



Dodatkowe znaczenia n-słów ze zdań 5b, 6b oraz 7b faktycznie nie wprowadzają negacji do zdania (np. *nikt* ma znaczenie *ktoś nieważny*) i jako nieciekawe zostały zignorowane.

Zamodelowane za to zostało zjawisko dopełniacza negacji. Można to robić automatycznie (jeżeli wymaganym przypadkiem jest biernik, zamień go na dopełniacz), lecz ze względu na prostotę rozwiązania leksykon posiada oddzielne wpisy dla czasownika i jego negacji.

3 Semantyka

Dla części słów z leksykonu (reszta okazała się być dokładnym powtórzeniem) podano semantykę w postaci formuł λ FOL. Dla słów wprowadzających negację w kontekście podmiotu znaczenie było analogiczne do znaczenia słów takich jak *każdy*, *wszystko* i podobnych, np.

Bartek nie dał książki.	$\neg\text{Dać}(\text{Bartek}, \text{Książka})$
Nikt nie dał książki.	$\forall x. \neg\text{Dać}(x, \text{Książka})$
Wszyscy nie dali książki.	$\forall x. \neg\text{Dać}(x, \text{Książka})$

W kontekście dopełnienia są dwie wersje, w zależności od tego, czy uznamy, że znak negacji obejmuje kwantyfikator, czy nie.

Bartek nie dał książki.	$\neg\text{Dać}(\text{Bartek}, \text{Książka})$	$\neg\text{Dać}(\text{Bartek}, \text{Książka})$
Bartek nie dał niczego.	$\forall x. \neg\text{Dać}(\text{Bartek}, x)$	$\neg\exists x. \text{Dać}(\text{Bartek}, x)$
Bartek nie dał wszystkiego.	$\exists x. \neg\text{Dać}(\text{Bartek}, x)$	$\neg\forall x. \text{Dać}(\text{Bartek}, x)$

Rozwiązania wydają się równoważne, ale ze względu na semantykę zdań postaci „A wskazuje/pokazuje/opisuje B” wybrano podejście, w którym znak negacji obejmuje kwantyfikator dopełnienia. W poniższym przykładzie przedstawiono jedynie intuicję, nie są to prawidłowe formuły λFOL .

- Dowody wskazują na nieprzestrzeganie niczego.
 - Wskazywać(Dowody, $\forall x. \neg\text{przestrzegać}(x)$)
 - Wskazywać(Dowody, $\neg\exists x. \text{przestrzegać}(x)$)
- Dowody nie wskazują na przestrzeganie niczego.
 - $\neg\text{Wskazywać}(\text{Dowody}, \forall x. \text{przestrzegać}(x))$ źle
 - $\neg\text{Wskazywać}(\text{Dowody}, \exists x. \text{przestrzegać}(x))$ dobrze
- Dowody nie wskazują na przestrzeganie wszystkiego.
 - $\neg\text{Wskazywać}(\text{Dowody}, \forall x. \text{przestrzegać}(x))$ źle
 - $\neg\text{Wskazywać}(\text{Dowody}, \exists x. \text{przestrzegać}(x))$ dobrze

Słowo *zaden* i formy pokrewne zachowują się analogicznie do słów *nikt* czy *nic*, a mianowicie

słowo	kontekst podmiotu	kontekst dopełnienia
nikt	$\lambda k. \forall x. k(x)$	$\lambda k. \exists x. k(x)$
nic	$\lambda k. \forall x. k(x)$	$\lambda k. \exists x. k(x)$
żadna Marysia	$\lambda k. \forall x. \text{Marysia}(x) \Rightarrow k(x)$	$\lambda k. \exists x. \text{Marysia}(x) \wedge k(x)$
zaden	$\lambda m. \lambda k. \forall x. m(x) \Rightarrow k(x)$	$\lambda m. \lambda k. \exists x. m(x) \wedge k(x)$

W ostatnim wierszu tabeli występuje aplikacja $m(x)$, z czego wynika, że wcześniejsze znaczenie frazy rzeczownikowej *Marysia* nie może być konkretną wartością. Ponadto mogą wyniknąć problemy, gdybyśmy chcieli w przyszłości uwzględnić przetwarzanie wyrażeń w stylu *zaden ktoś* lub *zaden nikt* (niezależnie od tego, co one miałyby znaczyć). W związku z tym, frazy rzeczownikowe zostały wzbogacone o dodatkowe znaczenie, znaczenie predykatowe — jest ono predykatem, który jest prawdziwy tylko i wyłącznie dla obiektów opisywanych przez to wyrażenie. W ten sposób słowu *zaden* można przypisać następujące znaczenia (odpowiednio podmiotowe i dopełnieniowe):

$$\lambda m. \lambda k. \forall x. \text{znaczeniePodmiotowePredykatowe}(m)(x) \Rightarrow k(x)$$

$$\lambda m. \lambda k. \exists x. \text{znaczenieDopelnieniowePredykatowe}(m)(x) \wedge k(x)$$

Kluczowym elementem przy opracowywaniu semantyki było zapewnienie, aby pomimo wielu fraz wprowadzających negację, w końcowym zdaniu negacja pojawiła się tylko raz. Wynik ten uzyskiwany jest poprzez odpowiednie znaczenia czasowników. Ponieważ słowa wprowadzające negację implikują negację któregoś czasownika (lub słów pochodnych), znaczenie tego właśnie wyrazu pozwoli na odpowiednie zarządzanie negacją. Zastosowaną strategię najłatwiej wyjaśnić na przykładzie: „Żaden Adam nie pisał niczego używając pióra”.

- Żaden Adam nie pisał niczego używając pióra.
 $\forall x. \text{Adam}(x) \Rightarrow \neg \exists y. \text{Pisać}(x, y, \text{Pióro})$
 - żaden Adam
 znaczenie podmiotowe: $\lambda k. \forall x. \text{Adam}(x) \Rightarrow k(x)$
 znaczenie dopełnieniowe: $\lambda k. \exists x. \text{Adam}(x) \wedge k(x)$
 - żaden
 zn. podmiotowe: $\lambda m. \lambda k. \forall x. \text{znPodmiotowePredykatowe}(m)(x) \Rightarrow k(x)$
 zn. dopełnieniowe: $\lambda m. \lambda k. \exists x. \text{znDopełnieniowePredykatowe}(m)(x) \wedge k(x)$
 - Adam
 znaczenie zwykłe: $\lambda k. k(\text{Adam})$
 znaczenie predykatowe $\lambda x. \text{Adam}(x)$
 - nie pisał niczego używając pióra
 $\lambda a. \text{znaczeniePodmiotoweZwykłe}(a)(\lambda a'. \neg \exists y. \text{Pisać}(a', y, \text{Pióro}))$
 - nie pisał
 $\lambda n. \lambda o. \lambda a.$
 $\text{znaczeniePodmiotoweZwykłe}(a)(\lambda a'.$
 $\quad \neg \text{znaczenieDopełnienioweZwykłe}(n)(\lambda n'.$
 $\quad \quad \text{znaczenieDopełnienioweZwykłe}(o)(\lambda o'.$
 $\quad \quad \quad \text{Pisać}(a', o', n')$
 $\quad \quad \quad)$
 $\quad \quad)$
 $\quad)$
 $)$
 - niczego
 znaczenie podmiotowe: $\lambda k. \forall x. k(x)$
 znaczenie dopełnieniowe: $\lambda k. \exists x. k(x)$
 - używając pióra
 znaczenie zwykłe: $\lambda k. k(\text{Pióro})$
 znaczenie predykatowe: $\lambda x. \text{Pióro}(x)$
 (w tym przypadku imiesłów *używając* zachowuje się podobnie do przyimka i jest usuwany wcześniej, przy parsowaniu przez reguły syntaktyczne)

Zastosowana strategia działa nawet w przypadku czasowników podwójnie przechodnich:

Marek nie dał Marysi Kwiatów	$\neg \text{Dać}(\text{Marek}, \text{Marysia}, \text{Kwiaty})$
Marek nie dał nikomu Kwiatów	$\neg \exists x. \text{Dać}(\text{Marek}, x, \text{Kwiaty})$
Marek nie dał Marysi niczego	$\neg \exists y. \text{Dać}(\text{Marek}, \text{Marysia}, y)$
Marek nie dał nikomu niczego	$\neg \exists x. \exists y. \text{Dać}(\text{Marek}, x, y)$

Niestety, zaproponowane podejście nie sprawdza się uniwersalnie. Po dodaniu słów w stylu *wszyscy*, zdanie *Marek nie dał wszystkim niczego* jest rozumiane nieregularnie (pojawia się dodatkowa negacja dla słowa *niczego*) jako

$$\neg\forall x. \neg\exists y. \text{Dać}(\text{Marek}, x, y)$$

zamiast przewidywanego

$$\neg\forall x. \exists y. \text{Dać}(\text{Marek}, x, y)$$

W prawdzie pozostałe osiem przypadków jest rozpoznawanych poprawnie, jednak ten jeden wymyka się spod kontroli i w tej chwili autor nie wie, czy da się go rozpatrzyć w spójny, jednolity sposób.

4 Podsumowanie

W artykule opisano najważniejsze zjawiska dotyczące uzgodnień i semantyki słów wprowadzających negację w języku polskim. Zaprezentowano szkic gramatyki zaimplementowanej w środowisku Separ[1] oraz budowę wyrażeń semantycznych otrzymywanych przez parser. Jako reprezentatywne wybrano słowa *nikt*, *nic* oraz *żaden*. Projekt uwzględnia propagację negacji, a także występujące dla niej bariery. Ponadto dodano obsługę zjawiska dopełniacza negacji, czyli zmianę przypadków z biernika na dopełniacz dla zanegowanych czasowników. Jedynym nierozwiązanym problemem jest nieregularne zachowanie się negacji w znaczeniu zdań postaci „ α (czasownikDwuprzechodni) wszyscy nic”. Inne możliwe rozszerzenia to m.in. dodanie alternatywnych znaczeń dla *nikt*, *nic*, etc., obsługa pytań, dodanie semantyki dla wyrażeń czasowych wprowadzających negację (*nigdy*, *żadnego wieczoru*) oraz dozwole nie *nic* w zdaniach postaci *Nic nie widzę* (zamiast *Niczego nie widzę*).

Literatura

- [1] D. LENIOWSKI, *Automatyczna analiza semantyczna tekstu w języku naturalnym, studium przypadku*, Master's thesis, Uniwersytet Warszawski, Warszawa, 2010.
- [2] A. PRZĘPIÓRKOWSKI AND A. KUPŚĆ, *Negative concord in Polish*, Tech. Rep. 828, Institute of Computer Science, Polish Academy of Sciences, feb 1997.