

Budowa dziedzinowej reprezentacji wiedzy
z pomocą rozszerzonej metody
ramowej na podstawie korpusu umów najmu
w języku polskim, angielskim i niemieckim

Agnieszka Pluwak
Instytut Slawistyki, PAN

Rama (frame) i FrameNet

Fall_asleep

[Lexical Unit Index](#)

Definition:

A **Sleeper** goes from wakefulness to an altered state of consciousness where they are largely unaware of their environment. The LUs of this frame are differentiated by how quickly and deeply the state is entered and what causes the change (especially whether it comes about naturally as a result of sleepiness or is brought about by injury).

I DOZED OFF at about 1 am.

FEs:

Core:

Sleeper [Slp] The **Sleeper** is the entity that falls asleep.

Non-Core:

Place [Place] This FE identifies the **Place** where the event occurs.

Semantic Type:

Locative_relation

Time [Time] This FE identifies the **Time** where the event occurs.

Semantic Type: Time

Frame-frame Relations:

Inherits from:

Is Inherited by:

Perspective on:

Is Perspectivized in:

Uses:

Is Used by:

Subframe of: [Sleep_wake_cycle](#)

Has Subframe(s):

Precedes: [Sleep](#)

Is Preceded by: [Being_aware](#)

Is Inchoative of:

Is Causative of:

See also:

Lexical Units:

black out.v, doze off.v, drift off.v, faint.v, konk out.v, nod off.v, pass out.v, zonk out.v

Anotacja FrameNet

Frame Element	Core Type
Compensation	Peripheral
Contract_basis	Peripheral
Employee	Core
Employer	Core
Field	Core
Instrument	Peripheral
Manner	Peripheral
Means	Peripheral
Period_of_iterations	Extra-Thematic
Place	Peripheral
Position	Core
Purpose	Peripheral
Task	Core
Time	Peripheral

[Turn Colors Off](#)

- 429-s20-rcoll-firm
 - Viscount Althorp claims a national newspaper HIRED a firm of private detectives to follow them round-the-clock .
 - The 31-year-old Canadian sprinter HIRED a Toronto law firm to issue a denial of the story in the Toronto Star .
 - THE Department of the Environment has HIRED the security firm dropped by Belfast City Council to guard an important depot near Enniskillen .
- 429-s20-rcoll-on
 - You can HIRE an architect or get advice on digging drains .
 - Thanks to Sybille two local women were HIRED to do extra cooking and cleaning on these occasions , when it seemed the concept of the extended family had turned the Ashleys into hoteliers . CNI
 - She 's HIRED a number of people on my say-so . "DNI
- 500-s20-vpto
 - There was a woman HIRED to do what she felt was a wife 's job . CNI

Zastosowania w lingwistyce komputerowej



THINK

ΣΚΕΨΟΥ

DEN

P

omīs

चिए

\$2,000

Ken

\$13,400

WATSON

\$5,000

BRAD





Albinism		98%
Albino		10%
Porphyria		7%

Pomysł na projekt:

- chęć nauczenia się opisu metodą ramową,
- praca jako analityk umów najmu (pomysł na zwiększenie wydajności pracy analityka o około 30% przy pomocy automatyzacji wybranych zadań),
- poszukiwanie materiału (korpusu) do pracy doktorskiej,
- mnożące się pytania nt. semantyki ramowej i projektu FrameNet.

Założenia projektu

1. zbudować korpus umów najmu przynajmniej w 3 znanych mi językach,
2. zdefiniować ramy odpowiadające informacjom, poszukiwanym przez analityka,
3. przyjąć, że ramy są na tak ogólnym poziomie wiedzy encyklopedycznej, że mogą być językiem pośrednikiem (metajęzyk, tertium comparationis) na poziomie schematu,
4. wykorzystać podręcznik anotatora do budowy ram najpierw w języku polskim (budowa ram semantycznych według Rupperhofera i innych), a potem zestawić je z ramami FrameNet,

Założenia projektu 2

1. przyjąć (choćby za Pado, 2007), że poziom składniowy jest specyficzny dla każdego języka i opis składniowy musi zostać opisany inaczej niż we FrameNecie (znaczniki parsera Gobio),
2. wybór tekstów ekwiwalentnych, a nie tłumaczeń (ekwiwalenty leksykalne używane w danym kontekście, zamiast tłumaczenie prawa innego kraju z pomocą terminologii drugiego języka),
3. badanie ram w kontekście (za Hansem Boasem, 2005), zamiast założenia, że ramy FN będą miały podobną realizację powierzchniową w każdym stylu, gatunku i tematyce.

Korpus umów najmu

- 100 tekstów (około 30 polskich, angielskich i niemieckich umów najmu powierzchni mieszkalnych i biurowych),
- autentyczne umowy lub wzory,
- dobór tekstów różnej długości (od 1 do 20 stron),
- pochodzenie umów: USA (różne stany), Wielka Brytania (różne regiony), kilka tekstów z Kanady.

Procedura Analizy

- Wybór około 30 jednostek leksykalnych, powtarzających się w umowach najmu i powiązanych z informacjami istotnymi dla analityka najmu, np. *making agreement, lease, give notice of termination, etc.* i na podstawie przykładów z NKJP oraz przykładów korpusowych, zdefiniowanie dla nich około 30 ram,
- przeprowadzenie dla każdej z ram porównania ich realizacji w kilku językach,
- porównanie struktury ramowej z ramami FN (definicja, elementy, realizacja).

Metody analizy ramowej FN

- Typowa analiza FrameNetu to analiza oddolna (bottom-up), która polega na anotacji (kilkudziesięciu) zdań zawierających leksemy należące do opisywanej ramy.
- Cechy charakterystyczne: opis na poziomie zdania, akcent położony na opis polisemii leksemów i wzorów walencyjnych.
- Analiza dziedzinowa jest analizą typu odgórnego (top-down), w której poszukuje się realizacji elementów ramy w tekście (slot filling) i anotuje często cały tekst.

Rozszerzenie metody analizy

Dlaczego rozszerzona?

- czasem trzeba wyjść poza schemat walencyjny lub poziom zdania na poziom akapitu, aby znaleźć realizację ramy,
- elementy niejądrowe są bardzo ważne, bo zawierają poszukiwane przez analityka informacje (w szczególności okoliczniki miejsca, czasu, etc.), dlatego trzeba dokładnie je zdefiniować,
- trzeba wziąć pod uwagę styl i analizowaną domenę
- i określić wszystkie sposoby realizacji danej ramy (czasownikowe, rzeczownikowe, przymiotnikowe, itd. itp.) w danej domenie tematycznej i danym korpusie, a nie wybrane leksemy.
- dzięki określeniu relacji między ramami możemy zbudować małą dziedzinową reprezentację wiedzy, która pomoże nam wpisać w nią leksemy z różnych języków.

ELEMENTY RAMY - ZAWARCIE UMOWY

JEDNOSTKI SEMANTYCZNE

*an agreement is made (make an agreement)
an agreement is entered into
(enter into an agreement)
an agreement is*

STRONY

Fraza przyimkowa: *between AGENT 1 and AGENT 2
by and between.....and*

PRZEDMIOT

Fraza nominalna z przydawkami rzeczownikowymi, przymiotnymi lub bezokolicznikiem:
*agreement to lease an unfurnished apartment
(Texas residential) lease agreement
(apartment/ residential) lease
(assured shorthold) tenancy agreement, residential tenancy agreement
basic rental agreement*

PUNKT W CZASIE

Fraza przyimkowa, okolicznik czasu: *on 13th January 2012, on the day of
13th January 2012, the__ day of_____ 20__,
(formularz: Lease Term: From: to:.....)*

ANOTACJA PRZYKŁADÓW - ROZWIĄZANIE UMOWY

{Bez podania przyczyn | SPOSÓB/MANNER}, {Pekpol | STRONA 1/ PARTY1}
{rozwiązał | TGT} {umowę | ZOBOWIĄZANIE/OBLIGATION}{z Polskim
Konsorcjum Finansowym | STRONA 2/ PARTY 2} {dotyczącą kupna akcji... |
PRZEDMIOT/SUBJECT}

{Bez podania przyczyn | FPRZY/OKOL.SPOS}, {Pekpol | R/PODM} {rozwiązał | TGT}
{umowę | FR/DOP}{z Polskim Konsorcjum Finansowym | FPRZY/DOP.DAL}
{dotyczącą kupna akcji... | FC/PRZYD}

Wyniki i efekty

Szczegółowo opracowano elementy i realizacje ram (20):

POROZUMIENIA, ZOBOWIĄZANIA, ZAWARCIA UMOWY, PLANOWANIA TERMINU SPOTKANIA, TRWANIA, SAMOISTNEGO i AGENTYWNEGO ZAKOŃCZENIA umowy (w tym: WYPOWIEDZENIA i ROZWIĄZANIA UMOWY), SAMOISTNEGO i AGENTYWNEGO PRZEDŁUŻENIA UMOWY, UPRAWNIENIA, OBOWIĄZKU, ODPOWIEDZIALNOŚCI, NAJMU, LOKALIZACJI, PŁACENIA, USTALANIA, TRANSAKCJI HANDLOWEJ i WARTOŚCI.

Ponadto mniej szczegółowo opisano ramy (12): REPREZENTACJI, PODPISANIA UMOWY, ZAWIERANIA, OBOWIĄZYWANIA, ODNIESIENIA, PRZYCZYNOWOŚCI, KOMUNIKACJI, WYPOWIEDZI, TRANSPORTU, LOKALIZACJI, PRZEJŚCIA ZMIANY oraz PRZELEWU.

Ramy FN

Ramy FrameNet, przydatne w definiowaniu moich ram (15):

MAKING AGREEMENT ON ACTION, BEING IN AGREEMENT ON ACTION, TRANSPORT, CAUSE TO CONTINUE, UNDERGO CHANGE, CHANGE EVENT DURATION, PERMITTING, GRANT PERMISSION, POSSIBILITY, BEING OBLIGATED, RESPONSIBILITY, BEING LOCATED, COMMERCE PAY, DECIDING, AMOUNTING TO.

Z tego możliwe do przyjęcia bez z lub z niewielką liczbą uzupełnień/z niewielkim dostosowaniem: 8.

Podsumowanie opisu

Liczba zdefiniowanych przez mnie ogółem ram	Liczba ram FN przydatnych przy definiowaniu ram	Liczba ram FN możliwych do użycia w badanym kontekście	Średnia liczba wariantów leksykalnych wyrażenia danej ramy w każdym z języków (x 3)
32	15	8	3

Podsumowanie opisu

**Liczba ram, w
których
wystąpiła
realizacja poza-
czasownikowa**

80-90%

**Liczba ram,
w których
wystąpiła
realizacja
w zdaniach
złożonych**

90-100%

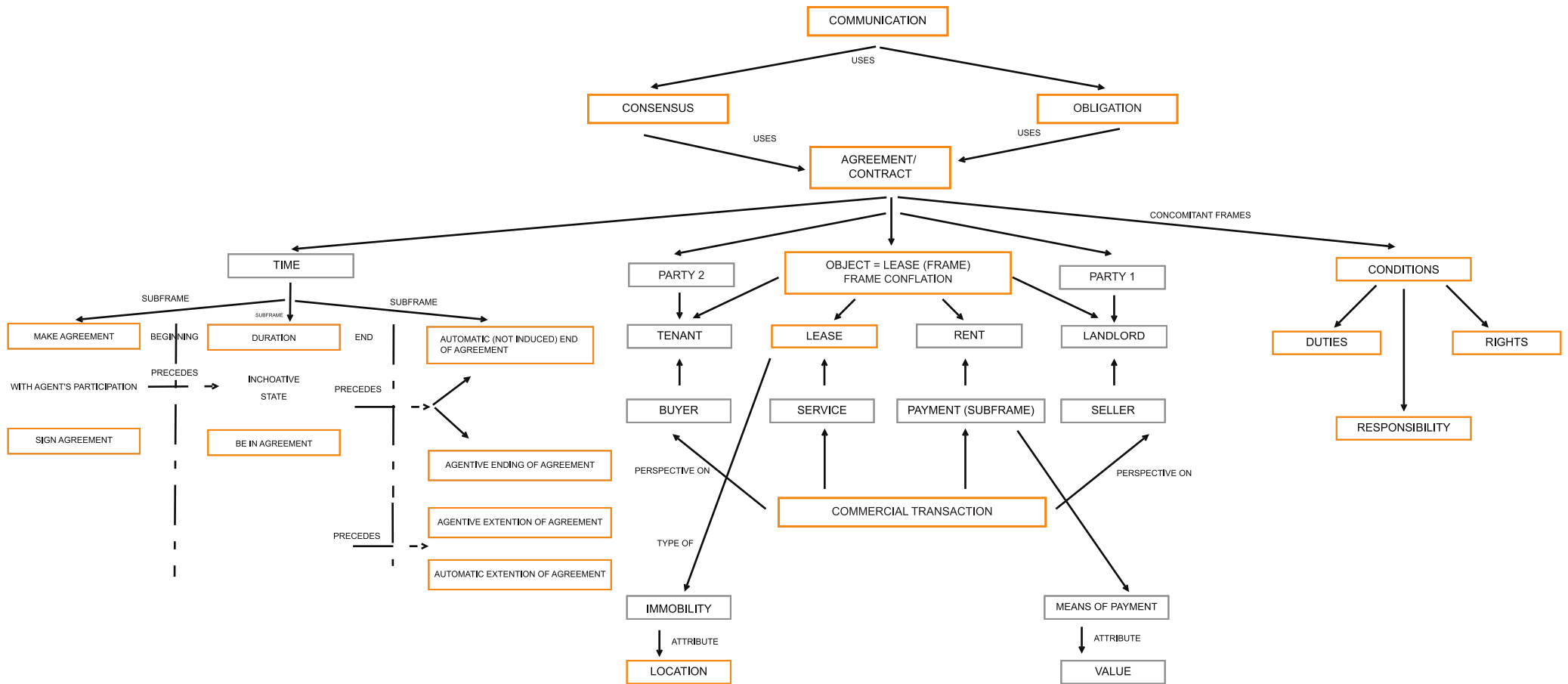
**Liczba ram,
w których
wystąpiła
realizacja
wykraczająca
poza poziom
zdania**

10-15%

**Liczba ram, w
których
konieczne było
określenie
relacji między
nimi**

50-60%

Relacje między ramami i dziedzinowa prezentacja wiedzy



Wnioski

- Opracowano model porównawczej analizy dziedzinowej metodą semantyki ramowej, w którym uwzględnić można realizacje leksykalne w różnych językach w obrębie jednego modelu reprezentacji wiedzy,
- Model ten realizuje postulat Boasa o badanie ram w danym kontekście,
- Analiza na podstawie NKJP pozwala pokazać bogactwo realizacji elementów ramowych, a analiza dziedzinowa – zbiór jednostek leksykalnych realizujących ramę w danej domenie słownictwa.
- Do budowy dziedzinowej reprezentacji wiedzy wystarczy definicja około 30 ram, jednak nie można do końca liczyć na to, że więcej niż $\frac{1}{4}$ z nich będzie już gotowa we FrameNecie do zastosowania przez nas. Należy raczej nastawić się na modyfikację istniejących i/lub budowę własnych ram.

Wnioski 2

- Dzięki określeniu relacji między ramami, możemy odnaleźć niektóre informacje dotyczące umowy, np. jej trwania przy rzeczownikach oznaczających najem jako metonimię umowy najmu.
- Dzięki opisowi ram zagnieżdżonych nie pominiemy elementów zawartych w przydawkach lub okolicznikach, np. *umowa została zawarta z możliwością jej rozwiązania w danym czasie*.
- Warto przyjrzeć się wszystkim realizacjom ramowym, bo niektóre z nich mogą wystąpić poza głównym schematem walencyjnym lub poziomem zdania, np. *data i miejsce podpisania umowy albo czas trwania umowy wyrażony okolicznikiem (zawarta na okres od ...do...) i inne*.

Systemy do ekstrakcji danych z tekstów umów

<http://prismlegal.com/improving-contract-management-analysis/>

<https://kirasystems.com/how-it-works/contract-analysis>

<https://www.lawgeex.com/pricing/>

<http://ebrevia.com/ebrevia-contract-manager/>

Pytania do Państwa:

1. Czy warto jeszcze rozmawiać o semantyce ramowej w obliczu dynamicznego rozwoju sieci neuronowych?
2. Jaką metodę analizy ról semantycznych Państwo wybieracie w swojej pracy?
3. Jak najlepiej wykorzystać moją analizę? Udzielić licencji na korpus w 3 językach, aby ktoś mógł na tej podstawie zbudować system do ekstrakcji danych z umów najmu ?

Bardzo dziękuję za uwagę!

agnieszka.pluwak@gmail.com

www.ispan.waw.pl