



Universal Discourse: w stronę wielojęzycznego modelu relacji dyskursywnych

Maciej Ogrodniczuk

Anna Latusek, Alina Wróblewska, Bartosz Żuk

Instytut Podstaw Informatyki PAN

Seminarium IPI | ZIL

Warszawa, 18 maja 2026

Grant NCN 2023/50/A/HS2/00559



NARODOWE CENTRUM NAUKI

Wprowadzenie

Relacje dyskursywne w pigułce

Relacje dyskursywne to związki między **sytuacjami** (zdarzeniami, stanami, faktami, twierdzeniami, aktami mowy) wyrażone w tekście w sposób jawny lub niejawny.

Argumenty relacji niesymetrycznych są nazwane — np. dla relacji przyczynowo-skutkowej CAUSE jeden argument to REASON (przyczyna), a drugi to RESULT (skutek).

Relacje jawne (eksplicytne) są wyraźnie sygnalizowane za pomocą **znaczników dyskursywnych** (np. *jednak, i, nie tylko... ale także*).

Relacje niejawne (implicitne) dają się wyróżnić mimo braku znacznika.

Znaczniki są wieloznaczne!

Jeden znacznik może odpowiadać różnym relacjom, np.

- Piotr spóźnił się na spotkanie, ponieważ utknął w korku. → CAUSE
- Jana nie ma dziś w pracy, ponieważ wysłał mi wiadomość, że jest chory. → EVIDENCE
- Pytam cię, co robisz wieczorem, ponieważ w kinie grają dobry film. → EXPLANATION

Dlaczego ten temat jest ważny?

- relacje dyskursywne zapewniają spójność tekstu na poziomie przekraczającym granice zdań,
- mimo semantycznej zgodności powstałych dotąd formalizmów opisujących te relacje, nie przyjęto dotąd ich jednolitej reprezentacji,
- temat rozwija moje wcześniejsze prace nad relacjami referencyjnymi,
- modele LLM wciąż mają trudności z rozpoznawaniem relacji tego rodzaju.

Kilka najpopularniejszych formalizmów:

- Rhetorical Structure Theory (RST)
- Hobbs' Theory of Discourse Coherence (HTDC)
- Segmented Discourse Representation Theory (SDRT)
- Cognitive Approach of Coherence Relations (CCR)
- Penn Discourse Treebank (PDTB)

Główne różnice dotyczą reprezentacji relacji (drzewa czy grafy?), podejścia teoretycznego (ważniejsza jest intencja retoryczna czy znaczniki?) oraz elastyczności co do wyznaczania granic argumentów.

Wybrane korpusy dyskursu

Nazwa korpusu	Język	Formalizm
Basque RST-DT	baskijski	RST
Prague Discourse Treebank	czeski	PDTB
Dutch RST-DT	niderlandzki	RST
Graphbank	angielski	HTDC
Penn Discourse Treebank	angielski	PDTB
GUM	angielski	RST + PDTB
ANNODIS	francuski	SDRT
MAZ Corpus	niemiecki	RST
Spanish RST-DT	hiszpański	RST
Polish Discourse Corpus	polski	ISO
Turkish Discourse Bank	turecki	PDTB
Copenhagen Dependency Treebanks	wielojęzyczny	CDT
TED Multilingual Discourse Bank	wielojęzyczny	PDTB

TED-MDB:

- korpus transkrypcji wystąpień TED w siedmiu językach (EN, DE, LT, PL, PR, RU, TR),
- anotacja zgodnie z formalizmem PDTB:
 - znaczniki jawne i niejawne,
 - powiązania tylko pomiędzy sąsiadującymi fragmentami tekstu,
 - dowolnie długie argumenty relacji,
- ok. 600 relacji / język, IAA 0,8,
- wstępne wnioski: język portugalski ma najwięcej relacji niejawnych, polski – najwięcej relacji referencyjnych (EntRel).

HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC.)



<https://xkcd.com/927/> + 

Standard ISO 24617-8:2016

„W specyfikacji DR-core przyjęto **podejście przedteoretyczne, polegające na niskopoziomym opisie relacji dyskursywnych** przy założeniu, że poszczególne relacje można będzie dodatkowo opisywać i łączyć w struktury drzew lub grafów wyższego poziomu, w zależności od osobistych preferencji teoretycznych. Z punktu widzenia interoperacyjności anotacja niskiego poziomu może również służyć jako reprezentacja pośrednia przy porównywaniu anotacji różnych zasobów opartych na różnych teoriach.”

INTERNATIONAL
STANDARD

ISO
24617-8

First edition
2016-12-15

**Language resource management —
Semantic annotation framework
(SemAF) —**

Part 8:
**Semantic relations in discourse, core
annotation schema (DR-core)**

*Gestion des ressources langagières — Cadre d'annotation sémantique
(SemAF) —*

*Partie 8: Relations sémantiques dans le discours, schéma d'annotation
de base (DR-core)*



Reference number
ISO 24617-8:2016(E)

© ISO 2016

Lista relaciji ISO 24617-8

	ISO 24617-8	Symmetry	Relations and argument role definitions
1.	Cause	Asymmetric	Arg2 is an explanation for Arg1.
2.	Condition	Asymmetric	Arg2 is an unrealized situation which, when realized, would lead to Arg1.
3.	Negative condition	Asymmetric	Arg2 is an unrealized situation which, when "not" realized, would lead to Arg1.
4.	Purpose	Asymmetric	Arg2 is the goal or purpose of the situation described by Arg1.
5.	Manner	Asymmetric	Arg2 describes how Arg1 comes about or occurs.
6.	Concession	Asymmetric	An expected causal relation between Arg1 and \neg Arg2 is cancelled or denied by Arg2.
7.	Contrast	Symmetric	One or more differences between Arg1 and Arg2 are highlighted with respect to what each predicates as a whole or to some entities they mention.
8.	Exception	Asymmetric	Arg2 indicates one or more circumstances in which the situation(s) described by Arg1 does not hold.
9.	Similarity	Symmetric	One or more similarities between Arg1 and Arg2 are highlighted with respect to what each predicates as a whole or to some entities they mention.
10.	Substitution	Asymmetric	Arg1 and Arg2 are alternatives, with Arg2 being the favoured or chosen alternative.
11.	Conjunction	Symmetric	Arg1 and Arg2 bear the same relation to some other situation evoked in discourse. Their conjunction indicates that they both hold with respect to that situation.

Lista relaciji ISO 24617-8

	ISO 24617-8	Symmetry	Relations and argument role definitions
12.	Disjunction	Symmetric	Arg1 and Arg2 bear the same relation to some other situation evoked in the discourse, explicitly or implicitly. Their disjunction indicates that they are alternatives with respect to that situation, with the disjunction being non-exclusive so that both Arg1 and Arg2 may hold.
13.	Exemplification	Asymmetric	Arg1 describes a set of situations; Arg2 describes an element of that set.
14.	Elaboration	Asymmetric	Arg1 and Arg2 are the same situation, but Arg2 contains more detail.
15.	Restatement	Symmetric	Arg1 and Arg2 describe the same situation, but from different perspectives.
16.	Synchrony	Symmetric	Some degree of temporal overlap exists between Arg1 and Arg2. All forms of overlap are included.
17.	Asynchrony	Asymmetric	Arg1 temporally precedes Arg2.
18.	Expansion	Asymmetric	Arg2 is a situation involving some entity/entities in Arg1, expanding the narrative forward of which Arg1 is a part, or expanding on the setting relevant for interpreting Arg1. The Arg1 and Arg2 situations are distinct.
19.	Functional dependence	Asymmetric	Arg2 is a dialogue act with a responsive communicative function; Arg1 is the dialogue act(s) that Arg2 responds to.
20.	Feedback dependence	Asymmetric	Arg2 is a feedback act that provides or elicits information about the understanding or evaluation by one of the dialogue participants of Arg1.

Nazwy argumentów relacji ISO 24617-8

	Discourse relation	Arg1	Arg2
1.	Cause	Result	Reason
2.	Condition	Consequent	Antecedent
3.	Negative Condition	Consequent	Negated-antecedent
4.	Purpose	Enablement	Goal
5.	Manner	Achievement	Means
6.	Concession	Expectation-raiser	Expectation-denier
7.	Exception	Regular	Exclusion
8.	Substitution	Disfavoured-alternative	Favoured-alternative
9.	Exemplification	Set	Instance
10.	Elaboration	Broad	Specific
11.	Asynchrony	Before	After
12.	Expansion	Narrative	Expander
13.	Functional dependence	Antecedent-act	Dependent-act
14.	Feedback dependence	Feedback-scope	Feedback-act

Mapowanie formalizmów

DR-core		RST	RST Treebank			
Cause		<i>Vol. cause, Non-vol. cause, Vol. result, Non-vol. result, Evidence, Justify</i>	<i>Cause, Consequence, Result, Evidence, Explanation-argumentative, Reason</i>			
Condi	DR-core	HTDC		GraphBank		
Negat condit	Cause	<i>Explanation, Cause, Evaluation</i>		<i>Cause-effect, Elaboration</i>		
Purpc	Condition	DR-core	PDTB		CCR/DiscAn	
Mann	Negative condition	Cause	Reason, Result, Justification		Causal-Semantic-Basic-Positive, Causal-Semantic-NonBasic-Positive, Causal-Pragmatic-Basic-Positive,	
Conce	Purpose					
Contr	Manner					
Excep	Concession	Conditio	DR-core	SDRT	DISCOR	ANNODIS
Simila	Contrast		Cause	<i>Explanation, Result</i>	<i>Explanation, Result</i>	<i>Explanation, Result</i>
Subst	Exception		Condition	<i>Consequence</i>	<i>Consequence</i>	<i>Conditional</i>
Conju	Similarity		Negative condition	<i>Consequence</i>	<i>Consequence</i>	<i>Conditional</i>
Disjui	Substitutio		Purpose	<i>Explanation</i>	<i>Explanation</i>	<i>Goal</i>
Exemj	Conjunctio		Manner	<i>Elaboration</i>	<i>Elaboration</i>	<i>Elaboration</i>
Elabo	Disjunctio		Concession	<i>Contrast</i>	<i>Contrast</i>	<i>Contrast</i>
	Exemplificatio		Contrast	<i>Contrast</i>	<i>Contrast</i>	<i>Contrast</i>
			Purpose	Exception	—	—
Restatement			Similarity	<i>Parallel</i>	<i>Parallel</i>	<i>Parallel</i>
	Manner	Substitution	—	—	—	
		Conjunction	<i>Continuation</i>	<i>Continuation</i>	<i>Continuation</i>	

Projekt ma na celu:

- stworzenie **ujednoliconego, wielojęzycznego modelu dyskursu opartego na korpusach,**
- w oparciu o **normę ISO 24617-8,**
- wyposażonego w model formalny w postaci **ontologii relacji dyskursywnych,** typów argumentów oraz znaczników dyskursywnych,
- oraz **zharmonizowanie zbioru istniejących korpusów dyskursu** poprzez konwersję i reanotację,
- a następnie wdrożenie zestawu **prototypowych parserów dyskursu** w celu komputerowego przetestowania modelu.

Czy to się uda?

Powinno:

- ponieważ działamy metodą oddolną:
 - nie anotujemy nowych danych zgodnie z nową teorią,
 - lecz harmonizujemy istniejące dane w celu stworzenia wspólnego opisu,
- ponieważ ta metoda sprawdziła się już wcześniej:
 - Universal Dependencies,
 - CorefUD,
- ponieważ nie zaczynamy od zera:
 - mamy do dyspozycji wiele istniejących korpusów z anotacjami dyskursywnymi,
 - opieramy się na normie ISO 24617-8.

Segmentacja

Składniowe podejście do segmentacji dyskursu

- Segmentacja jest pierwszym etapem parsowania dyskursu, a jej wynikiem są **segmenty dyskursywne** [discourse units, DUs].
- Motywacja – formalizmy mają różne definicje DU; segmentem wspólnym są zdania/frazy zdaniowe.
- Założenie – granice segmentów dyskursywnych pokrywają się z realizacją fraz zdaniowych, których podstawą jest **struktura predykatowo-argumentowa** (uniwersalium językowe).
- Implementacja – Universal Dependencies (UD):
 - zapewnia opis składniowy uniwersalny dla różnych języków,
 - koduje relacje predykat–argument,
 - umożliwia segmentację wielojęzyczną.

Reguły segmentacji oparte na UD (1)

- Granice DU są identyfikowane na podstawie relacji UD.
- Podstawowe typy relacji UD odpowiadające frazom zdaniowym:
 - 1.1. **root** – zdanie główne,
 - 1.2. **advcl** – zdanie okolicznikowe,
 - 2.1. **ccomp** – dopełnienie zdaniowe,
 - 2.2. **acl** – przydawka zdaniowa,
 - 2.3. **csbj** – podmiot zdaniowy,
 - 2.4. **SCONJ** + **xcomp** – complementizer + dopełnieniowa fraza bezokolicznikowa.

#1: ⟨You're so stupid_{root}⟩ ⟨thinking_{advcl}⟩ ⟨I spent_{ccomp} the night.⟩

Reguły segmentacji oparte na UD (2)

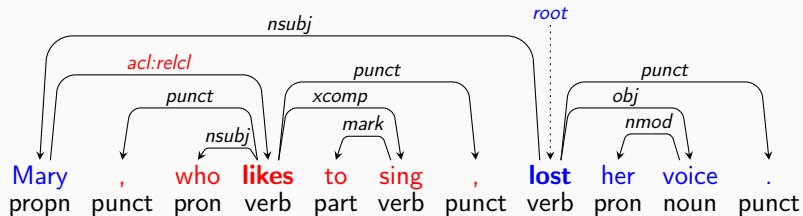
- Dodatkowe reguły wydzielenia osobnych DUs:
 3. *parataxis* – zdanie sąsiadujące,
 4. *acl* / *acl:relcl* – zdanie względne,
 5. koordynacja fraz zdaniowych,
 6. komentarze i nawiasy z dodatkową treścią,
 7. reguła specyficzna dla języka, np. angielskie frazy *including...* .
- Interpunkcja sama w sobie nie stanowi wystarczającego sygnału segmentacji.

#2: ⟨This is a trend_{root}⟩ ⟨that bears_{acl:relcl} more scrutiny⟩ ⟨than it has received_{acl}.⟩

#3: ⟨As each task becomes_{advcl} more specialized,⟩ ⟨Smith noted_{parataxis},⟩ ⟨it engages_{root} less of the person.⟩

- Etap 1: identyfikacja elementów głównych segmentów dyskursywnych na podstawie:
 - typów relacji zależnościowych,
 - kategorii POS (części mowy),
 - struktury drzewa UD.
- Etap 2: wyznaczenie zakresu DU \rightarrow poddrzewo UD.
- Obsługa struktur zagnieżdżonych:
 - segment podrzędny wyłączany jest z zakresu segmentu nadrzędnego,
 - możliwe jest modelowanie segmentów nieciągłych.

Segmenter regułowy



Segmenter regułowy:

- oparty na drzewach UD,
- implementuje 7 prostych i (prawie) uniwersalnych reguł.

Segmenter regułowy vs modele głębokie

Dataset	Our			DisCut			HITS		
	P	R	F1	P	R	F1	P	R	F1
eng.rst.rstdt	87.56	87.00	87.28	97.21	98.04	97.62	96.46	97.66	97.06
eng.rst.gum	90.76	93.14	91.93	94.59	96.42	95.50	95.08	95.29	95.19
eng.sdrst.stac	87.92	90.37	89.13	95.75	94.70	95.22	96.71	95.09	95.89
eng.dep.scidtb	90.84	90.88	90.86	94.96	95.18	95.07	94.77	95.09	94.93
eng.dep.covdtb*	89.60	87.02	88.29	94.04	90.31	92.13	90.22	90.38	90.30
Mean	89.34	90.20	89.50	95.51	94.93	95.11	94.65	94.70	94.67

Segmenter regułowy w porównaniu do modeli głębokich:

- jest średnio o ~5 pp. gorsze pod względem F1 (89.5 vs. 94.7-95.1)
- osiąga najmniejszą stratę na zbiorze OOD eng.dep.covdtb*

Niespójność anotacji segmentów

Anotacje segmentów nie są **spójne!**

Zdanie ze zbioru testowego eng.rst.umuc

The parties agreed on the need for a sustainable ceasefire, the resumption of the contact group, control of border crossing points monitored by the Organization for Security and Cooperation in Europe (OSCE) and an extension of the work of the OSCE observer mission to include border monitoring activities.

Zdanie ze zbioru testowego eng.rst.oll

In case for more complex systems (e.g., Word processors, web-authoring tools, etc.) that provide a wider range of functionality, users are willing to spend longer time learning these products.

Niespójność anotacji segmentów

Anotacje segmentów nie są **spójne!**

Zdanie ze zbioru testowego eng.rst.umuc

The parties agreed on the need for a sustainable ceasefire, the resumption of the contact group, control of border crossing points monitored by the Organization for Security and Cooperation in Europe (OSCE) and an extension of the work of the OSCE observer mission to include border monitoring activities.

Zdanie ze zbioru testowego eng.rst.oll

In case for more complex systems (e.g., Word processors, web-authoring tools, etc.) that provide a wider range of functionality, users are willing to spend longer time learning these products.

Niespójność anotacji segmentów

Anotacje segmentów nie są **spójne!**

Zdanie ze zbioru testowego eng.rst.umuc

The parties agreed on the need for a sustainable ceasefire, the resumption of the contact group, control of border crossing points monitored by the Organization for Security and Cooperation in Europe (OSCE) and an extension of the work of the OSCE observer mission to include border monitoring activities.

Zdanie ze zbioru testowego eng.rst.oll

In case for more complex systems (e.g., Word processors, web-authoring tools, etc.) that provide a wider range of functionality, users are willing to spend longer time learning these products.

Skuteczność modeli głębokich

Pomimo różnic w anotacji modele głębokie osiągają wysokie wyniki.

Dataset	Qwen2.5-3B	DiscUT	HITS	SeCoRel
	F1			
eng.dep.covdtb*	91.57	92.58	91.80	90.12
eng.dep.scidtb	94.16	95.26	94.77	94.38
eng.erst.gentle*	90.36	94.09	91.43	88.28
eng.erst.gum	91.87	93.70	92.93	90.27
eng.rst.oll	83.44	91.03	87.19	83.58
eng.rst.umuc	86.54	88.95	87.17	86.61
eng.rst.rstdt	96.77	96.69	97.18	95.10
eng.rst.sts	83.36	88.14	81.62	80.92
eng.sdrst.msdc	95.52	95.81	95.37	94.76
eng.sdrst.stac	93.02	92.93	92.81	91.36
Mean	90.66	92.82	91.43	89.54

Przetrenowanie modeli głębokich

Segmentery oparte na modelach głębokich są **przetrenowane!**

Oryginalne zdanie z eng.rst.011

In case for more complex systems (e.g., Word processors, web-authoring tools, etc.) that provide a wider range of functionality, users are willing to spend longer time learning these products.

Zdanie z eng.rst.011 w stylu "dyplomatycznym"

In case for more complex international issues (e.g., Israel-Palestine, India-Pakistan, etc.) that warrant a wider range of considerations, members are willing to spend longer time debating these points.

Przetrenowanie modeli głębokich

Segmentery oparte na modelach głębokich są **przetrenowane!**

Oryginalne zdanie z eng.rst.011

In case for more complex systems (e.g., Word processors, web-authoring tools, etc.) that provide a wider range of functionality, users are willing to spend longer time learning these products.

Zdanie z eng.rst.011 w stylu "dyplomatycznym"

In case for more complex international issues (e.g., Israel-Palestine, India-Pakistan, etc.) that warrant a wider range of considerations, members are willing to spend longer time debating these points.

Przetrenowanie modeli głębokich

Segmentery oparte na modelach głębokich są **przetrenowane!**

Oryginalne zdanie z eng.rst.011

In case for more complex systems (e.g., Word processors, web-authoring tools, etc.) that provide a wider range of functionality, users are willing to spend longer time learning these products.

Zdanie z eng.rst.011 w stylu "dyplomatycznym"

In case for more complex international issues (e.g., Israel-Palestine, India-Pakistan, etc.) that warrant a wider range of considerations, members are willing to spend longer time debating these points.

Wpływ błędów parsera składniowego

Dataset	Our	DiscUT	HITS	SeCoRel
	F1			
deu.rst.pcc	88.19	95.84	92.78	94.20
deu.rst.pcc.corrected	99.70 (↑11.51)	—	—	—

Po ręcznej poprawie drzew składniowych:

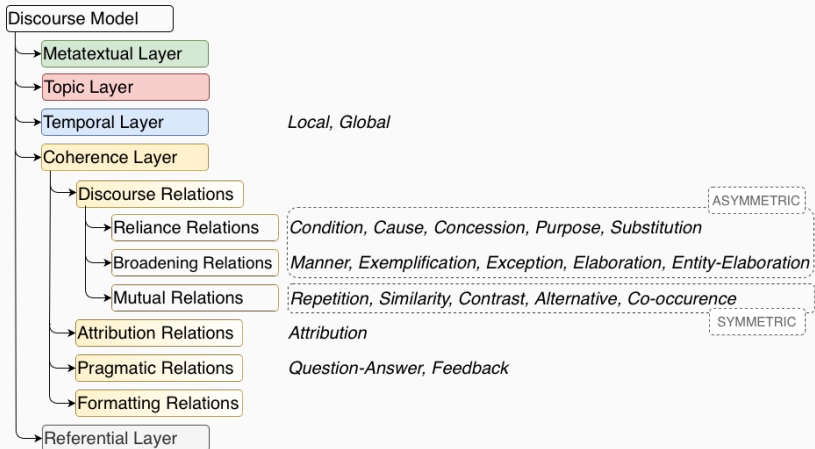
- skuteczność segmentera regułowego na zbiorze deu.rst.pcc wzrosła o 11.51 pp. F1,
- segmenter regułowy działa lepiej niż wiodące systemy oparte na modelach głębokich.

Relacje dyskursywne

Wielowarstwowy model analizy dyskursu (1)

- Proponowany model analizuje dyskurs jako **strukturę wielowarstwową**.
- Celem modelu jest uchwycenie złożoności tekstu poprzez integrację różnych poziomów organizacji dyskursu.
- Model umożliwia:
 - analizę relacji w poszczególnych warstwach i między warstwami,
 - opis spójności tekstu,
 - reprezentację organizacji dokumentu.

Wielowarstwowy model analizy dyskursu (2)



Warstwa relacji koherencyjnych [coherence layer]

- Rozróżniamy:
 - **spójność wewnątrzdanową** (poprawność gramatyczna, zgodność czasowa, spójność referencyjna),
 - **spójność międzudanową** (relacje semantyczne między segmentami tekstu).
- Segmenty dyskursywne powiązane relacjami koherencyjnymi tworzą spójną całość.
- Interpretując tekst, odkrywamy relacje między segmentami.
- Wyróżniamy cztery podtypy relacji koherencyjnych:
 - relacje dyskursywne,
 - relacje atrybucyjne,
 - relacje pragmatyczne,
 - relacje techniczne (formatowanie).

HOLMES HAD READ carefully a note which the last post had brought him. Then, with the dry chuckle which was his nearest approach to a laugh, he tossed it over to me.

"For a mixture of the modern and the mediaeval, of the practical and of the wildly fanciful, I think this is surely the limit," said he. "What do you make of it, Watson?"

I read as follows:

46, Old Jewry,
Nov. 19th.

Re Vampires

Sir:

Our client, Mr. Robert Ferguson, of Ferguson and Muirhead, tea brokers, of Mincing Lane, has made some inquiry from us in a communication of even date concerning vampires. As our firm specializes entirely upon the assessment of machinery the matter hardly comes within our purview, and we have therefore recommended Mr. Ferguson to call upon you and lay the matter before you. We have not forgotten your successful action in the case of Matilda Briggs.

We are, sir,
— Faithfully yours,
MORRISON, MORRISON, AND DODD.
per E. J. C.

"Matilda Briggs was not the name of a young woman, Watson," said Holmes in a reminiscent voice. "It was a ship which is associated with the giant rat of Sumatra, a story for which the world is not yet prepared. But what do we know about vampires? Does it come within our purview either? Anything is better than stagnation, but really we seem to have been switched on to a Grimms' fairy tale. Make a long arm, Watson, and see what V has to say."

I leaned back and took down the great index volume to which he referred. Holmes balanced it on his knee, and his eyes moved slowly and lovingly over the record of old cases, mixed with the accumulated information of a lifetime.

"Voyage of the *Gloria Scott*," he read. "That was a bad business. I have some recollection that you made a record of it, Watson, though I was unable to congratulate you upon the result. Victor Lynch, the forger. Venomous lizard or gila. Remarkable case, that! Vittoria, the circus belle. Vanderbilt and the Yeggman. Vipers. Vigor, the Hammersmith worker.

Hullo! Hullo! Good old index. You can't beat it. Listen to this, Watson. Vampirism in Hungary. And again, Vampires in Transylvania." He turned over the pages with eagerness, but after a short intense perusal he threw down the great book with a snarl of disappointment.

"Rubbish, Watson, rubbish! What have we to do with walking corpses who can only be held in their grave by stakes driven through their hearts? It's pure lunacy."

"But surely," said I, "the vampire was not necessarily a dead man? A living person might have the habit. I have read, for example, of the old sucking the blood of the young in order to retain their youth."

"You are right, Watson. It mentions the legend in one of these references. But are we to give serious attention to such things? This agency stands flat-footed upon the ground, and there it must remain. The world is big enough for us. No ghosts need apply. I fear that we cannot take Mr. Robert Ferguson very seriously. Possibly this note may be from him and may throw some light upon what is worrying him."

He took up a second letter which had lain unnoticed upon the table while he had been absorbed with the first. This he began to read with a smile of amusement upon his face which gradually faded away into an expression of intense interest and concentration. When he had finished he sat for some little time lost in thought with the letter dangling from his fingers. Finally, with a start, he aroused himself from his reverie.

"Cheeseman's, Lamberley. Where is Lamberley, Watson?"

"It is in Sussex, south of Horsham."

"Not very far, eh? And Cheeseman's?"

"I know that country, Holmes. It is full of old houses which are named after the men who built them centuries ago. You get Odley's and Harvey's and Carriton's—the folk are forgotten but their names live in their houses."

"Precisely," said Holmes coldly. "It was one of the peculiarities of his proud, self-contained nature that though he docketed any fresh information very quietly and accurately in his brain, he seldom made any acknowledgment to the giver. 'I rather fancy we shall know a good deal more about Cheeseman's, Lamberley, before we are through. The letter is, as I had hoped, from Robert Ferguson. By the way, he claims acquaintance with you.'"

"With me!"



Formating

Entry-Elaboration

T1

f1

f2

f3

T2

T2

T2

Question
Answer

Warstwa tematyczna [topic layer]

- Segmenty dyskursywne połączone relacjami tworzą **segmenty tematyczne**.
- Segmenty tematyczne są lokalnie spójne i powiązane leksykalnie.
- Długość segmentów tematycznych zależy od gatunku; mogą one odpowiadać:
 - całym rozdziałom artykułów naukowych (np. przeglądowi literatury, prezentacji wyników badań),
 - opisom scen, tła czy przyrody w tekstach narracyjnych,
 - artykułom prasowym itp.

HOLMES HAD READ carefully a note which the last post had brought him. Then, with the dry chuckle which was his nearest approach to a laugh, he tossed it over to me.

"For a mixture of the modern and the mediaeval, of the practical and of the wildly fanciful, I think this is surely the limit," said he. "What do you make of it, Watson?"

I read as follows:

46, Old Jewry,
Nov. 19th.

Re Vampires

Sir:

Our client, Mr. Robert Ferguson, of Ferguson and Muirhead, tea brokers, of Mincing Lane, has made some inquiry from us in a communication of even date concerning vampires. As our firm specializes entirely upon the assessment of machinery the matter hardly comes within our purview, and we have therefore recommended Mr. Ferguson to call upon you and lay the matter before you. We have not forgotten your successful action in the case of Matilda Briggs.

We are, sir,
— Faithfully yours,
MORRISON, MORRISON, AND DODD.
per E. J. C.

"Matilda Briggs was not the name of a young woman, Watson," said Holmes in a reminiscent voice. "It was a ship which is associated with the giant rat of Sumatra, a story for which the world is not yet prepared. But what do we know about vampires? Does it come within our purview either? Anything is better than stagnation, but really we seem to have been switched on to a Grimms' fairy tale. Make a long arm, Watson, and see what V has to say."

I leaned back and took down the great index volume to which he referred. Holmes balanced it on his knee, and his eyes moved slowly and lovingly over the record of old cases, mixed with the accumulated information of a lifetime.

"Voyage of the *Gloria Scott*," he read. "That was a bad business. I have some recollection that you made a record of it, Watson, though I was unable to congratulate you upon the result. Victor Lynch, the forger. Venomous lizard or gila. Remarkable case, that! Vittoria, the circus belle. Vanderbilt and the Yeggman. Vipers. Vigor, the Hammersmith worker.

Hullo! Hullo! Good old index. You can't beat it. Listen to this, Watson. Vampirism in Hungary. And again, Vampires in Transylvania." He turned over the pages with eagerness, but after a short intense perusal he threw down the great book with a snarl of disappointment.

"Rubbish, Watson, rubbish! What have we to do with walking corpses who can only be held in their grave by stakes driven through their hearts? It's pure lunacy."

"But surely," said I, "the vampire was not necessarily a dead man? A living person might have the habit. I have read, for example, of the old sucking the blood of the young in order to retain their youth."

"You are right, Watson. It mentions the legend in one of these references. But are we to give serious attention to such things? This agency stands flat-footed upon the ground, and there it must remain. The world is big enough for us. No ghosts need apply. I fear that we cannot take Mr. Robert Ferguson very seriously. Possibly this note may be from him and may throw some light upon what is worrying him."

He took up a second letter which had lain unnoticed upon the table while he had been absorbed with the first. This he began to read with a smile of amusement upon his face which gradually faded away into an expression of intense interest and concentration. When he had finished he sat for some little time lost in thought with the letter dangling from his fingers. Finally, with a start, he aroused himself from his reverie.

"Cheeseman's, Lamberley. Where is Lamberley, Watson?"

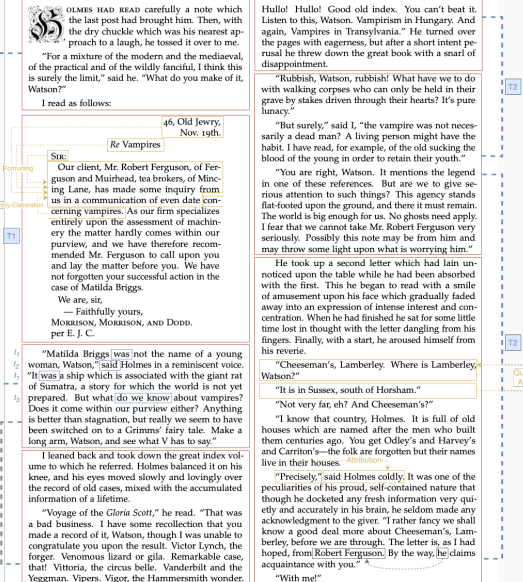
"It is in Sussex, south of Horsham."

"Not very far, eh? And Cheeseman's?"

"I know that country, Holmes. It is full of old houses which are named after the men who built them centuries ago. You get Odley's and Harvey's and Carriton's—the folk are forgotten but their names live in their houses."

"Precisely," said Holmes coldly. "It was one of the peculiarities of his proud, self-contained nature that though he docketed any fresh information very quietly and accurately in his brain, he seldom made any acknowledgment to the giver. 'I rather fancy we shall know a good deal more about Cheeseman's, Lamberley, before we are through. The letter is, as I had hoped, from Robert Ferguson. By the way, he claims acquaintance with you.'"

"With me!"



Warstwa temporalna [temporal layer]

- Warstwa temporalna organizuje dyskurs na osi czasowej.
- Dwa poziomy warstwy temporalnej:
 - **lokalny** – pojedyncze zdarzenia i punkty czasowe wewnątrz segmentów tematycznych,
 - **globalny** – ramy czasowe obejmujące wiele zdarzeń.
- Temporalność jest oddzielona od relacji koherencyjnych:
 - relacje koherencyjne odpowiadają za interpretację znaczenia,
 - globalne relacje temporalne porządkują przebieg wydarzeń w czasie.

HOLMES HAD READ carefully a note which the last post had brought him. Then, with the dry chuckle which was his nearest approach to a laugh, he tossed it over to me.

"For a mixture of the modern and the mediaeval, of the practical and of the wildly fanciful, I think this is surely the limit," said he. "What do you make of it, Watson?"

I read as follows:

46, Old Jewry,
Nov. 19th.

Re Vampires

Sir:

Our client, Mr. Robert Ferguson, of Ferguson and Muirhead, tea brokers, of Mincing Lane, has made some inquiry from us in a communication of even date concerning vampires. As our firm specializes entirely upon the assessment of machinery the matter hardly comes within our purview, and we have therefore recommended Mr. Ferguson to call upon you and lay the matter before you. We have not forgotten your successful action in the case of Matilda Briggs.

We are, sir,
— Faithfully yours,
MORRISON, MORRISON, AND DODD.
per E. J. C.

"Matilda Briggs was not the name of a young woman, Watson," said Holmes in a reminiscent voice. "It was a ship which is associated with the giant rat of Sumatra, a story for which the world is not yet prepared. But what do we know about vampires? Does it come within our purview either? Anything is better than stagnation, but really we seem to have been switched on to a Grimms' fairy tale. Make a long arm, Watson, and see what V has to say."

I leaned back and took down the great index volume to which he referred. Holmes balanced it on his knee, and his eyes moved slowly and lovingly over the record of old cases, mixed with the accumulated information of a lifetime.

"Voyage of the *Gloria Scott*," he read. "That was a bad business. I have some recollection that you made a record of it, Watson, though I was unable to congratulate you upon the result. Victor Lynch, the forger. Venomous lizard or gila. Remarkable case, that! Vittoria, the circus belle. Vanderbilt and the Yeggman. Vipers. Vigor, the Hammersmith worker.

Hullo! Hullo! Good old index. You can't beat it. Listen to this, Watson. Vampirism in Hungary. And again, Vampires in Transylvania." He turned over the pages with eagerness, but after a short intense perusal he threw down the great book with a snarl of disappointment.

"Rubbish, Watson, rubbish! What have we to do with walking corpses who can only be held in their grave by stakes driven through their hearts? It's pure lunacy."

"But surely," said I, "the vampire was not necessarily a dead man? A living person might have the habit. I have read, for example, of the old sucking the blood of the young in order to retain their youth."

"You are right, Watson. It mentions the legend in one of these references. But are we to give serious attention to such things? This agency stands flat-footed upon the ground, and there it must remain. The world is big enough for us. No ghosts need apply. I fear that we cannot take Mr. Robert Ferguson very seriously. Possibly this note may be from him and may throw some light upon what is worrying him."

He took up a second letter which had lain unnoticed upon the table while he had been absorbed with the first. This he began to read with a smile of amusement upon his face which gradually faded away into an expression of intense interest and concentration. When he had finished he sat for some little time lost in thought with the letter dangling from his fingers. Finally, with a start, he aroused himself from his reverie.

"Cheeseman's, Lamberley. Where is Lamberley, Watson?"

"It is in Sussex, south of Horsham."

"Not very far, eh? And Cheeseman's?"

"I know that country, Holmes. It is full of old houses which are named after the men who built them centuries ago. You get Odley's and Harvey's and Carriton's—the folk are forgotten but their names live in their houses."

"Precisely," said Holmes coldly. "It was one of the peculiarities of his proud, self-contained nature that though he docketed any fresh information very quietly and accurately in his brain, he seldom made any acknowledgment to the giver. 'I rather fancy we shall know a good deal more about Cheeseman's, Lamberley, before we are through. The letter is, as I had hoped, from Robert Ferguson. By the way, he claims acquaintance with you.'"

"With me!"



Formating
Entry-Elaboration

T1

f1

f2

f3

T2

T2

T2

T2

Question
Answer

Warstwa metatekstowa [metatextual layer]

- Warstwa metatekstowa organizuje strukturę tekstu, a nie jego treść.
- Obejmuje: tytuły, nagłówki, przypisy, numery stron, elementy edytorskie etc.
- Elementy metatekstowe:
 - nie należą do struktury semantycznej dyskursu,
 - pełnią funkcję porządkowania treści.

HOLMES HAD READ carefully a note which the last post had brought him. Then, with the dry chuckle which was his nearest approach to a laugh, he tossed it over to me.

"For a mixture of the modern and the mediaeval, of the practical and of the wildly fanciful, I think this is surely the limit," said he. "What do you make of it, Watson?"

I read as follows:

46, Old Jewry,
Nov. 19th.

Re Vampires

Sir:

Our client, Mr. Robert Ferguson, of Ferguson and Muirhead, tea brokers, of Mincing Lane, has made some inquiry from us in a communication of even date concerning vampires. As our firm specializes entirely upon the assessment of machinery the matter hardly comes within our purview, and we have therefore recommended Mr. Ferguson to call upon you and lay the matter before you. We have not forgotten your successful action in the case of Matilda Briggs.

We are, sir,
— Faithfully yours,
MORRISON, MORRISON, AND DODD.
per E. J. C.

"Matilda Briggs was not the name of a young woman, Watson," said Holmes in a reminiscent voice. "It was a ship which is associated with the giant rat of Sumatra, a story for which the world is not yet prepared. But what do we know about vampires? Does it come within our purview either? Anything is better than stagnation, but really we seem to have been switched on to a Grimms' fairy tale. Make a long arm, Watson, and see what V has to say."

I leaned back and took down the great index volume to which he referred. Holmes balanced it on his knee, and his eyes moved slowly and lovingly over the record of old cases, mixed with the accumulated information of a lifetime.

"Voyage of the *Gloria Scott*," he read. "That was a bad business. I have some recollection that you made a record of it, Watson, though I was unable to congratulate you upon the result. Victor Lynch, the forger. Venomous lizard or gila. Remarkable case, that! Vittoria, the circus belle. Vanderbilt and the Yeggman. Vipers. Vigor, the Hammersmith worker.

Hullo! Hullo! Good old index. You can't beat it. Listen to this, Watson. Vampirism in Hungary. And again, Vampires in Transylvania." He turned over the pages with eagerness, but after a short intense perusal he threw down the great book with a snarl of disappointment.

"Rubbish, Watson, rubbish! What have we to do with walking corpses who can only be held in their grave by stakes driven through their hearts? It's pure lunacy."

"But surely," said I, "the vampire was not necessarily a dead man? A living person might have the habit. I have read, for example, of the old sucking the blood of the young in order to retain their youth."

"You are right, Watson. It mentions the legend in one of these references. But are we to give serious attention to such things? This agency stands flat-footed upon the ground, and there it must remain. The world is big enough for us. No ghosts need apply. I fear that we cannot take Mr. Robert Ferguson very seriously. Possibly this note may be from him and may throw some light upon what is worrying him."

He took up a second letter which had lain unnoticed upon the table while he had been absorbed with the first. This he began to read with a smile of amusement upon his face which gradually faded away into an expression of intense interest and concentration. When he had finished he sat for some little time lost in thought with the letter dangling from his fingers. Finally, with a start, he aroused himself from his reverie.

"Cheeseman's, Lamberley. Where is Lamberley, Watson?"

"It is in Sussex, south of Horsham."

"Not very far, eh? And Cheeseman's?"

"I know that country, Holmes. It is full of old houses which are named after the men who built them centuries ago. You get Odley's and Harvey's and Carriton's—the folk are forgotten but their names live in their houses."

"Precisely," said Holmes coldly. "It was one of the peculiarities of his proud, self-contained nature that though he docketed any fresh information very quietly and accurately in his brain, he seldom made any acknowledgment to the giver. 'I rather fancy we shall know a good deal more about Cheeseman's, Lamberley, before we are through. The letter is, as I had hoped, from Robert Ferguson. By the way, he claims acquaintance with you.'"

"With me!"



Formating
Entry-Elaboration

T1

f1

f2

f3

T2

T2

T2

Question
Answer

- Warstwa referencyjna reprezentuje relacje referencyjne i ich ciągi.
- Odpowiada za modelowanie koreferencji.

Interakcje między warstwami

- Warstwy modelu są autonomiczne, ale wchodzą ze sobą w interakcje.
- Relacje koherencyjne pomagają wyznaczać granice segmentów tematycznych.
- Zmiany tematu często współwystępują ze:
 - zmianą osób uczestniczących w zdarzeniach,
 - zmianą perspektywy czasowej.
- Elementy metatekstowe mogą sygnalizować przejścia tematyczne.

Relacje koherencyjne – ISO jako baza

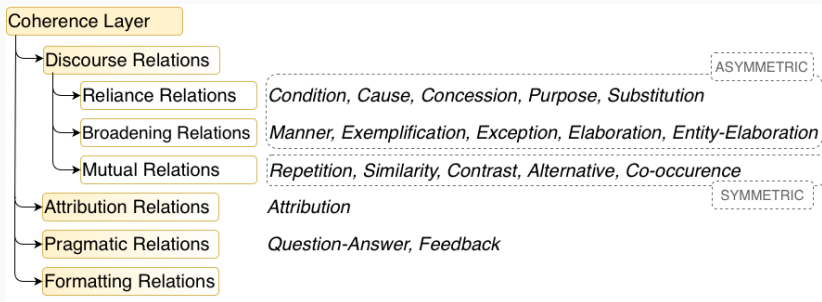
Zestaw relacji koherencyjnych w UDisc bazuje na normie ISO-24617-8, bo chcemy:

- **oznaczyć dane w wielu językach** – to wymaga globalnego spojrzenia i uchwycenia ogólnych tendencji (w tym celu lepiej sprawdza się lista 20 relacji niż np. 76),
- **uspójnić dotychczasowe dane** – zestaw z ISO to „parasol”, który obejmuje relacje występujące we wszystkich formalizmach.

Na podstawie schematu anotacyjnego ISO powstały już:

- Polish Discourse Corpus (PDC),
- Fables-DTR – korpus bajek dla języków: polskiego, portugalskiego i angielskiego.

Typy relacji koherencyjnych



Obecny zbiór relacji w UDisc wpisuje się w strukturę wielowarstwowego modelu analizy dyskursu, ale dokonaliśmy pewnych zmian względem ISO:

- relacje temporalne i *Expansion* przenieśliśmy do osobnych warstw,
- relacje *Condition* i *Negative Condition* zamknęliśmy w jednej: *Condition*,
- relację *Elaboration* podzieliliśmy na *Elaboration* i *Entity-Elaboration*,
- zmieniliśmy nazwy relacji *Conjunction*, *Disjunction*, *Restatement*, *Functional dependence* i *Feedback dependence* na bardziej adekwatne: *Co-occurrence*, *Alternative*, *Repetition*, *Question-Answer*, *Feedback*.

To relacje znaczeniowe pomiędzy wyodrębnionymi segmentami spójnego tekstu. Dzielimy je na trzy podtypy:

- **reliance (determinujące)** – sytuacja przedstawiona w jednym segmencie pełni jakąś funkcję względem drugiej, determinuje ją (jest warunkiem, przyczyną, celem itd.),
- **broadening (rozszerzające)** – dodanie, rozszerzenie informacji, doprecyzowanie jej,
- **mutual (obustronne)** – segmenty są wobec siebie symetryczne znaczeniowo $(REL, A, B) = (REL, B, A)$.

Relacje determinujące, asymetryczne [reliance]

RELIANCE RELATIONS

Relation Type	Definition / Paraphrase	Polish Example
Condition	<i>If X happens, then Y happens</i>	[Złamiesz nogę.] [jeśli będziesz dalej skakać.] [Gdyby kózka nie skakała,] [toby nóżki nie złamała.]
Cause	<i>X causes Y / Y results from X</i>	[Jacek złamał nogę,] [bo skakał.] [Nie palę,] [więc nie zanieczyszczam powietrza.]
Concession	<i>Y happens despite X</i>	[Choć bardzo się spieszyłam,] [nie zdążyłam kupić chleba.] [Przegrali ten mecz,] [mimo że dali z siebie wszystko.]
Purpose	<i>Y indicates the goal of X</i>	[Uczę się angielskiego,] [żeby móc studiować w Oxfordzie.]
Substitution	<i>Y replaces X</i>	[Zamiast pójść spać,] [czytał książkę do północy.]

Relacje rozszerzające, asymetryczne [broadening]

BROADENING RELATIONS

Relation Type	Definition / Paraphrase	Polish Example
Manner	<i>Y indicates the way in which X occurs</i>	[Rozwiązałem tę sprawę w taki sposób,] [że po prostu odpuściłem.]
Exemplification	<i>Y provides an example of X</i>	[Stalowe części amerykańskich wież trafiły do jedenastu państw.] [Pojechały na przykład do Indii, Chin i Malezji.]
Exception	<i>Y is an exception to the rule given in X</i>	[Wszyscy studenci oblali egzamin.] [Wyjątkiem była Iza z maksymalną liczbą punktów z testu.]
Elaboration	<i>Y expands or specifies the situation presented in the whole of X</i>	[Maria dostała prestiżową nagrodę naukową.] [Chodzi o grant Europejskiej Rady ds. Badań.]
Entity-Elaboration	<i>Y expands or specifies one element in X</i>	[Była niewielka odległość między bramkami a tunelem,] [w który wchodzili kończący bieg zawodnicy.]

Relacje obustronne, symetryczne [mutual]

MUTUAL RELATIONS

Relation Type	Definition / Paraphrase	Polish Example
Repetition	<i>Y restates or repeats X</i>	[Zniknął,] [przepadł bez wieści.]
Similarity	<i>Y is similar or comparable to X</i>	[Opowiadał długo i porywająco,] [tak samo jak przed laty robił to jego ojciec.]
Contrast	<i>Y contrasts with X</i>	[Ja lubię zimę,] [a ona lubi lato.]
Alternative	<i>X and Y are mutually exclusive; There is a choice between X and Y</i>	[Wieczorem będę oglądać film] [albo poczytam książkę.]
Co-occurrence	<i>X and Y co-occur in the discourse</i>	[Nie zanieczyszczam powietrza] [i nie rzucam śmieci.]

Atrybucja to **informacja o źródle wypowiedzi**, wprowadza kontekst.

ATTRIBUTION RELATIONS

Relation Type	Definition / Paraphrase	Polish Example
Attribution	<i>Y is attributed to a speaker or source expressed in X</i>	[Lubię chodzić do pracy –] [powiedział Piotr.] [Konstytucja mówi:] [(artykuł konstytucji)]

Relacje pragmatyczne

Relacje te pełnią funkcję pragmatyczną: nie budują spójności znaczeniowej między segmentami dyskursywnymi, nie wprowadzają nowych treści. **Organizują interakcję, wprowadzają komentarz, czasem pełnią funkcję fatyczną.**

PRAGMATIC RELATIONS

Relation Type	Definition / Paraphrase	Polish Example
Question–Answer	<i>Y answers the question raised in X</i>	[Czy na pewno dobrze zrobiłem?][Chyba tak.]
Feedback	<i>Y is a feedback response to X (not a question)</i>	[– Dobry jest ten makaron.] [– Mhm.] [– Pożyczę od ciebie samochód dziś wieczorem.] [– OK.]

Relacje techniczne (formatowanie)

To relacje, które **porządkują wypowiedź formalnie**. To też metatekst wewnątrz dyskursu.

FORMATTING RELATIONS

Relation Type	Definition / Paraphrase	Polish Example
Formatting	<i>Y is a footnote, reference, or list item relative to X</i>	[Karol Nowak pisał o tym w swojej książce] [(15).] [Dawno temu nie było telefonów] [(zob. "Telefony a nowoczesność" Nowaka).] [Żeby otrzymać dowód osobisty, należy:] [– dostarczyć kopię zdjęcia,] [– okazać dowód zapłaty.]

Anotacja – podstawy:

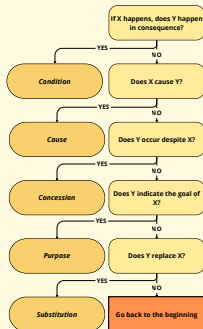
- Narzędzie: UDream.
- Cel: wskazanie relacji zachodzących między **wszystkimi segmentami dyskursywnymi** w tekście.
- Każda warstwa w modelu analizy dyskursu (tu: koherencyjna) pozwala na oznaczenie **tylko jednej relacji**.
- Relacje koherencyjne tworzą **drzewo skierowane**, które odzwierciedla **strukturę** opisywanego tekstu.

Drzewo decyzyjne – główne założenia

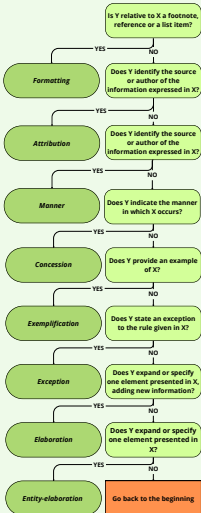
- Cel: powtarzalność decyzji anotacyjnych dzięki maksymalnemu ograniczeniu arbitralności.
- Obejmuje wszystkie relacje koherencyjne.
- Pierwszy krok: decyzja, jaki jest ogólny charakter relacji (determinująca, rozszerzająca narrację, symetryczna czy pragmatyczna).
- Dalej: identyfikacja relacji przebiega od najbardziej szczegółowej do najbardziej ogólnej.
- Każda relacja w wytycznych opatrzona jest szeregiem przykładów.
- Aby uprościć proces decyzyjny, relacje *Attribution* i *Formatting* zostały włączone do gałęzi relacji rozszerzających narrację (identyfikowane są na początku).

What is the relation between X and Y?

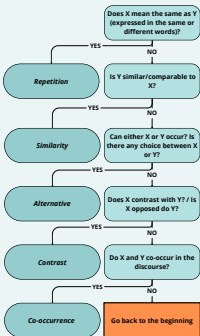
Relation expresses a reliance between X and Y such as condition, cause–result, purpose, substitution, or concession



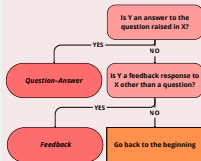
Relation expands or specifies X



Relation compares, contrasts, presents an alternative, or expresses similarity between X and Y



Relation is primarily pragmatic



Prace w toku

Po co nam nowe narzędzie?

- żeby uprościć anotację, która ma być maksymalnie wygodna i przyjazna anotatorowi,
- żeby wprowadzić dodatkowe funkcje weryfikacyjne, specyficzne dla naszego podejścia (np. łatwe sprawdzanie, które segmenty nie mają jeszcze przypisanych relacji),
- żeby wprowadzić etapowość anotacji z możliwością aktualizowania analizy,
- żeby zintegrować system anotacyjny z narzędziami automatycznymi.

Texts in project *PLCS* 90 items • strona 1 z 3

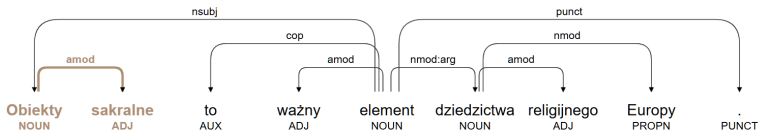
No.	Text ID	Filename	Progress	Annotator
1.	nauki humanistyczne_6	plcs_polish_100_segmented	100%	Anna Latusek
2.	nauki humanistyczne_21	plcs_polish_100_segmented	100%	Anna Latusek
3.	nauki humanistyczne_24	plcs_polish_100_segmented	93%	Anna Latusek
4.	nauki humanistyczne_38	plcs_polish_100_segmented		Anna Latusek
5.	nauki humanistyczne_39	plcs_polish_100_segmented		Anna Latusek
6.	nauki humanistyczne_48	plcs_polish_100_segmented	94%	Anna Latusek
7.	nauki humanistyczne_53	plcs_polish_100_segmented	100%	Anna Latusek
8.	nauki humanistyczne_73	plcs_polish_100_segmented	90%	Anna Latusek
9.	nauki humanistyczne_81	plcs_polish_100_segmented	88%	Anna Latusek
10.	nauki inzynieryjno-techniczne_5	plcs_polish_100_segmented	100%	Anna Latusek

sztuka_4

- 1.1 Obiekty sakralne to ważny element dziedzictwa religijnego Europy.
- 2.1 Obecnie z powodu laicyzacji i sekularyzacji Europy adaptacja obiektów sakralnych staje się rozpowszechnionym zjawiskiem.
- 3.1 Miejsca kultu nie powinny być traktowane wyłącznie w kategoriach funkcji użytkowej.
- 4.1 Celem artykułu jest przedstawienie adaptacji budynków sakralnych na kolumbaria jako jednej z możliwości zachowania dziedzictwa religijnego.
- 5.1 Taki typ rozwiązania łączy nową funkcję z tożsamością danego miejsca oraz jego wartościami: historycznymi, kulturowymi i sakralnymi.
- 6.1 Zastosowanie metody narracyjnej przy analizie tego zjawiska może być jednym z narzędzi poznawczych i krytyki architektury transformacji, 6.2 które pomoże znaleźć godne nowe przeznaczenie, 6.3 nie powodując profanacji miejsca o szczególnym znaczeniu religijnym.
- 7.1 Metoda narracyjna może być przydatna dla projektantów takich obiektów, 7.2 gdyż jest narzędziem holistycznym, 7.3 spajającym nie tylko kategorie estetyczne, lecz także filozoficzne, etyczne i semantyczne.

sztuka_4 Sentence 1

Obiekty sakralne to ważny element dziedzictwa religijnego Europy.



No.	Form	Lemma	UPos	Head	Deprel	Misc
1.	Obiekty	obiekt	NOUN	5	nsubj	Seg=609.1
2.	sakralne	sakralny	ADJ	1	amod	Seg=0
3.	to	to	AUX	5	cop	Seg=0
4.	ważny	ważny	ADJ	5	amod	Seg=0
5.	element	element	NOUN	0	root	Seg=0
6.	dziedzictwa	dziedzictwo	NOUN	5	nmod:arg	Seg=0
7.	religijnego	religijny	ADJ	6	amod	Seg=0
8.	Europy	Europa	PROPN	6	nmod	Seg=0 SpaceAfter=No
9.	.	.	PUNCT	5	punct	Seg=0

nauki inżynierijno- techniczne_5 Anna Latusek



- [1.2 Entity-elaboration →](#) 1.1 Omówiono sposób aproksymacji funkcją wykładniczą wyników badania procesu sprężania czynnika roboczego w cylindrach okrętowego tłokowego silnika spalinowego Sulzer 6AL20/24, [→ 2.1 Cause](#)

[1.3 Entity-elaboration →](#) 1.2 wyposażonego w czujnik ciśnienia [→ 1.1 Entity-elaboration](#) 1.3 umieszczony w głowicy. [→ 1.2 Entity-elaboration](#)
- [1:1 Cause →](#) [2.2 Entity-elaboration →](#) 2.1 Daje to możliwość wyznaczenia położenia GMP, ciśnienia początkowego i ciśnienia sprężania wtedy, 2.2 gdy zapłon następuje przed GMP. [→ 2.1 Entity-elaboration](#)

9 Artykuł 8 tej dyrektywy stanowi:

„1. Państwa członkowskie zapewniają, żeby używanie informacji handlowych, które są częścią lub stanowią usługę społeczeństwa informacyjnego świadczoną przez przedstawiciela zawodu regulowanego, było dozwolone pod warunkiem zgodności z zasadami wykonywania zawodu, dotyczącymi w szczególności niezależności, godności i prestiżu zawodu, tajemnicy zawodowej i rzetelności wobec klientów i innych przedstawicieli zawodu.

2. Bez uszczerbku dla autonomii organizacji i stowarzyszeń zawodowych państwa członkowskie i Komisja zachęcają stowarzyszenia i organizacje zawodowe do opracowywania kodeksów postępowania na poziomie wspólnotowym w celu ustalenia rodzajów informacji, które udzielane są do celów informacji handlowych zgodnie z regułami określonymi w ust. 1.

[...]”.

9 Article 8 of that directive provides:

‘1. Member States shall ensure that the use of commercial communications which are part of, or constitute, an information society service provided by a member of a regulated profession is permitted subject to compliance with the professional rules regarding, in particular, the independence, dignity and honour of the profession, professional secrecy and fairness towards clients and other members of the profession.

2. Without prejudice to the autonomy of professional bodies and associations, Member States and the Commission shall encourage professional associations and bodies to establish codes of conduct at Community level in order to determine the types of information that can be given for the purposes of commercial communication in conformity with the rules referred to in paragraph 1

...’

Dyskurs prawniczy (2)

- Korpus równoległy (języki UE):
 - wysokiej jakości tłumaczenia,
 - teksty w jednym języku tłumaczone na pozostałe języki UE,
 - otwarte dane.
- Szczególny typ tekstów:
 - dużo przypisów, zaznaczonych elips, wtrąceń, cytatów, obcych nazw własnych, itp.
 - często zawiera listy z wielozdaniowymi punktami.
- Cele badawcze:
 - czy nasz model nadaje się do analizy tekstów prawniczych – orzeczeń sądowych?
 - czy koncepcja segmentacji i ogólny zbiór relacji opracowany nadają się do anotacji tekstów dziedzinowych (prawniczych)?
 - czy możliwe jest zaanotowanie zbioru danych równoległych?
 - na ile anotacje dyskursywne w różnych językach są spójne?
 - czy można wytrenować parser dyskursu prawniczego?

Ontologia dyskursywna i aplikacja do jej przeglądania

Istnieją już:

- opisy formalizmów reprezentacji dyskursu w postaci ontologii,
- przeglądarki korpusów dyskursu,
- listy i przeglądarki znaczników dyskursywnych,

ale środowisko jest rozproszone, a formalne modele są trudne w bezpośrednim użyciu dla lingwistów (wymagają znajomości OWL i użycia edytorów takich jak Protégé).

To może:

- uzupełnimy ontologię danymi z korpusów,
- stworzymy przeglądarkę ontologii z intuicyjnym interfejsem.

Przyszły model dla skonwertowanych korpusów

Ontologia:

- osobne pliki dla modelu i instancji (`discourse.ISO.<korpus>.owl`),
- instancje z 3 korpusów testowych (instrukcja anotacji PDC, przykłady z normy ISO 24617-8, obecna wersja PDC),
- dodatki w postaci identyfikatora przykładu i kolejności elementów.

Przeglądarka:

- wyświetlanie wewnętrznej struktury relacji,
- filtrowanie względem języka, znacznika, typu relacji i symetrii,
- wyszukiwanie pełnotekstowe w przykładach.

Przeglądarka ontologii

ISO 24617-8 Ontology Viewer

Corpora

- PDC 17834
Language: pl
- ISO-24617-8 56
Language: en
- PDC-Guidelines 54
Language: pl

[Select All](#) [Deselect All](#)

Filters

Language: Connective: Relation type: Relation symmetry:

Full-text Search:

Examples

Concession #PDC 1 Asymmetric pl

W spotkaniu weźmie udział blisko 7 tysięcy braci z całej Europy, ale tylko 206 z nich będzie ubiegało się o tytuł Europejskiego Króla Kurkowego

TYPE	TEXT	ORDER
Arg1	ExpectationRaiser W spotkaniu weźmie udział blisko 7 tysięcy braci z całej Europy	1
Conn	Connective ale	2
Arg2	ExpectationDenier tylko 206 z nich będzie ubiegało się o tytuł Europejskiego Króla Kurkowego	3

Corpus: PDC • ID: http://purl.org...6554_12327533_1

<https://ontology-viewer.ipipan.waw.pl/>
<https://github.com/ipipan/ontology-viewer>

Mapowanie między formalizmami

Mapowanie relacji było/jest wyzwaniem:

- w repozytorium DISRPT-a (Shared Task on Discourse Unit Segmentation, Connective Detection, and Discourse Relation Classification) znalazło się **39 korpusów** zaanotowanych w kilku formalizmach (m.in. RST, PDTB, SDRT, ISO).
- pojawia się w nich łącznie **prawie 400 różnych etykiet** relacji.

Na razie wstępnie zmapowaliśmy wszystkie ww. relacje (w oparciu o dane z korpusów i o przypisane im relacje DISRPT) do naszego formatu.

Czy podejmiemy się mapowania relacji pomiędzy poszczególnymi korpusami / typami korpusów? Temat jest otwarty.

Mapowanie – przykład

key	disrpt_label	UIISO	files
restatement-partial	reformulation	repetition	eng.erst.gentle_test, eng.erst.gum_dev, eng.erst.gum_test, eng.erst.gum_train, zho.rst.godt_dev, zho.rst.godt_test, zho.rst.godt_train
restatement-repetition	reformulation	repetition	eng.erst.gentle_test, eng.erst.gum_dev, eng.erst.gum_test, eng.erst.gum_train, zho.rst.godt_dev, zho.rst.godt_test, zho.rst.godt_train
result	causal	cause	ces.rst.crdt_dev, ces.rst.crdt_train, deu.rst.pcc_dev, deu.rst.pcc_test, deu.rst.pcc_train, eng.dep.scidtb_dev, eng.dep.scidtb_test, eng.dep.scidtb_train, eng.rst.rstdt_dev, eng.rst.rstdt_test, eng.rst.rstdt_train, eng.sdrf.msdc_dev, eng.sdrf.msdc_test, eng.sdrf.msdc_train, eng.sdrf.stac_dev, eng.sdrf.stac_test, eng.sdrf.stac_train, fra.sdrf.annodis_dev, fra.sdrf.annodis_test, fra.sdrf.annodis_train, spa.rst.sctb_dev, spa.rst.sctb_test, spa.rst.sctb_train, zho.dep.scidtb_dev, zho.dep.scidtb_test, zho.dep.scidtb_train, zho.rst.sctb_dev, zho.rst.sctb_test, zho.rst.sctb_train
result-e	causal	cause	eng.rst.rstdt_train
resultado	causal	cause	spa.rst.rstsb_dev, spa.rst.rstsb_test, spa.rst.rstsb_train
resumen	organization	elaboration	spa.rst.rstsb_dev, spa.rst.rstsb_test, spa.rst.rstsb_train
rhetorical-question	query	question-answer	eng.rst.rstdt_test, eng.rst.rstdt_train
secuencia	temporal	temporal	spa.rst.rstsb_dev, spa.rst.rstsb_test, spa.rst.rstsb_train
sekuenzia	temporal	temporal	eus.rst.ert_dev, eus.rst.ert_test, eus.rst.ert_train
sequence	temporal	temporal	ces.rst.crdt_dev, ces.rst.crdt_test, ces.rst.crdt_train, deu.rst.pcc_dev, deu.rst.pcc_test, deu.rst.pcc_train, eng.rst.oil_dev, eng.rst.oil_test, eng.rst.oil_train, eng.rst.rstdt_dev, eng.rst.rstdt_test, eng.rst.rstdt_train, eng.rst.sts_dev, eng.rst.sts_test, eng.rst.sts_train, eng.sdrf.msdc_dev, eng.sdrf.msdc_test, eng.sdrf.msdc_train, nld.rst.nldt_dev, nld.rst.nldt_test, nld.rst.nldt_train, por.rst.cstn_dev, por.rst.cstn_test, por.rst.cstn_train, rus.rst.rrt_dev, rus.rst.rrt_test, rus.rst.rrt_train, spa.rst.sctb_dev, spa.rst.sctb_test, spa.rst.sctb_train, zho.rst.sctb_dev, zho.rst.sctb_test, zho.rst.sctb_train
similarity	mode	similarity	pol.iso.pdc_dev, pol.iso.pdc_test, pol.iso.pdc_train
solución	query	elaboration	spa.rst.rstsb_dev, spa.rst.rstsb_test, spa.rst.rstsb_train
solutionhood	query	elaboration	ces.rst.crdt_dev, ces.rst.crdt_test, ces.rst.crdt_train, deu.rst.pcc_dev, deu.rst.pcc_test, deu.rst.pcc_train, eng.rst.oil_dev, eng.rst.oil_test, eng.rst.oil_train, eng.rst.sts_dev, eng.rst.sts_test, eng.rst.sts_train, nld.rst.nldt_dev, nld.rst.nldt_test, nld.rst.nldt_train, rus.rst.rrt_dev, rus.rst.rrt_test, rus.rst.rrt_train

Podsumowanie

- Uniwersalne podejście do dyskursu – opis różnych języków za pomocą tych samych zasad.
- Segmentacja oparta na uniwersalnych drzewach zależnościowych.
- Nowa koncepcja wielowarstwowej analizy dyskursu.
- Uspójniona, powtarzalna anotacja bazująca na drzewie decyzyjnym.
- Ujednolicenie zbiorów danych.

Dziękujemy za uwagę!

Bibliografia

- Ogródniczuk M., Latusek A., Saputa K., Wróblewska A., Ziembicki D., Żuk B., Lewandowska M., Okraśiński A., Rosalska P., Śliwicka A., Tomaszewska A., Żurowski S. (2025). *Where frameworks (dis)agree: A study of discourse segmentation*. In M. Strube, C. Braud, C. Hardmeier, J. J. Li, S. Loaiciga, A. Zeldes, C. Li (eds.) *Proceedings of the 6th Workshop on Computational Approaches to Discourse, Context and Document-Level Inferences (CODI 2025)*, pp. 182–196, Suzhou, China. ACL.
- P. Silvano, A. Leal, A. Tomaszewska, M. Ogródniczuk, M. Lewandowska, A. Śliwicka, L. F. Cunha, E. Amorim, J. Gomes (2026). *Fables-DTR: A Corpus of Fables Annotated for Discourse and Temporal Relations*. In *Proceedings of the 15th Language Resources and Evaluation Conference (LREC 2026)*, Palma de Mallorca, Spain. ELRA.
- Ogródniczuk M., Czerski D. (2026). *Towards Corpus-Based Population and Visualization of ISO 24617-8 Ontology*. In *Proceedings of the 2026 Joint ACL – ISO Workshop on Interoperable Semantic Annotation (ISA-22) at LREC 2026*. Palma de Mallorca, Spain. ELRA.
- Latusek A., Ogródniczuk M., Wróblewska A., Żuk B. (2026). *Universal Discourse Relations: A Proposal*. In *Proceedings of 2nd Joint Workshop on Computational Approaches to Discourse, Context and Document-Level Inferences (7th CODI) and Computational Models of Reference, Anaphora and Coreference (9th CRAC)*, San Diego, United States. ACL.