

Das kombinatorische Wörterbuch von ETAP





Typen der Lexeminformation

- Im Gegensatz zu den allgemeinen Regeln werden die Schablonen- und Wörterbuchregeln nur auf die Phrasen angewandt, deren lexikalisches Potential das verlangen könnte.
- Die Wortinformation wird in zwei Typen unterteilt:
 - Klassifikationsinformation – Information über die Wortmerkmale, auf die Systemregeln verweisen können (Wortart, syntaktische und semantische Merkmale, Abhängigkeiten der Aktanten im RM)
 - Operationsinformation – Information über die Schablonen- oder Wörterbuchregeln.



Klassifikationsinformation im KW

- Ein Wortartikel beschreibt genau ein Lexem (eine Bedeutung des Wortes). Er fängt mit einer Kopfzeile an, die Artikelnummer und den Lexemnamen enthält.
- Wenn das Wort lexikalisch-grammatikalische Homonyme besitzt, enthält der Name jedes Lexems einen Index
 - ЧТО1 (союз) vs. ЧТО2 (союзное слово)
- Klassifikationsinformation wird in einem Artikel in vier Bereichen wiedergegeben. Jeder Bereich wird mit einem speziellen Zeichen eingeführt:
 - POR: Wortart
 - SYNT: syntaktische Merkmale
 - DES: semantische Merkmale
 - $D_{i,j}$: Zeilen, die das RM beschreiben
 - (TRANS: Triviale Übersetzung eines russischen Lexems in Englisch)



Wortartbereich

- In diesem Bereich nach dem POR-Zeichen wird das Symbol eines Wortarts angegeben.
 - **S** (Substantiv), **A** (Adjektiv), **ADV** (Adverb), **NUM** (Zahlwort), **V** (Verb), **CONJ** (Konjunktion), **PR** (Präposition), **PART** (Partikel), **COM** (die erste Komponente der zusammengesetzten Wörter wie *франко-русский*)
- Klasse A enthält außer den kanonischen Adjektiven
 - Pronominaladjektive: *ЭТОТ, ТОТ*
 - Possesivpronominale Adjektive: *МОЙ, ТВОЙ, ВАШ, НАШ*
 - Interrogativpronomen: *КАКОЙ, ЧЕЙ*
 - Indefinitpronominale Adjektive: *ДРУГОЙ, НЕКОТОРЫЙ, ЛЮБОЙ*
 - Ordnungszahlwörter: *ПЕРВЫЙ, ВТОРОЙ*
 - Komplexe Pronominaladjektive: *КАКОЙ-ТО, КОЕ-КАКОЙ*



Wortartbereich

- Klasse S enthält außer den kanonischen Substantiven
 - Personalpronomen *что2, кто*
 - Interrogativpronomen *который*
 - Relativpronomen *некто, нечто*
 - Reflexivpronomen *себя*
 - Negativpronomen *никто, ничто*
 - Demonstrativpronomen *это1, то1*
 - Komplexe Substantivpronomen *кое-кто, кое-что, кто-либо, что-то*
- Klasse ADV enthält außer den Adverbien
 - Einführungswörter und Wendungen *естественно, вероятно, по всей видимости*
 - Pronominaladverbien *где, когда2, там, туда*
 - Prädikativa *жаль, можно, некогда*
 - Quantifikatoren *мало, много, несколько*



Bereich der syntaktischen Merkmale

- Unter syntaktischen Merkmalen versteht man Abkürzungen der Wortmerkmale, die die Worteinbindung in die einen syntaktischen Konstruktionen erlauben und nicht in die anderen

количест

две большой важности проблемы

количест

**две большой важности проблемы*

- Zusammenhang zwischen syntaktischen Merkmalen und morphologischen Charakteristiken (m, f, n, belebt):
 - Klassifikationsrolle bei den Substantiven (syntaktische Merkmale)
 - Flexion bei den Adjektiven und Verben (morphologische Merkmale)



Syntaktische Merkmale der Verben

- ВВОД-1: Verben, die einen Nebensatz eingeführt mit der Präposition *что, чтобы* als ersten Aktanten haben und die mit der Konjunktion *как* in der Einführungskonstruktion verwendet werden können

- ВОДИТЬСЯ, ВЫЯСНЯТЬСЯ, ОКАЗАТЬСЯ, ПОЛАГАТЬСЯ...

Некоторые сотрудники, как выяснилось в ходе расследования, не соблюдали элементарных правил техники безопасности

- ВВОД-2: Verben, die einen Nebensatz eingeführt mit der Präposition *что, чтобы* als zweiten Aktanten haben und die mit der Konjunktion *как* in der Einführungskonstruktion verwendet werden können

- ВИДЕТЬ, ВЫЯСНЯТЬ, ГОВОРИТЬ, ДУМАТЬ, СЧИТАТЬ...

Некоторые сотрудники, как выяснила группа в ходе расследования, не соблюдали элементарных правил техники безопасности



Syntaktische Merkmale der Verben

- **ГРАД:** Verben, die eine Steigerungssituation beschreiben und Adverbien wie *очень, весьма, немного, чуть-чуть* anschließen können
 - **ВЫРАСТИ, ЛЮБИТЬ, УСТАВАТЬ, ХОТЕТЬ, ОБИДЕТЬСЯ...**
Он немного обиделся на меня
- **ИЗМЕН:** Verben, die Veränderung des Zustandes oder der Lage im Raum bedeuten und mit der adverbialen Bestimmung wie *на три дня* verwendet werden können
 - **ВСТАВАТЬ, ПРИХОДИТЬ, УЕЗЖАТЬ, ЗАМОЛЧАТЬ...**
Он уехал на неделю в командировку
- **МГНОВ:** Verben der augenblicklichen Handlung, die nicht mit den adverbialen Dauerbestimmung wie *долго, неделю, три дня* verwendet werden können
 - **ВЕЛЕТЬ, ДОСТИГАТЬ, КАСАТЬСЯ, НАХОДИТЬ, ПРИХОДИТЬ**
**судно целый день приходило в порт*



Syntaktische Merkmale der Verben

- **МОД: Modalverben**
 - МОЧЬ, УМЕТЬ, ХОТЕТЬ...

- **РДОП: transitive Verben, die den Wechsel von Akkusativ des Objekts gegen Genitiv im Negierungskontext zulassen**
 - ДЕЛАТЬ, ЕСТЬ, ПИТЬ, ЧИТАТЬ, ВИДЕТЬ...
 - Все читали эту книгу -> Никто не читал этой книги*

- **РДОПЛОТР: transitive Verben, die den Wechsel von Nominativ des Subjekts gegen Genitiv im Negierungskontext zulassen**
 - БЫТЬ, БЫВАТЬ, ИМЕТЬСЯ, СУЩЕСТВОВАТЬ, ПОСТУПАТЬ...
 - Такие сведения в институт еще не поступали ->*
 - Таких сведений в институт еще не поступало*



Syntaktische Merkmale der Substantive

- ЖЕНСК; МУЖСК; СРЕДН
- АГЕНС: Substantive, die agentive Objekte sein können bei der Passivform des Verbes
 - ЧЕЛОВЕК, АЛГОРИТМ, СОЛНЦЕ
Запрос обрабатывается компьютером
- ВРЕМ: Substantive mit der Bedeutung eines Zeitpunkts oder -intervals
- ГЕОГР: Substantive mit der Bedeutung eines geographischen Objekts, die keine Eigennamen sind
 - ГОРА, ГОРОД, РЕКА, СТРАНА...
- ДЕН: Währungsamen
- ИЗМЕР: Maßeinheiten
- ИСЧИСЛ: zählbare Substantive, die in Mengen- und Sammelgruppen verwendet werden können
- ЛИЧН: Personalpronominale Nomen



Syntaktische Merkmale der Substantive

- ПРОФ: Substantive in Maskulinum, die einen Beruf bezeichnen und ein Verb oder Adjektiv in Femininum zulassen
 - ВРАЧ, ДИРЕКТОР, ПИЛОТ...
В комнату вошла администратор гостиницы
- СОВОК: Substantive mit der Bedeutung einer Gesamtheit, die prädikative Konstruktionen bilden können
 - БОЧКА, ГРУППА, ЗАВОД, ЯЩИК...
Коров было огромное стадо
- ТВОРОБСТ: Substantive mit der zeitlichen Bedeutung, die im Instrumental als adverbiale Bestimmung auftreten können
 - ВЕСНА, ВРЕМЯ, ПОРА...
приехать зимой, встречаться летней порой



Semantische Merkmale

- Richtigkeit der syntaktischen Konstruktionen hängt oft von den semantischen Kongruenz der Wörter ab.
- Beiordnende (сочинительные) Konstruktionen sind ein anderer Teil des semantischen Sprachsystems, wo semantische Kongruenz geprüft werden muss.
**Оборудование поступило на завод и в феврале*
- In der beschriebenen ETAP-Version werden 18 semantische Merkmale (Deskriptoren) verwendet
- Es werden hauptsächlich semantische Merkmale der Substantive und Verben beschrieben.



Semantische Merkmale

- Für die Prüfung der semantischen Kongruenz werden zwei Prädikate benutzt: CORDES und CODES
 - Bedingung CORDES der deskriptiven Kongruenz zwischen dem Hauptwort X und dem regierten Wort Y ist erfüllt, falls Y eins der Descriptoren besitzt, die in der entsprechenden Spalte des RMs von X aufgeführt sind, oder Y hat gar keinen Descriptor oder in der RM-Spalte von X sind keine Descriptoren aufgeführt.
 - Bedingung CODES der deskriptiven Kongruenz zwischen den Wörtern X und Y, die Bestand einer beiordnenden Konstruktion sind, ist erfüllt, falls X und Y mind. einen gemeinsamen Deskriptor haben oder einer von den beiden gar keinen Deskriptor besitzt



Deskriptoren in ETAP-3

- ВЕЛИЧИНА: Maßeinheiten wie *год, доллар, метр*
- ВЕЩЕСТВО: z.B. *воздух, газ, масло*
- ВРЕМЯ: Zeitwörter wie *будущее, десятилетие*
- ДЕЙСТВИЕ: Handlung, z.B. *работа, анализ, измерять*
- ИНФОРМАЦИЯ: z.B. *данные, опыт, уравнение, смысл*
- КРИСТАЛЛ: Substantive mit der Bedeutung eines Teils, einer Menge oder einer Klasse der Objekte z.B. *порция, слой, часть, КОМПОНЕНТ*
 - *Покрасить дверь голубой краской*
 - *Покрасить дверь двумя слоями голубой краски*
- ЛИЦО: eine Person oder Organisation – *агентство, оператор*
- МЕХАНИЗМ: Bezeichnungen der Geräte, Mechanismen wie z.B. *детектор, компьютер*



Deskriptoren in ETAP-3

- ОТНОШЕНИЕ: *контраст, граничить*
- ПАРАМЕТР: messbare Parameter wie *температура, цена, масса*
- ПРЕДМЕТ: unbelebtes Objekt
- ПРОСТРАНСТВО: ein Objekt, für das seine physische Größe relevant ist, z.B. *Аэропорт, область*
- ПРОЦЕСС: Situationen mit einem passiven Subjekt, dessen Zustand sich verändert – *рост, таять*
- СВОЙСТВО: langfristige Eigenschaften von Objekten, Prozessen, Handlungen wie *опасность, удобство*
- СОСТОЯНИЕ: возможность, нужда, лежать
- СФЕРА: Fachgebiete wie *биология, промышленность*
- ФАКТ: Ereignisse, z.B. *авария, происшествие*
- ЭНЕРГИЯ: Substantive wie *ток, ион, напряжение*



RM-Bereich

- Im RM-Bereich werden Rektionsmerkmale der Prädikate beschrieben, d.h. welche Anforderungen stellt das gegebene Prädikat an die Kandidaten für seine Argumente (Aktanten).
- Ein Rektionsmodell wird als eine Zeilenmenge dargestellt. Jede Zeile wird mit dem Zeichen $D_{i,j}$ eingeführt und enthält eine Term- oder Lexemmenge.

- RM für das Wort СРАВНЕНИЕ

D1.1:РОД, 'ЛИЦО'

D1.2:ТВОР

D2.1:РОД

D3.1:С3

D4.1:ПО1



Operationsinformation

- In einen Wortartikel vom KW können zwei Arten der Operationsinformation eingeschlossen werden:
 - Ein Verweis auf eine Schablonenregel oder
 - Eine ganze Regel
- Regeltypen:
 - Syntaktische Analyse
 - Normalisierung
 - Semantisierung



Beispiel zum Wortartikel

03918

ДАВЛЕНИЕ

POR S

SYNT СРЕДН, ЕД!, ПАРАМ, ХАРАКТТВОР
DES 'ДЕЙСТВИЕ', 'ПРОЦЕСС', 'ФАКТ'

D1 1 РОД

D2 1 НА1

D3 1 В1, 'ВЕЛИЧИНА'

D3 2 ИМ, 'ИЗМЕР'

ТРАФ АТРИБ 15



Literatur

- Ю.Д. Апресян et al.: Лингвистический процессор для сложных информационных систем. Москва 1992 (Kapitel 6.1)