

СЕМАНТИКА

I. Логическа форма и квази-логическа форма

I.1. Общ увод

Семантика - значение, теория за значението. Да разгледаме диалога:

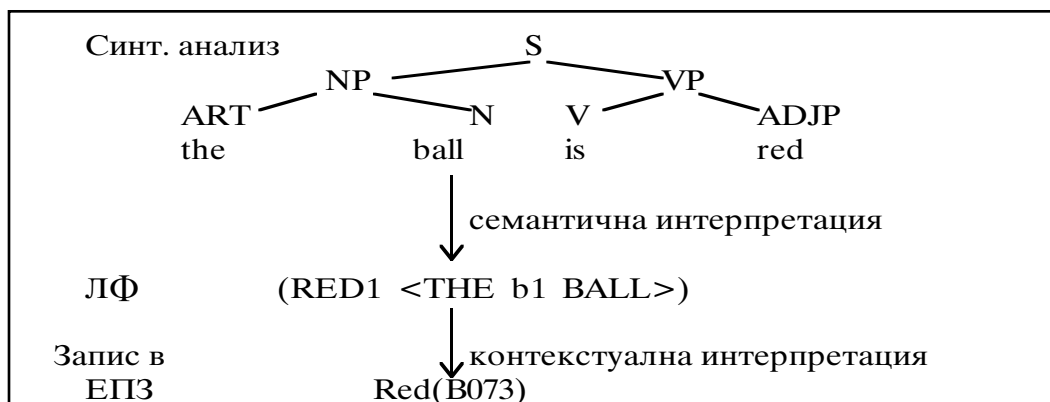
1a: А: Знаеш ли колко е часът?/Извинете, имате ли часовник? Б: Да.

1b: А: Колко е часът? Б: 2,15.

Очевидно въпроси (1a) на А са целяли получаване на отговор “2,15”. Виждаме, че значението на изреченията зависи от контекста на диалога.

Можем ли да дефинираме **значение на изречението**, което да не зависи от контекста? Т.е., има ли езиково ниво, на което изреченията 1a от горния диалог да имат по едно единствено значение, а то да се **употребява** по различни начини в различни контексти? Няма еднозначен теоретичен отговор, но КЛ дава практически отговор: разумно е да решим “ДА”, понеже имаме полза от такъв избор. *Модуларността е голямо улеснение*. Оказва се, че макар ЕЕ да е силно контекстно-зависим, има значителна част от семантиката, която се поддава и на самостоятелно третиране. Освен това, така отделяме изучаването на ЕЕ от общите теории за човешко поведение и умозаклучения в дадени контексти. Важни понятия:

Значение (meaning) ще се използва в контекстно-независим аспект, а **употреба (usage)** - в рамките на контекста. Представянето на контекстно-независимото значение се нарича **логическа форма ЛФ (logical form)**, а процесът на **изобразяване (mapping)** на изречението в неговата ЛФ - **семантична интерпретация (semantic interpretation)**. Изобразяването на ЛФ в крайния Език за Представяне на Знанията (ЕПЗ), в термините на който се осъществява “разбирането на ЕЕ”, се нарича **контекстуална интерпретация (contextual interpretation)**. Най-грубо ще имаме схемата:

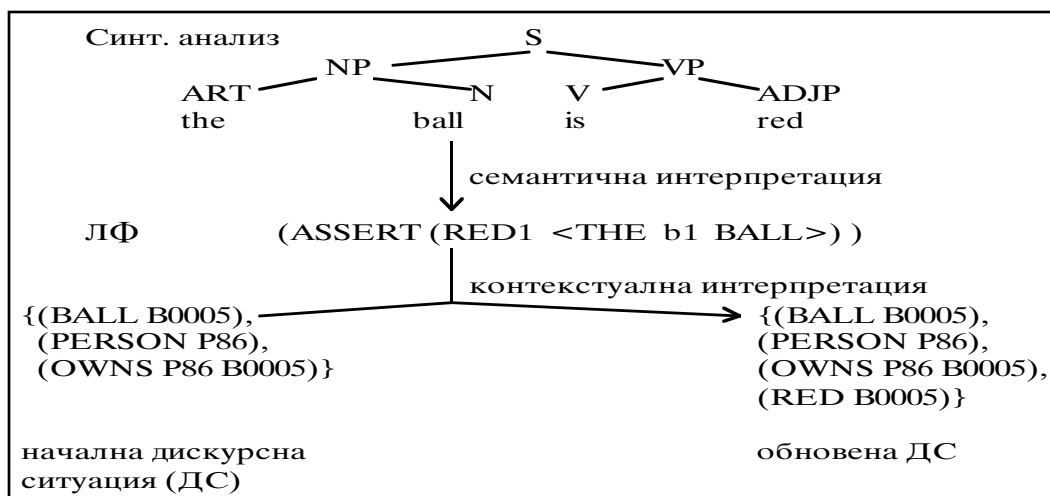


Дали езикът за представяне на ЛФ е ПСПР (Предикатно Смятане от Първи Ред), и какъв трябва да бъде ЕПЗ? Дали двата езика съвпадат? Във всеки случай, ние искаме в ЛФ да се кодират различни сложни явления, които предстои да бъдат “разрешени” в крайния ЕПЗ. Например, искаме ЛФ да кодира т. нар. **indexicals**, т.е. константи, определени от контекста. Такива са местоименията *АЗ*, *ТИ* и т. н. На практика всички членувани групи на съществителното (definite descriptions) като напр. *червената топка* са **indexicals**, тъй като са контекстно-зависими референции към даден обект в света. Изобщо, има много моменти (вкл. обхватът на кванторите), които ни се иска да кодираме в ЛФ, но да ги “отложим” за разрешаване при контекстната интерпретация. Така че не е необходимо ЕПЗ да бъде толкова сложен, както езикът на ЛФ. В повечето подходи ЕПЗ съвпада с ПСПР. Както ще видим по-долу обаче, езикът на ЛФ е по-сложен.

Но ако езикът на ЛФ не е ЕПЗ, тогава какъв е неговият формален статус? Ще отговорим на този въпрос с използване на понятието **ситуация** (situation): *множество от обекти и връзки между тях*. Пример за много проста ситуация е да имаме 2 обекта, човек P86 и топка B0005, и човекът да притежава топката (OWNS P86 B0005). Кодираме ситуацията като множество: SITUATION1 = {(B0005), (P86), (OWNS P86 B0005)}.

Чрез предаваната информация ЕЕ създава специфични ситуации. Нека предположим, че при всеки разговор има ситуация /на разговора/ (discourse situation), в която се записва предадената информация. Тогава всяко ново изречение се интерпретира в рамките на тази “стара” ситуация и произвежда “нова” ситуация, в която са записани старата и новата информация заедно. Тогава ЛФ е функция, която изобразява

“ситуацията, в която е направено изказването” в
 “ситуация-резултат на направеното изказване”.



Например, ако е налице SITUATION1 и “се случи” изречението *The ball is red*, може да се произведе нова ситуация, която се състои от старата SITUATION1 плюс още един факт, а именно (вж. и картинката по-горе):

SITUATION2 = {(B0005), (P86), (OWNS P86 B0005), (RED B0005)}.

Сега започваме да се занимаваме с компютърна обработка на тази част от значението, която е контекстно-независима.

I.2. Значения на думите и многозначност (ambiguity)

И лингвистиката отделя някои значения като контекстно-независими. Например за глагола забивам в Тълковен речник (1994) имаме:

- забивам**
1. Започвам да бия. *Камбаните забиха посред ноц.*
 2. Втъквам, забучвам нещо с биене, с удар; забучвам, забождам.
В село е влязла, байрак забила. Заби нож в сърцето ми.
 3. Хващам, тръгвам.
Дай си, мъжо, жената, а ти забий в гората!

Виждаме три различни значения, имат и различни синоними според значението. Така че “забивам” е семантично многозначна дума (semantically ambiguous), тъй като има повече от едно значения. От друга страна, съществителното **кон** според същия речник има две значения:

- кон** ...
1. Едро еднокопитно домашно животно за езда и впрягане.
 2. В шахмата - фигура, която

Но “кон” означава “жеребец”, “кобила”, “впрегатен кон”, “товарен кон”, “състезателен кон”. Стигаме до важното наблюдение, че някои **значения на думи (word senses)** са по-широки, а други - по-специфични. Това свойство на ЕЕ се нарича **неопределеност (vagueness)**. Освен това в многоезиков аспект, думите от различните ЕЕ “фрагментират” семантиката на света според различна скала на “грануларност”. Например, някои индианци от джунглите имат десетки думи за “червено”, а някои ескимоси - десетки думи за видове сняг. Ние нямаме толкова, за нас тези обекти не са важни и ние не ги различаваме, поради което не сме ги нарекли с отделни названия.

Има ли лингвистични тестове, с помощта на които да различаваме **многозначността** (повече отделни значения) от **неопределеността**? Имаме нужда от такива тестове, за да можем да определим *колко смисъла има дадена дума*, а тогава бихме знаели например как да си строим семантичната част на лексиконите. Предложени са някои тестове (за английския език). Използва се фактът, че в определени синтактични позиции се налага употреба на едно или друго значение или е възможна употребата на по-широко значение. Български примери по тези идеи:

Аз виждам две вериги, а ти - три.

Аз виждам два коня, а ти - три.

Първото изречение не може да бъде съкратена форма на “аз виждам две вериги на колело, а ти - три вериги от войници”. Но второто изречение е вярно като обобщение на “аз виждам два жребеца, а ти - три впрегатни коня”. Така *верига* е многозначна с поне два смисъла “1. верига на колело” (ВЕРИГА1) и “2. верига от хора” (ВЕРИГА2), а *кон* - по-обща от “жеребец” и “впрегатен кон”.

Друг примерен тест за глаголи: “Аз оставям Иванка и ти правиш същото с Любка” не може да се прочете като “аз се *развеждам* с Иванка, а ти *завеждаш* Любка до работата и *тя остава там*”. Значи *оставям* е поне двузначна дума. От друга страна, “аз целувам Иванка и ти правиш същото с Любка” може да означава, че аз целувам Иванка по бузата, а ти целуваеш ръка на Любка. Така *целувам* е неопределен по отношение на мястото, на което се извършва действието.

Освен семантичната многозначност на думите, която трябва да се обработи при семантичната интерпретация, има и структурна многозначност. Отчасти тя е производна от синтактичната многозначност:

Аз видях момичето с телескопа.

Във фермата живеят щастливи кучета и котки.

Но понякога семантична многозначност настъпва и при единствена синтактична структура: *Едно куче влизаше с всяко дете*.

Има и неопределени значения, напр. кванторите или думи, които играят ролята на квантори: *Много хора видяха инцидента*. (колко и кой?).

Да споменем отново случаите, в които дадени значения не са (все още) специфицирани (*аз, вчера, тук, червената топка* и т.н. indexicals). Но **неспецифицираност** не значи, че са **неопределени** в горния смисъл.

Важен аспект на контекстно-независимите значения е фактът, че понякога тяхното съвместно случване води до фиксиране на едно от значенията. Например, ако в лексикона имаме ВЕРИГА1 и ВЕРИГА 2, срещането на “войници” след “верига” веднага би органичило {ВЕРИГА1, ВЕРИГА 2} до ВЕРИГА 2 в ЛФ. Друг път синтактичната структура забранява употребата на едно или друго значение: “*аз забих ножа* (ЗАБИВАМ2 е преходен глагол), *а ти заби в гората* (ЗАБИВАМ3 е непреходен глагол)”. При семантичната интерпретация е много важно да се намерят подходящи филтри, чрез които да се редуцира броят на разглежданите значения в даденото изречение. В примера {ЗАБИВАМ1, ЗАБИВАМ2, ЗАБИВАМ3}, при синтактичния анализ стойността на признака SUBCAT на 3-те значения ще бъде използвана като филтър.

Така стигаме до идеята, под значение на думата в лексикона да разбираме идентификатор, произхождащ от самата дума (с цифра на края,

съответстваща на номера значение). Например 3-те значения на “забивам” в нашия лексикон са {ЗАБИВАМ1, ЗАБИВАМ2, ЗАБИВАМ3}. Колкото обаче и да ограничаваме значенията, с които трябва да работим при всяка дума, пак обаче остават твърде много случаи за комбиниране. Така че някои от значенията ще формулираме в по-общи класове, организирани в **онтология**. Важни понятия са **действия** и **събития**, както и **ситуация**. Има и **състояния**. По-нататък ще говорим за това по-подробно.

1.3. Основни елементи в езика за записване на ЛФ

На този етап на разглеждане, примитивните значения са тези на думите (word senses). От тях изграждаме значение на изречението (sentence meaning). Значенията на думите са **атоми** или **константи** в езика на ЛФ.

Константите, които описват обекти в света, са **терми**. **Предикати** са константите, които описват връзки между обектите или свойства. **Израз** е предикат, следван от подходящ (според ариостта) брой терми-**аргументи**. Пример: изречение “Джафчо е куче”. Имаме терма-константа “Джафчо1” и терма едноместен предикат КУЧЕ1 (DOG1). Получаваме израза:

(DOG1 Джафчо1)

И така свойствата са едноместни предикати. Има и двуместни предикати:

(ОБИЧА1 Иван1 Мария1)

Различните класове думи играят различна роля в езика на ЛФ. *Собствените имена* стават константи-терми; *съществителните нарицателни* - едноместни предикати; *глаголите* - n-местни предикати, където n е броят на категориите, за които глаголят има дефиниран признак “subcategorization” плюс подлога.

За по-сложни изрази, въвеждаме нови класове константи: имена на **логически оператори**. За повечето от тях имаме думи на ЕЕ, така че както досега можем да означаваме фрагменти значение с идентификатори-думи:

Джак не обича Сю: (NOT1 (LOVES1 JACK1 SUE1))

Джак обича Сю или Мери:

(OR1 (LOVES1 JACK1 SUE1) (LOVES1 JACK1 MERY1))

За конюнкцията имаме ЕЕ-дума (съюз “и”), но за импликацията например нямаме една дума. Ще прибягваме и до повече думи.

Квантори: в ПСПР имаме два квантора (за съществуване и всеобщност), но в ЕЕ редица думи се държат като квантори: *много*, *малко*, *няколко* и т.н. Въвеждаме променливи, които ще се менят в обхвата на кванторите, но с една важна разлика. В ПСПР променливите в две формули се третират като

различни, например x в две формули е име на две променливи. Но в ЕЕ забелязваме по-специално поведение на т. нар. **дискурсни променливи**. Например: *Един човек влезе в стаята. Той мина до масата и ...* Първото изречение въвежда “променлива-човек”, т.е. променлива, която се мени в “област” ЧОВЕК. Искаме тази променлива да продължи да съществува и във второто изречение, където е цитирана с думата *той*.

За да позволим това, решаваме, че всяка нововъведена дискурсна променлива (discourse variable) получава единствено име. По-нататък, в следващите изречение, ще го цитираме.

В ПСПР, $\forall x P(x)$ е вярно само когато за всяка стойност x от областта на x имаме $P(x)$. Изказвания в такъв стил са редки в ЕЕ. Казваме “всички кучета лаят”, но също така често казваме “повечето хора се смеят”. Имаме нужда от нови конструкции: **обобщени квантори** (generalized quantifiers). Дефинираме наредената четворка:

(<име-квантор> <променлива>: <израз - област> <израз-тяло>) и така:
(MOST1 d1: (DOG1 d1) (BARKS1 d1))

“повечето обекти d1, които удовлетворяват (DOG1 d1), удовлетворяват и (BARKS1 d1)”. Забележете, че размяната на 3-ти и 4-ти елемент в наредената четворка би довела до изречението “повечето лаещи обекти са кучета”. Така че позицията е от значение при тези дефиниции.

Важни обобщени квантори са пълният и краткият член, THE и A:

Кучето лае: (THE x: (DOG1 x) (BARKS1 x))

Това ще бъде вярно само ако има еднозначно определено куче в дискурсна ситуация (и то лае). Ето таблица на кванторите-членове:

Квантор	Използване	Пример
THE	определена референция	the dog
A	неопределена референция	a dog
BARE	празен за NP - mass term (ед. ч-ло)	water, food
BARE	празен за множ. число NP - generic	dogs

Примери на по-сложни изречения: The happy dog barks.

(THE x: (& (DOG1 x) (HAPPY1 x) (BARKS1 x))

Променливата x се мени в областта на тези кучета, които са щастливи.

Множественото число също заслужава специално внимание. Имаме нужда от оператор, който от дадена област (напр. КУЧЕ) да дефинира област от множества от нейни членове. Въвеждаме нова конструкция - т.нар. **предикатен оператор** (predicate operator), който има за аргумент предикат и за резултат друг предикат:

(PLUR DOG1) - това е предикат, който е истина за множество кучета. Тогава “кучетата лаеха” се записва като

(THE x: ((PLUR DOG1) x) (BARKS1 x)).

Предикатният оператор за “множествено число на дискурсната променлива” ни кара да забележим нов вид многозначност: т. нар. *колективни и дистрибутивни четения*:

Двама мъже купиха видео. (Двамата едно видео или две?)

За някои изречения е възможно само колективно четене: “*те се срещнаха на ъгъла*”. Можем да си представим и *респективно* четене при някои изречения: “*Иван и Стоян обичат Пенка и Марийка*”, т.е. имаме две дискурсни променливи с множествени числа.

Завършваме дефиницията на езика на ЛФ с т. нар. **модални оператори**. Те приличат на логическите оператори, но с едно важно ограничение - не може свободно да се заместват еквивалентни термове, когато са в обхвата на модален оператор. Това явление се нарича **невъзможност за заместване** (failure of substitutivity). Разглежда се като определяща характеристика на модалните оператори. Примери:

Нека JACK1=JOHN22, и нека поне някои хора го знаят. За тях, (HAPPY1 JACK1) е вярно винаги когато е вярно (HAPPY1 JOHN22). Така е и с по-сложните изрази:

(OR (HAPPY1 JACK1) (SAD1 JACK1)) е вярно винаги когато е вярно
(OR (HAPPY1 JOHN22) (SAD1 JOHN22)).

Нека обаче разгледаме модалния оператор
(BELIEVE1 SUE1 (HAPPY1 JACK1)).

Не можем свободно да сменим JACK1 с JOHN22, понеже не сме наясно дали Сю знае за тяхната идентичност (трябва да погледнем нейния модел на света, за да сме сигурни, че имаме право на такава субституция).

Подобни свойства показват и **операторите за глаголно време** (tense operators). Дефинираме ги като оператори, които стоят пред израза:

(PRES (SEE1 JOHN1 SUE1)), т.е. Джон вижда Сю,
(PAST (SEE1 JOHN1 SUE1)), т.е. Джон видя Сю,
(FUT (SEE1 JOHN1 SUE1)), т.е. Джон ще види Сю.

Те са модални, понеже и при тях наблюдаваме невъзможност за заместване. Например, ако сега JOHN1=PRESIDENT1, съгласно горното щом Джон вижда Сю, то и президентът я вижда. Но не можем да напишем същото твърдение за PAST или FUT вместо за PRES - т.е. не можем да заместим JOHN1=PRESIDENT1 в PAST или FUT-изразите. Освен това нека отбележим интересния факт, че е възможно да са едновременно верни

(PAST (HAPPY1 JACK1)) и (PAST (NOT1 (HAPPY1 JACK1))), понеже в миналото има различни моменти.

I.4. Кодирание на многозначността в ЛФ

Ще компресируем някои многозначности в ЛФ, за да ги запазим за по-късно разрешаване, така ще избегнем твърде голямото нарастване на броя на възможните ЛФ за дадено изречение (сериозен проблем). Често ЛФ с кодирана многозначност се нарича **квази-логическа форма (КЛФ)**.

Първи вид компресия: при наличие на повече значения, които изглеждат еднакво възможни на етапа на синтактичния анализ на изречението. Например, Sue watched the ball. Нека BALL1 е топка, а BALL2 - бал (в зала). Така че като област на дискурсна променлива можем да напишем:

(THE b1: ({BALL1 BALL2} b1) (PAST (WATCH1 SUE1 b1)))

Сега дискурсна променлива b1 се мени в множество от значения (нещо подобно имаме и при признаковите структури). Така че КЛФ съдържа {} (чупени скоби).

Друг вид многозначност възниква поради обхвата на кванторите, например в изречението *всяко момче обича едно куче*. Двете възможни ЛФ са:

(EVERY b1: (BOY1 b1) (A d1: (DOG1 d1) (LOVES1 b1 d1))) Æ
(A d1: (DOG1 d1) (EVERY b1: (BOY1 b1) (LOVES1 b1 d1)))

Ще съкратим двата записа в един в КФЛ чрез символите <>:

(LOVES1 <EVERY b1 (BOY1 b1)> <A d1 (DOG1 d1)>)

После, при контекстната интерпретация, ще разрешаваме обхвата. Има най-различни примери за многозначност поради обхвата на кванторите, като “във всеки хотел портиерът беше готин” (членът THE на портиера трябва да влезе в обхвата на “всеки хотел”, но това не би се случило при “всяко момче обича кучето”).

По-надолу ще използваме по-опростени формати винаги когато може: напр. вместо <EVERY b1 (BOY1 b1)> може <EVERY b1 BOY>.

Както отрицанието, така и операторите за глаголно време имат обхват.

Например: every boy did not run има двете възможни ЛФ

(NOT (EVERY b1: (BOY b1) (RUN1 b1))) и
(EVERY b1: (BOY1 b1) (NOT (RUN1 b1)))

Съкращаваме като:

(<NOT RUN1> <EVERY b1 BOY1>).

На края ще разгледаме представяне на собствените имена и местоименията. В контекста на много ситуации, собствените имена могат да се окажат недостатъчни за идентификация. Въвеждаме предикат-име:

(NAME <променлива> <константа-име>).

Тогава *John ran* се записва: (<PAST RUN1> (NAME j1 "John"))

Местоимения: (PRO <променлива> <израз>). Пример:

Every man liked him

(<PAST LIKE1> <EVERY m1 MAN1> (PRO m2 (HE1 m2)))

Триковете за компресия, които описахме тука, доста намаляват броя на КЛФ, които ще подлежат на разглеждане в процеса на контекстуалната интерпретация за всяко отделно изречение.

I.5. Глаголи и състояния в ЛФ

Досега присвоявахме на глаголите имена на предикати, но не сме разглеждали по-подробно въпроса за “арността” на предикатите. Има един много важен аспект, на който се спираме в момента: че “арността”/”валентността” може да се мени, или поне някои изречения на пръв поглед ни водят към тази мисъл. Пример:

John broke the window with a hammer.

The hammer broke the window.

The window broke.

Интуитивно е ясно, че трите изречения описват едно и също събитие - но с различна степен на общност. Тогава, дали става дума за едно и също значение на глагола BREAK, или имаме 3 различни, които го описват като едноместен, двуместен и триместен предикат?

Едно решение е да приемем трите възможни ЛФ, а именно:

1. (<PAST BREAK1>

(NAME j1 "John") <THE w1 WINDOW > <THE h1 HAMMER1>)

2. (<PAST BREAK2> <THE h1 HAMMER1> <THE w1 WINDOW >)

3. (<PAST BREAK3> <THE w1 WINDOW >)

и да дефинираме т.нар. аксиоми за значението (meaning postulates): Ако е вярно 1, то е вярно 2; ако е вярно 2, то е вярно 3. Обаче това е твърде тромаво решение. През 1967 Дейвидсън предлага следната идея: да разгледаме изреченията като дефиниции на **събития (events)** или **състояния (states)**; тогава можем да въведем променлива за събитието или състоянието и да имаме (в сегашно време, без излишно усложняване на примера):

($\exists e1$: (BREAK1 e1 (NAME j1 "John") <THE w1 WINDOW >))

или

($\exists e1$: (BREAK1 e1

(NAME j1 "John") <THE w1 WINDOW > <THE h1 HAMMER1>))

Сега, към събитието “счупване на прозорец с чук от Джон” ще добавяме или махаме повече елементи - но имаме само едно събитие и само едно значение на BREAK.

Така сме отговорили на въпроса колко значения на BREAK са ни необходими, но не сме обяснили, какво става всъщност с “арността” на BREAK като предикат в трите начални примера? Отговор ни дават т. нар. **тематични роли (семантични роли, дълбинни падежи, падежни роли, валенции и актанги)**, които за първи път се споменават от Филмор като стройна лингвистична теория през 1968. Интуитивно, в трите изречения Джон, чукът и прозорецът играят винаги една и съща роля: Джон е агент, счупващ прозореца, прозорецът е обект, понасящ действието, чукът е инструмент, извършващ действието. Това предполага представяне на изречението във вид, подобен на семантичните мрежи, където връзките между обектите се разпадат до връзки между един и два обекта. Имаме:

(∃ e: (& (BREAK e) (AGENT e (NAME j1 “John”))
(THEME e <THE w1 WINDOW >)
(INSTR e <THE h1 HAMMER1>)))

тоест въвеждаме запис:

(∃ e: (& (Event-predicate e) (Relation1 e Object1) (Relation2 e Object2)
..... (Relation_k e Object_k)))

за k-местен предикат Event-predicate. По-съкратено ще го записваме като:

(Event-predicate e [Relation1 Object1] [Relation_k Object_k]),
а в КЛФ за горното конкретно изречение:

(<PAST BREAK> e [AGENT (NAME j1 “John”)]
[THEME e <THE w1 WINDOW >])).

Състояния (states): едноместни предикати. Пример: *Мери беше нещастна:*

(∃ s: (<PAST UNHAPPY> s) (THEME s (NAME m1 “Mary”))))

или *Мери беше нещастна на срещата* в по-съкратения запис от по-горе:

(<PAST UNHAPPY > s [THEME (NAME m1 “Mary”)]
[IN-LOC <THE p1 MEETING >])).

Досега сме разглеждали записи на ЛФ на три нива на грануларност:

a) (PRES (SEES1 (NAME j1 “John”) (NAME m1 “Mary”))))

b) (PRES (SEES1 e1 [AGENT (NAME j1 “John”)]
[THEME (NAME m1 “Mary”)]))))

c) (PRES (∃ e: (& (SEES1 e) (AGENT e (NAME j1 “John”))
(THEME e (NAME m1 “Mary”))))))

Ще използваме и трите вида запис, според удобството в конкретните случаи. Различаването на тематичните роли е основа на няколко формализма за семантично представяне и се счита за удобна форма на фрагментиране на връзките между обектите.

I. 6. Тематични роли

Дотук въведохме неформално AGENT, THEME и INSTR. Сега ще разгледаме по-систематично някои роли.

AGENT изглежда лесен за дефиниране. Фраза NP “запълва ролята AGENT” в даден глагол, ако описва одушевен извършител на съзнателно действие. Ето някои тестове как да разпознаем ролята AGENT: допустимо ли е изречението, образувано с добавяне на “*съзнателно*” или “*с цел да, за да*”:

- (1) Иван съзнателно счупи прозореца (допустимо), или
- (2) Иван счупи прозореца, за да проветри стаята (допустимо).
- (3) Чукът счупи съзнателно прозореца (недопустимо), или
- (4) Чукът счупи прозореца, за да проветри стаята (недопустимо).

Значи, Иван запълва ролята AGENT в изречения 1 и 2, но чукът - не в 3 и 4.

Не всички одушевени подлози са в дълбинната роля AGENT. Например:

- (5) Иван съзнателно умря. (недопустимо).
- (6) Мария си спомни за рождения си ден, с цел да получи подаръци (не).
- (7) Петър съзнателно мисли, че земята е плоска. (недопустимо)

Това е ролята EXPERIENCER - лице, което е подложено на действието на възприемане или се намира в дадено психологическо/физическо състояние.

Да разгледаме THEME: запълва я обект, който е подложен на действието или го понася. За преходни глаголи това е прякото допълнение:

Дадох книгата на Иван. (книгата запълва THEME на *давам*)

Камъкът се счупи. (камъкът запълва THEME на *счупвам се*)

Иван е висок. (Иван запълва THEME на *съм*)

Група роли описват местоположението във време, място, стойност, притежание: групирани са в ролята AT и нейните под-роли (вж. таблицата).

При движение, имаме начално положение (FROM) и крайно (TO).

Аз вървя от тук (FROM-LOC) до училище (TO-LOC) по улицата (PATH).

Взех кола в 3 ч. (AT-TIME) и я карах от 4 (FROM-TIME) до 6 (TO-TIME).

Купих книга за Петър (TO-POSS) от Иван (FROM-POSS).

Пращам колет на Мария (BENEFICIARY, получател).

Интересна роля е INSTR (виж таблицата), често я срещаме като дълбинна роля на повърхнинния подлог:

Слънцето (INSTR) изсуши прането.

Още няколко роли:

Иван вдигна пианото с Петър (CO-AGENT).

Иван плати на Петър 5 лв. за книгата. (5 лв. - THEME, за книгата - CO-THEME).

Иван купи книгата от Петър за 5 лв. (книгата - THEME, 5 лв. - CO-THEME).

По-общо решение е да “разложим” сложната ситуация “давам пари и вземам книга” на две по-елементарни ситуации, но тази задача е доста встрани от нашите разглеждания за разбиране на ЕЕ.

Както виждаме на таблицата (във файл **lect5_table.pdf**), има известна зависимост между синтактичната роля на дадена фраза в изречението и нейната дълбинна семантична роля (стълб 4).

I. 7. Речеви актове и вложени изречения

Както знаем, изреченията се използват по различни начини. Всяка употреба обозначава определена връзка между говорителя и информационното съдържание на изказването му. В ЛФ ще въведем оператори, които имат за област интерпретацията на изреченията и за резултат т. нар. **повърхнинни речеви актове** (surface speech acts). Тези оператори задават какво е намерението на говорителя за използване на описаното в КЛФ изказване, и как това намерение ще повлияе (обнови) текущата ситуация на дискурса. Ще различаваме 4 повърхнинни речеви акта:

ASSERT - изказването да се отрази (включи) в дискурсната ситуация.

Y/N QUERY - въпрос да/не, т.е. вярно ли е, че

COMMAND - изказването описва действие, за което говорителят иска да се изпълни

WH-QUERY - изказването описва обект за идентификация (кой, кого, но не ЗАЩО!)

Така за разказни изречения пълната ЛФ е (вече сме я виждали на една картинка): *The man ate a peach:*

```
(ASSERT (<PAST EAT1> e1 [AGENT <THE m1 MAN1> ]
        [THEME <A p1 PEACH1>] ))
```

За въпросителни изречения “вярно ли е, че?”

Did the man eat a peach?

```
(Y/N QUERY (<PAST EAT1> e1 [AGENT <THE m1 MAN1> ]
            [THEME <A p1 PEACH1>] ))
```

Императив: *Eat a peach:*

```
(COMMAND (EAT1 e1 [THEME <A p1 PEACH1>] )) и т.н.
```

При вложени изречения влагаме ЛФ: *The man who ate a peach left*

```
(ASSERT (<PAST LEAVE1> e1 [AGENT <THE m1
```

(& (MAN1 m1)
 (<PAST EAT1> e2 [AGENT m1] [THEME <A p1 PEACH1>])
 >))

Сега сме дефинирали ЛФ достатъчно подробно за всякакъв вид изречения.

II. Връзка между синтаксиса и семантика

Ще композираме значението на фразовите конституенти така, както композираме синтактичната им структура. Всяко синтактично правило в граматиката ще има съответно семантична интерпретация, поради което този подход се нарича често “**семантична интерпретация правило по правило**” (rule-by-rule semantic interpretation).

II. 1. Композиция и λ -смятане

Кои са предимствата на композиционния подход? (1) Интерпретацията на цялото се строи с последователни нараствания от интерпретацията на частите; (2) Оказва се, че композиционните модели правят граматиката по-лесна за разширяване и поддържане.

Но не е лесно да се създаде композиционен модел. Първи проблем: за много изречения, така дефинираната ЛФ изглежда твърде различна от синтактичната им структура. Вече говорихме за класическия пример: кванторите. Да разгледаме пак изречението *Иван обича всяко куче*. Синтактичните конституенти групират думите във фрази както е показано със скобите:

((Иван) (обича (всяко куче))).

Тоест, кванторът се държи “като прилагателно” на синтактично ниво. Но единствената ЛФ би била:

(EVERY d: (DOG1 d) (LOVES1 e1 (NAME j1 "John") d)).

(за всяко куче d има събитие e1, което е “обичане на d от Иван”).

Няма прилика в линейната последователност и влагането на термовете в синтактичното и семантичното представяне. В ЛФ, кванторът е изразен като най-външен оператор, в чиято област стои глаголният предикат LOVES1 (а в синтактичната конструкция включването е обратно); освен това променливата d в LOVES1 е свързана с EVERY (тоест, кванторът е предизвикал появата на два - доста отдалечени - елемента от ЛФ). Затова решаваме, че композиционно ще строим само КЛФ, тоест т. нар. **unscoped logical form**, при която има повече прилика между синтактичната и семантичната структури:

(LOVES1 e1 (NAME j1 "John") <EVERY d: DOG1>).

Втори проблем са идиомите, които присвояват значение на повече от една дума. Например, *Иван хвърли топа*. Тука “хвърлям топа” има значение УМИРАМ1. Отбележете, че това изречение е двузачно (вторият му “дума-по-дума” смисъл се вижда веднага в страдателен залог: *топът беше хвърлен от Иван*, при което идиомът вече не се разпознава). Идиомите са важна част от ЕЕ, но ние разглеждаме само присвояване на примитивни значения на ЕДНА ДУМА (а не на цели фрази, което е начинът за обработка на идиоми).

Остава да споменем (и да решим) следния проблем: глаголните предикати имат различна “арност” според тематичните си роли, но на нас може да ни се наложи да ги комбинираме във фрази с друга арност. Например, LOVES1 е двуместен предикат. Но от него ние трябва да композираме едноместната фраза на глагола “X обича Мария”, в която е запълнена само ролята THEME, а ролята AGENT предстои да бъде запълнена от подлога. Значи трябва да се справим с променливата “арност” на конституентите, за да можем да строим конституенти на различни нива на композиция. Удобен формализъм за това предлага т. нар. λ -смятане. Пример за λ -израз:

$$(\lambda x (\text{LOVES1 } l1 \ x \ (\text{NAME } m1 \ \text{"Mary"}))).$$

(X обича Мери, x е като параметър).

Този предикат е истина за всеки обект o , който заместен вместо x дава израз-истина. Както за всеки друг предикат, можем да конструираме предикатно-аргументен израз, т.е. предикат последван от аргумент:

$$(\lambda x (\text{LOVES1 } l1 \ x \ (\text{NAME } m1 \ \text{"Mary"}))) (\text{NAME } j1 \ \text{"John"}).$$

Този израз е истина, ако аргументът (NAME j1 "John") удовлетворява λ -израза, тоест ако

$$(\text{LOVES1 } l1 \ (\text{NAME } j1 \ \text{"John"}) \ (\text{NAME } m1 \ \text{"Mary"}))$$

е истина. Казваме, че последният израз е получен чрез **прилагане** на λ -израза над аргумента (NAME j1 "John"). Тази операция е **λ -редукция**.

По-формално: в λ -смятането

$((\lambda x \ P x) a) = P \{x/a\}$, където P е произволна формула, съдържаща x и $P \{x/a\}$ е същата формула, в която x е заместено от a . От тука имаме двете основни операции: λ -редукция (\rightarrow) и λ -абстракция (\leftarrow).

Тези λ -изрази много ще ни помагат да поддържаме четлив запис на процесите на композиране на КЛФ, но нямат отношение към самата идея за композиционна семантика.

При по-големи граматики композиционният подход става все по-атраکتивен, тъй като започват да се обработват по-големи групи изречения. Да разгледаме едно първо улеснение от въвеждането на λ -изрази като модел на глаголни предикати:

Сю пее и отваря вратата.

Тука имаме една сложна глаголна група, “пее и отваря (вратата)”. Това са два едноместни λ -израза “някой пее” и “някой отваря вратата”:

$(\lambda a (\text{SINGS1 } e1 a))$ и $(\lambda a (\text{OPENS1 } o1 a \text{ <THE d DOOR1>}))$.
 От тези два израза с проста трансформация произвеждаме
 $(\lambda a (\& (\text{SINGS1 } e1 a) (\text{OPENS1 } o1 a \text{ <THE d DOOR1>})))$,
 което може да бъде семантиката на споменатото по-горе сложно VP. Така
 че с въвеждането на подходящ запис (а именно λ -изрази) започваме да се
 възползваме и от предимствата на трансформациите на този запис.

Всяка главна синтактична фраза съответства на определена семантична конструкция. VP и PP са обикновено едноместни предикати, най-често сложни λ -изрази. NP са термове, производни на съществителни, имена и местоимения. По-малките категории имат за семантична интерпретация изрази, които изграждат семантиката на “по-големите” категории. Сега в момента на свързване например на NP с VP вече не ни е нужно да знаем структурата на VP, а само ще заместим NP-терма в едноместния λ -израз.

II.2. Примерна граматика и лексикон със семантични признаци

Ще илюстрираме как КЛФ се изчислява чрез признакови структури в процеса на синтактичния анализ. За да запазим примерите по-прости, КЛФ няма да съдържа тематични роли. Към лексикона добавяме признак SEM, както и към всяко правило на граматиката (вж. примерния лексикон и граматика). Да обсъдим правило 1 на граматиката:

$$\begin{aligned} (S \text{ SEM } (?semvp \ ?semdp)) &\longrightarrow \\ (NP \text{ SEM } ?semdp) & \\ (VP \text{ SEM } ?semvp) & \end{aligned}$$

Нека NP има SEM терма (NAME m1 "Mary"), VP - израза

$$(\lambda a (\text{SEES1 } e8 a (\text{NAME } j1 \text{ "John"}))$$

Тогав SEM на S се изчислява като просто ги написваме един след друг и извършваме λ -редукцията с цел опростяване:

$$\begin{aligned} (\lambda a (\text{SEES1 } e8 a (\text{NAME } j1 \text{ "John"})) (\text{NAME } m1 \text{ "Mary"}) &= \\ (\text{SEES1 } e8 (\text{NAME } m1 \text{ "Mary"}) (\text{NAME } j1 \text{ "John"})) & \end{aligned}$$

Примерната проста граматика (в която сме пропуснали другите признаци, които разгледахме в по-ранни примери), освен SEM съдържа и VAR. VAR е нов признак, който запазва дискурсната променлива, съответстваща на конституентата. VAR се генерира автоматично от парсера и се поддържа единствена за дадена конституента.

Правила 2 и 3 се отнасят за VP-групи. При тях във VAR се поддържа променлива за събитието или състоянието (знаем, че глаголната група описва събитието или състоянието в едно изречение). В правила 2 и 3 е илюстрирано как се “появяват” λ -изразите: тяхното конструиране се управлява от граматиката. Правила 2 и 3 “вземат” семантиката на глагола ?semv и конструират λ -изрази като едноместни термове за прилагане към подлога. Правила 4 и 5 са за местоимения и имена. Правило 6 дефинира израз, който съдържа квантор с неразрешен обхват. Задават се вида на

квантора λ semart, променлива λv и полето на действие на квантора. Правило 7 построява CNP (common NP от съществително нарицателно).

В лексикона SEM задава идентификаторите, кодиращи възможните значения на всяка дума. Имаме различни значения на думите: например имаме FISH в ед. число и FISH в множествено число, както и MAN и MEN (забележете предиката за множествено число в лексикона). Имаме една дума без SEM: това е “to”, което е частицата за свързване на два инфинитива. Тя не “внося” своето значение при построяване семантиката на изречението, а само управлява синтактичната правилност на съчетанието от два глагола.

За глаголите често различните значения имат различни SUBCAT стойности. Например DECIDE1 в лексикона (решавам) има SUBCAT _none, и след него следва инфинитив. Това е значението, което се използва в изречения като “решавам да пия”. Друго значение DECIDE2 изисква SUBCAT с предложна фраза, започваща с предлога on: това е значението, което се употребява в изречения като “we have decided on the new computer” (ние решихме /избрахме, спряхме се/ за новия компютър).

По-долу е дадено едно примерно дърво на разбора в тази граматика и лексикона. Забележете, че “листата” на дървото се формират отчасти с правила на граматиката, отчасти с преформатиране на данните от лексикона. В дървото са отразени за всяка конституента нейните SEM и VAR стойности.

За да извършваме семантична интерпретация с признака SEM, сега са ни необходими две прости промени в разглежданите до тук синтактични анализатори:

- при всяко първо (ново) прилагане на правило, VAR се инициализира с нова дискурсна променлива;

- винаги когато се построява конституента, се прави опит за λ -редукция на получения семантичен израз.

Един примерен лексикон с признак SEM:

```

a (art AGR 3s SEM INDEF1)
can (aux SUBCAT base SEM CAN1)
car (n SEM CAR1 AGR 3s)
cry (v SEM CRY1 VFROM base SUBCAT _none)
decide (v SEM DECIDES1 VFORM base SUBCAT _none)
decide (v SEM DECIDES-ON1 VFORM base SUBCAT _pp:on)
dog (n SEM DOG1 AGR 3s)
fish (n SEM FISH1 AGR 3s)
fish (n SEM (PLUR FISH1) AGR 3p)
house (n SEM HOUSE1 AGR 3s)
has (aux VFORM pres AGR 3s SUBCAT pastpart SEM PERF)
he (pro SEM HE1 AGR 3s)
in (p PFORM {LOC MOT} SEM IN-LOC1)
Jill (name AGR 3s SEM “Jill”)

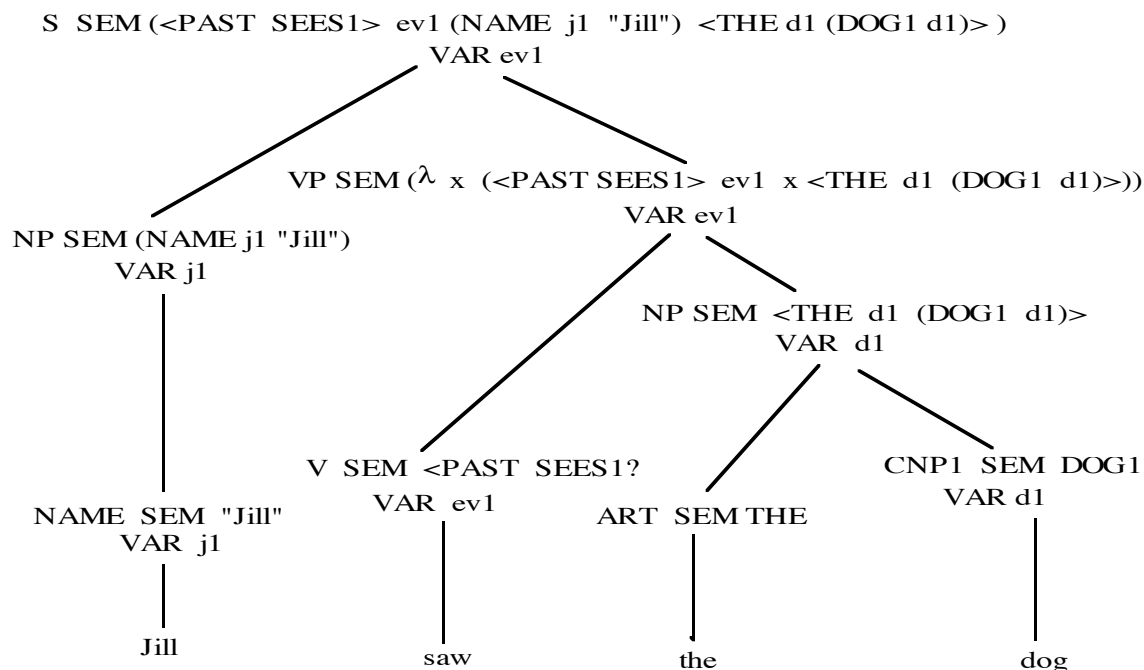
```

man (n SEM MAN1 AGR 3s)
 men (n SEM (PLUR MAN1) AGR 3p)
 on (p PFORM LOC SEM ON-LOC1)
 saw (v SEM SEES1 VFORM past SUBCAT _np AGR ?a)
 see (v SEM SEES1 VFORM base SUBCAT _np IRREG-PAST+ EN-PASTPRT +)
 she (pro SEM SHE1 AGR 3s)
 the (art SEM THE AGR {3s 3p})
 to (to AGR - VFORM inf)

БКГ, разширена със признаци за семантика:

1. (S SEM (?semvp ?semnp)) → (NP SEM ?semnp) (VP SEM ?semvp)
2. (VP VAR ?v SEM (λ a2 (?semv ?v a2))) → (V[_none] SEM ?semv)
3. (VP VAR ?v SEM (λ a3 (?semv ?v a3 ?semnp))) →
 (V[_np] SEM ?semv) (NP SEM ?semnp)
4. (NP WH- VAR ?v SEM (PRO ?v ?sempro)) → (PRO SEM ?sempro)
5. (NP VAR ?v SEM (NAME ?v ?semname)) → (NAME SEM ?semname)
6. (NP VAR ?v SEM < ?semart ?v (?semcnp ?v) >) →
 (ART SEM ?semart) (CNP SEM ?semcnp)
7. (CNP SEM ?semcnp) → (N SEM ?semn)

(Забележка: V[_np] е съкратен запис на V със SUBCAT _np, V[_none] е съкратен запис на V със SUBCAT _none)



II.3. Колко семантика в лексикона и колко в правилата на граматиката?

В горните примери лексиконът съдържа само буквено-цифров идентификатор за кодиране на значението на отделните думи, а λ-изразите се конструират чрез правилата на граматиката. Възможен е и друг подход:

да поставим λ -изрази в лексикона и така да работим с граматика с по-прости правила. В контекста на разгледания по-горе пример, бихме могли да имаме в лексикона

decide ... SEM (λy (DECIDES1 e1 y)), и съответно
decide ... SEM (λo (λy (DECIDES-ON1 e1 y o))).

Каква е ползата от такова решение да усложним лексикона за сметка на граматиката? За прости граматика (като горната) изглежда по-удобно λ -изразите да се конструират чрез правилата. Но при включването на описания за повече видове конституенти граматиката става сложна и съответно се усложняват ЛФ на конституентите, които трябва да се “изчисляват” чрез правила. Например, ако искаме нашата граматика да работи с преходни глаголи с по 1, 2 или 3 тематични роли, няма трябва да имаме 3 вида правила, които да управляват извода за отделните случаи чрез конструиране на λ -изрази с необходимия брой аргументи? *Решението “къде колко семантика да се дефинира” е винаги компромисно. Виждаме, че граматиката (програмата) не може да се конструира независимо от лексикона (данните), а двете са взаимодействащи си компоненти на едно цяло.*

Понеже не е лишено от смисъл да поставим информацията за тематичните роли в лексикона при самия глагол, нека видим как се прави. Пример:

see: (v VAR ?v SEM (λo (λa (SEES1 ?v [AGENT a] [THEME o]))))).
break: (v VAR ?v SEM (λo (λi (λa (BREAKS1 ?v [AGENT a] [THEME o] [INSTR i])))))).

Тогава познатото ни правило за пряко допълнение

(VP SEM (?semv ?semnp)) \rightarrow (V SEM ?semv) (NP SEM ?semnp)

би ни помогнало за получим за фразата *see the book* нейната КЛФ
(λa (SEES1 b1 [AGENT a] [THEME <THE b1 (BOOK b1)>])),

а за *break the book* съответната КЛФ
(λi (λa (BREAKS1 b1 [AGENT a] [THEME <THE b1 (BOOK b1)>] [INSTR i]))).

II.4. Йерархични лексикони

Нека отбележим, че при дефинирането в лексикона на тематичните роли на *see* и *break* сме били принудени да съобразим кой е “най-външният” параметър на λ -израза с линейността на правилото за преходен глагол, т.е. при даденото по-горе правило се изисква в лексикона да има подредба на λ -параметрите според предполагаемата поява на прякото допълнение на

първо място след глагола. Обаче на практика ситуацията е по-сложна: например, глаголът *give* позволява три “повърхнинни” употреби:

- I gave the money.
- I gave John the money.
- I gave the money to John.

Обработката на тези “синтактични варианти” би изисквала в лексикона да имаме три записа на семантиката на *give*:

- (v SUBCAT _np SEM $\lambda o \lambda a$ (GIVE1 ?v [AGENT a] [THEME o]))
- (v SUBCAT _np_np SEM $\lambda r \lambda o \lambda a$ (GIVE1 ?v [AGENT a] [THEME o] [TO-POSS r]))
- (v SUBCAT _np_pp:to SEM $\lambda o \lambda r \lambda a$ (GIVE1 ?v [AGENT a] [TO-POSS r] [THEME o]))

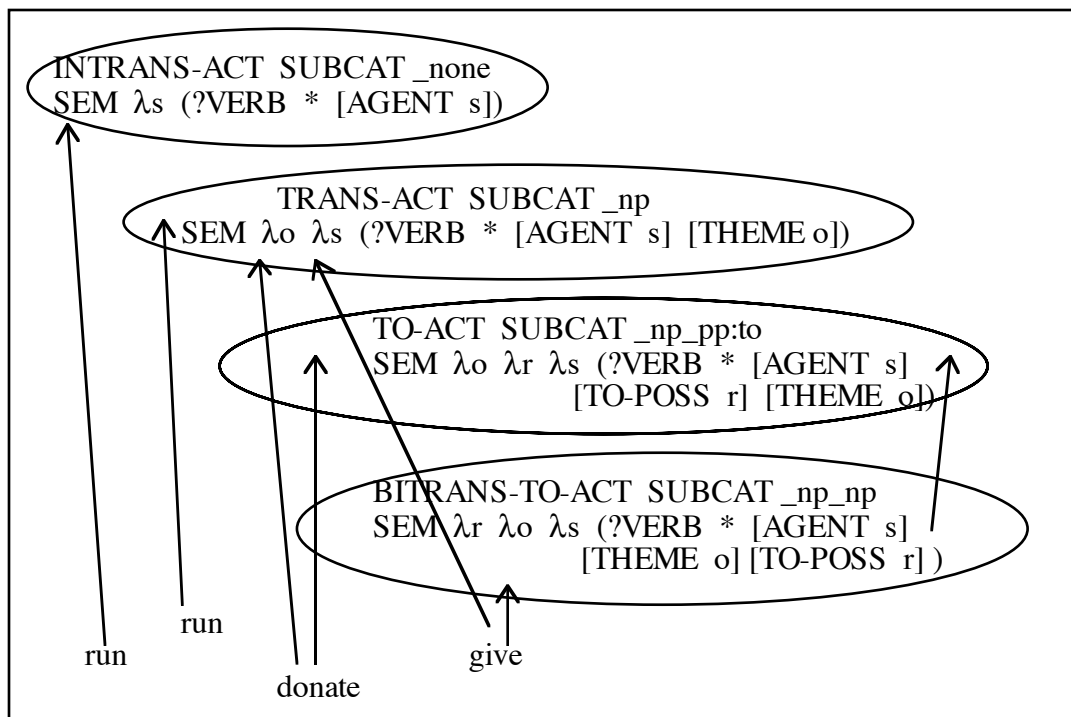
с цел да отреагираме по правилния “семантичен” начин, независимо от конкретната синтактична форма на фразата за анализ.

Разбира се, невъзможно е за всеки глагол да дефинираме в лексикона многобройните варианти на линейно разположение на “пълнителите на тематичните роли”. Затова си служим с т. нар. “йерархични лексикони”, в които правим **наследяване (inheritance)**.

Организираме глаголите в абстрактни класове, където се описват общите им свойства. Например, всички *глаголи задаващи действия* са или

- непреходни, 1-аргументни INTRANS-ACT като *тичам, смея се, седя ...* със SUBCAT _none и семантика λs (?VERB * [AGENT s]), или
- преходни, 2-аргументни TRANS-ACT като *пия, пея, рисувам, ...* със SUBCAT _np и SEM: $\lambda o \lambda s$ (?VERB * [AGENT s] [THEME o])
- и т.н.

Получаваме лексикон с така наречената *лексикална йерархия* (вж. картинката). Показани са три типа транзитивни глаголи за английския език, като е отразен фактът, че даден глагол може да принадлежи към няколко класа. Освен това е показано как даден тип глаголи (т. нар. двойно-транзитивни, клас BITRANS, при непряко допълнение с TO) допускат и синтактична конструкция от класа TO-ACT. Сега няма нужда да изброяваме всички възможни синтактични фрази за дадени значение на



глагола - както започнахме в тази част с примера за *давам*, а е достатъчно да групираме глаголите във възможни абстрактни класове и да зададем техните зависимости в лексикона. Така *give* влиза в 3 класа.

Забележете, че на различните значения съответстват различни върхове (*run* като “тичам” и като “управлявам /машина/”).

III. Разрешаване на многозначността на значенията на думите

Ще видим, че в редица изречения е възможно със сравнително прости средства да се ограничи броят на комбинациите от значения, които трябва да се разглеждат при семантичната интерпретация (НО НЕ ВИНАГИ!).

Например, за многозначната дума МОСТ

- МОСТ1 - пътно съоръжение ПЪТНО_СЪОР5,
- МОСТ2 - зъболекарски мост ЗЪБ_МОСТ37,
- МОСТ3 - контакт, “мост между партии”, АБСТР_ВРЪЗКА18,
- МОСТ4 - гимнастическо уптажнение, ГИМ_УПР4,

в изречението “Мостът свързва Варна с квартал Аспарухово” трябва да разгледаме само МОСТ1, понеже става дума за географски обекти. Да строим интерпретация за МОСТ2, 3 и 4 е загуба на ресурси. Но досега не знаехме как да ограничим разглежданията до “по-правдоподобните” значения. Сега ще се занимаем (донякъде) с този въпрос. Пълното му разрешаване е възможно само при контекстуалната интерпретация.

III.1. Ограничения на избора чрез йерархия (таксономия) от значения на думи

Йерархия от значения: групираме нещата и индивидите в типове (класове) обекти. КУЧЕ и КОТКА са класове, които не се пресичат, а и двете са подклас на БОЗАЙНИК. Ако класифицираме ЧОВЕК като МЪЖ и ЖЕНА двата типа не се пресичат, но при класификация МЪЖ, ЖЕНА, СТУДЕНТ, ДЕТЕ подтиповете се пресичат. Класификациите на подкласове биват:

- **изчерпващи (exhaustive partitions)** или не (напр. “ОБЕКТ е МАТЕРИАЛЕН-ОБЕКТ и АБСТРАКТЕН-ОБЕКТ” е изчерпваща, а “БОЗАЙНИК е КУЧЕ и КОТКА” е неизчерпваща),

- **изключващи се (disjoint partitions)** или не (напр. “БОЗАЙНИК е КУЧЕ и КОТКА” е изключваща, а “ЧОВЕК е ЛЕКАР и СТУДЕНТ и МЪЖ и ЖЕНА” не е изключваща).

Нормално е в една йерархия да има смесен тип класификация на различните нива. Зависи с каква цел е правена таксономията (класификацията). Обръща се повече внимание на по-важните за задачата върхове. Йерархиите са решетки, а не дървета. Отгоре е TOP (ALL), отдолу ВОТТОМ (NOTHING).

За какво ни е всичко това в ОЕЕ? Това знание служи за ограничаване на възможните значения при семантичната интерпретация. Например:

- цветът ЗЕЛЕН може да бъде съчетан само с PHYS-ОБЪЕСТ;

- обектите от типа ПРЕЦИЗЕН са модификатори само на ИДЕЯ или ДЕЙСТВИЕ и т.н.

Това са ограничения върху съчетаемостта на значенията, задават се в рамките на по-големи групи значения. Ето примерна идея как се прави:

Ако нашата система за разбиране на ЕЕ разполага с долната йерархия на значенията, могат да й се зададат **ограничители при избора** (selectional restrictions). Това са спецификации на “позволените” (т.е. случващите се заедно) комбинации от значения. Например:

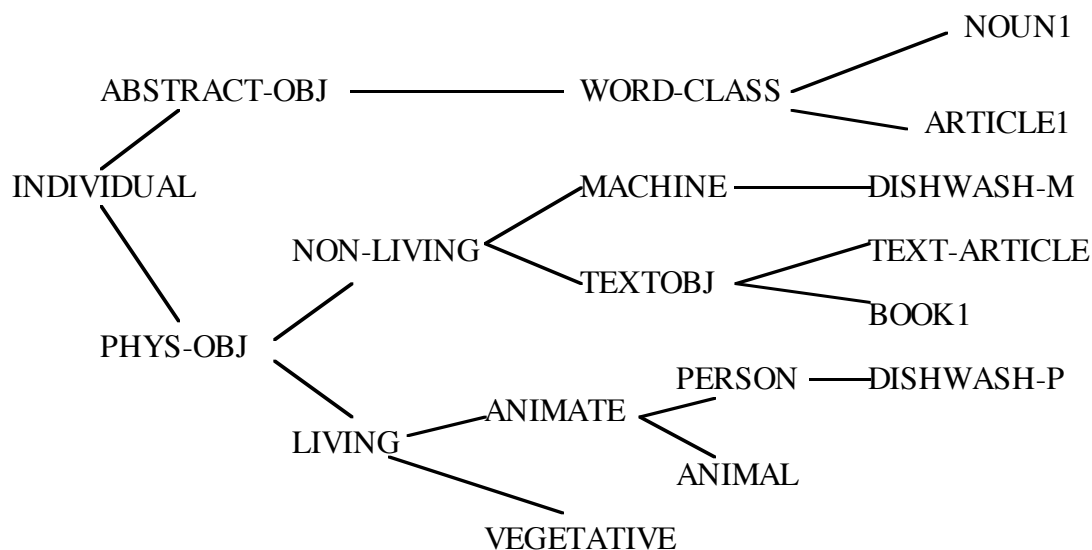
(AGENT READS1 PERSON) - агентът на “чета” е лице

(THEME READS1 TEXTOBJ) - това, което се чете, е ТЕХТОВЪ

Нека си представим, че в лексикона READ2 е значението “виждам ясно”, тогава

(AGENT READS2 PERSON)

(THEME READS2 MENTAL-STATE)



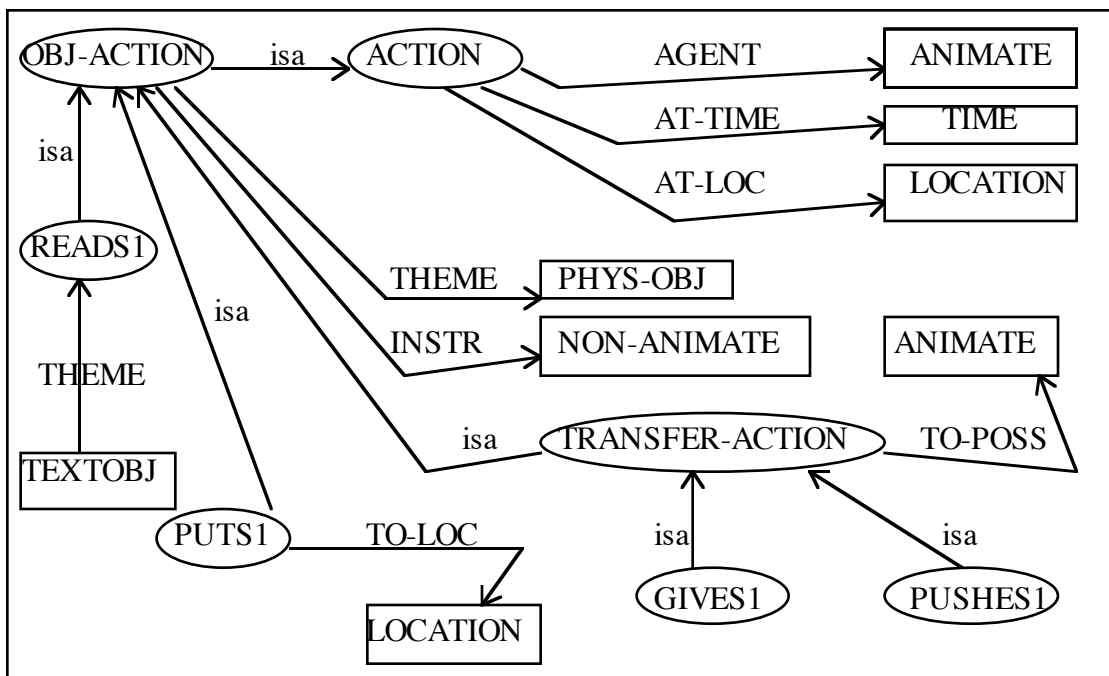
Ако такава система трябва да разбере изречението

The dishwasher reads the article

очевидно ограничения от горния вид ще ни помогнат да сведем 8-те възможни комбинации от значения (2 за READ, 2 за dishwasher и 2 за article) до единственото правдоподобно значение.

Общи спецификации за възможна съчетаемост ни помагат и в други случаи: **първо**, да разграничим *буквалното* от *метафоричното* четене на изречения като “*моята кола пие бензин*”. **Второ**, те са важен инструмент за “предсказване” на референции: например, в изречението Аз го чета излиза, че го е или текстов обект (ако е употребено READ1), или абстрактен обект (ако е употребено READ2). Така знаем какъв вид референт да търсим из текста.

Почти всички системи за разбиране на ЕЕ използват йерархии и спецификации от горния вид. В някои системи така се прави избора, в други спецификациите задават *преференциални четения*. Даваме и една по-обща картинка (отдолу) на **семантична мрежа** (semantic network), в която думите от лексикона са свързани със знанието в йерархията. С кръгчета са дадени единиците от лексикона, с правоъгълници - елементите на знанието. По дъгите винаги стоят имена на семантични релации. Обърнете внимание, че макар да сме разделили **лингвистичните данни (в лексикона)** от **общото знание за света (в базата знание)**, те в долната илюстрация са свързани помежду си и е показано как може да стане това. Възниква въпросът, защо да ги делим? Едната причина е модуларността, прегледността и т.н. Другата важна причина е, че признакът SEM в лексикона използва *код на значението във вид на идентификатор*, но в лексикона не се знае нищо повече за този код, освен че е свързан с ЕЕ-дума.



Не се знае как този код се съотнася с другите кодове на значенията, според приписаното му концептуално съдържание. Знаем, че традиционното решение в Изкуствения Интелект е концептуалната информация да се задава в **база знание БЗ (knowledge base, KB)**.