Корпус от синтактични описания на българския език - BulTreeBank

Кирил Симов и Петя Осенова

www.bultreebank.org

28.01.2005

Програма

12.30-14.45 Представяне на корпуса от синтактични описания на българския език - Лингвистични параметри на анотационната схема:

- Основни понятия на Опорната фразова граматика (ОФГ)(HPSG)
- Основни принципи на проектирането на анотационната схема
- Представяне на лингвистичните явления.
- Лингвистично търсене на езикови примери с CLaRK системата.

- **14.45-15.00** Почивка
- **15.00-16.45** Поканени изказвания:

  проф. дфн Руселина Ницолова
  проф. дфн Йордан Пенчев
  ст. н.с. д-р Елена Паскалева
  доц. д-р Радка Влахова
  доц. д-р Йовка Тишева
- **16.45-17.00** Почивка
- **17.00-17.30** Свободна дискусия

Участници в проекта

Кирил Симов – ръководител на проекта от българска страна

Лингвисти:

Петя Осенова, Сия Колковска, Милена Славчева, Елисавета Балабанова, Димитър Дойков, Илиана Гаравалова, Магделена Паунова

Програмисти:

Александър Симов, Милен Куйлеков, Красимира Иванова, Христо Ганев, Илко Григоров

Финансова подкрепа

BulTreeBank (2001-2004)

беше съвместен проект между

Seminar für Sprachwissenschaft,

Eberhard-Karls-Universität, Тюбинген, Германия

и

Лабораторията за лингвистично моделиране, БАН, София

Проектът беше финансиран от

Volkswagen-Stiftung, Германия

Целта

- Създаване на електронен ресурс от синтактично анотирани текстове за българския език, който да бъде:
  - Основа на формална компютърна граматика на българския език (parser)
  - Консистентно анотиран, за да бъде лесно и удобно използваем
  - Достатъчно детайлно анотиран, за да позволява по-специфични употреби

Кое НЕ БЕШЕ наша цел ☺

- Да решим проблемите на българския синтаксис
- Да удовлетворим всички вкусове в лингвистичните интерпретации
- Да смятаме, че това е единственият възможен поглед върху синтактичните структури

Какво е синтактично анотиран корпус (treebank)

- Терминът е въведен от Джефри Лийч – синтактичен анализ, представен чрез дървовидна структура
- Видове корпуси като качество:
  - "Ботаническа градина" – проверена ръчно
  - "Гора" – частично проверена
  - "Джунгла" – само автоматичен анализ

Статусът на синтактичния корпус в момента

Синтактичният корпус в момента съдържа около **15 000** изречения, от които:

- Примери, извадени от граматики
  - **1500 от 4 български граматики**
- Примери, извадени от корпуса
  - **1500** изречения, случайно избрани от корпуса
  - **12 000** изречения от пълни статии от вестници, както и от други източници (художествена литература, законови текстове, публицистика и др.)

Текстовете, извадени от граматиките

Кои са граматиките?

- Академична граматика, БАН, 1983
- Ст. Брезински, Български синтаксис, 2001
- Й. Пенчев, Синтаксис – управление и свързване, 1993
- К. Попов, Синтаксис, издание 1993

Основни характеристики:

- Те са т. нар. *ядрено множество* в корпуса
- Следват предварителна класификация
- Реализацията на кореференции и елипси е ограничена в рамките на изречението

Текстовете, извадени от корпуса

Основни характеристики:

- Анотирани са цели параграфи и статии
- Има разнообразие на синтактични отношения, което е типично за свързания текст
- Синтактичните отношения са по-сложни, но същевременно е по-лесно да се разрешат елипсите и кореференциите

Използване на корпуса извън проекта

За лингвистични изследвания:

- Дисертации
- Курсови работи
- Обучение на студенти
- Речници – нови думи, нови употреби, изграждане на словници

За автоматично извличане на лингвистично знание – тагери, речникова информация, парсер

Опорна фразова граматика HPSG

Лексикална лингвистична теория, разглеждаща лингвистичните анализи като атрибутни структури (Feature Structure) (езикови обекти, графи)

Всяка граматика в ОФГ се състои от две части:

- Лингвистична онтология (sort hierarchy)

  Определя основните типове обекти и техни свойства

- Граматични принципи

  Представляват ограничения върху възможните обекти

Лингвистична онтология

root

FEAT-A : *subsort1*

FEAT-B : *subsort2*

subsort1

subsort2

FEAT-B : *subsort4*

FEAT-C : *subsort2*

subsort4

subsort5

subsort3

Граматични принципи

Логически описания, които се интерпретират като верни или неверни в/у лингвистичните обекти

Най-често се задават като импликации

$$\mathbf{A} \rightarrow \mathbf{B}$$

Интерпретират се като универсални принципи, т.е. всеки обект, който удовлетворява **A**, трябва да удовлетворява и **B**

Лексиконът също е представен чрез принципи

ОФГ сорт йерархия - *sign*

Основният тип лингвистични обекти са от сорт знак (*sign*):

sign

PHON : *phonlist*

SYNSEM : *synsem*

word

ARG-ST : *list-of-synsem*

phrase

DTRS : *con-struc*

ОФГ сорт йерархия - *con-struc*

Конституентната структура се дефинира чрез йерархията:

con-struc

headed-phrase

HEAD-DTR : *sign*

COMP-DTRS : *list-of-phrases*

head-complement

HEAD-DTR : *word*

COMP-DTRS : *ne-l-of-phrase*

con-struc (2)

head-subject

HEAD-DTR : *phrase*

SUBJ-DTR : *phrase*

COMP-DTRS : *empty-list*

head-adjunct

HEAD-DTR : *phrase*

ADJUNCT-DTR : *phrase*

COMP-DTRS : *empty-list*

head-sem-adjunct

head-pragmatic-adjunct

con-struc (3)

head-only

head-filler

HEAD-DTR : *phrase*

FILLER-DTR : *phrase*

COMP-DTRS : *empty-list*

non-headed-phrase

coordination-phrase

CONJ-DTRS : *set(sign)*

CONJUNCTION-DTR : *word*

Видове опори - *head*

head

substantive (subst)

PRD : *boolean*

MOD : *mod-synsem*

noun

verb

adj

adv

prep

Принципи на ОФГ (1)

- Принцип на опората (Head Feature Principle)

Стойността на атрибута **HEAD** на една фраза с опора (*headed-phrase*) съвпада със стойността на атрибута **HEAD** на опората (**HEAD-DTR**)

Опорните характеристики на една фраза с опора се определят от опорните характеристики на опората ☺

Принципи на ОФГ (2)

- Валентен принцип (Valence Principle)

Стойността на всеки валентен атрибут (**SUBJ** или **COMP**) на една фраза с опора (*headed-phrase*) е равна на стойността на този атрибут на опората (**HEAD-DTR**) без другите депенденти

Опората определя вида на комплементите или подлога

Принципи на ОФГ (3)

- Адюнктен принцип (Head-Adjunct Principle)

Стойността на атрибута **SYNSEM** на една фраза с опора (*head-adjunct*) е равна на стойността на **MOD** атрибута на адюнкта

Адюнктът определя вида на опората, към която се свързва

Принципи на ОФГ (4)

- Семантичен принцип (Semantic Principle)

Стойността на атрибута **CONTENT** съвпада със стойността на същия атрибут на адюнкта, ако фразата е със семантичен адюнкт (*head-sem-adjunt*), или е равна на стойността на този атрибут на опората

Семантиката се определя от сематичната опора, която съвпада със синтактичната опора с изключение на семантичните адюнкти

Принципи на анотация

- Адекватно представяне на лингвистичните факти

  съблюдаване на лингвистичната теория

- Удобство при ръчно анотиране

  въвеждане на минимално количество информация

- Стабилност на анотиране

  всяко изречение в корпуса трябва да получи адекватен анализ на същото ниво на детайлност

Елементи на анотацията

- Конституентност
- Категория
- Отношения на зависимост
- Линейна подредба
- Кореферииране
- Неизразени синтактични елементи

Пълният ОФГ анализ може да бъде възстановен на базата на тази информация

Този избор на елементи улеснява прехода към други лингвистични теории

Анотационна схема

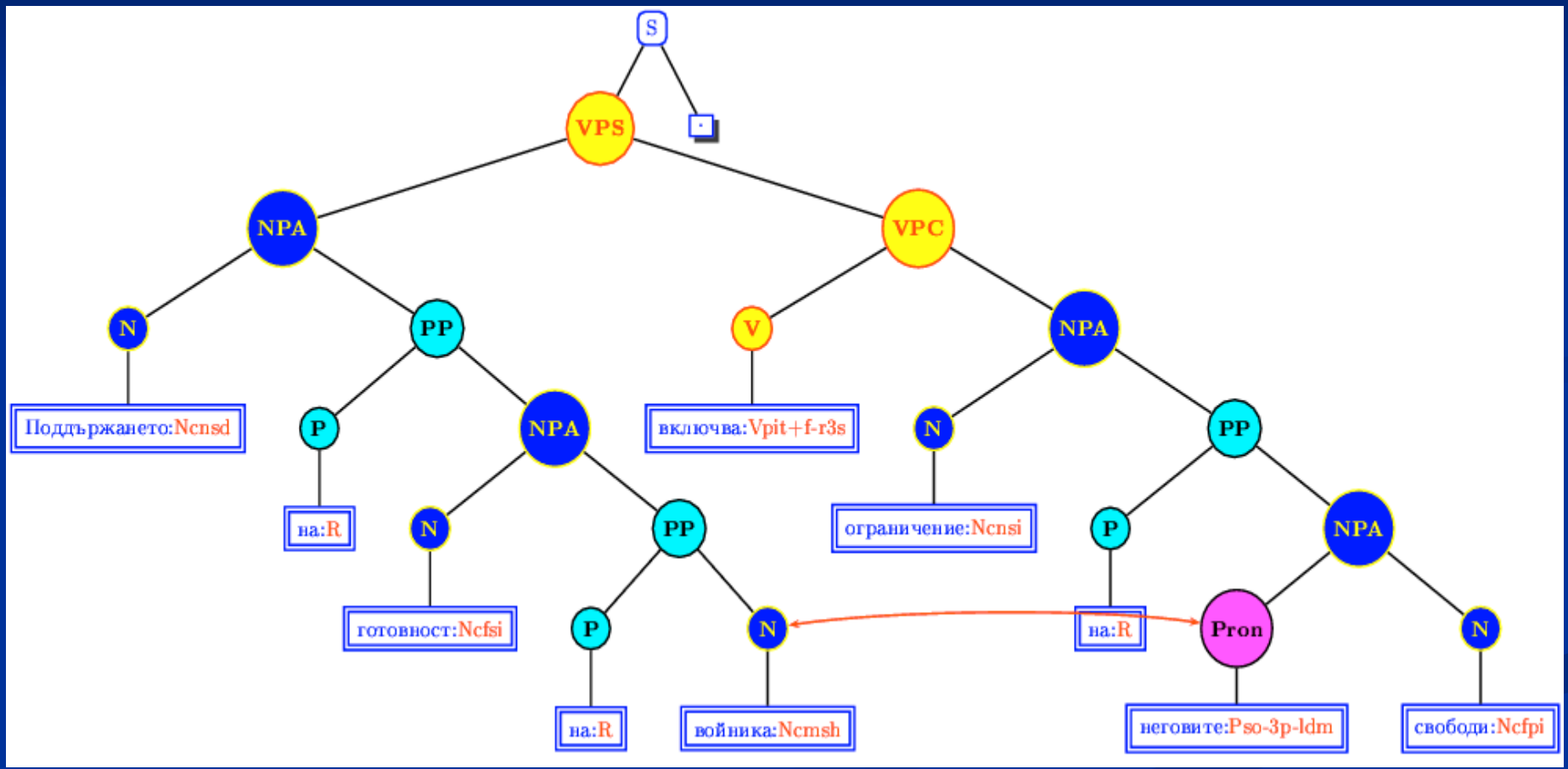
- **Следва езиковия модел на ОФГ (HPSG)**
- **Йерархия от фрази, депендентни отношения и кореференции**
- **Имплементация в XML**
  - Текстови елементи
  - Лексикални елементи

    (N, V, Prep)
  - **Синтактични елементи и депендентна информация**

    (VPA(djunct), NPC(omplement))
  - Функционални елементи

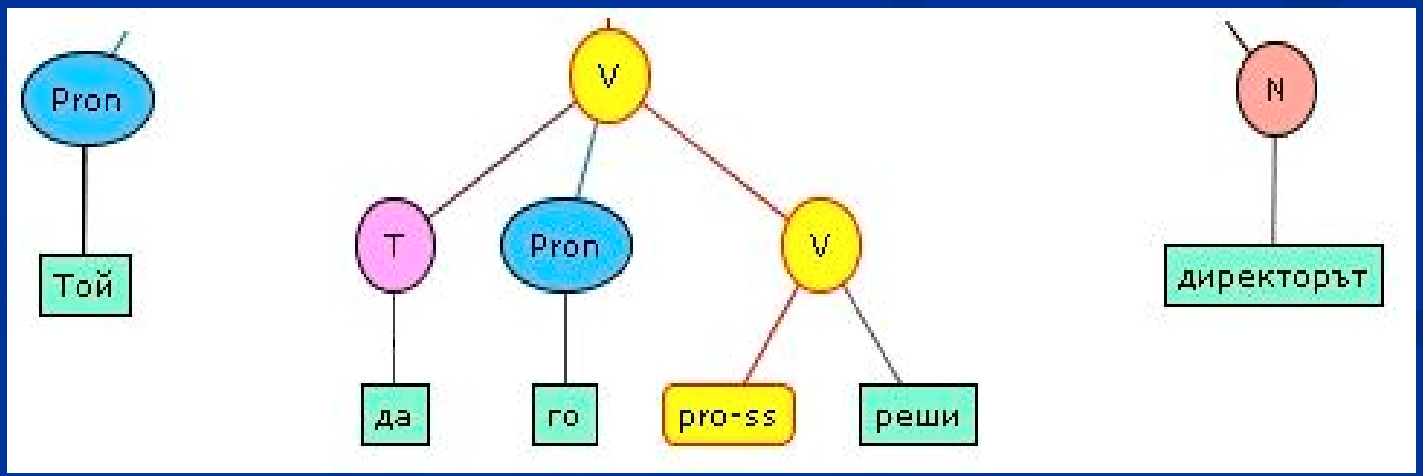
    (Disc(ontinous), E(xtracted)

Конституентност и депендентност



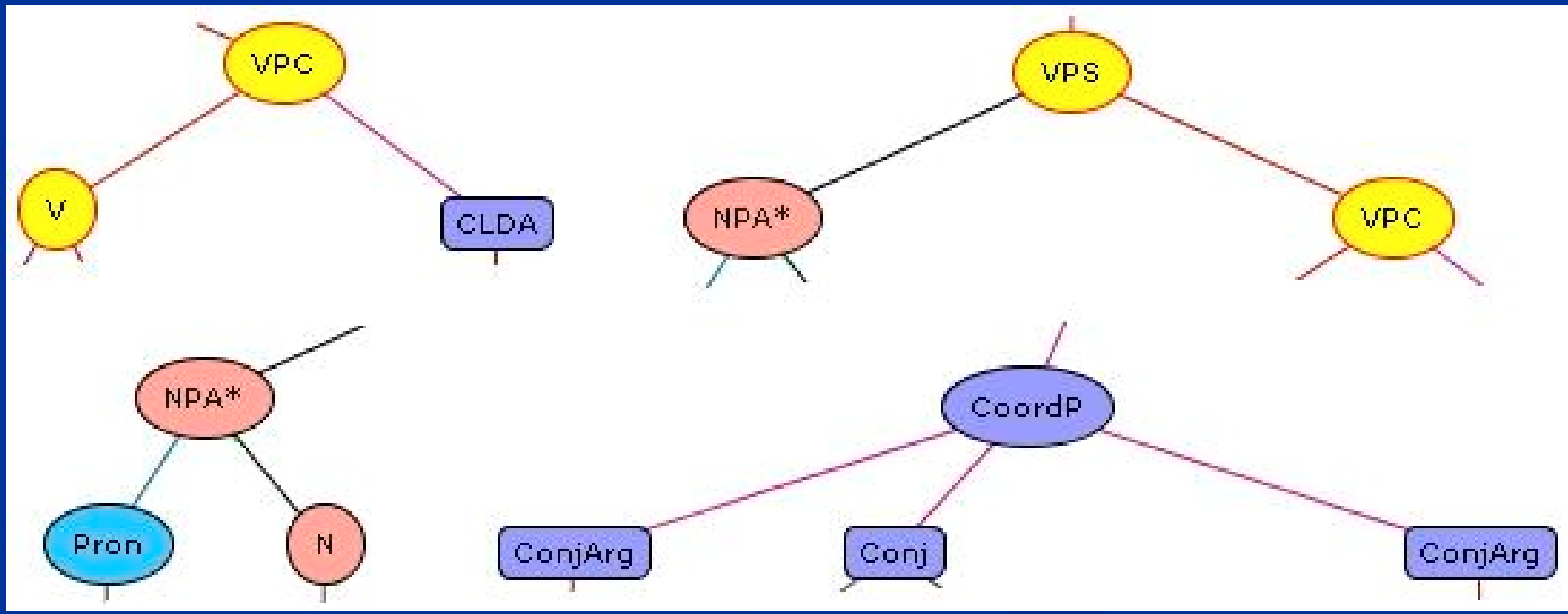
Лексикални елементи

- Прости (Pron, M, T, V, N, C, Ger и др.)
- Сложни (V, C, Part и др.)



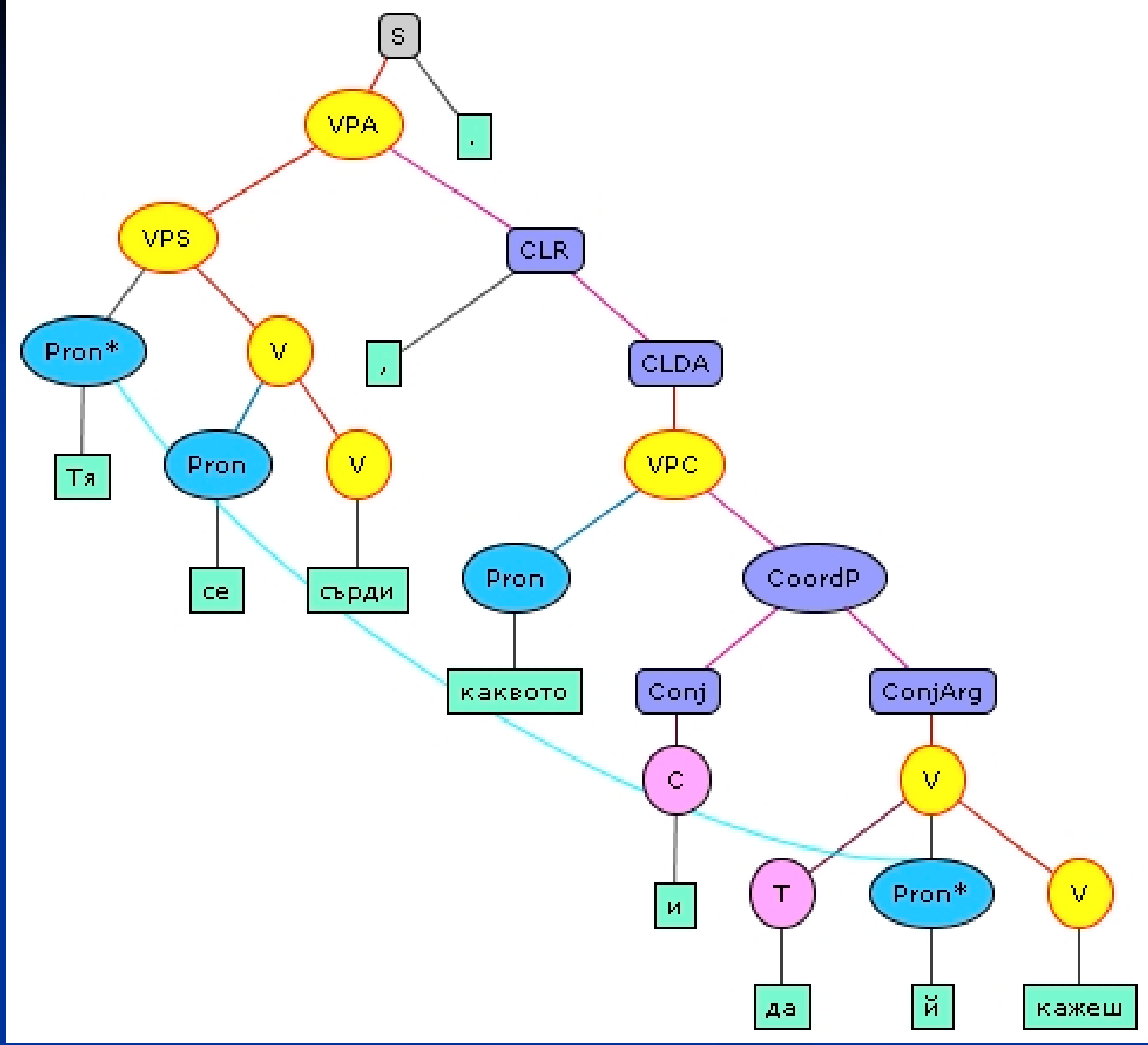
Фразови елементи

NPA, NPC, VPC, VPA, VPS, PP, APA, APC, APA, AdvPA, AdvPC, CoordP и др.



Функционални елементи

- Клаузи – *CL, CLDA, CLCHE, CLZADA, CLR, CLQ*
- Изречение – *S*
- Маркери на координацията - *Conj, ConjArg*
- Маркери на елипсата – *V-Elip, N-Elip*
- Маркери на дистантно разположени конституенти – *DiscA, DiscE, DiscM*
- Прагматични елементи - *Pragmatic*
- НЕнепосредствено доминиране - *nid*


S
VPA
.
VPS
CLR
Pron*
V
,
CLDA
Тя
Pron
V
VPC
се
сърди
Pron
CoordP
каквото
Conj
ConjArg
с
V
и
Т
Pron*
V
да
й
кажеш

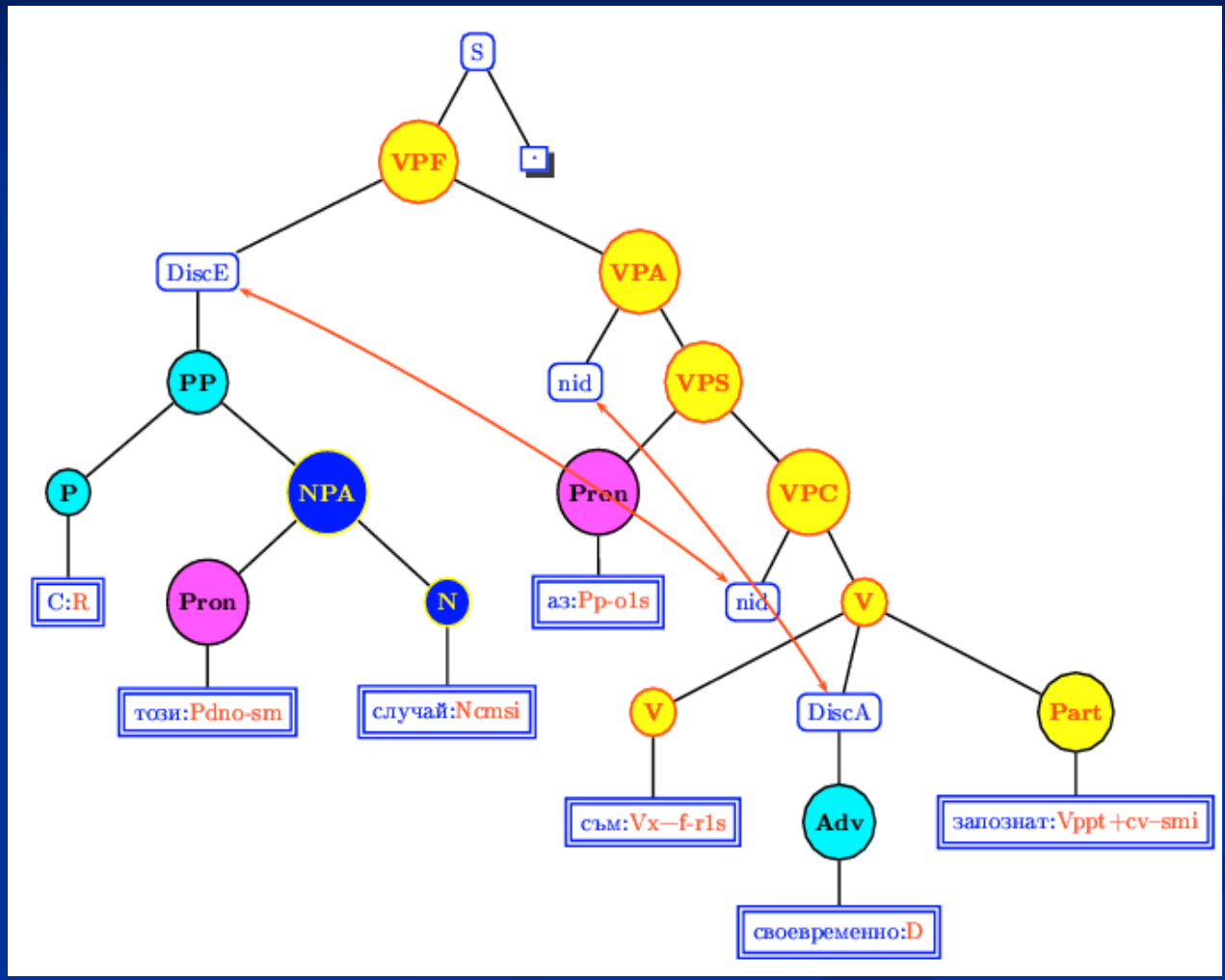
Реализация на депендентите

- В каква конституентна йерархия?

Комплемент(и) -> Субект -> Адюнкт(и)

- Колко депенденти наведнъж?
  - Комплементите се взимат наведнъж (Immediate Dominant Schemata 3 и Valence list Principle в HPSG94)
  - Адюнктите се взимат един по един (Mod Principle в HPSG94)

Реализация на депендентите



Конституентност и словоред (1)

- Конституентната структура е отделена от линейното подреждане
- Елементите на един конституент обикновено се реализират в близост. Но не приемаме определен словоред за основен → → → →

Конституентност и словоред (2)

Всички възможни пермутации се разглеждат като една и съща структура:

Мъжът целува момичето

Целува момичето мъжът

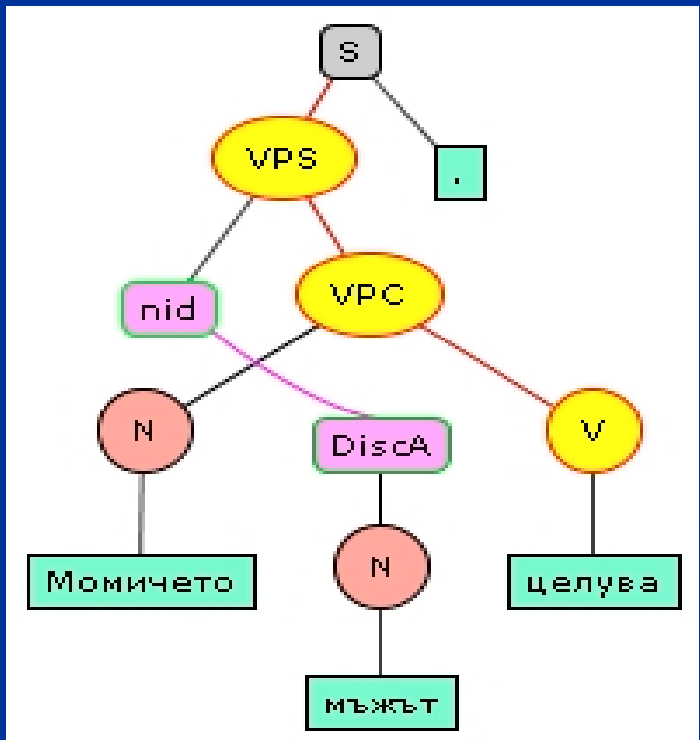
Момичето целува мъжът

Мъжът момичето целува

……..

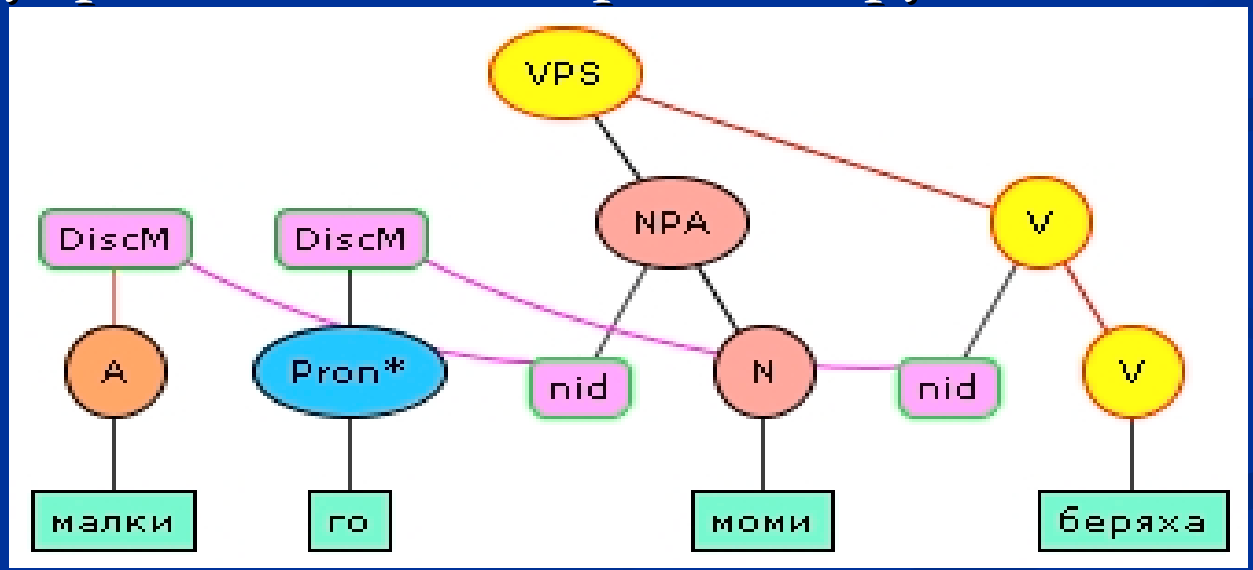
Типове дистантно разположени конституенти (1)

- **Пермутация на депенденти на една опора** (*DiscA* функционален елемент). По-висок депендент се реализира между опората и по-нисък депендент.



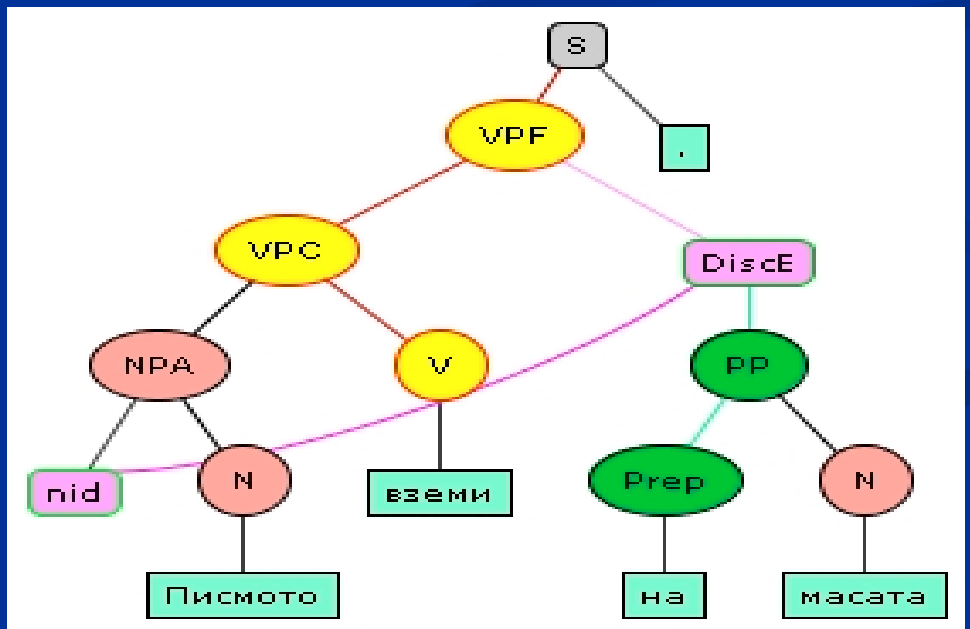
Типове дистантно разположени конституенти (2)

- **Смесване на два конституента** (*DiscM* фунционален елемент). Елементите на две конституентни структури са смесени, без някоя от тях да е управителна категория на другата.



Типове дистантно разположени конституенти (3)

- **Външна реализация на вътрешен конституент** (*DiscE* функционален конституент). Това обикновено се нарича *екстракция*. Т.е. елементът се управлява от опора, която е разположена по-ниско.



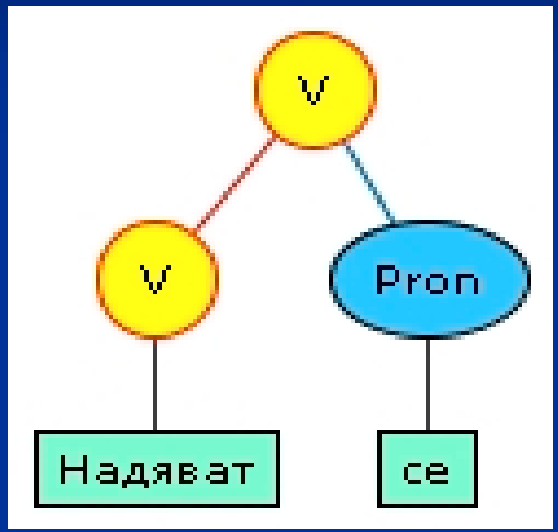
Типове лексикални елементи

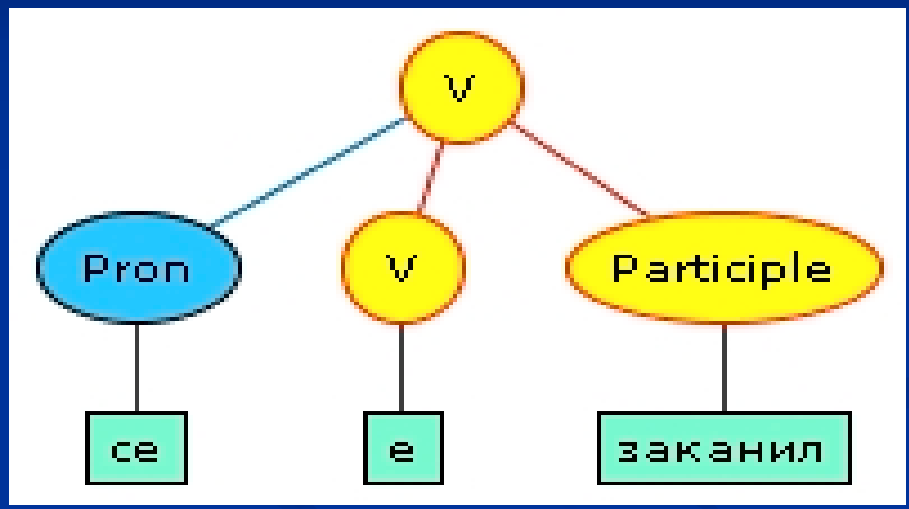
- Глагол **(V),** Причастие **(Part),** Деепричастие **(Ger)**
- Съществително **(N),** Местоимение **(Pron),** Прилагателно **(A),** Числително **(M),** Фамилно име **(H)**
- Наречие **(Adv)**
- Съюз **(C)**
- Предлог **(Prep)**
- Частица **(T)**
- Междуметие **(I)**

Лексикалният елемент: Глагол

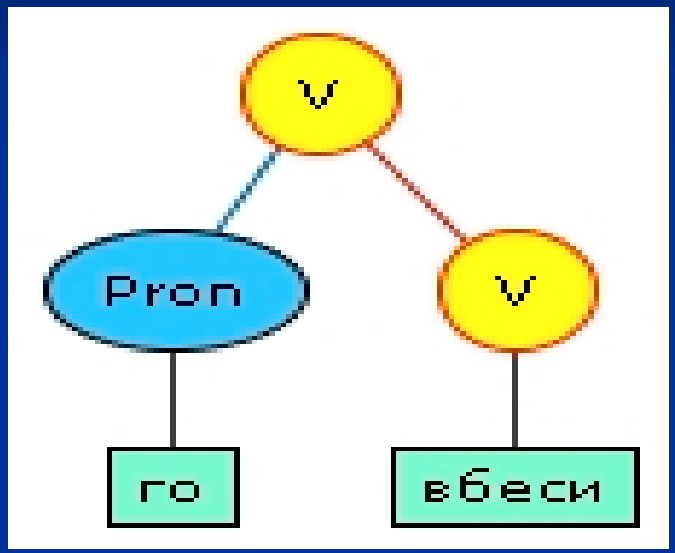
- Глаголният лексикален елемент (*V*) отговаря на отделен глагол или на аналитична глаголна група (пасив, наклонение, глаголни времена и др.)
- За многоелементни глаголи смятаме глаголен комплекс, в който опората е личен глагол, придружен от клитики, спомагателни частици, емфатични наречия, причастия

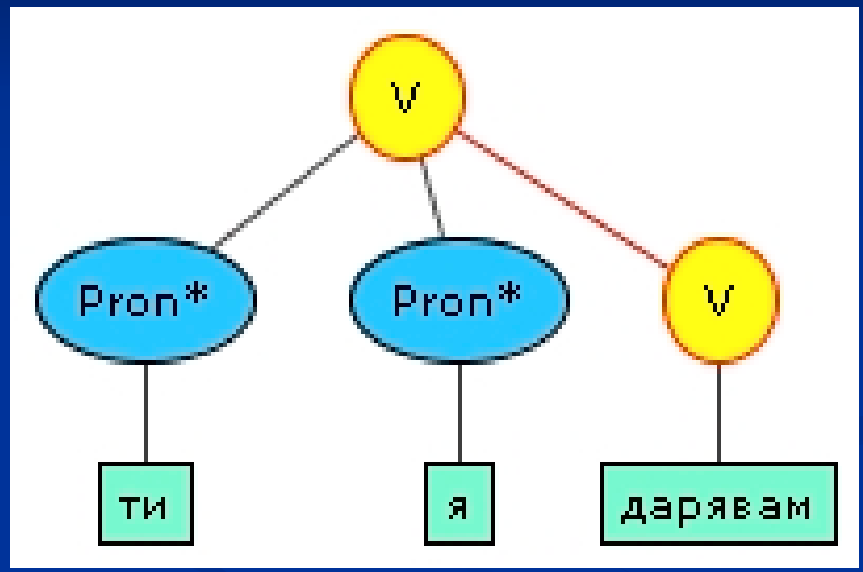
Клитични форми (се, си)





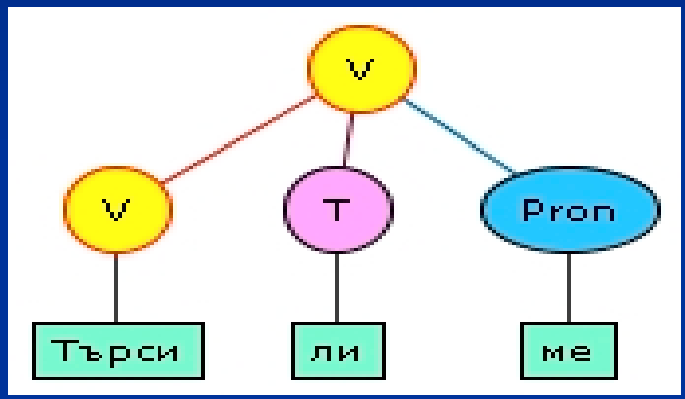
Винителни и дателни местоименни клитики



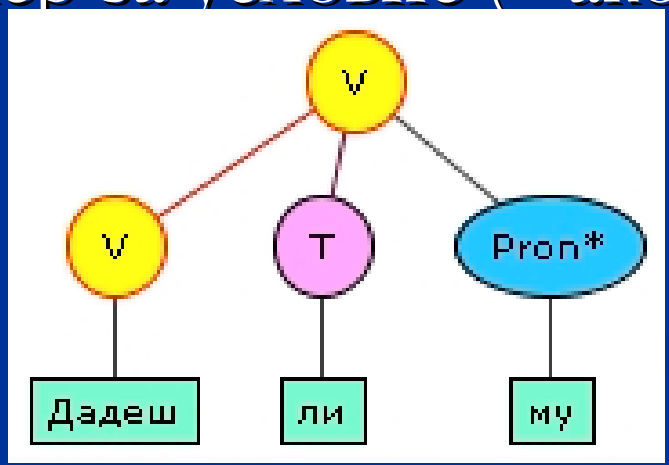


Въпросителна клитика – въпросителната частица 'ли'

- Маркер за въпросителност

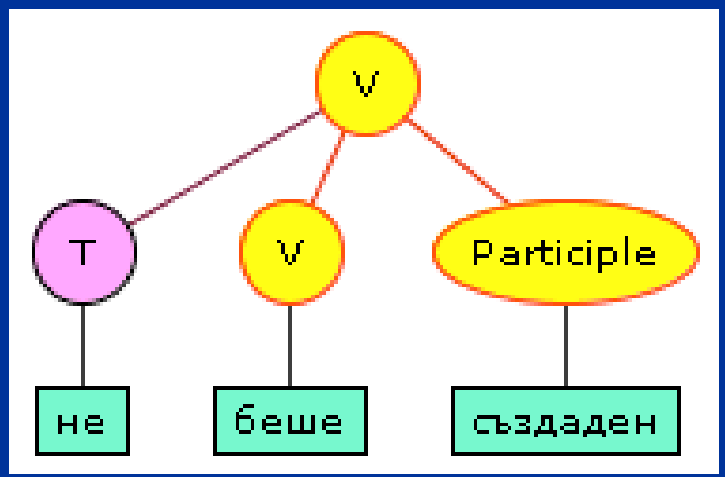


- Маркер за условие (=ако)



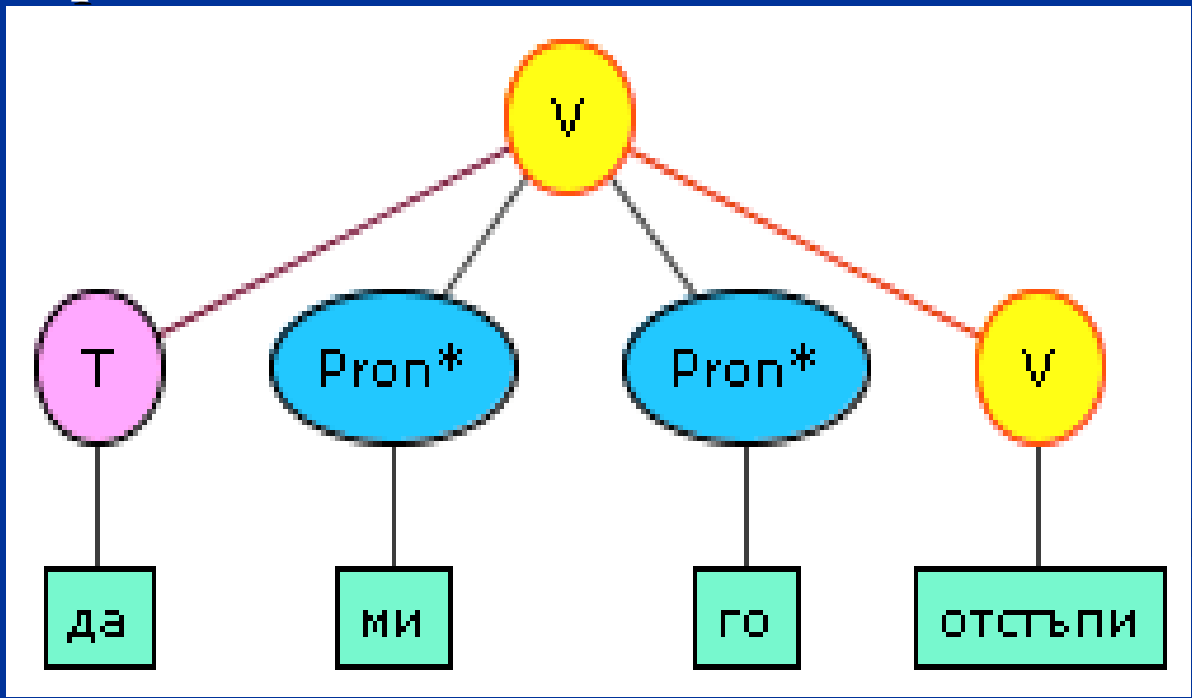
Отрицателна клитика –отрицателната частица 'не'

Отрицателната частица 'не' винаги се реализира първа във глаголния комплекс освен случаите, когато има ‘да’



Спомагателна частица 'да'

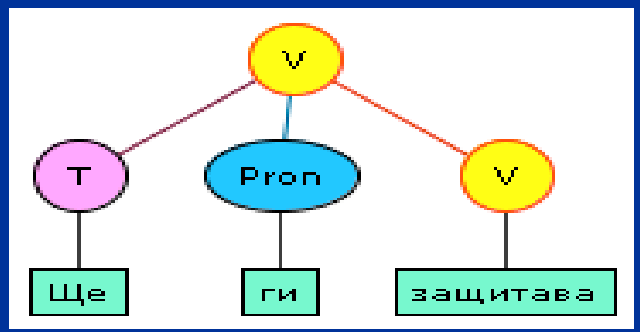
Спомагателната частица 'да' в инфинитивното си значение избира глагол в сегашно време и образува глаголен комплекс, който наследява *ARG-ST* листа на избрания глагол



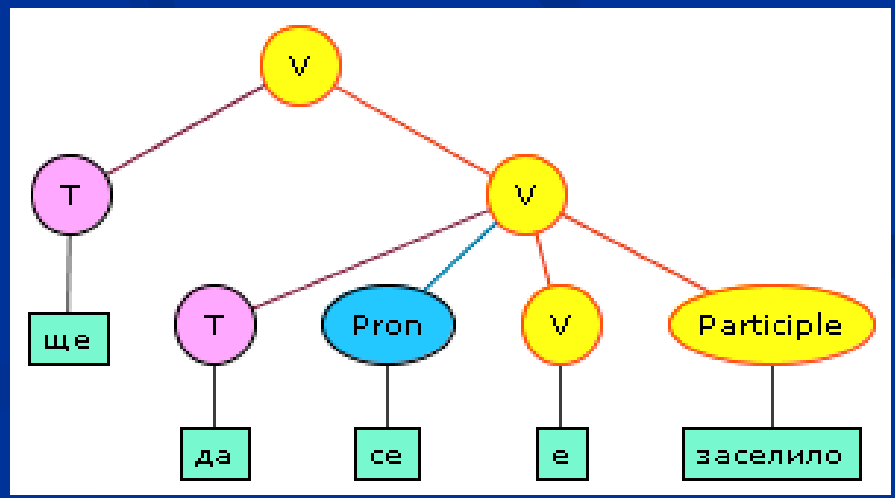
Спомагателна частица 'ще'

Спомагателната частица ***ще*** също има няколко роли във формирането на глаголния комплекс:

- Образуване на бъдеще време и бъдеще предварително време



- Презумптивни форми



Други аналитични форми

- Аналитични форми със спомагателните глаголи: 'съм', 'бъда', 'бивам‘

бях дошъл, взел е.....

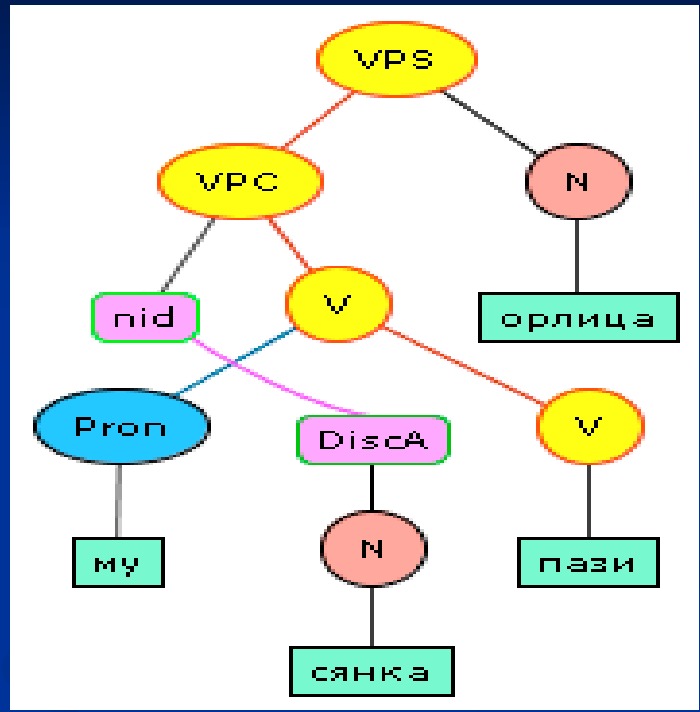
- Аналитични форми с глагола 'ща‘

-> те образуват фрази

щях [да дойда]

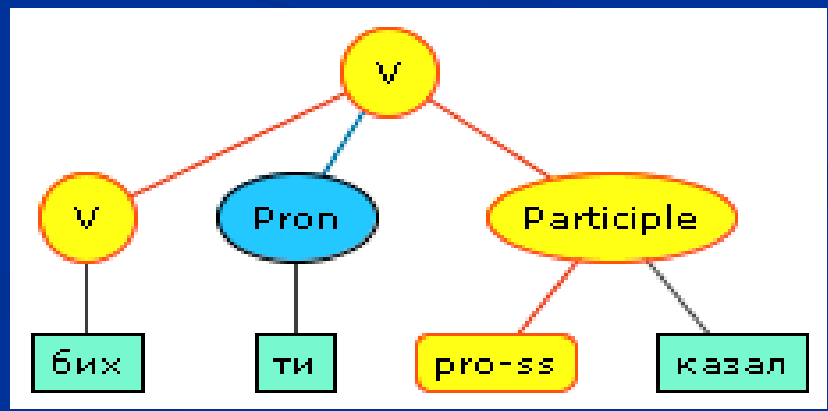
Специфики

- Дистантно разположени конституенти



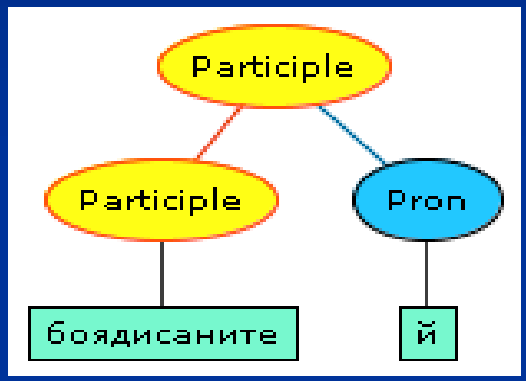
- Нулева субектност

- Случаи на елипси
- V проекция на *нека, ето*

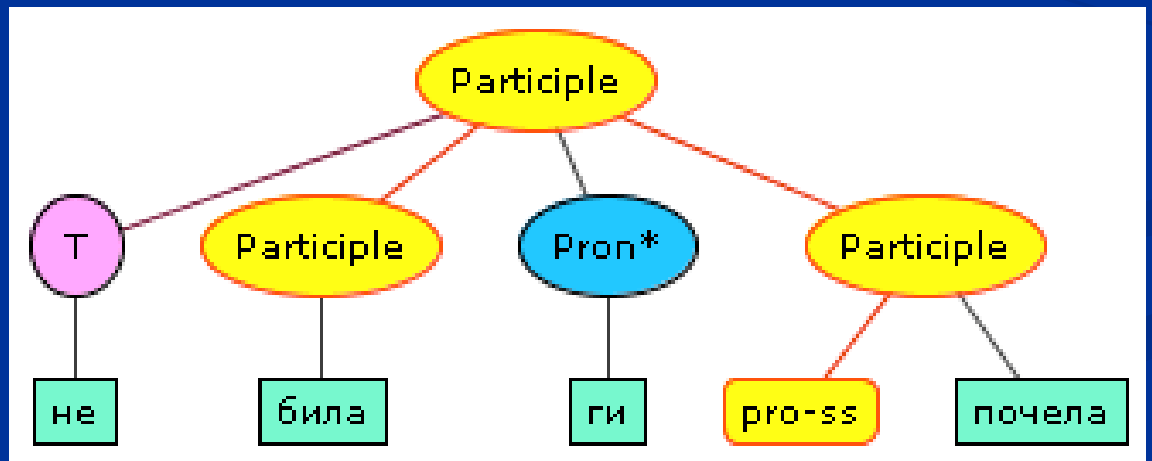


Лексикален елемент: причастие

- Причастия като прилагателни

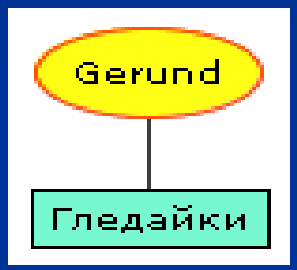


- Причастия като глаголи

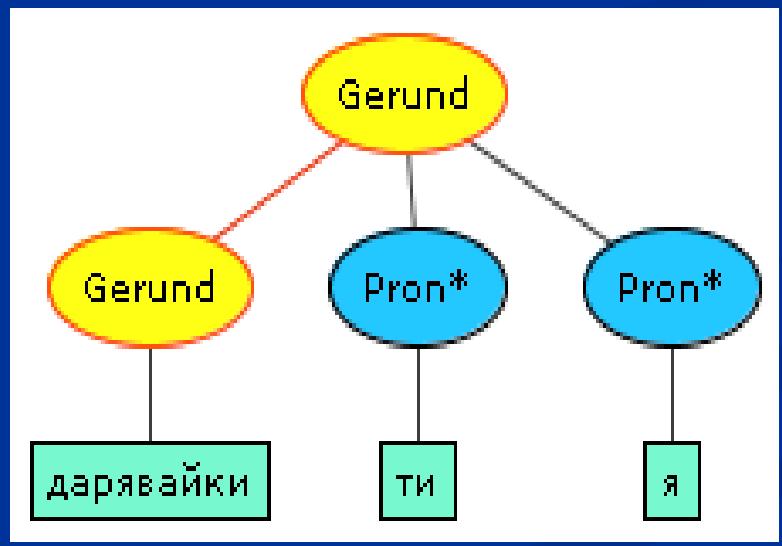


Лексикален елемент: деепричастие

- Без клитики

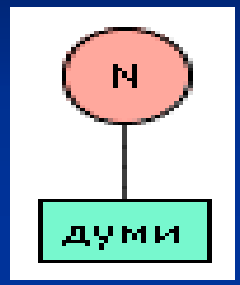


- С клитики

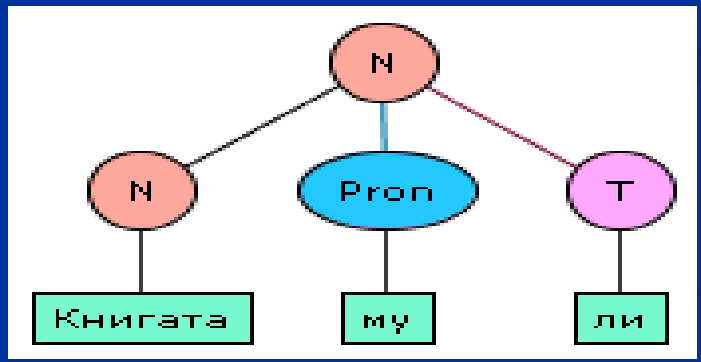


Лексикален елемент: съществително

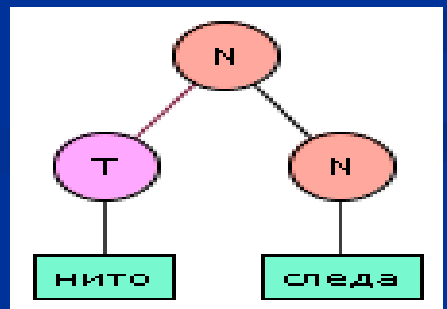
- Само съществително



- С клитики



- С емфатична дума

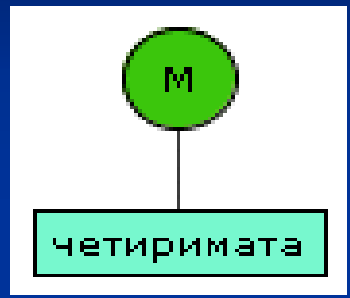


Лексикален елемент: Местоимение

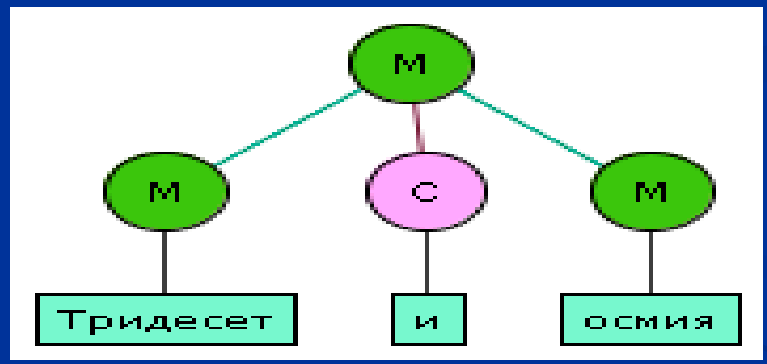
- Проекции на съществително (и прилагателно) (някой, кой, този, това)
- Проекции на прилагателно (някакъв, такъв)
- Проекции на наречие (тук, някак...)

Лексикален елемент: числително

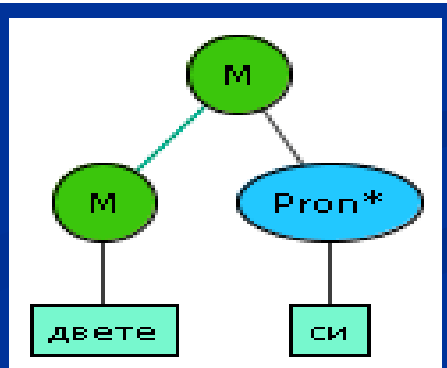
- Само елемент



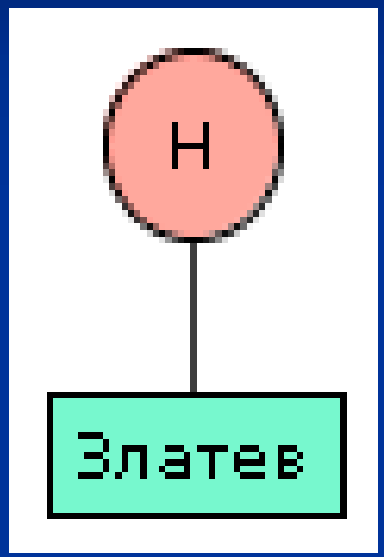
- Аналитичен елемент

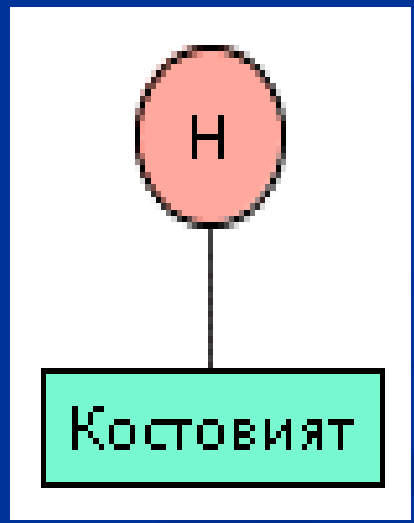


- С клитики



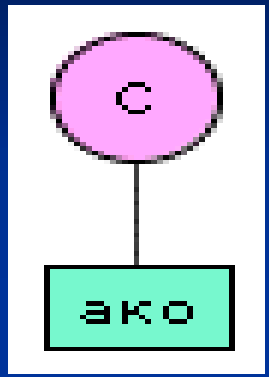
Лексикален елемент: фамилно име



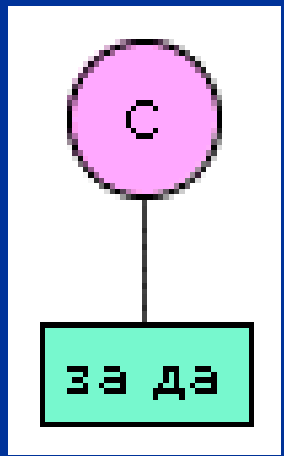


Лексикален елемент: съюз

- Прости

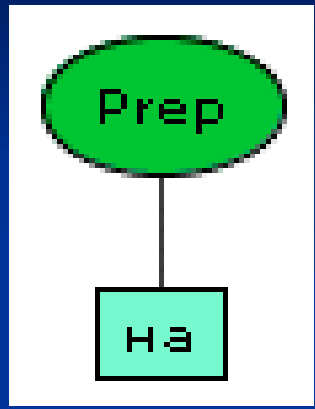


- Сложни (*след като, за да, тъй като...*)

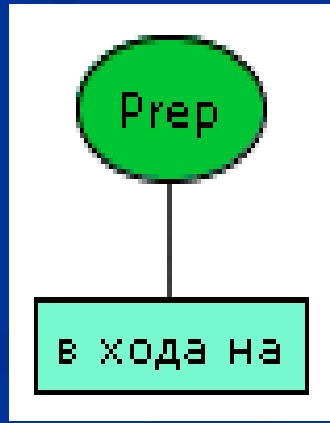


Лексикален елемент: предлог

- Прости



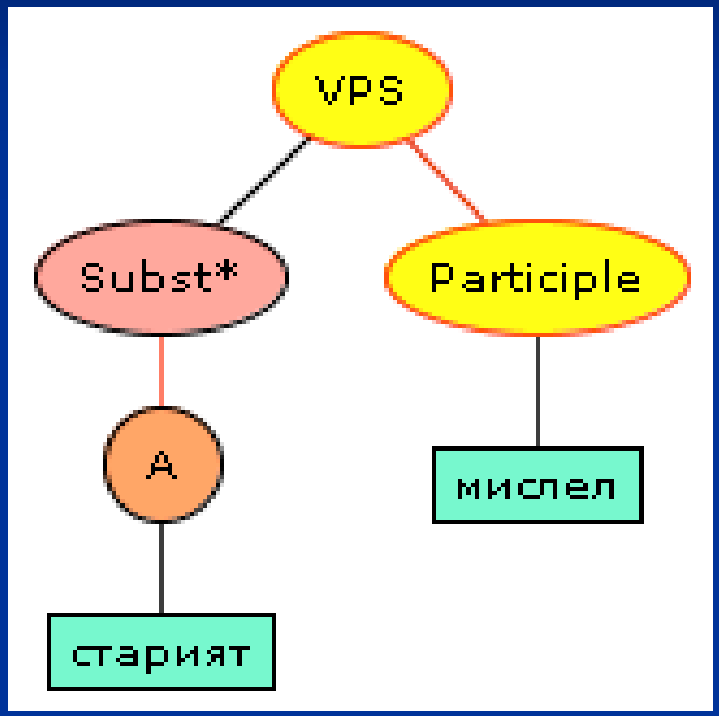
- Сложни (*в съгласие с, по отношение на, в полза на*)

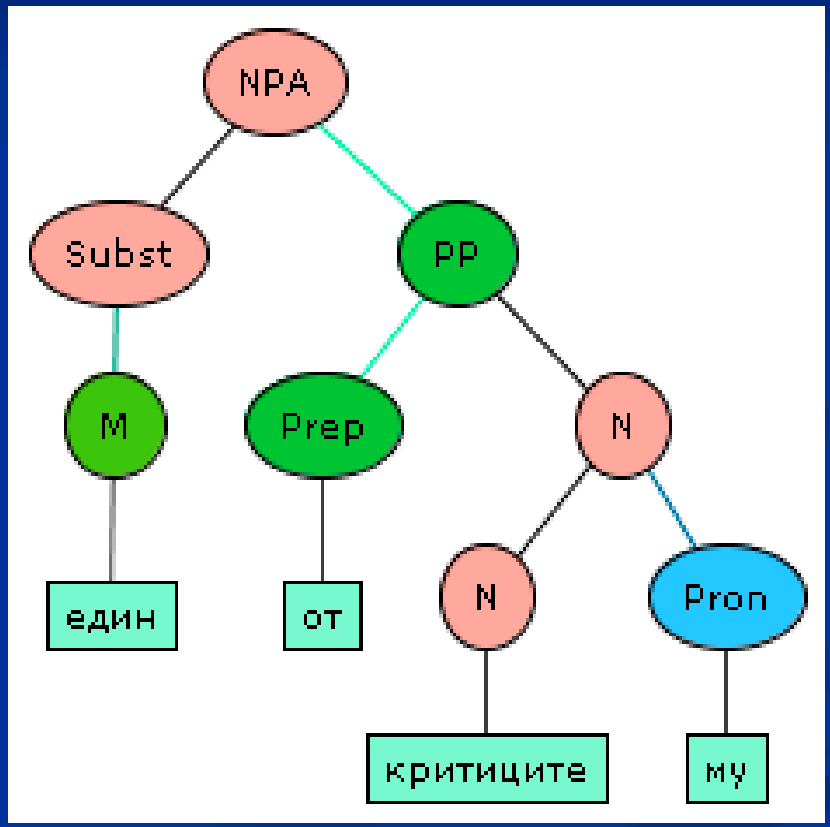


Промяна на типа (Type shifting)

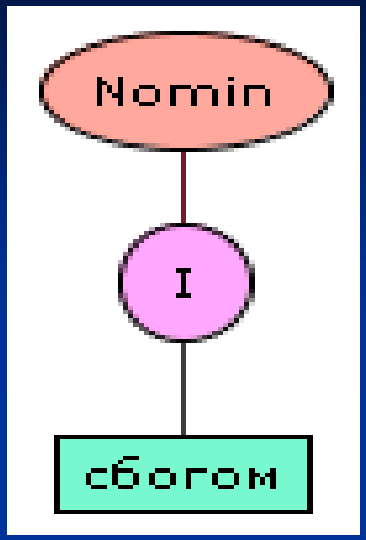
- Субстантивация - променя номинални депенденти в опори: прилагателни, числителни, причастия, местоимения
- Номинализация - променя НЕноминални елементи в номинални: предикати, междуметия и др.
- Вербализация – когато междуметия, частици или наречия изразяват предикатна функция. Един пример са междуметните сказуеми

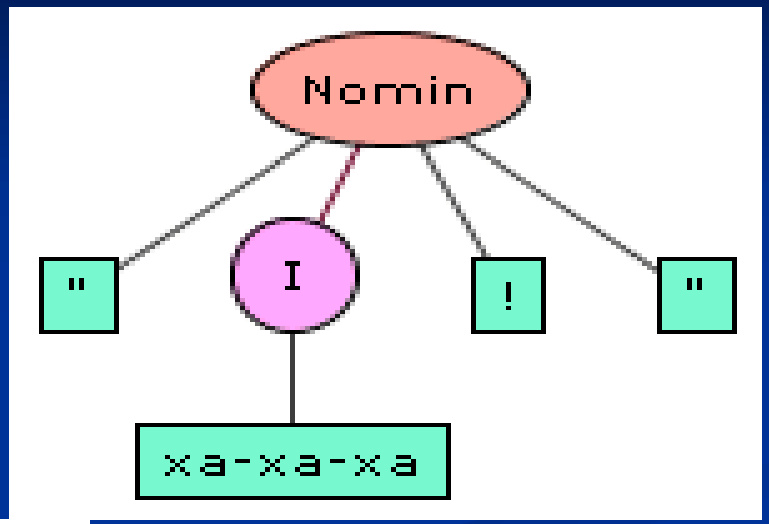
Субстантивация

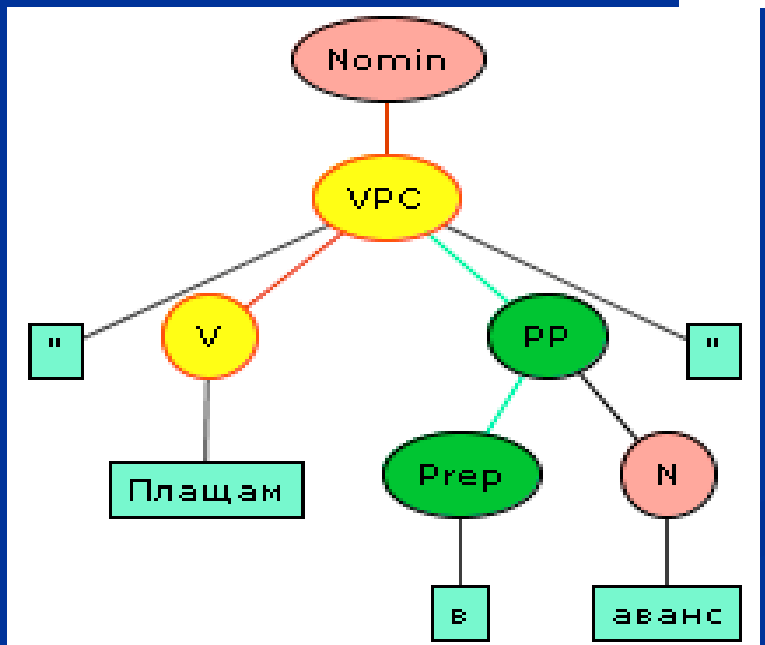


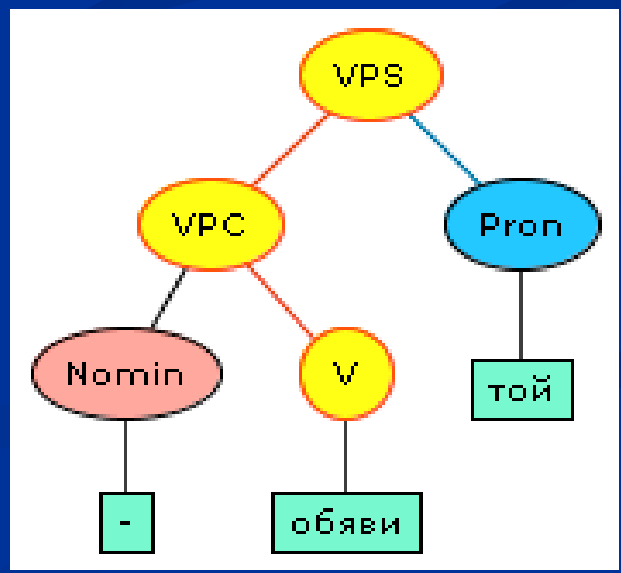


Номинализация

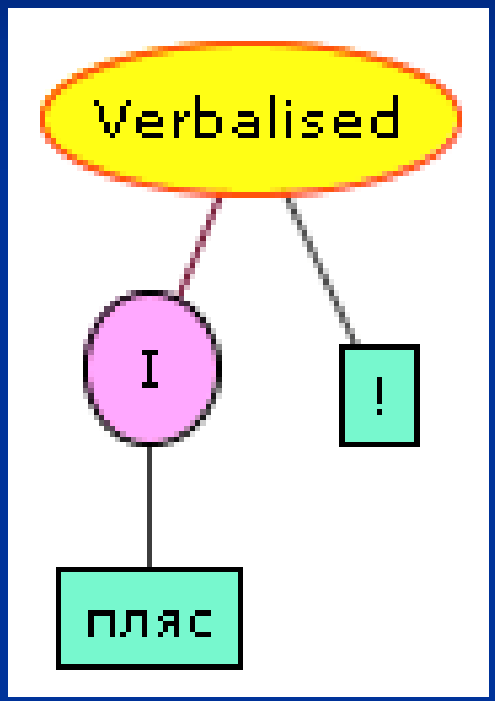


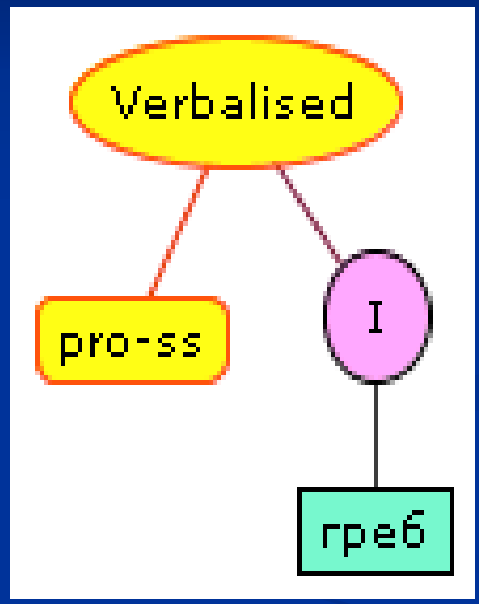






Вербализация

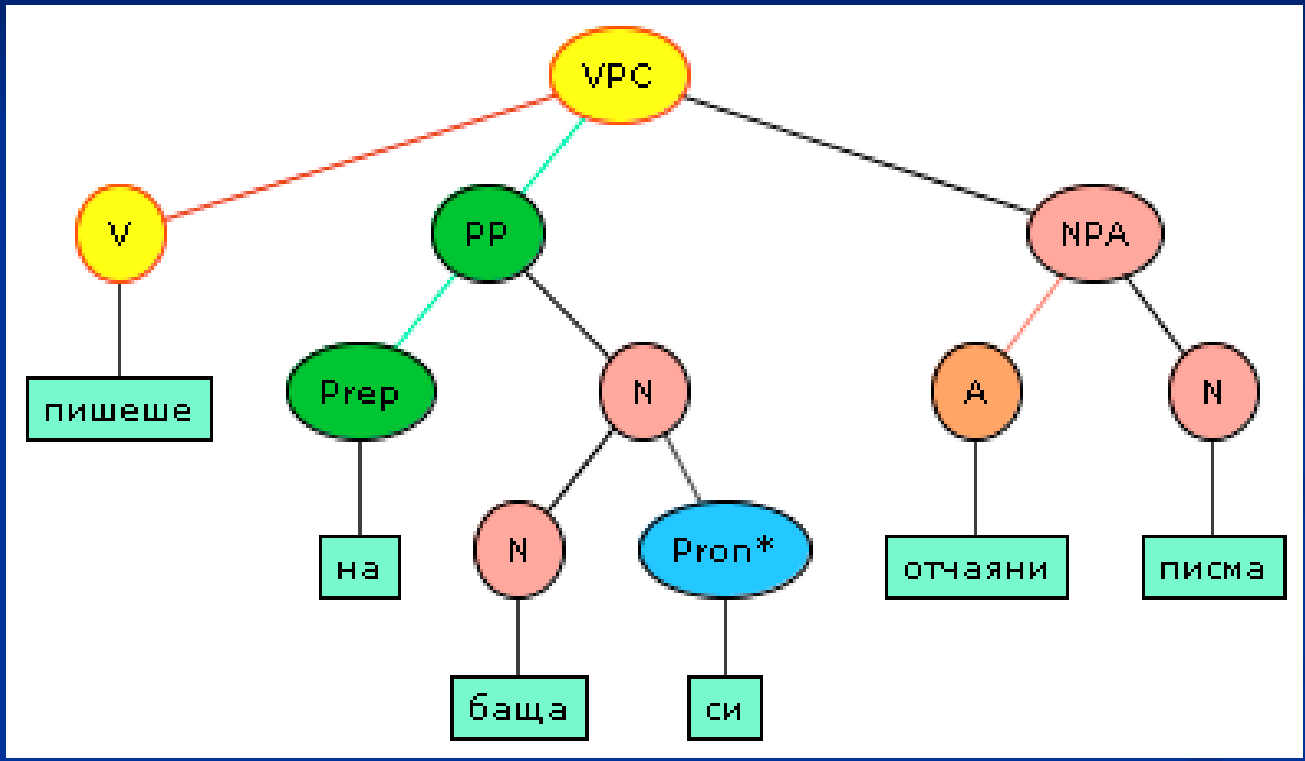




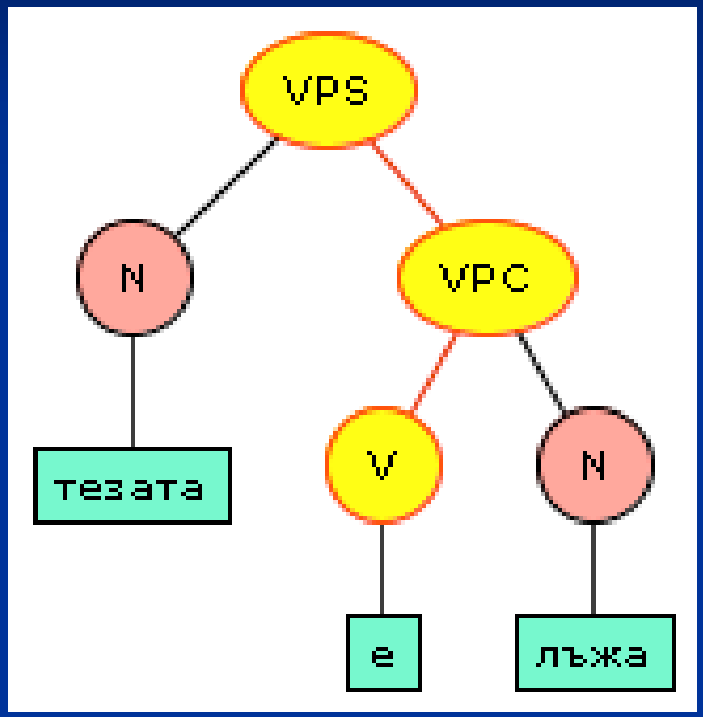
Вербална фраза: VPC

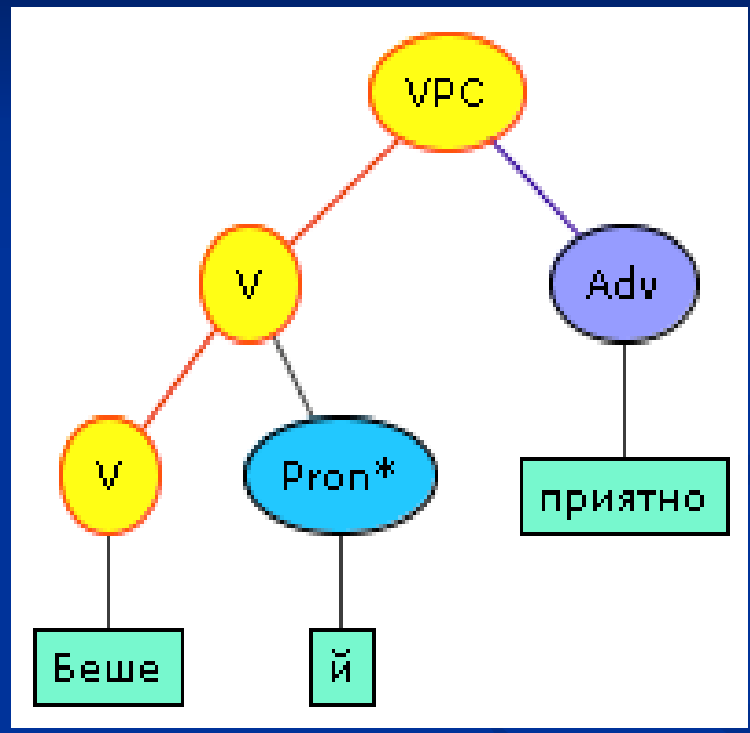
- **Опори:** V, Participle, V-Elip, VD-Elip, CoordP, Verbalized
- **Комплементи:** номинали, адвербиали, предложни групи, клаузи, координационни фрази
- **Типове:**
  - **Канонични VPC**
  - **VPC**
    - **С копула**
    - **С пасиви**
    - **С аналитични глаголни форми**
  - **VPC с комплементи-клаузи (контрол)**
  - **VPC с поддържащи и “леки” глаголи**

Канонични VPC

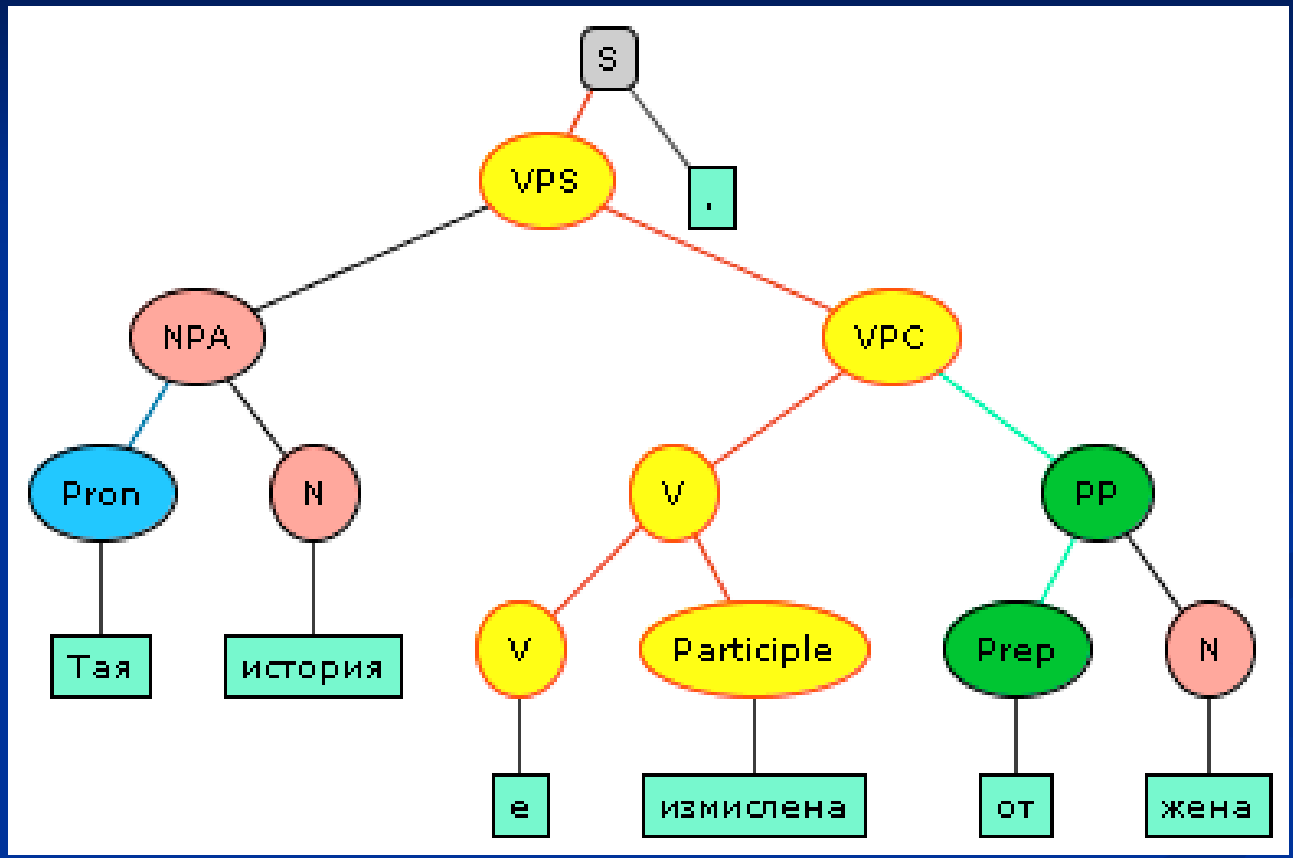


VPC с копула

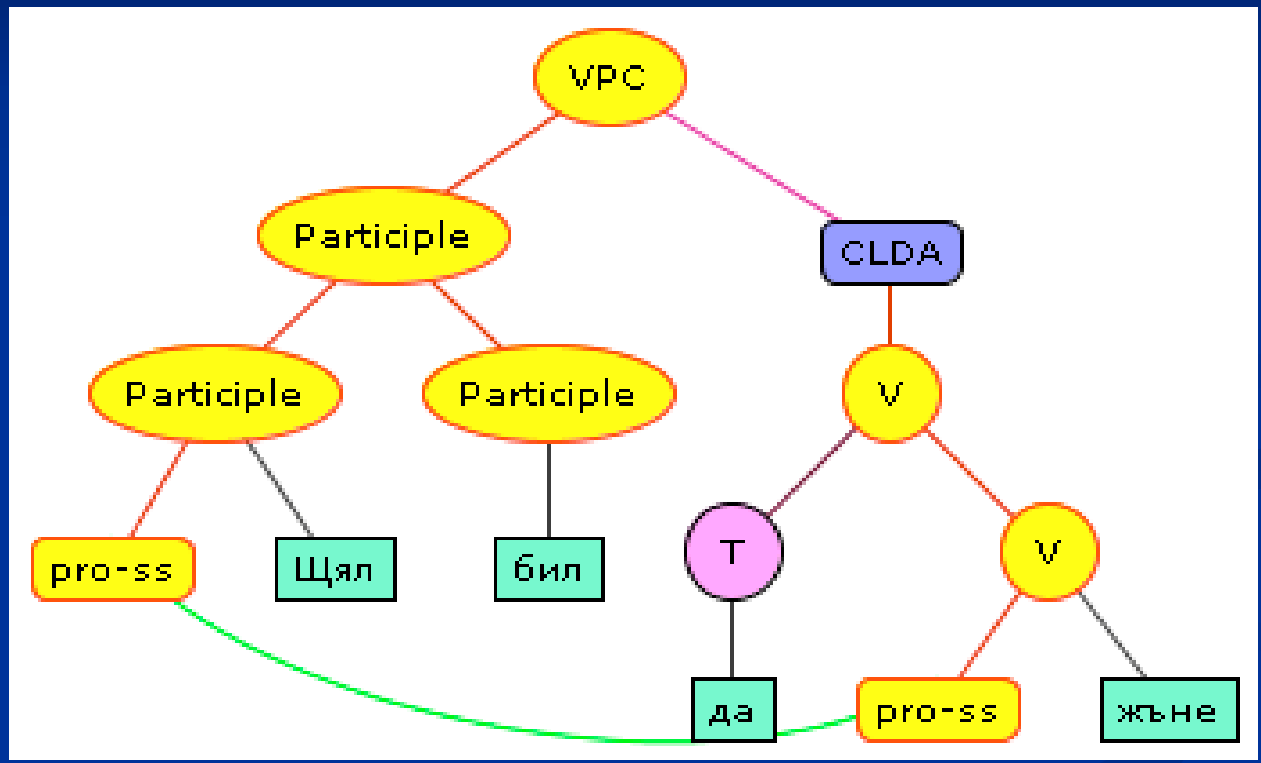




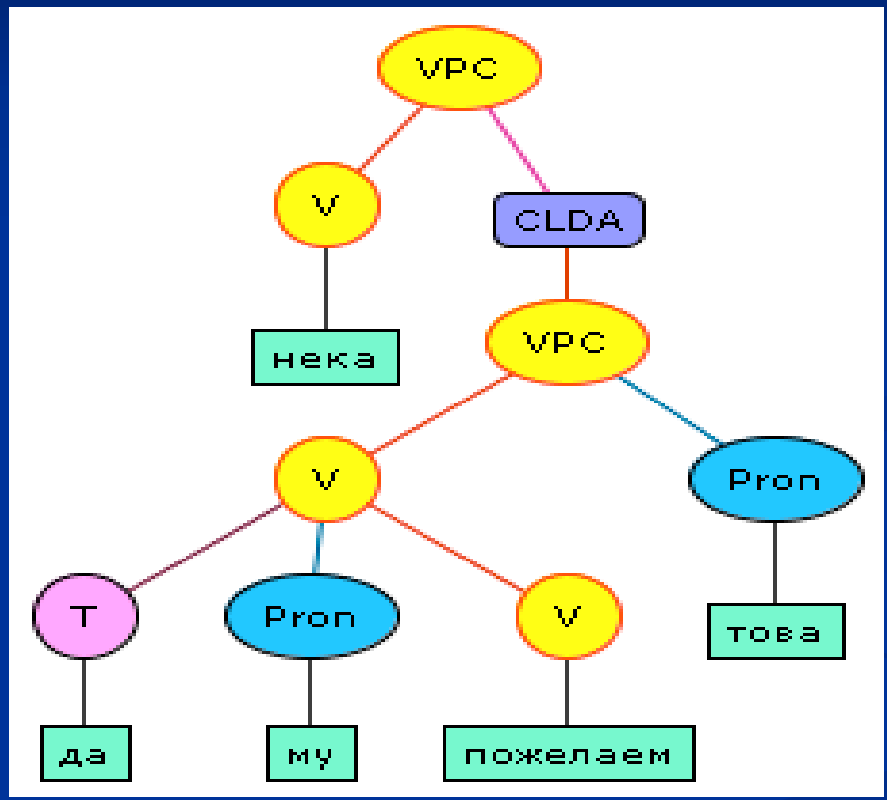
Пасиви


S
VPS
.
NPA
VPC
Pron
N
V
PP
V
Participle
Prep
N
Тая
история
е
измислена
от
жена

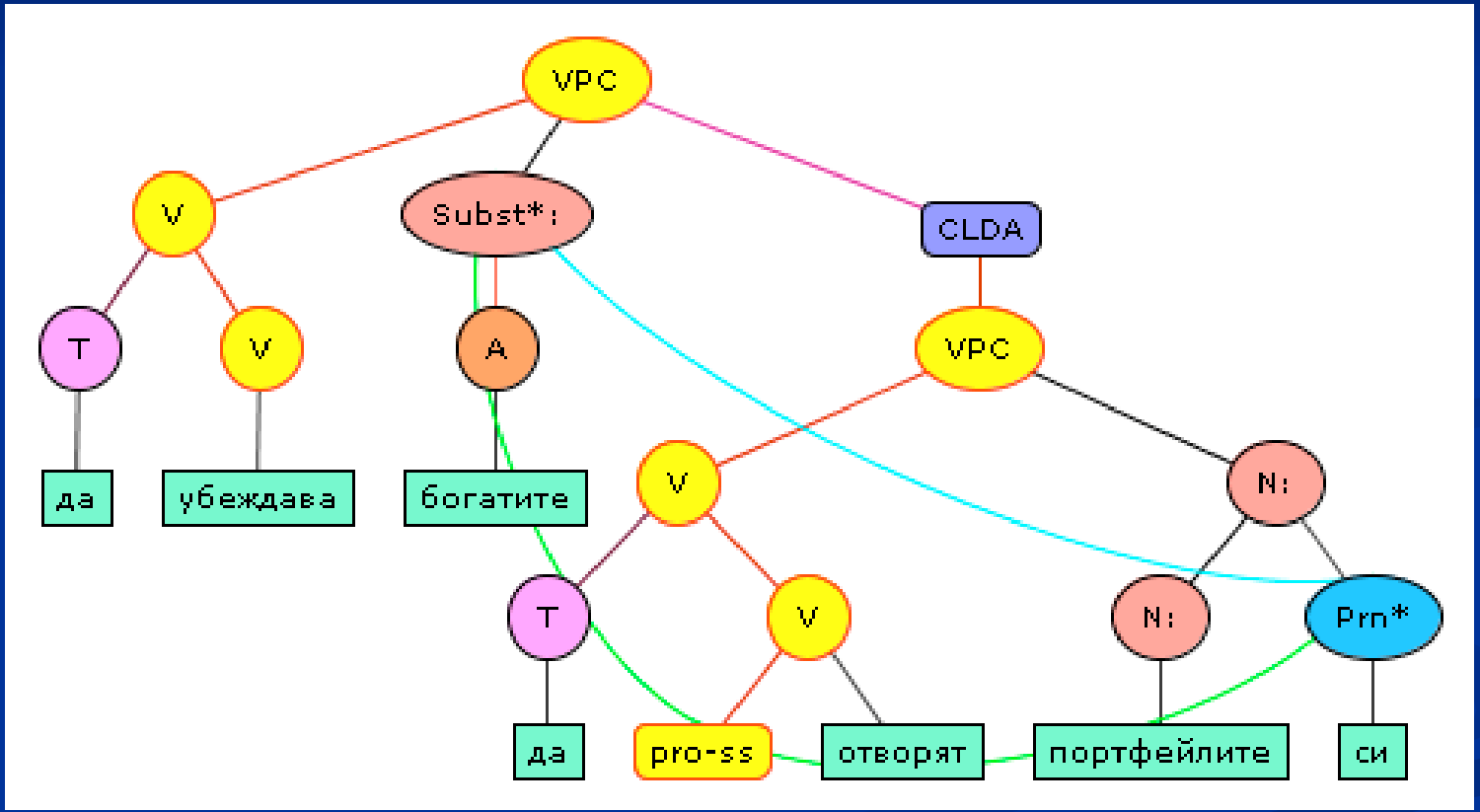
Аналитични вербални форми като VPC



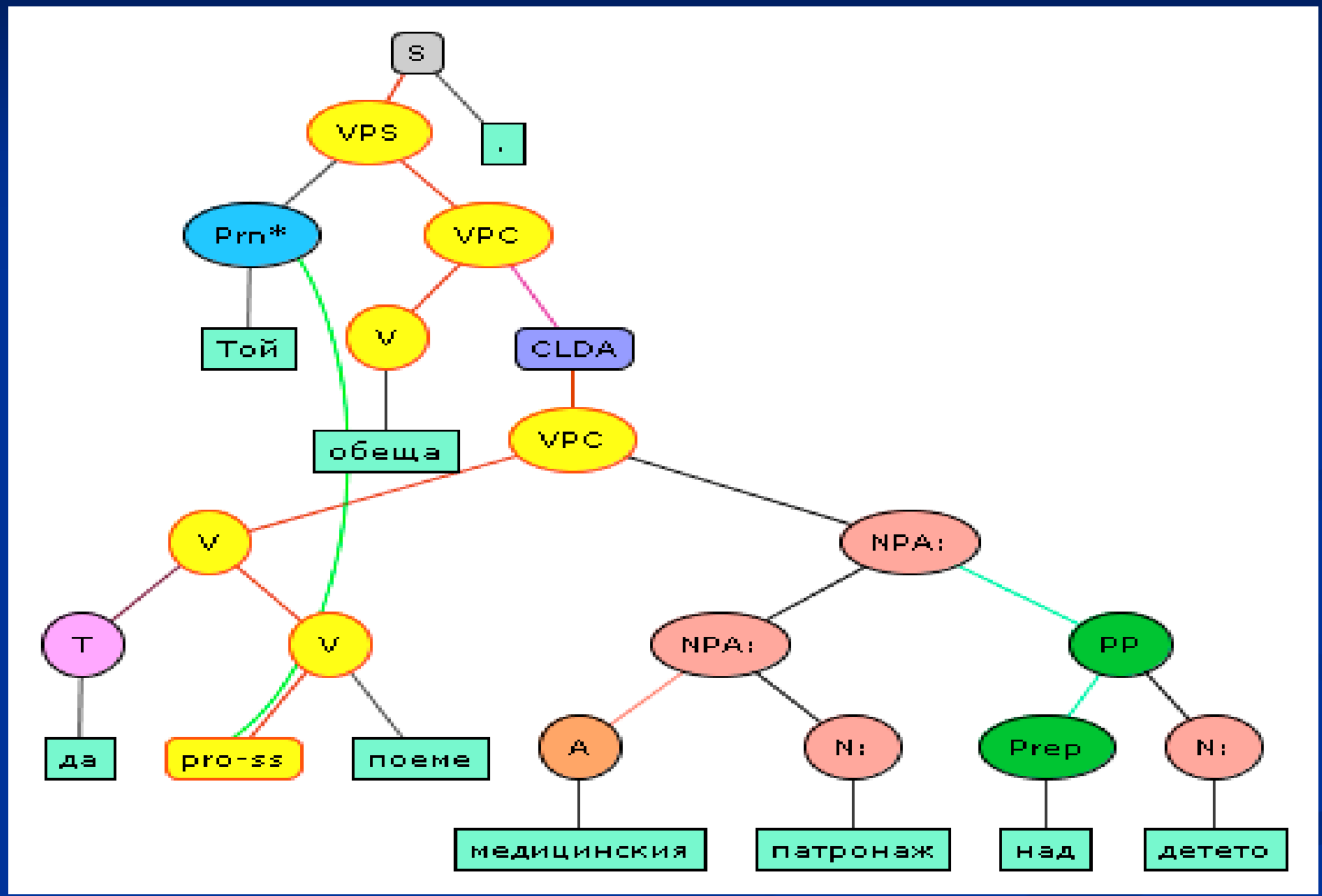
С комплемент-клауза



Контрол (1)


VPC
V
Subst*:
CLDA
T
V
A
VPC
да
убеждава
богатите
V
N:
T
V
N:
Prn*
да
pro-ss
отворят
портфейлите
си

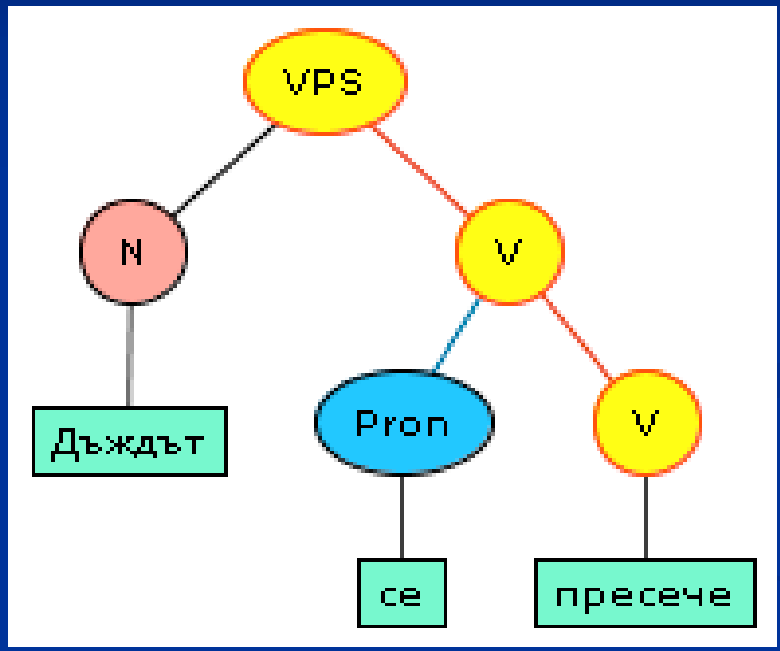
Контрол (2)


S
VPS
.
Prn*
VPC
Той
V
CLDA
обеща
VPC
V
NPA:
T
V
NPA:
PP
A
N:
Prep
N:
да
pro-ss
поеме
медицинския
патронаж
над
детето

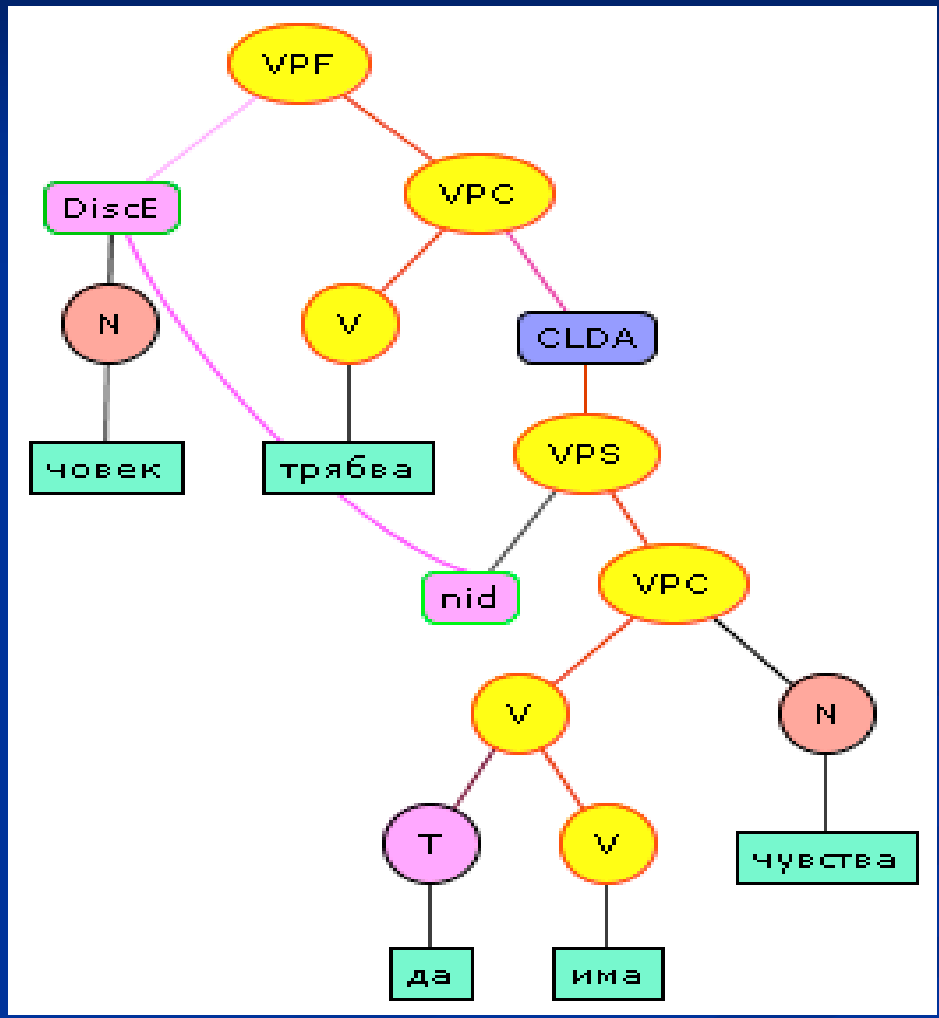
Вербална фраза: VPS

- **Опора:** V, VPC, Participle, V-Elip, VD-Elip, CoordP, Verbalized
- **Подлог:** номинали, адвербиали, предложни групи, клаузи, координационни фрази
- **Типове:**
  - **Канонични VPS**
  - **Екстрактнат подлог**
  - **PP като подлог**
  - **CL като подлог**

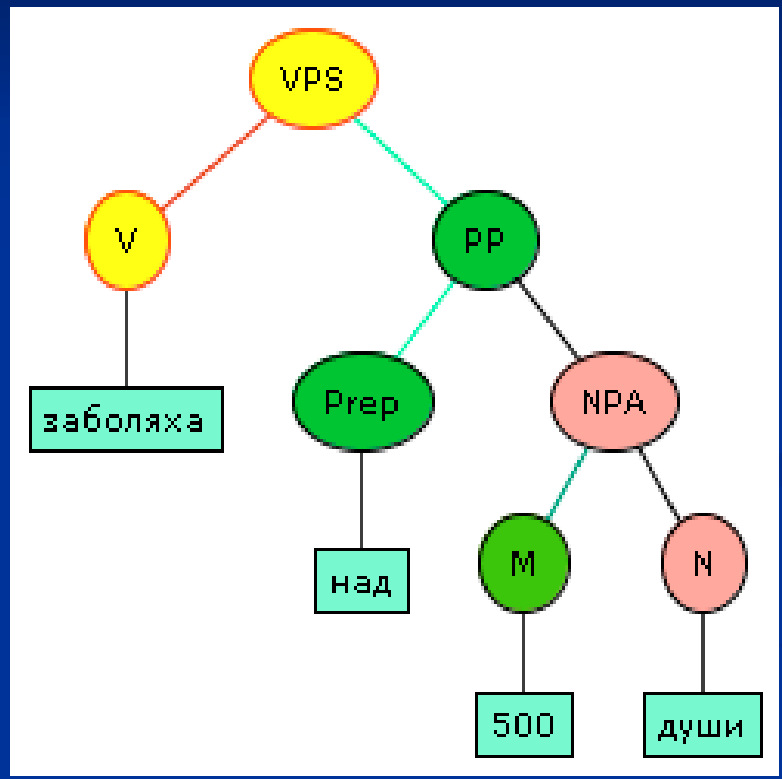
Канонично VPS



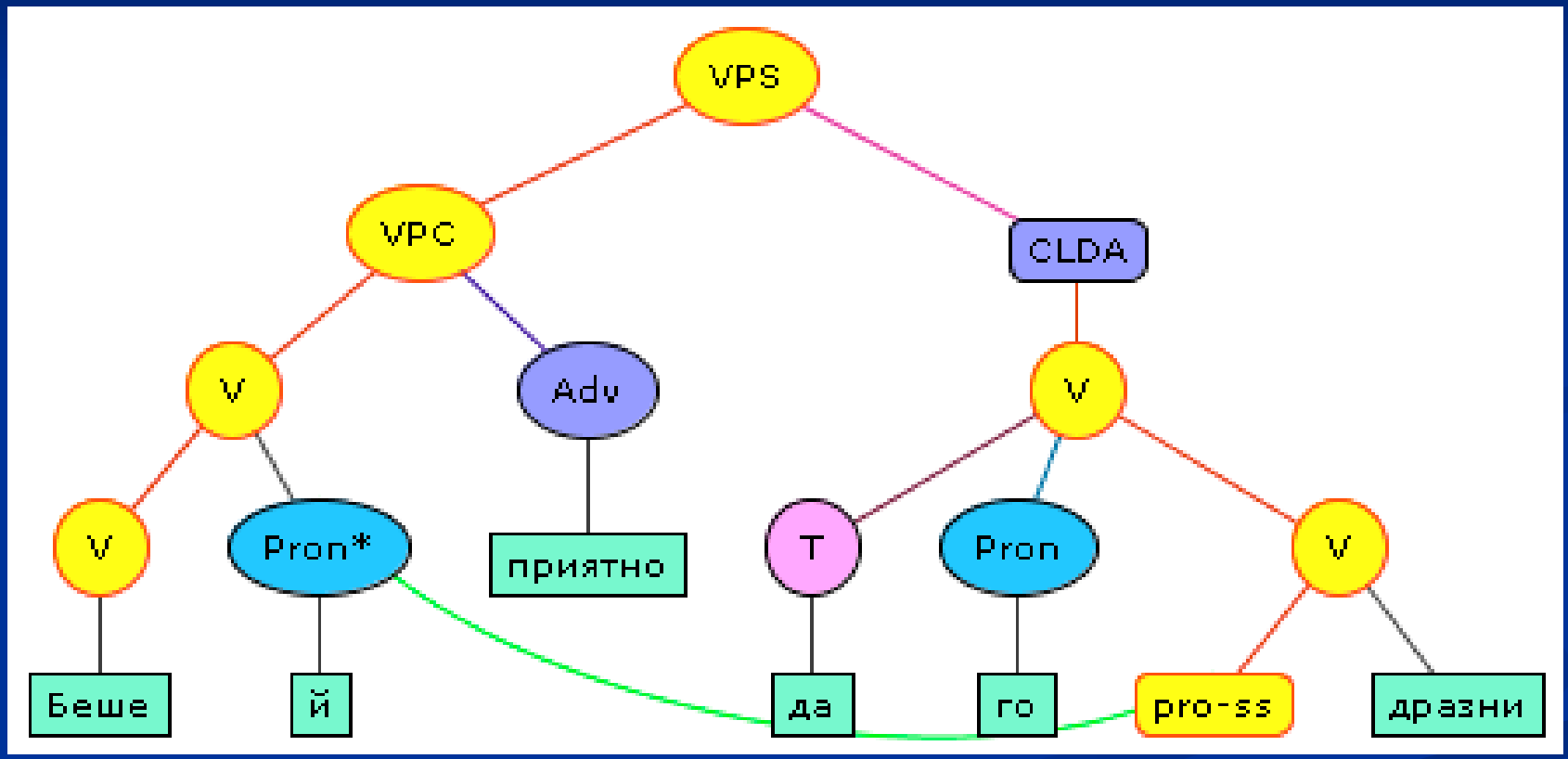
Екстрактнат подлог



PP като подлог



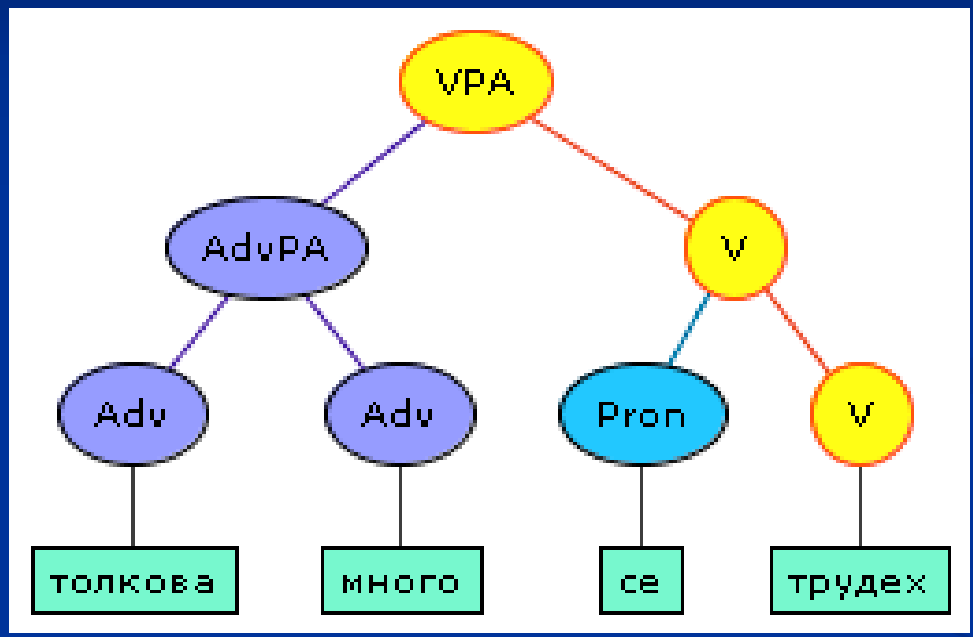
CL като подлог



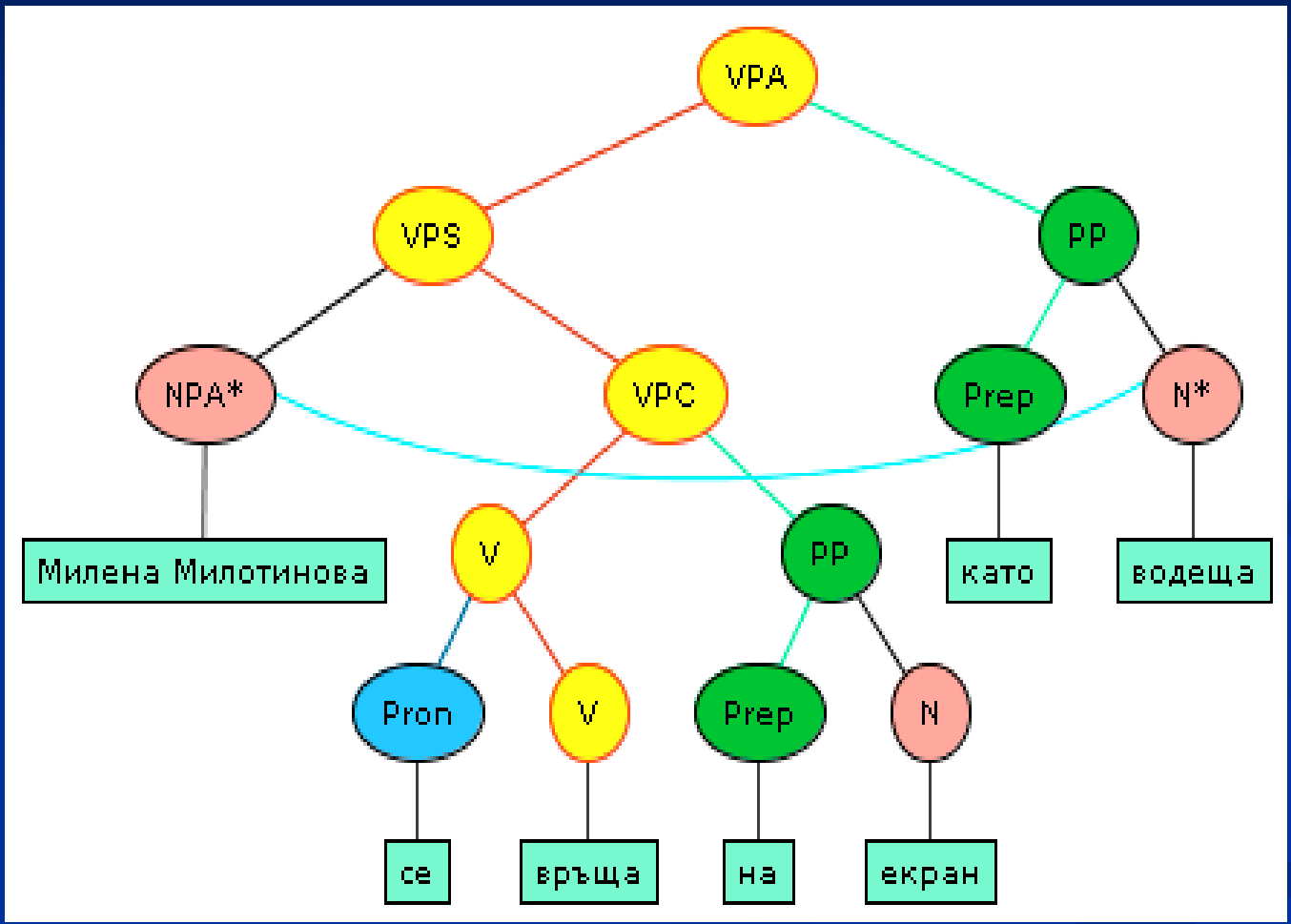
Вербална фраза: VPA

- **Опори:** V, VPC, VPS, VPA, Participle, V-Elip, VD-Elip, CoordP, Verbalized
- **Адюнкти:** номинали, адвербиали, предложни групи, клаузи, координационни фрази
- **Типове:**
  - **Канонични VPA**
  - **Малки изречения като VPA**
  - **Клаузи като VPA**
  - **VPA с въпросителни частици**

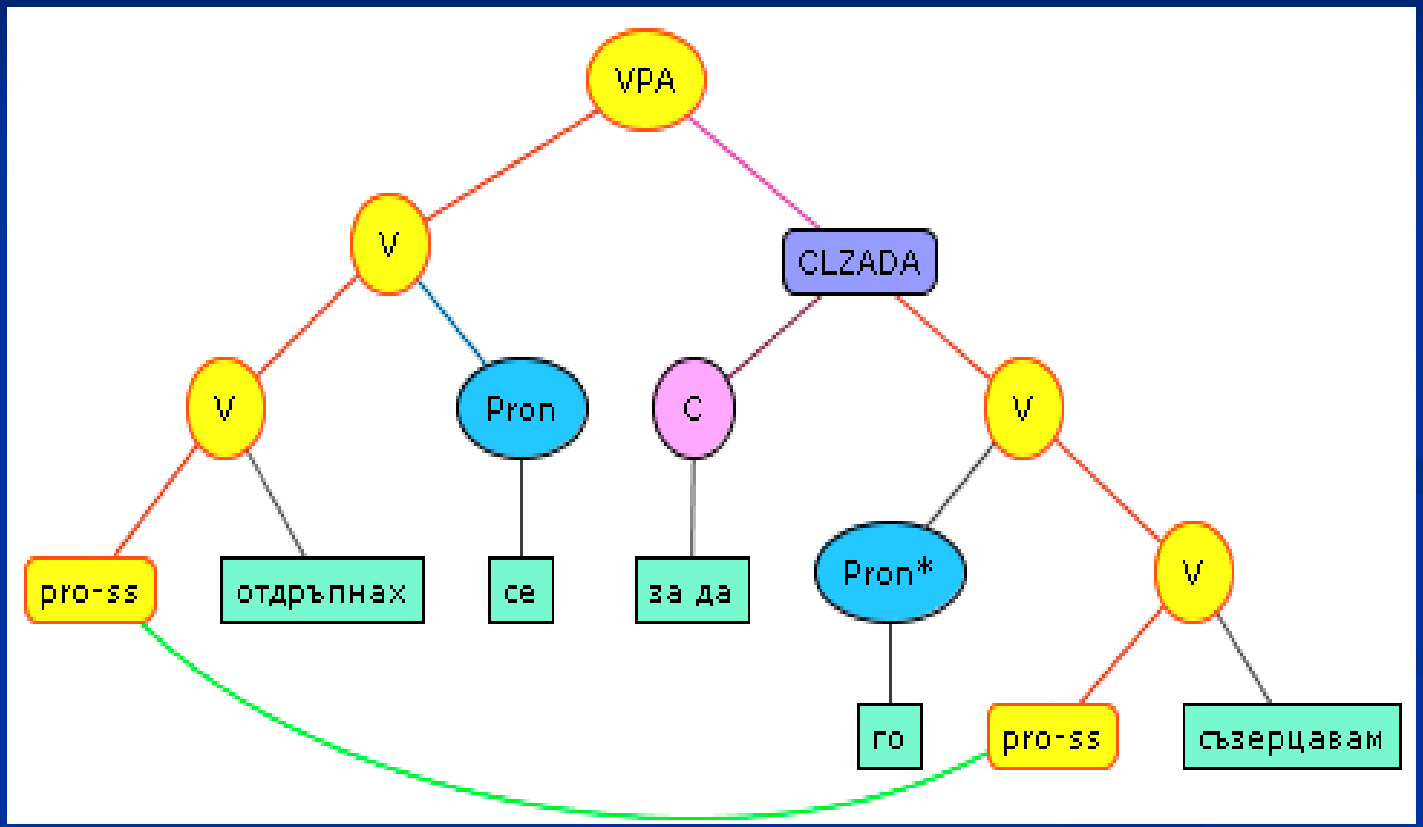
Канонично VPA



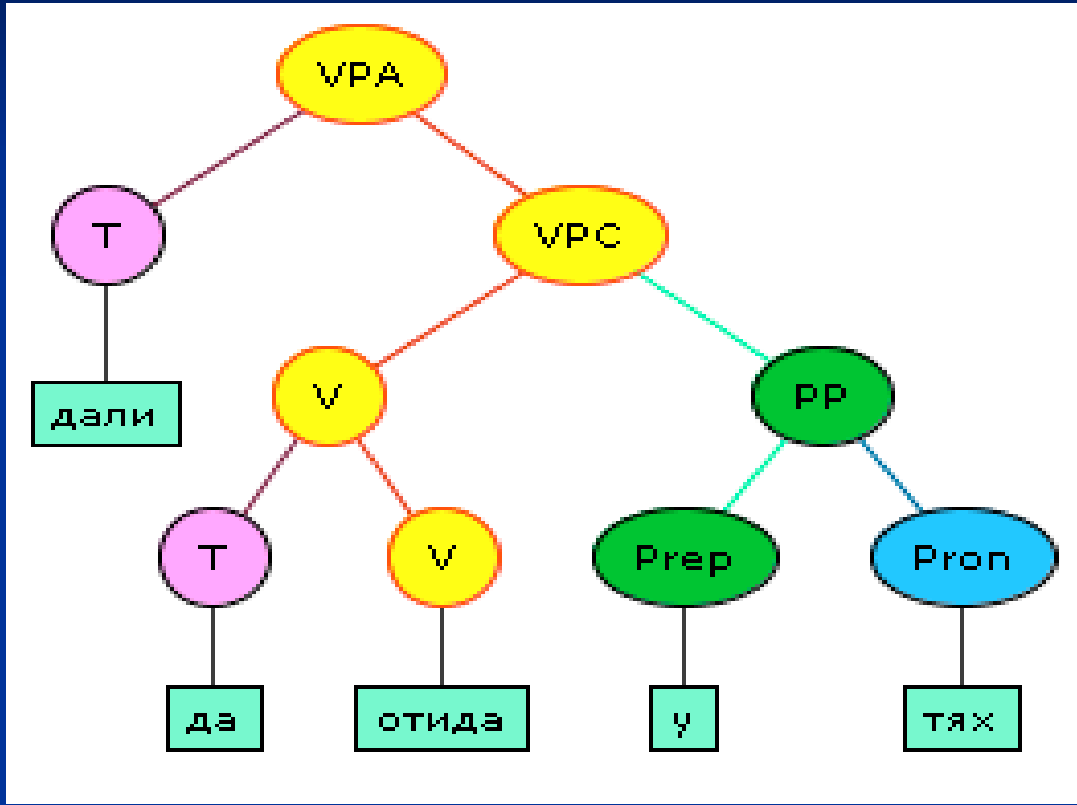
Малко изречение като VPA



CL като VPA


VPA
V
CLZADA
V
Pron
C
V
pro-ss
отдръпнах
се
за да
Pron*
V
го
pro-ss
съзерцавам

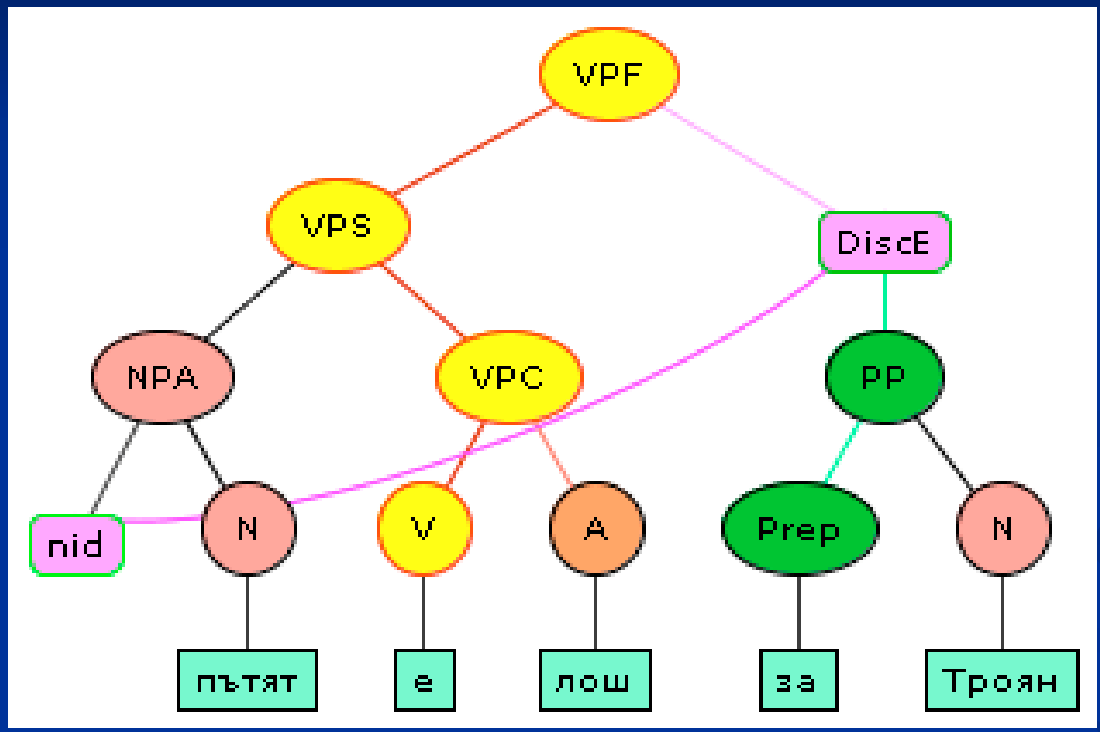
VPA с въпросителни частици



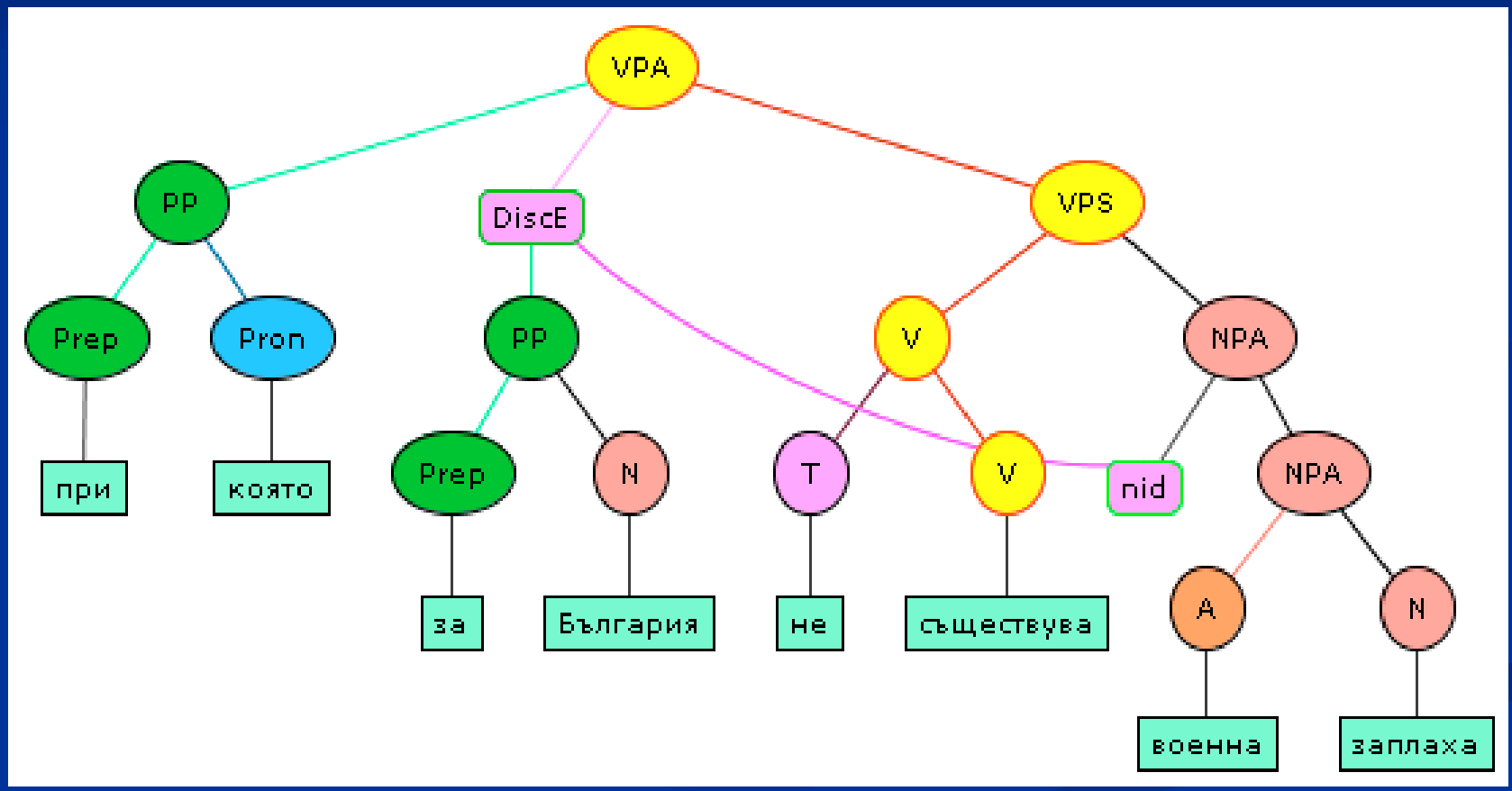
Вербална фраза: VPF

- **Опори:** V, VPC, VPS, VPA, Participle, CoordP
- **Запълващи фрази (fillers):** DiscE елементи
- **Типове:**
  - Свързани към VPF
  - Свързани към друг родител

Пример



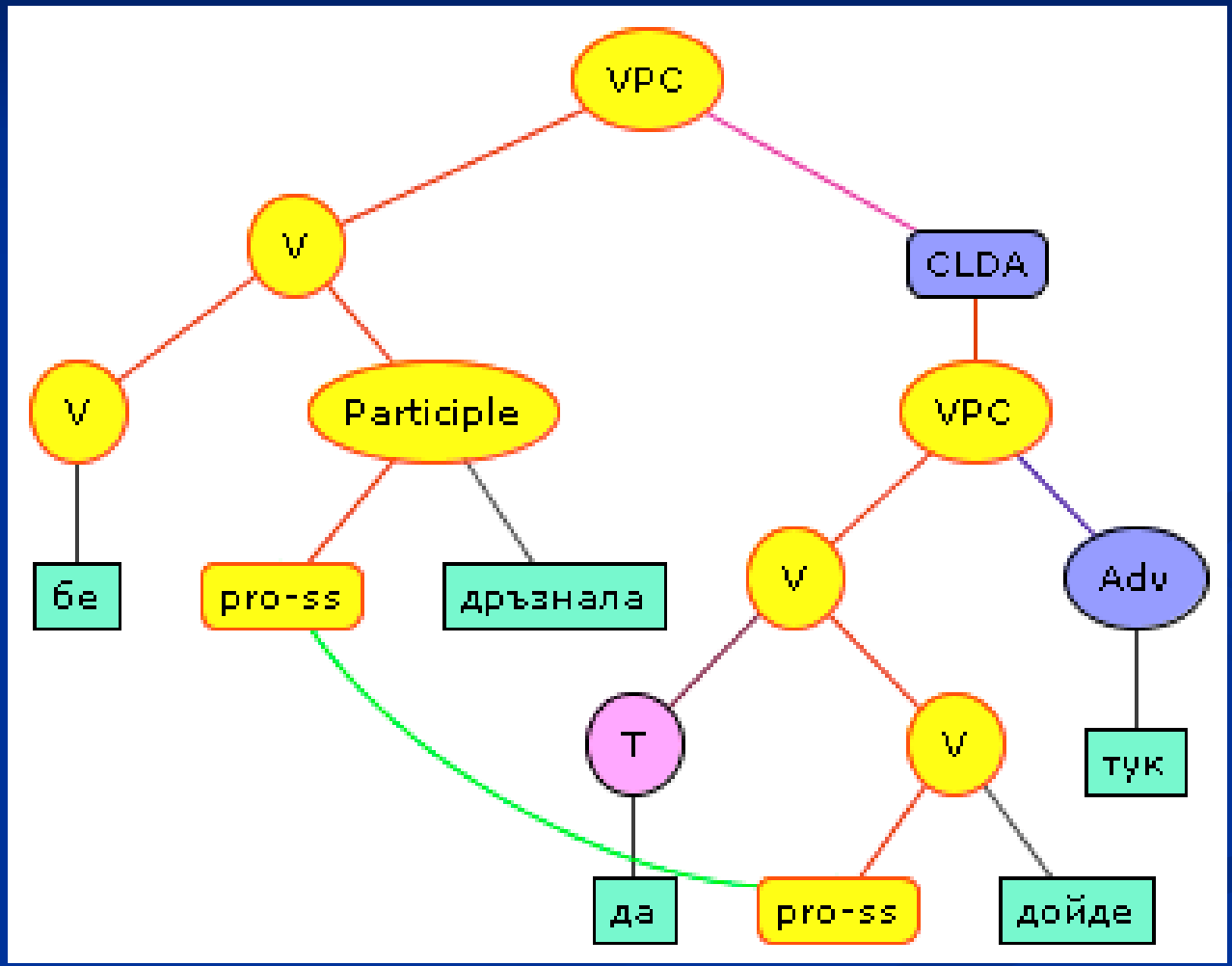
Пример


VPA
PP
DiscE
VPS
Prep
Pron
PP
V
NPA
при
която
Prep
N
T
V
nid
NPA
за
България
не
съществува
A
N
военна
заплаха

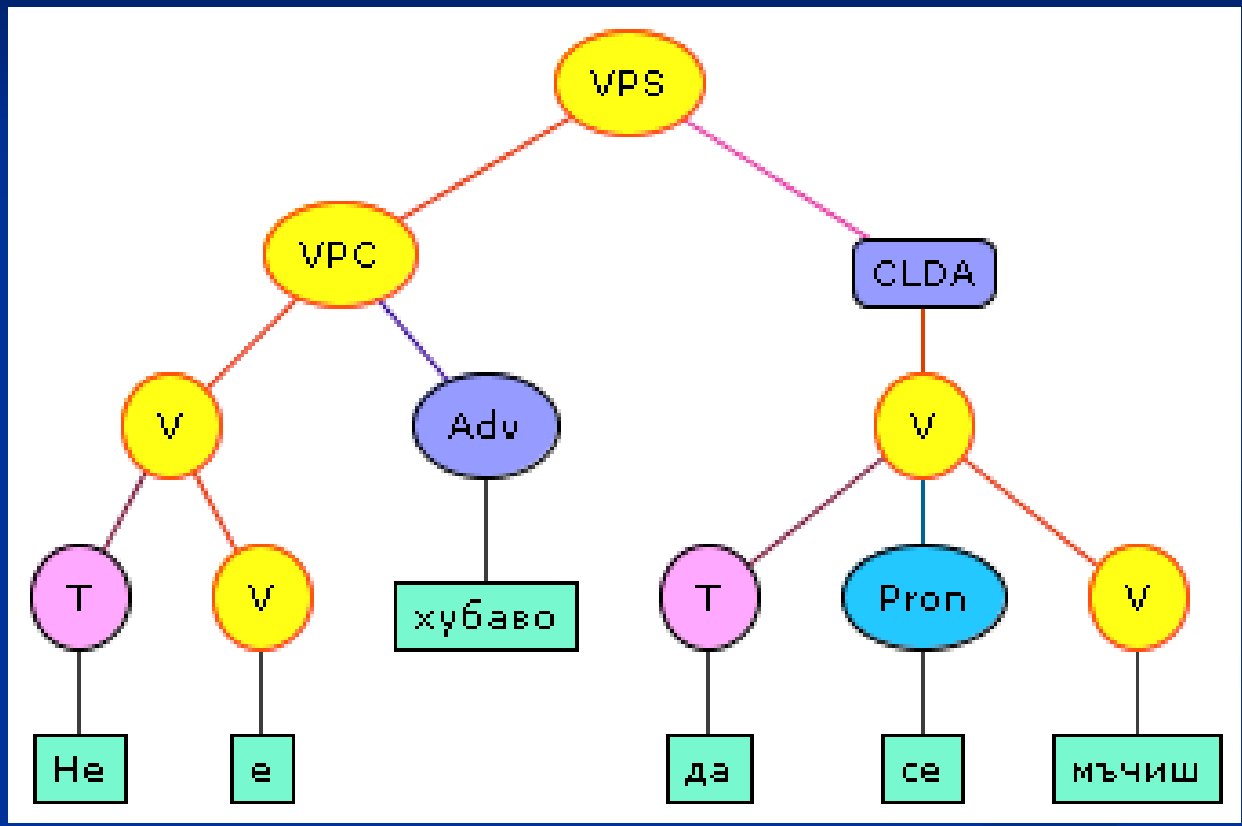
Клаузи

- Дефиниция – наситена вербална фраза
- **Роли** – комплемент, подлог, адюнкт
- Типове според въвеждащата дума (не по функция) – CLR, CLQ, CLDA, CLZADA, CLCHE, CL (общ тип)

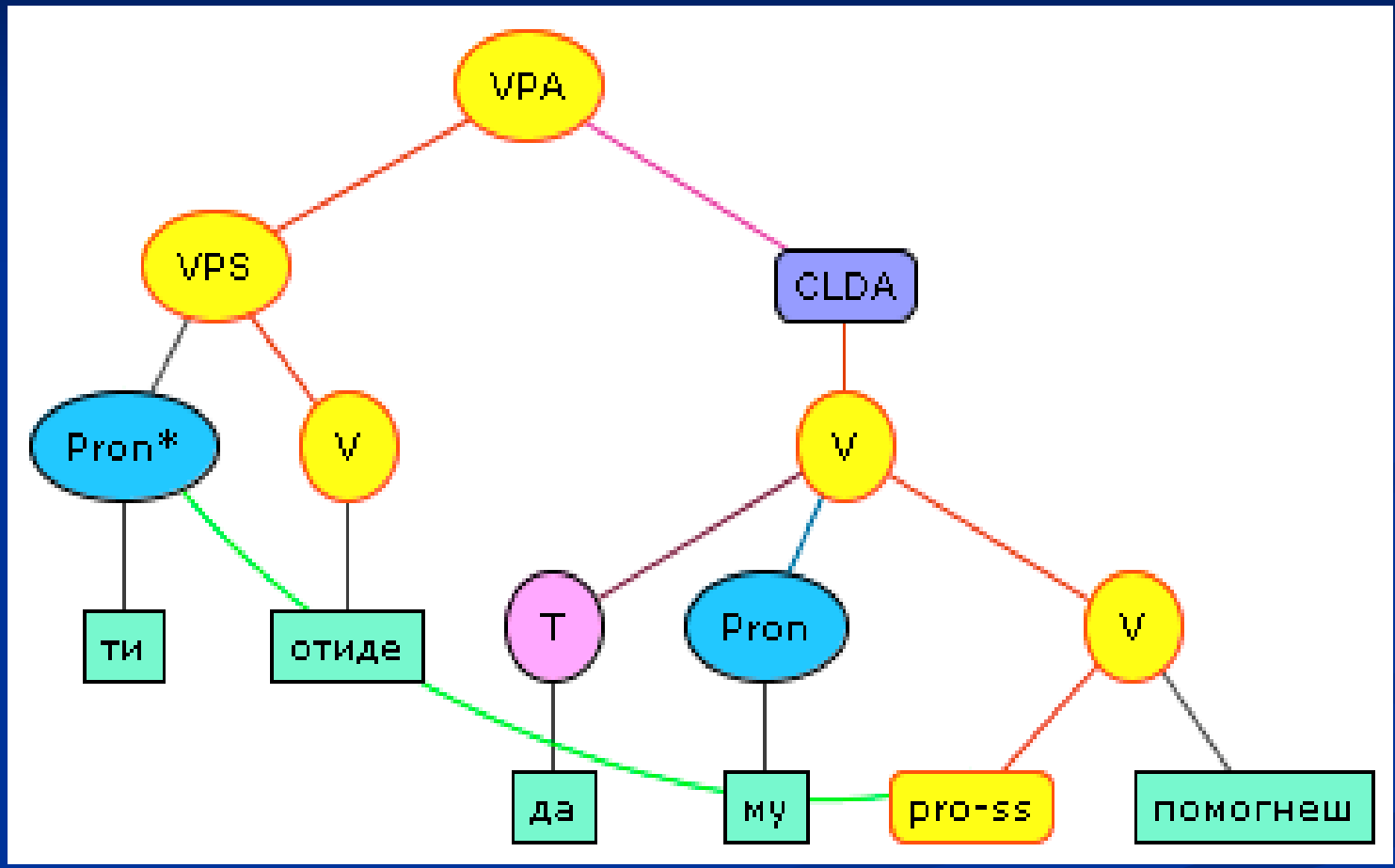
CLDA като комплемент



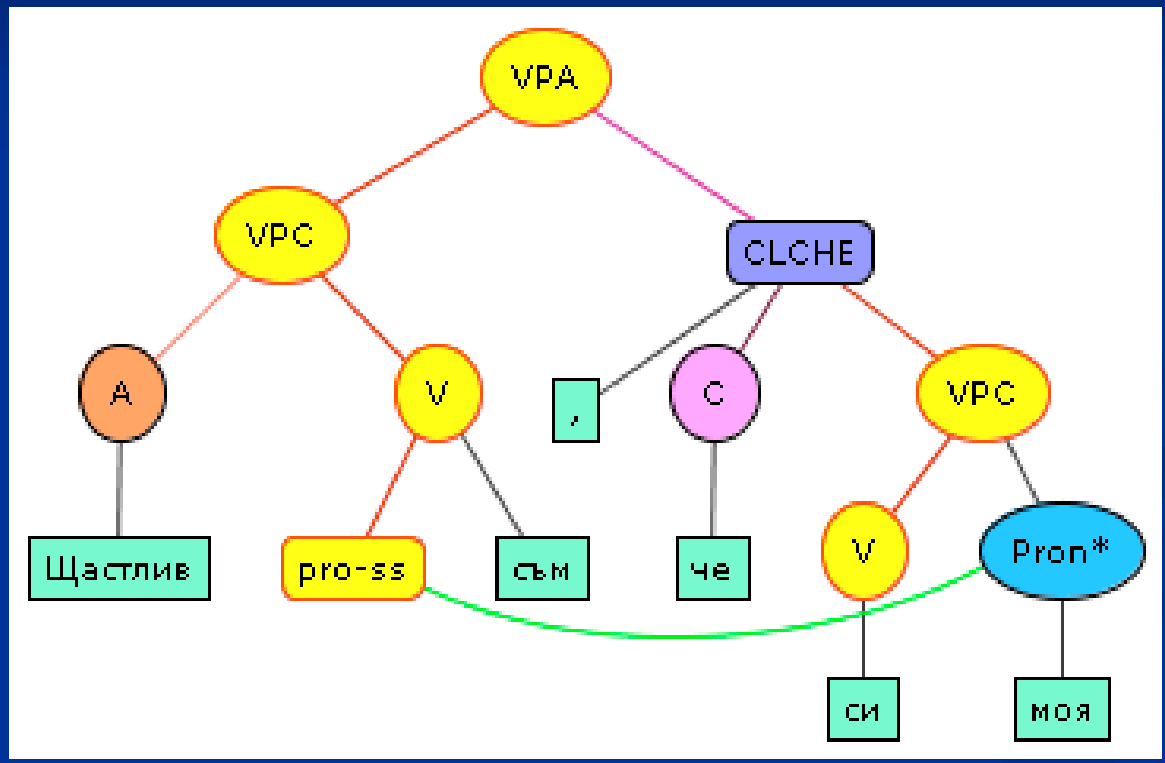
CLDA като подлог



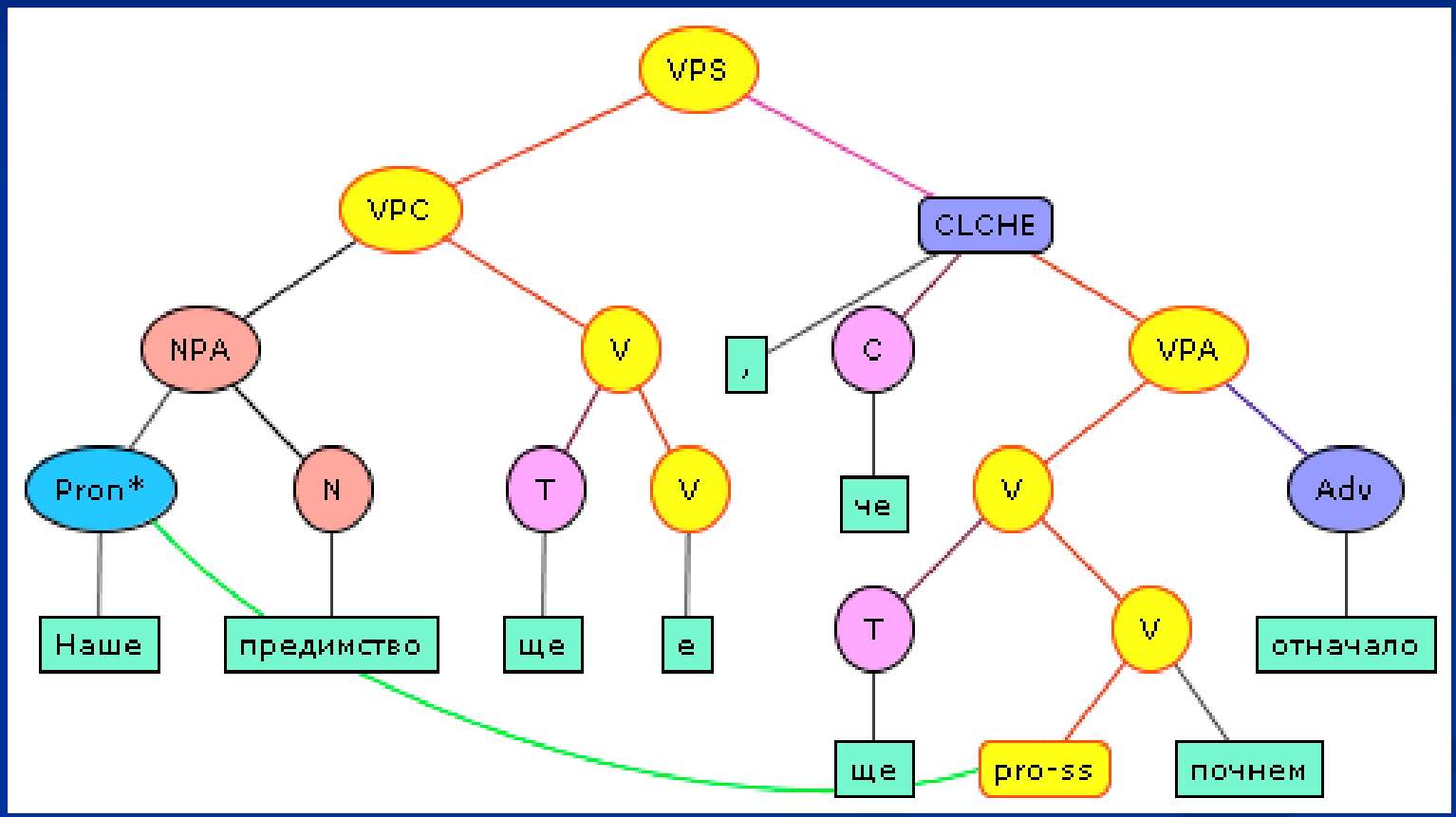
CLDA като адюнкт



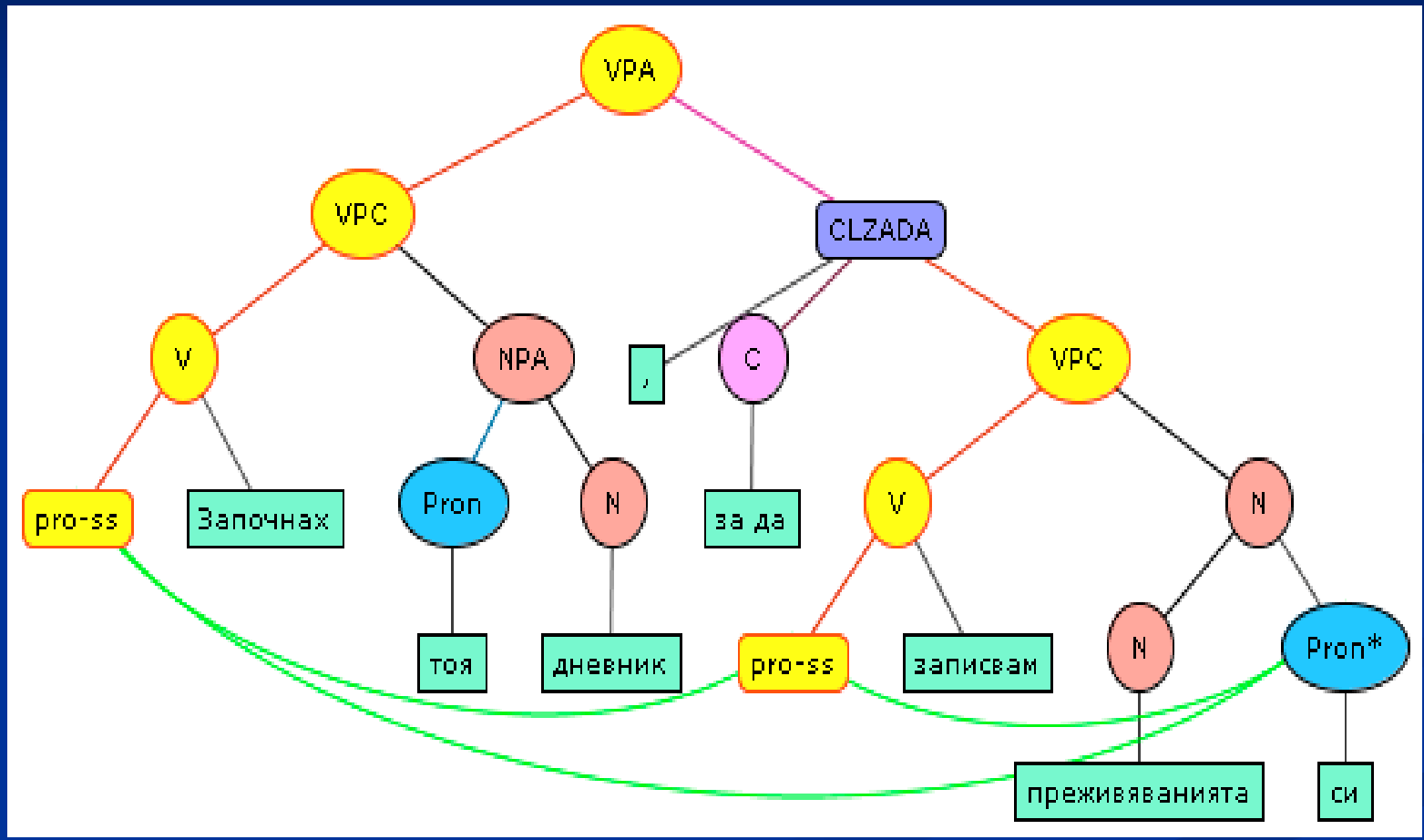
CLCHE като адюнкт



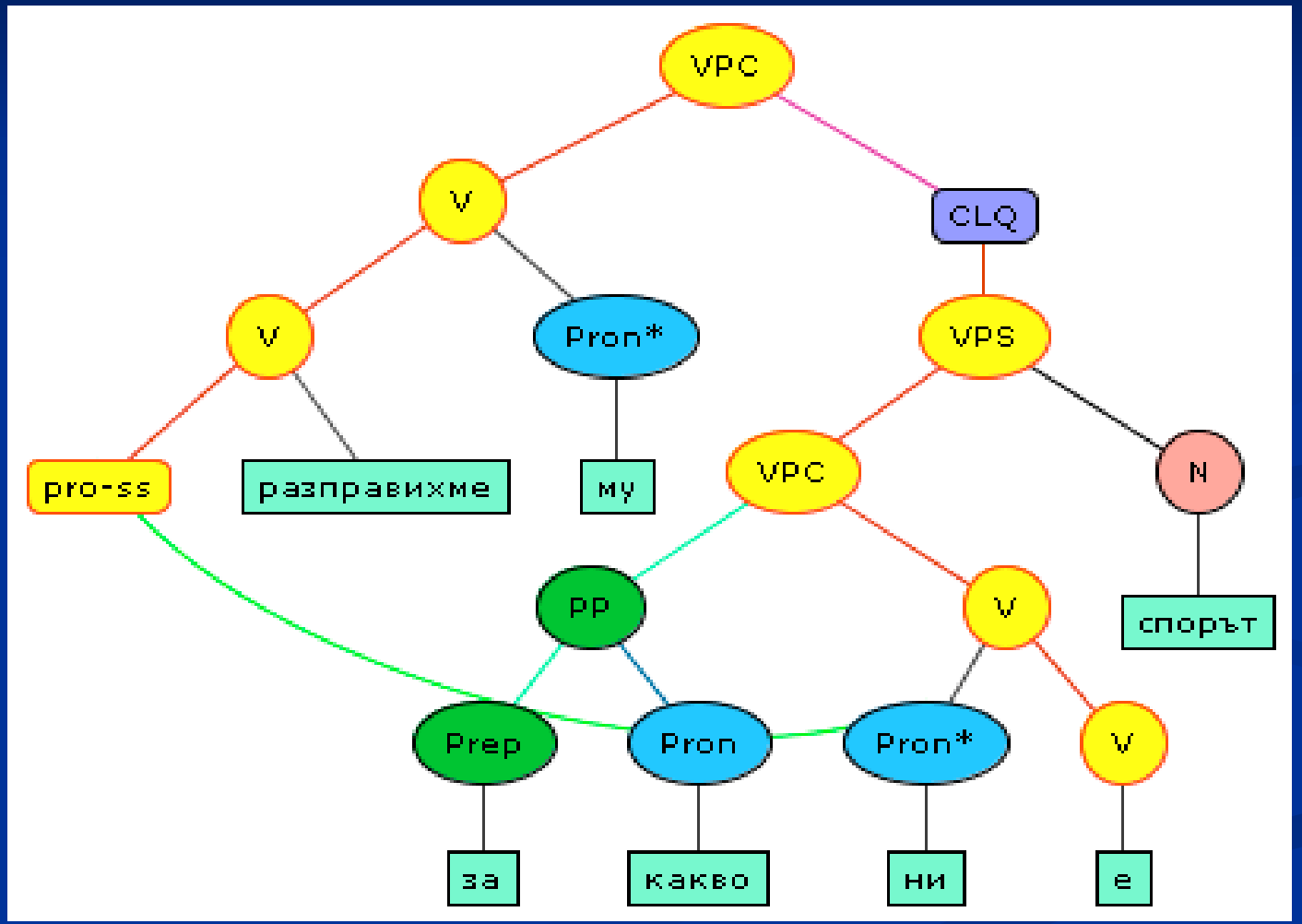
CLCHE като подлог



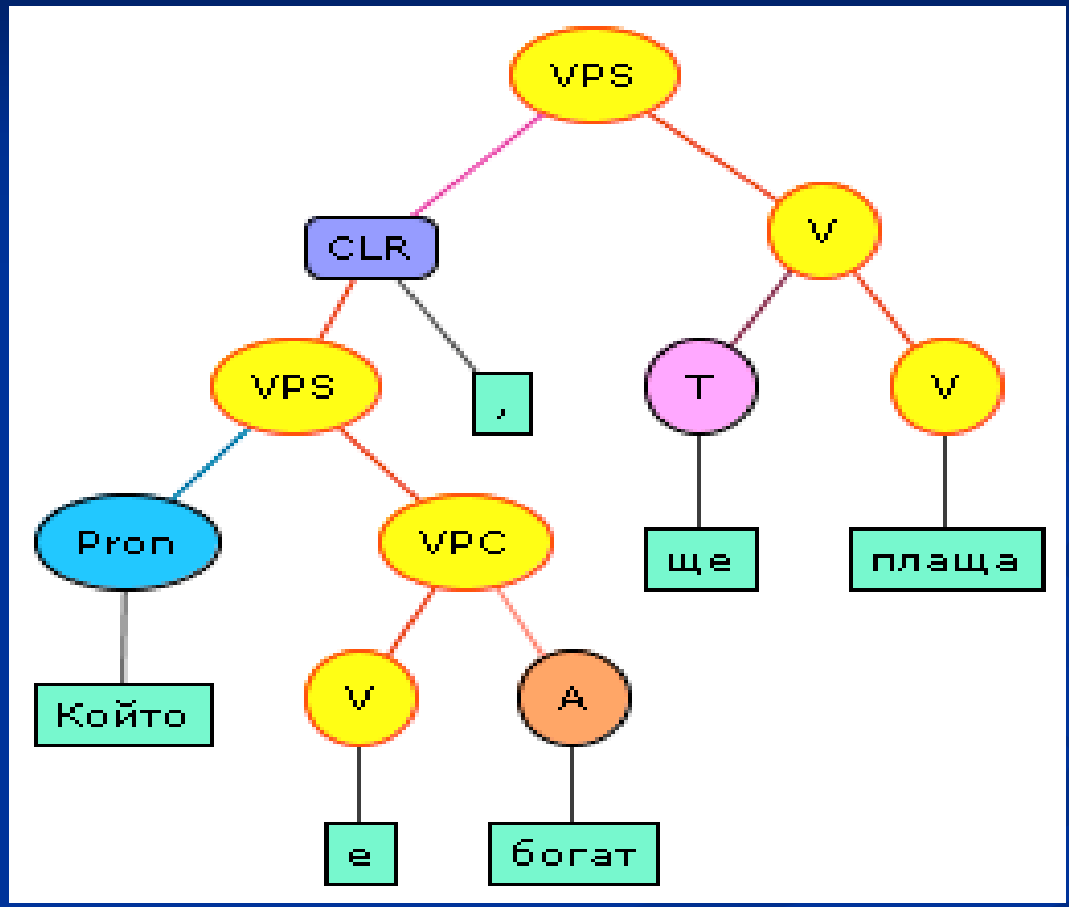
CLZADA


VPA
VPC
CLZADA
V
NPA
,
C
VPC
pro-ss
Започнах
Pron
N
за да
V
N
тоя
дневник
pro-ss
записвам
N
Pron*
преживяванията
си

CLQ като комплемент



CLR като подлог



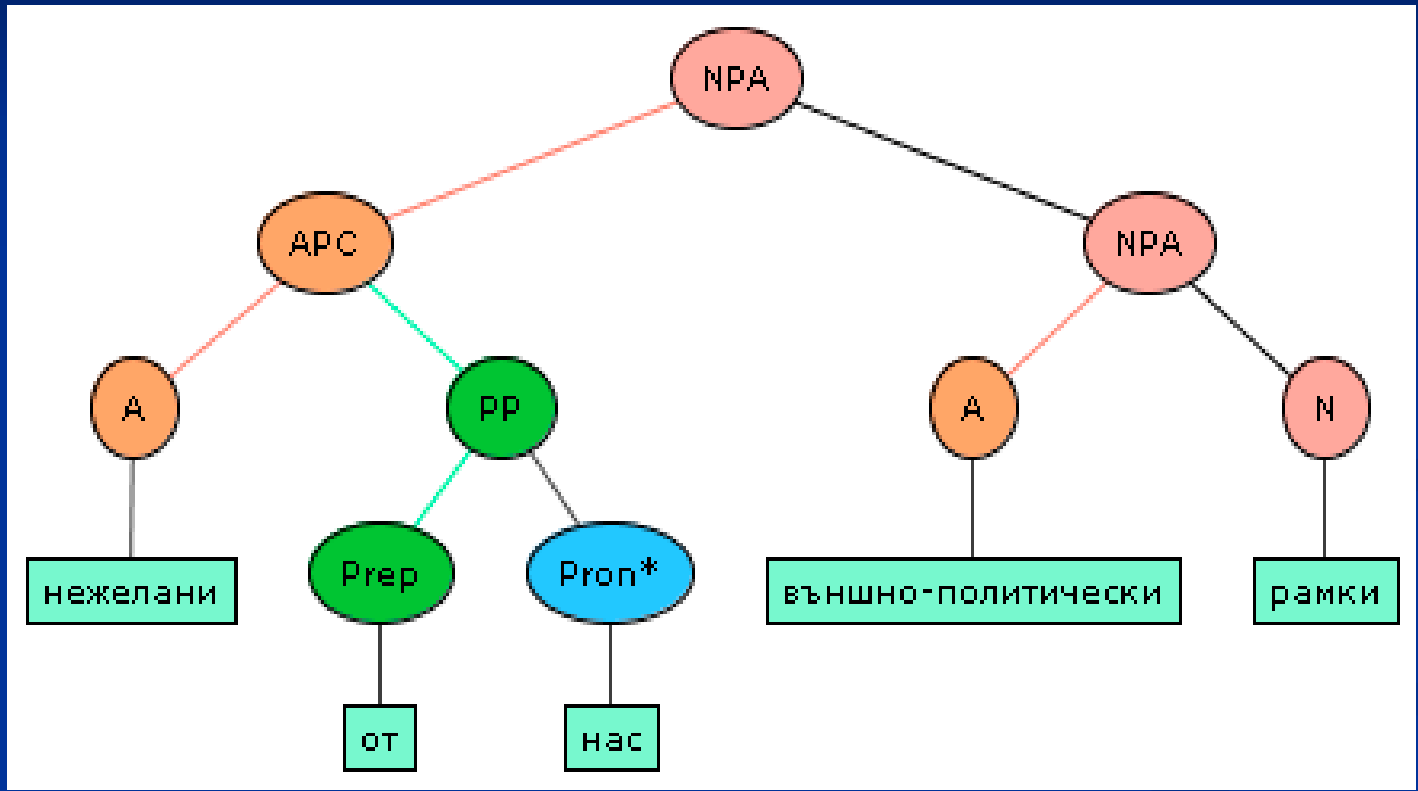
Именна фраза: NPC

- Опора– N, Subst, Nomin, N-Elip, ND-Elip, CoordP
- Елементи
  - Отношение *Количество-Субстанция* (литър мляко)
  - Отношение *Съдържащо-Съдържимо* (чаша мляко)
  - Отношение *Групиране- Същини* (група студенти)

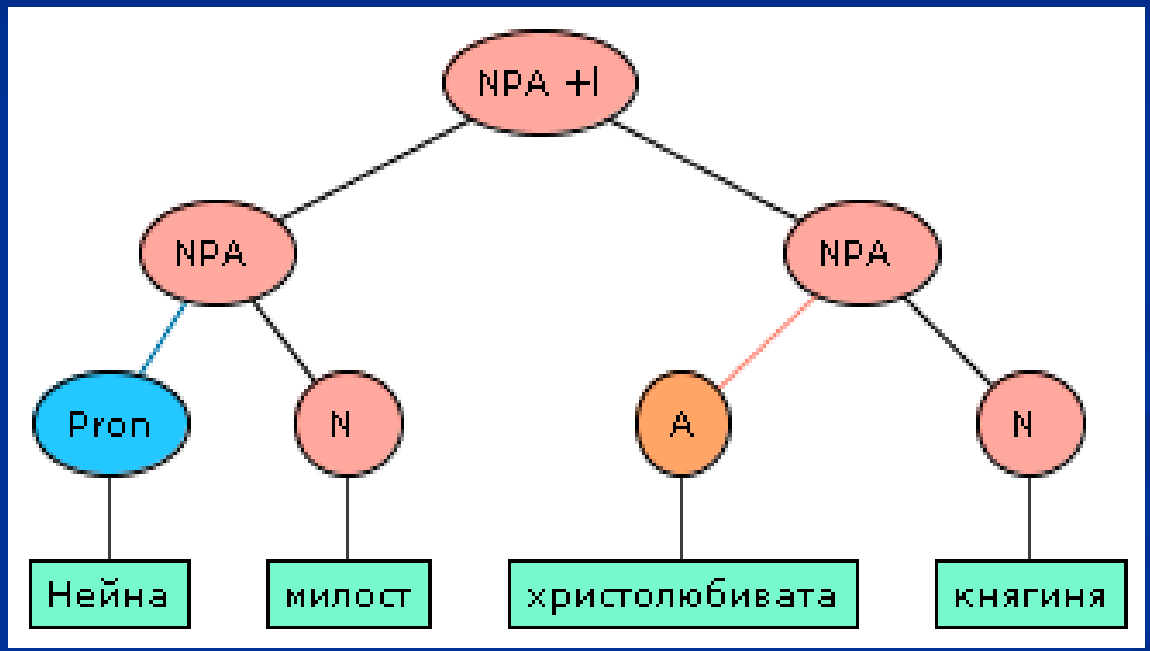
Именна фраза: NPA

- **Опора** – N, NPC, NPA, Subst, Pron, H, Nomin, N-Elip, ND-Elip, CoordP
- **Типове:**
  - Адективни модификатори
  - Адвербиални модификатори
  - Предложни модификатори
  - NP модификатори
  - Модификатори-клаузи
  - Реципрочни NP

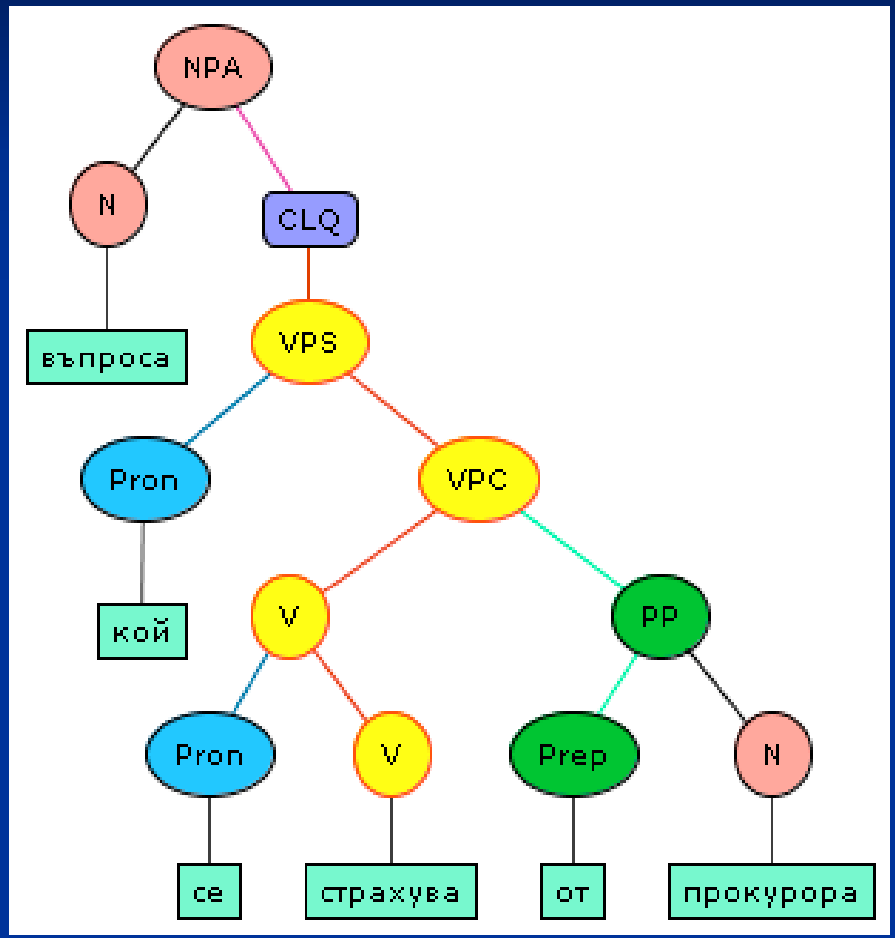
Пример (1)

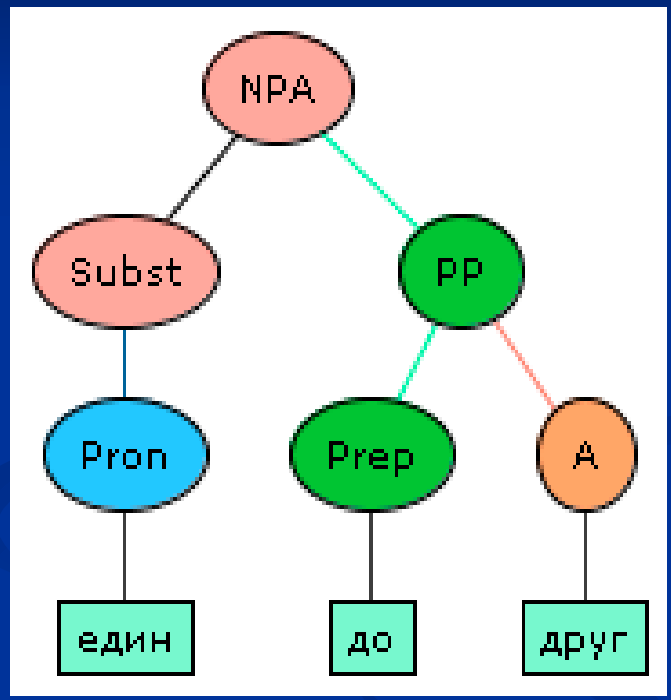

NPA
APC
NPA
A
PP
A
N
нежелани
Prep
Pron*
външно-политически
рамки
от
нас

Пример (2)



Пример (3)


NPA
N
CLQ
въпроса
VPS
Pron
VPC
кой
V
PP
Pron
V
Prep
N
се
страхува
от
прокурора


NPA
Subst
PP
Pron
Prep
A
един
до
друг

Адективна фраза: APC и APA

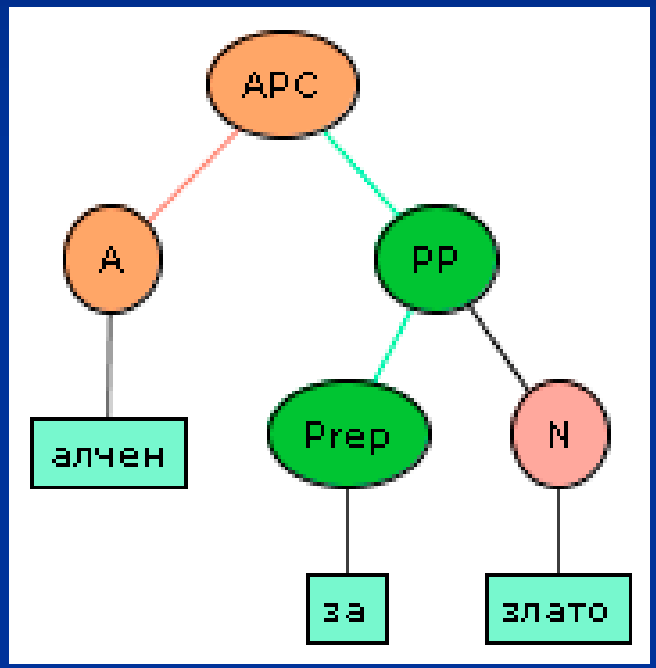
APC

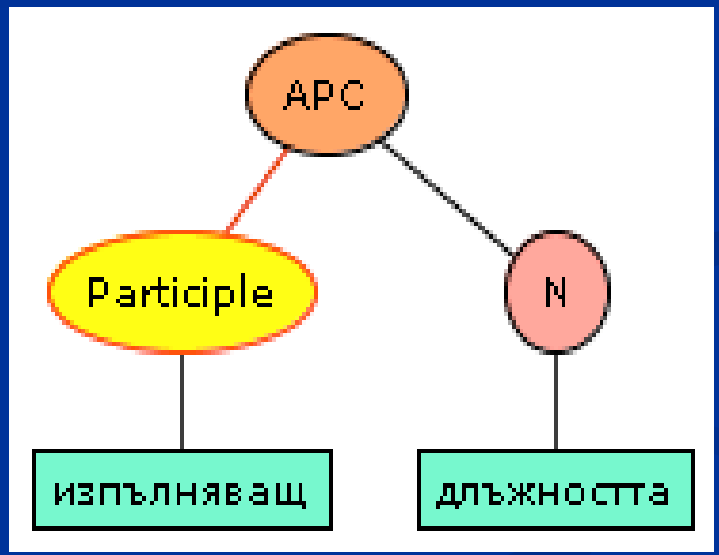
- Адективна опора
- Компаративни и суперлативни форми

APA

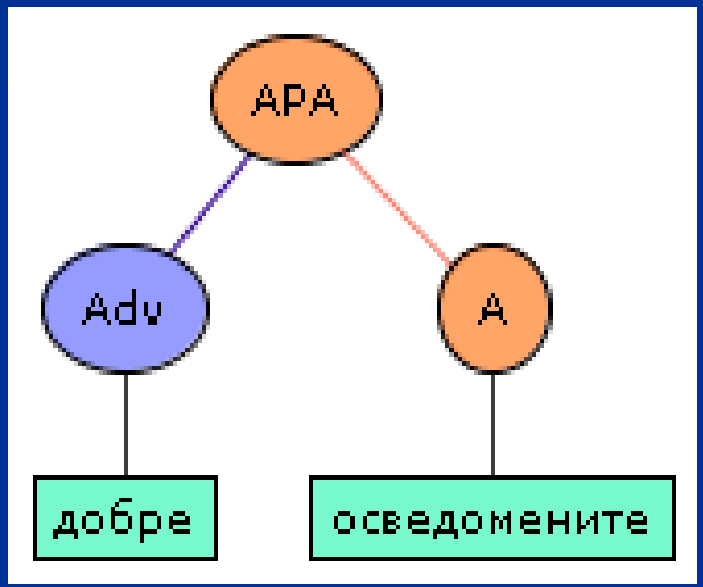
Подобно е за AdvP

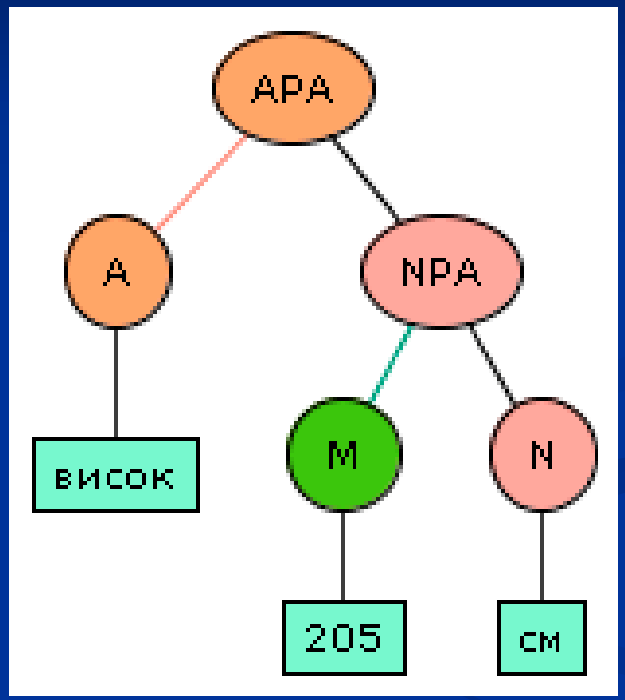
Примери


APC
A
PP
Prep
N
алчен
за
злато


APC
Participle
N
изпълняващ
длъжността

Примери

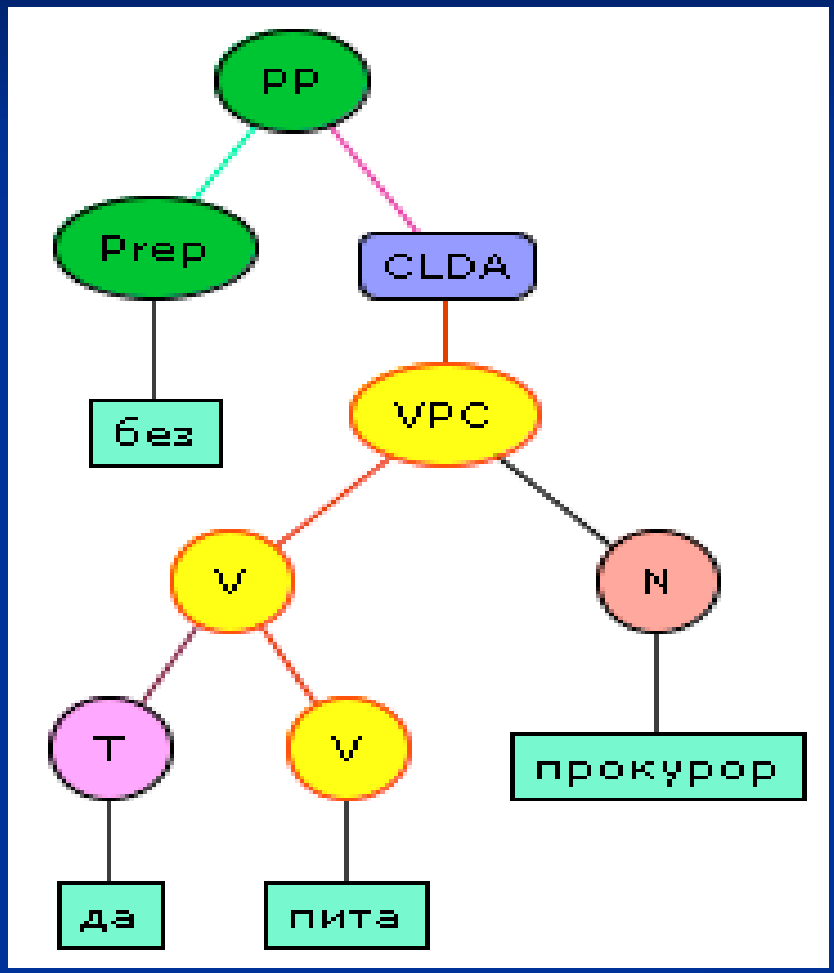


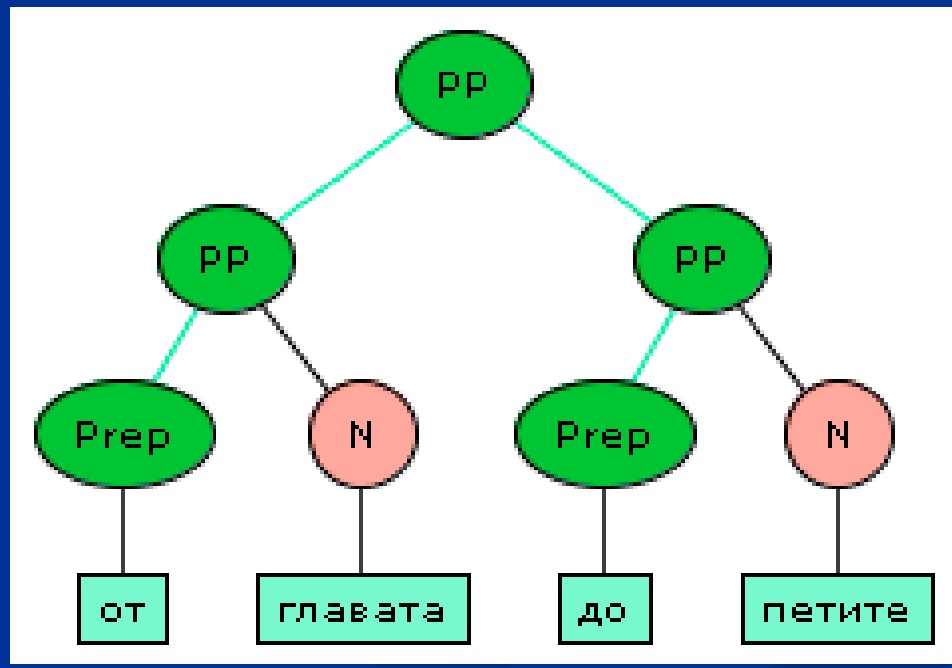


Предложни фрази

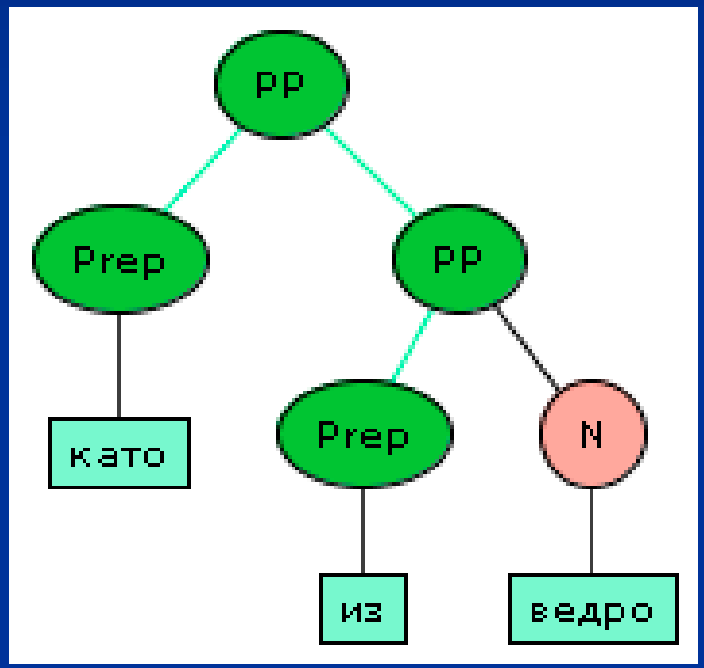
- Номинални комплементи
- Комплементи клаузи
- PP комплементи
- Един PP комплемент
- Адвербиални комплементи

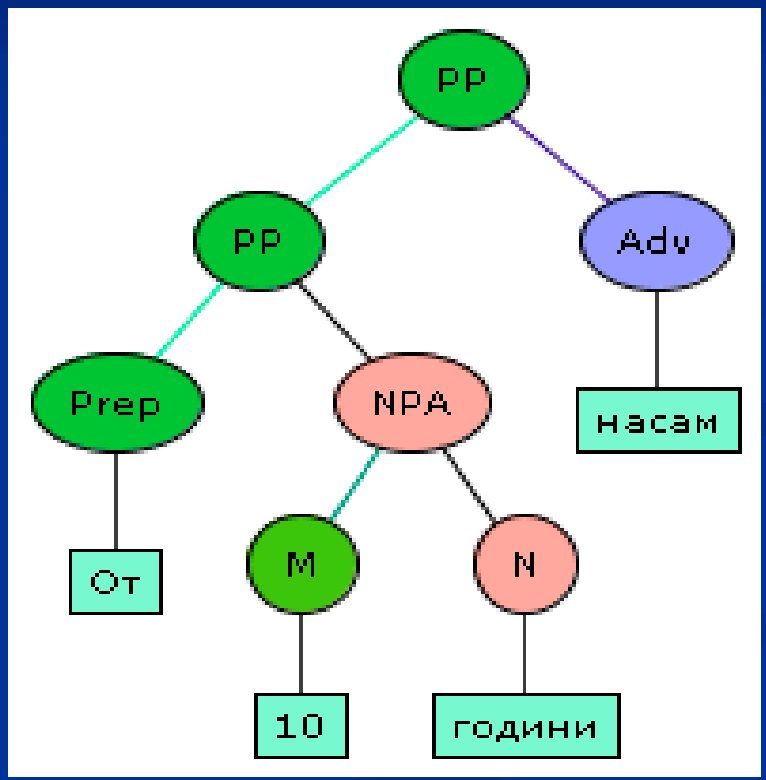
Примери (1)


PP
Prep
CLDA
без
VPC
V
N
T
V
прокурор
да
пита


PP
PP
PP
Prep
N
Prep
N
от
главата
до
петите

Примери (2)


PP
Prep
PP
като
Prep
N
из
ведро


PP
PP
Adv
Prep
NPA
насам
От
M
N
10
години

Специализиране на анотационната схема

- Нови сортове на фрази в сорт йерархията на ОФГ (HPSG)
- Прилагане на т. нар. **Преференциални правила**

Преференциални правила : **мотивация**

- Има смесени категории, които в различни контексти се държат различно
- Има явления, които са с повече от една възможност за лингвистично мотивиран анализ

Примери за такива правила

- **Координация**
  - Предпочитай сентенциалната координация пред пре-координацията
  - Предпочитай конституентната координация пред елипсата
- **Елипса**

  Предпочитай да свързваш елипсата в изречението, ако е възможно, а не в дискурса
- **Модални глаголи**

  Ако има две възможни четения: лично и безлично, предпочитай личното

Кодиране на специални явления

- Координация
- Елипса
- Прагматични изрази
- Фокусиращи елементи
- Кореференция

Координация: Нашата хипотеза (1)

- Граматичната роля на конюнктите е решаваща
- Синтактичните категории имат второстепенна роля
- Основният фактор е валенцията на конюнктите

Нашата хипотеза (2)

Координацията се смята за безопорна фраза, където:

- Конюнктите трябва да се съгласуват по валентността си: *Valency lists* и *Mod feature*
- Те могат да се генерализират над конкретната категория: *coordination*

Типове координация

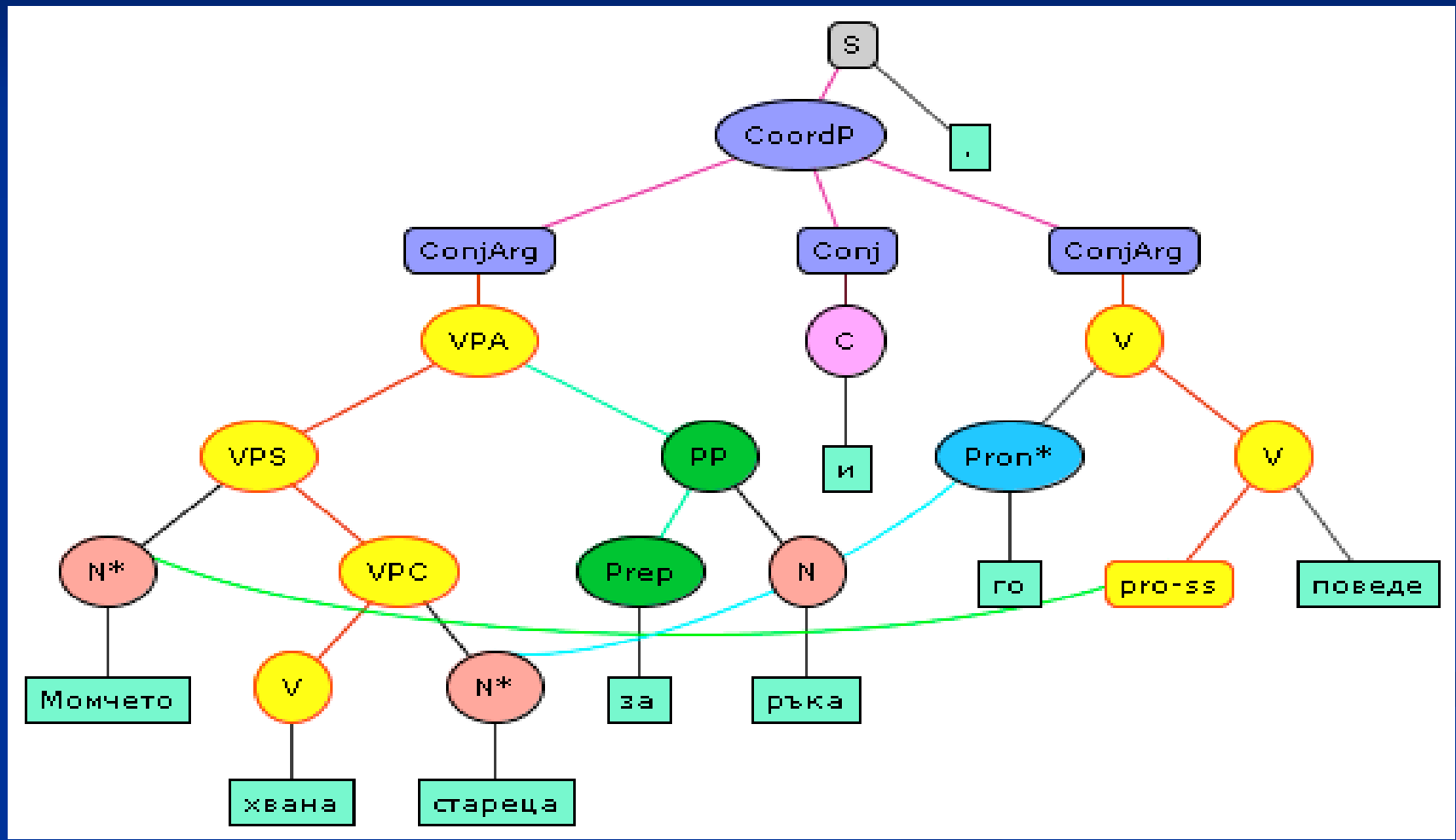
- Наситена координация

  Клаузална (изреченска), NP, AP, AdvP, PP
- Адюнктна координация

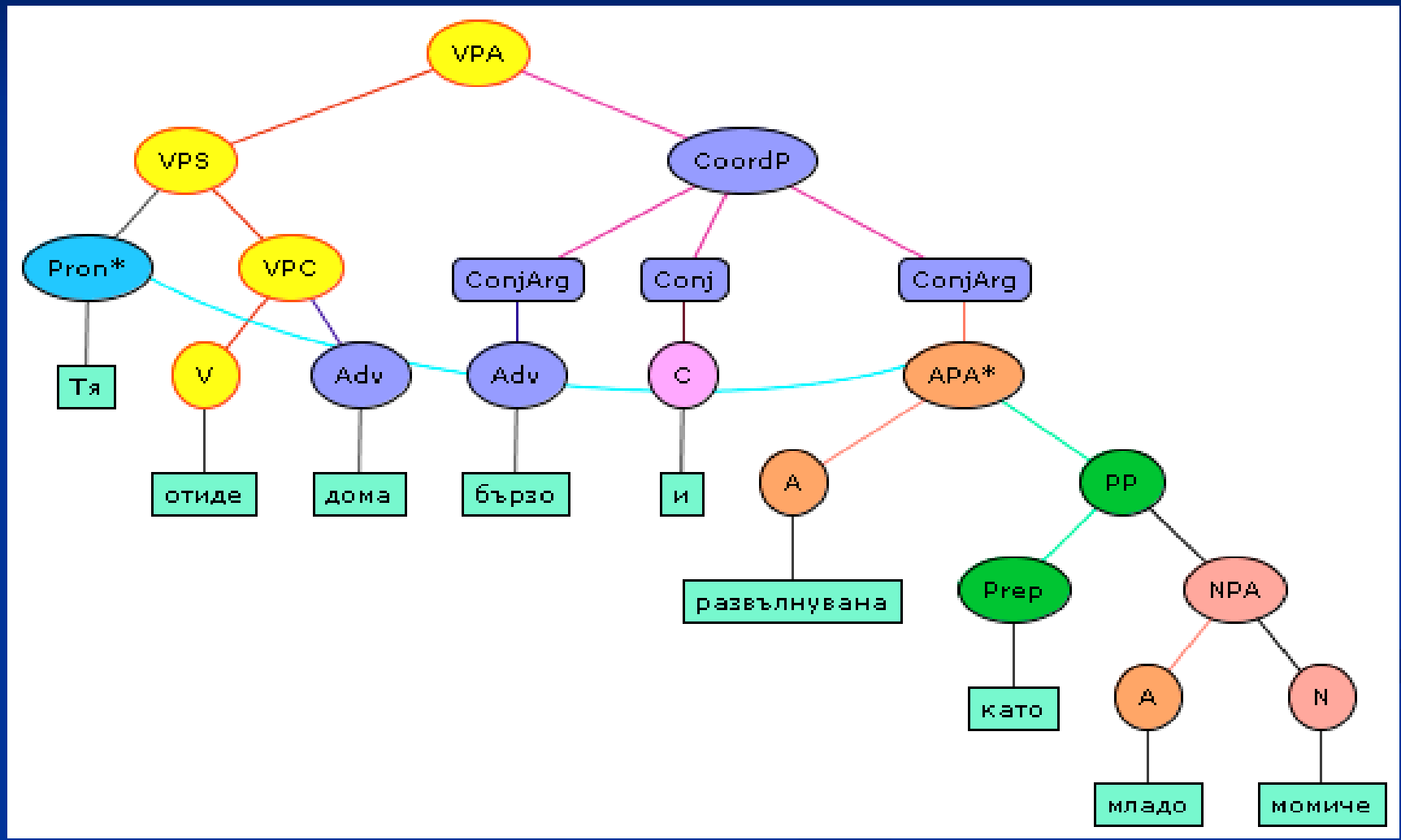
  Всички горни плюс *unlike categories*
- Ненаситена координация

  Лексикална координация

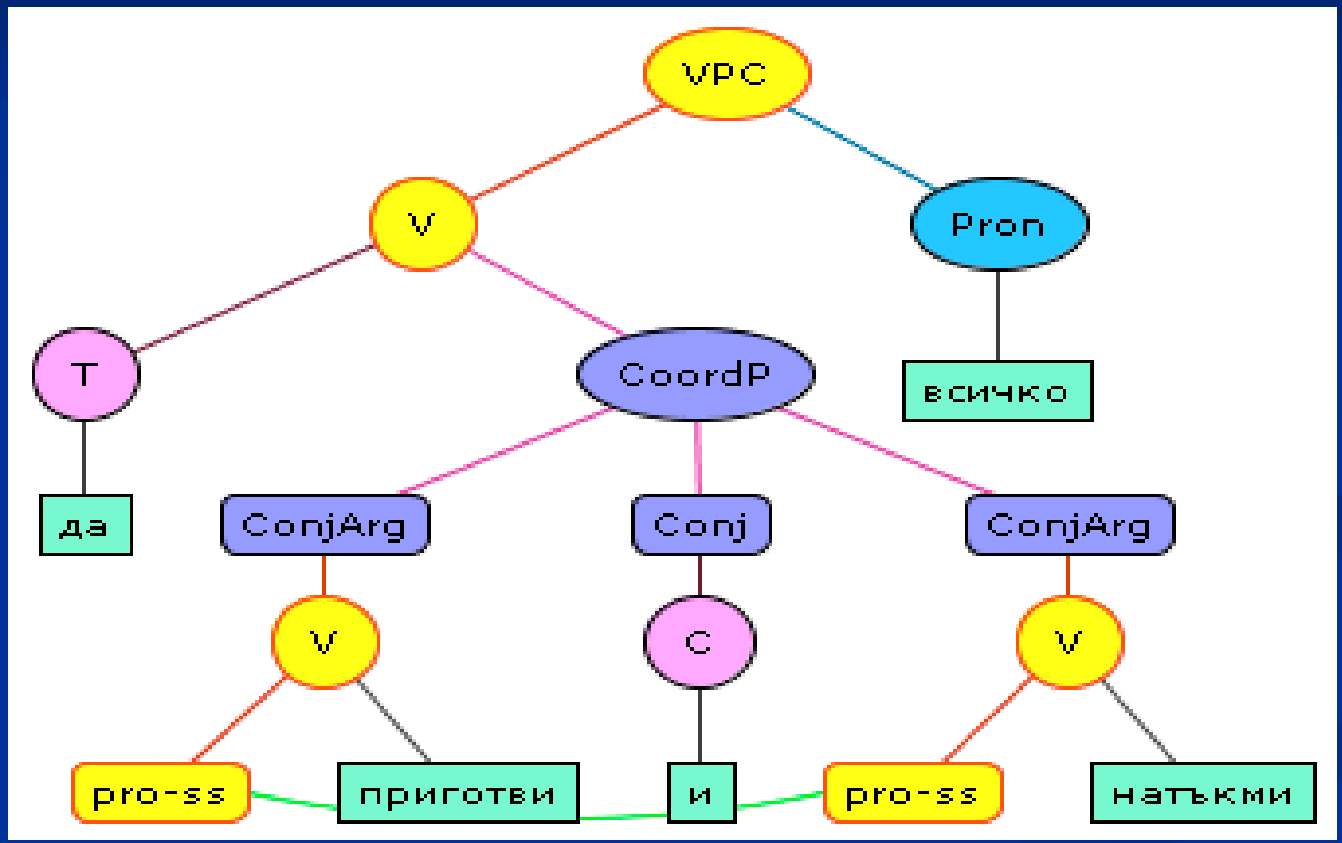
Изреченска координация



Адюнктна координация



Лексикална координация



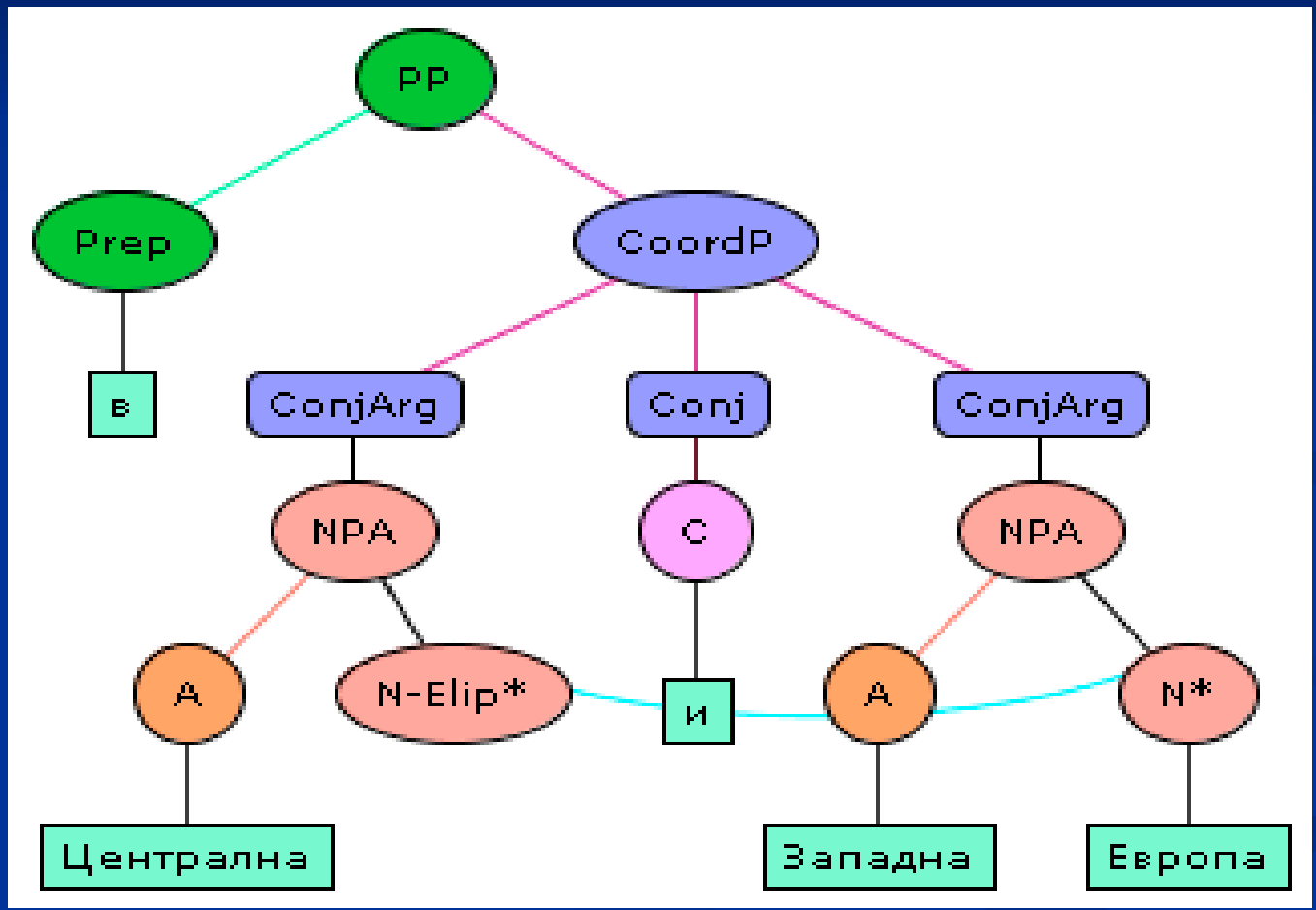
Елипса (1)

- Която се възстановява в изречението
  - V-Elip (за глагол или глаголна фраза)
  - N-Elip (за име или именна фраза)
  - Prep-Elip (за предлог)
  - PP-Elip (за предложна фраза)

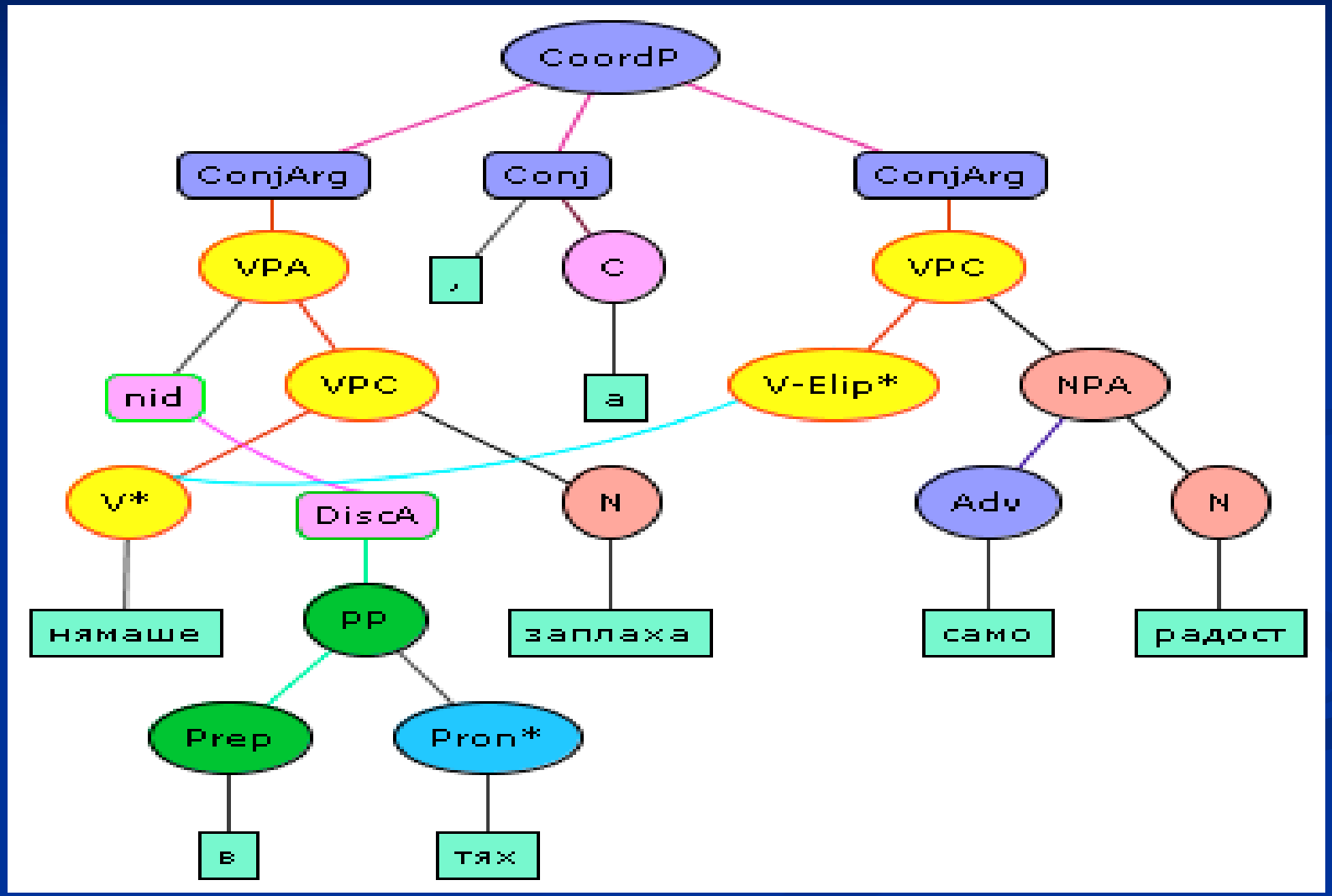
Атрибути: type: equal, variant, negation

grammar: конкретната форма

Пример с елипса на име



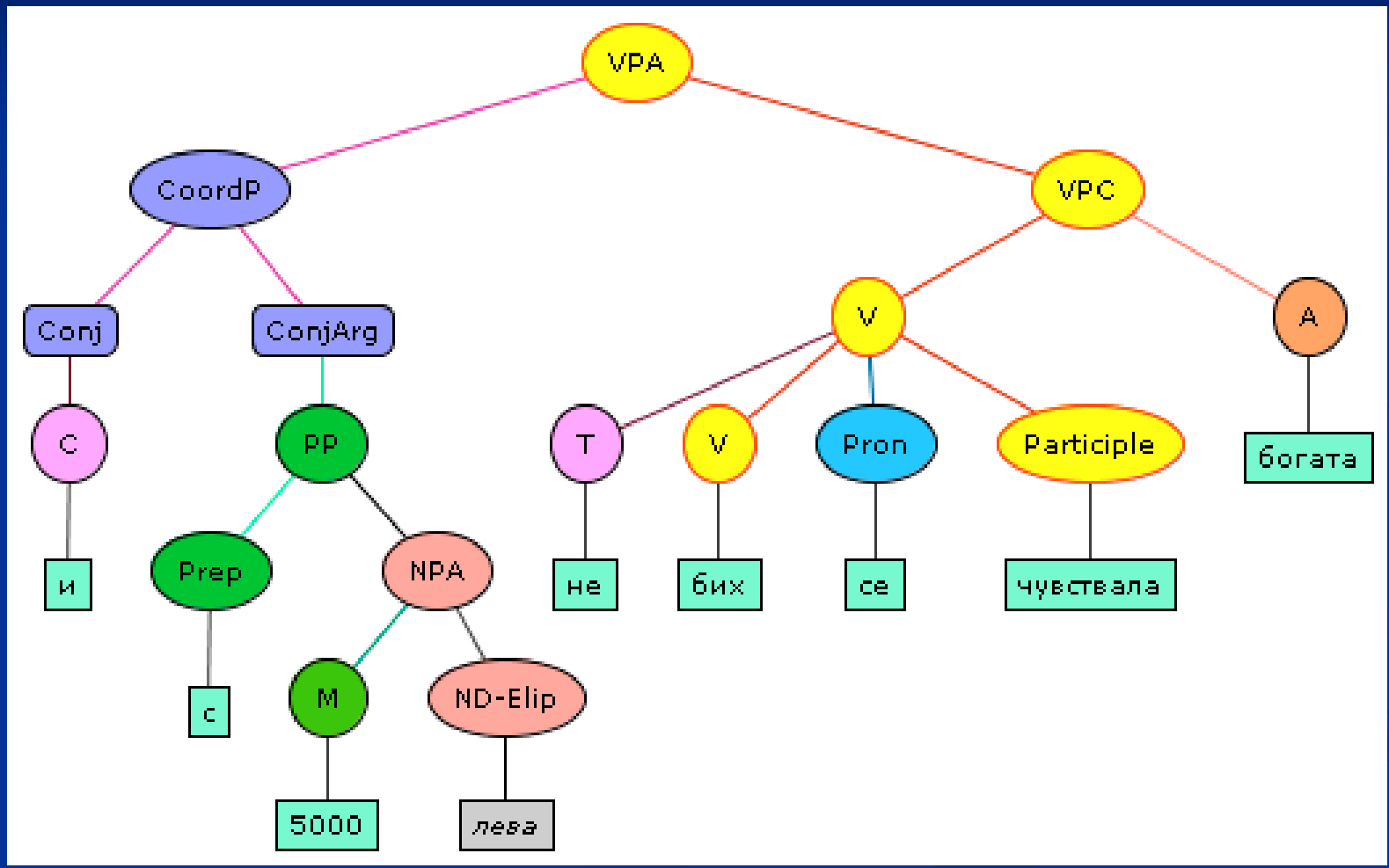
Пример с елипса на глагол



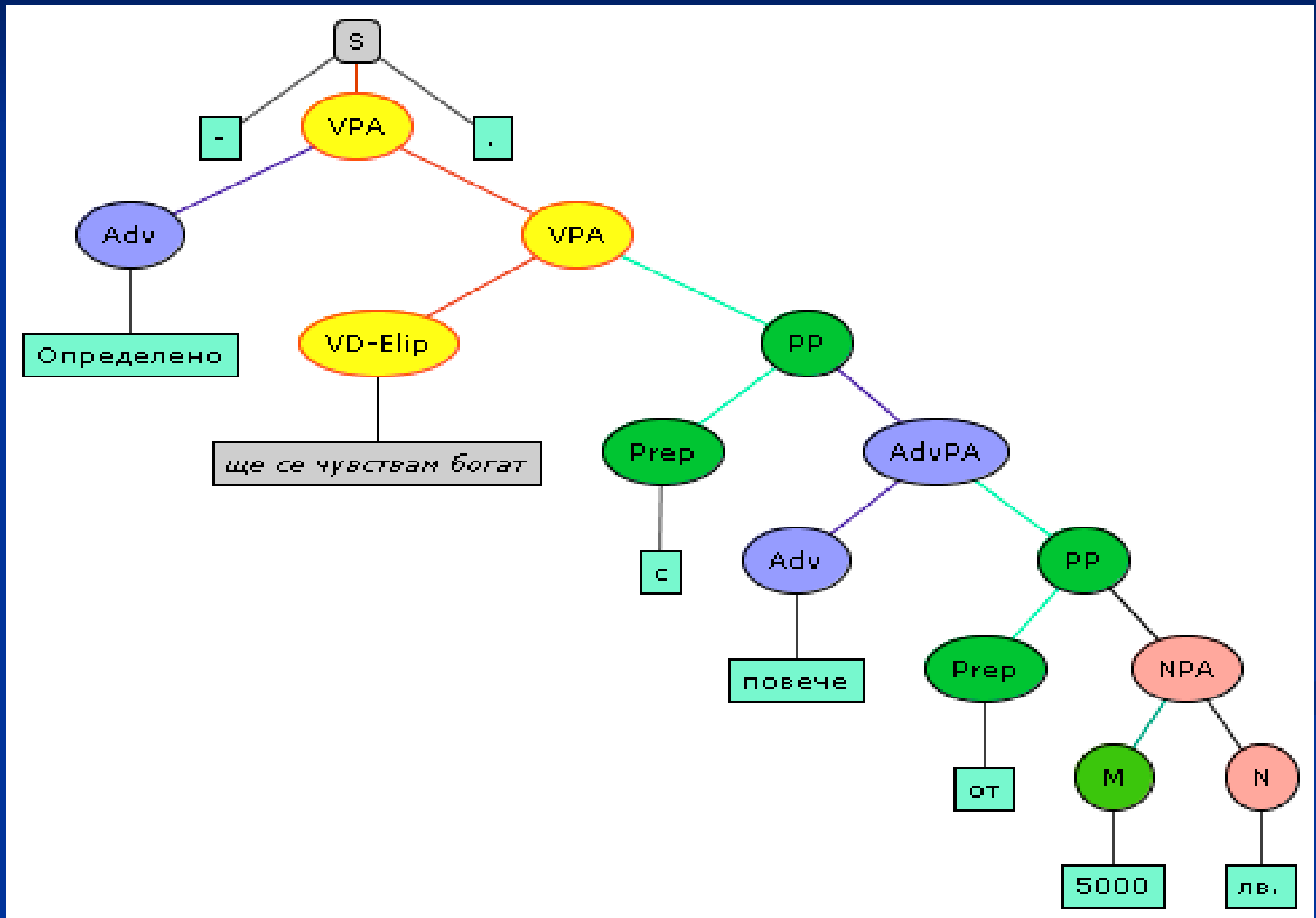
Елипса (2)

- Която НЕ се възстановява в изречението
  - VD-Elip (за глагол или глаголна фраза)
  - ND-Elip (за име или именна фраза)
  - PPD-Elip (за предложна фраза)
- Атрибути:
  - type: world knowledge, discourse, exists (само за VD-Elip)
  - grammar: за морфосинтактичните характеристики
  - Form: за конкретната форма

Пример с елипса на име



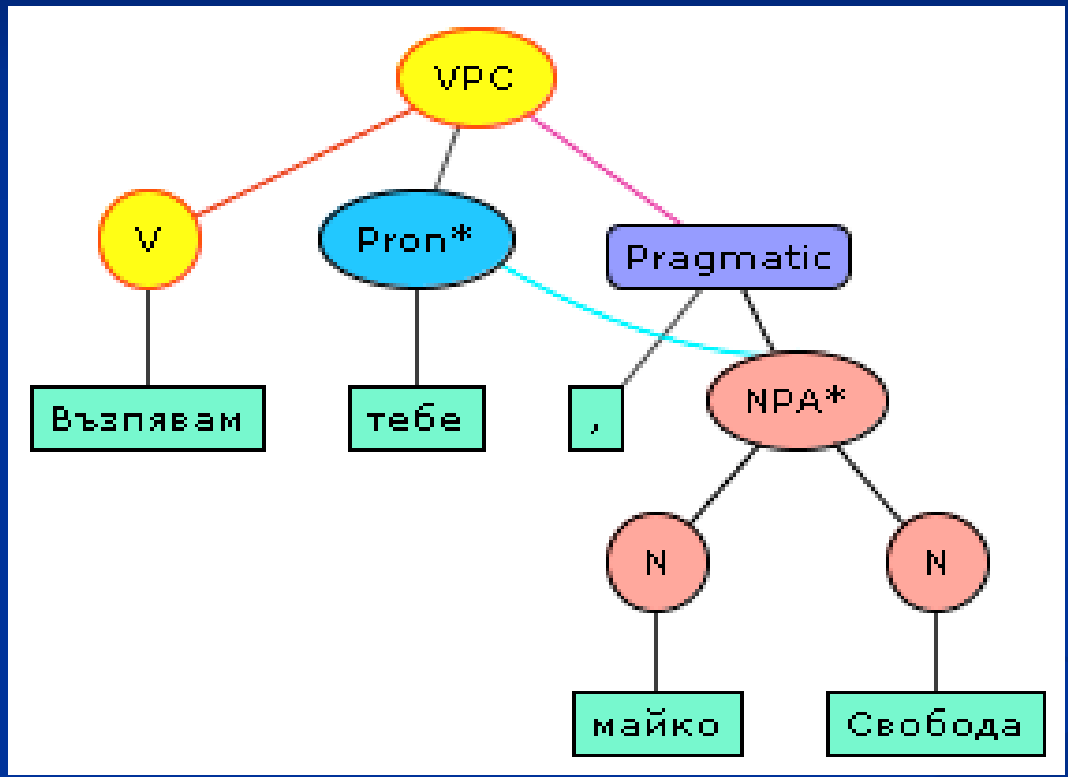
Пример с елипса на глаголна фраза



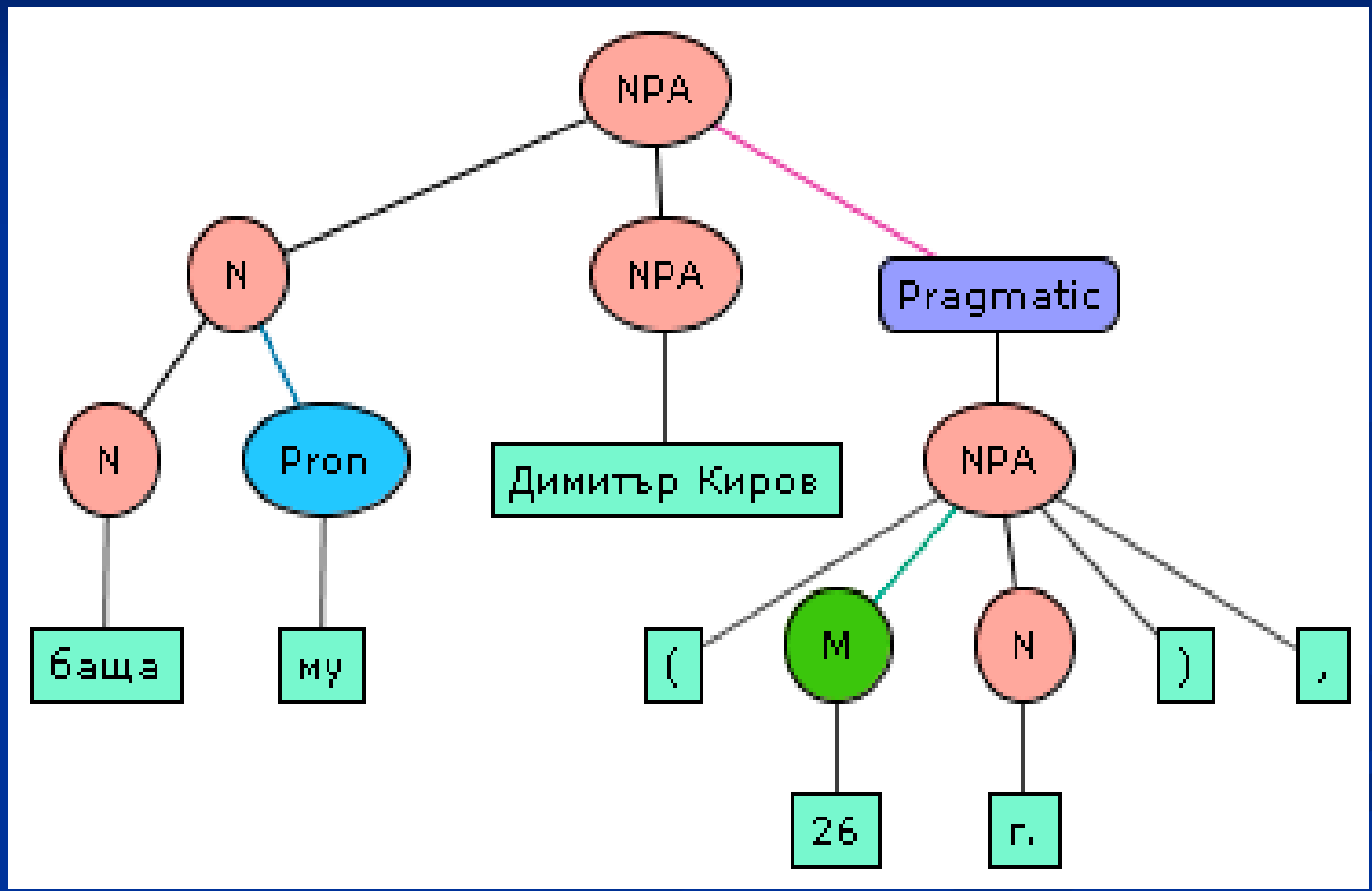
Прагматични елементи

- Звателни форми
- Оценъчни адвербиали
- Парентези
- Фокусиращи елементи

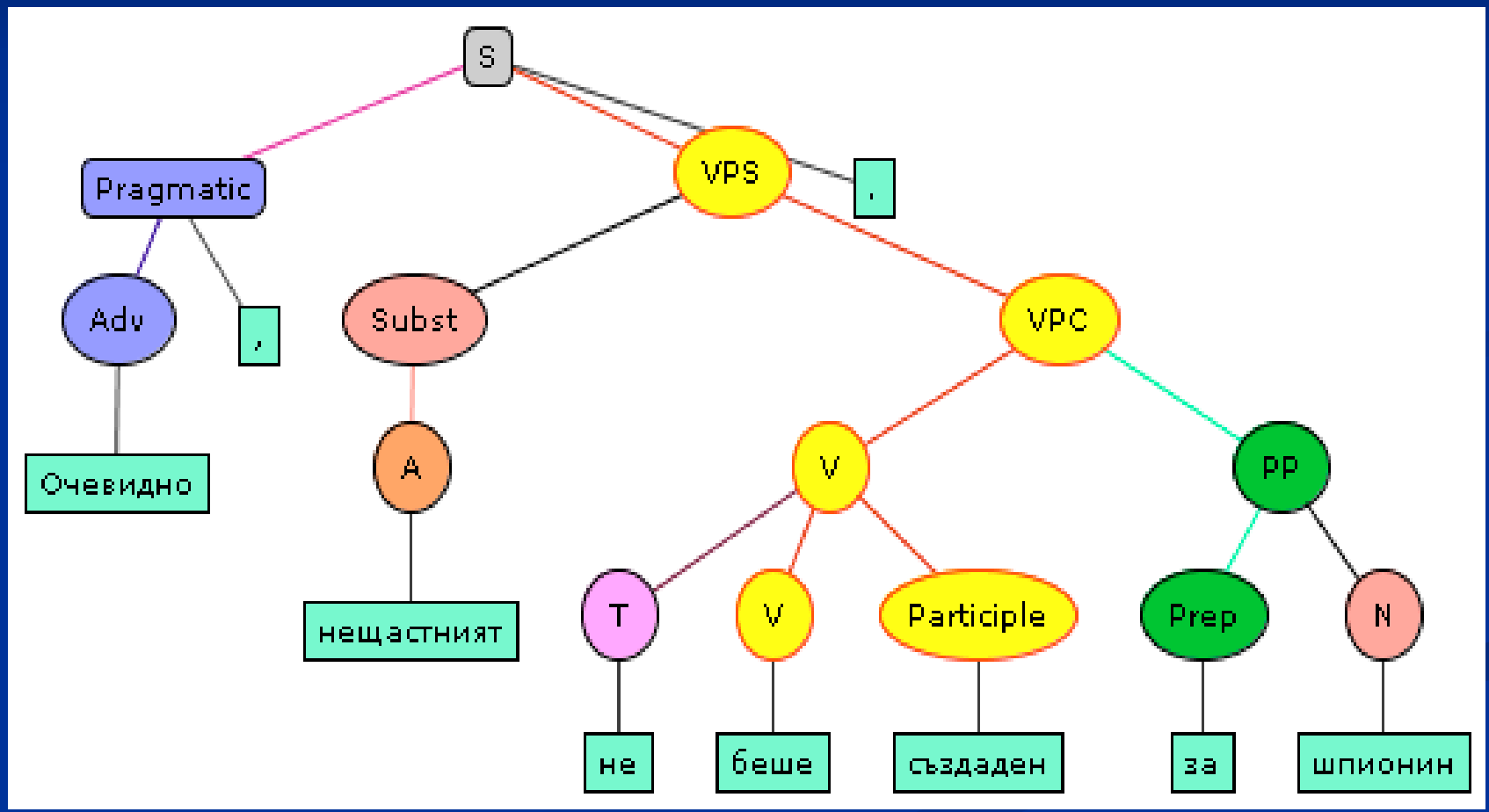
Звателни форми



Парентези

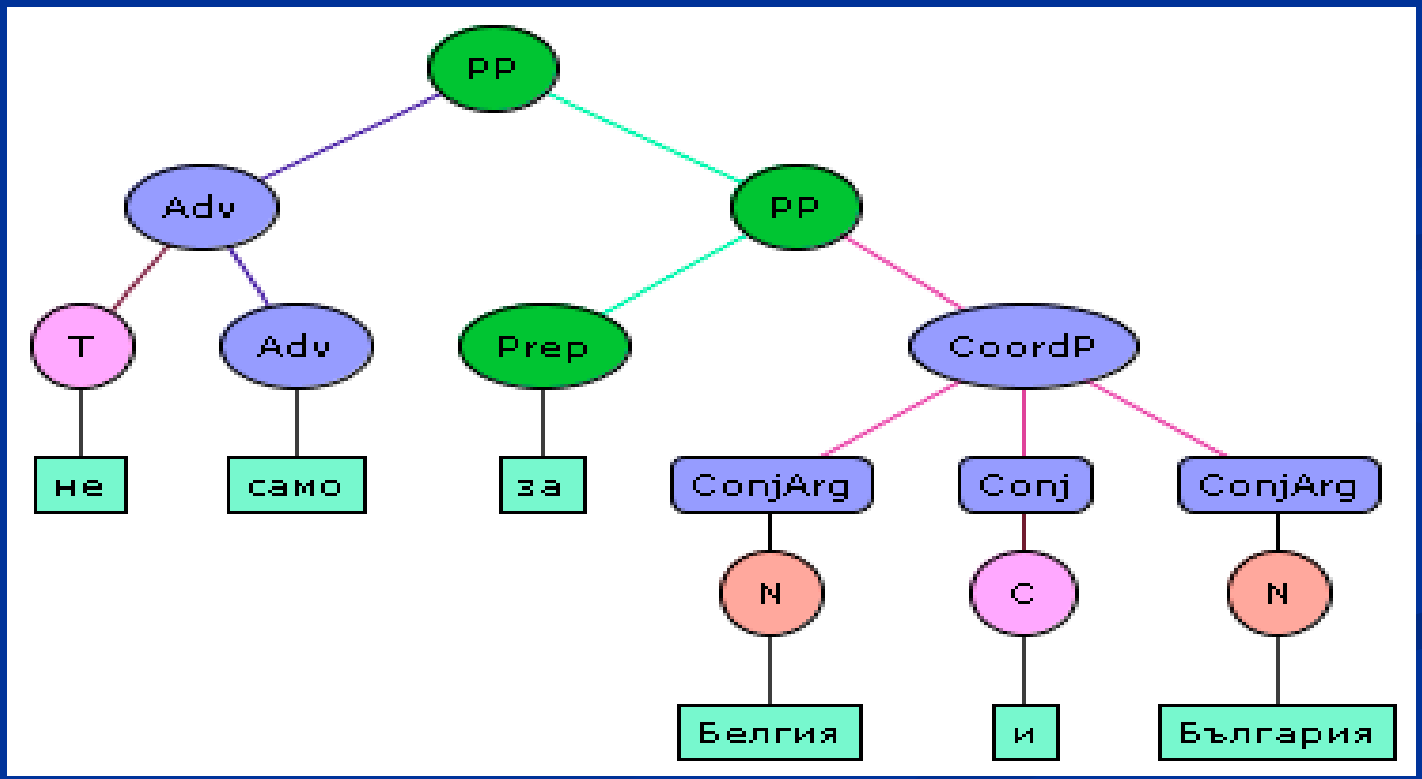


Оценъчни адвербиали

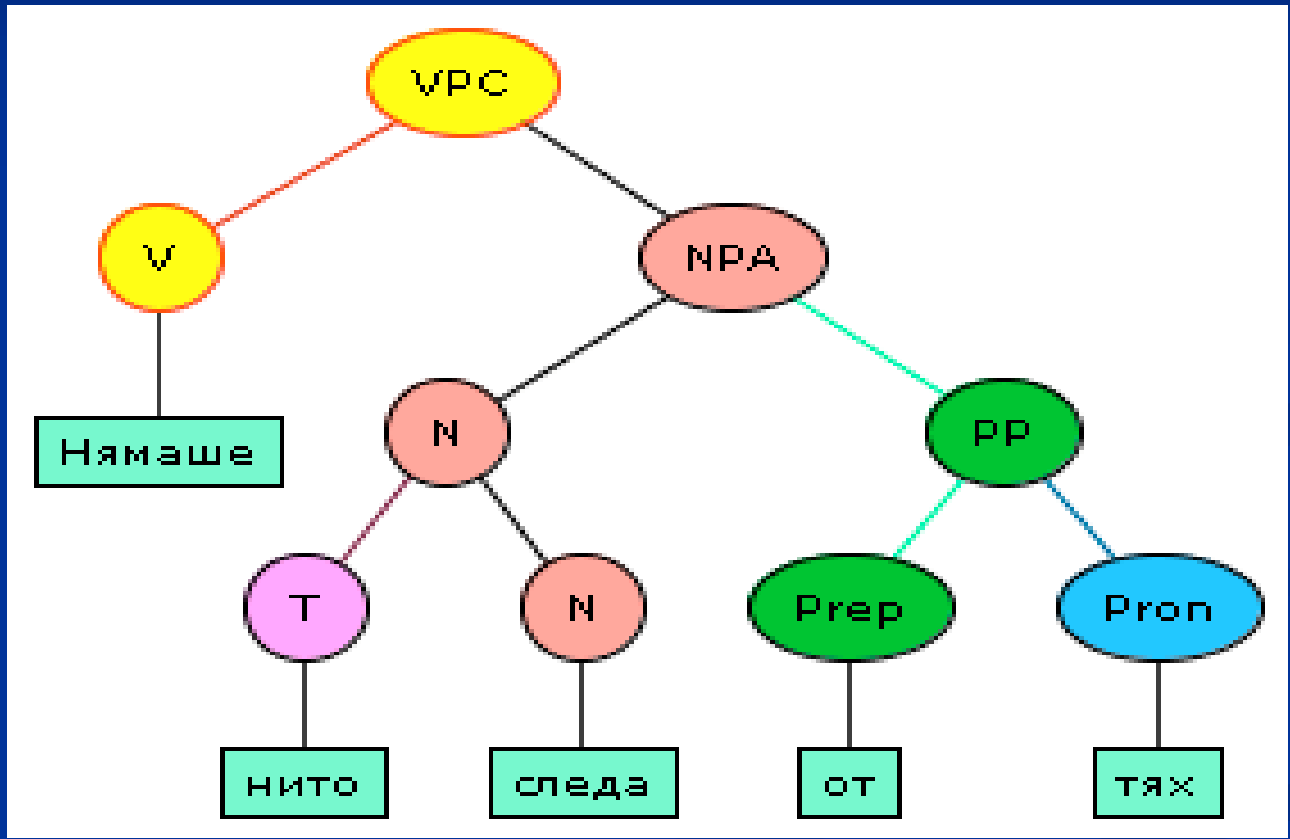


Фокусиращи елементи

- Не променят лексикалната или синтактичната категория на езиковия елемент



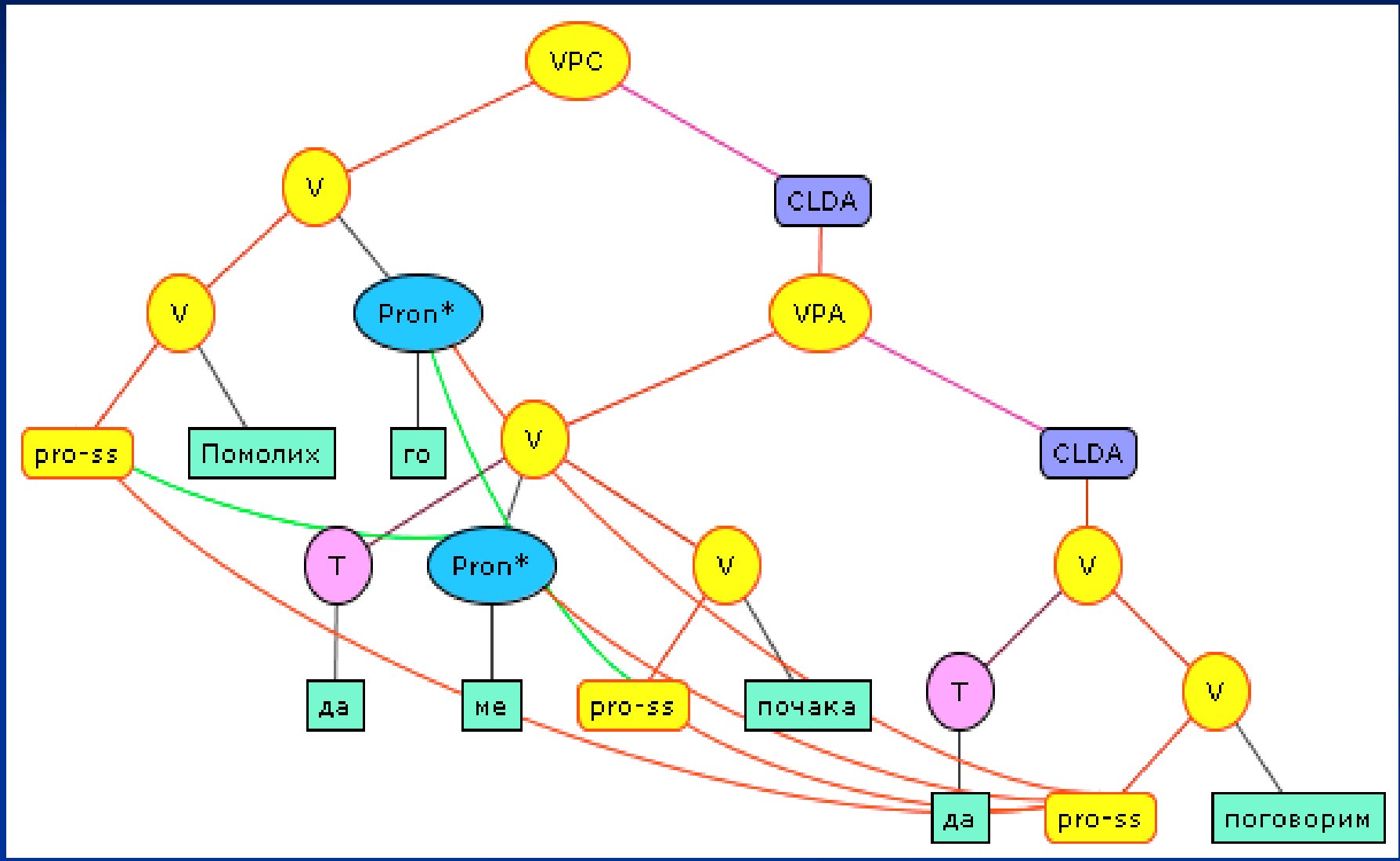
Пример



Кореферииране

- Видове:
  - Равенство на обекти (множества)
  - Елемент на множество
  - Подмножество на множество
- Какви явления?
  - Нулева субектност
  - Притежателност
  - Малки изречения
  - Номинализация
  - Кохезийни вериги (повторения, синоними)

Пример


VPC
V
CLDA
V
Pron*
VPA
pro-ss
Помолих
го
V
CLDA
T
Pron*
V
V
да
ме
pro-ss
почака
T
V
да
pro-ss
поговорим

ДЕМО с CLaRK

- Как да направим лингвистично търсене?