

ЛЕКСИКОН-ГРАМАТИКА И СЕМАНТИЧКИ ВЕБ

Анибале Елиа^{*}, Симонета Виетри^{}**

Универзитет у Салерну,

Одсек за комуникационе науке

Превод с енглеског: Бојана Ђорђевић

Апстракт: Термин “семантички веб” означава скуп програма који користе семантичке етикете за проналажење информација унутар текстова расутих по вебу. Уметање семантичких етикета омогућавају онтологије, као огромне колекције дефиниција о садржају, организованих у класификацијске схеме. Имплементација онтологија омогућава интеракцију различитих домена, тако да је могуће вршити разноврсне претраге које повезују различите типове текстова. Семантика укључена данас у семантички веб односи се, пре свега, на семантичке дефиниције и сродне онтологије које се тичу властитих и заједничких имена. Друго, семантика је замишљена као формални логички апарат за хијерархијску организацију концептуалног система класе, поткласе и својстава. Овај рад даје допринос расправи о семантичком вебу с нове тачке гледишта која се заснива на могућности развоја семантичких предиката, полазећи од конкретних лексичких облика и синтаксичких структура у којима се уобичајено јављају унутар информативних и комуникативних контекста. Овакви предикати се унутар текстова аутоматски генеришу помоћу лексичко-граматичког “интелигентног агента”. Текстови етикетиран на овај начин могу се претраживати по садржају помоћу упита на природном језику.

^a Анибале Елиа (Annibale Elia) је аутор делова 1, 2 и увода, Симонета Виетри (Simonetta Vietri) је аутор дела 3 и закључка.

* elianni@tin.it

** simona.vietri@tiscali.it

Увод: семантика семантичког веба

Термин “семантички веб” означава скуп програма који користе семантичке етикете за проналажење информација унутар текстова расутих по вебу. Тим Бернерс-Ли, оснивач веба, покренуо је ову информатичку и филозофску провокацију пре неколико година.¹ Традиционални веб се још увек заснива на употреби HTML-а (*Hyper Text Mark-up Language*), графичко-типографског језика за обележавање; семантички веб се заснива на XML-у (*eXtensible Mark-up Language*)², новом језику за обележавање који подржава семантичке упите.

Према Тим Бернерс-Лију и W3C-у (интернационални конзорцијум истраживача чија је мисија развој веб стандарда), семантички веб (СВ) не би заменио традиционални веб већ би га укључио у себе помоћу додатног скупа упита и новог, “интелигентног” модела структуре, који би олакшао приступ информацијама на вебу захваљујући софтверским агентима способним да решавају специфичне проблеме са свакодневним информацијама.

Чак ни најнапредније машине за претраживање не омогућавају разликовање господина Ђоакина Росинија, аутомеханичара из Форлија и славног италијанског композитора опера прве половине XIX века. Према заступницима СВ-а, када бисмо Ђоакина Росинија могли да етикетирамо са <КОМПОЗИТОР> при претраживању веба, сва документа која не садрже ову етикету била би аутоматски елиминисана.

Како би се дошло до ових резултата, сва документа о композитору Ђоакину Росинију требало је да људи који су их поставили на веб *ручно* етикетирају. Штавише, требало би употребити и етикете везане за музички садржај опера, за друге композиторе или било који тип информација из овог одређеног домена.

¹ Видети Berners-Lee et Alii 2001.

² Видети Fuiano 2004 (додатак) за кратко поређење HTML-а и XML-а.

Према Бернерс-Лију, милиони и милиони дефиниција о садржају, организованих у класификацијске схеме, требало би да се сакупе након неколико година. Ове схеме се називају “онтологијама”, у част много сложенијој филозофској онтологији, чији је циљ да дефинише суштину “ентитета”.

У случају Росинија, требало би развити подскуп “музичке онтологије”, при чему би овај подскуп био повезан са другим онтологијама преко термина који би се могли појавити у тексту. Узорак текста какав је овај:

Quando Gioacchino Rossini scrisse il Barbiere di Siviglia, non immaginava che la prima di quell'opera a Napoli sarebbe stato un vero fiasco.

Када је Ђоакино Росини написао Севилског берберина, није ни помишљао да ће премијера тог дела у Напуљу доживети прави фијаско.

у XML-у би могао бити обележен на следећи начин:

Quando <compositore>Gioacchino Rossini</compositore> scrisse <opera>il Barbiere di Siviglia</opera>, non immaginava che la prima di quell'opera a <capoluogo_regione_campania>Napoli</capoluogo_regione_campania> sarebbe stato un vero fiasco.

Машина за претраживање семантичког веба одабрала би само текстове који садрже ниску *Gioacchino Rossini* етикетирану са <КОМПОЗИТОР>.

Имплементација «онтологија» омогућила би интеракцију различитих домена, тако да би било могуће вршити додатне претраге које би повезале различите типове текстова и упућивале на URL-ове који садрже све композиторе, опере, али и позоришта у којима су опере биле извођене, градове у којима се та позоришта налазе итд. На тај начин би било могуће добити одговоре на упите типа «У ком граду је извођен ‘Севилски берберин’?» «У којој продавници у Напуљу могу да купим CD ‘Севилски берберин’?» итд.

Настајање семантичког веба подразумева јак дух сарадње међу добављачима информација и онима који претражују Интернет, усвајање заједничког класификационог система и глобални систем ажурирања информација.

Тренутни експерименти су још увек везани за веома уске домене, а спроводе их заједнице Интернет корисника посебно осетљивих на дух сарадње³, али СВ ипак напредује.

Са лингвистичке тачке гледишта, желимо да направимо нека запажања о типу семантике укључене у приступ СВ-а. Пре свега, семантичке дефиниције и сродне онтологије најпре се тичу властитих и заједничких имена. Друго, семантика је замишљена као формални логички апарат за хијерархијску организацију концептуалног система класе, поткласе и својстава.

Ова имена су речи које се схватају као аргументи чији су предикати ван света речи, будући да их представљају логички оператори који треба да повежу, изведу и успоставе однос између аргумената.

Овај текст је допринос расправи о СВ-у с нове тачке гледишта која се заснива на могућности развоја семантичких предиката, полазећи од конкретних лексичких облика и синтаксичких структура у којима се уобичајено јављају унутар информативних и комуникативних контекста.⁴

Овакви предикати би унутар текстова били аутоматски генерисани помоћу лексичко-граматичког «интелигентног агента». Текстови етикетиран на овај начин могли би бити претраживани по садржају помоћу упита на природном језику.⁵

³ Екипа из KSL-а са Стенфордског универзитета направила је апликацију Wine Agent која може да упари вино са изабраним јелом. За више информација погледати: <http://www.ksl.stanford.edu/people/dlm/webont/wineAgent/>.

⁴ Видети М. Gross, 1978, 1981, који је први говорио о семантичким предикатима у оквиру лексикон-граматике.

⁵ Видети Fuiano 2004.

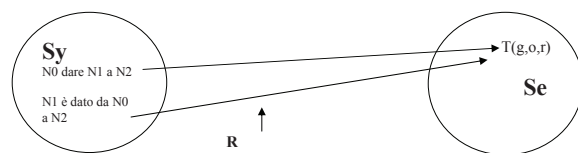
Ослањамо се на теорију лексикон-граматике, која даје исцрпан опис лексичко-синтаксичких структура језика.⁶

1. Од интуитивног појма “размене” до предиката трансфера

Скуп лексичко-синтаксичких језичких структура доступан је у оквиру теорије лексикон-граматике, што значи да је могуће доделити статус семантичког предиката глаголу, а статус аргумента јединицама које бира тај предикат (укључујући и субјекат).⁷

У овом тексту ћемо се усредсредити на оне предикате који се односе на интуитивни појам “размене” између особа, који се могу дефинисати као “предикати трансфера” (Т) који узимају три аргумента: “даваоца” (g), “објекат” (o) и “примаоца” (r). Ова релација се може формализовати као T(g,o,r).

Ако је Sy скуп лексичко-синтаксичких језичких облика, а Se скуп значењских јединица, онда се елементи Sy могу повезати са елементима Se помоћу интерпретативних правила (R).



У горе датој схеми, проста активна реченица с глаголом *dare* ‘дати’ (N0 V N1 a N2) и одговарајућа пасивна реченица N1 essere Vpp da N0 a N2 (N1 је Vpp од N0 N2 по моделу *Поклон је дат Марији од (стране) Милана*) повезане су јединственим семантичким предикатом T(g,o,r) према интерпретативном

⁶ Иновирани приказ италијанске лексикон-граматике представља Vietri (2004). За општу слику о примени лексикон-граматике у свету погледати сајт Универзитета Marn-la-Vallée: <http://infolingua.univ-mlv.fr/>.

⁷ Термин “актант”, који је увео Лисјен Тезнијер (Lucien Tesnière) 1953, доста се употребљава у лексикон-граматици.

правилу које повезује N0 са “g”, N1 са “o”, а N2 са “r”.⁸

Ови предикати представљају посебан случај ширег интуитивног појма који означава трансфер неког објекта (живог или неживог) са једног места (L1) на друго (L2). Многи глаголи који изражавају трансфер с места на место могу да изразе и трансфер између две особе, као у реченицама:

(1) *Max porta un topolino dal negozio (a casa di Ale + da Ale)*

Max доноси мишића из радње (у Алеову кућу + Алеу)

(2) *Max porta un topolino ad Ale*

Max доноси мишића Алеу

Можемо приметити да је Але «прималац» управо као и (*a casa di Ale + da Ale*) који се реализује као “место” L2, при чему је *dal negozio* “место” L1.

Лексичко-синтаксичке структуре и класе лексикон-граматике су следеће:

Класа 28 N0 portare N1 Loc N2 Loc N3
Класа 23D N0 portare N1 а N2

Предикати месног трансфера су типа TL(am,m,L1,L2), где је am =: “агенс покрета”, m =: “покретни објекат”, L1 =: “полазиште”, L2 =: “одредиште”. У складу са тиме, реченица (1) може бити схематизована на следећи начин:

TL =: *portare*

am =: N0 =: *Max*

⁸ Систем обележавања у лексикон-граматици детаљно је описан код Elia A., Martinelli M., D’Agostino E. 1981. Овде желимо да подсетимо да се *N* односи на именичку групу, бројеви 0, 1 и 2 односе се на позицију именичке групе у гећеничној структури: N0 =: “subjekat”, N1, N2, N3 =: “dopune”, Loc означава било који локативни предлог, Ch F означава клаузу, *di Ch F означава да дати низ не постоји. Класе лексикон-граматике су нумерисане: класе од 1 до 15 обухватају непрелазне глаголе који не примају реченичне допуне; класе од 16 до 40 обухватају прелазне глаголе који не примају реченичне допуне; класе од 41 до 60 обухватају прелазне и непрелазне глаголе који примају реченичне допуне. Испитали смо 53 класе, као и око 200 својстава за сваки појединачни глагол.

m =: N1 =: *un topolino*

L1 =: Loc N2 =: *dal negozio*

L2 =: Loc N3 =: *a casa di Ale*

Сем на «објекте», размена се може односити и на «поруке», па бисмо у овом другом случају имали **предикате трансфера порука** који улазе у лексичко-синтаксичке структуре чији се глаголи јављају са реченичном допуном и «људским» актантама.

(3) *Max dice ad Ale che parte*

Мак каже Алеу да одлази

(4) *Max informa Ale che parte*

Мак обавештава Але да одлази

(5) *Max chiacchiera con Ale del fatto che partirà*

Мак ћаска са Алом о чињеници да ће отићи

У овом случају ће предикати имати схему TM(dm,m,rm), где је dm =: “давалац поруке”, m =: “порука”, rm =: “прималац поруке”.

TM =: *dire*

dm =: N0 =: *Max*

m =: N1 =: Ch F =: *che parte (Max parte)*

rm =: N2 =: *Ale*

TM =: *informare*

dm =: N0 =: *Max*

rm =: N1 =: *Ale*

m =: di N2 =: *di Ch F =: *che parte (Max parte)*

TM =: *chiacchierare*

dm =: N0 =: *Max*

rm =: N1 =: *Ale*

m =: di N2 =: del fatto Ch F =: *del fatto che partirà (Max o Ale parte)*

Лексичко-синтаксичке структуре и одговарајуће предикатске схеме у линеарном редоследу дефинисане су овако:

TM(dm,m,rm)

Класа 47 N0 V Ch F N2 =: *Max dice che parte ad Ale*

TM(dm,rm,m), али и TM(dm,m,rm)

Класа 50 N0 V N1 *di Ch F =: *Max informa Ale che parte*

Класа 53 N0 V *di Ch F Prep N2 =: *Max chiacchiera con Ale del fatto che partirà*

Узмимо поново најједноставније предикате трансфера, оне који укључују размену објеката међу особама: глагол *dare* ‘дати’, чија лексичко-синтаксичка структура одговара класи 23D, дозвољава додељивање следећих вредности својим актантима:

N0 *dare* N1 a N2 N0 =: “давалац”, N1 =: “објекат”, N2 =: “прималац”

Узевши у обзир реченицу

(6) *Max ha dato un libro ad Ale*
Макс је дао књигу Алеу

имаћемо:

T =: *dare*
g =: N0 =: *Max*
o =: N1 =: *un libro*
r =: N2 =: *Ale*

Иста схема T(g,o,r) валидна је и за трансформисане пасивне реченице:

(7) *Un libro è stato dato da Max ad Ale*
Књига је дата од стране Макса Алеу
(8) *Un libro è stato dato ad Ale da Max*
Књига је дата Алеу од стране Макса

Реченица

(9) *Ale riceve un libro da Max*
Але прима књигу од Макса

је семантички слична оној под (7) и (8), али се однос између актаната и аргумената разликује:

Класа 23R N0 *ricevere* N1 da N2 N0 =: “прималац”, N1 =: “објекат”, N2 =: “давалац”

схема је ипак идентична:

T =: *ricevere*
g =: N0 =: *Max*
o =: N1 =: *un libro*
r =: N2 =: *Ale*

Узевши у обзир линеарни след аргумената, примећујемо да је схема глагола *dare* T(o,g,r), а глагола *ricevere* ‘примати’ T(r,o,g).

T =: *dare* T =: *ricevere*
g =: N0 =: *Max* r =: N0 =: *Ale*
o =: N1 =: *un libro* o =: N1 =: *un libro*
r =: N2 =: *Ale* g =: N2 =: *Max*

Слично томе, за пасивне реченице под (7) и (8) линеарно уређене схеме су типа T(o,g,r) или T(o,r,g).

Ако посматрамо класе лексикон-граматике, ситуација је следећа:

T(g,o,r)	Класа 23D
N0 V N1 a N2 =: <i>Max ha dato un libro ad Ale</i> (<i>Макс је дао књигу Алеу</i>)	

Примери употребе глагола⁹: *dare*, *passare*₁, *vendere*, *cedere*, *regalare*, *offrire*, *firmare*₁, *trasferire*₁ – дати, проћи, продати, уступити, регулисати, нудити, изнајмити, пребацити

T(r,o,g)	Класа 23Rda
N0 V N1 da N2 =: <i>Ale riceve un libro da Max</i> (<i>Але прима књигу од Макса</i>)	

Примери употребе глагола: *ricevere*, *avere*, *ottenere*, *comprare*, *accettare*, *acquistare*, *firmare*₂ – примити, имати, понудити, купити, примити, купити, изнајмити

T(r,o,g)	Класа 23Ra
N0 V N1 a N2 =: <i>Ale ruba un libro a Max</i> (<i>Але краде књигу Максy</i>)	

Примери употребе глагола: *rubare*, *strappare*, *sfilare*, *sottrarre*, *togliere* – украсти, поцепати, извући, отети, одузети

T(r,g,o)	Класа 24
N0 V N1 di N2 =: <i>Ale priva Max di un libro</i> (<i>Але Макса лишава књиге</i>)	

Примери употребе глагола: *derubare*, *privare* – покрасти, лишити

T(o,g,r)	Класа 7DP
N0 V da N1 a N2 =: <i>Il libro passa da Max ad Ale</i> (<i>Књига прелази од Макса Алеу</i>)	

Примери употребе глагола: *passare*₂, *transitare* – прелазити

⁹ Бројеви уз глаголе означавају да постоје различите употребе за исти глаголски уносак, као у случају глагола *firmare*: Mario fitta una casa a Lia ‘Марио изнајмљује кућу Лији’, Lia fitta una casa da Mario ‘Лија изнајмљује кућу од Марка’.

Класа 30
N0 V N1 da N2 a N3 =: <i>Lia trasferisce un libro da Max ad Ale (Луја преноси књигу од Макса Алеу)</i>

Примери употребе глагола: *trasferire*² – преносити

1.1. Корелирани предикати

Могуће је открити још предиката повезаних са предикатом T. На пример, гледано с правне стране, могли бисмо имати предикат поседовања. Грађански кодекс (Civil Code) јасно разликује “власништво” од “поседовања”.

Власништво је правило закона и означава право да се поседује. Члан 832 Грађанског кодекса тврди да «Приватна лица могу слободно да располажу власништвом које им припада, подложно изменама утврђеним законом («il proprietario ha diritto di godere [949] e disporre delle cose in modo pieno ed esclusivo, entro i limiti e con l’osservanza degli obblighi stabiliti dall’ordinamento giuridico»).

Поседовање је правило силе а не закона (?). Члан 1140 Грађанског кодекса тврди: «Поседовање је присвајање неке ствари или права или уживање ствари или права које имамо и сами користимо или које неко користи у наше име.

Када се то примени на предикат T(g,o,r), могу се направити две нове предикатске схеме, једна за власништво (Pr), друга за поседовање (Ps):

Pr(p1,o), где је p1 =: “власник”, o =: “објекат”
Ps(p2,o), где је p2 =: “поседник”, o =: “објекат”

и аутоматски можемо доделити аргументе p1 и p2 следећим реченицама:

(10) *Ale ha acquistato un libro da Max*
Ale је позајмио књигу од Макса

(11) *Ale ha rubato un libro a Max*
Ale је украо књигу Максу

чије су лексичко-синтаксичке структуре, класа лексикон-граматике и предикатска схема следеће:

Класа 23Rda N0 acquistare N1 да N2 T (r,o,g)
Класа 23Ra N0 rubare N1 a N2 T (r,o,g)

N0 =: Ale (10) =: +p1,+ p2
N0 =: Ale (11) =: +p2, -p2
N2 =: Max (10) =: -p1, -p2¹⁰
N2 =: Max (11) =: +p1, -p2

Претходне схеме можемо да проширимо са:

Класа 23Rda N0 *acquistare* N1 да N2 T (r,o,g)
Pr(r = p1, o) Pr(r = p2, o)
Класа 23Ra N0 *rubare* N1 a N2 T (r,o,g)
Pr(d = p1, o) Ps(r = p1,o)

1.2. Предикати трансфера кроз језике

Када лексичко-синтаксички опис датог језика постоји, могуће је поредити га са другим језицима проналажењем еквивалентне класе. Осим тога, како свака класа садржи употребе глагола које улазе у одређене лексичко-синтаксичке структуре, могуће је за дати семантички предикат доделити вредност одговарајућег семантичког аргумента свакој допуни глагола.

На пример, узевши лексичко-синтаксичку структуру француског, енглеског, италијанског и немачког, можемо успоставити однос међу реченицама и доделити аргументе предиката T(g,o,r) у линеарном редоследу.

Француски:

N0 V N1 a N2 =: *Max a donné un livre à son amie*
T(g,o,r)

Италијански:

N0 V N1 a N2 =: *Max ha dato un libro alla sua amica*
T(g,o,r)

Енглески:

N0 V N1 N2 =: *Max gave his friend a book*
T(g,r,o)
(N0 V N1 to N2 =: *Max gave a book to his friend*
T(g,o,r))

Немачки:

N0(nom) V N1(dat) N2(acc) =: *Max gab seiner Freundin ein Buch* **T(g,r,o)**

Макс је дао књигу својој пријатељици

¹⁰ У реченици *Ale ha acquistato un libro da Max* ‘Ale је набавио књигу од Макса’, Мах није власник књиге будући да је, продајући је, пребацио власништво на Алеа.

Ако узмемо класификацију глагола у италијанском и француском¹¹, која је описана до детаља, добићемо следеће схеме:

Француски	Италијански	Превод
N0 V N1 à N2	N0 V N1 a N2	
36DT donner T(gor)	23D dare T(gor)	дати
38LD lancer T(gor)	28 lanciare T(gor)	бацити
38LR abandonner T(gor)	28SC abband- donare T(gor)	уступити
36DT voler T (rog)	23D rubare T (rog)	украсти
38LS enlever T(rog)	23D togliere T(rog)	одузети
36R abonner T(gro)	23R abbonare T(gro)	претплатити се
N0 V N1 de N2	N0 V N1 di/da N2	
36DT dédom- mager T(gro)	23D dispensare da T(gro)	поштедети
37M1 nantir T(gro)	24 dotare di T(gro)	опскрбити
36DT recevoir T(rog)	23R ricevere da T(rog)	примити
38LS soutirer T(rog)	29 travasare da T(rog)	измамити
36DT arnaquer T (rgo)	23D scroccare T(rog)	измамити
37E gruger T(rgo)	24 spogliare T(rog)	опељешити

2. Семантички веб и лексикон-граматика

Већ низ година се описи лексикон-граматике користе унутар софтвера намењених аутоматској анализи текстова. Ови програми, направљени под управом Мориса Гроса¹²,

¹¹ За француски видети M.Gross 1975, Ch.Leclère 2002, 2005.

¹² Прве програми за конструисање електронских речника и табела лексикон-граматике направили су најпре Филип Васо (Philippe Vasseur) и Натали Бели (Nathalie Bely) седамдесетих, а затим Блондине Куртоа (Blandine

омогућавају проналажење података и информација унутар текстова, као и додељивање етикета и морфосинтаксичких анотација подацима. Ови програми се заснивају на уграђеном систему морфосинтаксичких речника, лексичко-синтаксичких класификација и локалних граматика.

На тај начин је у одређеном електронском корпусу (у рачунару или на мрежи) могуће аутоматски доделити вредности и екстраховати вредности семантичких аргумената на основу предиката придружених лексичко-синтаксичким структурама лексичких уносака. Ова врста аутоматске анализе може се проширити и на текстове на језицима за које постоји комплетна анализа у оквиру лексикон-граматике.

Одавде следи могућност хомогеног и *аутоматског* етикетања семантичких предиката и аргумената пре обраде милијарди текстова присутних на мрежи, као и могућност приступа овом типу информација не само преко именичких облика (као што је то могуће сада) већ и преко глаголских и суплетивних глаголских облика тј. предиката.

Courtois)уCNRS-Laboratoire d'Automatique Documentaire et Linguistique на Универзитету Парис 7. Касније су дуги низ година лингвисти и рачунарски стручњаци радили заједно како би направили нове компоненте за обраду података. Деведесетих је Макс Силберштајн (Max Silberstein) развио први софтвер, INTEX. Након смрти Мориса Гроса 2001, тим рачунарских стручњака на Универзитету Marne La Vallée, под управом Ерика Лапорта (Eric Laporte), развио је UNITEX, софтвер врло сличан INTEX-у. Макс Силберштајн, који је на челу истраживачког тима на Универзитету Franche-Comté, почео је последњих година да развија NOOJ, нови софтвер за аутоматску обраду језика; NOOJ се развио из INTEX-а. Погледати сајт <http://perso.orange.fr/rosavram/pages/noojpag.html> за преузимање NOOJ-а и демо верзија за различите језике, између осталог и италијански. За сваки од ових софтвера су потребни електронски речници, табеле лексикон-граматике и електронске граматике које су конструисали тимова лингвиста за различите језике (француски, италијански, енглески, али и српски и корејски).

Узмимо узорак текста:

Lia aveva lanciato dal finestrino del treno la borsa nera a Luciano, che, dopo una capriola, era riuscito a prenderla, rimettendosi in piedi a fatica. Luciano corse fino alla macchina parcheggiata poco lontano, si sedette alla guida e con il cuore in gola estrasse dalla borsa il pacchetto che gli interessava. Mise in moto e arrivò velocemente al baretto, dove incontrò Carlo, al quale diede il prezioso pacchetto, ricevendo da lui una busta gialla stretta e lunga. Carlo lo pregò di andarsene via di corsa perché c'erano in giro personaggi pericolosi. Luciano ripartì in tutta fretta e arrivò in piazza Merli verso le 23.

(Лија је била бацила с прозора воза црну торбу Лучану, који је, после скока, успео да је ухвати, једва се дочекавши на ноге. Лучано отрча до аутомобила паркираног мало даље, седе на место возача и са срцем у грлу извуче из торбе пакет који га је занимао. Крену и стиже брзо у мали бар, где се нашао с Карлом коме даде драгоцени пакет, добивши од њега једну жуту коверту, узану и дугачку. Карло га замоли да оде одатле јер су у близини биле неке опасне особе. Лучано се врати највећом брзином и стиже на Трг Мерли двадесет тројком.)

На глаголске облике који се јављају у тексту можемо применити раније показане формализације семантичких предиката.

T(g,o,r)	Предикат трансфера
TL(am,m,L1,L2)	Предикат месног трансфера
TM(dm,m,rm)	Предикат трансфера поруке
M(am,L1,L2)	Предикат покрета

На основу неких експерименталних истраживања у NOOJ-у, програмском пакету који такође користи електронске речнике и лексикон-граматичке описе италијанског, могуће је аутоматски доделити корелиране предикате и аргументе, као што је и приказано у следећем програмском излазу у коме су актанти и глаголи детектовани и етикетирани.

Lia <g> aveva lanciato <T> dal finestrino del treno la borsa nera <o> a Luciano <r>, che, dopo una capriola, era riuscito a prenderla, rimettendosi in piedi a fatica. Luciano <am> corse <M> fino alla macchina

parcheeggiata poco lontano dalla stazione <L1>, si sedette alla guida e con il cuore in gola (Luciano) <am> estrasse <TL> dalla borsa <L2> il pacchetto <m> che gli interessava. Mise in moto e (Luciano) <am> arrivò <M> velocemente al baretto <L1>, dove incontrò Carlo al quale <r> (Luciano) <g> diede <T> il prezioso pacchetto <o> (Luciano)<r> ricevendo <T> da lui <g> una busta gialla stretta e lunga <o>. Carlo <gm> lo <rm> pregò <TM> di andarsene via <m> di corsa perché c'erano in giro personaggi pericolosi. Luciano <am> ripartì in tutta fretta e arrivò <M> in piazza Merli <L1> verso le 23.

У складу са овим излазом, можемо направити следеће семантичке упите над текстом:

Ко су даваоци објеката?

Quali sono i soggetti di azione datori di oggetti?

Lia, Luciano, Carlo

Ко су и даваоци и примаоци објеката?

Quali sono i soggetti di azione datori e riceventi di messaggi?

Carlo, Luciano

Ко су агенси покрета?

Quali sono i soggetti di azione agenti di movimento?

Luciano, Luciano

Ко су агенси месног трансфера?

Quali sono i soggetti di azione agenti di trasferimento locativo?

Luciano

Које је одредиште Лучанових кретања?

Qual è il luogo di destinazione dei movimenti di Luciano?

la macchina parcheggiata poco lontano dalla stazione
il baretto
piazza Merli

Шта прима Лучано и од кога?

Che cosa riceve Luciano e da chi?

la borsa nera da Lia
una busta gialla stretta e lunga da Carlo

Шта Лучано даје Карлу ?

Che cosa dà Luciano a Carlo ?

il prezioso pacchetto

Шта Лучано пребацује и одакле?

Che cosa trasferisce Luciano e da dove?

il pacchetto dalla borsa

Можемо да замислимо и многе друге упите, а захваљујући усклађености између лексикон-граматика различитих језика, могуће је истовремено постављати упите за текстове на различитим језицима.

У следећем делу ћемо приказати неопходне међукораке – на нивоу морфосинтаксичке анализе – како би се открили конкретни лексички елементи за које се аргумент и предикати могу везати.

3. Стратегије и тактике за ископавање из текстова

Виетри (2004) даје иновирани приказ италијанске лексикон-граматике и језичких и рачунарских алата развијених на одсеку за Комуникациону науку Универзитета у Салерну¹³. Овде желимо само да поменемо неке основне принципе:

Постоје три типа алата неопходних за анализу текста:

- Електронски морфосинтаксички речници простих и сложених речи. Ти речници садрже основне и флективне облике.
- Табелу глагола, предикативних именица, придева и прилога у оквиру лексикон-граматике
- Локалне граматике у облику коначних аутомата (флективне, морфолошке и синтаксичке граматике)

Примери локалних граматика наведени су у наредним пасусима. Овде ћемо дати пример електронског речника флективних простих речи:

amo,amare,V+t+a+PR+1+s	волим, воleti
amo,N+m+s	удица
cortese,A+f+s	љубазна
cortese,A+m+s	љубазан
di,PREP	од
dottore,N+m+s	доктор

¹³ Видети <http://infolingu.univ-mlv.fr/>.

dottorressa,dottore,N+f+s	докторка, доктор
dottorresse,dottore,N+f+p	докторке, доктор
lentamente,AVV	полако

и пример табеле лексикон-граматике:

			V										
	N0 =: Num	N0 =: il fatto Ch F	N0 =: V1 Comp		N0 V che F a N2	F0 V che Fcong a N2	N0 V di V0 Comp a N2	N0 V di V2 Comp a N2	il fatto Ch F a N2	N1 =: se F o se F	N1 =: N1 V1-inf Comp	N0 V a N2 di N3	Passivo
-	+	+		apportare (донети)	+	-	-	+	+	-	-	-	+
+	-	-		articolare (артикулисати)	+	+	+	+	-	+	-	-	+
+	-	-		assegnare (подвући)	-	+	-	+	+	-	-	-	+
+	-	-		asserire (тврдити)	+	-	+	-	+	-	-	-	+
+	+	+		assicurare (осигурати)	+	-	+	-	+	-	-	-	+
+	-	-		attribuire (доделити)	+	-	-	+	+	-	-	-	+

Ова табела се састоји од формалног описа глагола у форми матрице. Свака врста одговара једном глаголу, а свака колона формалном својству. Знак “+” или “-” значи да глагол може тј. не може да прихвати дато својство. Прве три леве колоне показују својства N0, док се девет десних колоне односи на различите конструкције које глагол прихвата тј. не прихвата.

Предикати трансфера могу да се употребе не само у активним реченицама, већ и у многим врстама трансформисаних реченица, нпр. реченицама добијеним применом пасивизације или проминализације. Једна од реченица из узорка текста:

- (1) *Luciano diede il pacchetto a Carlo*
Лучано даде пакет Карлу

може имати пасивну структуру:

Il pacchetto fu dato a Carlo da Luciano
 Пакет би дат Карлу од (стране) Лучана

или прономинализовану структуру:

Luciano lo diede a Carlo
 Лучано га даде Карлу

У наредним пасусима ћемо разматрати могућност аутоматског корелирања активних, пасивних и прономинализованих облика пре етикетирања семантичких предиката.

3.1. Пасив

У тексту можемо имати пасивне реченице у којима један или више аргумената може бити изостављено. Узимајући у обзир реченицу (1), све ниже наведене реченице су прихватљиви пасивни облици:

- (2) *Il pacchetto fu dato a Carlo da Luciano*
 Пакет би дат Карлу од (стране) Лучана
- (3) *Il pacchetto fu dato da Luciano*
 Пакет би дат од стране Лучана
- (4) *Il pacchetto fu dato a Carlo*
 Пакет би дат Карлу
- (5) *Il pacchetto fu dato*
 Пакет би дат

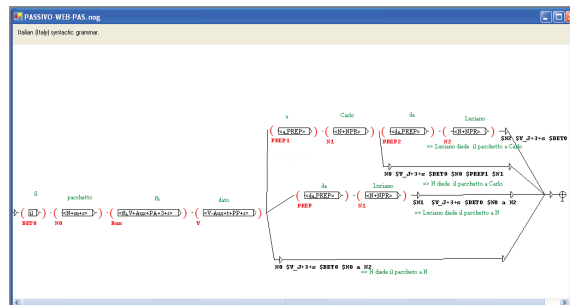
Активни облик глагола *dare* ‘дати’ у (1) има предикатски схему типа T(g,o,r) која додељује вредност ДАВАЛАЦ елементу N0, вредност ОБЈЕКАТ елементу N1, а вредност ПРИМАЛАЦ елементу N2. У одговарајућим пасивним реченицама (2)-(5), предикат *dare* може имати све аргументе, али и само неке од њих, док линеарног уређена схема не мора да буде T(g,o,r):

- у (2): T(o,r,g)
- у (3): T(o,g)
- у (4): T(o,r)
- у (5): T(o)

Неопходно је открити пасивну реченицу и повезати је са одговарајућом активном.

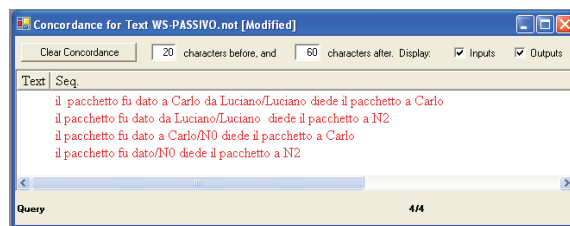
Тек након ове анализе биће могуће исправно етикетирање предиката. NOOЈ омогућава стварање не само флективних и морфолошких, већ и синтаксичких грамика које могу успоставити однос међу реченицама насталим трансформацијама (манипулацијама) које подразумевају замену или померање лексичких елемената, као у случају пасива (погледати белешку 14). Може се конструисати грамика која лоцира пасивне реченице и производи одговарајуће активне реченице.

Следећа слика приказује упрошћену граматику у којој су ДАВАОЦИ и ПРИМАОЦИ лична имена, али аргументи могу да буду и сложенији:



Слика 1. Пример грамике пасива

Ако ову граматику применимо на реченице (2)-(5), добићемо следећи излаз у коме су активне и пасивне реченице повезане:



Слика 2. Пример веза пасивна-активна

У реченици (3), неизражени ПРИМАЛАЦ ће бити пронађен у одговарајућој активној реченици под нотацијом N2; у реченици (4)

ће ДАВАЛАЦ, празан аргумент бити означен као N0, док ће у реченици (5) и ДАВАЛАЦ и ПРИМАЛАЦ бити означени у активном облику нотацијама N0, односно N2. На основу активних структура биће могуће исправно етикетирати предикат *dare*.

3.2. Прономинализација

У реченицама се могу јавити различити облици прономинализације; у овом случају ћемо се позабавити клитицизацијом. За дату реченицу (1):

- (1) *Luciano diede il pacchetto a Carlo*
Лучано даде пакет Карлу

клитицизација се може применити било на ОБЈЕКАТ (= *il pacchetto*):

- (1a) *Luciano lo diede a Carlo*
Лучано га даде Карлу

било на ПРИМАОЦА (= *Carlo*):

- (1b) *Luciano gli diede il pacchetto*
Лучано му даде пакет

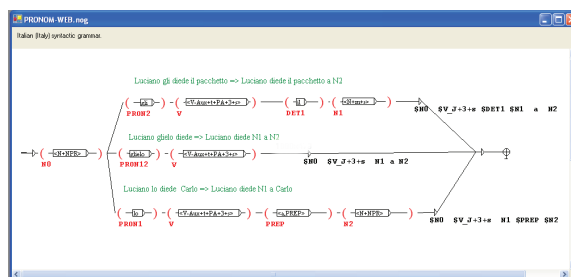
било на оба:

- (1c) *Luciano glielo diede*
Лучано му га даде

У овим реченицама, линеарно уређене схеме предиката *dare* су следеће:

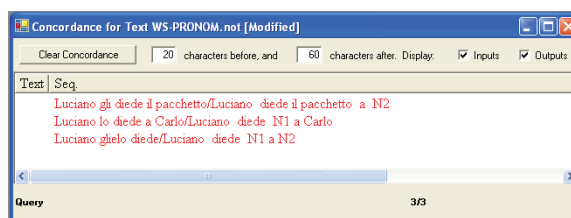
- у (1a): T (g,o,r), где ОБЈЕКАТ долази пре глагола
 у (1b): T (g,r,o), где ПРИМАЛАЦ долази пре глагола
 у (1c): T (g,r,o), где и ПРИМАЛАЦ и ОБЈЕКАТ долазе пре глагола

Што се пасива тиче, пре него што се текст етикетира неопходно је повезати реченице добијене применом клитицизације са одговарајућим полазним реченицама. Овде је приказана граматика која омогућава лоцирање тако трансформисаних реченица и производњу основних реченица:



Слика 3. Пример граматике прономинализације

Примена овакве граматике на реченице (1a)-(1c) даје излаз који садржи следећи однос међу реченицама:



Слика 4. Пример повезивања при прономинализацији

С оваквим излазом се могу исправно етикетирати предикати. Текст који садржи следеће реченице:

Il pacchetto fu consegnato a Carlo. Luciano lo diede a Carlo
Пакет би достављен Карлу. Лучано га даде Карлу

Након примене пасива и прономинализације биће овако етикетиран:

Il pacchetto (o) fu consegnato (T) a Carlo (r). Luciano (g) lo (o) diede (T) a Carlo (r)

3.3. Још о прономинализацији

У италијанском клитицизација примењена на одређене глаголске облике као што су инфинитив, герунд (глаголски прилог садашњи), прошли партицип (глаголски радни придев) и императив нешто је другачија, будући да се клитика мора наћи иза глагола и

семантичких предиката и потоњу аутоматску обраду садржаја текста који није унапред етикетиран.

Језички алати које треба направити и имплементирати су сложени и веома разноврсни, али када се једном направе, пружају могућност примене на милијарде текстова.

Верујемо да развој семантике изграђене на премисама лексикон-граматике нуди добру перспективу развоја за решавање проблема које је покренуо покрет семантичког веба.

Напомена: За додатке овом чланку погледати верзију на енглеском језику, стр. 25а.

Литература

- Berners-Lee T., Hendler J., Lassila O. (2001), *The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities*, "Scientific American", 284(5)
- Boons J.P., Guillet A., Leclère C. (1976). *La structure des phrases simples en français. Constructions intransitives*, Droz: Genève
- Courtois B. (1990a), *Un système de dictionnaires électroniques pour les mots simples du français*, "Langue Française" n.87, Paris: Larousse
- Courtois B. (1990b), *Dictionnaire DELAS par Classes flexionnelles*. Rapport Technique du LADL n. 26, Université Paris 7 : Paris.
- Courtois B., Silberstein M. (1990a), *Développement d'outils de navigation dans les grandes documentations*, Rapport Technique du LADL n. 28, Université Paris 7 : Paris
- Courtois B., Silberstein M. (eds.), (1990b), *Les dictionnaires électroniques du français*. "Langue Française" n. 87, pp. 11-22, Larousse: Paris.
- D'Agostino E. (1992), *Analisi del discorso*, Loffredo: Napoli
- D'Agostino, E. (1995), cura di, *Tra sintassi e semantica. Descrizione e metodi di elaborazione automatica della lingua d'uso*, ESI: Napoli
- D'Agostino E., Elia A.(1998), *Il significato delle frasi: un continuum dalle frasi semplici alle forme polirematiche*, in AAVV, "Ai limiti del linguaggio", Laterza: Bari
- Elia A. (1984), *Le verbe italien*. Schena-Nizet: Bari - Paris
- Elia A. (1995), *Per una disambiguazione semi-automatica di sintagmi composti: i dizionari elettronici lessico-grammaticali*, in Cipriani R., Bolasco S., a cura di, "Ricerca qualitativa e computer", Franco Angeli: Milano
- Elia, A. (1995b), *Standards informatici per la lessicografia computazionale*, in AAVV, "Lessico e grammatica", Bulzoni: Roma
- Elia A. (1995c), *Dizionari elettronici e applicazioni informatiche*, in Bolasco S., Lebart L., Salem A., a cura di "JADT 1995, III Giornate internazionali di Analisi Statistica dei Dati Testuali", CISU: Roma
- Elia A., Renzi L. (1995d), *Il problema delle reggenze*, in AAVV, "Lessico e grammatica", Bulzoni: Roma 1995
- Elia A. (1996), *Elogio dell'imperfezione*, in Gambarrara D., Gensini S., Pennisi A., a cura di, "Language Philosophies and the Language Sciences", Nodus Publikationen: Münster.
- Elia A. (2002), *Discorso scientifico e linguaggio settoriale. Un esempio di analisi lessico-grammaticale di un testo neuro-biologico*, in "Quaderni del Dipartimento di Scienze della Comunicazione – Università di Salerno", Cicalese A., Landi A., a cura di, "Simboli, linguaggi e contesti", n. 2, Carocci: Roma
- Elia A., Martinelli M., D'Agostino E. (1981), *Lessico e Strutture sintattiche*, Liguori: Napoli
- Fuiano L. (2004), *Semantic Web e approccio lessico-grammaticale*, in "Studi Italiani di Linguistica Teorica e Applicata", n.1, Pacini: Roma
- Gross M. (1975), *Méthodes en syntaxe*, Hermann: Paris
- Gross M. (1978), *Observations on Semantic Theories*. "Theoretical Linguistics" 5:1, p. 3-17
- Gross M. (1979), *On the failure of Generative Grammar*, in "Language" 55 :4
- Gross M. (1981), *Les bases empiriques de la notion de prédicat sémantique*, in "Langages" n. 63, Larousse: Paris
- Gross M. (1990), *Sur la notion harissienne de transformation et son application au français*, in "Langages" n. 99, Larousse: Paris
- Gross M. (2002), *Le temps grammatical et le temps qui s'écoule*, LADL: Paris
- Guillet A., La Fauci N. (1984), eds, *Lexique-Grammaire des Langues Romanes*, John Benjamins: Amsterdam/Philadelphia
- Guillet A., Leclère C. (1992), *La structure des phrases simples en français. Constructions transitives locatives*, Droz: Genève
- Harris Z.S. (1976), *Notes du cours de syntaxe*, Editions du Seuil: Paris
- Harris Z.S. (1978), *Operator-grammar of English*, in "Linguisticae Investigationes", II, 1, John Benjamins: Amsterdam/Philadelphia.

- Harris Z.S. (1982), *A Grammar of English on Mathematical Principles*, Wiley: New York
- Harris Z.S. (1988-95), *Linguaggio e informazione*, Adelphi: Milano
- Leclère C. (2002), *Organization of the Lexicon-Grammar of French verbs*, "Linguisticae Investigationes" 25:1, John Benjamins: Amsterdam/Philadelphia.
- Leclère C. (2005), *The lexicon-grammar of French verbs: a syntactic database*, In "Linguistic Informatics - State of the Art and the Future", Kawaguchi Y. et alii (eds.), Tokyo University of Foreign Studies, UBLI 1, John Benjamins: Amsterdam/Philadelphia.
- Silberztein M. (1993.), *Dictionnaires électroniques et analyse automatique de textes*, Masson: Paris.
- Silberztein M. (1994), *INTEX: a corpus processing system*. In "Proceedings of COLING 1994": Kyoto.
- Silberztein M. eds. (2004), *INTEX pour la Linguistique et le Traitement Automatique des Langues*. "Proceedings of the 4th and 5th INTEX workshop", Bordeaux, May 2001, Marseille, May 2002, Presses Universitaires de Franche-Comté: Besançon.
- Silberztein M. (2004), *NooJ: a cooperative object oriented architecture for NLP*, "Cahiers de la MSH Ledoux 1, INTEX pour la linguistique et le traitement automatique des langues", Muller C., Royauté et M. Silberztein (éds.), Presse Universitaire de Franche-Comté: Besançon.
- Silberztein M. (2005), *NooJ's dictionaries*, in Zygmunt Vetulani (ed.), "Proceedings of the 2nd Language & Technology Conference, April 21-23, 2005, Poznań, Poland", Wydawnictwo Poznańskie Sp. z o.o.
- Tesnière L. (1953), *Esquisse d'une syntaxe structurale*, Kincksieck: Paris
- Vietri S. (1984), *Lessico e sintassi delle espressioni idiomatiche. Una tipologia tassonomica in italiano e in inglese*, Liguori: Napoli
- Vietri S. (1990), *La sintassi delle frasi idiomatiche*, in "Studi italiani di linguistica teorica e applicata", XIX:1, Liviana Ed: Padova
- Vietri S. (1996), *The syntax of the Italian verb essere Prep.* in "Linguisticae Investigationes", XX:2 John Benjamins: Amsterdam/Philadelphia
- Vietri S. (2001), *Navigare nei testi. Teorie e applicazioni informatiche per la linguistica testuale*. Editoriale Scientifica Italiana: Napoli.
- Vietri S., Elia A. (2001), *Analisi automatica dei testi e dizionari elettronici*, in Cordeschi R., Burattini E., ed. 2001. "Intelligenza Artificiale. Manuale per le discipline della comunicazione". Carocci: Roma.
- Vietri S., Elia A., D'Agostino E. (2004), *Lexicon-grammar, Electronic Dictionaries and Local Grammars in Italian*, in Laporte, E., Leclère, C., Piot, M., Silberztein M., eds.. "Syntaxe, Lexique et Lexique-Grammaire " Volume dédié à Maurice Gross. Linguisticae Investigationes Supplementa 24, John Benjamins: Amsterdam/Philadelphia
- Vietri S. (2004), *Lessico-grammatica dell'italiano. Metodi, descrizioni, applicazioni*. UTET: Torino
- Vietri S. (2004), *Lemmatization of Idioms in Italian*, in Muller C., Royauté J., Silberztein M., eds. 2004. "INTEX pour la Linguistique et le Traitement Automatique des Langues. Proceedings of the 4th and 5th INTEX workshop", Bordeaux, May 2001, Marseille, May 2002, Presses Universitaires de Franche-Comté: Besançon
- Vasseux P. (1982), *Le système LEXSYN de classification des données du LADL*. Rapport interne du