

Risorse lessicali per lo studio della struttura argomentale dei verbi

Elisabetta Jezek

Dipartimento di Studi Umanistici
Sezione di Linguistica Teorica e Applicata
Università di Pavia

CLUB
Università di Bologna
7 febbraio 2017

Outline of the talk

- What is a lexical resource?

Outline of the talk

- What is a lexical resource?
- Traditional vs. computational lexicography.

Outline of the talk

- What is a lexical resource?
- Traditional vs. computational lexicography.
- Introducing **Typed Predicate Argument Structures** (T-PAS resource).

Outline of the talk

- What is a lexical resource?
- Traditional vs. computational lexicography.
- Introducing **Typed Predicate Argument Structures** (T-PAS resource).
- Methodology to acquire T-PASs from corpora.

Outline of the talk

- What is a lexical resource?
- Traditional vs. computational lexicography.
- Introducing **Typed Predicate Argument Structures** (T-PAS resource).
- Methodology to acquire T-PASs from corpora.
- **Linguistic issues.**

Outline of the talk

- What is a lexical resource?
- Traditional vs. computational lexicography.
- Introducing **Typed Predicate Argument Structures** (T-PAS resource).
- Methodology to acquire T-PASs from corpora.
- **Linguistic issues.**
- Proposed solutions.

Outline of the talk

- What is a lexical resource?
- Traditional vs. computational lexicography.
- Introducing **Typed Predicate Argument Structures** (T-PAS resource).
- Methodology to acquire T-PASs from corpora.
- **Linguistic issues.**
- Proposed solutions.
- Conclusion and ongoing work.

What is a lexical resource?

- A lexical resource is a repository of information associated with words, such as **word senses**, **part(s) of speech**, **typical combinations**, **idiomatic expressions**, **semantic relations**, **argument structure** ...

What is a lexical resource?

- A lexical resource is a repository of information associated with words, such as **word senses**, **part(s) of speech**, **typical combinations**, **idiomatic expressions**, **semantic relations**, **argument structure** ...
- Typically, they are available in **digital format** and can be queried through a **query system**.

What is a lexical resource?

- A lexical resource is a repository of information associated with words, such as **word senses**, **part(s) of speech**, **typical combinations**, **idiomatic expressions**, **semantic relations**, **argument structure** ...
- Typically, they are available in **digital format** and can be queried through a **query system**.
- They are for both computational applications and linguistic analysis.

- WordNet focuses on the representation of semantic relations between words and word senses.

- WordNet focuses on the representation of semantic relations between words and word senses.
- Entry points in WN are **synsets** - which are sets of cognitive synonyms - each expressing a distinct concept, compiled on the base of **psychological assumptions** regarding the semantic relations holding among words in the mental lexicon (Miller 1995, Fellbaum, 2008).

Verb

- **S: (v) increase** (become bigger or greater in amount) *"The amount of work increased"*
 - [direct troponym](#) / [full troponym](#)
 - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
 - [antonym](#)
 - **W: (v) decrease** [Opposed to: [increase](#)] (decrease in size, extent, or range) *"The amount of homework decreased towards the end of the semester"; "The cabin pressure fell dramatically"; "her weight fell to under a hundred pounds"; "his voice fell to a whisper"*
 - [derivationally related form](#)
 - [sentence frame](#)
- **S: (v) increase** (make bigger or more) *"The boss finally increased her salary"; "The university increased the number of students it admitted"*

Traditional and computational lexicography

- Traditional lexicography tends to rely on **introspective data**.

Traditional and computational lexicography

- Traditional lexicography tends to rely on **introspective data**.
- Computational lexicography uses evidence found in large collections of digitalized texts (**corpora**) that can be quickly accessed through sophisticated **corpus query tools** such as the Sketch Engine (cf. Kilgarriff et al. [2004]).

Goals of computational lexicography

- Apply computer technology to dictionary-making: use software tools (**dictionary writing systems, DWS**) that are designed to manage the entire process of producing a dictionary, from the compilation of entries to final publication.

Goals of computational lexicography

- Apply computer technology to dictionary-making: use software tools (**dictionary writing systems, DWS**) that are designed to manage the entire process of producing a dictionary, from the compilation of entries to final publication.
- Build **computational lexicons** designed primarily not for human users but as components of computational systems intended to process natural language with the goal of enhancing human–machine interaction.

Goals of computational lexicography

- Apply computer technology to dictionary-making: use software tools (**dictionary writing systems, DWS**) that are designed to manage the entire process of producing a dictionary, from the compilation of entries to final publication.
- Build **computational lexicons** designed primarily not for human users but as components of computational systems intended to process natural language with the goal of enhancing human-machine interaction.
- The first electronic lexicons in the 60's were digitalized versions of ordinary dictionaries and thesauri (so called **machine-readable dictionaries**).

The T-PAS resource

- T-PAS is a repository of Typed Predicate Argument Structures for Italian, acquired from corpora by **manual clustering of distributional information about Italian verbs.**

The T-PAS resource

- T-PAS is a repository of Typed Predicate Argument Structures for Italian, acquired from corpora by **manual clustering of distributional information about Italian verbs**.
- Under development at the Dept. of Humanities of the University of Pavia, in collaboration with the Human Language Technology Group of Fondazione Bruno Kessler (FBK), Trento.

The T-PAS resource

- T-PAS is a repository of Typed Predicate Argument Structures for Italian, acquired from corpora by **manual clustering of distributional information about Italian verbs**.
- Under development at the Dept. of Humanities of the University of Pavia, in collaboration with the Human Language Technology Group of Fondazione Bruno Kessler (FBK), Trento.
- Technical support of the Faculty of Informatics at Masaryk University in Brno (CZ).

The T-PAS resource

- T-PAS is a repository of Typed Predicate Argument Structures for Italian, acquired from corpora by **manual clustering of distributional information about Italian verbs**.
- Under development at the Dept. of Humanities of the University of Pavia, in collaboration with the Human Language Technology Group of Fondazione Bruno Kessler (FBK), Trento.
- Technical support of the Faculty of Informatics at Masaryk University in Brno (CZ).
- For linguistic research and to be exploited in semantic processing tasks.

- The resource consists of three components:

Resource architecture

- The resource consists of three components:
- A repository of T-PASs linked to lexical units (verbs).

Resource architecture

- The resource consists of three components:
- A repository of T-PASs linked to lexical units (verbs).
- An inventory of about **230 corpus-derived semantic classes for nouns**, relevant for assigning a specific interpretation to the verb in context;

Resource architecture

- The resource consists of three components:
- A repository of T-PASs linked to lexical units (verbs).
- An inventory of about **230 corpus-derived semantic classes for nouns**, relevant for assigning a specific interpretation to the verb in context;
- A corpus of sentences that instantiate T-PAS, tagged with lexical unit (verb) and pattern number.

- The first T-PAS release contains 1000 analyzed **average polysemy verbs** (for a total of 4241 patterns), selected on the basis of random extraction of 1000 lemmas out of the total set of fundamental lemmas of Sabatini Coletti Dictionary 2008.

- The first T-PAS release contains 1000 analyzed **average polysemy verbs** (for a total of 4241 patterns), selected on the basis of random extraction of 1000 lemmas out of the total set of fundamental lemmas of Sabatini Coletti Dictionary 2008.
- Verbs have been selected according to the following proportions: 10% 2-sense verbs, 60% 3-5-sense verbs, 30% 6-11-sense verbs.

What counts as a Typed Predicate-Argument Structure?

A Typed Predicate Argument Structure (T-PAS) is a corpus-derived predicate argument structure with the specification of the expected semantic type for each argument position, assigned by generalizing over the lexical set (Hanks 1986) that typically fill each argument position in the corpus.

[[Human]-subj] partecipa a [[Event]-iobj_a]

- Lexical set [[Event]] = {gara, riunione, selezione, manifestazione, seduta, cerimonia, conferenza, votazione, elezione, celebrazione, esequia, competizione, maratona, discussione, messa, festa, marcia, fiaccolata, trattativa, missione, commemorazione, incontro, concorso, convegno, raduno, iniziativa, stage, evento, seminario, torneo, attività, corso, asta, dibattito, progetto, festival. . . }

Span of context considered as T-PAS

- Complements are included in T-PAS if they contribute to the way the verb is interpreted in the context of use.

Span of context considered as T-PAS

- Complements are included in T-PAS if they contribute to the way the verb is interpreted in the context of use.
- Expressions that are instantiated by T-PAS are disambiguated.

Span of context considered as T-PAS

- Complements are included in T-PAS if they contribute to the way the verb is interpreted in the context of use.
- Expressions that are instantiated by T-PAS are disambiguated.
- This has important consequences for the use of T-PAS in NLP tasks (question answering, textual entailment etc.).

Syntax and Semantics in T-PAS

T-PAS are semantically determined. Different syntactic realizations are encoded as alternating subcategorization frames within the same T-PAS.

[[Human]-subj] finisce [[Event]-obj | di INF [V]

- Finisce **l'allenamento**, la gente l'applaude.
- Non faccio in tempo a finire **di bere la mia birra** che viene un mio amico.

Acquisition Procedure: Background

- 1 A reduced version of ITWaC corpus (Baroni & Kilgarriff 2006).

Acquisition Procedure: Background

- 1 A reduced version of ITWaC corpus (Baroni & Kilgarriff 2006).
- 2 CPA procedure (Corpus Pattern Analysis), Hanks - Pustejovsky 2005.

Acquisition Procedure: Background

- 1 A reduced version of ITWaC corpus (Baroni & Kilgarriff 2006).
- 2 CPA procedure (Corpus Pattern Analysis), Hanks - Pustejovsky 2005.
- 3 A model of compositional mechanisms that are active in V-ARG selection (type coercion, etc.; Pustejovsky 2006).

Acquisition Procedure: Background

- 1 A reduced version of ITWaC corpus (Baroni & Kilgarriff 2006).
- 2 CPA procedure (Corpus Pattern Analysis), Hanks - Pustejovsky 2005.
- 3 A model of compositional mechanisms that are active in V-ARG selection (type coercion, etc.; Pustejovsky 2006).
- 4 A “shallow” list of semantic type labels (HUMAN, ARTEFACT, EVENT, etc.) borrowed from the English resource.
(<http://deb.fi.muni.cz/cpa/#onto>)

Acquisition Procedure: Background

- 1 A reduced version of ITWaC corpus (Baroni & Kilgarriff 2006).
- 2 CPA procedure (Corpus Pattern Analysis), Hanks - Pustejovsky 2005.
- 3 A model of compositional mechanisms that are active in V-ARG selection (type coercion, etc.; Pustejovsky 2006).
- 4 A “shallow” list of semantic type labels (HUMAN, ARTEFACT, EVENT, etc.) borrowed from the English resource.
(<http://deb.fi.muni.cz/cpa/#onto>)
- 5 A suite of corpus tools: Manatee, Bonito, Sketch Engine (Kilgarriff, Rychly, Smrz, Tugwell 2004).

Corpus Pattern Analysis procedure (CPA, Hanks 2004)

- Choose a target verb and create a sample concordance of 250 actual uses.

Corpus Pattern Analysis procedure (CPA, Hanks 2004)

- Choose a target verb and create a sample concordance of 250 actual uses.
- Identify the relevant structures (typical syntagmatic patterns corresponding to the minimal contexts where all words are disambiguated).

Corpus Pattern Analysis procedure (CPA, Hanks 2004)

- Choose a target verb and create a sample concordance of 250 actual uses.
- Identify the relevant structures (typical syntagmatic patterns corresponding to the minimal contexts where all words are disambiguated).
- Associate a typing constraint to each argument position in the patterns.

Corpus Pattern Analysis procedure (CPA, Hanks 2004)

- Choose a target verb and create a sample concordance of 250 actual uses.
- Identify the relevant structures (typical syntagmatic patterns corresponding to the minimal contexts where all words are disambiguated).
- Associate a typing constraint to each argument position in the patterns.
- Assign the instances of the sample to one of the patterns. Examples of each verb which are considered either as “not pertinent” or as “undecidable” are excluded.

Corpus Pattern Analysis procedure (CPA, Hanks 2004)

- Store the patterns (with the associated corpus instances) in the resource.

Corpus Pattern Analysis procedure (CPA, Hanks 2004)

- Store the patterns (with the associated corpus instances) in the resource.
- Associate each pattern with at least one sense, expressed in the form of one (or more) implicatures linked to the typing constrains specified in the pattern.

Corpus Pattern Analysis procedure (CPA, Hanks 2004)

- Store the patterns (with the associated corpus instances) in the resource.
- Associate each pattern with at least one sense, expressed in the form of one (or more) implicatures linked to the typing constrains specified in the pattern.
- [[Human]-subj] **essere presente a** [[Event]-iobj_a].

Unclassified sample from ITWaC for the verb *arrestare* 1/2

Concordance

Word List

Word Sketch

Thesaurus

Find X

Sketch-Diff

CPA

?

Save

View options

KWIC/Sentence

Sort

Left | Right

Node

References

Shuffle

Sample

Filter

Frequency

Node tags

Node forms

Collocations

ConcDesc

Original Conc.

?

Corpus: ITWAC (reduced)

Hits: 28952 (30.9 per million) sampled to 250 Annotating: arrestare [Info](#) [Sort](#) [Finish](#) New pattern: Add Number globally:

[First](#) | [Previous](#) Page of 3 [Go](#) [Next](#) | [Last](#)

#329254756

vostra maestà, in nome del mio governo, di

arrestare

questi tre uomini e di farli tradurre immedia-

#329382915

8 mesi di reclusione il tifofo vicentino

arrestato

ieri prima di Verona-Vicenza. Processato

#333053505

non pochi minuti al giorno non potevano

arrestare

la valanga di parole che ogni giorno la

#333334035

riuscire per poco a non essere, diciamo,

arrestato

, per me quello è stato un miracolo. Poi

#341298959

manutenzione delle tombe, o ai turisti musulmani

arrestati

perché filmavano la basilica di san Petronio

#343354957

ex detenuto politico saharawi, è stato

arrestato

ieri appena sceso dall'aereo all'aeroporto

#345644392

Quindi le strutture investigative che avevano

arrestato

Curcio, smantellato le colonne torinesi

#356196998

ERA A ROMA Roma, 18:50 LONDRA: HUSSAIN

ARRESTATO

A CASA DEL COGNATO A ST. TERMINI Roma

#356197118

TRE FERMI Londra, 16:30 LONDRA: BBC,

ARRESTATI

PRESUNTI TERRORISTI OVAL E AUTOBUS Hous

#375439893

redini del suo destriero. La carovana s'

arrestò

come rispondendo ad un riflesso condizionat

#381550505

del carcere di Pavia, dopo essere stato

arrestato

e pestato a sangue dai Carabinieri per

#384150711

ambito dell'operazione "Download", hanno

arrestato

- in esecuzione di o.c.c. - 41 affiliati

#389129479

07 Ago 2005, 15:44 Non nel senso che mi

arrestano

, ma nel senso che dopo l'ultima settimana

#391433189

conseguenze. Due sono scappati, il terzo è stato

arrestato

Da "Corriere Adriatico" del 27 giugno

#394800942

Seluce: Aria: Mi volgo ad ogni fronda, m'

arresta

7 - Elisa: Aria: Se talor miri un fior

#402598525

E' un flusso lento ma incessante, che si

arresterà

solo a notte inoltrata dopo aver ubriacato

#408159170

Carabinieri della Compagnia di Tolentino hanno

arrestato

4 persone, responsabili di spaccio di

#413769789

dice - alla frontiera siriana sono stati

arrestati

1300 terroristi. Custodire le frontiere

#414434423

già. Perché il "famoso" clandestino,

arrestato

alla frontiera, o comunque arrestato all'

#415400469

è stato decapitato a Jeddah. Era stato

arrestato

mentre contrabbandava "una certa quantità

#418049003

un appunto in cui si diceva che era stato

arrestato

, tra gli altri, il brigatista Lauro Azzolini

#420202409

pressione sul piano. Il disco si sposta senza

arrestarsi

sulla superficie da levigare. Pulire regolarm

#423235353

coloni che si credono più pii: la polizia ha

arrestato

quattro terroristi ebrei che avevano pianific

Unclassified sample from ITWaC for the verb *arrestare* 2/2

Concordance

Word List

Word Sketch

Thesaurus

Find X

Sketch-Diff

CPA

?

Save

Change options

Clustering

Sorting

Gramrels

MW links

More data

Less data

arrestare (-v)

ITWAC (reduced) freq = 28952 (30.9 per million)

Annotating: arrestare [Info](#) [Finish](#) New label:

postV_N	7770	4.0
pregiudicato	47	42.11
latitante	40	37.31
scafista	21	33.59
spacciatore	31	32.91
declino	61	32.62
emorragia	36	31.75
persona	496	30.74
ladro	48	29.87
marocchino	34	29.47
dimostrante	19	28.01
terrorista	54	27.15
rapinatore	17	25.81
sospetto	51	25.78
trafficante	18	25.45
corsa	68	25.34
colpevole	26	24.08
marcia	44	23.57
caduta	41	23.3
sospettato	12	23.26
manifestante	21	23.05

preV_N	4538	3.4
polizia	407	46.02
carabiniere	112	37.28
etiopio	20	31.46
manifestante	32	30.91
brigatista	16	26.67
giornalista	67	26.32
poliziotto	33	25.53
terrorista	34	24.48
romanista	11	21.33
marocchino	14	21.08
spacciatore	10	20.72
no	19	20.37
tifoso	24	19.96
pachistano	7	19.7
dissidente	10	19.25
latitante	8	18.68
persona	127	18.26
uomo	98	18.0
minorenne	15	17.97
scafista	5	17.15

pp_per	1332	5.1
furto	58	42.44
rapina	42	40.84
detenzione	53	39.46
favoreggiamento	20	36.96
spaccio	25	35.67
omicidio	47	35.09
resistenza	52	32.63
concussione	12	32.36
reato	46	28.35
pedofilia	13	27.3
rissa	15	27.16
truffa	18	27.05
estorsione	11	26.67
violenza	40	26.22
oltraggio	10	26.05
corruzione	17	24.96
vagabondaggio	7	24.3
possessione	27	23.08
omissione	11	22.23
danneggiamento	8	20.56

pp_dal	684	6.2
carabiniere	343	71.43
agente	49	33.46
nazista	18	27.81
tedesco	28	26.68
cc	11	25.14
doganiere	5	22.73
soldato	11	17.07
fascista	7	16.23
poliziotto	7	16.08
militare	12	14.38
americano	11	13.43
regime	7	10.0
uomo	13	9.85

pp_in	608	1.6
flagranza	78	65.12
seguito	37	26.59
occasione	16	16.79
relazione	21	16.16
appartamento	7	14.57
precedenza	7	14.42
esecuzione	9	13.69
base	15	12.29
blocco	6	12.1
fondo	13	11.99
attesa	6	11.42
massa	6	10.32
aula	5	10.27
mezzo	10	10.23
operazione	8	10.14
modo	16	9.61
momento	10	9.42
tempo	16	8.75
passato	5	8.05
condizione	8	7.75

Patterns 1 and 2 and associated senses for the verb *arrestare*

The screenshot shows a window titled "arrestare: CPA Patterns". The main area is a light green header with the text "Patterns for: arrestare" and an "Add" button. Below this is a control bar with a "Save" button, a "Sample size" input field, a "Semantic class" input field, and the text "Erlangen No". The main content area contains two numbered patterns:

- 1 **[[Human 1]] arrestare [[Human 2]]**
[[Human 1]] seize [[Human 2]] by legal authority; [[Human 1]] take [[Human 2]] into custody
- 2 **[[Entity]] arrestare [[Process]]**
[[Entity]] cause [[Process]] to stop

Tagged contexts Pattern 1 and Pattern 2 *arrestare*

Concordance

Word List

Word Sketch

Thesaurus

Find X

Sketch-Diff

CPA

?

Save

View options

KWIC/Sentence

Sort

Left | Right

Node

References

Shuffle

Sample

Filter

Frequency

Node tags

Node forms

Collocations

ConcDesc

Original Conc.

?

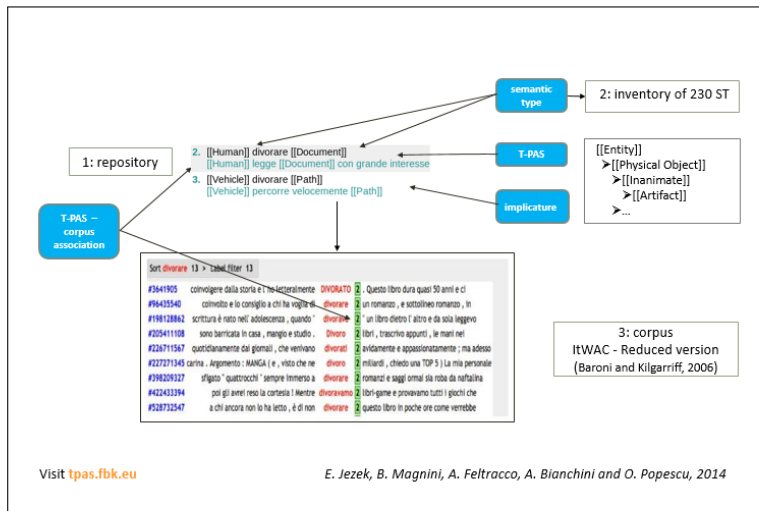
Corpus: itWAC (reduced)

Hits: 28952 (30.9 per million) sampled to 250 Annotating: arrestare [Info](#) [Sort](#) [Finish](#) New pattern: [Add](#) Number globally:

Page 1 of 3 [Go](#) [Next](#) | [Last](#)

#17133668	rimasto coinvolto . L' immigrato è stato	arrestato	1	per violazione della Bossi-Fini , per non
#17133879	. L' uomo è stato condotto in questura e	arrestato	1	. Da " Brescia Oggi " del 7 marzo 2005
#32141614	Secondo Dávalos , gli accusati sono stati	arrestati	1	in modo violento e senza previa notifica
#33957699	italiani implicati fra i golpisti furono	arrestati	1	e condannati e , qualche anno dopo , scambiati
#277198846	sono rispettati , i sindacalisti vengono	arrestati	1	, sulla pena di morte Pechino ha un triste
#282698528	quartieri di Ostia e della Cassia , hanno	arrestato	1	un 40 enne , imprenditore edile Incensurato
#282844992	operazione denominata " Caracas Express " , hanno	arrestato	1	un pregiudicato trovato in possesso di
#307731833	guerra del '15 - 18 . Assieme a Gianì furono	arrestate	1	la madre ottantenne Gisella e la moglie
#308159426	manifestazioni , durante le quali vengono	arrestati	1	molti dimostranti . 1989 - 9 - 10 febbraio
#310111899	attentato di ieri , nel settembre scorso fece	arrestare	1	numerosi militanti islamici libanesi che
#319856328	criminalità alla radice , senza cercare di	arrestare	1	i delinquenti , con i pessimi risultati
#328940320	del paese . Circa 50 persone sono state	arrestate	1	, mentre sette sono morte ed una decina
#381550505	del carcere di Pavia , dopo essere stato	arrestato	1	e pestato a sangue dai Carabinieri per
#413769789	dice - alla frontiera siriana sono stati	arrestati	1	1300 terroristi . Custodire le frontiere
#423235353	coloni che si credono più pii : la polizia ha	arrestato	1	quattro terroristi ebrei che avevano pianificato
#477484027	per molto meno , la procura di Cosenza ha	arrestato	1	quaranta persone . Quelle antenne criminali
#575524911	adoratori di Satana , ha fatto sì che la polizia	arrestasse	1	un noto uomo politico mentre celebrava
#838723077	in due bar del villaggio di Lazerat per	arrestare	1	un ricercato . Impunità Non sempre i procuratori
#26251532	che barriere climatiche o fisiche possano	arrestare	2	nel corso di migliaia di anni i flussi
#92019511	" dei mercati finanziari potrebbe quindi	arrestare	2	questo sviluppo . 6) La proposta del governo
#97177376	totale delle vendite di avorio , anche per	arrestare	2	la ripresa del braccionaggio . Dopo la vendita
#262306676	contribuire alla rivitalizzazione dell' area ed	arrestare	2	il degrado di cui da alcuni decenni è stata
#272370412	quattro sì , perché non è giusto che si	arrestati	2	il cammino della scienza . Che può presentare

Architecture of the T-PAS resource



Issue: Type mismatches

- The paradigmatic sets of words that populate specific argument slots within the same verb sense do not map neatly onto the “expected” type (selected by V) (Pustejovsky and Jezek 2008).

Issue: Type mismatches

- The paradigmatic sets of words that populate specific argument slots within the same verb sense do not map neatly onto the “expected” type (selected by V) (Pustejovsky and Jezek 2008).
- **Mismatches** between “pattern” type (expected by V) and “instance” type (inherent in N) within the same grammatical relation.

Issue: Type mismatches

- The paradigmatic sets of words that populate specific argument slots within the same verb sense do not map neatly onto the “expected” type (selected by V) (Pustejovsky and Jezek 2008).
- **Mismatches** between “pattern” type (expected by V) and “instance” type (inherent in N) within the same grammatical relation.
- The phenomenon is pervasive (Jezek and Quochi, 2010) and spread over several of Levin's 1993 verb classes (aspectual verbs, communication verbs, perception verbs, directed motion verbs).

Type mismatch: Aspectual verbs

[[Human]-subj] interrompe [[Event]-obj]

- Arriva Mirko e interrompe **la conversazione**.
'Mirko arrives and interrupts the conversation' (matching)
- Il presidente interrompe **l'oratore**.
'The president interrupts the speaker' (**HUMAN** as **EVENT**)

Type mismatch: Communication verbs

[[Human]-subj] annuncia [[Event]-obj]

- Lo speaker annuncia **la partenza**.
'The speaker announces the departure' (matching)
- Il maggiordomo annuncia **gli invitati**.
'The butler announces the guests' (**HUMAN** as **EVENT**)
- **L'altoparlante** annunciava l'arrivo del treno.
'The loudspeaker announces the arrival of the train' (**ARTIFACT** as **HUMAN**)
- **Una telefonata anonima** avvisa la polizia.
'An anonymous telephone call alerted the police' (**EVENT** as **HUMAN**)

Type mismatch: Perception Verbs

[[Human]-subj] ascolta [[Sound]-obj]

- Rilassarsi ascoltando **il rumore della pioggia**.
'Relax while listening to the sound of rain' (matching)
- Ascoltava **la radio** con la cuffia.
'He listened to the radio with his earphones' (**ARTIFACT** as **SOUND**)
- Rimasi a lungo ad ascoltare **il suo respiro**.
'I stayed for a long while listening to his breath' (**EVENT** as **SOUND**)
- Non ho potuto ascoltare **tutti i colleghi**
'I could not listen to all colleagues' (**HUMAN** as **SOUND**)

Type mismatch: Directed motion verbs

[[Human]-subj] raggiunge [[Location]-obj]

- Abbiamo raggiunto **l'isola** alle 5.
'We reached the island at 5' (matching)
- Ho raggiunto **il semaforo** e ho svoltato a destra.
'I reached the traffic light and turned right' (**ARTIFACT** as **LOCATION**)

Type mismatch: Directed motion verbs

[[Human]-subj] arriva (Adv [[Location]])

- Alla fine, ormai col buio, sono arrivata **a una radura**.
'Finally in the dark I came upon a clearing.' (matching)
- Gli invitati arrivano **al concerto** in ritardo.
'The guests arrived late at the concert' (**EVENT** as **LOCATION**)

Type mismatch: Motion using a vehicle

[[Flying Vehicle]-subj] atterra ([[Location]]-adv)

- Il **nostro aereo** atterra alle 21.
'Our plane lands at 9pm' (matching)
- Il **pilota** e' regolarmente atterrato senza problemi.
'The pilot landed regularly with no problems' (**HUMAN** as **VEHICLE**)
- **Tutti i voli civili** sono atterrati.
'All civilian flights landed' (**EVENT** as **VEHICLE**)

Type mismatch: Vehicle Verbs

[[Human]-subj] parcheggiare ([[Vehicle]-obj])

- **Luca** ha parcheggiato sotto casa.
'Luca parked near the house' (matching)
- **L'ambulanza** ha parcheggiato lontano.
'The ambulance parked far away' (**VEHICLE** as **HUMAN**)

Shimmering Lexical Sets

- Lexical sets populating a node in the ontology (i.e. a semantic type) tend to **“shimmer”** (Jezek and Hanks 2010) – that is, the membership of the lexical set tends to vary when one moves from verb to verb: some words drop out while other come in, just as predicated by Wittgenstein (**family resemblances**).

Shimmering Lexical Sets

- Lexical sets populating a node in the ontology (i.e. a semantic type) tend to **“shimmer”** (Jezek and Hanks 2010) – that is, the membership of the lexical set tends to vary when one moves from verb to verb: some words drop out while other come in, just as predicated by Wittgenstein (**family resemblances**).
- Different verbs select different prototypical members of a semantic type even if the rest of the set remains the same.

Issue: Shimmering Lexical Sets

lavare [[Body Part]-obj]

- Lexical set [[Body Part]] = {denti, mano, piede, viso, faccia, schiena, testa, orecchio, volto ... }

amputare [[Body Part]-obj]

- Lexical set [[Body Part]] = {arto, gamba, braccio, dito, orecchio, mano, piede ... }

- Regular choices of types within an overall pattern are coded as **type alternations** at the pattern level. Common alternations in subject position are for instance `[[Human|Institution]]` and `[[Human|Body Part]]`”.

- Regular choices of types within an overall pattern are coded as **type alternations** at the pattern level. Common alternations in subject position are for instance `[[Human|Institution]]` and `[[Human|Body Part]]`.
- Non-canonical lexical items which still convey the same verb sense are **“coerced”** into ‘honorary’ membership of a semantic type in particular contexts.

- Regular choices of types within an overall pattern are coded as **type alternations** at the pattern level. Common alternations in subject position are for instance `[[Human|Institution]]` and `[[Human|Body Part]]`.
- Non-canonical lexical items which still convey the same verb sense are **“coerced”** into ‘honorary’ membership of a semantic type in particular contexts.
- Honorary members are tagged as **“a”** = anomalous arguments.

- Regular choices of types within an overall pattern are coded as **type alternations** at the pattern level. Common alternations in subject position are for instance `[[Human|Institution]]` and `[[Human|Body Part]]`.
- Non-canonical lexical items which still convey the same verb sense are **“coerced”** into ‘honorary’ membership of a semantic type in particular contexts.
- Honorary members are tagged as **“a”** = anomalous arguments.
- Pustejovsky et al. 2004, Hanks and Jezek 2008, Jezek and Hanks 2010.

Contexts for Pattern1 *arrestare*

[[Human1 | Institution]-subj] arresta [[Human2]-obj]

Sketch Engine

Search

User: jezek Corpus: itWAC (reduced)

Search

In itWAC (r

Concordance

Word List

Word Sketch

Thesaurus

Find X

Sketch-Diff

CPA

?

Save

View options

KWIC/Sentence

Sort

Left | Right

Node

References

Shuffle

Sample

Filter

Frequency

Node tags

Node forms

Collocations

ConcDesc

Original Conc.

?

Corpus: itWAC (reduced)

Hits: 28952 (30.9 per million) sampled to 250 Annotating: arrestare [Info](#) [Sort](#) [Finish](#) New pattern: Add Number globally:

Page 1 of 3 [Go](#) [Next](#) [Last](#)

#6641425

#8956597

#10660502

#16595036

#17133668

#17133879

#25686546

#27685593

#32141614

#33957699

#39160144

#42474029

#61694394

#277198846

#282698528

#282844992

#307731833

#308159426

#310111899

#319856328

#328940320

#381550505

#413769789

#477484027

si è costituito un giovane di Paratico , eventualmente da confutare , non un brigante da semplicemente perché tali persone non sono state carabinieri con calci e pugni , è stato rimasto coinvolto . L' immigrato è stato . L' uomo è stato condotto in questura e accusa Ermione di essergli infedele e li fa liberando alcuni militanti integralisti Secondo Dávalos , gli accusati sono stati Italiani implicati fra i golpisti furono giornalista Sina Motallebi è stato infatti strana accusa di cospirazione BADOGLIO fa febbraio 2001 - Nigeriano " re " della coca - sono rispettati , i sindacalisti vengono quartieri di Ostia e della Cassia , hanno operazione denominata " Caracas Express " , hanno guerra del '15 - 18 . Assieme a Giani furono manifestazioni , durante le quali vengono attentato di ieri , nel settembre scorso fece criminalità alla radice , senza cercare di del paese . Circa 50 persone sono state del carcere di Pavia , dopo essere stato dice - alla frontiera siriana sono stati per molto meno , la procura di Cosenza ha

arrestato

arrestare

arrestate

arrestato

arrestato

arrestato

arrestare

arrestati

arrestati

arrestati

arrestato

arrestare

arrestato

arrestati

arrestato

arrestato

arrestate

arrestati

arrestare

arrestare

arrestate

arrestato

arrestati

arrestato

dai carabinieri . Il 23 enne nell' interrogatorio . Ma allora perché arrestarlo di notte ; se lo fossero state , ci sarebbe un referto a Torre del Greco (Napoli) con l' accusa per violazione della Bossi-Fini , per non . Da " Brescia Oggi " del 7 marzo 2005 entrambi . Toante promette di risparmiare dalla polizia di Arafat - nessuno sembra in modo violento e senza previa notifica e condannati e , qualche anno dopo , scambiati lo scorso aprile con l' accusa di - aver il suo ex collega generale UGO CAVALLERO dalla polizia . Aim Esajobor , 19 anni , sulla pena di morte Pechino ha un triste un 40 enne , imprenditore edile incensurato un pregiudicato trovato in possesso di la madre ottantenne Gisella e la moglie molti dimostranti . 1989 - 9 - 10 febbraio numerosi militanti islamici libanesi che i delinquenti , con i pessimi risultati i , mentre sette sono morte ed una decina e pestato a sangue dai Carabinieri per 1300 terroristi . Custodire le frontiere quaranta persone . Quelle antenne criminali

Contexts with anomalous arguments for Pattern1 *atterrare*

[[Flying Vehicle]-subj] atterra ([[Location]]-adv))

Corpus: itWAC (reduced)

Hits: 4892 (5.2 per million) sampled to 250 Annotating: **atterrare** [Info](#) [Sort](#) [Finish](#) New pattern: Number globally:

[First](#) | [Previous](#) Page of 3 [Next](#) | [Last](#)

#70190621	cercavo tutt ' altro) e mi sembra di essere	atterrato	1.a	su uno strano pianeta : gente che raccoglie
#88500741	BALDAZZI e dal Colonnello Roberto ROCCHI , è	atterrato	1.a	con un aereo C - 130 J all' aeroporto di
#103748746	operatori della sala radar di Linate) sono	atterrato	1.a	all' aereoporto di San damiano (piacenza
#103749868	sicurezza . Per quasi due anni sono sempre	atterrato	1.a	a Linate ... e questo inconveniente non
#103752373	alla minima e fai il circuito a destra e	atterri	1.a	(anche di notte) Per il decollo per la
#129081889	privato e penso che forse sarebbe stato meglio	atterrarci	1.a	e attendere che passasse tutto questo tempaccio
#145305030	10 settembre 2004 alle 22:57 dipende dove	atterri	1.a	e che genere di vacanza vuoi ? se vuoi
#146773219	regolarità durante tutto l' anno . Naturalmente	atterrerete	1.a	e scenderete a Manaus , in cui sosterete
#155929206	mongolfiera ad elio " , con la quale riusci ad	atterrare	1.a	davanti alla palazzina reale di caccia
#160931401	poi non ho capito , ieri pomeriggio sono	atterrata	1.a	io alla Malpensa e non ho trovato nessuno
#206393108	non si era svegliato nemmeno quand ' erano	atterrati	1.a	con qualche scossone , e ciò le faceva
#215314872	costa azzurra ... che dall ' aereo quando	atterri	1.a	a nizza vedi tutti questi occhi celesti
#219649517	Oriente e l' Europa : quasi tutti i voli	atterrano	1.a	all' aeroporto Mehrbad di Teheran . La
#232621527	emergenza del genere . Gli aiuti italiani sono	atterrati	1.a	all' aeroporto di Colombo domenica mattina
#245768623	di Zugdidi . Il luogo in Marocco dove è "	atterrato	1.a	" il grosso asteroide che ha distrutto
#254133888	a Livorno - Il Messagero - James Forrest	atterrerà	1.a	in Italia con un giorno di ritardo . Il

- A node in the ontology (i.e. a semantic type) is not to be thought of as an address for 'all and only' the lexical items that belong to that node.

- A node in the ontology (i.e. a semantic type) is not to be thought of as an address for 'all and only' the lexical items that belong to that node.
- Rather, it is an address for lexical items that typically belong to that node.

- A node in the ontology (i.e. a semantic type) is not to be thought of as an address for 'all and only' the lexical items that belong to that node.
- Rather, it is an address for lexical items that typically belong to that node.
- “The ontology is not conceived as a rigid yes/no structure, but as **a statistically based structure of collocational preferences**, which we call **“shimmering lexical sets”** .

- A node in the ontology (i.e. a semantic type) is not to be thought of as an address for 'all and only' the lexical items that belong to that node.
- Rather, it is an address for lexical items that typically belong to that node.
- “The ontology is not conceived as a rigid yes/no structure, but as **a statistically based structure of collocational preferences**, which we call **“shimmering lexical sets”** .
- Hanks and Jezek 2008.

- ... book <read 772/9037; write 933/13015; publish 416/7230, borrow 43/1358 ... >

- ... book <read 772/9037; write 933/13015; publish 416/7230, borrow 43/1358 ... >
- ... novel <write 182/13015; read 88/9037; publish 45/7230; set 19/14528 ... >

- ... book <read 772/9037; write 933/13015; publish 416/7230, borrow 43/1358 ... >
- ... novel <write 182/13015; read 88/9037; publish 45/7230; set 19/14528 ... >
- ... article <write 263/13015; publish 174/7230; read 156/9037; contribute 28/1313 ... >

- ... book <read 772/9037; write 933/13015; publish 416/7230, borrow 43/1358 ... >
- ... novel <write 182/13015; read 88/9037; publish 45/7230; set 19/14528 ... >
- ... article <write 263/13015; publish 174/7230; read 156/9037; contribute 28/1313 ... >
- ... letter <write 1032/13015; send 540/12011; receive 544/18381; post 77/451 ... >

- ... meeting <attend 663/5355; hold 953/24798; arrange 200/3581; adjourn 36/424, organize 32/2055 ... >

[[Event]]

- ... meeting <attend 663/5355; hold 953/24798; arrange 200/3581; adjourn 36/424, organize 32/2055 ... >
- ... conference <attend 267/5355; hold 382/24798; organize 81/2055; arrange 29/3581 ... >

[[Event]]

- ... meeting <attend 663/5355; hold 953/24798; arrange 200/3581; adjourn 36/424, organize 32/2055 ... >
- ... conference <attend 267/5355; hold 382/24798; organize 81/2055; arrange 29/3581 ... >
- ... lecture <attend 75/5355; deliver 65/3949; give 226/75759; organize 5/2055; hold 12/24798; arrange 5/3581... >

[[Event]]

- ... meeting <attend 663/5355; hold 953/24798; arrange 200/3581; adjourn 36/424, organize 32/2055 ... >
- ... conference <attend 267/5355; hold 382/24798; organize 81/2055; arrange 29/3581 ... >
- ... lecture <attend 75/5355; deliver 65/3949; give 226/75759; organize 5/2055; hold 12/24798; arrange 5/3581... >
- ... concert <stage 18/1157; attend 29/5355; play 27/17832; organize 13/2055; hold 21/24798; arrange 6/3581 ... >

Conclusion

- We have introduced T-PAS, a resource of corpus-derived typed predicate argument structures for Italian language, distributed under a Creative Common license at <http://tpas.fbk.eu/>.

Conclusion

- We have introduced T-PAS, a resource of corpus-derived typed predicate argument structures for Italian language, distributed under a Creative Common license at <http://tpas.fbk.eu/>.
- The current T-PAS release includes typed patterns for 1000 Italian verbs and 4241 patterns.

Conclusion

- We have introduced T-PAS, a resource of corpus-derived typed predicate argument structures for Italian language, distributed under a Creative Common license at <http://tpas.fbk.eu/>.
- The current T-PAS release includes typed patterns for 1000 Italian verbs and 4241 patterns.
- We have described the lexicographic process of manual acquisition of T-PAS.

Conclusion

- We have introduced T-PAS, a resource of corpus-derived typed predicate argument structures for Italian language, distributed under a Creative Common license at <http://tpas.fbk.eu/>.
- The current T-PAS release includes typed patterns for 1000 Italian verbs and 4241 patterns.
- We have described the lexicographic process of manual acquisition of T-PAS.
- We have examined specific linguistic phenomena that must be addressed when we try to extract generalizations from the analysis of the corpus.

Ongoing work

- Systematic annotation of type mismatches.

Ongoing work

- Systematic annotation of type mismatches.
- Analysis of the lexical sets of transitive Object and Intransitive Subjects in causative/inchoative alternating verbs using distributional semantic methods (“Grounding the Lexical Sets of Causative-Inchoative Verbs with Word Embedding”, Ponti, Jezek and Magnini 2016 at D-SALT ESSLI 2016; best paper award at CliC 2016).

Ongoing work

- Systematic annotation of type mismatches.
- Analysis of the lexical sets of transitive Object and Intransitive Subjects in causative/inchoative alternating verbs using distributional semantic methods (“Grounding the Lexical Sets of Causative-Inchoative Verbs with Word Embedding”, Ponti, Jezek and Magnini 2016 at D-SALT ESSLLI 2016; best paper award at CliC 2016).
- Experiments for semi-automatically extending the coverage of the resource by finding new examples of existing patterns and predict new patterns from corpora.

Ongoing work

- Systematic annotation of type mismatches.
- Analysis of the lexical sets of transitive Object and Intransitive Subjects in causative/inchoative alternating verbs using distributional semantic methods (“Grounding the Lexical Sets of Causative-Inchoative Verbs with Word Embedding”, Ponti, Jezek and Magnini 2016 at D-SALT ESSLLI 2016; best paper award at CliC 2016).
- Experiments for semi-automatically extending the coverage of the resource by finding new examples of existing patterns and predict new patterns from corpora.
- Development of a friendly interface for linguistic analyses.

Acknowledgments

We acknowledge Patrick Hanks (University of Wolverhampton) and his collaborators of the CPA, PDEV and DVC projects for scientific support for T-PAS, and Vitek Baisa (Masaryk University) for technical support.

Basic References

- Jezek, E. *The Lexicon: An Introduction*, Oxford, Oxford University Press.

Basic References

- Jezek, E. *The Lexicon: An Introduction*, Oxford, Oxford University Press.
- Jezek E., Magnini B., Feltracco A., Bianchini A., Popescu O. (2014). “T-PAS: A resource of Typed Predicate Argument Structures for linguistic analysis and semantic processing”. In: Calzolari N. et al. (eds) *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014)*, Paris: European Language Resources Association.



T-PAS Typed Predicate Argument Structures for Italian



[Home](#)
[Resource](#)
[Download](#)
[People](#)
[Acknowledgments](#)

Resource

The first T-PAS release (T-PAS 1.0) contains 1000 Italian verbs with average polysemy, selected on the basis of a random extraction of 1000 verbs out of the set of fundamental lemmas of the Sabatini Coletti 2008 dictionary, according to the following proportions: 10% 2-sense verbs, 60% 3-5-sense verbs, 30% 6-11-sense verbs.

The resource consists of three components:

- 1) a repository of corpus-derived T-PAS linked to lexical units (verbs);
- 2) an inventory of about 230 corpus-derived semantic classes for nouns, relevant for disambiguation of the verb in context;
- 3) a corpus of sentences that instantiate T-PAS, tagged with lexical unit (verb) and pattern number.

For a complete description of the resource as well as of the acquisition procedure, please refer to the following paper:

Elisabetta Jezek, Bernardo Magnini, Anna Feltracco, Alessia Bianchini, Octavian Popescu: T-PAS: A resource of corpus-derived Typed Predicate Argument Structures for linguistic analysis and semantic processing, Proceedings of LREC 2014.

The Lexicon

An Introduction

By **Elisabetta Ježek**, Associate Professor of Linguistics,
University of Pavia

Oxford Textbooks in Linguistics

9780199601530 | Hardback | £65.00 | January 2016

9780199601547 | Paperback | £24.99 | January 2016

