

# Acquisizione di strutture argomentali “tipate” da corpora

Elisabetta Jezek (Pavia)

Workshop Senso Comune  
Roma, 23 marzo 2011

- Lo scopo della presentazione è di illustrare la metodologia adottata per l'acquisizione di strutture argomentali "tipate" da corpora, utilizzata per la compilazione di un repertorio di pattern verbali corpus-based per l'italiano.
- Illustrare i problemi principali problemi che emergono nell'identificazione dei pattern.
- Esplorare i possibili utilizzi lessicografici e computazionali della risorsa, oltreché l'integrazione con altre risorse esistenti.

- Risorsa in costruzione presso l’Università di Pavia con il supporto della Facoltà di Informatica dell’Università Masaryk a Brno (CZ), consistente allo stato attuale in un piccolo nucleo di 50 unità lessicali (verbi). Collaboratori: Francesca Frontini (ILC-CNR Pisa), Francesca Mazzariello (Paris 4).
- Nella risorsa, ogni unità lessicale (V) è associata a un set di patterns/frames estratti da corpora secondo la metodologia della Corpus Pattern Analysis (CPA, Hanks - Pustejovsky 2005) e a una serie di contesti che rappresentano casi di istanziazione dei frames.
- <http://nlp.fi.muni.cz/projects/cpa/>

# Che cosa è una struttura argomentale “tipata”?

Una struttura argomentale “tipata” è un frame verbale con specificazione del tipo semantico atteso per le posizioni argomentali - ‘popolati’ da set lessicali.

[[Persona]] partecipa a [[Evento]]

- Set lessicale [[Evento]] = {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. . . }

- Un corpus bilanciato della lingua.  
ITWaC (Baroni & Kilgarriff 2006).
- Una metodologia di analisi del corpus.  
CPA, Corpus Pattern Analysis, Hanks - Pustejovsky 2005.  
Per l'analisi statistica, Church - Hanks 1989.
- Un modello dei meccanismi compositivi nella selezione V-ARG (type coercion ecc.).  
Pustejovsky 2006.
- Una lista "shallow" di etichette di tipi semantici (HUMAN, ARTEFACT, EVENT, ecc.) mutuata dalla analoga risorsa inglese.  
<http://deb.fi.muni.cz/pdev/>
- Una suite di corpus tools: Manatee, Bonito, Sketch Engine.  
Kilgarriff, Rychly, Smrz, Tugwell 2004.

- Creazione di un campione di concordanze random (KWIC index) per il verbo target.
  - 250 esempi d'uso per ogni verbo target.
- Analisi del set lessicale per ogni relazione grammaticale (SUBJ\_OF, OBJ\_OF, IOBJ\_OF ...).
- Clustering manuale dei set lessicali.
- Identificazione dei frames.
- Assegnazione di ogni istanza del campione a uno dei frames individuati.
- Catalogazione dei frames nella risorsa.
- Associazione di un senso a ciascun frame; il senso è “ancorato” ai tipi semantici specificati nel frame.

# Campione non classificato da ITWaC (Baroni and Kilgarriff 2006)

- Gianfranco , ti aggiorno che oltre ai dolci ... **divoro** anche le verdure.
- la condanna viene eseguita, un incendio **divora** Mahagonny.
- si incolla, il gelato squagliato, cibi **divorati** distrattamente, senza un pensiero.
- quinta tappa, da Blois a Montargis, è **stata divorata** alla incredibile media di 48,584.
- capace di acchiappare al volo i neofiti (che **divorati** dalla voglia di leggere i numeri passati...
- rimanendo di guardia affinché la grafica non **divori** la fantasia artistica
- Una volta uccisi i nemici, tendono a **divorarli** per acquisirne il potere.
- invece direttamente nella polpa, raggiunge e **divora** i semi. Gli attacchi del fitofago su melo
- posso approvare che la bestia più grande **divori** la bestia più piccola
- coinvolgere dalla storia e l' **ho** letteralmente **divorato**. Questo libro dura quasi 50 anni

## divorare

- [[Animate]-subj] divora [[Food]-obj]
- [[Human]-subj] divora [[Document]-obj]
- [[Vehicle]-subj] divora [[Path]-obj]
- [[Emotion]-subj] divora [[Human]-obj]
- .....

# Pattern per il verbo divorare

The screenshot shows a window titled "divorare: CPA Patterns". The window has a green header bar with the text "Patterns for: divorare" and several buttons: "Add", "Copy", "Corpora", "Preview", "Renumber", "Delete", and "Close". Below the header bar is a grey bar with a "Save" button, two input fields for "Sample size" and "Semantic class", and the text "Erlangen No". The main area of the window contains a list of five patterns, each with a checkbox on the left:

- 1 **[[Animate]] divorare [[Food]]**  
[[Animate]] ingest and consume [[Food]] quickly and eagerly
- 2 **[[Human]] divorare [[Document]]**  
[[Human]] read [[Document]] quickly and eagerly
- 3 **[[Vehicle]] divorare [[Path]]**  
[[Vehicle]] cover [[Path]] quickly
- 4 **[[Emotion]] divorare [[Human]]**  
[[Emotion]] consume, absorb [[Human]]
- 5 **[[Event]] divorare [[Physical Object]]**  
[[Event]] destroy [[Physical Object]] completely

# Pattern1 [[Animate]-subj] divorare [[Food]-obj]

- quando nessuno lo vedeva, **divorò** il salame da solo
- Mentre **divoravano** la cena
- **l'aquila** che **divora** il **serpente**
- Cinzia non **divorò** il panettone
- Durante la festa i cittadini **hanno divorato** in pochi secondi una torta di 8
- E **le cavallette**, buone solo a **divorare** i raccolti?
- E' molto magra, **divora** i biscotti con gran voracità
- Nascostamente **le larve avevano** divorato **i bordi delle tele**
- davvero molto bellina!! Oltre a **divorare** gelati (d' estate me ne mangio
- Bevono caffelatte, **divorano** cornetti, comprano
- **Il leone** che assalta e **divora l'uomo** è un' offesa per la ragione
- sorseggiare un caffè' in tranquillita' o **divorare** un panino
- **bruchi**, che minacciano di **divorare le foglie** delle vostre piante
- a guardare la televisione a bere birra e **divorare** patatine.

## Pattern2: [[Human]-subj] divora [[Document]-obj]

- di mettersi a studiare, a **divorare** libri come non mai.
- e appena ho avuto la possibilità **ho divorato** il libro.
- L' amica **divorò** il romanzo in meno di tre giorni.
- Durante le medie **ho divorato** i racconti della "Golden Age".
- Nelle ore libere **divoravo** i grossi e piccoli volumi di storia.
- Legge soprattutto romanzi e **divora** i fumetti, sua vera passione.
- Spinti da genitori e insegnanti a **divorare** intere enciclopedie
- **un autore** da **divorare** a cui affidare il nostro recondito.
- ma soprattutto **ho** scoperto e **divorato** **la Deledda**, mi piace da impazzire.

## Pattern3: [[Vehicle]-subj] divora [[Path]-obj]

- sono migliorati a ogni uscita, **hanno divorato** i chilometri.
- **Ha divorato** la pista, la conosceva bene.
- la Maserati **divora** i rettilinei di Adria.
- Il poderoso incrociatore che **divorava** miglia su miglia.
- Il gommone **divora** 71 chilometri sulle onde e lì
- in certe gare **il vincitore divora** 90 chilometri in sette ore.
- la Skoda divora chilometri.

## Pattern4: [[Emotion]-subj] divorava [[Human]-obj]

- frenare la gelosia che ormai lo **divorava**.
- È meglio così. Il tuo amore ti **avrebbe divorato**.
- altrimenti l'ansia e l'angoscia vi **divorerà**.
- Un atroce dubbio mi **divora** e nessuno riesce a scioglierlo.
- sotto lo stimolo della cupidigia che lo **divorava**, sarebbe sicuramente arrivato
- il senso di colpa **divora** il boia Kaliyappan
- con i sensi di colpa che ci **divorano**.
- a combattere la depressione che la **divorava** da anni.
- La tensione mi **sta divorando**.

# Contesti Pattern1 divorare

Annotating: divorare	<a href="#">Info</a> <a href="#">Sort</a> <a href="#">Finish</a>	New pattern: <input type="text"/>	<input type="button" value="Add"/>	Number globally: <input checked="" type="checkbox"/>	
#76962398		piatto di pasta con i funghi , mi guardavi e	divoravi	il	quel piatto , stavo andando via quando
#94270451		scampare al mare , il pericolo di finire	divorati	il	dai selvaggi sarebbe stato superiore alla
#121260390		passione è mangiare ; sei come una termite :	divori	il	tutto il commestibile che ti capita a portata
#125028431		attorno come leone ruggente , cercando di	divorare	il	ma la sua attività è frenata durante
#126504501		mangiare , poi pokerino , poi in 6 s' è	divorato	il	9 chili di cocomero ( anche se nn mi piace
#126517710		fantasia e nell' orrore nel vedere una persona	divorare	il	i resti di un animale con piacere e indemoniata
#130157369		vita della crosta terrestre . I batteri che	divorano	il	il petrolio rendono inutilizzabili la maggior
#140386566		. S' accampano ai piedi di alcuni pini	divorando	il	radici e licheni per ingannare la fame
#163932620		per raggiungerle , così poi da ucciderli e	divorarli	il	. Nell' Odissea le Sirene sono legate a
#173602762		dalle carni significano allora cessare di	divorare	il	la carne del povero , lottare perché tutti
#176746065		Qualche prete , vedendolo sbrindellato e	divorato	il	dagli insetti , non ardiva dargli l' ostia
#177788496		. La sera compriamo delle provviste che	divoriamo	il	sul nostro terrazzino . Ed eccoci alla
#179514820		posture , gesti grotteschi dei cadaveri	divorati	il	sotto il sole da schiere di mosche " .
#201358385		sufficienti a sfamare la ragazza , ma lei le	divorò	il	in un baleno . " Le hai dato tutte le nostre
#219191349		delle mani , nessuno del cucchiaino . Chi	divorava	il	più in fretta riceveva di più ; il più
#219581855		guardano mentre quello preso in pescheria lo	divorano	il	!!! In bocca al lupo DeIorenziDanilo Inserito
#239376048		lei siamo degli spietati assassini , che	divorano	il	anatre , conigli , piccioni , rane , pesci
#281775601		... Guarda quel brutto giaguaro che sta	divorando	il	la testa della tua vittima . " " Prutta
#281775803		erano gettate anche sul giaguaro , e lo	divoravano	il	ancora agonizzante . Piene fino a scoppiare
#339203416		reticella che impedisca ai riproduttori di	divorare	il	le uova ; ci sono da tenere presente 2
#341664477		noi si sono estese a vista d' occhio : ora	divoravano	il	l' erba su di un circolo irregolare di

- Il set paradigmatico di parole che popola specifici slot argomentali nell'ambito dello stesso senso di un verbo raramente coincide perfettamente con il tipo "target" (richiesto da V).
- Mancate corrispondenze tra tipo "target" (assegnato da V) e tipo "source" (del Nome) nella stessa relazione grammaticale.
- In molti casi si osserva una regolare alternanza di tipo semantico nella stessa posizione argomentale.

# Type mismatches

- [[Vehicle]–subj] atterra  
L'aereo | **il pilota** | **il Papa** | **il volo** stava atterrando.
- [[Human]–subj] commenta, pensa, crede, annuncia ...  
Il Presidente | **il Quirinale** | **il governo** non ha commentato.  
Lo speaker | **l'altoparlante** | **il campanello** annunciava l'arrivo del treno.  
I negozianti | **I cartelli colorati** annunciano sconti.
- [[Human]–subj] sanguina  
L'uomo | **la mia mano** ha smesso di sanguinare.
- [[Human]–subj] divora [[Documento]-obj]  
divorare un romanzo | **un autore** | **Freud**.

- “Regular choices of types within an overall pattern are coded as type alternations. Common alternations in subject position are for instance [[Human|Institution]] and [[Human|Body Part]]”.
- “Non-canonical lexical items breaking a particular statistical threshold are “coerced” into ‘honorary’ membership of a semantic type in particular contexts.”

Alcuni dei potenziali candidati del tipo “target” non fanno parte del set lessicale.

[[Persona]] partecipa a [[Evento]]

- Set lessicale [[Evento]] = {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. . . }
- . . . ma non [[Evento]] = {temporale, suicidio. . . }

- La composizione dei set lessicali che popolano un nodo dell'ontologia (un tipo semantico) tendono a variare da verbo a verbo: alcune parole “entrano” e altre “escono” dal set, come predetto da Wittgenstein (*family resemblances*).

## lavare [[Body Part]-obj]

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

## amputare [[Body Part]-obj]

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

# Che cosa è un [[Documento]]?

## leggere

{opinione, commento, libro, avvertenza, recensione, giornale, cronaca, intervista, articolo, blog, messaggio, poesia, notizia, resoconto, racconto, fumetto, romanzo ... }

## pubblicare

{articolo, intervista, nota, avviso, bando, studio, libro, sondaggio, notizia, volume, saggio, testo, giornale, comunicato, monografia, rivista, documento, inesattezza, annuncio ... }

## spedire

{cartolina, pacco, e-mail, fax, messaggio, lettera, raccomandata, foto, telegramma, copia, sms, merce curriculum, pacchetto, invito, vaglia, posta, file, libro, coupon, modulo ... }

## tradurre

{testo, frase, bibbia, parola, brano, libro, poesia, vocabolo, concetto, opera, versetto, termine, idea, nome, spiegazione, romanzo, canzone, pensiero, vangelo, espressione ... }

- “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 - Jezek 2008.

- ... 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 ... >

# Tipi complessi

VERBI possono selezionare tipi “target” complessi.

NOMI possono essere associati a tipi “source” complessi.

[[Human]-subj] legge [[Phys•Info]-obj]

- [[Phys•Info]]: {libro, lettera ...}
- [[Info]]: {intervista, discorso, musica ...}
- [[Phys]]: {contatore (?)}

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

- [[Location]]: {riva del mare, vetta, località turistica, radura ...}
- [[Artefact]]: {cancello, macchina, semaforo ...}
- [[Artefact•Location]]: {casa, albergo, aeroporto ...}

- Costruzione di un repertorio di pattern verbali corpus-based intesi come distinzioni sintagmatiche motivate semanticamente.
- Costruzione di un'ontologia di nomi raggruppati in clusters in base alle posizioni argomentali, in grado di predire il significato di V in contesto.
- Costruzione di un repertorio di spostamenti metonimici.

- Utilizzo delle strutture argomentali “tipate” estratte da corpora per arricchire risorse lessicografiche esistenti e renderle più compatibili con applicazioni computazionali.
- Per es. distinzione tra sensi dei VERBI motivate empiricamente aumentano le probabilità di successo in WSD.
- Interazione delle strutture argomentali “tipate” con informazioni tematiche (Agent, Patient, Experiencer. . . ) – tipi e ruoli sono informazioni ortogonali.
- Raffinamento di ontologie di NOMI, in relazione alle strutture argomentali in cui i NOMI entrano.

- Baroni, M. and A. Kilgarriff 2006. “Large Linguistically-Processed Web Corpora for Multiple Languages”, *EACL 2006 Proceedings*, 87-90.
- Church, K.W. and P. Hanks 1990. “Word Association Norms, Mutual Information and Lexicography”. In *Computational Linguistics*, vol. 16, n. 1, 22-29.
- Hanks, P. and E. Jezek 2008. “Shimmering Lexical Sets”. In E. Bernal and J. De Cesaris (eds) *Proceedings of the XIII EURALEX International Congress*, Barcelona, Universidad Pompeu Fabra, 391-402.
- Hanks, P. and J. Pustejovsky 2005. “A Pattern Dictionary for Natural Language Processing”. In *Revue française de linguistique appliquée*, 10 (2).

- Kilgarriff A., P. Rychly, P. Smrz & D. Tugwell 2004. "The Sketch Engine". In G. Williams and S. Vessier (eds.). *Proceedings of the XI EURALEX International Congress* (July 6-10, 2004). Lorient. France. 105-11.
- Pustejovsky, J. 2006. "Type theory and lexical decomposition". In *Journal of Cognitive Science* 6, 39-76.
- Pustejovsky, J. and E. Jezek 2008. "Semantic Coercion in Language: Beyond Distributional Analysis". In *Italian Journal of Linguistics* 20:1, 181-214.