

# Validation of Corpus Pattern Analysis

## Assigning pattern numbers to random verb samples

Silvie Cinkova, Patrick Hanks

Pan Shiqi, Wang Yingyan, Donati Lorenzo  
Linguistica Applicata-a A.A. 2012/13

# Motivazioni

- I significati di molte parole sono ambigui
- Mentre i *PATTERN* dell'uso di una parola, raramente presentano ambiguità
- SCOPI:
  1. Identificare i *pattern* d'uso comune di ogni parola
  2. Associare significati a ogni *pattern*
- *Pattern Dictionary of English Verbs*

# Prototipi e pattern

- **Teoria dei prototipi:** sostiene che alcuni oggetti di una categoria sono più centrali (prototipici) di altri.
- **Teoria delle norme e delle eccezioni** (*Theory of norms and exploitations*): non si hanno solo usi prototipici delle parole (**norme**), ma si possono trovare anche espressioni che non rispettano un *pattern* prototipico, ma sono ricollegabili a un uso nella norma (**eccezioni**)

- Per ogni verbo vengono analizzate 250 occorrenze casuali prese a campione (*sample*)
- Se vengono classificati più di 25 *pattern* per una parola viene considerato un numero doppio di occorrenze.
- La maggior parte dei verbi ha un numero ristretto di *pattern*
- I *phrasal verbs* sono trattati come *pattern* diversi dello stesso verbo base

# Struttura di un *pattern*

Secondo la **grammatica sistemica**, si identificano degli **argomenti** (*clause roles*). Essi sono:

- **S-Subject**: ciò che è semanticamente soggetto, “Agente” nella grammatica delle dipendenze
- **P-Predicator**: il verbo e i suoi eventuali ausiliari
- **O-Object**: sia diretto (incluso il soggetto delle frasi passive) che indiretto
- **C-Complement**: sintagma coreferenziale o con il soggetto (es.: *He is happy; he is the President*) o con l’oggetto diretto (es.: *They elected him President; it made him happy*)
- **A-Adverbial**: normalmente un sintagma preposizionale o un avverbio (es.: *She drove to London*)

- L'assenza di un oggetto diretto può essere parte di un *pattern*: [NO OBJ]

P: [[Human]] attend [NO OBJ] {to [[Anything]]}

I: [[Human]] pays attention to and deals as necessary with [[Anything]]

- Mentre l'assenza di un sintagma avverbiale normalmente non è espressa, perché non influisce sul significato

# Alternanze

- Le alternanze sintattiche sono classificate in *pattern* separati
- Ma se in un campione si trovano solo una o due occorrenze di una alternanza, queste possono essere trattate come *exploitation*
- Le principali eccezioni sono:
  1. Le alternanze attivo/passivo
  2. Alternanze dell'oggetto indiretto
  3. *Reciprocal verbs*

1. Alternanze attivo/passivo: i verbi passivi sono normalmente inseriti nello stesso *pattern* della loro corrispondente forma attiva. Il predicato è inserito nella forma passiva solo quando un particolare uso del verbo è normalmente passivo
2. Alternanze dell'oggetto indiretto: con i *verbs of giving* in inglese c'è una regolare alternanza fra un costrutto con il doppio oggetto (es.: *He gave her an apple*) e un costrutto con oggetto e sintagma avverbiale (es.: *He gave an apple to her*)
3. *Reciprocal verbs*. Es.: *John met Alice; Alice met John; John and Alice met; John met with Alice; Alice met with John*

# Implicatura

- È una parafrasi di un *pattern*, usando un verbo e una fraseologia diversi.
- Ci può essere una implicatura secondaria che aggiunge informazioni complementari
- La combinazione di un *pattern* e della sua implicatura è detta **categoria** (*category*).

PATTERN: [[**Human** | **Action** | **Drug**]] alleviate {**pain** | **anxiety** | **illness** | ...}

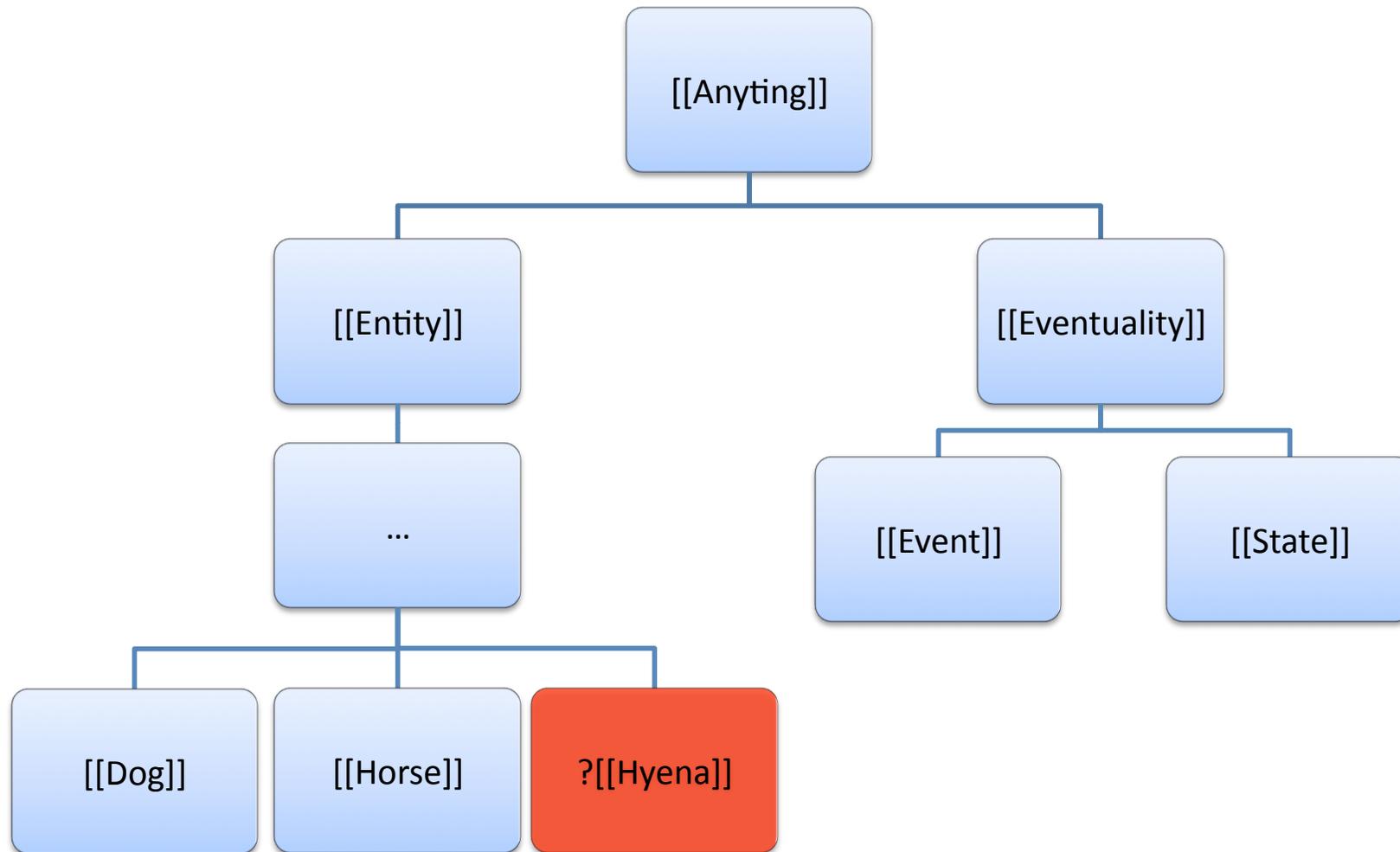
IMPLICATURE: [[{Human | Action | Drug}]] causes {pain | anxiety | illness | ...} to become less intense

SECONDARY IMPLICATURE: [[Human]] is typically a Health Professional

# *Semantic type*

- ***Collocations***: parole che occorrono regolarmente in un argomento in relazione a un particolare verbo
- Le *collocations* di un verbo in un argomento sono solitamente nomi che condividono aspetti base del significato → ***semantic type***
- Un *semantic type* rappresenta concetti cognitivi come *Human, Institution, Animal,...*
- I *semantic type* sono inseriti in un inventario finito (***shallow ontology***) strutturato gerarchicamente
- Nella revisione dei *pattern* possono essere aggiunti nuovi *semantic type*

# Gerarchia dei *semantic type*



Ogni *semantic type* possiede le proprietà del *semantic type* sovrastante

- Quando due o più argomenti hanno lo stesso *semantic type*, sono distinti da numeri

**PATTERN:** **[[Human 1 | Animal 1 | Institution 1 | Document]] signal ([[Human 2 | Animal 2 | Institution 2]]) [[Information | Eventuality]]**

**IMPLICATURE:** **[[Human 1 | Animal 2 | Institution 1 | Document]] communicates [[Information]] about [[Eventuality]] (to [[Human 2 | Animal 2 | Institution 2]]) by gestures, language, or other means.**

- Un argomento può essere costituito da più *semantic type*

# Set lessicali

- **Set lessicale:** elenco delle *collocations* di un argomento, non riconducibili a un *semantic type*
- È rappresentato tra parentesi graffe
- In un set lessicale sono indicate solo due o tre parole, se ce ne son di più vengono indicate con i punti di sospensione

PATTERN: **[[Human]] take {risk | chance | ...}**

IMPLICATURE: **[[Human]]** does something that may have bad consequences, in the hope of obtaining some **[[Benefit]]**

# Set lessicali

- Un Set lessicale può consistere di una sola parola

PATTERN: **[[Human]] take responsibility for [[Anything]]**

IMPLICATURE: **[[Human]]** accepts the duty of doing whatever is necessary to ensure that **[[Anything]]** is OK

SECONDARY IMPLICATURE: If **[[Anything]]** turns out to be bad or not OK, **[[Human]]** may be fired or otherwise punished

- Un set lessicale può alternarsi con un *semantic type*

PATTERN: **[[Human | Animal]] bleed [NO OBJ] {from [[Body Part]] | from {wound | injury | laceration}}**

IMPLICATURE: **[[Human | Animal]]** loses blood from a wound or injured **[[Body Part]]**

# Ruolo semantico

- Un *semantic type* è talvolta complementato da un **ruolo semantico**
- Il *semantic type* rappresenta le proprietà formali dell'argomento, intrinseche al nome
- Il ruolo semantico esprime proprietà estrinseche

PATTERN: [[Human 1 = Judge]] sentence [[Human 2 = Criminal]] to [[Event = Punishment]]

*Mr. Woods sentenced Bailey to 3 years.*

*The judge sentenced the old villain to a term of imprisonment.*

- **Argomenti verbali:** componenti verbali come infinitive, that- clauses, wh- clauses etc.

[[Human | Document | Institution]] acknowledge {that-CLAUSE}

- **Argomenti opzionali:** l'omissione di questi argomenti non influisce sull'implicature. Sono inclusi tra parentesi tonde

# Ambiguità dei *semantic type*

- A volte è possibile scegliere quale *semantic type* assegnare e questo influisce sull'implicatura
- Es.: *The AAA launched their education programs.*  
→ [[Event]]/[[Product]]?

*Pattern 1: begin or initiate an endeavour*

*Pattern 2: begin to produce or distribute; start a company*

- In questi casi, si assegna il *semantic type* che implica il *pattern* più frequente.

**The task: assigning concordance  
lines to patterns**

# *Introduzione*

- Si assortiscono concordanze assegnando i numeri dei pattern come i tag sulla base della somiglianza percepita del significato dei pattern.
- I tag sono assegnati
  - utilizzando un campione casuale di concordanze.
  - utilizzando concordanze pre-ordinate da Sketch Engine.

Ogni linea in un campione selezionato a caso deve essere:

1. assegnata a un pattern
2. dichiarata di essere un “tag error” o “unclassifiable”

Il campione casuale è la base per le frequenze comparative, non tutte le linee etichettate di un corpus.

# ***Procedura***

## **reference random sample (RRS) :**

il campione casuale di riferimento

RRS sarà diviso in due parti:

---S1 (200 concordanze)

per l'orientamento

--- S2 (50 concordanze)

senza tag

- Il compito del validatore :

---- etichettare ogni linea di concordanza in S2 con il numero di pattern appropriato.

## Per i disaccordi

1. I pattern verranno modificati
2. un nuovo campione di 50-concordanze sarà generato per i validatori da etichettare secondo i pattern modificati
3. SC modificherà il RRS originale secondo i pattern migliorati.

# ***Norm-conformant concordances***

--Una norma-conforme linea di concordanza riceve il numero di pattern corrispondente.

Una linea di concordanza è norma-conforme quando:

- 1) ha l'implicatura uguale al pattern;
- 2) ha lo stesso numero di argomenti come il pattern presenta. (eccezione: ellissi grammaticale)

- 3) gli argomenti preposizionali corrispondono alle preposizioni nel pattern
- 4) i sostantivi si conformano ai tipi semantici elencati  
----Eccezione, e.g. [[Drink]] => [[Container]] of a [[Drink]]
- 5) gli argomenti hanno lo stesso ruolo semantico
- 6) i non-sostantivo argomenti corrispondono alla descrizione della forma definita dal pattern (ad esempio, that-CLAUSE)

# Grammatical ellipsis

- In alcuni casi, il soggetto del verbo è mancante sotto osservazione per ragioni grammaticali.
- Ellissi grammaticale di soggetto in imperative e control verb, i quali sono chiamati “phased predicators” nella grammatica sistematica, è ancora considerata come l’uso normo-conforme.

# Esempi

- Imperative (2nd person):

----***Speak** to your friend!*

- Control Verb :

----*Peter decided to **speak**.*

----Questi casi sono ancora considerati come norma-conforme linea di concordanza.

# Control verb:

È un verbo con 2 argomenti

----un complemento sostantivale

---- un complemento verbale.

Il complemento sostantivale deve essere l'argomento sia del control verb che il complemento verbale.

[back](#)

# Contextual clues, contextual ellipsis

I suggerimenti contestuali possono essere utilizzati:

- 1) per risolvere l'ellissi contestuale
- 2) per specificare il dominio o registrare e disambiguare tipi/ ruoli semantici

# Esempi

*Mr. Mao (un politico) abstained again*

abstain:

1. refrain from voting
2. choose no to consume

----e.g. "I abstain from alcohol"

----con il corretto pattern *abstain dà vote.*

# Esempi

- “ *I will ride and you will follow me*”

*----ride a bicycle*

*----ride a horse*

- **Pronouns as arguments**
- **that-clauses**
- **Direct speech and indirect speech**

# that-clauses

- La congiunzione subordinativa “that” che introduce una clause di sostantivo è spesso omessa in inglese.

- Esempi:

*---I told them that they should stay.*

*---I told them they should stay.*

# Direct speech and indirect speech

Con [i reporting verbs](#), il discorso diretto è etichettato come [QUOTE], anche se le virgolette non sono presenti.

- *“I won’t do it”, said Mr. Smith, “unless you accept my conditions.”* – [QUOTE]
- *Mr. Smith said, “I won’t do it unless you accept my conditions.”* – [QUOTE]
- *“I won’t do it unless you accept my conditions,” said Mr. Smith.* – [QUOTE]
- *I won’t do it, Mr. Smith said, unless you accept my conditions.* – [QUOTE]

# reporting verbs

verbi che descrivono o segnalano un'attività o un'azione che ha già avuto luogo e sono associati a specifiche preposizioni o vengono usati con il pronome relativo “that”.

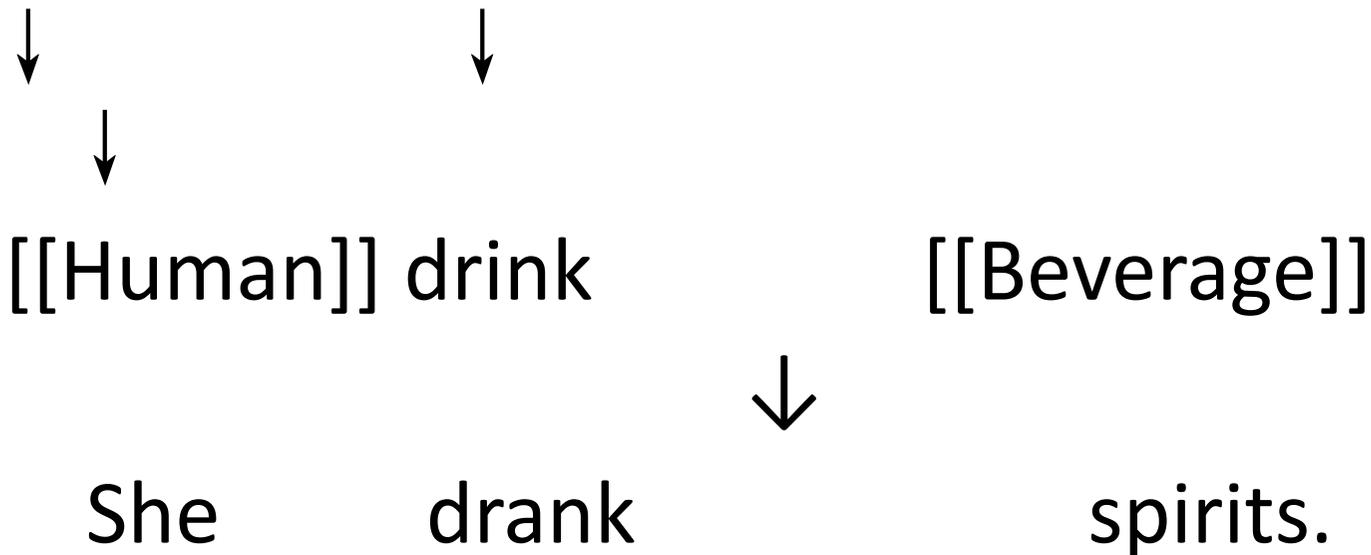
e.g. claim, demand, deny ecc

[back](#)

- *Mr. Smith says he won't do it unless they accept his conditions. – that-CLAUSE*
- *He won't do it, Mr. Smith says, unless they accept his conditions. – that-CLAUSE*
- *He wouldn't do it, Mr. Smith said, unless they accepted his conditions. – that-CLAUSE*
- *He won't do it unless they accept his conditions, Mr. Smith said. – that-CLAUSE*

# Semantic type coercion between a verb argument and its modifier

- Drink a [[Beverage]]
- Drink a **cup**, a **glass**, a **bottle** of a beverage
- [[**Container**]]+ “drink” → **quantitatives**. *She drank 8 glasses of spirits.*



# Semantic type coercion between a verb argument and its modifier

- Partitivo:
- “slice”

*He ate four slices of toast.*

↓ ↓

[[Human]] eat

↓

[[Food]]

# Semantic type coercion between a verb argument and its modifier

- *She drank 8 glasses.*

*She drank 8 pints.*

*He ate four slices.*

→ “ellipsis”, “coercion”

→ “verb coercion”:N.

{ predicato implicato  
predicato presente

# Semantic type coercion between a verb argument and its modifier

- “ellissi”:
  - \*non è un tipo di “exploitations”
  - \*conoscenza comune
    - (“*glass*”, “*pints*” -----[[Beverage]]
    - “*slice*”-----[[Food]])

→ “Pattern-comformant”

# Semantic type coercion between a verb argument and its modifier

- “coercion”:

*the 28-year-old stockbroker was riding his first ever winner*

\* “winner”=[[Horse]]

(domain-specific, horse-riding)

→\*non è un tipo di “exploitations”

\*estensione del semantic type [[Horse]]

→polisemia----diversi “semantic types”

# Semantic Type Coercion between a verb and its argument

- “coercion”:

*He **enjoyed** his soup.*

----- (default) eating the soup

→ He **did something** to the soup **with enjoyment**.

# Semantic Type Coercion between a verb and its argument

- I **enjoyed** the soup.

=(default) I **ate** the soup **with enjoyment**.

→ I **enjoyed eating** the soup.

=(da un cuoco) I **cooked** the soup **with enjoyment**.

→ I **enjoyed cooking** the soup.

# Semantic Type Coercion between a verb and its argument

- *I **enjoyed** the performance.*

=(da un spettatore) I **watched** the performance  
**with enjoyment.**

→ I **enjoyed watching** the performance.

=(da un attore o un'attrice) I **made** the performance  
**with enjoyment.**

→ I **enjoyed making** the performance.

# Semantic Type Coercion between a verb and its argument

- “coercion” nel verbo



.C.

# Exploitations

- -----".**c.**" (Coercion in verb)
- -----".**a.**" (Anomalous argument)
- -----".**f.**" (Figurative uses)
- -----".**s.**" (Unusual syntax)

# Exploitations

## (.a.)

- “.a.” → “anomalous argument” o “honorary member”

\* “honorary member”:

\* un N. come membro anormale di un lexical set;

\* non è adattato a nessun semantic type nel pattern

# Exploitations (.a.)

- *Rashid Solh, the prime minister, spoke darkly of the arrival in Lebanon of several hundred Israeli agents provocateurs whose mission was to destroy the republic. **The plot had arrived at Beirut.***

→ [[Human | Vehicle | Animal]] arrive  
[ NO OBJ ]{at[[Location]]}

\*non è una metafora, ma “.a.”

# Exploitations (.f.)

- “.f.” → “figurative uses”, non-conventional metaphor

\*metafora convenzionale → pattern normale

*Jane and I quickly **arrived** at joint decisions about the project.*

→ [[Human | Institution]] arrive [NO OBJ]  
{at[[Concept= Considered Opinion]]}

→ convenzionale → un pattern normale

# Exploitations (.f.)

- Non-conventional metaphor:

*Nobody **arrives** at ICI board level without some steel and determination in his character.*

→ ".f." figurative

\* "board level" ≠ [[Location]]

# Exploitations (.s.)

- “.s.” → “unusual syntax”:
  - \*exploitation of phraseological patterns;
  - \*exploitation of grammatical constructions

# Exploitations (.s.)

- Exploitation of phraseological patterns:

\*omissione di parola → il “focus”

*We **punish** too much---and in particular, we **imprison** too much.*

----(BNC)J. Dignan, 1992. *The Penal System*

\*V.tr → V.intr.

- \*esempi di “syntactic exploitations” (l.s.)

# Exploitations (.s.)

- Exploitation of grammatical constructions:

\*si tratta di costruzione completa

*"I would also like to apologize about losing you and Ema the house."*

---Marian Keyes, The Other Side of the Story, p.632

\*non è un uso normale di "lose"

\* "lose"                      "give" → si cambia la costruzione  
grammaticale 

# Tagging errors and other noise (.x.)

- Il verbo relativo e gli altri (noise)
- errori: misprints in the corpus;  
tagging errors in the Sketch Engine grammar  
→ verbo--- uso come n. e agg.

# Tagging errors and other noise (.x.)

- Noise:

---- “Now let us look at idioms such as ‘grasp the nettle’.”

\* “grasp” → noise → “.x.”

--- “Frankie *goes* to Hollywood”

\* nome di una banda → “.x.”

# Tagging errors and other noise (.x.)

- Participio presente o passato di un verbo:
  - \*si interpreta come un' occorrenza del verbo:
    - l'occorrenza della forma di participio
    - presente pattern del verbo

# Tagging errors and other noise (.x.)

- Participio presente o passato di un verbo:  
\* si considera come n. o agg. → “.x.”:

*1) Which/ some/no/any/ reading;*  
*There was (no) ---ing*

→ determinante + “---ing” (participio presente)

↓  
n.

# Tagging errors and other noise (.x.)

- 2) *I hate him **coming** late.* ①  
*I hate his **coming** late.* ②

\*il primo → v.

\* il secondo → n.



possessivo determinante + “---ing”

(participio presente)

# Tagging errors and other noise (.x.)

- 3) \* “---ing”(participio presente)+ oggetto diretto  
→ v.

es. *building bridges*

- \* “---ing”(participio presente)+ oggetto diretto

↑

preposizione(*of*)

→n.

es. *building of bridges*

# Tagging errors and other noise (.X.)

- 4) participio (attributivo) + n.



agg.

*es. surprising evidence, an interesting article,  
a wicked guy...*

# Unclassifiables (.u.)

- Non è adattato a nessun pattern normale



.u.

- \* Nei casi con gran incertezza → .u.

# Noun uses that don't fit

- contiene una parola non adatta a nessun semantic type nel pattern:

---- “.a.” (anomalous argument)

---- “.u.” (unclassifiable)

---- “.c.” (coercion)