
Panoramica sul senso del **T**rattamento del **L**inguaggio **N**aturale

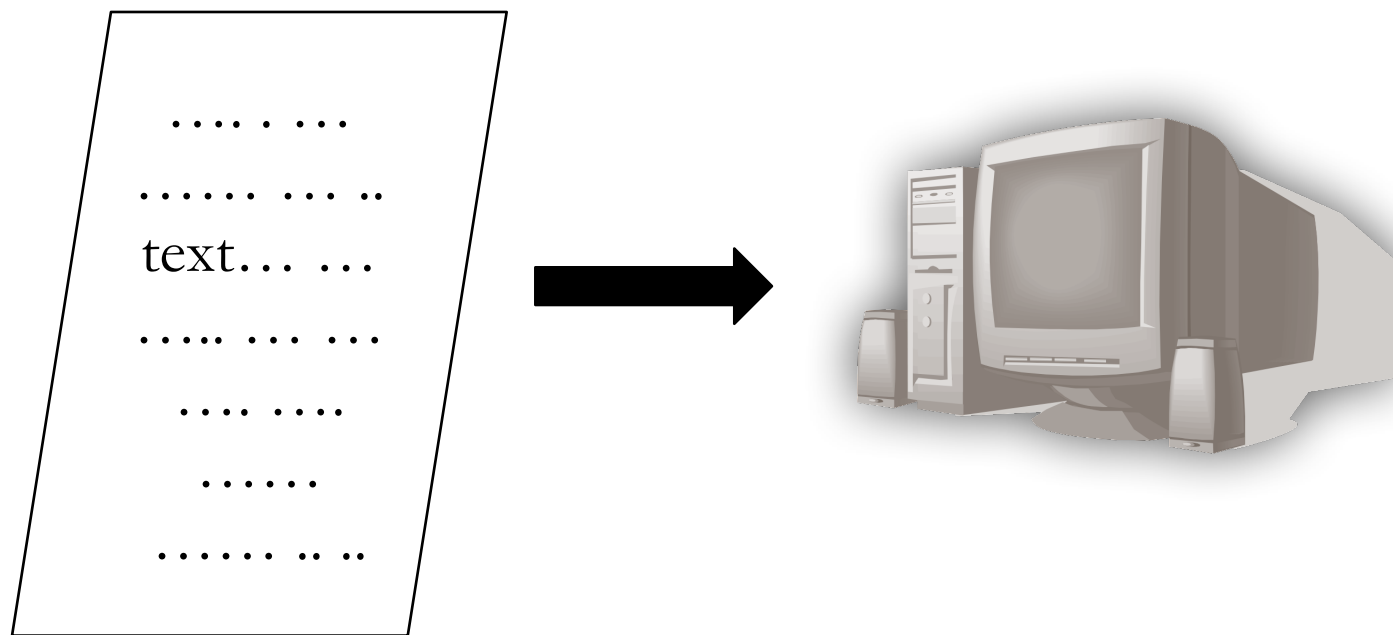
Luigi Di Caro

Panoramica veloce sul senso del TLN

- Cosa significa *comprendere* il linguaggio naturale
- Bisogna risalire alla problema originario del TLN,
 - ovvero la comunicazione tra macchina e uomo
(*one of the dreams of AI*)

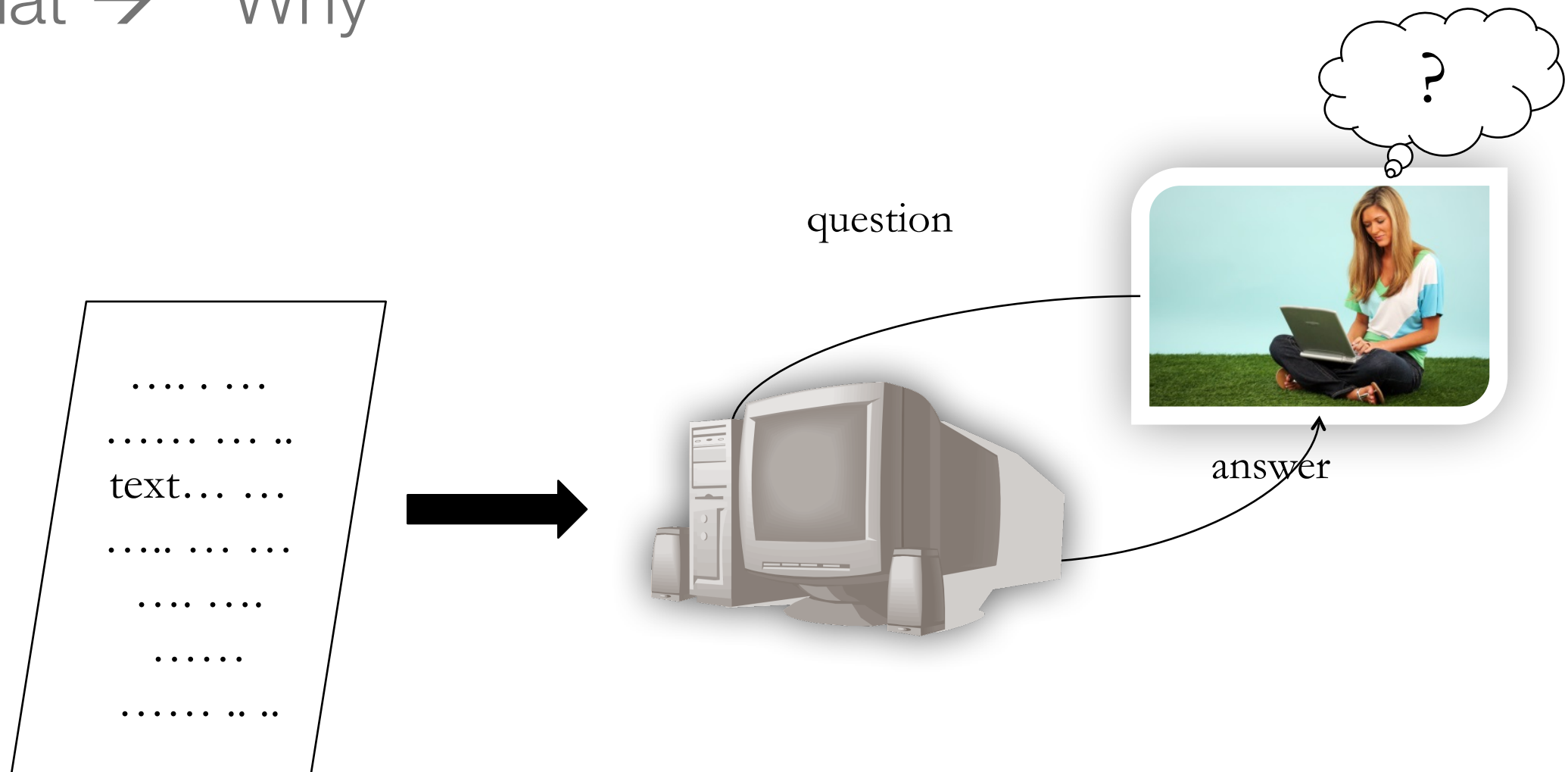
Panoramica veloce sul senso del TLN

- What



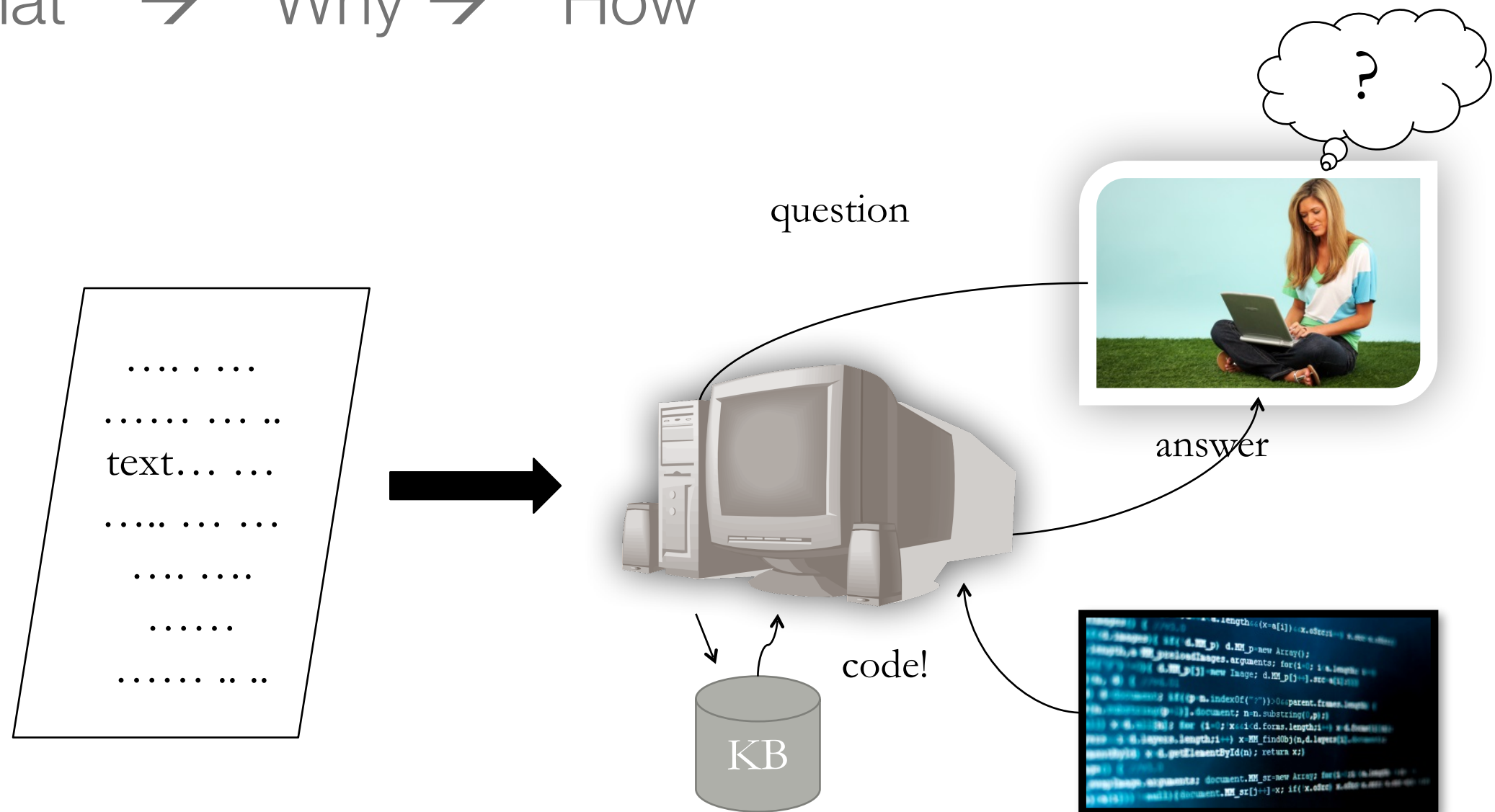
Panoramica veloce sul senso del TLN

- What → Why



Panoramica veloce sul senso del TLN

- What → Why → How



Panoramica veloce sul senso del TLN

- Primo esercizio
 - Pensare a possibili tipi di domande

Panoramica veloce sul senso del TLN

- Primo esercizio
 - Pensare a possibili tipi di domande
 - *fact, list, definition, how, why, hypothetical, semantically constrained, cross-lingual questions, who, where, how many, how much, when, yes/no, ...*

Panoramica veloce sul senso del TLN

what		64	54			
	basic what	40	34	MONEY/NUMBER/DEFINITION/TITLE/NNP/UNDEFINED	<i>What was the monetary value of the Nobel Peace Prize in 1989?</i>	monetary value
	what-who	7	7	PERSON/ORGANIZATION	<i>What costume designer decided that Michael Jackson should only wear one glove?</i>	costume designer
	what-when	3	2	DATE	<i>In what year did Ireland elect its first woman president?</i>	year
	what-where	14	12	LOCATION	<i>What is the capital of Uruguay?</i>	capital
who		47	37	PERSON/ORGANIZATION	<i>Who is the author of the book "The Iron Lady: A Biography of Margaret Thatcher"?</i>	author
how		31	21			
	basic how	1	0	MANER	<i>How did Socrates die?</i>	Socrates
	how-many	18	13	NUMBER	<i>How many people died when the Estonia sank in 1994?</i>	people
	how-long	2	2	TIME/DISTANCE	<i>How long does it take to travel from Tokyo to Niigata?</i>	—
	how-much	3	2	MONEY/PRICE	<i>How much did Mercury spend on advertising in 1993?</i>	Mercury
	how-much-<modifier>	1	0	UNDEFINED	<i>How much stronger is the new vitreous carbon material invented by the Tokyo Institute of Technology compared with the material made from cellulose?</i>	new vitreous carbon material
	how-far	1	1	DISTANCE	<i>How far is Yaroslavl from Moscow?</i>	Yaroslavl
	how-tall	3	3	NUMBER	<i>How tall is Mt. Everest?</i>	Mt. Everest
	how-rich	1	0	UNDEFINED	<i>How rich is Bill Gates?</i>	Bill Gates
	how-large	1	0	NUMBER	<i>How large is the Arctic refuge to preserve unique wildlife and wilderness value on Alaska's north coast?</i>	Arctic refuge
where		22	16	LOCATION	<i>Where is Taj Mahal?</i>	Taj Mahal
when		19	13	DATE	<i>When did the Jurassic Period end?</i>	Jurassic Period
which		10	8			
	which-who	1	1	PERSON	<i>Which former Klu Klux Klan member won an elected office in the U.S.?</i>	former Klu Klux Klan member
	which-where	4	3	LOCATION	<i>Which city has the oldest relationship as sister-city with Los Angeles?</i>	city
	which-when	1	1	DATE	<i>In which year was New Zealand excluded from the ANZUS alliance?</i>	year
	which-what	4	3	NNP/ORGANIZATION	<i>Which Japanese car maker had its biggest percentage of sale in the domestic market?</i>	Japanese car maker
name		4	4			
	name-who	2	2	PERSON/ORGANIZATION	<i>Name the designer of the show that spawned millions of plastic imitations, known as "jellies"?</i>	designer
	name-where	1	1	LOCATION	<i>Name a country that is developing a magnetic levitation railway system?</i>	country
	name-what	1	1	TITLE/NNP	<i>Name a film that has won the Golden Bear in the Berlin Film Festival?</i>	film
why		2	0	REASON	<i>Why did David Koresh ask for a word processor?</i>	David Koresh

Panoramica veloce sul senso del TLN

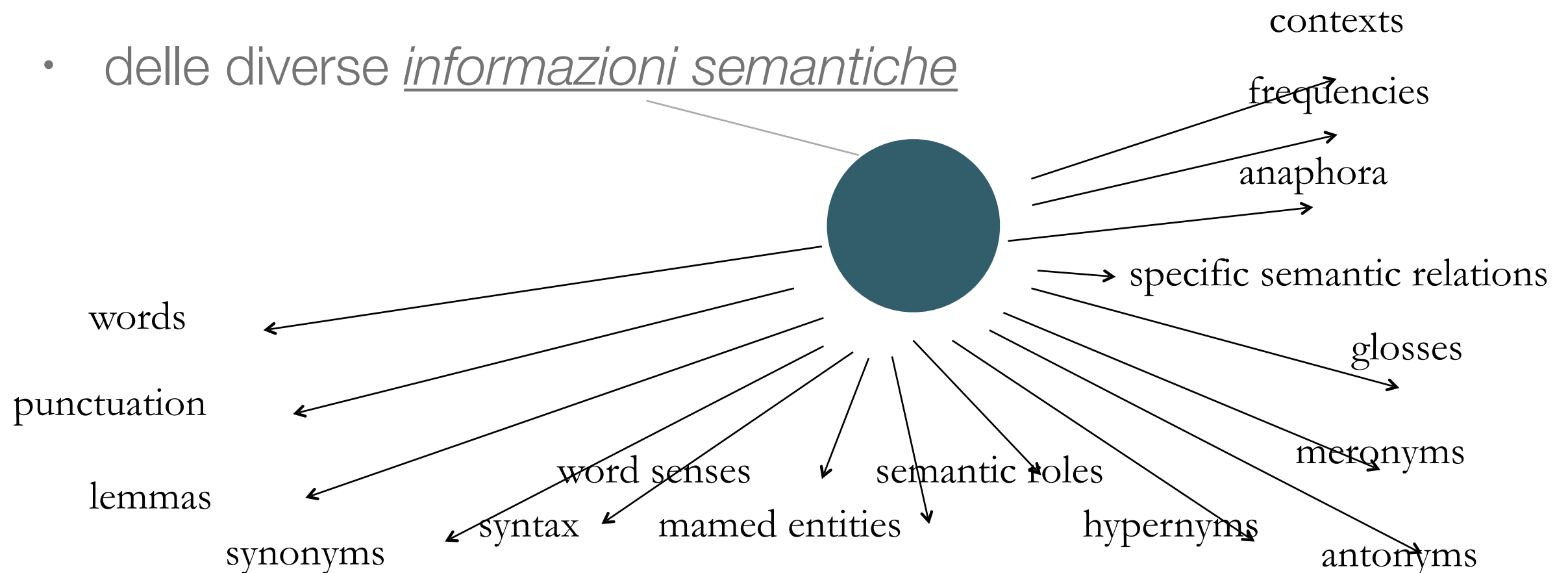
	what-where	14	12	LOCATION
who		47	37	PERSON/ ORGANIZATION
how		31	21	
	basic how	1	0	MANER
	how-many	18	13	NUMBER
	how-long	2	2	TIME/DISTANCE
	how-much	3	2	MONEY/PRICE
	how-much- <modifier>	1	0	UNDEFINED
	how-far	1	1	DISTANCE
	how-tall	3	3	NUMBER
	how-rich	1	0	UNDEFINED
	how-large	1	0	NUMBER
where		22	16	LOCATION
when		19	13	DATE
which		10	8	
	which-who	1	1	PERSON

Panoramica veloce sul senso del TLN

- Il punto è che...
 - ...per ogni domanda, è necessario una specifica analisi, e
 - delle diverse informazioni semantiche

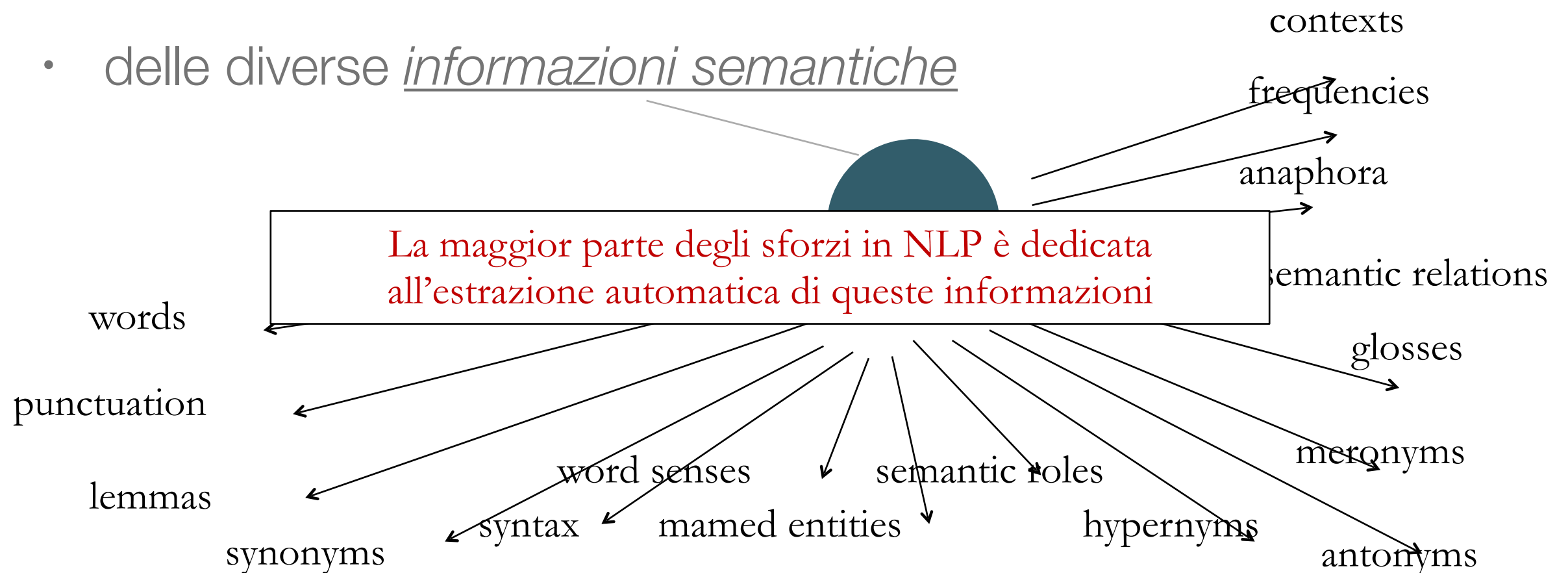
Panoramica veloce sul senso del TLN

- Il punto è che...
- ...per ogni domanda, è necessario una specifica analisi, e
- delle diverse informazioni semantiche



Panoramica veloce sul senso del TLN

- Il punto è che...
- ...per ogni domanda, è necessario una specifica analisi, e
- delle diverse informazioni semantiche



Panoramica veloce sul senso del TLN

- Esempio: hypernym extraction
 - Algoritmo:
 - Avendo a disposizione una serie di pattern, ad es.
 - X is a Y
 - X is a kind of Y
 - X such as Y
 - ...
 - Trovare un match tra frasi in input e tali pattern, ritornando X e Y

Panoramica veloce sul senso del TLN

- Esempio 2: word sense disambiguation

- Algoritmo:

- Avendo a disposizione una lista di definizioni di sensi

- Play.1 – participate in games or sport

- Play.2 – play on an instrument

- ...

- Disambiguare parole in frasi in input trovando un match diretto tra il contesto e le definizioni dei sensi

- Yesterday I ***played*** my favourite *instrument*, which is the guitar

Panoramica veloce sul senso del TLN

- Esempio 2: word sense disambiguation

- Algoritmo:

- Avendo a disposizione una lista di definizioni di sensi

- Play.1 – participate in games or sport

- Play.2 – play on an instrument

- ...

- Disambiguare parole in frasi in input trovando un match diretto tra il contesto e le definizioni dei sensi

- Yesterday I *played* my favourite instrument, which is the guitar.

- Yesterday I *played* the guitar.

Panoramica veloce sul senso del TLN

- Esempio 3: prompter (suggeritore automatico)
 - Algoritmo:
 - Avendo a disposizione un corpus di frasi in input
 - Calcolare le probabilità di avere una parola X, data una parola Y, per ogni parola
 - Ogni volta che un utente digita X, ritornare la parola con la massima probabilità

Panoramica veloce sul senso del TLN

- Esempio 3: prompter (suggeritore automatico)
 - Algoritmo:
 - Avendo a disposizione un corpus di frasi in input
 - Calcolare le probabilit' di avere una parola X, data una parola Y, per ogni parola
 - Ogni volta che un utente digita X, ritornare la parola con la massima probabilità
 - come integrare la history personale nel prompter?

Panoramica su semantica computazionale

- Semantica
 - lessicale,
 - formale,
 - statistica,
 - linguistico-distribuzionale

Significato del “significato”

- Word Sense Disambiguation
 - Problemi:
 - *specificità*
 - *copertura*
 - *soggettività*
 - Esempi
 - *WordNet*

Significato del “significato”

- Word Sense Disambiguation

- Problemi:

- *specificità*

- *copertura*

- *soggettività*

- Esempi

- *WordNet*



approfondimento

Palmer, M., Dang, H. T., and Fellbaum, C. (2007). Making fine-grained and coarse-grained sense distinctions, both manually and automatically. *Nat.Lan.Eng.*, 13(02):137–163.

Significato del “significato”

- Word Sense Disambiguation
- Word Sense Induction
- Inventory / No inventory
- human-based / data-based
- grammar-based / usage-based
- Evaluation: semplice ma criticabile / più complicato

Significato del “significato”

- Word Sense Disambiguation
- **Word Sense Induction**
 - Inventory / No inventory
 - human-based / data-based
 - grammar-based / use-based
 - **Evaluation:** semplice ma criticabile / più complicato

?

Significato del “significato”

- Word Sense Disambiguation
- Word Sense Induction
 - Metodo della *pseudo-word*
 - *Merging -> concatenazione di parole (random)*
 - *clustering -> applicazione del meccanismo WSI per l'identificazione di clusters (i sensi identificati automaticamente)*
 - *cluster-to-class evaluation -> confronta i cluster ottenuti automaticamente con le parole originali delle pseudo-word*

Comprensione del linguaggio naturale

- Digressione su Lexical Semantics e tipi di risorse
 - Dizionari elettronici: WordNet, BabelNet, ...
 - Risorse linguistico-cognitive: Property norms
 - Common-sense knowledge: ConceptNet
 - Visual Attributes <http://groups.inf.ed.ac.uk/calvin/Publications/silberer13ACL.pdf>
 - Word (o sense) embeddings (Word2Vec, Glove, ...)
 - Corpus managers (ad es. SketchEngine)

Comprensione del linguaggio naturale

- **Property norms**
 - Catturano informazioni simili a CSK, ma con alcune differenze:
 - Percezione
 - Immediatezza

Semantic feature production norms for a large set of living and nonliving things
<https://link.springer.com/article/10.3758/BF03192726>

The Centre for Speech, Language and the Brain (CSLB) concept property norms
<https://link.springer.com/article/10.3758/s13428-013-0420-4>

Comprensione del linguaggio naturale

- Property norms
 - ...altre caratteristiche:
 - Mischiano attributi e valori
 - Contengono variabilità linguistiche -> sparsità
 - Creazione costosa
 - Informazioni CSK ad alto valore semantico in termini di similarità percepita

Table 1 A subset of the preprocessed features for the concept *turtle* (only a sample of uncollapsed features is shown)

PF	Relation	Feature	Participant list
23	has	a shell	p15 17 18 24 28 30 39 45 48 50 52 55 56 58 59 60 61 63 64 88 132 133 135
18	does	swim	p15 18 28 30 45 48 52 55 56 58 59 60 62 88 113 131 131 133
16	does	lay	p15 17 24 28 39 55 56 59 59 60 62 87 88 113 132 133
14	does	live	p18 24 30 45 45 52 52 52 55 59 60 62 64 133
10	is	a reptile	p18 45 50 56 58 60 64 113 132 134
10	is	an animal	p28 39 45 48 50 53 132 133 134 135
9	does	lay egg+s	p15 24 60 62 87 88 113 132 133
8	is	green	p24 30 45 48 55 59 60 134
7	is	slow	p24 28 48 50 57 61 64
5	has	four leg+s	p52 55 62 88 133
5	does	have	p62 87 87 87 87
5	does	swim in sea+s	p28 56 59 131 133
4	is	endanger+ed	p39 45 55 58
4	has	a tail	p39 63 64 88
4	has	flipper+s	p60 113 131 134
4	has	skin	p24 28 58 63
3	does	eat	p52 60 88
3	does	live in sea+s	p18 59 60
3	does	live in water	p24 52 62
3	does	move	p17 52 62
3	has	a beak	p18 24 45
2	does	crawl	p52 60
2	has	small head	p30 57
2	does	lay egg+s on the beach	p28 39
2	has	scaly skin	p24 58

Comprensione del linguaggio naturale

- **Visual Attributes**

- Catturano aspetti “osservabili” di oggetti concreti
- ...organizzati in una tassonomia
- Annotazione di più di 500 concetti legati a circa 688K immagini in ImageNet.

- Separazione attributo - valore
- Contiene variabilità linguistica
- Solo aspetti visivi e per concetti concreti

Attribute Categories		Example Attributes
color_patterns	(25)	is_red, has_stripes
diet	(35)	eats_nuts, eats_grass
shape_size	(16)	is_small, is_chubby
parts	(125)	has_legs, has_wheels
botany;anatomy	(25;78)	has_seeds, has_fur
behavior (in)animate	(55)	flies, waddles, pecks
texture_material	(36)	made_of_metal, is_shiny
structure	(3)	2_pieces, has_pleats

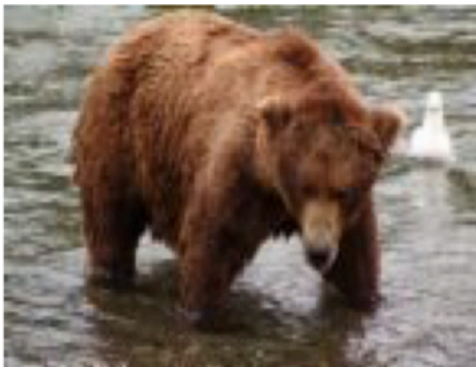
-

Models of Semantic Representation with Visual Attributes

<http://groups.inf.ed.ac.uk/calvin/Publications/silberer13ACL.pdf>

Comprensione del linguaggio naturale

- Visual Attributes



behavior	eats, walks, climbs, swims, runs
diet	drinks_water, eats_anything
shape_size	is_tall, is_large
anatomy	has_mouth, has_head, has_nose, has_tail, has_claws, has_jaws, has_neck, has_snout, has_feet, has_tongue
color_patterns	is_black, is_brown, is_white



botany	has_skin, has_seeds, has_stem, has_leaves, has_pulp
color_patterns	purple, white, green, has_green_top
shape_size	is_oval, is_long
texture_material	is_shiny



behavior	rolls
parts	has_step_through_frame, has_fork, has_2_wheels, has_chain, has_pedals has_gears, has_handlebar, has_bell, has_breaks has_seat, has_spokes
texture_material	made_of_metal
color_patterns	different_colors, is_black, is_red, is_grey, is_silver

Models of Semantic Representation with Visual Attributes

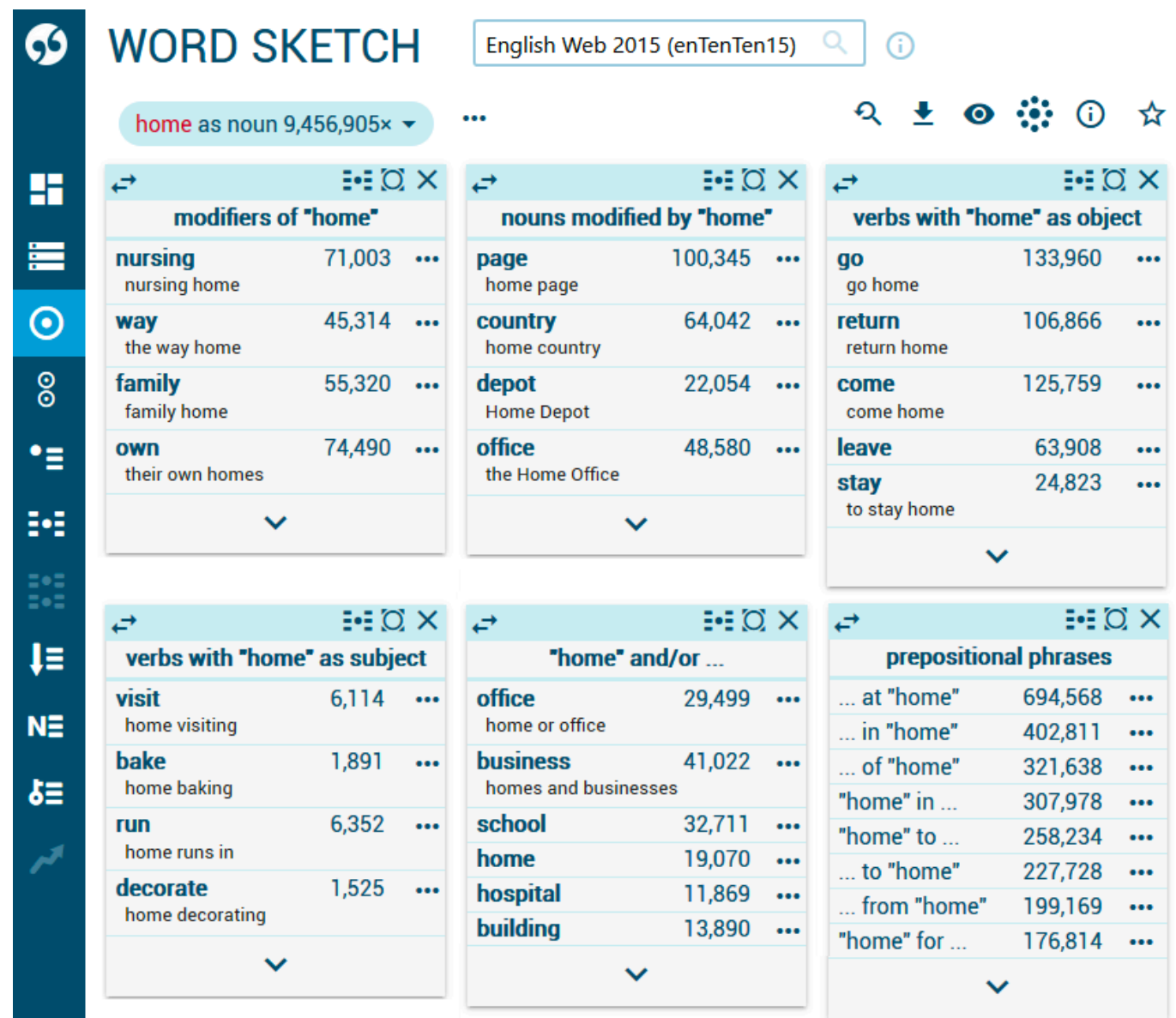
<http://groups.inf.ed.ac.uk/calvin/Publications/silberer13ACL.pdf>

Comprensione del linguaggio naturale

- SketchEngine




- Corpus manager
- Funzioni di I/O
- Aggregazione statistica su base linguistica
- Proprietà di ricerca, selezione, trasformazione

The screenshot shows the 'WORD SKETCH' interface for the word 'home' in the 'English Web 2015 (enTenTen15)' corpus. The interface has a dark blue sidebar with navigation icons. The main area displays six panels of linguistic data for 'home' as a noun.

WORD SKETCH		
English Web 2015 (enTenTen15)		
home as noun 9,456,905x		
modifiers of "home"		
nursing	71,003	...
nursing home		
way	45,314	...
the way home		
family	55,320	...
family home		
own	74,490	...
their own homes		
▼		
nouns modified by "home"		
page	100,345	...
home page		
country	64,042	...
home country		
depot	22,054	...
Home Depot		
office	48,580	...
the Home Office		
▼		
verbs with "home" as subject		
visit	6,114	...
home visiting		
bake	1,891	...
home baking		
run	6,352	...
home runs in		
decorate	1,525	...
home decorating		
▼		
"home" and/or ...		
office	29,499	...
home or office		
business	41,022	...
homes and businesses		
school	32,711	...
home	19,070	...
hospital	11,869	...
building	13,890	...
▼		
verbs with "home" as object		
go	133,960	...
go home		
return	106,866	...
return home		
come	125,759	...
come home		
leave	63,908	...
stay	24,823	...
to stay home		
▼		
prepositional phrases		
... at "home"	694,568	...
... in "home"	402,811	...
... of "home"	321,638	...
"home" in ...	307,978	...
"home" to ...	258,234	...
... to "home"	227,728	...
... from "home"	199,169	...
"home" for ...	176,814	...
▼		

Comprensione del linguaggio naturale

- SketchEngine 
 - Word sketch difference (differenze tra termini su ogni categoria)
 - Distributional Thesaurus (termini simili, con contesti simili, etc.)
 - Concordance search (esempi di occorrenza per forme anche complesse)
 - Word lists (statistiche su uso di un termine)
 - N-grams (freq. ngrams)
 - Keyword (parole chiave per multiwords)

Comprensione del linguaggio naturale

- Digressione su Lexical Semantics
 - Dizionari elettronici
 - Property norms
 - Common-sense knowledge
 - Visual Attributes
 - Word (o sense) embeddings

Potere espressivo	Scalabilità computazionale	Sorgente	Ambiguità, soggettività

Comprensione del linguaggio naturale

- Digressione su Lexical Semantics attraverso dizionari elettronici
 - Come si descrive un concetto?
 - Quali caratteristiche più importanti?
 - Che relazione con un gruppo semantico più generale?
 - E se il gruppo semantico fosse sconosciuto all'interlocutore?
 - Come si scrive una definizione? Come si valuta la qualità di una definizione? Quanto si è d'accordo?

Comprensione del linguaggio naturale

- Digressione su Lexical Semantics
 - Come si scrive una definizione? Come si valuta la qualità di una definizione? Quanto si è d'accordo?
- Lab

Comprensione del linguaggio naturale

- Esercizio di laboratorio “*Defs*”
 - Creazione di definizioni per 4 concetti:
 - 2 concreti (1 generico, 1 specifico)
 - person, brick
 - 2 astratti (1 generico, 1 specifico)
 - emotion, revenge
 - Breve analisi condivisa
 - Assegnazione task

Comprensione del linguaggio naturale

- Esercizio di laboratorio “*Defs*”
 - Calcolo similarità tra le definizioni create (cardinalità intersezione normalizzata su lunghezza minima tra le due, o varianti)
 - Aggregazione sulle due dimensioni (concretezza / specificità), e analisi risultati

Esercizio Lab - *Defs*

	Astratto	Concreto
Generico	Emotion	Person
Specifico	Revenge	Brick

- Esperimento usando filtri, preprocessing e metriche di sovrapposizione a vostra scelta

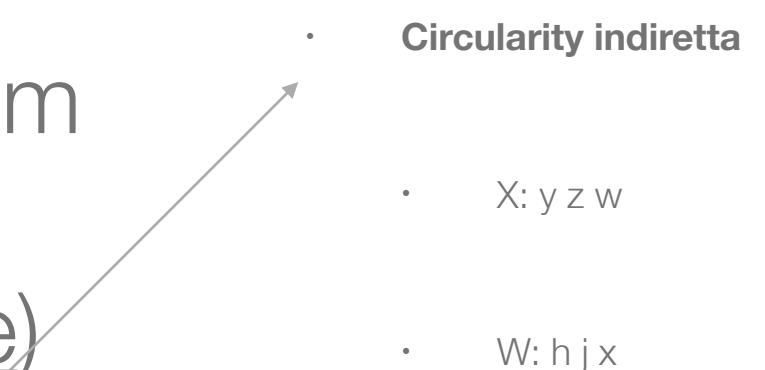
(es. usando stemming e stopword removal come preprocessing, calcolo frequenze top-k nelle definizioni, ecc.)

Esercizio Lab - *Defs*

	Astratto	Concreto
Generico	coraggio 0.13	carta 0.17
Specifico	inquietudine 0.09	temperino 0.33

Da edizioni
passate...

Comprensione del linguaggio naturale

- Quanto è “complesso” identificare un concetto data una sua definizione?
 - Onomasiologic search
 - Dizionari analogici
 - Tip-of-the-tongue problem
 - Genus-differentia mechanism
 - Circularity (dirette e indirette)
- 
- **Circularity indiretta**
- X: y z w
 - W: h j x

Laboratorio - *content2form*

- Lab
 - Esperimento content-to-form
 - Usando i dati sulle definizioni
 - Per ogni concetto,
 - prendere le definizioni a disposizione,
 - Cercare in WordNet il synset corretto
 - Suggerimento: usate il principio del “genus” per indirizzare la ricerca