



Come dare senso a un termine?

Caratteristiche, potenzialità e opportunità dello strumento LexO

A. Bellandi, E. Giovannetti, S. Marchi, S. Piccini, F. Sciolette
Istituto di Linguistica Computazionale “A. Zampolli”
CNR - Pisa



Approccio *knowledge-driven*

- ❖ LexO: uno strumento per un approccio alla terminologia *lexicon-driven*
- ❖ Due approcci adottati oggi in terminologia (L'Homme 2019):
 - approccio *knowledge-driven*
 - approccio *lexicon-driven*
- ❖ Approccio *knowledge-driven*, tipico della terminologia tradizionale
 - concetto punto di partenza dell'attività del terminologo
 - definizione del concetto *a priori*, indipendentemente dalla sua realizzazione linguistica
 - strutture concettuali considerate universali e *language-independent*
 - corrispondenza biunivoca perfetta tra concetto e termine: ad ogni concetto corrisponde uno ed un solo termine

Approccio onomasiologico concetto-centrico

Approccio *lexicon-driven*

- ❖ Crisi dell'approccio *knowledge-driven* (Bourigault e Slodzian 1999, Cabré 2003, Diki-Kidiri 2000, Faber 2012, Gaudin 2003, Temmermann 2000), sotto la spinta propulsiva della linguistica dei corpora

Le concept est le signifié d'un mot dont on décide de négliger la dimension linguistique (Rastier 1995: 55)

- ❖ Approccio *lexicon-driven*
 - termini considerati in tutto e per tutto dei segni linguistici, delle unità lessicali: mostrano il loro funzionamento all'interno di un contesto e nella interazione con le altre unità lessicali (approccio relazionale)
 - approccio semasiologico del terminologo

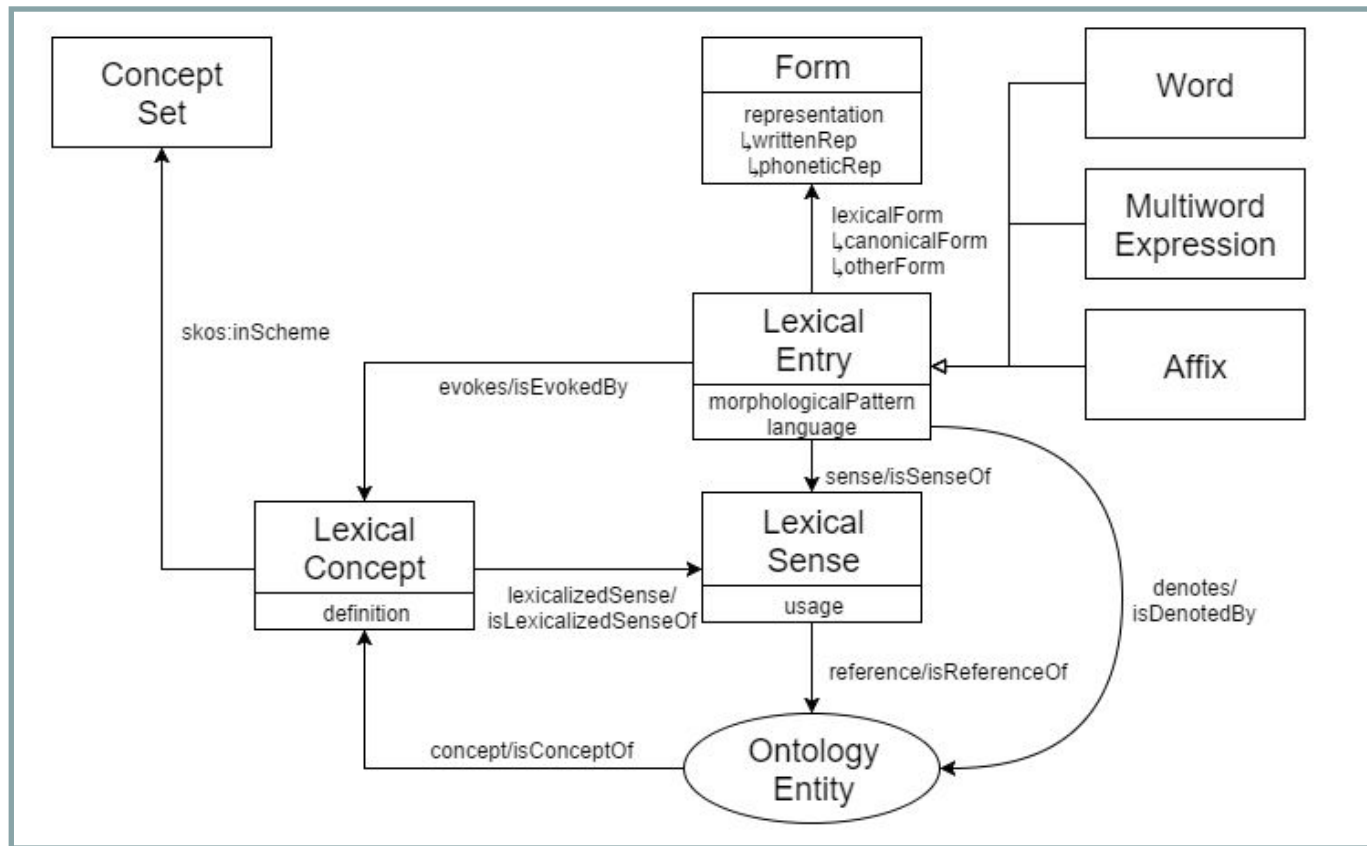
Terminologia e semantica lessicale

- ❖ Importanza, nella creazione di risorse terminologiche, delle acquisizioni della linguistica teorica e della semantica lessicale
- ❖ Applicazione di diverse teorie per l'analisi dei termini tratti da corpora specialistici:
 - *Corpus Pattern Analysis* (Hanks and Pustejovsky 2005)
 - *Distributional Semantics* (Harris 1968)
 - *Explanatory Combinatorial Lexicology* (Mel'čuk *et al.* 1995)
 - *DicoInfo* e *DicoEnviro*
 - *Frame Semantics* (Fillmore 1976)
 - *EcoLexicon*
 - *Generative Lexicon* (Pustejovsky 1995)
- ❖ Carenza di strumenti che aiutino il terminologo nella sua attività di descrizione e strutturazione del senso di un termine secondo il *framework* teorico che di volta in volta viene ritenuto più opportuno

LexO, il web semantico ed i *Linked Data*

- ❖ LexO: strumento per creare e pubblicare risorse costruite secondo le tecnologie del Web Semantico ed i principi dei *Linked Data*
- ❖ Strumento pensato per terminologi non esperti delle tecnologie del web semantico (RDF, OWL, SPARQL)
- ❖ Strumento per creare risorse terminologiche che possano essere messe a disposizione dell'intera comunità scientifica (filosofia della scienza aperta e dei principi FAIR)
- ❖ Strumento basato sul modello *OntoLex-Lemon* (McCrae *et al.* 2017), standard *de facto* per la rappresentazione di risorse lessicali e terminologiche in RDF nel contesto del web semantico

Il modello *Ontolex-Lemon*



lexInfo
['lɛksɪnfəʊ]



links

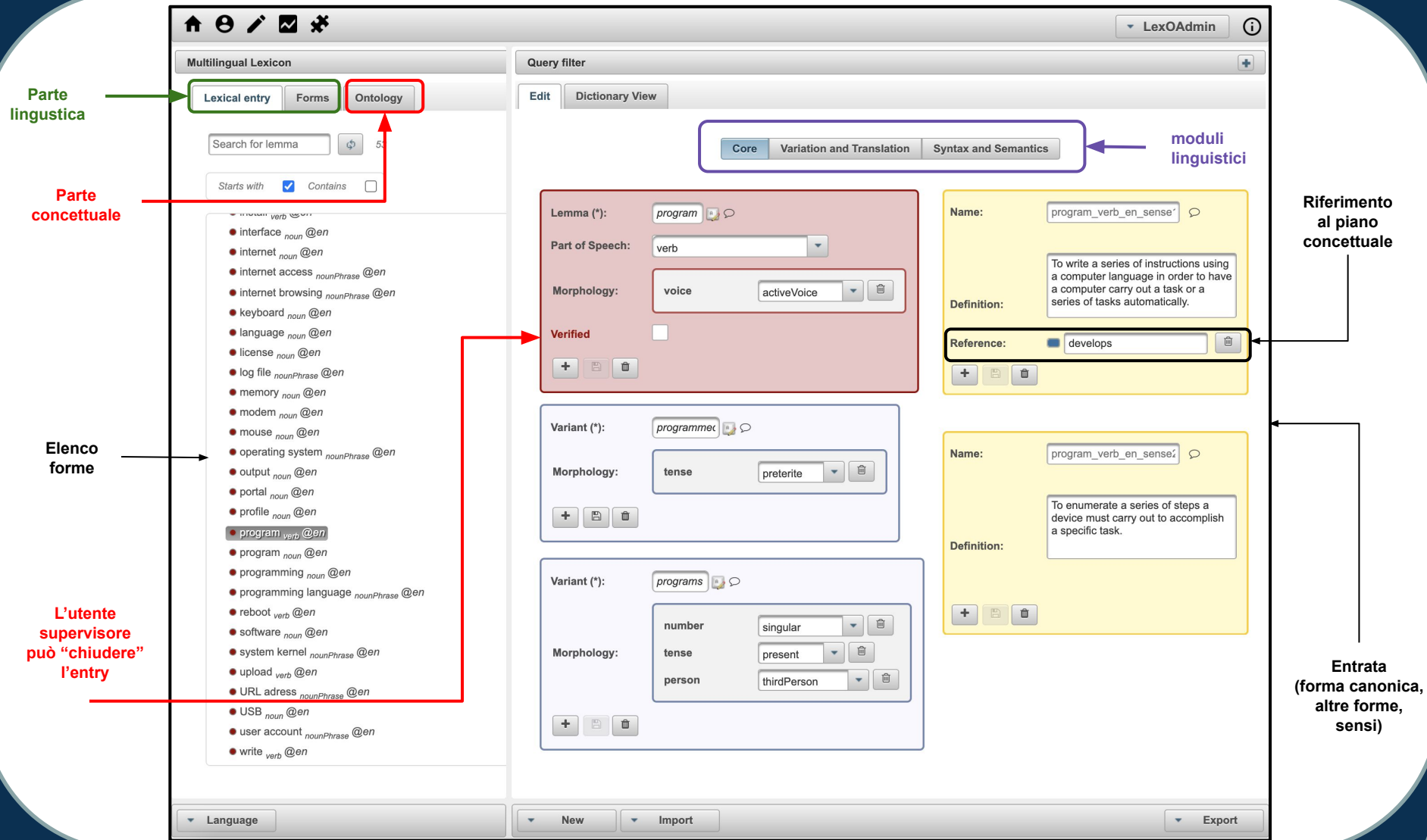


tecnologie

LexO: codifica nel modello *OntoLex-lemon*

- ❖ LexO (Bellandi *et al.* 2017, 2018): editor web collaborativo per la costruzione di terminologie e lessici computazionali secondo il modello *OntoLex-lemon*
- ❖ Applicazione di LexO in numerosi progetti
 - *DiTMAO*: progetto per la creazione della terminologia medico-botanica dell'antico Occitano (<https://www.uni-goettingen.de/en/487498.html>)
 - *FdS*: progetto per la creazione del lessico multilingue diacronico della terminologia di F. de Saussure (<http://www.ilc.cnr.it/en/content/saussure>)
 - *Totus Mundus*: progetto per la creazione di una risorsa bilingue cinese-italiano della mappa di Ricci (<http://www.totusmundus.it>)
 - *Talmud*: progetto di traduzione italiana del Talmud Babilonese (<http://www.talmud.it>)
- ❖ LexO facilita la codifica delle risorse, astruendo dalla complessità del modello di rappresentazione
- ❖ Scaricabile liberamente da <https://github.com/andreabellandi/LexO-lite>  GitHub

LexO: parte linguistica - core



LexO: parte linguistica - variazioni e traduzioni

The screenshot displays the LexO software interface, which is used for managing multilingual lexicons. The interface is divided into several sections:

- Multilingual Lexicon:** This section contains three tabs: **Lexical entry**, **Forms**, and **Ontology**. The **Lexical entry** tab is currently selected, showing a list of terms (e.g., interface, internet, internet access, etc.) with their respective parts of speech and language codes (e.g., @en).
- Query filter:** This section allows users to filter the lexicon data. It includes a search bar and a dropdown menu for selecting the language (currently set to @en).
- Core, Variation and Translation, Syntax and Semantics:** These are the main modules of the software, each represented by a button. The **Variation and Translation** module is currently selected.
- Elenco forme:** This is a list of forms (e.g., program, programmer, programs) that are associated with the selected term. It includes fields for the lemma, part of speech, morphology, and a verified checkbox.
- Definition:** This section provides a detailed definition of the term, including a text box for the definition and a dropdown menu for selecting the type of definition (e.g., quasi equivalent, write_verb_en_sense2).
- Direct relation:** This section shows the direct relations between the term and other terms in the lexicon, including a dropdown menu for selecting the relation type (e.g., associativeRelation, language_noun_en_ser) and a text box for the target term.
- Translation:** This section shows the translations of the term into other languages, including a dropdown menu for selecting the translation type (e.g., directEquivalent, programmer_verb_fr_si) and a text box for the target term.

Annotations in the image highlight the following elements:

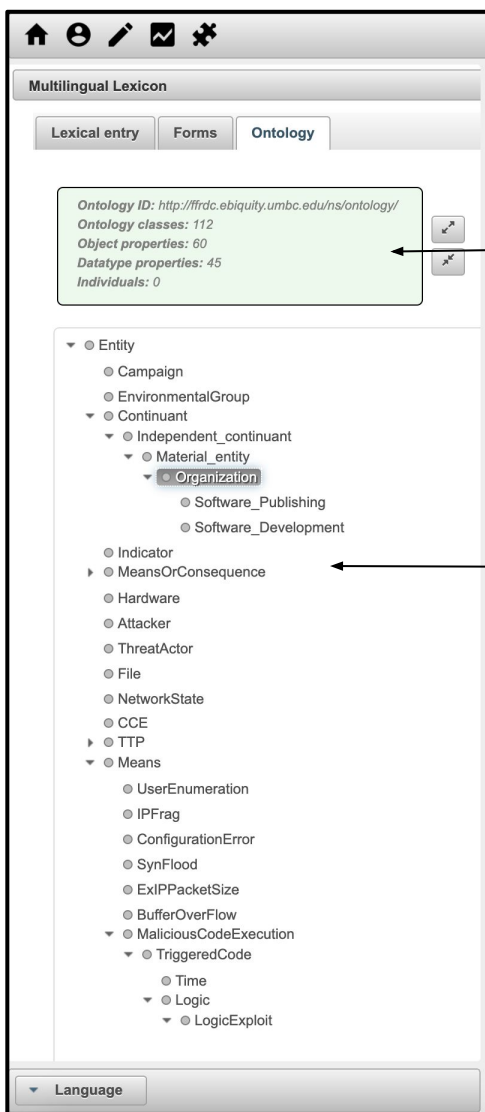
- Parte linguistica:** Points to the **Lexical entry** tab.
- Parte concettuale:** Points to the **Ontology** tab.
- Elenco forme:** Points to the list of forms (e.g., program, programmer, programs).
- moduli linguistici:** Points to the **Core**, **Variation and Translation**, and **Syntax and Semantics** modules.

LexO: parte linguistica - sintassi e semantica

The screenshot displays the LexO web interface, which is divided into several sections:

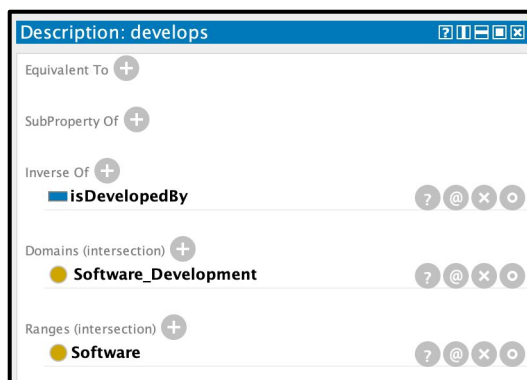
- Multilingual Lexicon:** Contains tabs for 'Lexical entry', 'Forms', and 'Ontology'. The 'Ontology' tab is highlighted with a red box. Below the tabs is a search bar labeled 'Search for lemma' and a list of lemmas. A red arrow points from the text 'Parte concettuale' to the 'Ontology' tab. A green arrow points from the text 'Parte linguistica' to the 'Lexical entry' tab. A black arrow points from the text 'Elenco forme' to the list of lemmas.
- Query filter:** Contains tabs for 'Edit' and 'Dictionary View'. The 'Dictionary View' tab is highlighted with a purple box. Below the tabs are three buttons: 'Core', 'Variation and Translation', and 'Syntax and Semantics'. A purple arrow points from the text 'moduli linguistici' to the 'Syntax and Semantics' button.
- Lemma details:** A red box highlights the 'Lemma (*)' section, which includes fields for 'Part of Speech' (set to 'verb') and 'Morphology' (set to 'voice' and 'activeVoice'). Below this is a 'Verified' checkbox.
- Frame type:** A green box highlights the 'Frame type' section, which includes a dropdown menu set to 'transitive frame' and a 'Frame example' field containing 'He programs iPhone apps'. Below this is an 'Arguments' section with two arguments: '#1' (Type: 'subject', Marker: 'Optional') and '#2' (Type: 'object', Marker: 'Optional'). Green arrows point from the text 'Il soggetto sintattico del frame corrisponde al soggetto della relazione concettuale "develops"' to the '#1' argument field, and from the text 'L'oggetto sintattico del frame corrisponde all'oggetto della relazione concettuale "develops"' to the '#2' argument field.
- Definition and Reference:** A yellow box highlights the 'Definition' and 'Reference' sections. The 'Definition' field contains the text 'To write a series of instructions using a computer language in order to have a computer carry out a task or a series of tasks automatically.' The 'Reference' field contains the text 'develops'. Below this is an 'Ontology mapping' section with a dropdown menu set to 'TransitiveFrame - program_verb_it_frame1' and two dropdown menus for 'Subj of property' and 'Obj of property'.
- Export:** A red box highlights the 'Export' button at the bottom right of the interface. A red arrow points from the text 'Export della risorsa pronta alla pubblicazione sul web' to the 'Export' button.

LexO: parte concettuale



**Statistiche
sull'ontologia di
dominio
(classi, proprietà,
individui)**

**Browsing della
struttura
dell'ontologia**



LexO non intende
essere un editor di
ontologie di dominio

Software come
Protégé permettono di
costruire ontologie

LexO offre un
meccanismo per
importare un'ontologia
di dominio e collegare
quindi il piano
concettuale a quello
linguistico

LexO: dictionary view

Stato della risorsa

program verb activeVoice

programmed *preterite*

programs *singular present thirdPerson*

Lemma e tratti morfologici

Forme e tratti morfologici

sense 1

definition: To write a series of instructions using a computer language in order to have a computer carry out a task or a series of tasks automatically.

directEquivalent: programmare (it)

directEquivalent: programmer (fr)

Equivalent: develop (en), write (en), code (en)

associativeRelation: programming (en), language (en), programming language (en)

SEMANTIC REFERENCE: DEVELOPS (OBJECTPROPERTY)

traduzioni

Relazioni semantiche

Entità concettuale denotata

Descrizione del senso

sense 2

definition: To enumerate a series of steps a device must carry out to accomplish a specific task.

synonym: configure (en)

SEMANTIC REFERENCE: (NONE)

Print

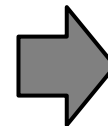
LexO: query

Filtri di supporto alla ricerca dei termini

The screenshot shows the 'Query filter' window with two tabs: 'Advanced search' and 'CNL queries'. Under 'Advanced search', there are several filters: 'All' (checked), 'Word' (unchecked), 'Multiword' (unchecked), and '@' (unchecked). A dropdown menu for 'All languages' is set to 'All languages'. Below this, there is a 'Part of speech' dropdown set to 'any'. At the bottom, there are checkboxes for 'Verified' (unchecked) and 'Unverified' (unchecked). To the right of the filters are 'Filter' and 'Reset' buttons.

Ricerca semantica dei termini

This screenshot shows the 'Query filter' window with the 'Advanced search' tab selected. The text 'Show the entries which refer to the' is followed by a dropdown menu showing a list of concepts: 'develops', 'Means', 'CVSSScoreType', 'Time', 'DomainExpert', 'Process', 'Logic', 'Indicator', and 'LogicExploit'. The 'develops' option is currently selected. To the right of the dropdown is the word 'concept'. 'Filter' and 'Reset' buttons are also visible.



The screenshot shows the 'Lexical entry' tab selected. There is a search input field with a refresh button. Below it, there are checkboxes for 'Starts with' (checked) and 'Contains' (unchecked). At the bottom, a search result is displayed: a red dot followed by 'program' and 'verb @en'.

Le funzioni lessicali in LexO

The screenshot shows the 'Dictionary View' of the LexO interface. At the top, there are two tabs: 'Edit' and 'Dictionary View'. The main content area displays the entry for 'geṭ' as a noun. It is divided into two senses. Sense 1 has a definition and two lexical functions: 'Oper₁(geṭ) = natan' and 'Anti(geṭ) = šeṭar', which are grouped by a blue bracket labeled 'Funzioni lessicali'. Sense 2 has a definition and four lexical functions: 'Syn_{narrower}(geṭ) = sefer keritut', 'Oper₁(geṭ) = natan', 'S₂(geṭ) = 'iša', and 'Anti(geṭ) = šeṭar', also grouped by a blue bracket labeled 'Funzioni lessicali'. An arrow points from the 'Oper₁(geṭ) = natan' entry in sense 2 to the text 'formalizza la nozione di verbo supporto' located below the interface. A 'Print' button is visible at the bottom left of the interface.

geṭ noun

sense 1
definition: Official written document attesting the juridical relationship between X and Y.
Oper₁(geṭ) = natan
Anti(geṭ) = šeṭar } *Funzioni lessicali*

sense 2
definition: Official written document stating that X is no longer tied to Y by the marriage bond.
Syn_{narrower}(geṭ) = sefer keritut
Oper₁(geṭ) = natan
S₂(geṭ) = 'iša
Anti(geṭ) = šeṭar } *Funzioni lessicali*

Print

formalizza la nozione di verbo supporto



Grazie per l'attenzione!



Andrea Bellandi & Silvia Piccini
Istituto di linguistica Computazionale
“A. Zampolli” ILC-CNR Pisa



Bibliografia (1/2)

- ❖ Bellandi A., Giovannetti E., Piccini S., Weingart A., 2017, “Developing LexO: a Collaborative Editor of Multilingual Lexica and Terminological Resources in the Humanities”, in *Proceedings of Language, Ontology, Terminology and Knowledge Structures Workshop (LOTKS 2017)*, Montpellier, France, 19 September 2017.
- ❖ Bellandi A., Giovannetti E., Weingart A., 2018, “Multilingual and Multiword Phenomena in a lemon Old Occitan Medico-Botanical Lexicon”. *Information*.
- ❖ Bourigault D., Slodzian M., 1999, Pour une terminologie textuelle. *Terminologies nouvelles* 19, pp. 29-32.
- ❖ Cabré M. T., 2003, “Theories of Terminology: Their Description, Prescription and Explanation”, *Terminology* 9(2), pp. 163-200.
- ❖ Diki-Kidiri M., 2000, “Une approche culturelle de la terminologie”, *Terminologies nouvelles* 21, pp. 27-41.
- ❖ Faber P. (ed.), 2012, *A Cognitive Linguistics View of Terminology and Specialized Language*, Berlin/New York, Mouton de Gruyter.
- ❖ Fillmore C.J., 1976, “Frame Semantics and the Nature of Language”, *Annals New York Academy of Sciences: Conference in the Origin and Development of language and Speech* 280, pp. 20-32.

Bibliografia (2/2)

- ❖ Gaudin F., 2003, *Socioterminologie: une approche sociolinguistique de la terminologie*, Bruxelles, Duculot.
- Hanks P., Pustejovsky J., 2005, "A Pattern Dictionary for Natural Language Processing", *Revue française de linguistique appliquée* 10(2), pp. 63-82.
- ❖ Harris Z., 1968, *Mathematical Structures of Language*, New York, Wiley.
- ❖ L'Homme M. C., 2019, *Lexical Semantics for Terminology. An Introduction*, Amsterdam/Philadelphia, John Benjamins Publishing Company.
- ❖ McCrae J. P., Bosque-Gil J., Gracia J., Buitelaar P., Cimiano P., 2017, "The Ontolex-Lemon model: development and applications", in *Proceedings of eLex 2017 conference*, pp. 19-21.
- ❖ Mel'čuk I. A. , Clas A., Polguère A., 1995, *Introduction à la lexicologie explicative et combinatoire*, Duculot, Bruxelles.
- ❖ Pustejovsky J., 1995, *The Generative Lexicon*, 1st ed. Cambridge MA, The MIT Press.
- ❖ Rastier F., 1995, "Le terme: entre ontologie et linguistique", *La Banque des mots* 7 (Terminologie et Intelligence Artificielle).
- ❖ Temmermann R., 2000, *Towards New Ways of Terminological Description: The Sociocognitive Approach*, Amsterdam/Philadelphia, John Benjamins Publishing Company.