



Applying the OntoLex Model to a Multilingual Terminological Resource

**Julia Bosque-Gil, [Jorge Gracia](#),
Guadalupe Aguado-de-Cea, Elena Montiel-Ponsoda**

Ontology Engineering Group (OEG)
Universidad Politécnica de Madrid (UPM)

jgracia@fi.upm.es

Acknowledgments: LIDER and 4V projects

**4th Multilingual Semantic Web Workshop
(MSW'15), Portoroz (Slovenia) 1/06/2015**




- Motivation and contribution
- The lemon-Ontolex model
- The Terminesp database
- Representing core lexical information
- Representing translations and variations
- Conclusions

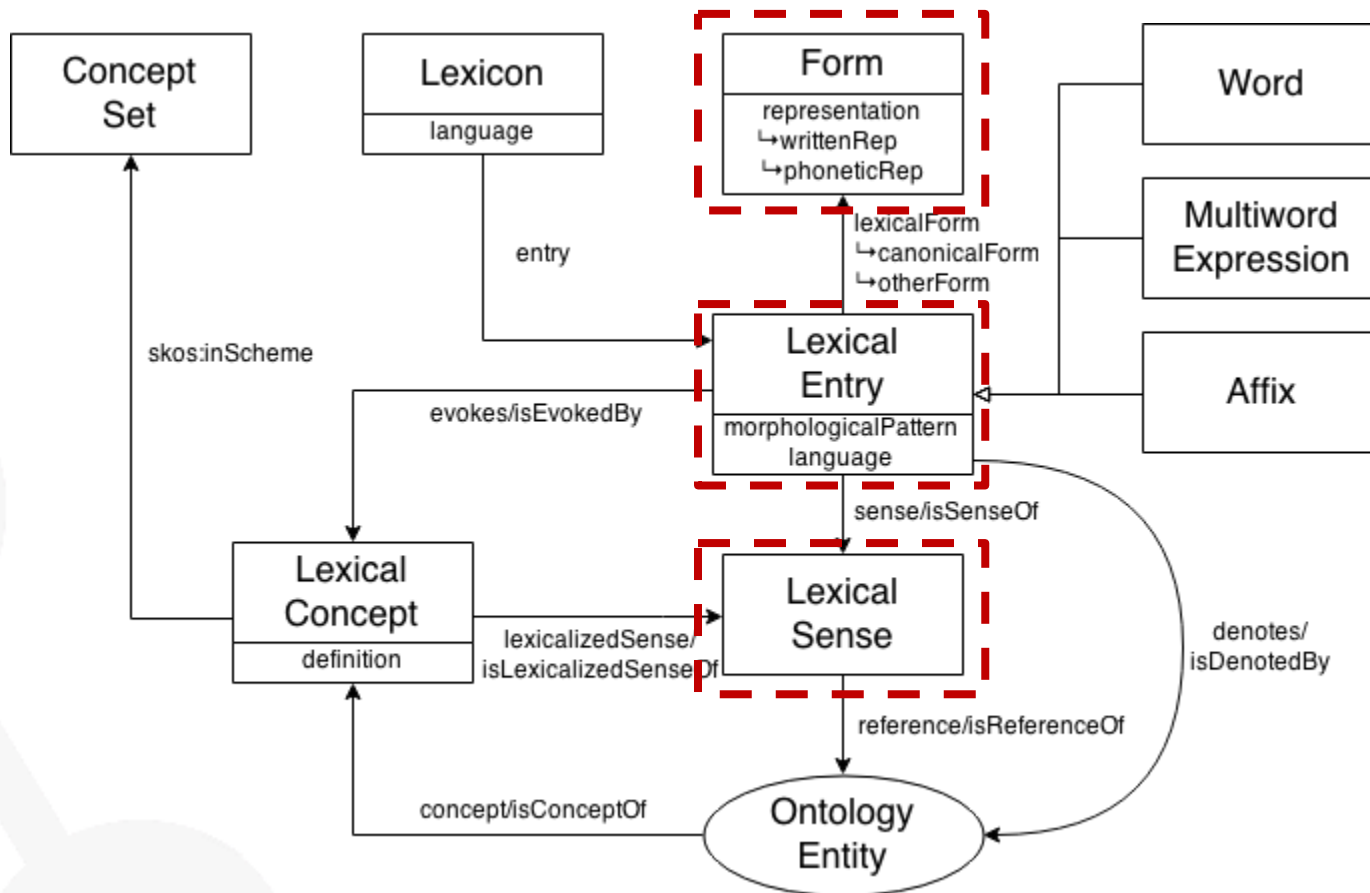
- Motivation and contribution
- The lemon-Ontolex model
- The Terminesp database
- Representing core lexical information
- Representing translations and variations
- Conclusions

- **RDF** generation of Terminesp, a multilingual database with highly specialized technical content
- **Extension** of the previous LD version with much more information (definitions, PoS, provenance, ...)
- A pool of atomic and multi-word **terminological expressions** labelled with syntactic information (PPs, NPs, etc.) and part-of-speech tags
- One of the first showcases of **lemon-Ontolex**
- First validating test-bench for its **vartrans** module

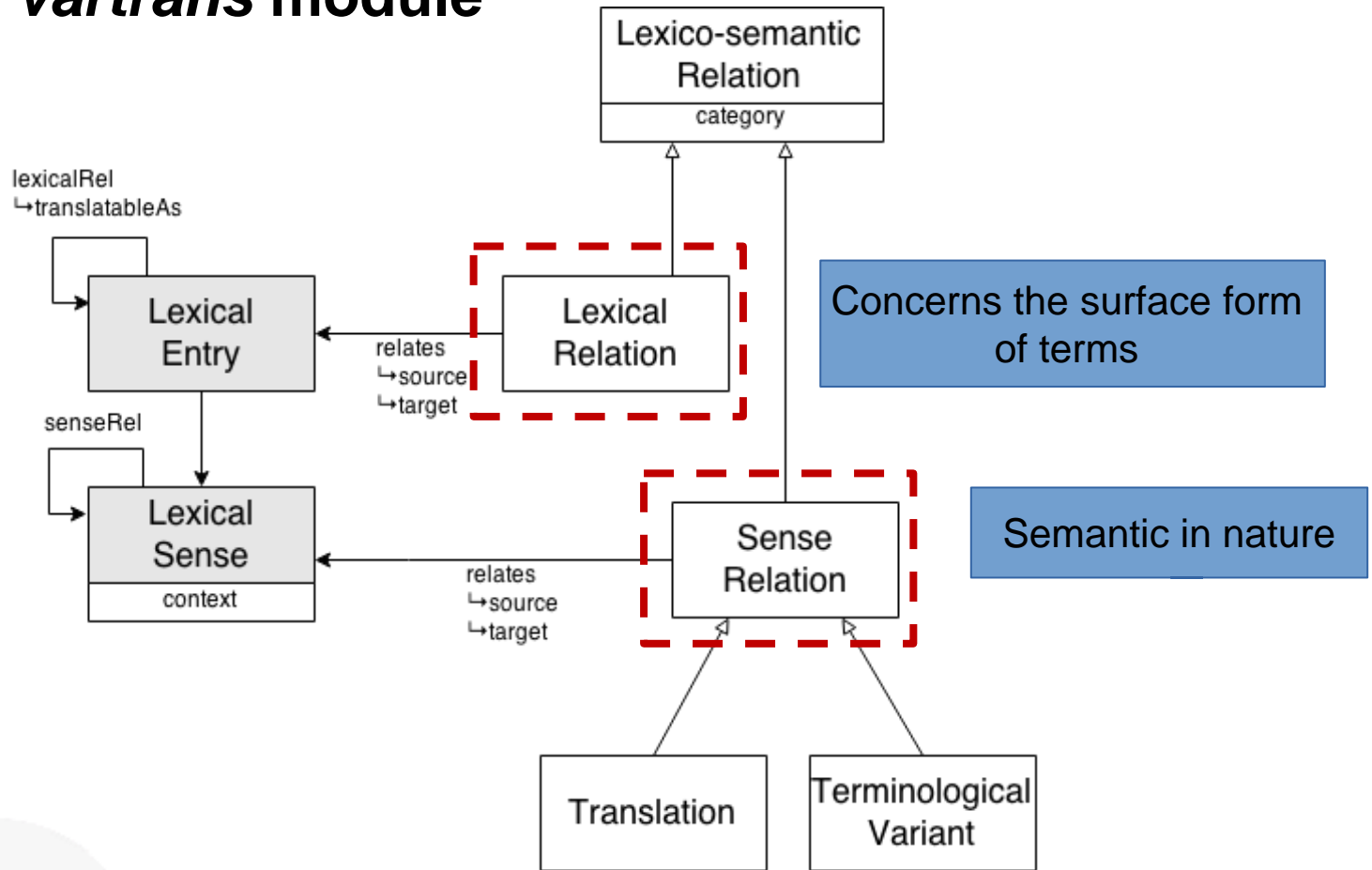
- Motivation and contribution
- **The lemon-Ontolex model**
- The Terminesp database
- Representing core lexical information
- Representing translations and variations
- Conclusions

- Rich model to represent the [ontology-lexicon](#) interface
- Developed by the [W3C Ontology-Lexica](#) (OntoLex) Community Group
- Largely based on the **lemon**  model
- Still [under development](#), final model to be released soon

The core concepts



The lemon *vartrans* module



- Motivation and contribution
- The lemon-Ontolex model
- **The Terminesp database**
- Representing core lexical information
- Representing translations and variations
- Conclusions

- Created by the Spanish Terminological Association (AETER)
- Terminological data from UNE documents (standards):
 - +30,000 technical terms with norms and definitions
 - A range of domains: aeronautics, electrical engineering, agriculture, ...
 - Translations to English, French, German, Italian Swedish, and Latin

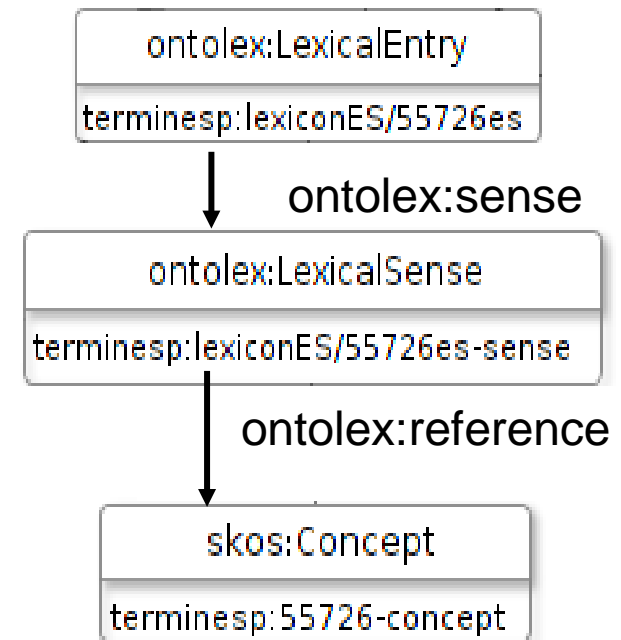
Norm	UNE_21302-806
Norm title	<i>Electrotechnical Vocabulary: Chapter 806. Sound and image recording and reading. (*)</i>
Spanish term	<i>aguja de lectura</i>
German term	<i>Abtastnadel</i>
English term	<i>reproducing stylus tip</i>
French term	<i>pointe de lecture</i>
Definition	<i>Mechanical transmission element consisting of an appropriate needle whose function is to follow the modulation of a registered groove, transmitting the resulting vibrations to the reading head (*)</i>
Note	<i>Note 1 - This term is commonly used to designate the group formed by the needle and its rim [...] (*)</i>

(*) In Spanish in the original, translated to English for illustrative purposes

- Motivation and contribution
- The lemon-Ontolex model
- The Terminesp database
- **Representing core lexical information**
- Representing translations and variations
- Conclusions

Representing core lexical information

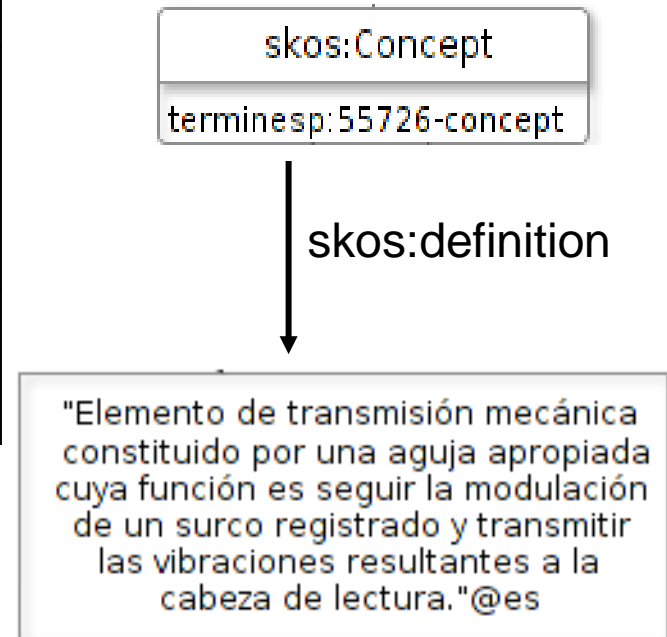
Norm	UNE_21302-806
Norm title	<i>Electrotechnical Vocabulary: Chapter 806. Sound and image recording and reading. (*)</i>
Spanish term	<i>aguja de lectura</i>
Definition	<i>Mechanical transmission element consisting of an appropriate needle whose function is to follow the modulation of a registered groove, transmitting the resulting vibrations to the reading head (*)</i>
Note	<i>Note 1 - This term is commonly used to designate the group formed by the needle and its rim [...] (*)</i>



Representing core lexical information

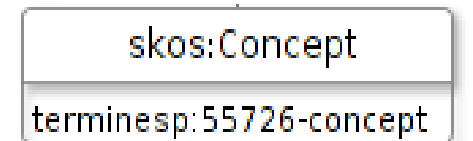
Norm	UNE_21302-806
Norm title	<i>Electrotechnical Vocabulary: Chapter 806. Sound and image recording and reading. (*)</i>
Spanish term	<i>aguja de lectura</i>
Definition	<i>Mechanical transmission element consisting of an appropriate needle whose function is to follow the modulation of a registered groove, transmitting the resulting vibrations to the reading head (*)</i>
Note	<i>Note 1 - This term is commonly used to designate the group formed by the needle and its rim [...] (*)</i>

- lemon:SenseDefinition not included in OntoLex
- ontollex:definition applies to ontollex:LexicalConcept



Representing core lexical information

Norm	UNE_21302-806
Norm title	<i>Electrotechnical Vocabulary: Chapter 806. Sound and image recording and reading. (*)</i>
Spanish term	<i>aguja de lectura</i>
Definition	<i>Mechanical transmission element consisting of an appropriate needle whose function is to follow the modulation of a registered groove, transmitting the resulting vibrations to the reading head (*)</i>
Note	Note 1 - This term is commonly used to designate the group formed by the needle and its rim [...] (*)

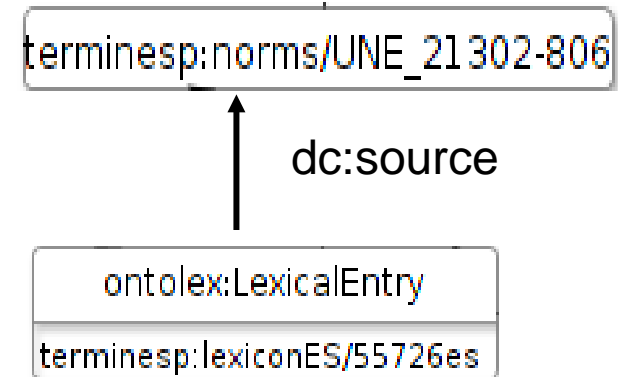


rdfs:comment

"Nota 1 - Esta expresión se emplea comúnmente para designar el conjunto formado por la aguja de lectura y su montura. Nota 2 - Véase la figura 806-3."@es

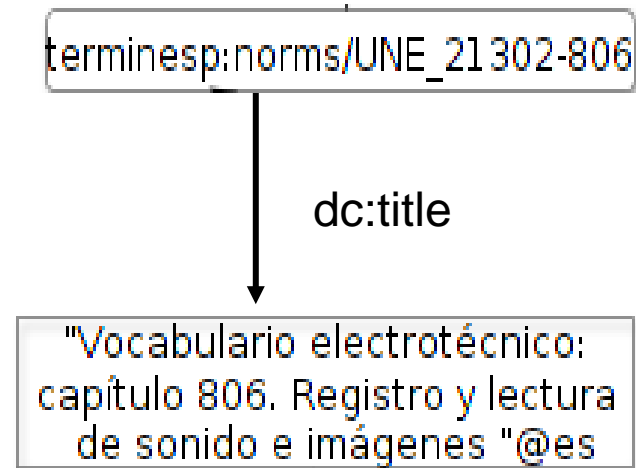
Representing core lexical information

Norm	UNE_21302-806
Norm title	<i>Electrotechnical Vocabulary: Chapter 806. Sound and image recording and reading. (*)</i>
Spanish term	<i>aguja de lectura</i>
Definition	<i>Mechanical transmission element consisting of an appropriate needle whose function is to follow the modulation of a registered groove, transmitting the resulting vibrations to the reading head (*)</i>
Note	<i>Note 1 - This term is commonly used to designate the group formed by the needle and its rim [...] (*)</i>



Representing core lexical information

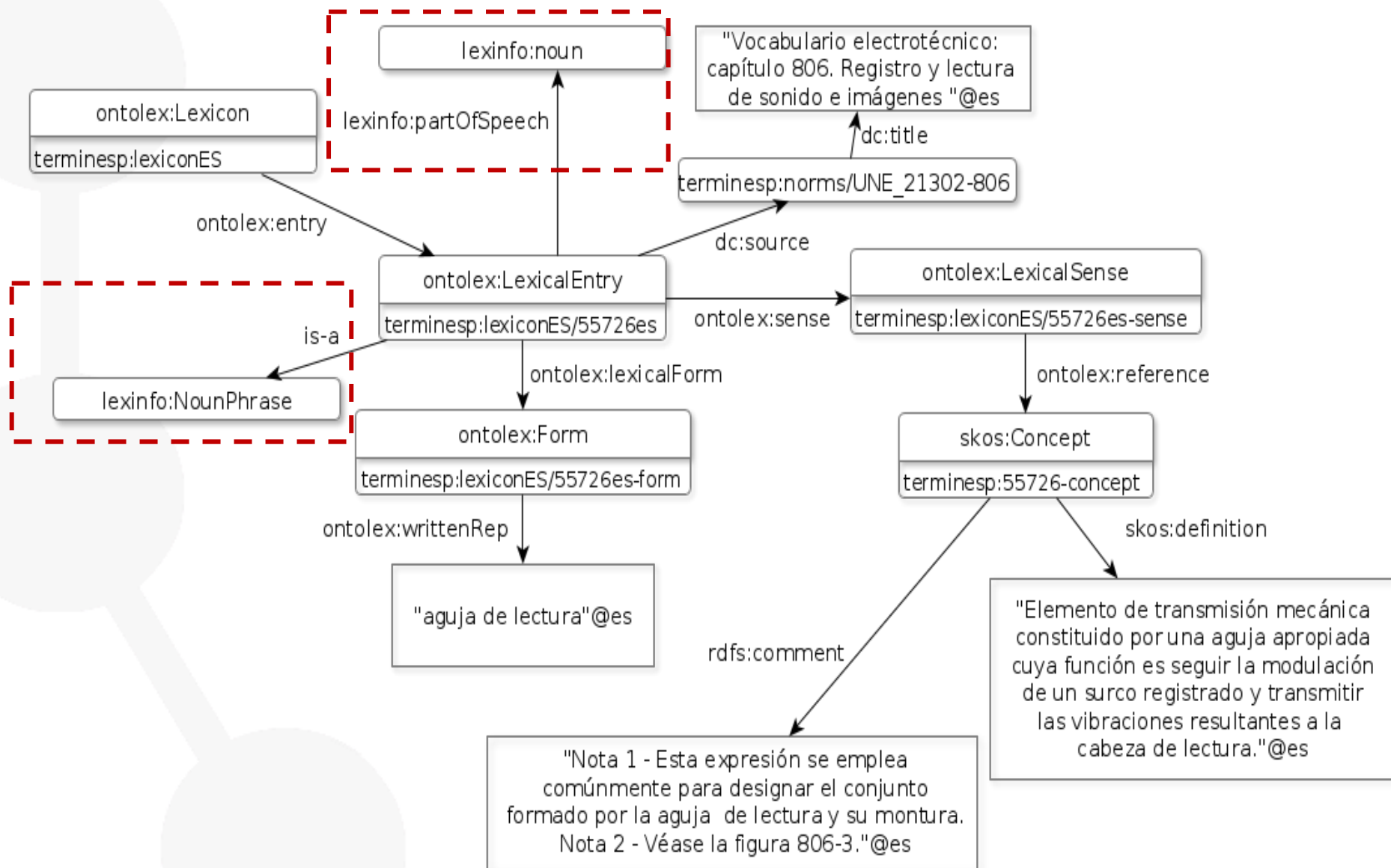
Norm	UNE_21302-806
Norm title	<i>Electrotechnical Vocabulary: Chapter 806. Sound and image recording and reading. (*)</i>
Spanish term	<i>aguja de lectura</i>
Definition	<i>Mechanical transmission element consisting of an appropriate needle whose function is to follow the modulation of a registered groove, transmitting the resulting vibrations to the reading head (*)</i>
Note	<i>Note 1 - This term is commonly used to designate the group formed by the needle and its rim [...] (*)</i>



- Part-of-Speech tags added (TreeTagger)
 - Total number of entries: 32491
 - Most of the terms are constituted by a single noun (13777) or are complex NPs (18552)
 - Scarce number of single adjectives and verbs
 - Tags were manually checked before RDF conversion

- Syntactic behaviour:
 - Type of syntactic phrase in multi-word entries:
lexinfo:NounPhrase/AdjectivePhrase/PrepositionPhrase
 - A prepositional phrase may function as an adjective or as an adverb in Terminesp (indicated in *lexinfo:partOfSpeech*): E.g., “En reposo” [“idle”]
lexinfo:PrepositionPhrase, lexinfo:adjective

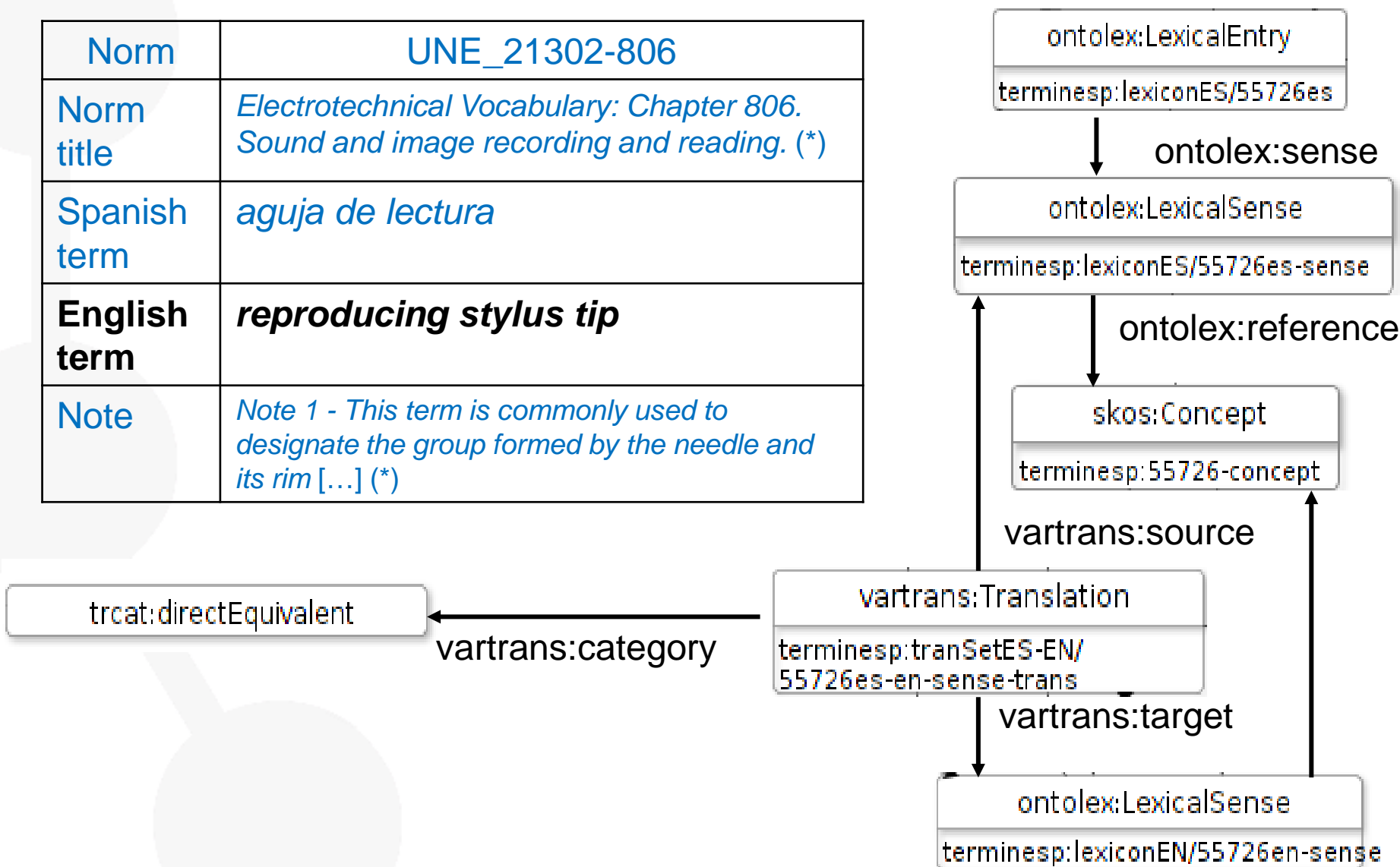
Representing core lexical information



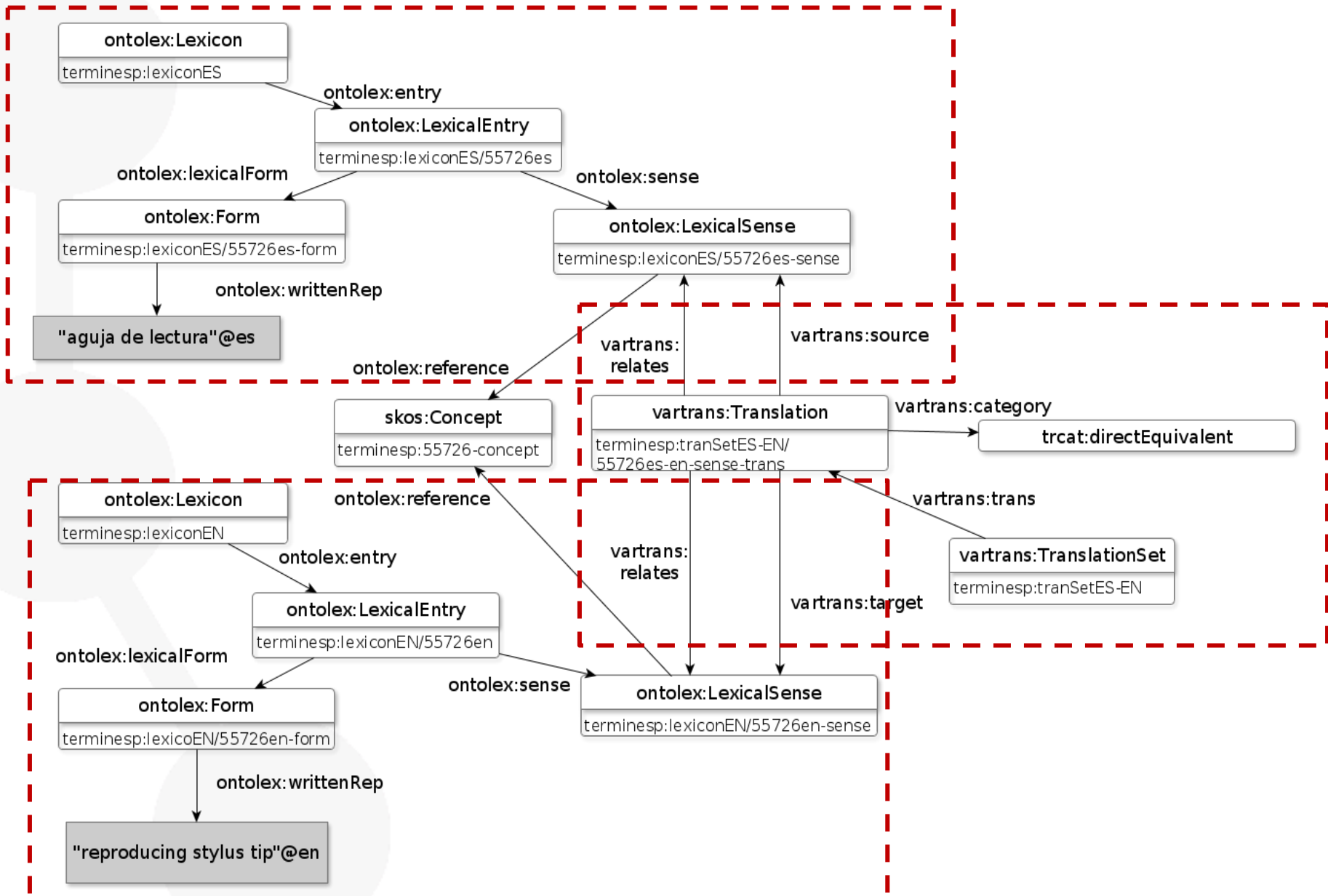
- Motivation and contribution
- The lemon-Ontolex model
- The Terminesp database
- Representing core lexical information
- **Representing translations and variations**
- Conclusions

Representing translations and variations

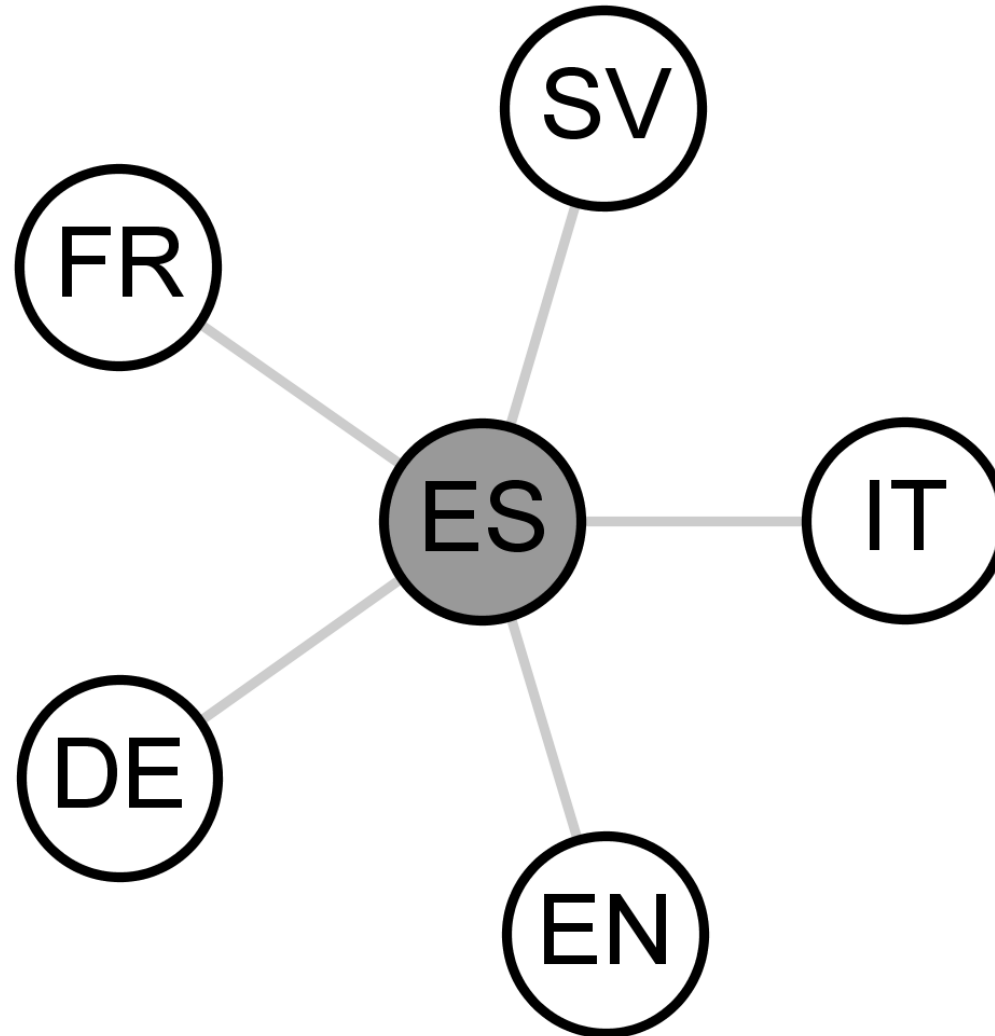
Norm	UNE_21302-806
Norm title	<i>Electrotechnical Vocabulary: Chapter 806. Sound and image recording and reading. (*)</i>
Spanish term	<i>aguja de lectura</i>
English term	reproducing stylus tip
Note	<i>Note 1 - This term is commonly used to designate the group formed by the needle and its rim [...] (*)</i>



Representing translations and variations

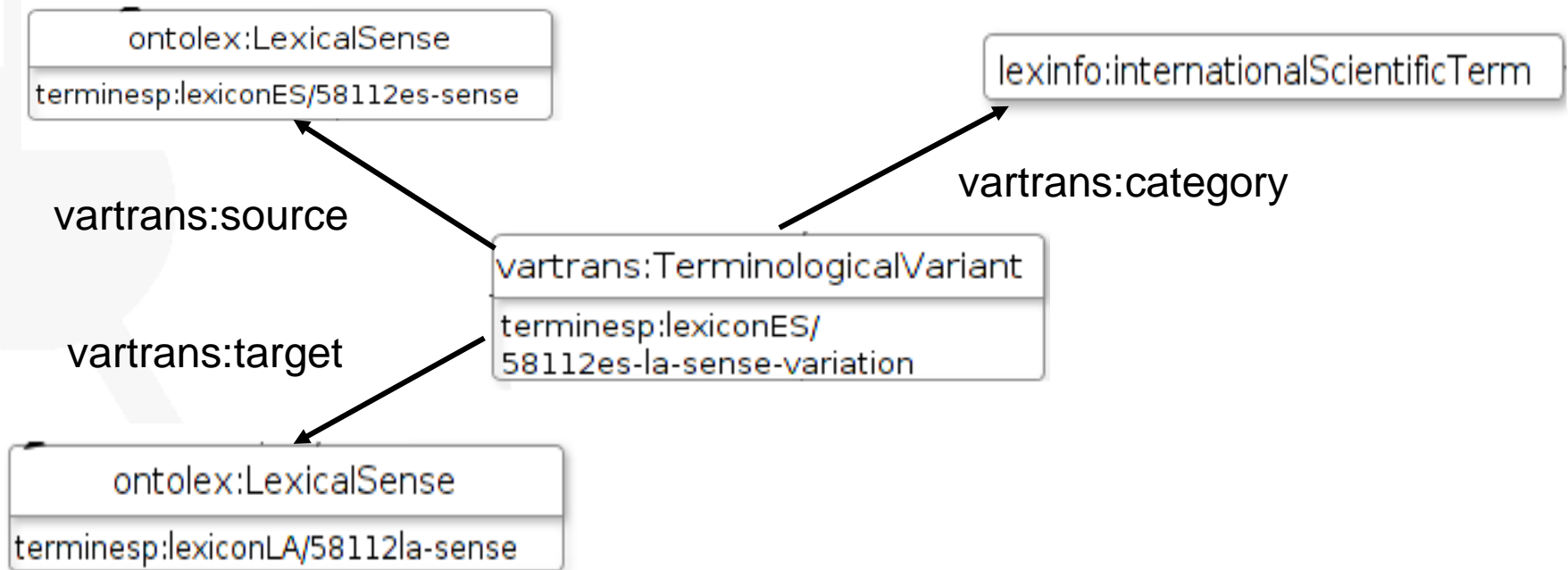


Representing translations and variations



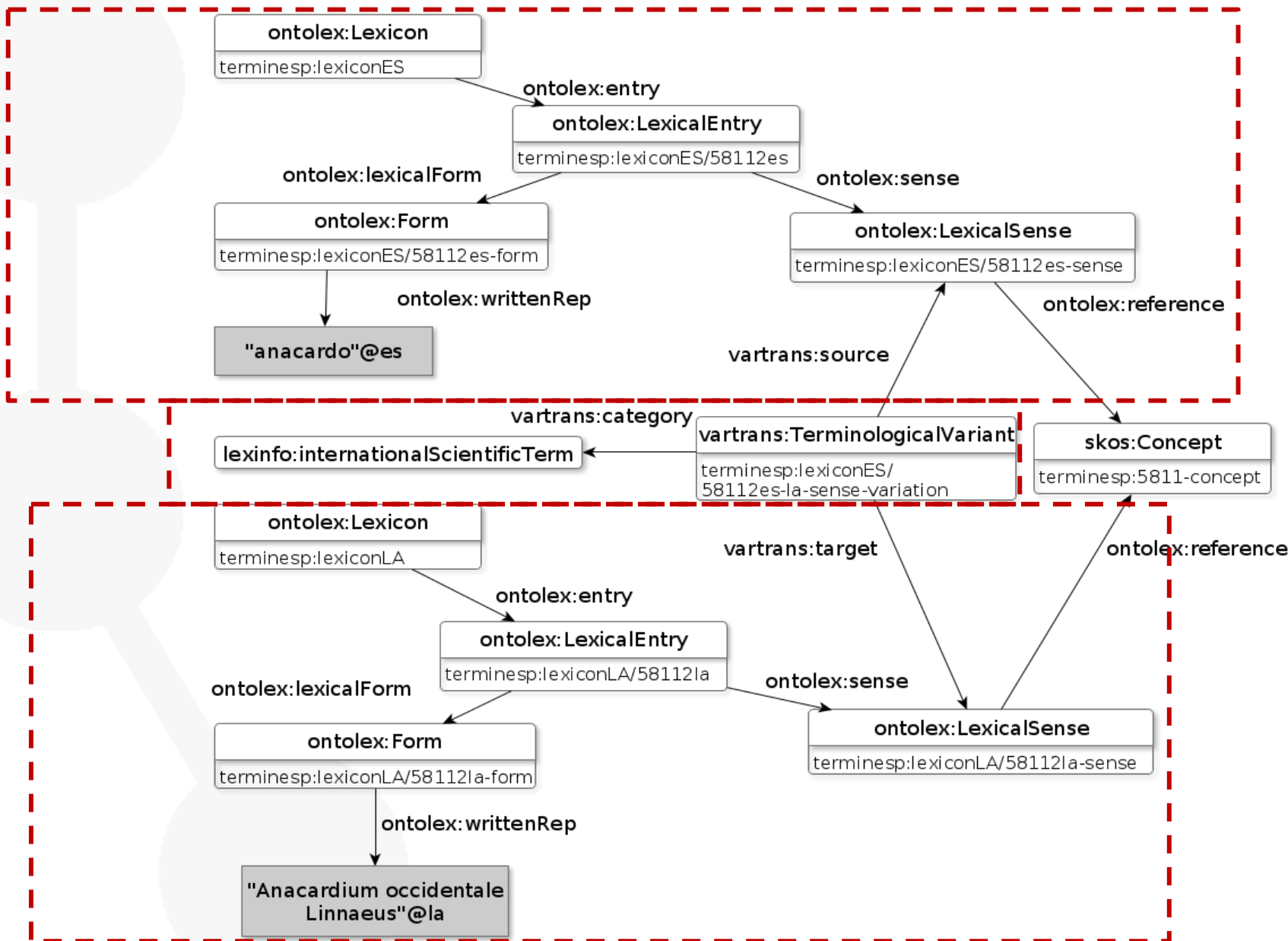
Representing translations and variations

Spanish term	<i>Anacardo</i> *
Latin term	<i>Annacardium occidentale</i> Linnaeus



(*) 'cashew'

Representing translations and variations



- Motivation and contribution
- The lemon-Ontolex model
- The Terminesp database
- Representing core lexical information
- Representing translations and variations
- **Conclusions**

- Terminesp data offered a suitable **testing bench** for **lemon-Ontolex** core and lemon vartrans module.
- We extended the resource with part-of-speech and additional **syntactic information**, which was not explicitly declared in the original data
- Potential significant resource for **NLP-based applications**. E.g.,
 - Future exploration of the use of the **pool of syntactic phrases** in multilingual Information Extraction
 - Use of Spanish as inter-lingua for **translations**



Thanks for your attention !