Atti del IX Convegno Annuale
dell'Associazione per l'Informatica
Umanistica e la Cultura Digitale
(AIUCD)

LA SVOLTA INEVITABILE:
SFIDE E PROSPETTIVE PER L'INFORMATICA UMANISTICA

15 – 17 gennaio 2020
Milano
Università Cattolica del Sacro Cuore

A CURA DI:
Cristina Marras
Marco Passarotti
Greta Franzini
Eleonora Litta

ISBN: 978-88-942535-4-2

Cristina Marras, Marco Passarotti, Greta Franzini, Eleonora Litta (a cura di),
Atti del IX Convegno Annuale AIUCD. La svolta inevitabile:
sfide e prospettive per l'Informatica Umanistica.

Available online as a supplement of Umanistica Digitale: https://umanisticadigitale.unibo.it

The Digital *Lexicon Translaticium Latinum*: Theoretical and Methodological Issues

Chiara Fedriani
University of Genoa
chiara.fedriani@unige.it

Irene De Felice
University of Genoa
irene.defelice@edu.unige.it

William Michael Short
University of Exeter
w.short@exeter.ac.uk

Abstract

English. In this paper, we present some theoretical and methodological issues involved with the creation of the *Lexicon Translaticium Latinum*, a new digital resource for the study of Latin metaphors. This resource is based on the ontology of the Latin WordNet '2.0' (https://latinwordnet.exeter.ac.uk) but extends this specification so as to be able to capture the kinds of large-scale metaphorical patterns that are becoming increasingly documented in Latin's semantic system. In particular, we discuss 1) the theory and method underpinning the revised and expanded ontology; 2) the tagset adopted for classifying metaphors and their relations; and 3) the procedure for annotating conceptual metaphors and linking these with other semantic structures within the WordNet (synsets). In line with the recognition in the 'second wave' cognitive sciences that metaphor is a fundamental mechanism of human cognition (as well as key to the structuring of cultural understanding), our aim is to establish the Lexicon not merely as a comprehensive repository of figurative usage in Latin, but as an accurate model of the conceptual system that its speakers relied upon in thinking, speaking and behaving in diverse contexts of symbolic expression.

Italiano. In questo articolo vengono presentati gli assunti teorici e metodologici alla base del *Lexicon Translaticium Latinum*, una nuova risorsa digitale per lo studio della metafora nella lingua latina, che si basa su Latin WordNet "2.0", un'ontologia semantica della lingua latina liberamente accessibile dal web (https://latinwordnet.exeter.ac.uk), ma estende significati-vamente i dati in esso contenuti rappresentando gli schemi metaforici documentati nel sistema semantico latino a diversi livelli della struttura lessicale. In particolare, si presentano: 1) gli assunti teorici e metodologici alla base della progettazione e della realizzazione dell'ontologia revisionata ad ampliata; 2) il tagset adottato per classificare le metafore e le loro relazioni; 3) la procedura per annotare le metafore concettuali e per collegarle con la struttura semantica dei synset coinvolti. Riconoscendo nella metafora, in linea con le scienze cognitive della "seconda ondata", un meccanismo fondamentale della cognizione umana, sia nella concettualizzazione delle esperienze che nella strutturazione di categorie culturali, il nostro obiettivo è quello di mostrare come il *Lexicon* non sia da intendersi come un mero archivio degli usi figurati attestati in latino, ma piuttosto come un accurato modello di rappresentazione del sistema concettuale alla base del pensiero e del comportamento, anche linguistico, dei parlanti di questa lingua antica.

1 Introduction

The *Lexicon Translaticium Graecum et Latinum* is a collaborative international project aimed at developing an on-line, extensible, open-access lexicon of metaphors in the ancient languages – beginning, in reverse chronological order, with Latin. Unlike existing electronic dictionaries for Latin, which simply re-create their printed counterparts in machine-readable form, the *Lexicon Translaticium* incorporates

insights from up-to-date theories of meaning and, in particular, the view developed in cognitive linguistics of metaphor as a key structuring device of language and thought. In capturing deeply entrenched and highly conventionalized metaphoric and metonymic patterns that organize meanings pervasively throughout this language and at different orders of linguistic encoding, the *Lexicon Translaticium* is meant as a psychologically realistic model of the conceptual system underpinning Latin. Built on top of the ontology provided by the Latin WordNet, the *Lexicon* will be interoperable with existing electronic corpora and thus capable of delivering rich figurative data for integration into natural language processing applications. The project, directed by William Short and Chiara Fedriani and staffed by an international five-member team, is currently on-going. However, its intellectual rationalization is well established and its technical design and implementation have progressed to the point where preliminary 'test' data is already publicly available, with about 20 metaphors presently annotated. We aim to launch the *Lexicon* officially in Spring of 2020 with a fuller dataset consisting of about 100 conceptual metaphors.

2 Theoretical background

Recognizing the all-pervasive character of certain metaphorical patterns in language, George Lakoff and Mark Johnson (1980) argued that the frequent clustering of metaphorical linguistic expressions around abstract, intellectual or otherwise intangible concepts in fact reflects the inherently metaphorical workings of cognition itself. People talk about most abstract concepts metaphorically, that is, because – it is claimed – they actually conceive of them metaphorically in terms of other (usually more concrete) concepts. On this view, metaphors are the projections of conceptual structure and content from one domain to another that occur as a way of mentally representing and reasoning about experiences not directly grounded in the physico-spatial world. Cognitive linguists argue, moreover, that it is the systematic nature of metaphors – in other words, that metaphors characterize regular mappings between organized domains of knowledge – that allows people to think and reason (and therefore also to speak) meaningfully about experiences that may be difficult to comprehend in and of themselves.

In line with this theory, the 'entries' of our metaphor dictionary – unlike those of a traditional lexicon – will therefore consist of large-scale patterns of metaphorical understanding that link together *concepts*, rather than the semantic structures of words *per se* and so structure the meanings of words across the lexicon and at different levels of linguistic encoding. The kinds of metaphors that constitute the data of our *Lexicon* are those that are so conventionalized and so entrenched in the shared linguistic and cognitive habits of Latin speakers that they seem not to have been perceived as figurative at all – and indeed deliver Latin speakers' entirely regular, 'everyday' ways of conceptualizing certain experiences. Of course, as 'imaginative' or 'creative' (or more narrowly 'literary') metaphors most often derive in some way from more conventionalized metaphors, these kinds will also be represented. In this way, the *Lexicon Translaticium Latinum* will form a comprehensive catalogue of the range of metaphorical themes that structure meaning in Latin.

3 Technical implementation

Technically, the *Lexicon* will be realized as a computerized relational database, whose data model combines aspects of the architecture of the MetaNet Project of the International Computer Science Institute in Berkeley, California, with the WordNet framework. The Berkeley MetaNet is an electronic repository, viewable interactively on the Internet as a Wiki, that contains records for hundreds of attested conventional and imaginative metaphors in English, including time metaphors, mind metaphors, and emotion metaphors, as well as metaphors relating to government, disease, and violence. Most importantly for our purposes, the MetaNet provides a set of high-level ontologies for annotating and organizing figurative language data under the theory of conceptual metaphor in cognitive linguistics. In particular, the MetaNet provides a theoretically-grounded formal specification for encoding *kinds* of conceptual metaphors as well as *kinds of relations* between metaphors. For example, the 'type' of a metaphor can be tagged with values such as 'primary', 'composed', or 'entailed', which correspond to well defined theoretical categories. A primary metaphor is one that emerges directly from correlations in experience, as in MORE IS UP or PURPOSES ARE DESTINATIONS, while complex metaphors are those built up out of at

least two more basic primary ones. Entailed metaphors are specialized submappings that can be inferred through experiential knowledge from a primary or complex metaphor, and which often form the basis of coherence between metaphors.

Likewise, metaphors can be organized into hierarchies through simple relations of super- or sub-ordination, or into more intricate systems according to different kinds of (again theoretically grounded) relationships, such as 'extension' (where one mapping takes advantage of conceptual material left unused by another), 'elaboration' (where one mapping embellishes another with additional conceptual material), 'combination', or 'questioning'. 'Reciprocality' is another common feature of metaphor systems and is available to capture 'orientational' metaphors that involve body-based experiential polarities such UP *vs* DOWN, LEFT *vs* RIGHT, CENTER *vs* PERIPHERY, IN *vs* OUT, and so on.

Whereas the MetaNet specification provides the foundation for encoding metaphors (as mappings between concepts) and their relations, the ontologies and data structures of the WordNet deliver the core repertoire of concepts that participate in these relations. As a semantic database, the WordNet represents lexical meaning in terms of *synsets*, which are uniquely identifiable 'definitions' for hypothetically all the senses capable of being expressed in a given language (thus organizing the lexicon into discrete 'synonym sets'). In other words, a WordNet synset – which pairs a unique identifier, consisting of a part-of-speech tag and a string of between six and eight integers, with a descriptive gloss and possibly higher-order 'domain'-level tags – should be seen as representing a distinct *concept* that may constitute the meaning of a word or words in the language under scrutiny. A WordNet for Latin was developed by Stefano Minozzi for the Fondazione Bruno Kessler's MultiWordNet Project (see Minozzi, 2008), consisting of about 9,000 lemmas tagged with synsets drawn from English and Italian. This is now being expanded through an international collaboration directed by the University of Exeter, to include over 70,000 words covering the archaic through classical periods of this language, as well as language-specific synsets defining meanings that are peculiar to Latin and not represented among the 100,000 or so synsets originally defined for English.

4 Innovations of design

Because the Latin WordNet (and indeed the WordNet specification generally) does not presently distinguish between literal and figurative sense attributions, it is being re-architected to accommodate the encoding of metonymic and metaphoric as well as literal senses of words. Annotation *at the level of the lemma* of specific sense (synset) assignments as being either literal, metonymic, or metaphorical is in fact one of the major new 'layers' at which figurative information is represented within the *Lexicon*. Consider, for example, the database entry for the word *baculum*, which can be accessible and marked-up by project participants through our bespoke on-line curation and annotation interface. In classical Latin, this word meant 'walking stick' and thus has been tagged with synset n#03585559, 'a stick carried in the hand for support in walking' as one of its literal senses (and indeed also its prototypical sense). Over time, however, and particularly in the early Christian period, the word came to be used more abstractly in the sense of any 'support' and in ecclesiastical texts regularly exhibits this meaning. This chronologically circumscribed figurative meaning of the word (n#04399253, 'something providing immaterial support or assistance to a person or cause or interest') is therefore annotated as a metaphorical sense. Differentiating between literal, metonymic, and metaphorical signification introduces an entirely new dimension of semantic structure into the WordNet framework, validated by modern linguistic theory.

Along with annotations at the level of lexical semantic structure distinguishing between a word's literal, metonymic, and metaphorical senses (represented by synsets), conceptual metaphors themselves will be coded as a relationship between synsets, understood as discrete concepts. For example, the FEAR IS A WEAPON metaphor, known in Latin in expressions such as the one in (1), is represented as a mapping between the synset that means 'fear' (n#05590260) and the one that means 'weapon' (n#03601056). In turn, the ANXIETY IS A SUBSTANCE metaphor, again illustrated by the passage in (1), is structured as a mapping between the synsets meaning 'anxiety' (n#04491326) and 'substance' (n#00010572), respectively.

1. *ipsius regis non tam subito pavore perculit pectus, quam anxiis inplevit curis* (LIV. 1, 56) 'As for the

king himself, his heart was not so much struck with sudden terror as filled with anxious forebodings'

Accordingly, any lemma annotated with one of these synsets as a literal, metonymic, or metaphorical sense is automatically linked (and accessible) via the metaphor by virtue of those sense attributions. In other words, as the theory posits, the metaphor operates as a *supralexical* structuring device of meaning in Latin: it helps determine, and motivate, the specific semantic developments of words and explains why the vocabulary of 'weapons' (not only the word corresponding to *weapon* but the whole conceptual domain relating to weapons and their use) can be used to talk about FEAR. Without the conceptual metaphor, there is no way to explain why WEAPON concepts are so regularly used to represent FEAR concepts and these would have to remain isolated, and – worse – arbitrary – facts of Latin's semantics. Crucially, moreover, the layer of more global conceptual-metaphorical information is tightly integrated with the more local layer of lexical-semantic information. In other words, the two layers of annotation – 1) the conceptual metaphor itself, as a mapping between synsets (concepts) and 2) the attribution of synsets to lemmas as specifically *metaphorical* senses – work hand in hand. When a lemma is tagged as 'having' a synset as one of its literal, metonymic, or metaphorical sense, the annotator is also able to indicate the specific metaphor that underpins the given sense.

This is to recognize within the relational structure of the database – and thus of the organization of Latin's semantic system – the theoretical claim that metaphors operate supra-lexically and provide motivating conceptual frameworks for the figurative extension of word meaning. In other words, rather than belonging to the semantic structure of any particular word (or determining, wholesale, the possible figurative meaning of a word), metaphors provide the *specific* pathways of figurative development that *specific* word senses may undergo in the course of a language's history. For instance, *baculum*'s metaphorical sense of 'something providing immaterial support or aid', would be tagged with the metaphor AN EMOTIONAL SUPPORT IS A PHYSICAL SUPPORT (or even more generally, THE EMOTIONAL IS THE PHYSICAL). This metaphor operates independently of this word's semantic structure – it very likely also determines the metaphorical usage of, e.g., *fulcio* – literally, 'to prop up' – in the sense of 'to uphold (emotionally)', as in CIC. Rab. 16, 43, *veterem amicum suum (. . .) labentem excepit, fulsit et sustinuit re, fortuna, fide* ('he supported his old friend – who was slipping downward – with his goods, his fortune and his confidence') – and so provides a powerful mechanism of bringing together otherwise disparate aspects of Latin's semantic system and discovering relationships that otherwise might remain hidden, obscured by outmoded principles of lexicographic organization.

Finally, the ability to organize metaphors into highly articulated networks or groupings via different kinds of mapping relations recognizes that, at a higher level of conceptual structure, metaphors participate in systems. Besides the relations mentioned above, another 'organizing' mechanism of metaphors is that of the image schema. In conceptual metaphor theory, an image schema is "a recurring dynamic pattern of our perceptual interactions and motor programs that gives coherence and structure to our experience" (Johnson 1987: xiv). Metaphorical mappings are usually encoded at a quite specific level of semantic granularity, and can be seen as detailed instantiations of more superordinate metaphors relying on general image schemas (e.g., FORCE, CONTAINER, OBJECT). In turn, mappings can give rise to further subordinate figurative patterns, with more semantic details filled in. These hierarchical relationships are all annotated within each metaphor record and give rise to a dense network of interconnected figurative meanings.

5 Annotation procedures and tagging scheme

Annotators first identify (a set of) documented metaphor(s) used by Latin writers to express an abstract concept, corresponding to a given synset, by analysing all occurrences of a relevant (set of) lemma(s) included in the synset within a selected corpus of literary texts. Encoding of metaphors, conceived as mappings between two synsets, is manually conducted through an annotation layer which has been designed expressly for this purpose. Very specifically, a metaphor is annotated according to its **status** (conventional, literary, or imaginative), **type** (primary, complex, orientational, ontological, one-shot image) and **period** of documentation. Moreover, it is labelled with a **shorthand expression** (e.g. 'ideas are food') and an **adjectival descriptor** (e.g. 'alimentary') following conventions in cognitive linguistics. The mapping itself is represented as a unidirectional relationship between two synsets, identified as the

source and **target**. Additional information includes relationships between two or more metaphors at higher or lower levels of semantic specificity, namely through **superordinate** and **subordinate** mappings. Annotators can also catalogue relations between mappings (e.g. extension, elaboration, reciprocity, derivation, combination, and entailment) that may characterize complex metaphor systems.

To exemplify this methodology, we present a case study of metaphor annotation pertaining to the semantic field of fear. A preliminary step identified synset n#05590260, 'an emotion experienced in anticipation of some specific pain or danger' (which pertains to five lemmas pointing to the concept of fear in Latin: *formido*, *metus*, *pavor*, *terror*, *timor*), as the primary target domain of the mapping. We scrutinized all occurrences of these lemmas (4,995) in the 'Antiquitas' section of the *Bibliotheca Teubneriana Latina* (3 BCE to 4 CE), distinguishing between literal (ex. 2) and figurative (ex. 3) usages. We counted but discarded literal usages, and further subclassified figurative usages into more fine-grained metaphorical subschemas.

2. *prae metu ubi sim nescio* (PLAUT. *Cas*. 413) 'I don't know where I am for fear'

3. *huic aliquem in pectus iniciam metum* (PLAUT. *Cas*. 589) 'I'll inject some fear into his heart'

Through careful analysis of the literal wording of the contexts in which these words appear, we identified 23 metaphorical mappings which instantiate three main superordinate image schemas, namely FORCE, CONTAINER, and OBJECT. An example of a metaphor actualizing the FORCE schema is FEAR IS A MILITARY FORCE (ex. 4); whereas FEAR IS A SUBSTANCE THAT FILLS THE EXPERIENCER (ex. 5) exemplifies the OBJECT schema.

4. *tum vero ingens metus nostros invadit* (SALL. *Iug*. 106, § 6) 'at last a great fear assailed the Romans'

5. *vidi hominem XIIII Kal. Febr. plenum formidinis* (CIC. *Att*. 9, 10) 'I saw him on January 17, thoroughly cowed [lit. filled up with]'

Once the catalogue of subschemas appeared to cover all possible metaphorical expressions involving the relevant lexical field, a generalized annotation template was used to record details about each mapping. For example, the annotation record for FEAR IS A MILITARY FORCE is as follows:

status ‹conventional›
type ‹ontological›
period Naev.+ ‹*Pun*. fr. 57, *magnae metus tumultus pectora possidit*›
shorthand expression ‹fear is a military force›
adjectival descriptor 'military'
source ‹n#06088783 | 'an opposing military force'›
target ‹n#05590260 | 'an emotion experienced in anticipation of some specific pain or danger'›
derives from ‹fear is a hostile force›

And it is annotated as follows in the *Lexicon* interface (Figure 1):

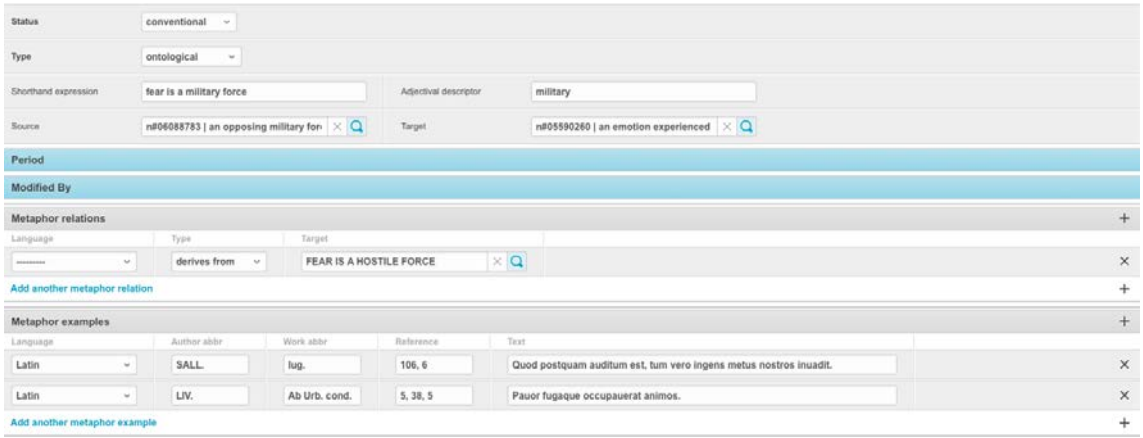

Figure 1. The annotation layer of the FEAR IS A MILITARY FORCE metaphor.

Finally, the metaphor entry is enriched with illustrative examples drawn from literature (ex. 6).

6. *olim iam adversus hunc metum emunivit animum* (SEN. *Con*. 3, 17, 10) 'but he has long since fortified his mind against fear of that'

According to this annotation procedure, users will be able to search the database using a variety of query types. For example, it will be possible to search for a single lemma (like *amor*), for a specific figurative source (like 'fire') or target domain ('love'), for an image schema (COUNTERFORCE), and thus to view all the metaphorical concepts built up from any of these elements. This will make it straightforward to discover certain features of figurative structuration within Latin's semantic system, such as the set of source domains that characterize the understanding of a given concept (what cognitive linguists called the 'range of the target') or, conversely, the set of target domains that are structured by a concept (the 'scope of the source'). It could also help shed light on the ways in which presumably human-universal aspects of cognition (sensorimotor gestalts) provide the scaffolding for culture-specific conceptualizations. What is more, because the metaphorical information of the *Lexicon Translaticium Latinum* piggybacks on the ontology provided by the WordNet, users will automatically be able to take advantage of the rich lexical and semantic knowledge already present in this database, enabling highly complex figuratively-aware queries. The *Lexicon* therefore portends to have significant implications for corpus search, text-processing and other natural language understanding applications.

6 Conclusions

The theoretical and methodological underpinnings of this project, along with the practical annotation procedure it has implemented, suggest that the *Lexicon Translaticium Latinum* could contribute significantly not only to cognitive and semantic approaches and to metaphor theory, but also to linguistic, literary, and cultural research in Classical Studies, especially as part of this field's wider ecosystem of natural language understanding applications. Indeed, we hope to position the Lexicon not merely as a repository of figurative usages in Latin, but as an interface to the system of knowledge itself that Latin speakers relied upon in thinking and speaking in diverse contexts of symbolic expression, and thus as a resource for better understanding how members of Roman society 'made sense' in, and of, their world.

References

Vyvyan Evans and Melanie C. Green. 2006. *Cognitive Linguistics: an Introduction*. Edinburgh University Press, Edinburgh.

Chiara Fedriani. 2016. Ontological and Orientational Metaphors in Latin: Evidence from the Semantics of Feelings and Emotions. In William M. Short (ed.), *Embodiment in Latin Semantics*, 115–140. John Benjamins, Amsterdam.

Dirk Geeraerts and Hubert Cuyckens (eds). 2010. *The Oxford Handbook of Cognitive Linguistics*. Oxford University Press, Oxford.

Raymond W. Gibbs. 2005. *Embodiment and Cognitive Science*. Cambridge University Press, Cambridge.

Evelyn Gius and Janina Jacke. 2017. The Hermeneutic Profit of Annotation: On Preventing and Fostering Disagreement in Literary Analysis. *International Journal of Humanities and Arts Computing* 11(2):233–254.

Mark Johnson. 1987. *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. University of Chicago Press, Chicago.

George Lakoff. 1987. *Women, Fire and Dangerous Things: What Categories Reveal about the Mind*. University of Chicago Press, Chicago.

Jerome McGann. 2016. Marking Texts of Many Dimensions. In Susan Schreibman, Ray Siemens, and John Unsworth (eds.), *A New Companion to Digital Humanities*, 358–376. Wiley Blackwell, Malden, MA.

Stefano Minozzi. 2008. La costruzione di una base di conoscenza lessicale per la lingua latina: Latin-Wordnet. In Giuseppe Sandrini (ed.), *Studi in Onore di Gilberto Lonardi*, 243–258. Fiorini, Verona.

Egle Mocciaro and William M. Short (eds.). 2019. *Towards a Cognitive Classical Linguistics: The Embodied Basis of Constructions in Greek and Latin*. De Gruyter, Berlin.

Samuel D. Neill. 1987. The Dilemma of the Subjective in Information Organisation and Retrieval. *Journal of Documentation* 43(3):193–211.

Douglas Raber and John M. Budd. 2003. Information as Sign: Semiotics and Information Science. *Journal of Documentation* 59(5):507–522.

William M. Short. 2008. Thinking Places, Placing Thoughts: Spatial Metaphors of Mental Activity in Roman Culture. *Quaderni del Ramo d'Oro* 1:106–29.

William M. Short. 2013. 'Transmission' Accomplished? Latin's Alimentary Metaphors of Communication. *American Journal of Philology* 134(2):247-275.

William M. Short (ed.). 2016. *Embodiment in Latin Semantics*. John Benjamins, Amsterdam.