

# NUNC EST AESTIMANDUM

## TOWARDS AN EVALUATION OF THE LATIN WORDNET

Franzini, G.\*, Peverelli, A.\*, Ruffolo, P.\*, Passarotti, M.\*, Sanna, H.^\, Signoroni, E.^\, Ventura, V.^\, Zampedri, F.^\

\*CIRCSE Research Centre, Università Cattolica del Sacro Cuore - Milan, Italy

^\Università degli Studi di Pavia - Pavia, Italy

### INTRODUCTION

Since the release of the Princeton WordNet, interest in providing WordNets for modern languages has far exceeded that for historical languages. The only two historical language WordNets in existence today are the Latin and the Ancient Greek WordNets, both of which have quantitative and qualitative limitations.

The study described here falls within the scope of the LiLa: Linking Latin project. In its wider effort to connect linguistic resources and NLP tools for Latin in a Linked Data Knowledge Base, LiLa is conducting a first assessment of the Latin WordNet.

### THE LATIN WORDNET

The LWN was first created in 2004 from MultiWordNet (MWN) following the *Expand Method*. It comprises 9,378 lemmas distributed across 8,973 synsets. The criteria behind the selection of LWN lemmas remain unclear, and there are some noticeable gaps, both lexical (*amo, amare*, 'to love') and relational (the adjectives *inaequalis* 'unequal' and *aequalis* 'equal' are placed in a relation of derivation only but could also count as antonyms). Moreover, as shown in Table 1, the Expand Method made incorrect sense associations.

Table 1: Example senses/synsets to be removed from LWN.

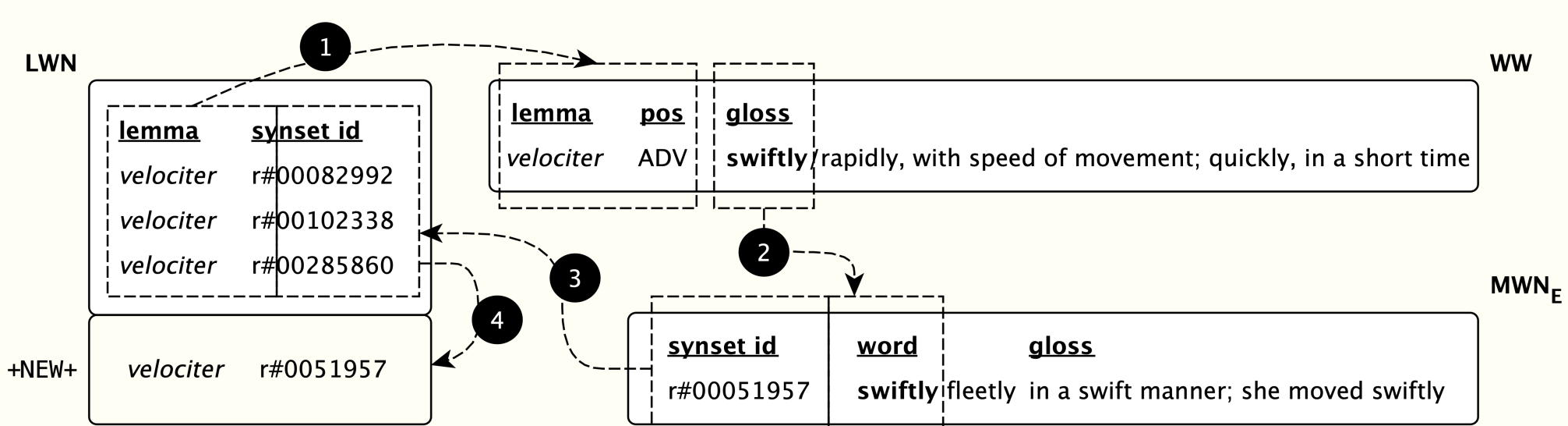
lemma	synset ID	definition
ager	n#W0021124	in un database, ogni area in cui vengono registrate le singole informazioni che compongono il record
capitolium	n#06188340	the federal government of the United States
genetrix	n#W0021113	titolo e appellativo che si dà alle suore professe o a quelle che hanno cariche particolari: sono venuta a fare atto d'obbedienza alla madre badessa di questo convegno
voco	v#00720710	send a message or attempt to reach someone by radio, phone, etc; make a signal in order to transmit a message

### EXPANSION AND EVALUATION METHOD

We formulated a first experiment combining a small, automated extension of the sense coverage of 100 lemmas (25 per PoS) in the LWN with a follow-up manual revision of their corresponding synsets. We formalised a rudimentary algorithm to automatically recommend new synsets for these 100 lemmas using the English data in MWN (MWN<sub>E</sub>) and the Whitaker's Words (WW) English-Latin dictionary as a reference. Figure 1 exemplifies the algorithmic process: for the LWN adverb *velociter* 'swiftly, quickly', the algorithm:

1. searched for joint lemma and PoS overlaps between LWN and WW;
2. where there was a match, it then looked for overlaps between the single-word WW glosses and MWN<sub>E</sub> lemmas;
3. where these also matched, it checked the lemma's corresponding synset(s) in MWN<sub>E</sub> for that PoS against existing LWN synsets to...
4. ...label machine recommendations as NEW (machine-suggested and not already present in LWN), COM (for "common", i.e., machine-suggested but already present in LWN) or OLD (synsets present in the LWN only).

Figure 1. Synset recommender system.



This process resulted in 3,746 machine-suggested lemma-synset entries to be manually evaluated. The evaluation was performed by five raters. We measured inter-rater reliability (IRR) using percentage agreement without chance correction. Lower agreements are not a reflection of the raters' inability to distinguish word meanings but, rather, of their difficulty in selecting the synsets that best fit their subjective opinion.

### RESULTS

As expected, our synset recommender system produced many false positives, with only 0.18% machine suggestions approved by all five raters. Even if the precision of the synset-recommendation algorithm were to be improved, recall would likely still be high due to the unavoidable assignment of modern senses to a historical resource.

Table 2 provides minimum (*m\_v*), maximum (*M\_v*) and average values of agreement (*A\_v*) per type of synset assignment as well as standard deviations (*S\_v*). The average values reveal that almost 1/3 of all synsets was not reliably rated.

Table 2: Inter-rater agreement values grouped by type of synset assignment.

type	n	<i>m_v</i>	<i>M_v</i>	<i>A_v</i>	<i>S_v</i>
OLD	35	0.200	1.000	0.691	0.345
COM	876	0.200	1.000	0.654	0.320
NEW	2835	0.200	1.000	0.702	0.329

Low agreement values might be caused by incorrect assignments or, more problematically, differences of opinion on subtle semantic differences. A close examination of the data, and, specifically, of the adverbs with agreement values below 60% (6 out of 25), points to the latter.

Synsets carrying temporal meanings tend to show lower agreement rates than those associated with space. The higher agreement rate on the spatial dimension resonates with cognitive linguistic theories on spatial semantics, according to which the concreteness of space over temporal or more abstract meanings induces us to map its structure onto other semantic domains.

### CURRENT WORK

Manual removal of modern senses inherited from MWN for a clean LWN.

### FUTURE WORK

1. Attach the clean LWN to the LiLa Knowledge Base of linguistic resources for Latin (Figure 2).
2. Extend LWN by extracting hypernyms, synonyms and bags of words from dictionary definitions, as well as lemma groups from three Latin synonym dictionaries. If combined, these dictionaries can supply the LWN with some 1,050 additional lemmas.

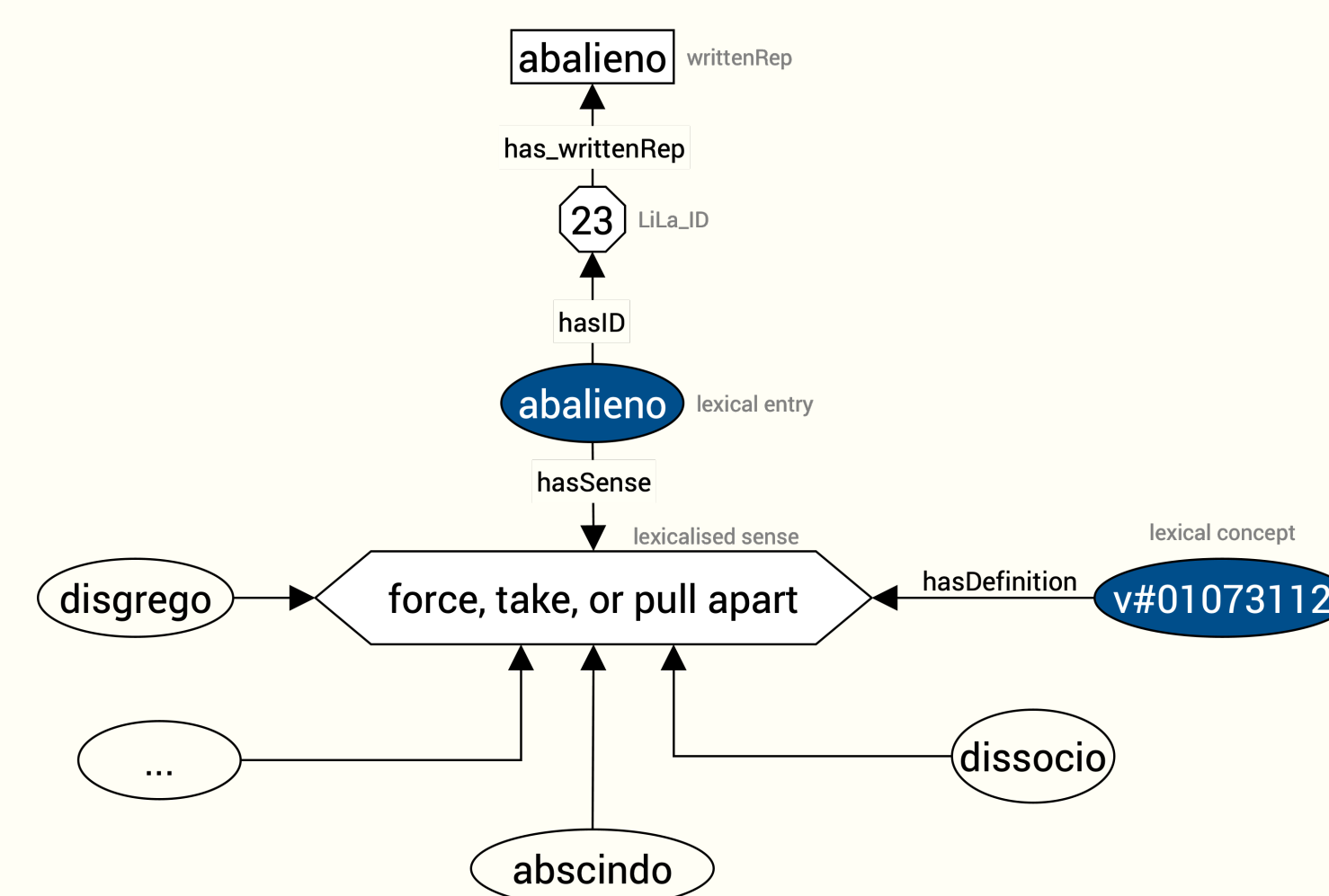


Figure 2. Ontological model of the Latin WordNet for Linked Data purposes (LiLa Knowledge Base).

### CONTACT

greta.franzini@unicatt.it  
https://lila-erc.eu

### ACKNOWLEDGEMENTS

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme – Grant Agreement No. 769994.



Sixth Italian Conference on Computational Linguistics (CLiC-it 2019)  
13-15 November 2019 - Bari, Italy

