Of Nodes and Cells.

Two Perspectives on (and from) *Word Formation Latin*

Marco Budassi, **Eleonora Litta** and **Marco Passarotti**

ICLL 2019 | Universidad de Las Palmas de Gran Canaria | 19 June 2019

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

Word Formation Latin

Some background

WFL: Word formation-based lexicon for Classical Latin

Word Formation Latin

Some background

WFL: Word formation-based lexicon for Classical Latin

- Lexical base => **Lemlat**: morphological analyser and lemmatiser for Classical and Late Latin (OLD, Georges, Gradenwitz)

Word Formation Latin

Some background

WFL: Word formation-based lexicon for Classical Latin

- Lexical base => **Lemlat**: morphological analyser and lemmatiser for Classical and Late Latin (OLD, Georges, Gradenwitz)
- WFRs are modelled as **directed one-to-many input-output** relations between lemmas (based on I&A models of grammatical description)

Word Formation Latin

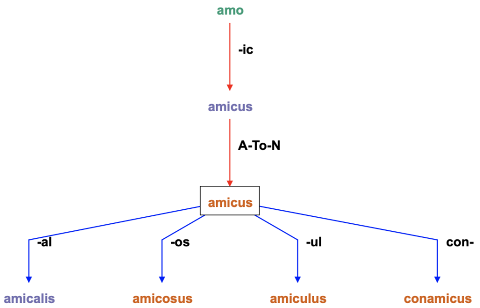
Some background

WFL: Word formation-based lexicon for Classical Latin

- Lexical base => **Lemlat**: morphological analyser and lemmatiser for Classical and Late Latin (OLD, Georges, Gradenwitz)
- WFRs are modelled as **directed one-to-many input-output** relations between lemmas (based on I&A models of grammatical description)
- **Morphotactic** approach: each WF process is treated individually as the application of one single rule in a certain order.

WFL Online

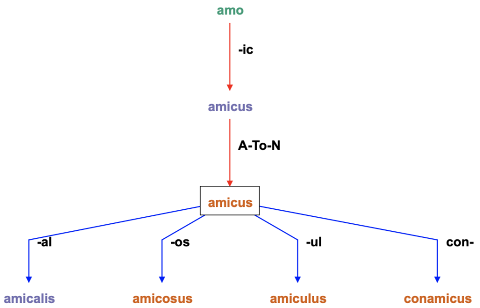
https://wfl.marginalia.it



- Relationships between lemmas of the same "word formation family" are represented as the edges in a **directed graph** with a **hierarchical tree-like** structure

WFL Online

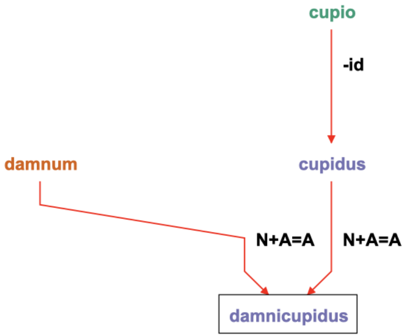
https://wfl.marginalia.it



- Relationships between lemmas of the same “word formation family” are represented as the edges in a **directed graph** with a **hierarchical tree-like** structure
- A **node** is a lemma, and an **edge** is the WFR used to derive the output lemma from the input one, together with any affix

WFL Online

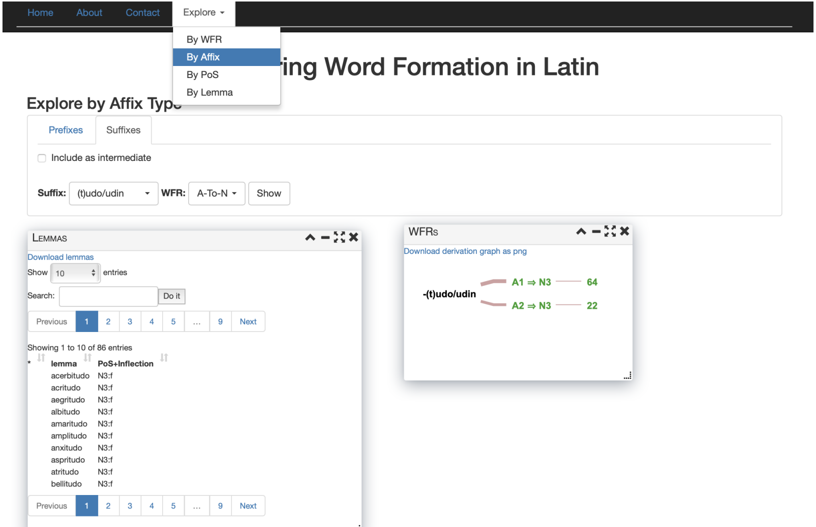
https://wfl.marginalia.it



Compounding is also shown as an intersection between word formation families.

WFL Online

https://wfl.marginalia.it



…ing Word Formation in Latin

Explore by Affix Type

Prefixes | Suffixes

□ Include as intermediate

Suffix: (t)udo/udin **WFR:** A-To-N Show

LEMMAS

Download lemmas

Show 10 entries

Search: Do it

Previous 1 2 3 4 5 … 9 Next

Showing 1 to 10 of 86 entries

*	lemma	PoS+Inflection
	acerbitudo	N3:f
	acritudo	N3:f
	aegritudo	N3:f
	albitudo	N3:f
	amaritudo	N3:f
	amplitudo	N3:f
	anxitudo	N3:f
	aspritudo	N3:f
	atritudo	N3:f
	bellitudo	N3:f

Previous 1 2 3 4 5 … 9 Next

WFRs

Download derivation graph as png

WFL I&A

Problems

But: **directed graphs** are not completely satisfactory in representing the full range of relationships included within a word formation family.

WFL I&A

Problems

But: **directed graphs** are not completely satisfactory in representing the full range of relationships included within a word formation family.

Main problems:

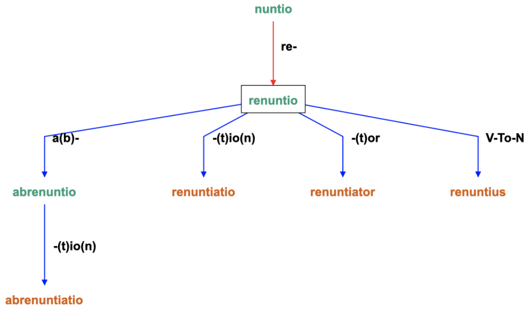
WFL I&A

Problems

But: **directed graphs** are not completely satisfactory in representing the full range of relationships included within a word formation family.

Main problems:

- Directionality

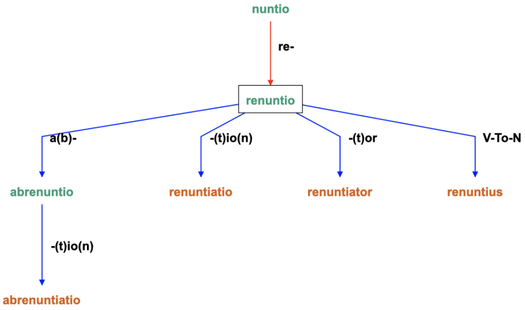


WFL I&A

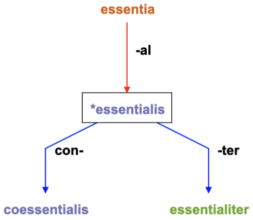
Problems

But: **directed graphs** are not completely satisfactory in representing the full range of relationships included within a word formation family.
Main problems:

- Directionality



- Non-linear derivations



Paradigmatic approach to WF: Advantages

Paradigmatic approach to WF: Advantages

The appeal of a **derivational paradigm** is the possibility of stacking recurring word formations into a model, carrying descriptive and predictive powers.

Paradigmatic approach to WF: Advantages

The appeal of a **derivational paradigm** is the possibility of stacking recurring word formations into a model, carrying descriptive and predictive powers.

- Relationships are not limited to base-derivative pairs (Jackendoff, 1975) => Non linear derivations

Paradigmatic approach to WF: Advantages

The appeal of a **derivational paradigm** is the possibility of stacking recurring word formations into a model, carrying descriptive and predictive powers.

- Relationships are not limited to base-derivative pairs (Jackendoff, 1975) => Non linear derivations
- No need for directionality (Jackendoff, 1975) => Diachronic discrepancies

Paradigmatic approach to WF: Advantages

The appeal of a **derivational paradigm** is the possibility of stacking recurring word formations into a model, carrying descriptive and predictive powers.

- Relationships are not limited to base-derivative pairs (Jackendoff, 1975) => Non linear derivations
- No need for directionality (Jackendoff, 1975) => Diachronic discrepancies
- Potentiality (Štekauer: 2014) and Availability of Slots more important than the form filling them (Bauer, 1997)

Paradigmatic approach to WF: Advantages

The appeal of a **derivational paradigm** is the possibility of stacking recurring word formations into a model, carrying descriptive and predictive powers.

- Relationships are not limited to base-derivative pairs (Jackendoff, 1975) => Non linear derivations
- No need for directionality (Jackendoff, 1975) => Diachronic discrepancies
- Potentiality (Štekauer: 2014) and Availability of Slots more important than the form filling them (Bauer, 1997)
- Regularity and predictability (Bauer, 1997)

Paradigmatic approach to WF: Advantages

The appeal of a **derivational paradigm** is the possibility of stacking recurring word formations into a model, carrying descriptive and predictive powers.

- Relationships are not limited to base-derivative pairs (Jackendoff, 1975) => Non linear derivations
- No need for directionality (Jackendoff, 1975) => Diachronic discrepancies
- Potentiality (Štekauer: 2014) and Availability of Slots more important than the form filling them (Bauer, 1997)
- Regularity and predictability (Bauer, 1997)

BUT… not many examples of fuller paradigms (in any language!).

Derivational Paradigms: Requirements

Derivational Paradigms: Requirements

- No directionality: necessary to accommodate those lemmas for which the derivational process is not of the simplex (or simpler) > complex type

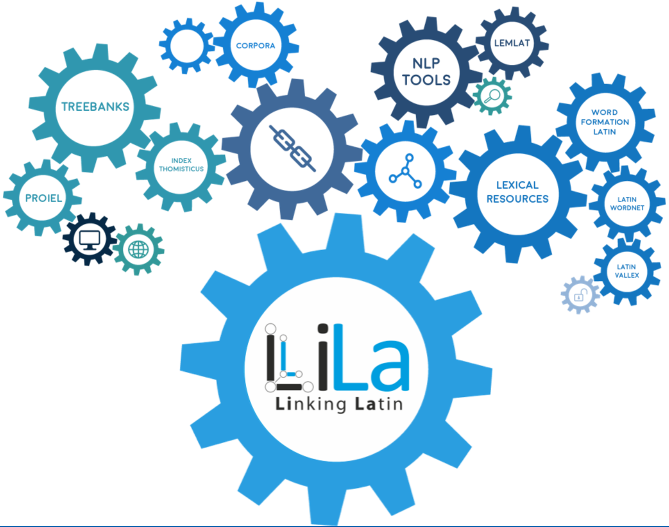
Derivational Paradigms: Requirements

- No directionality: necessary to accommodate those lemmas for which the derivational process is not of the simplex (or simpler) > complex type
- The CELL has a central role in the paradigm (predictability and regularity)

Derivational Paradigms: Requirements

- No directionality: necessary to accommodate those lemmas for which the derivational process is not of the simplex (or simpler) > complex type
- The CELL has a central role in the paradigm (predictability and regularity)
- Each cell must be described in both its morphological characteristics and its semantic features, due to the underlying role of semantics in accounting for derivational processes

LiLa: Linking Latin



LiLa: Linking Latin

The LiLa project: Knowledge Base of Linguistic Resources and Natural Language Processing (NLP) tools for Latin.

LiLa: Linking Latin

The LiLa project: Knowledge Base of Linguistic Resources and Natural Language Processing (NLP) tools for Latin. LiLa collects and connects:

- Linguistic Resources: corpora, lexica, ontologies, dictionaries, thesauri

LiLa: Linking Latin

The LiLa project: Knowledge Base of Linguistic Resources and Natural Language Processing (NLP) tools for Latin. LiLa collects and connects:

- Linguistic Resources: corpora, lexica, ontologies, dictionaries, thesauri
- NLP tools: tokenisers, lemmatisers, PoS-taggers, morphological analysers and dependency parsers

LiLa: Linking Latin

The LiLa project: Knowledge Base of Linguistic Resources and Natural Language Processing (NLP) tools for Latin. LiLa collects and connects:

- Linguistic Resources: corpora, lexica, ontologies, dictionaries, thesauri
- NLP tools: tokenisers, lemmatisers, PoS-taggers, morphological analysers and dependency parsers

The Knowledge Base consists of different kinds of objects connected via an explicitly-declared vocabulary for knowledge description.

The LiLa Knowledge Base

Conceptual and Structural Interoperability

LiLa is based on an ontology made of:

LiLa is based on an ontology made of:

- **Individuals**: instances of objects (one specific token, lemma etc.)

The LiLa Knowledge Base

Conceptual and Structural Interoperability

LiLa is based on an ontology made of:

- **Individuals**: instances of objects (one specific token, lemma etc.)
- **Classes**: types of objects/concepts (token, lemma, PoS etc.)

The LiLa Knowledge Base
Conceptual and Structural Interoperability

LiLa is based on an ontology made of:

- **Individuals**: instances of objects (one specific token, lemma etc.)
- **Classes**: types of objects/concepts (token, lemma, PoS etc.)
- **Attributes**: properties that objects can/must have (morphological features for lemmas/tokens). Each attribute can be a class or an individual

The LiLa Knowledge Base

Conceptual and Structural Interoperability

LiLa is based on an ontology made of:

- **Individuals**: instances of objects (one specific token, lemma etc.)
- **Classes**: types of objects/concepts (token, lemma, PoS etc.)
- **Attributes**: properties that objects can/must have (morphological features for lemmas/tokens). Each attribute can be a class or an individual
- **Relations**: ways in which classes and individuals can be related to one another: RDF triples. Labels from a restricted dictionary of knowledge description: *has_lemma*, *has_PoS*

Word Formation in LiLa

Different approach to Word Formation:

Word Formation in LiLa

Different approach to Word Formation:

- Structure: **declarative** rather than procedural

Word Formation in LiLa

Different approach to Word Formation:

- Structure: **declarative** rather than procedural
- No directionality

Word Formation in LiLa

Different approach to Word Formation:

- Structure: **declarative** rather than procedural
- No directionality
- No morphotaxis.

Construction Morphology (CxM)

Most unassuming, declarative way of describing word formation processes.

Construction Morphology (CxM)

Most unassuming, declarative way of describing word formation processes.

- Construction: [co(n) [stell](a)(t)io]$_N$ (more specific)

Construction Morphology (CxM)

Most unassuming, declarative way of describing word formation processes.

- Construction: [co(n) [stell](a)(t)io]$_N$ (more specific)
- Schema [co(n)[x](t)io]$_N$ (more generalised)

Construction Morphology (CxM)

Most unassuming, declarative way of describing word formation processes.

- Construction: [co(n) [stell](a)(t)io]$_N$ (more specific)
- Schema [co(n)[x](t)io]$_N$ (more generalised)
- Constructions and schemas are word-based and declarative

Construction Morphology (CxM)

Most unassuming, declarative way of describing word formation processes.

- Construction: [co(n) [stell](a)(t)io]$_N$ (more specific)
- Schema [co(n)[x](t)io]$_N$ (more generalised)
- Constructions and schemas are word-based and declarative
- Perfect for LiLa => words are described in their formative elements, which can be organised into connected classes of objects into an ontology.

Word formation in LiLa

Three classes of objects:

Word formation in LiLa

Three classes of objects:

1. Lemmas

Word formation in LiLa

Three classes of objects:

1. Lemmas
2. Affixes (prefixes and suffixes)

Word formation in LiLa

Three classes of objects:

1. Lemmas
2. Affixes (prefixes and suffixes)
3. Bases (connectors between lemmas of the same WF family)

Word formation in LiLa

Three classes of objects:

1. Lemmas
2. Affixes (prefixes and suffixes)
3. Bases (connectors between lemmas of the same WF family)

Connected by three possibile relationships:

Word formation in LiLa

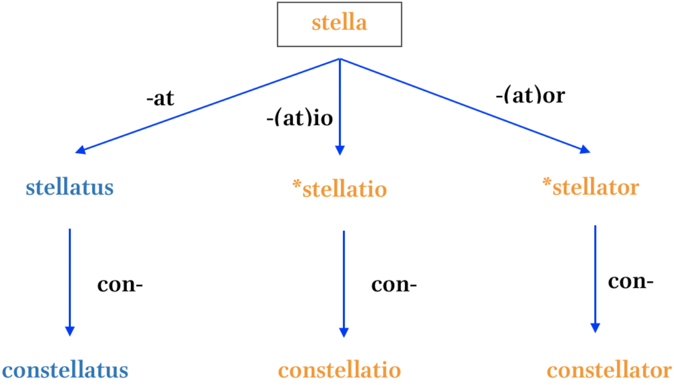
Three classes of objects:

1. Lemmas
2. Affixes (prefixes and suffixes)
3. Bases (connectors between lemmas of the same WF family)

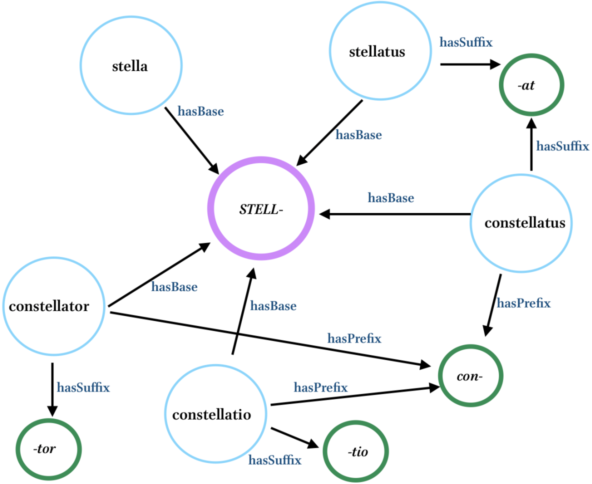
Connected by three possibile relationships:

1. hasPrefix
2. hasSuffix
3. hasBase

Stella - WFL


stella
-at
-(at)io
-(at)or
stellatus
*stellatio
*stellator
con-
con-
con-
constellatus
constellatio
constellator

STELL - LiLa


stella
stellatus
hasSuffix
-at
hasBase
hasBase
hasSuffix
STELL-
hasBase
constellatus
constellator
hasBase
hasBase
hasPrefix
hasPrefix
con-
hasSuffix
constellatio
hasPrefix
-tor
hasSuffix
-tio

WFL in LiLa

Prototype

LiLa triplestore available at:
https://lila-erc.eu/data/

More sophisticated ways of querying the data:

WFL in LiLa

Prototype

LiLa triplestore available at:
https://lila-erc.eu/data/

More sophisticated ways of querying the data:

- Start from a lemma (e.g. *formalis*), find its lexical base, retrieve all lemmas and affixes linked to that lexical base (word formation family)

WFL in LiLa

Prototype

LiLa triplestore available at:
https://lila-erc.eu/data/

More sophisticated ways of querying the data:

- Start from a lemma (e.g. *formalis*), find its lexical base, retrieve all lemmas and affixes linked to that lexical base (word formation family)
- Find lemmas with a specific prefix (ad-) or suffix (-bil)

WFL in LiLa

Prototype

LiLa triplestore available at:
https://lila-erc.eu/data/

More sophisticated ways of querying the data:

- Start from a lemma (e.g. *formalis*), find its lexical base, retrieve all lemmas and affixes linked to that lexical base (word formation family)
- Find lemmas with a specific prefix (ad-) or suffix (-bil)
- List distinct lemmas with suffix -bil, prefix ad- ending in -bilis only

WFL in LiLa
Prototype

LiLa triplestore available at:
https://lila-erc.eu/data/

More sophisticated ways of querying the data:

- Start from a lemma (e.g. *formalis*), find its lexical base, retrieve all lemmas and affixes linked to that lexical base (word formation family)
- Find lemmas with a specific prefix (ad-) or suffix (-bil)
- List distinct lemmas with suffix -bil, prefix ad- ending in -bilis only
- Count the frequency of the 15 most used affixes attached to nouns.

Your Questions

Do you have research questions?
Contact me at eleonoramaria.litta@unicatt.it

Conclusion and Future Plans

Summing up

Conclusions:

- WP models have a better way of demonstrating derivational relationships between words
- CxM was useful to develop a new model to account for Latin word formation that could fit into the LiLa Knowledge Base
- WFL in LiLa does not contain info on directionality.

Future plans:

- Find a way of defining and naming all "base" nodes
- More investigation is needed to add word formation specific semantic information to the LiLa knowledge base
- Enlarge the lexical basis for which WF is provided to Medieval Latin lemmas contained in Lemlat.

Credits and acknowledgements

Thanks to…

- Marco Pappalepore (for extracting triples from the old WFL)
- Francesco Mambrini (for creating the ontology behind the new WFL)

Thanks!

Get in touch

Eleonora Litta

Università Cattolica del Sacro Cuore, Milano

✉ eleonoramaria.litta@unicatt.it

https://github.com/CIRCSE/WFL

https://wfl.marginalia.it

Largo Gemelli 1, 20123 Milan, Italy

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