UNIVERSITÀ DI PARMA

NM STATE UNIVERSITY

UNIVERSITÀ DEGLI STUDI DI UDINE
hic sunt futura

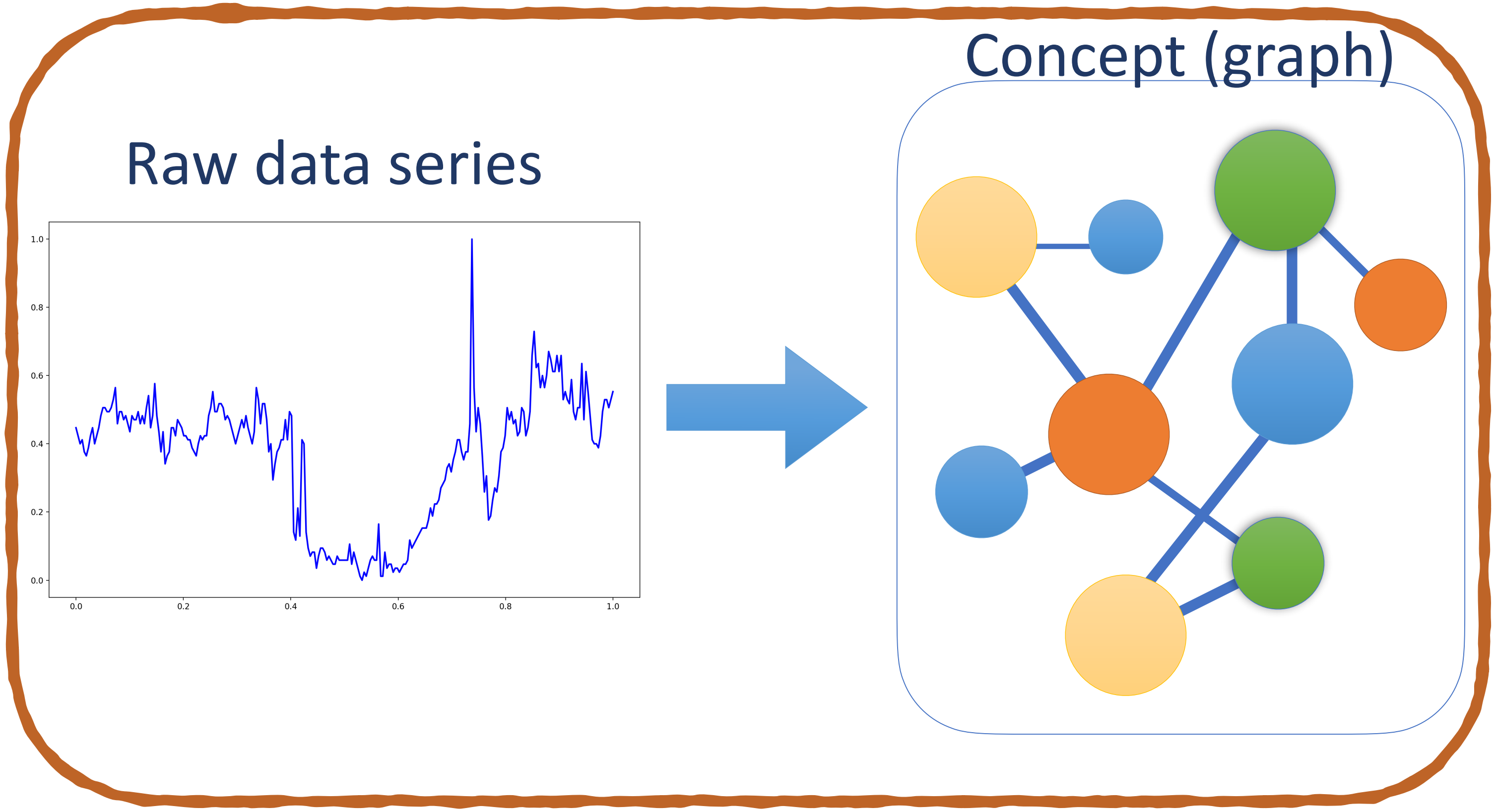
Concept2Text
an explainable multilingual rewriting of concepts into natural language

F. Bertini, A. Dal Palù, F. Fabiano, A. Formisano, F. Zaglio
UniPR + UniUD + NMSU

28/06/2024

Data 2 Concept (prev work)

- Identify emerging properties and represent them into concepts



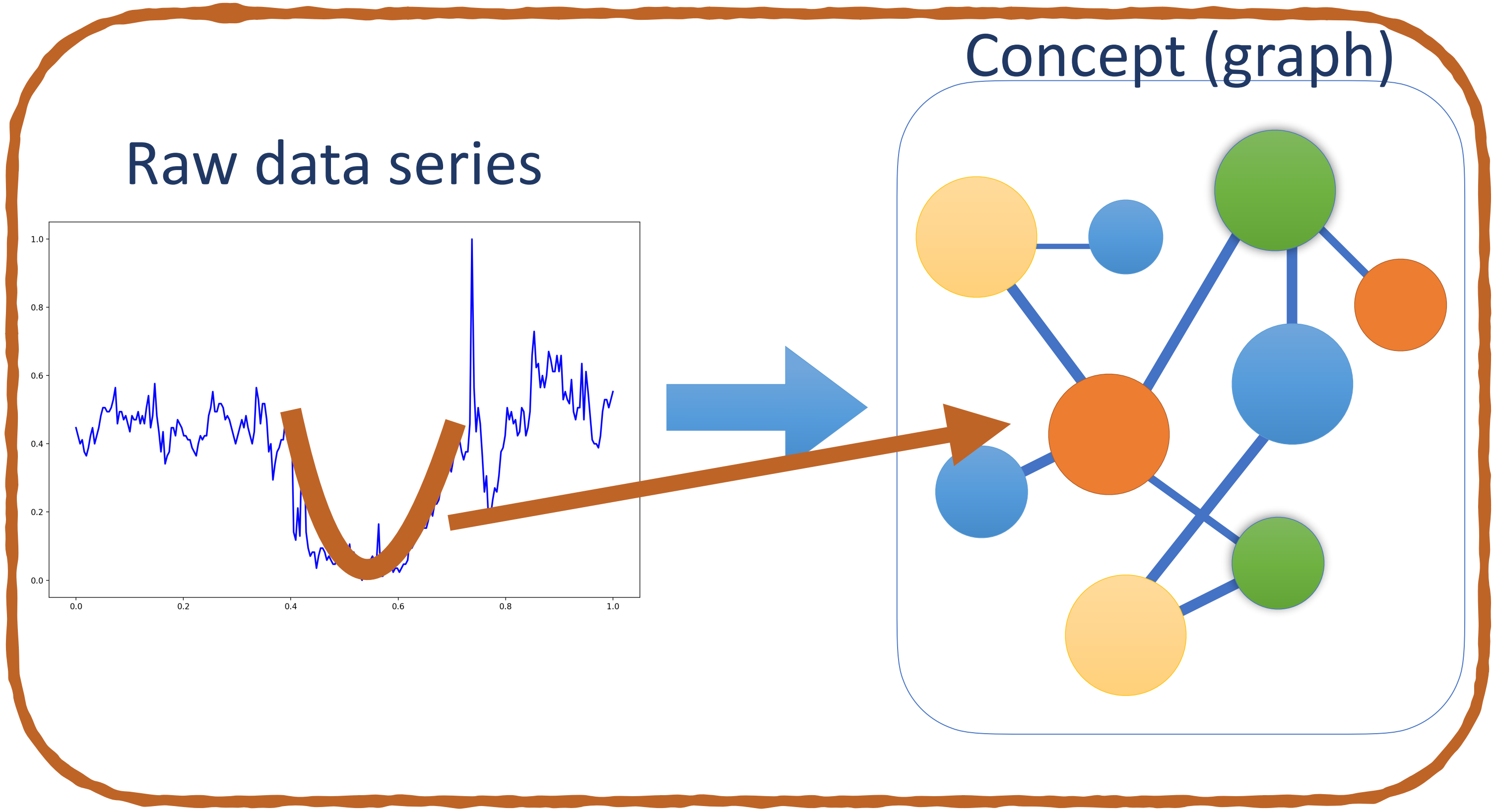
Raw data series

Concept (graph)

CILC23, ICLP23

Data 2 Concept (prev work)

- Identify emerging properties and represent them into concepts (ASP based)



Raw data series

Concept (graph)

- fact(valley,

  from(10),

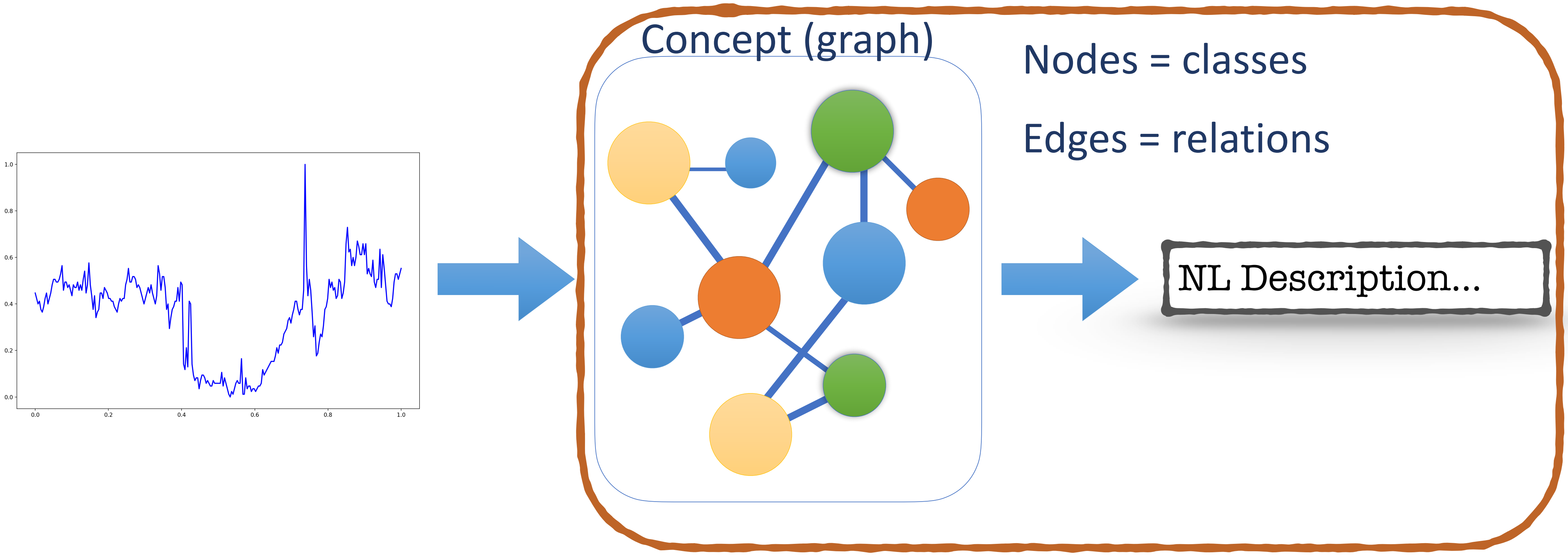
  to(20),

  depth(40))

CILC23, ICLP23

Concept2Text: an explainable multilingual rewriting of concepts into natural language

Concept 2 Text



Concept (graph)

Nodes = classes

Edges = relations

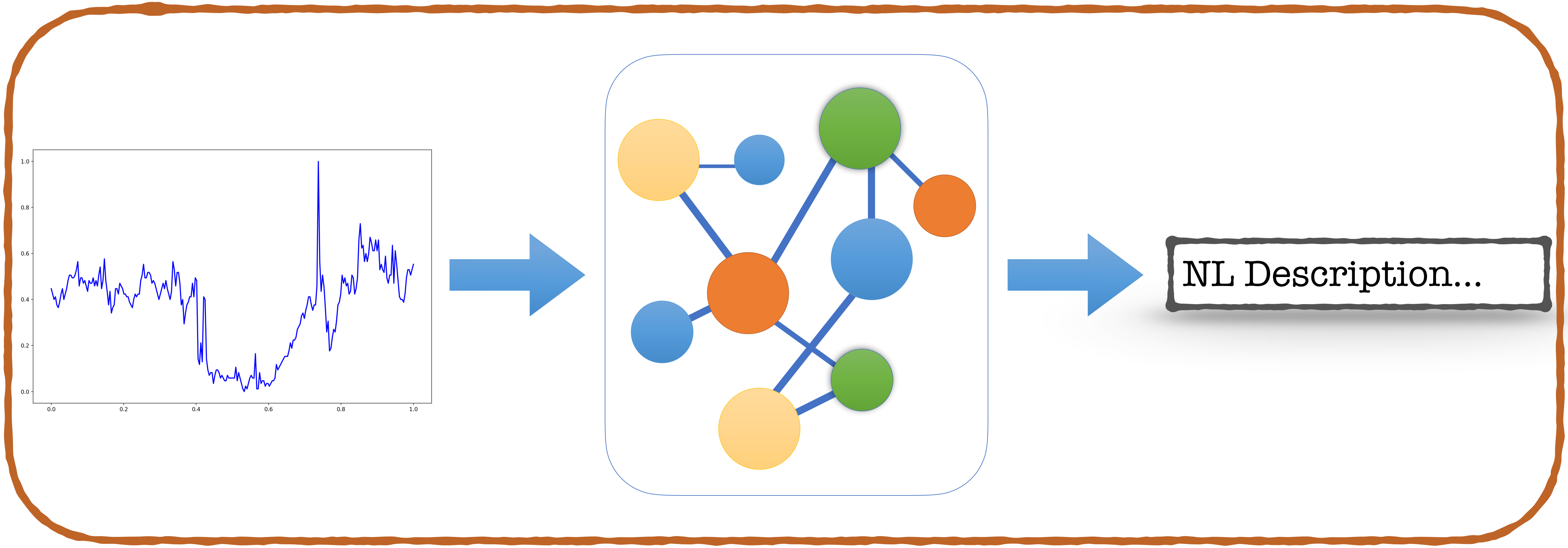
NL Description...

CILC24

Concept2Text: an explainable multilingual rewriting of concepts into natural language

Data 2 Text

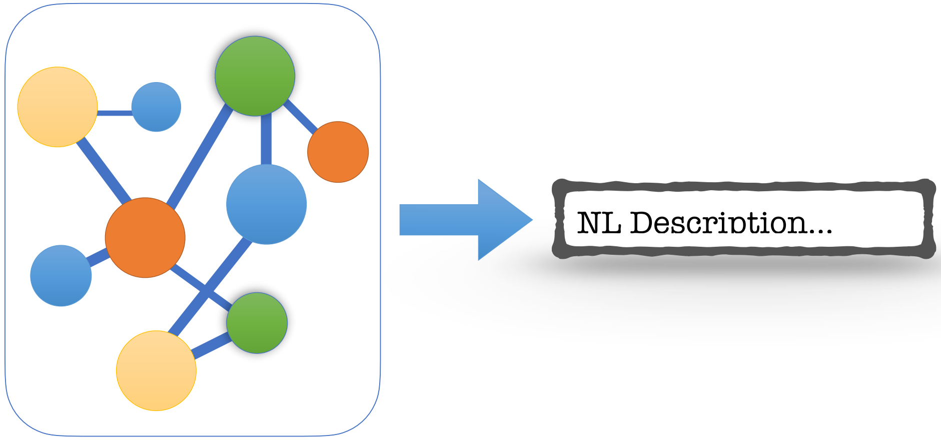
- Complete xAI pipeline for describing a set of raw data in Natural Language



NL Description...

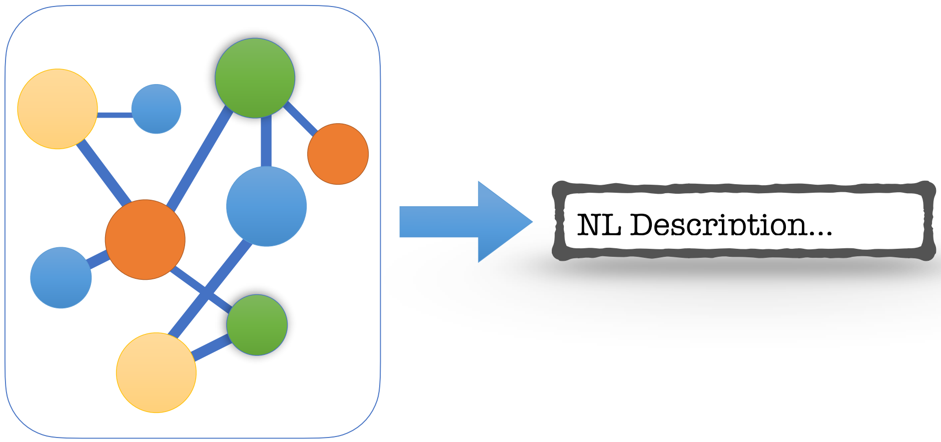
Concept2Text

- Concepts can be represented by a graph:
  - **Ontologies (classes + hyper-relations)**
  - **Domain specific concepts** from data **(our Data2Concept approach)**
  - Merge of both graphs



Explainable Concept2Text

- Why explainability?
- **Data Analysis + Large Language Models**
    - Black box (no high risk applications in EU)
    - No control on handling of concept/relations towards text
    - Summarization, bias, accuracy, reliability…
- **LP based**
    - Explicit model
    - Flexibility and modularity
    - Rather rich expressivity

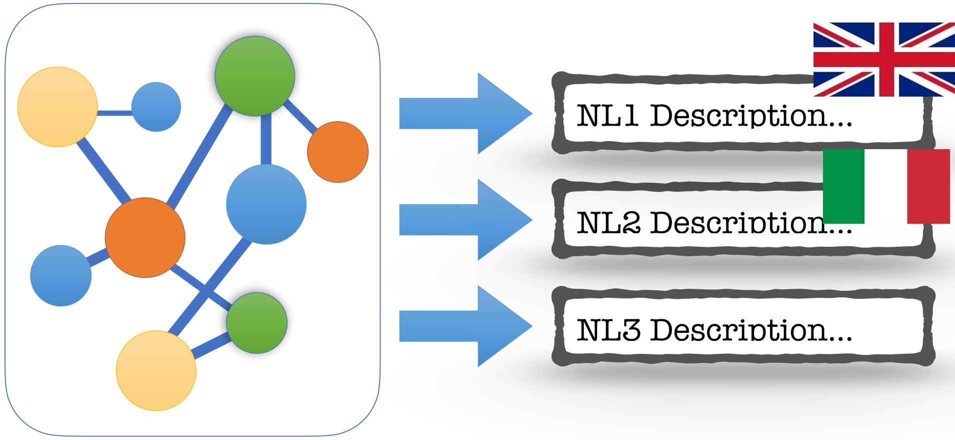


NL Description…

Concept2Text: an explainable multilingual rewriting of concepts into natural language

Multilingual Concept2Text

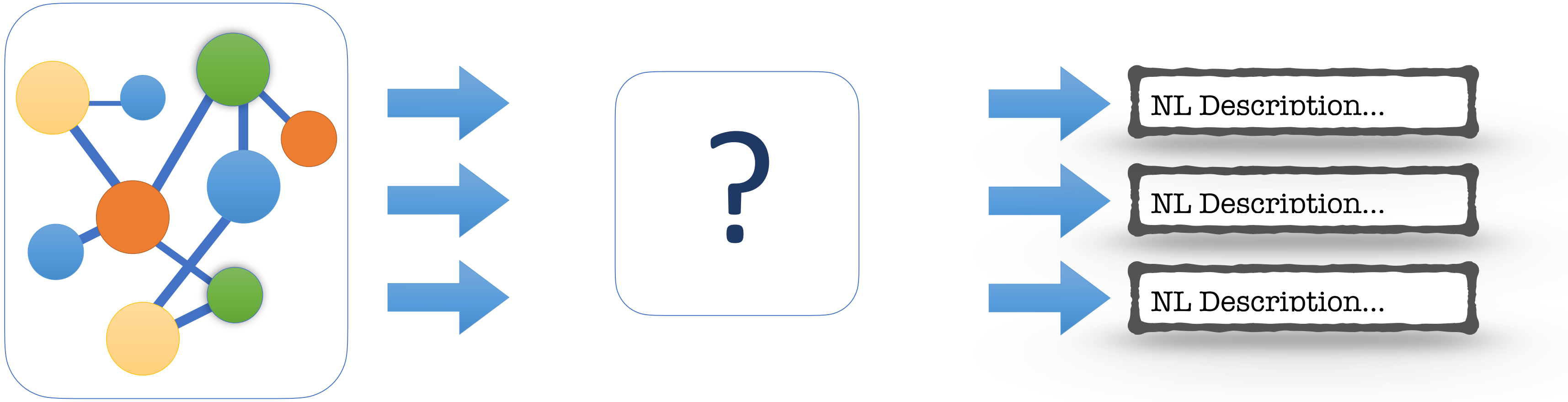
- Why multilingual?
- Concepts are language independent
- Different target languages
  - No language2language translations
  - No ambiguity at concept level
  - Concept2language rules
- Shared and Modular approach
- Variants (LLM-like *safe* richness)



Concept2Text: an explainable multilingual rewriting of concepts into natural language

Concept2Text in Prolog

- Typical NLP application: from text to abstraction (parsing, semantics…)
- **Little literature about concept2text with LP (‘80s [Pereira])**
- Definite Clause Grammars (Prolog extension for BNF grammars)? Too simple
    - Used for creation/parsing of a single grammar tree
    - How to handle the translation of concept —> BNF grammar?



?

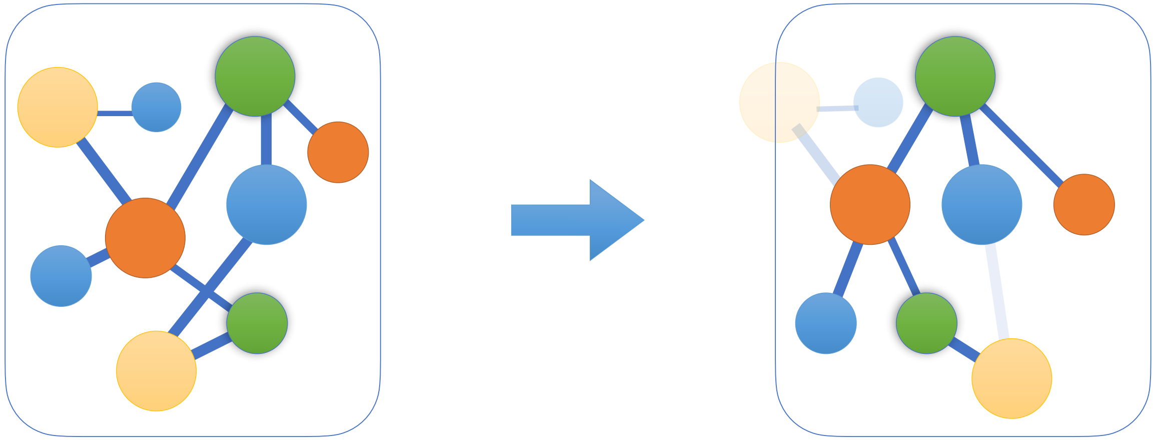
NL Description…

NL Description…

NL Description…

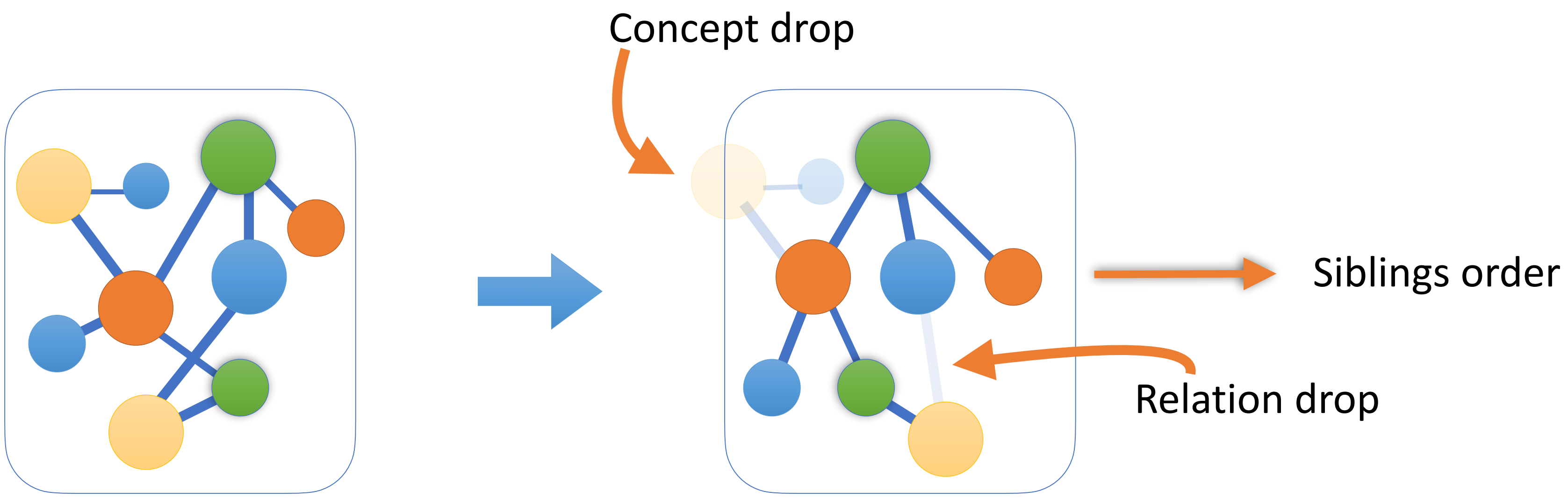
Idea: work with trees

- Concept converted to a tree (future work for graph ontologies —> trees)
  - Now done manually / assume a tree as input
  - Graph edges to be narrated, presentation order, synthesis



Idea: work with trees

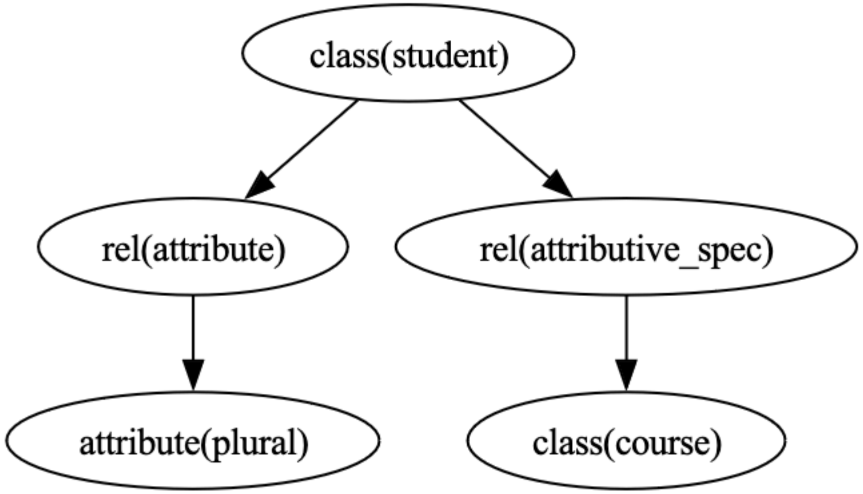
- Concept converted to a tree (future work for graph ontologies —> trees)
  - Now done manually / assume a tree as input
  - Graph edges to be narrated, presentation order, synthesis



Example: work with trees

- Concept trees as Prolog nested Lists
- [Root, Child_1, ..., Child_n]

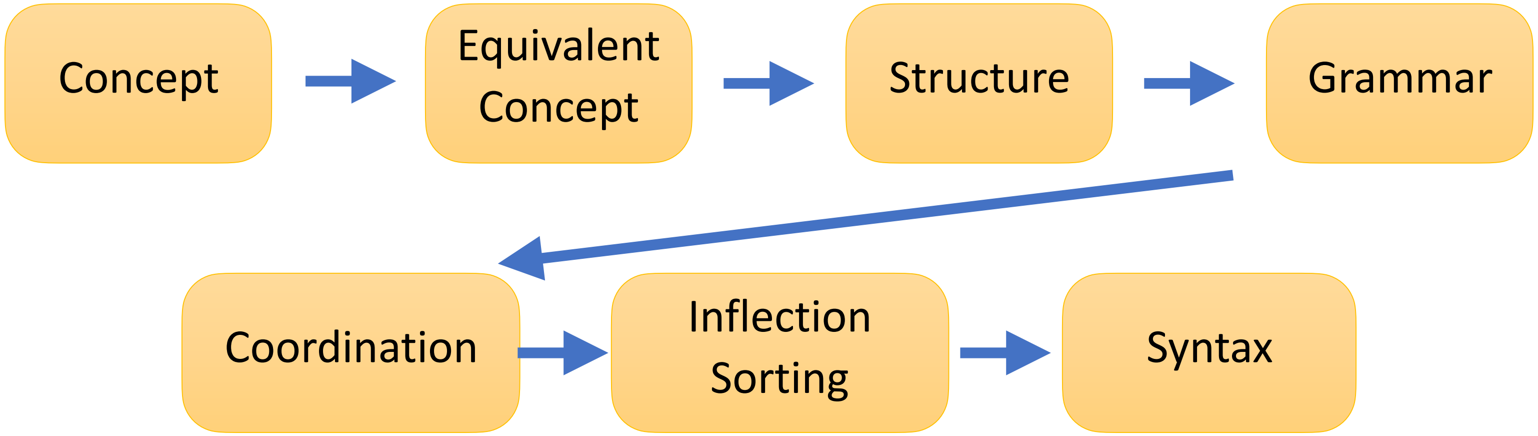
```
[class(student),
   [rel(attribute),attribute(plural)],
    [rel(attributive_spec),class(course)]
```



- Represented concept: Student + attr. plural + of a course

Idea: multi-step concept2text

- Intermediate tree **representations** from concept to text
- Rewrite a tree into another one, semantics preserving



Concept

Equivalent Concept

Structure

Grammar

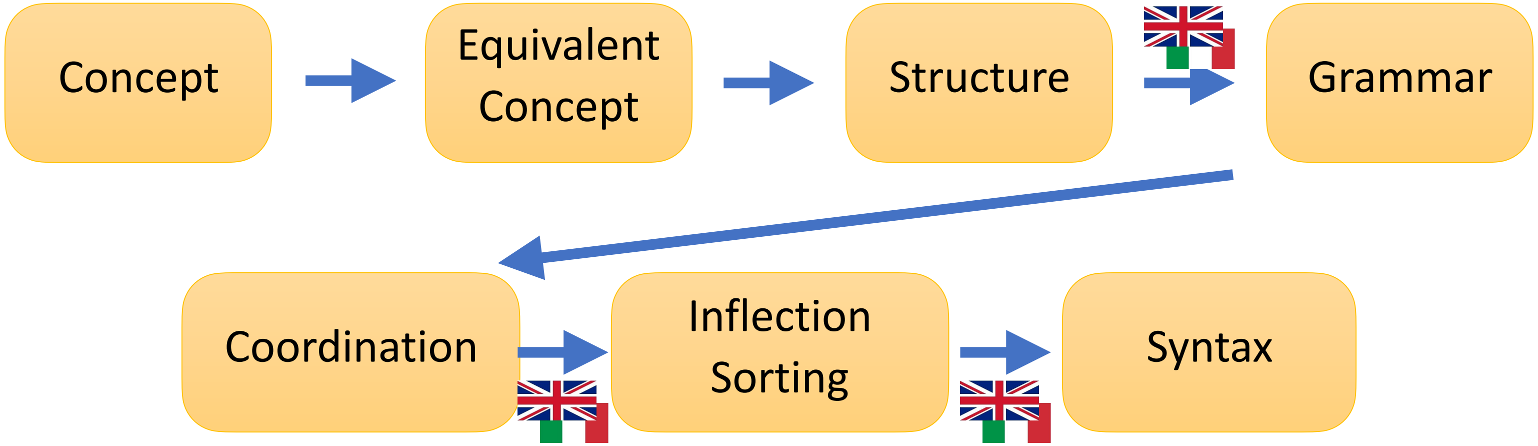
Coordination

Inflection Sorting

Syntax

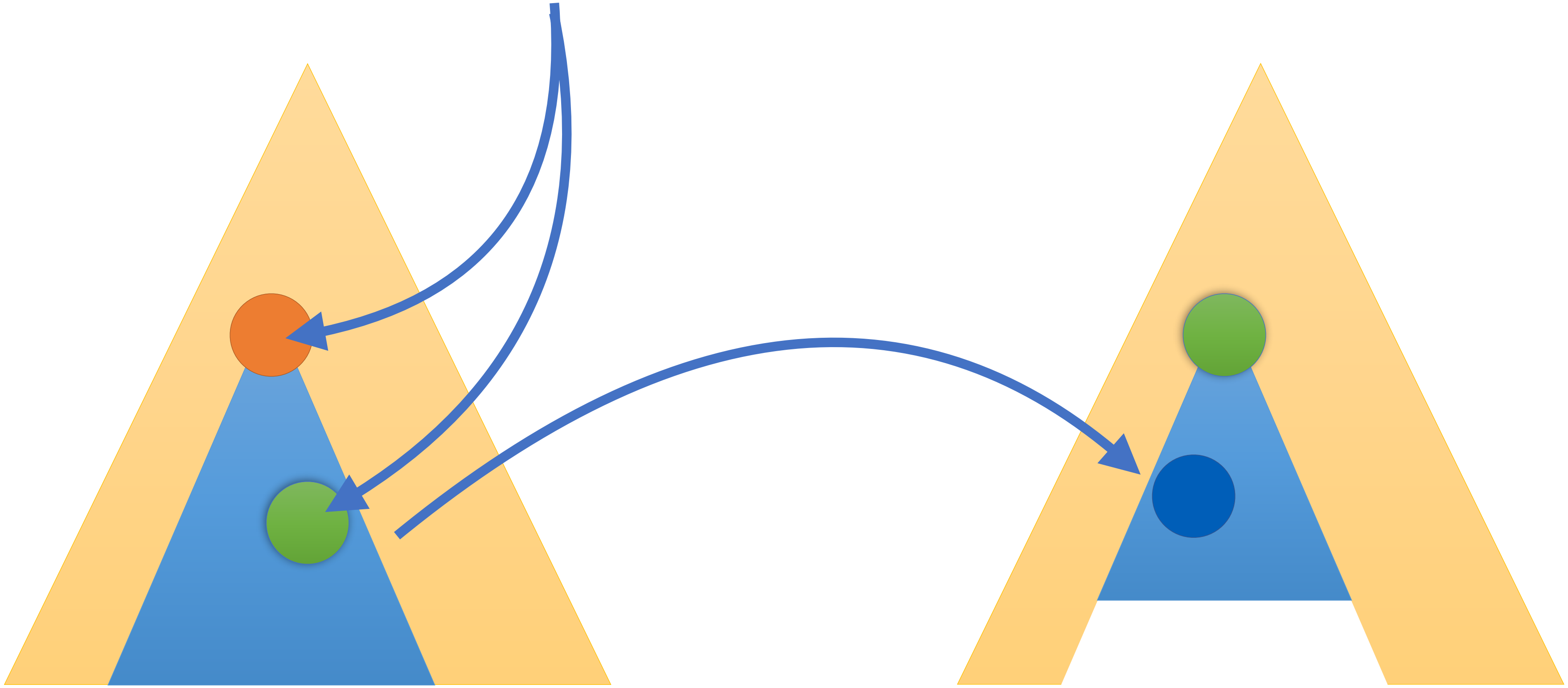
Idea: multi-step concept2text

- Some rewritings are language dependent
- Single program, modular language specialisation



Rules for Tree rewriting

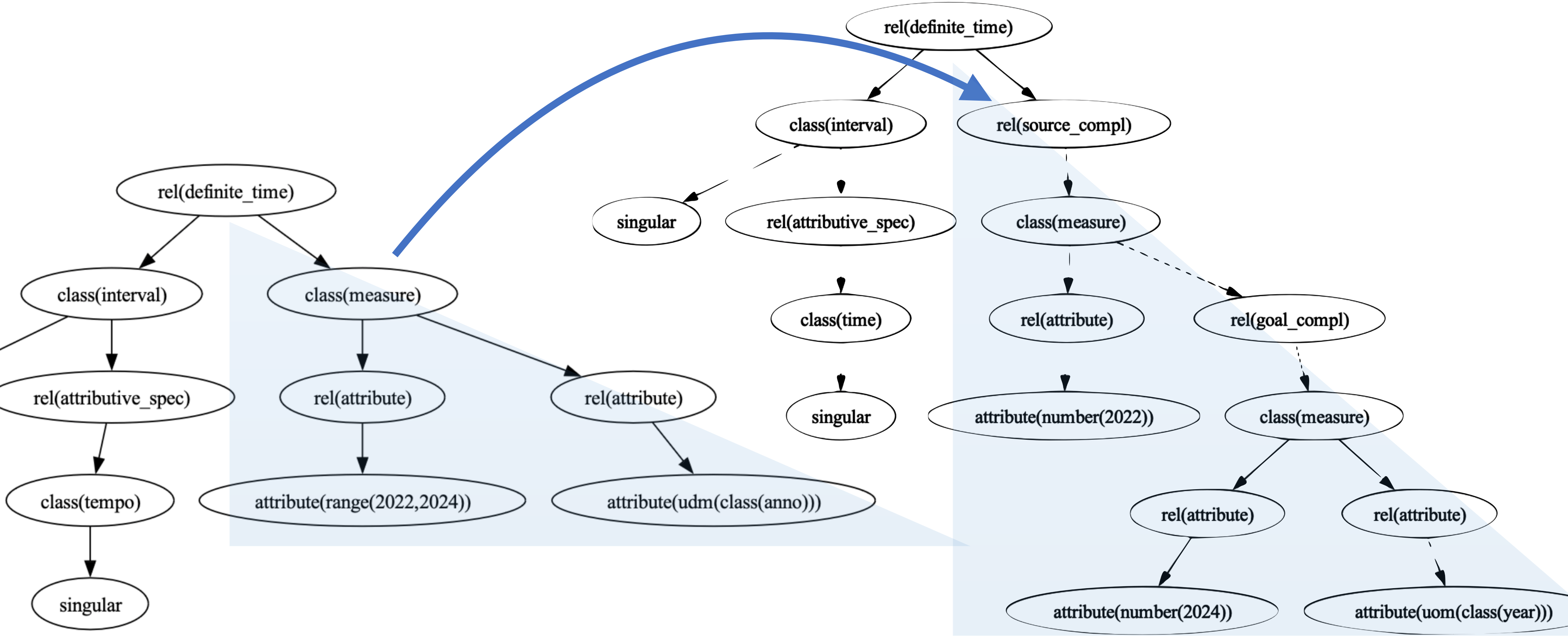
- From representation to the next one: set of rules applied to fix point
- Each rule defines a firing condition and a procedural tree substitution



Equivalent Concept

Representation

- Equivalent concept: depending on classes, ontology relations

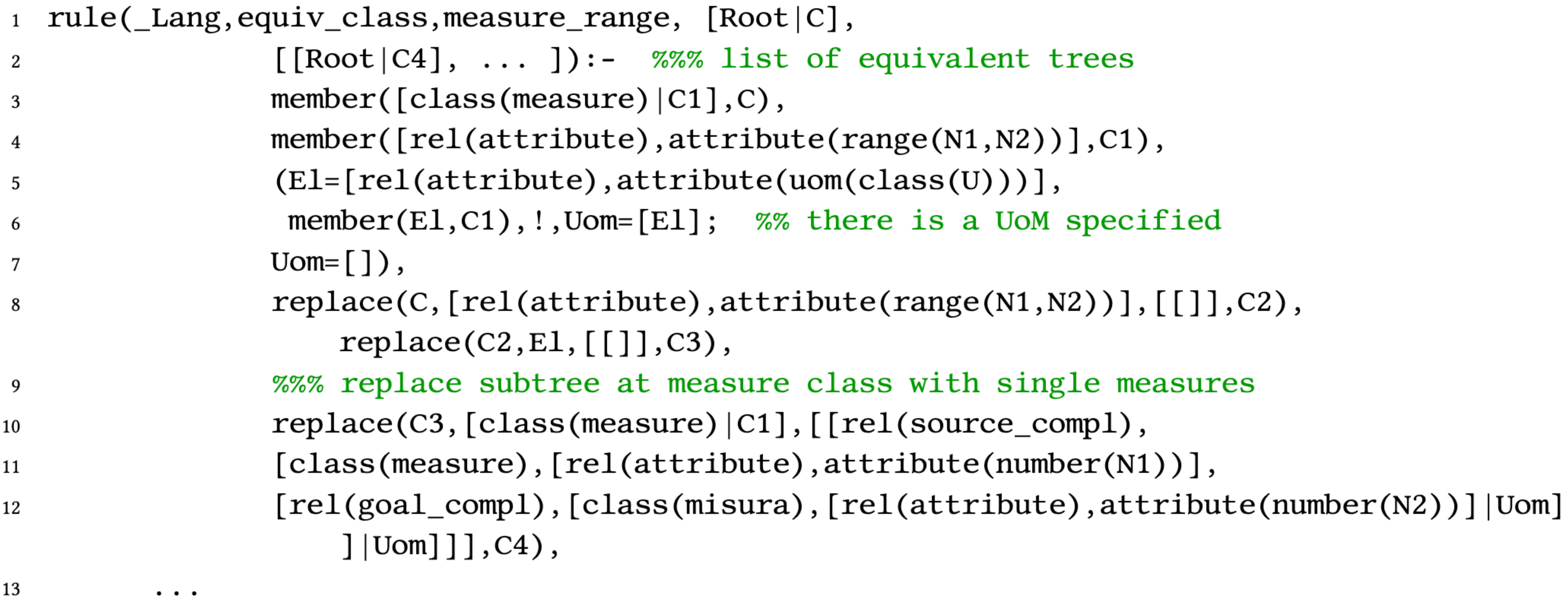


Concept2Text: an explainable multilingual rewriting of concepts into natural language

Equivalent Concept

Representation

- Equivalent concept: depending on classes, ontology relations

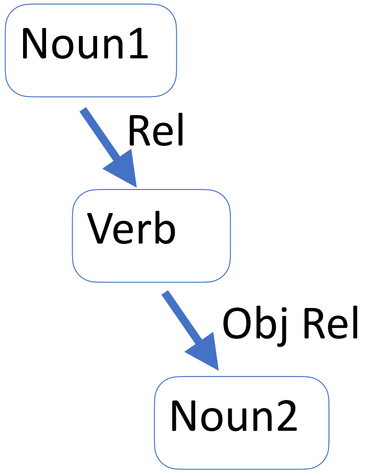


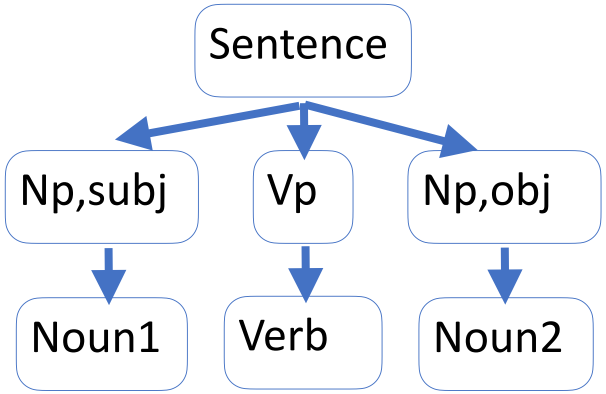
```
rule(_Lang,equiv_class,measure_range, [Root|C],
              [[Root|C4], ... ]):-  %%% list of equivalent trees
              member([class(measure)|C1],C),
              member([rel(attribute),attribute(range(N1,N2))],C1),
              (El=[rel(attribute),attribute(uom(class(U)))],
               member(El,C1),!,Uom=[El];  %% there is a UoM specified
              Uom=[]),
              replace(C,[rel(attribute),attribute(range(N1,N2))],[[]],C2),
                  replace(C2,El,[[]],C3),
              %%% replace subtree at measure class with single measures
              replace(C3,[class(measure)|C1],[[rel(source_compl),
              [class(measure),[rel(attribute),attribute(number(N1))],
              [rel(goal_compl),[class(misura),[rel(attribute),attribute(number(N2))]|Uom]
                  ]|Uom]]],C4),
      ...
```

Structure

Representation

- Internal node handling
- Hosting service information:
  - future type of phrase (noun, verbal, relative, etc)
  - Subtype (which complement etc)
  - Gender
  - Number
- Nested relations as siblings


Noun1
Rel
Verb
Obj Rel
Noun2


Sentence
Np,subj
Vp
Np,obj
Noun1
Verb
Noun2

Representation

Grammar

- From classes and relations to **lexemes**
- Non ground Gender and Number of nouns / verbs not yet assigned
- Choice of lexems can determine Gender

Coordination

- Unification of Gender and Number for related nodes in the tree
- Special requests for propagating information (e.g. relative clauses)

Inflection
Sorting

Representation

- Inflection = standard dictionary lookup
- Sorting -> compute the proper ordering of phrasal components
  - Noun phrase: article, pronoun, adjective, noun, absolute/ordinal number
  - Verbal phrase: auxiliary verb, verb, adverbs, tenses handling
- Define a CSP to compute the correct total order for phrases
  - Constraints specify local order between pairs of words type
  - Search for CSP solution

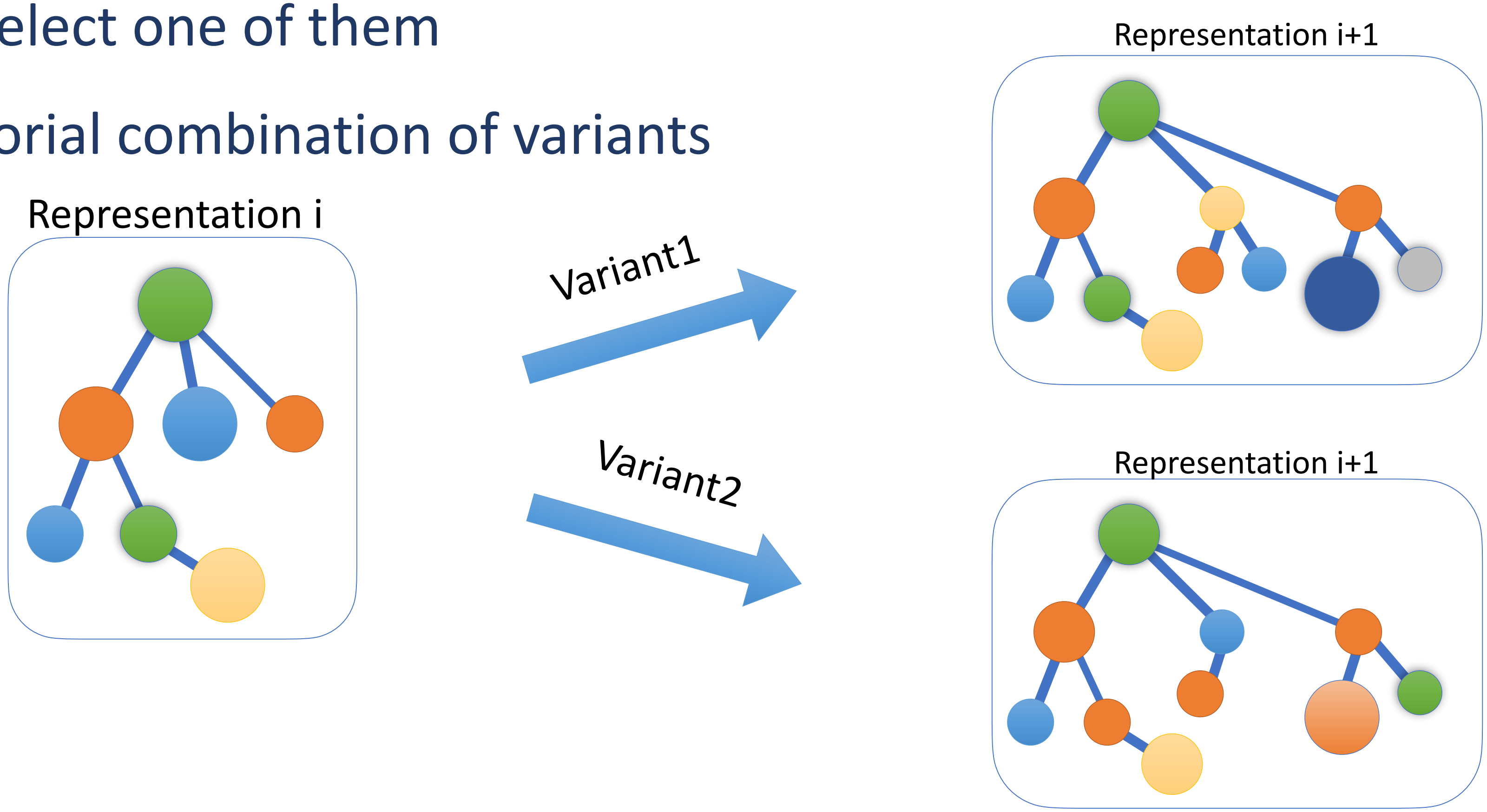
Syntax 🇬🇧🇮🇹

Representation

- Leaves of the the tree contain the words in correct order
- DFS visit collects the sentence
- Rules for correct character modifications, based on consecutive words
    - It + is -> It’s 🇬🇧
    - In + il -> Nel 🇮🇹

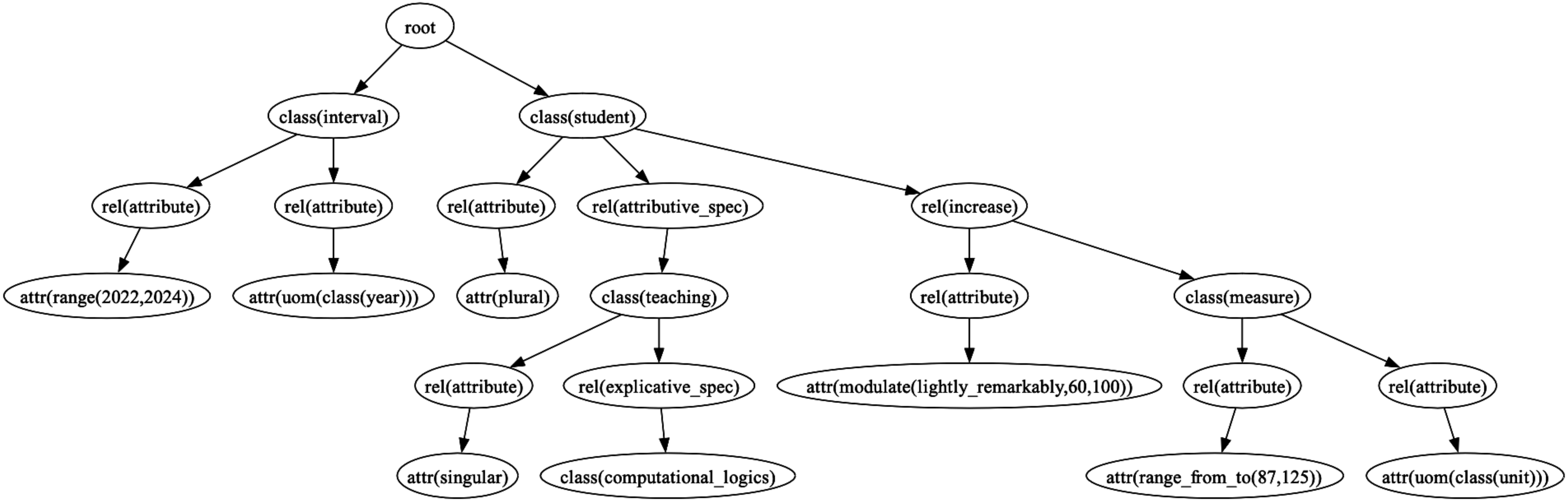
Idea: rewriting and variants

- Variants: each rule has multiple equivalent rewritings as output
- Random select one of them
- Combinatorial combination of variants



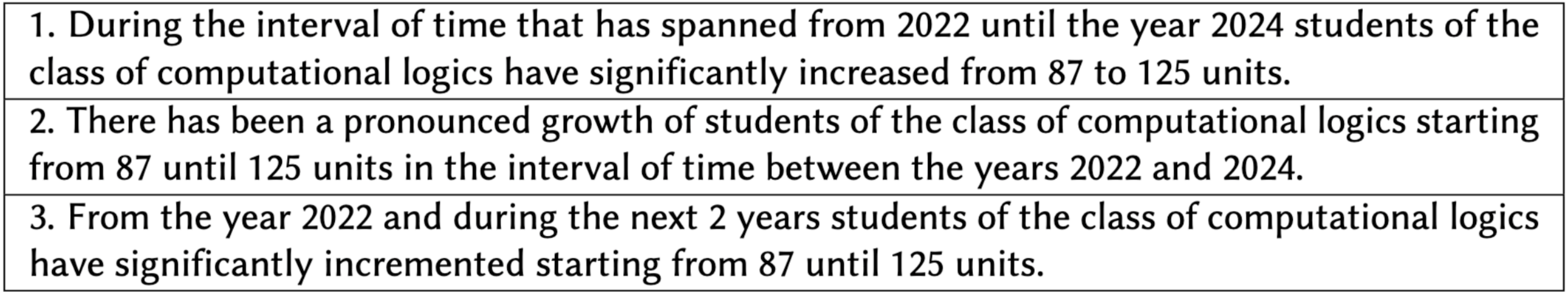
Example

- Input tree concept



Example

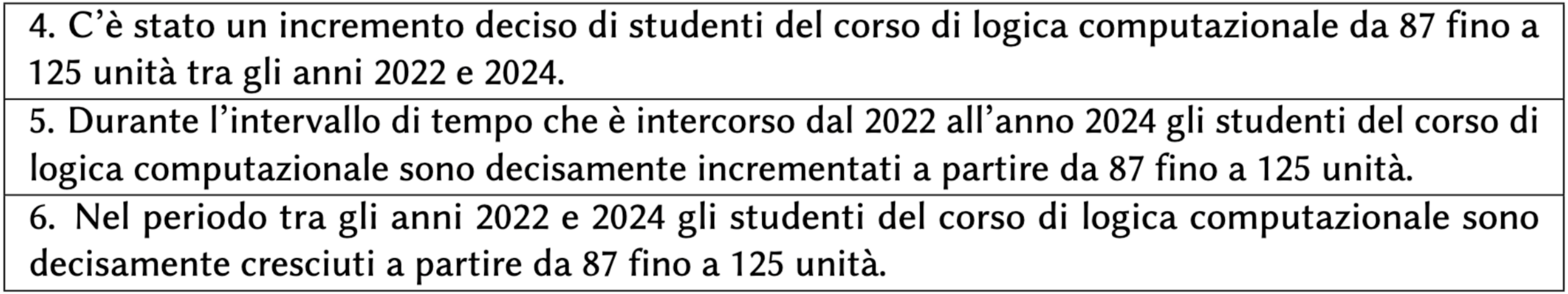
- 3 out of 500+ equivalent variants for both English and Italian



1. During the interval of time that has spanned from 2022 until the year 2024 students of the class of computational logics have significantly increased from 87 to 125 units.

2. There has been a pronounced growth of students of the class of computational logics starting from 87 until 125 units in the interval of time between the years 2022 and 2024.

3. From the year 2022 and during the next 2 years students of the class of computational logics have significantly incremented starting from 87 until 125 units.



4. C'è stato un incremento deciso di studenti del corso di logica computazionale da 87 fino a 125 unità tra gli anni 2022 e 2024.

5. Durante l'intervallo di tempo che è intercorso dal 2022 all'anno 2024 gli studenti del corso di logica computazionale sono decisamente incrementati a partire da 87 fino a 125 unità.

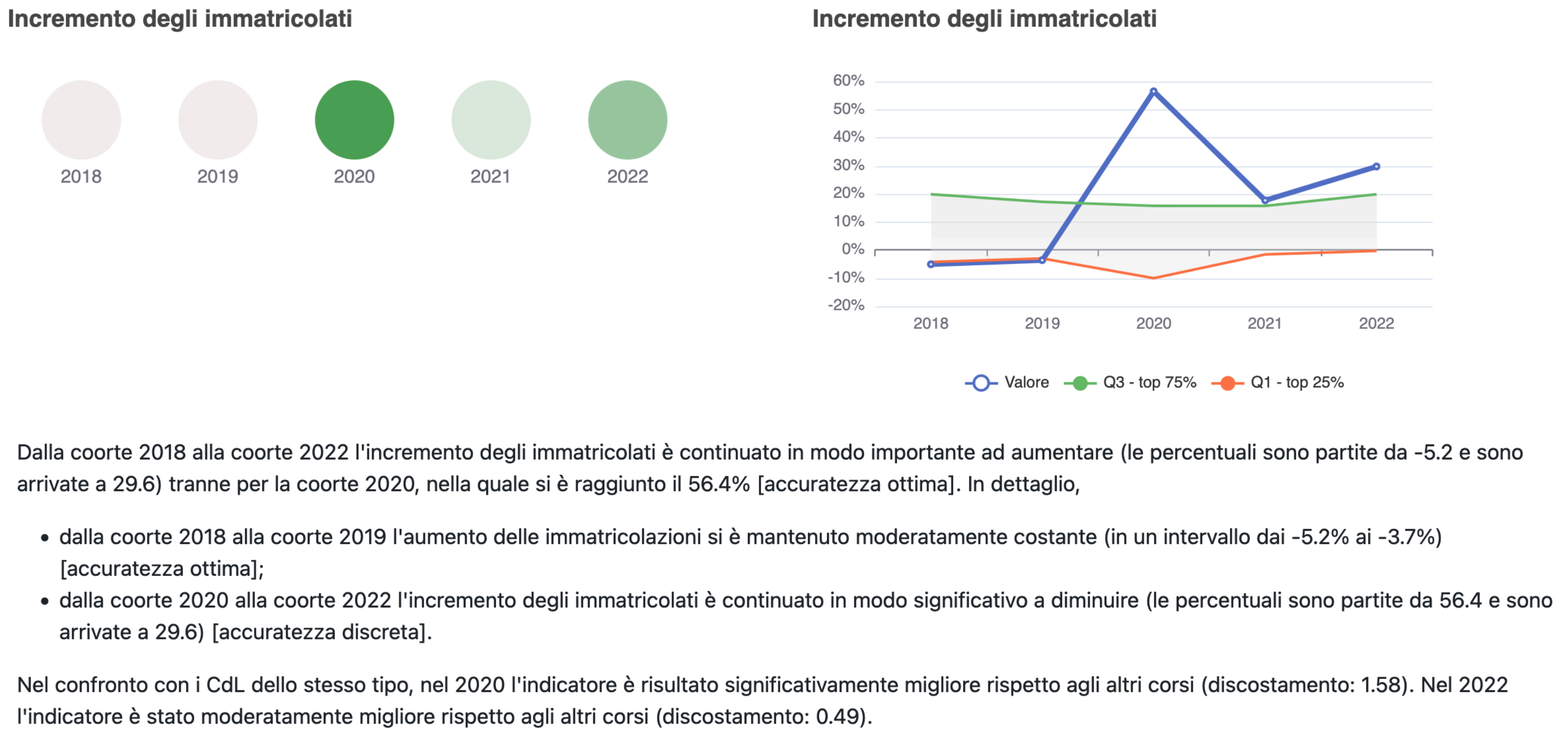
6. Nel periodo tra gli anni 2022 e 2024 gli studenti del corso di logica computazionale sono decisamente cresciuti a partire da 87 fino a 125 unità.

Conclusions

- The tree representation easily hosts typical concepts
- Language independent rules require initial work for building infrastructure
- Expressivity depends on variant rules -> this can be time consuming
- For domain specific applications —> by hand
- Rules are not encoded in online grammar resources!
    - Automated extraction of context dependent synonyms pairs of noun-verb
    - Bi-grams statistics
    - Use LLMs to synthetize rules?

Applications

- Currently: academic reports for students’ careers for Parma University Degrees
- Feedback on sentence accuracy and relevance



Dalla coorte 2018 alla coorte 2022 l'incremento degli immatricolati è continuato in modo importante ad aumentare (le percentuali sono partite da -5.2 e sono arrivate a 29.6) tranne per la coorte 2020, nella quale si è raggiunto il 56.4% [accuratezza ottima]. In dettaglio,

- dalla coorte 2018 alla coorte 2019 l'aumento delle immatricolazioni si è mantenuto moderatamente costante (in un intervallo dai -5.2% ai -3.7%) [accuratezza ottima];
- dalla coorte 2020 alla coorte 2022 l'incremento degli immatricolati è continuato in modo significativo a diminuire (le percentuali sono partite da 56.4 e sono arrivate a 29.6) [accuratezza discreta].

Nel confronto con i CdL dello stesso tipo, nel 2020 l'indicatore è risultato significativamente migliore rispetto agli altri corsi (discostamento: 1.58). Nel 2022 l'indicatore è stato moderatamente migliore rispetto agli altri corsi (discostamento: 0.49).

Applications

- The method can be safely deployed in high risk activities
  - Medical applications (automatic reports for ECG monitoring, interactive dialogue for rehab sessions)
  - Trustworthy Business Intelligence (explainable automated reporting)

Thank you for your attention