

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

“Easy peasy *term* squeezy”

A terminological perspective on text simplification

Federica Vezzani
federica.vezzani@unipd.it

Department of Linguistic and Literary Studies
Università degli Studi di Padova

SimpleText@CLEF-2023 - Automatic Simplification of Scientific Texts
September 18th 2023, Thessaloniki, Greece



“Nothing is more practical than a good theory”

-Vladimir Vapnik

Outline

- ▶ Background to Terminology
- ▶ On the Concept of Term *Difficulty*
- ▶ Towards Text Simplification
- ▶ Plain language and Controlled Natural Languages
- ▶ Terminology Management



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Federica Vezzani | SimpleText@CLEF-2023, September 18th 2023, Thessaloniki, Greece



Background to Terminology

Eugen Wüster (1898-1977)

- 1931 “Internationale Sprachnormung in der Technik, besonders in der Elektrotechnik”
- 1936 foundation of a technical committee dealing with terminological principles (ISA/TC37, later ISO/TC37)
- 1967/1968 elaboration of the systematic dictionary “The Machine Tool”

The Machine Tool An Interlingual Dictionary of Basic Concepts

comprising

An Alphabetical Dictionary and
A Classified Vocabulary
with Definitions and Illustrations

English-French Master Volume

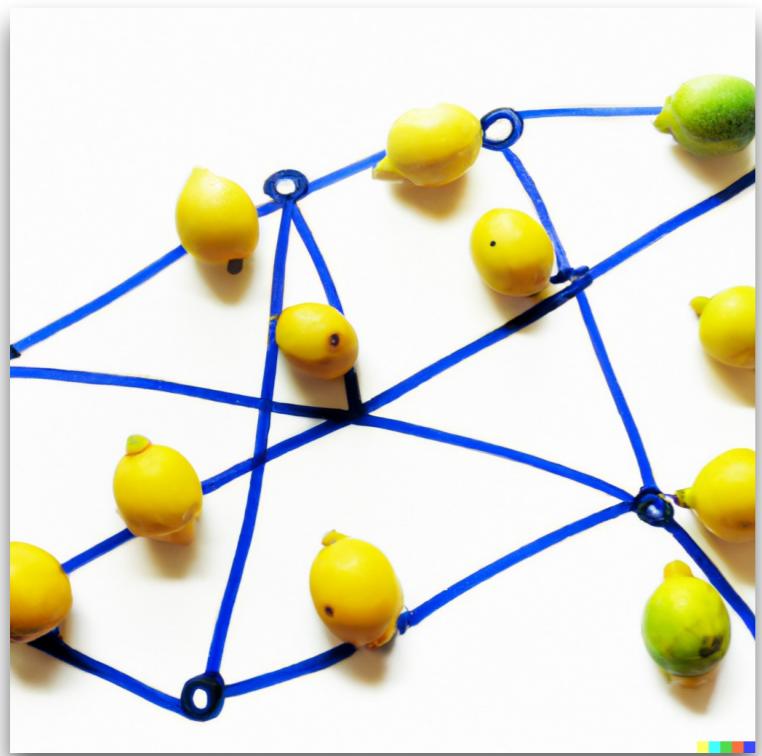
*Prepared under the auspices of
The United Nations
Economic Commission for Europe
and under the direction of
Eugen Wüster*



TECHNICAL PRESS
LONDON

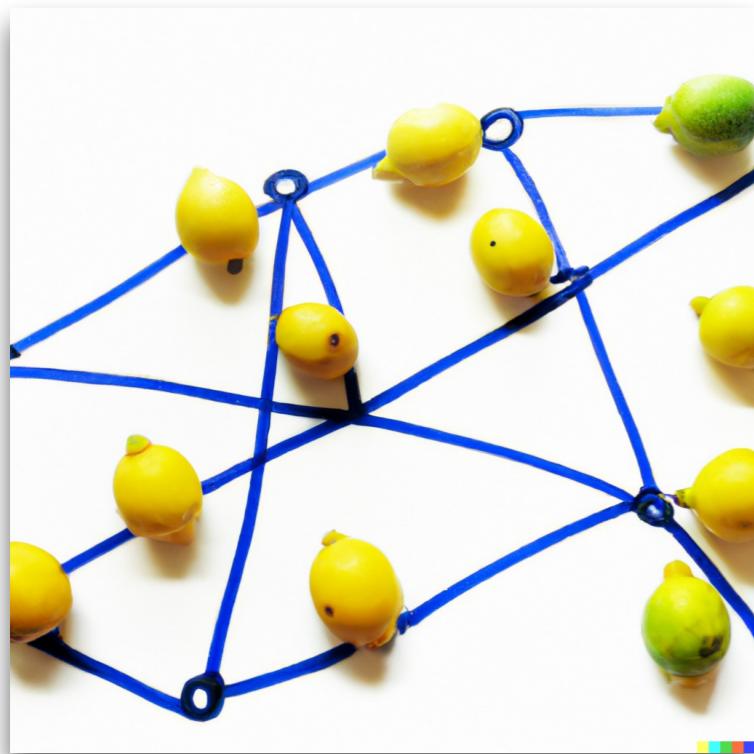
The Role of Terminology

The Role of Terminology



Knowledge organisation

The Role of Terminology



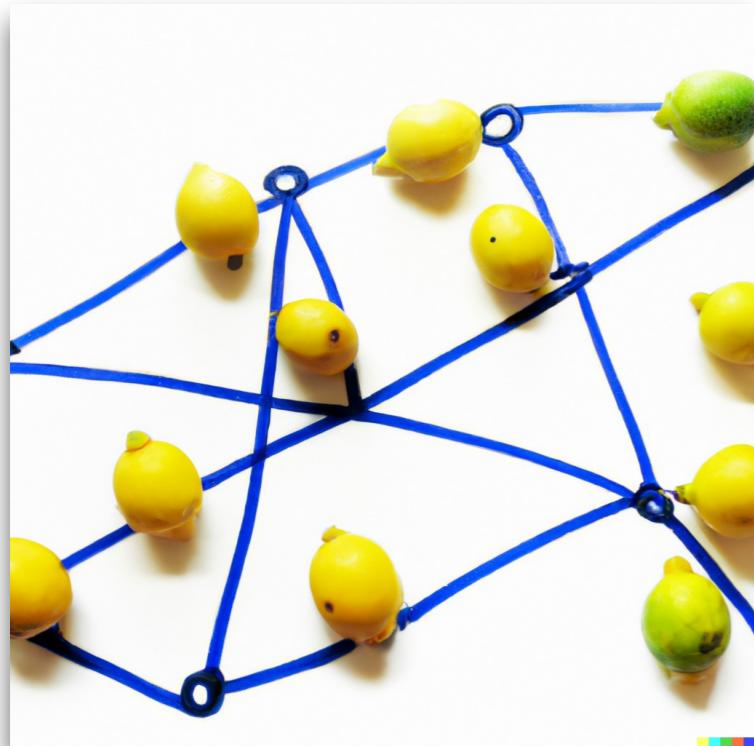
Knowledge organisation



Knowledge representation



The Role of Terminology



Knowledge organisation



Knowledge representation



Knowledge sharing

The Role of Terminology



Some Basic Concepts

Double Dimension of Terminology

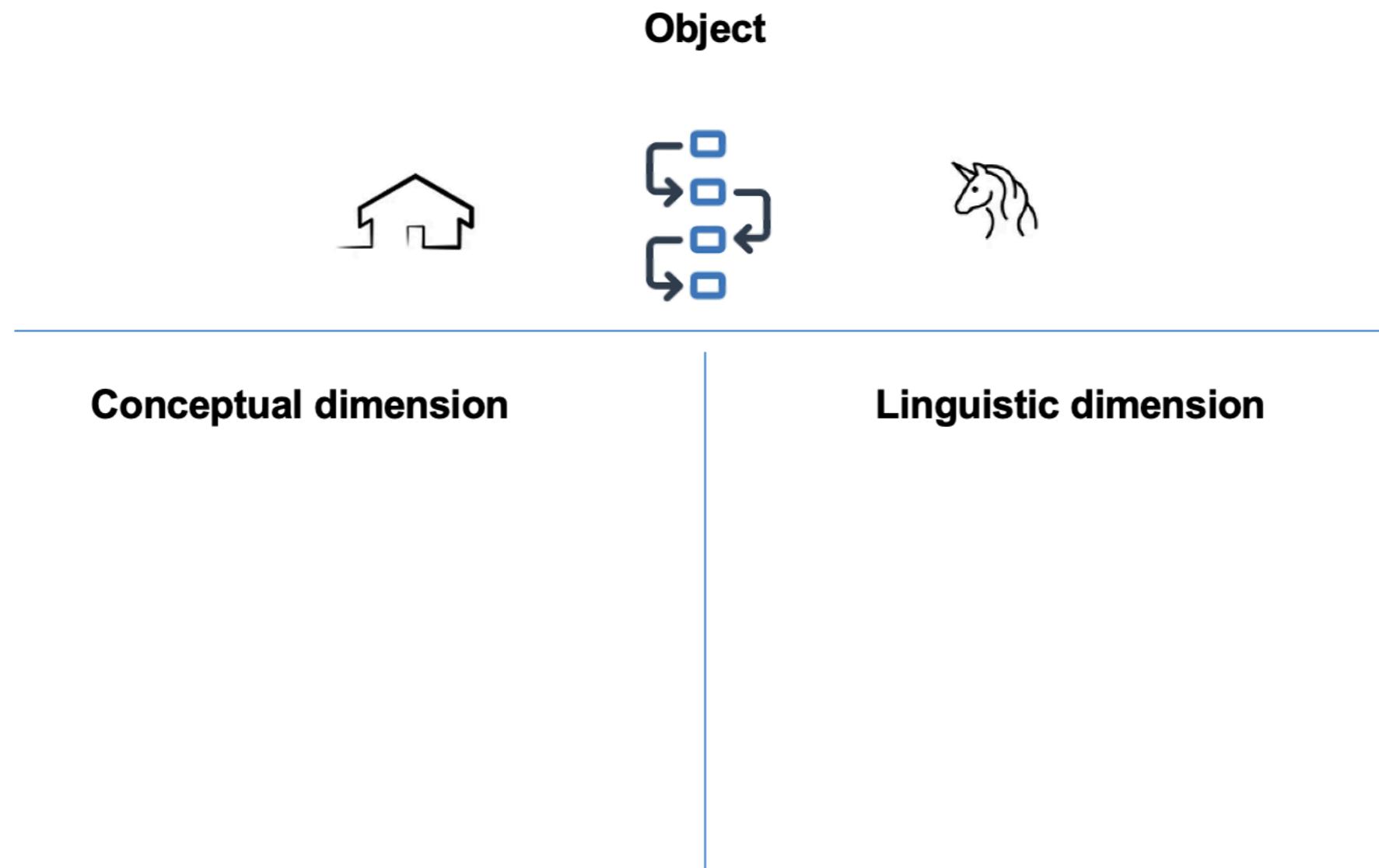
Conceptual dimension

Linguistic dimension

Costa, R. (2013). Terminology and Specialised Lexicography: two complementary domains. *Lexicographica*, (29), 29-42.

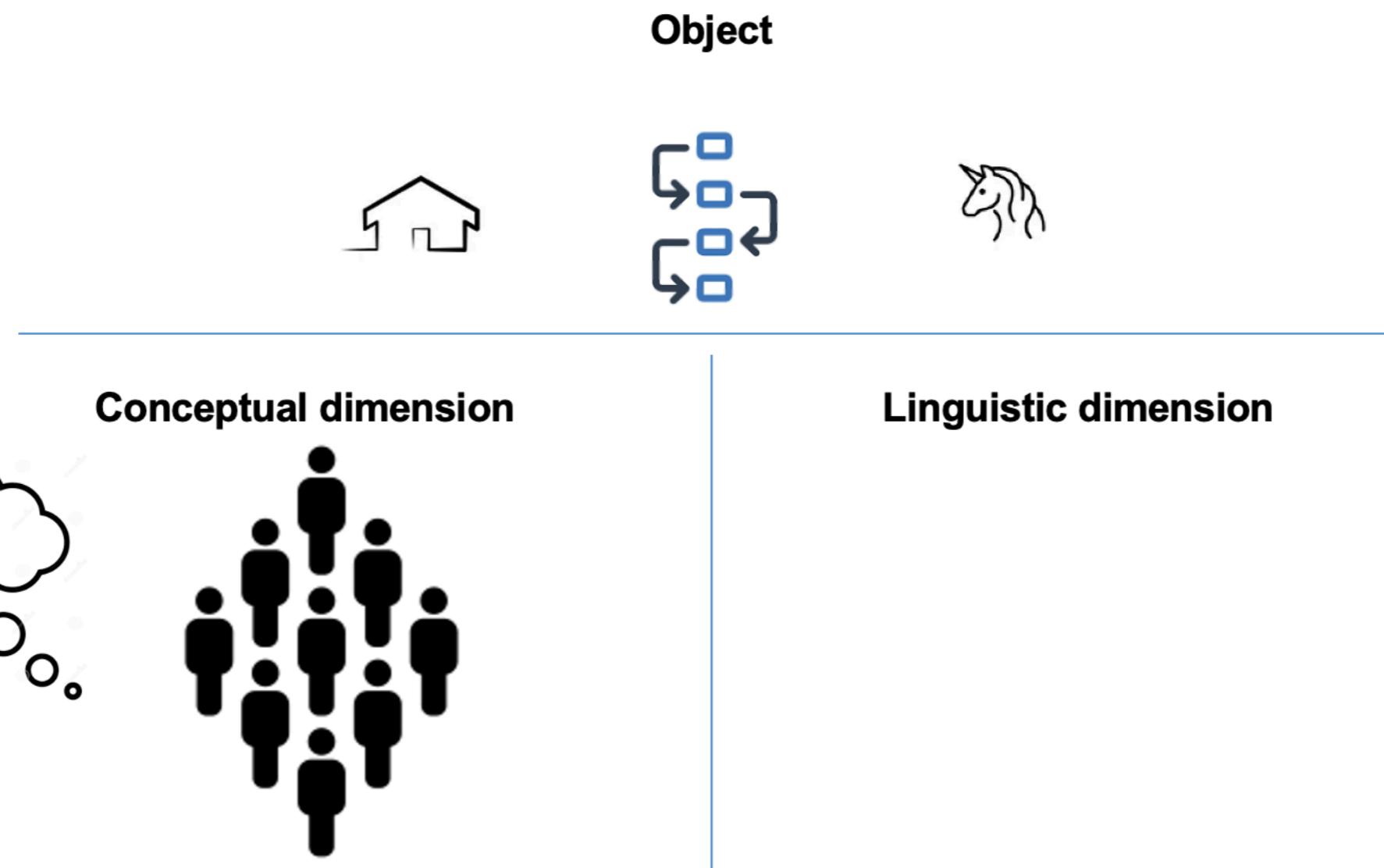
ISO 1087: 2019 Terminology Work and Terminology Science - Vocabulary

Double Dimension of Terminology



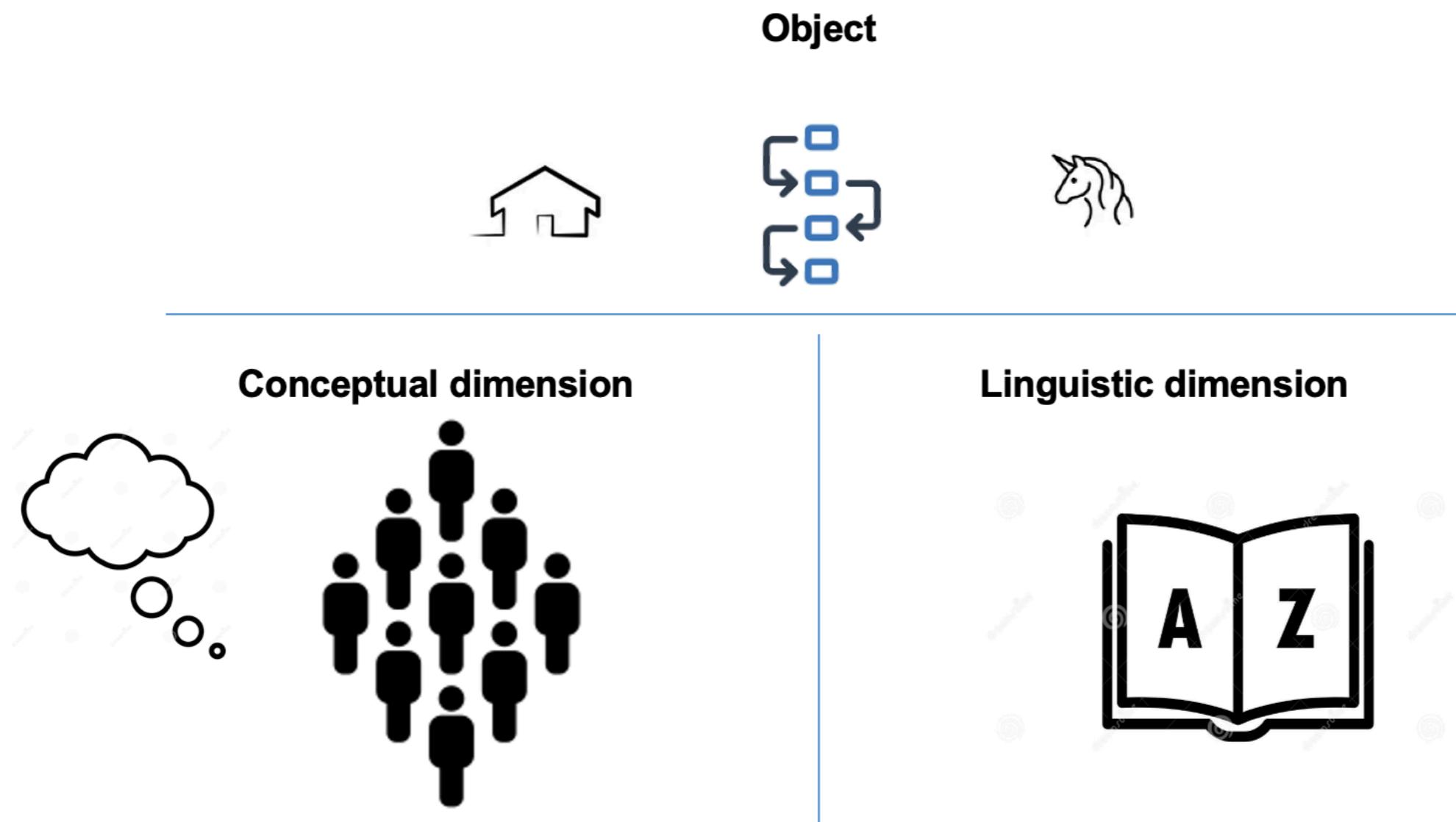
Costa, R. (2013). Terminology and Specialised Lexicography: two complementary domains. *Lexicographica*, (29), 29-42.

Double Dimension of Terminology



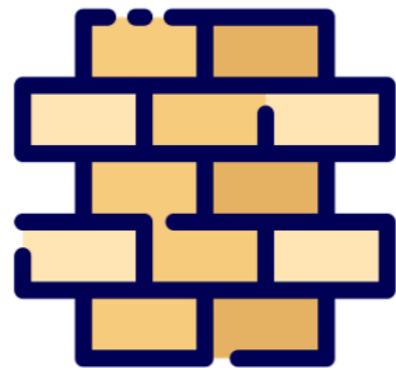
Costa, R. (2013). Terminology and Specialised Lexicography: two complementary domains. *Lexicographica*, (29), 29-42.

Double Dimension of Terminology



Costa, R. (2013). Terminology and Specialised Lexicography: two complementary domains. *Lexicographica*, (29), 29-42.

Double Dimension of Terminology



<Wall>

“wall”

“muro”

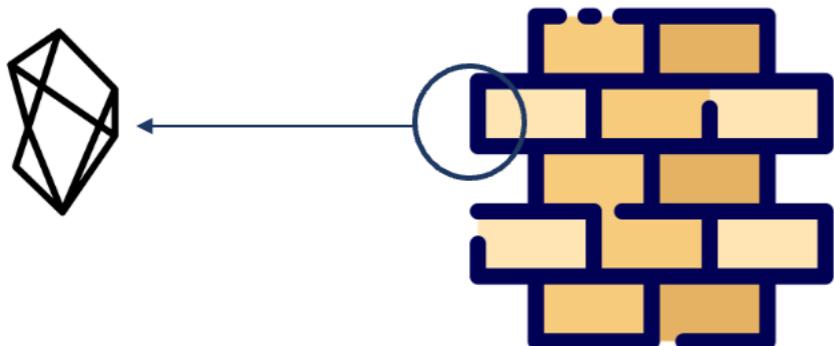
“mur”

“τείχος”

Double Dimension of Terminology



Object has *properties*



Concept has *characteristics*

<Wall>

“wall”

“muro”

“mur”

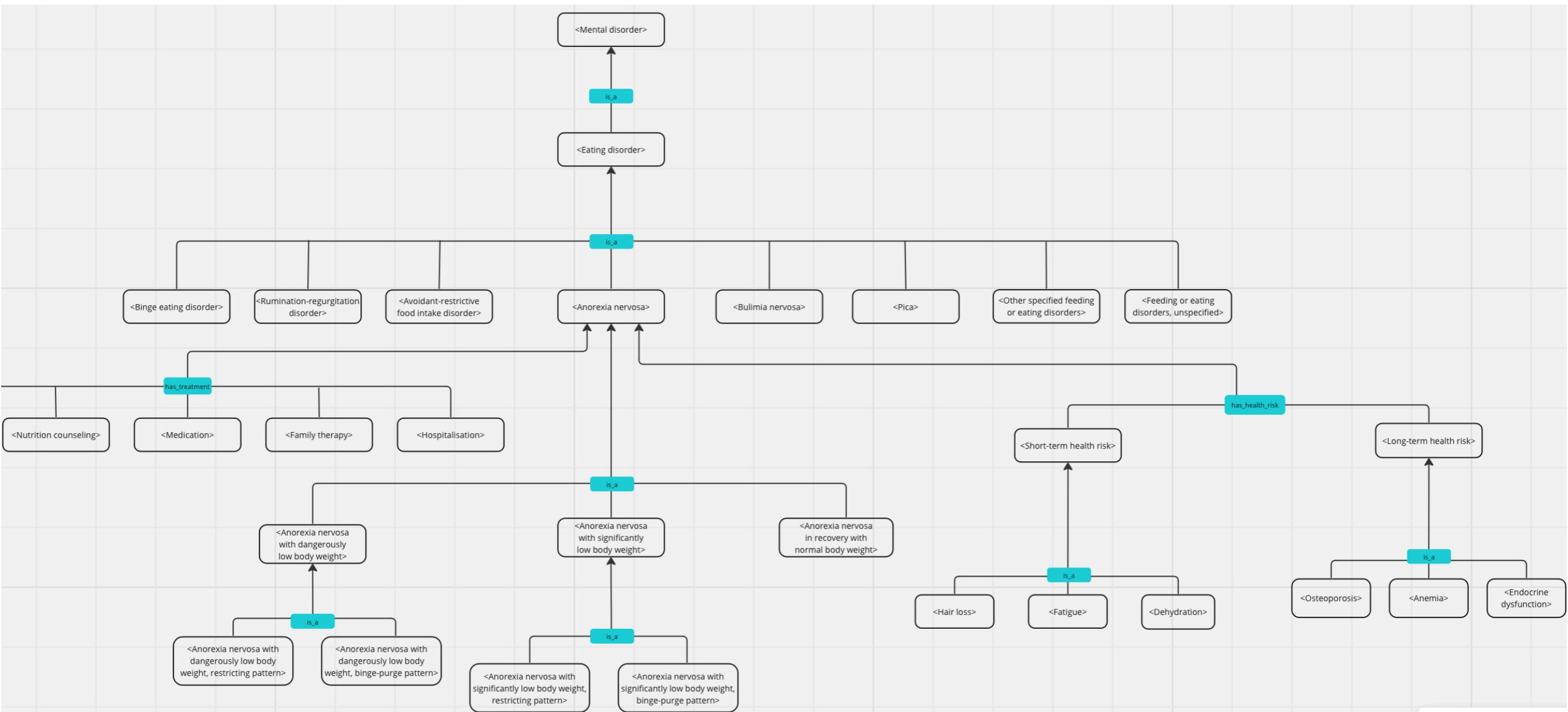
“τείχος”

Term has *semes*

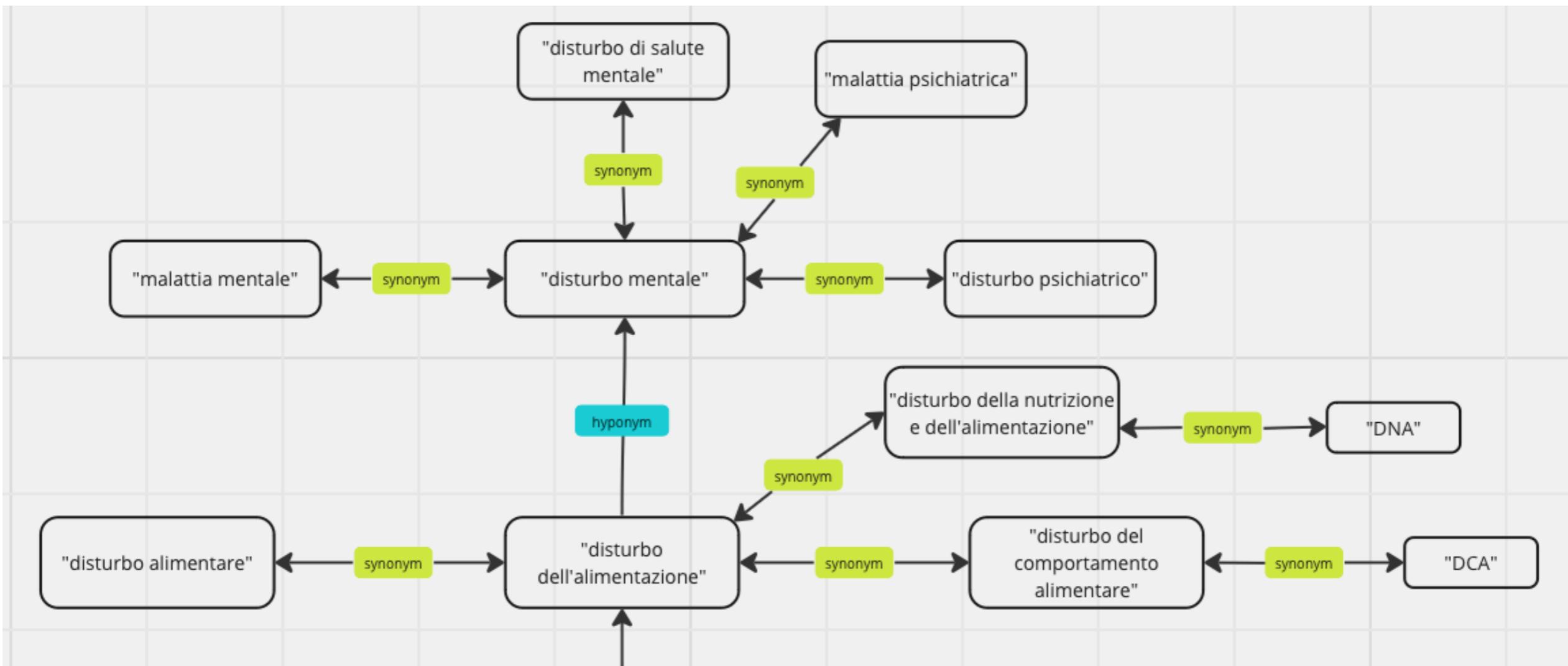
Conceptual System



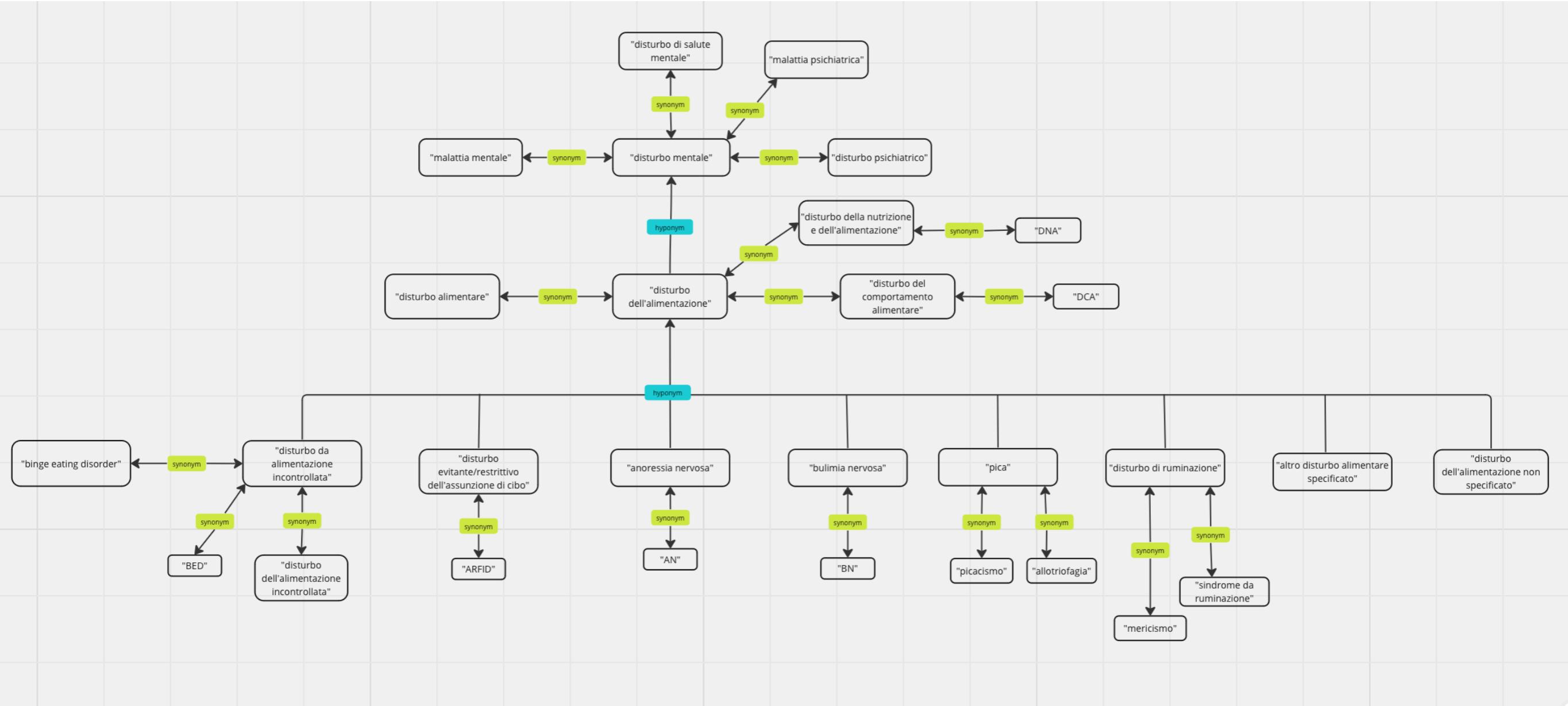
Conceptual System



(Multilingual) Lexical Networks



(Multilingual) Lexical Networks

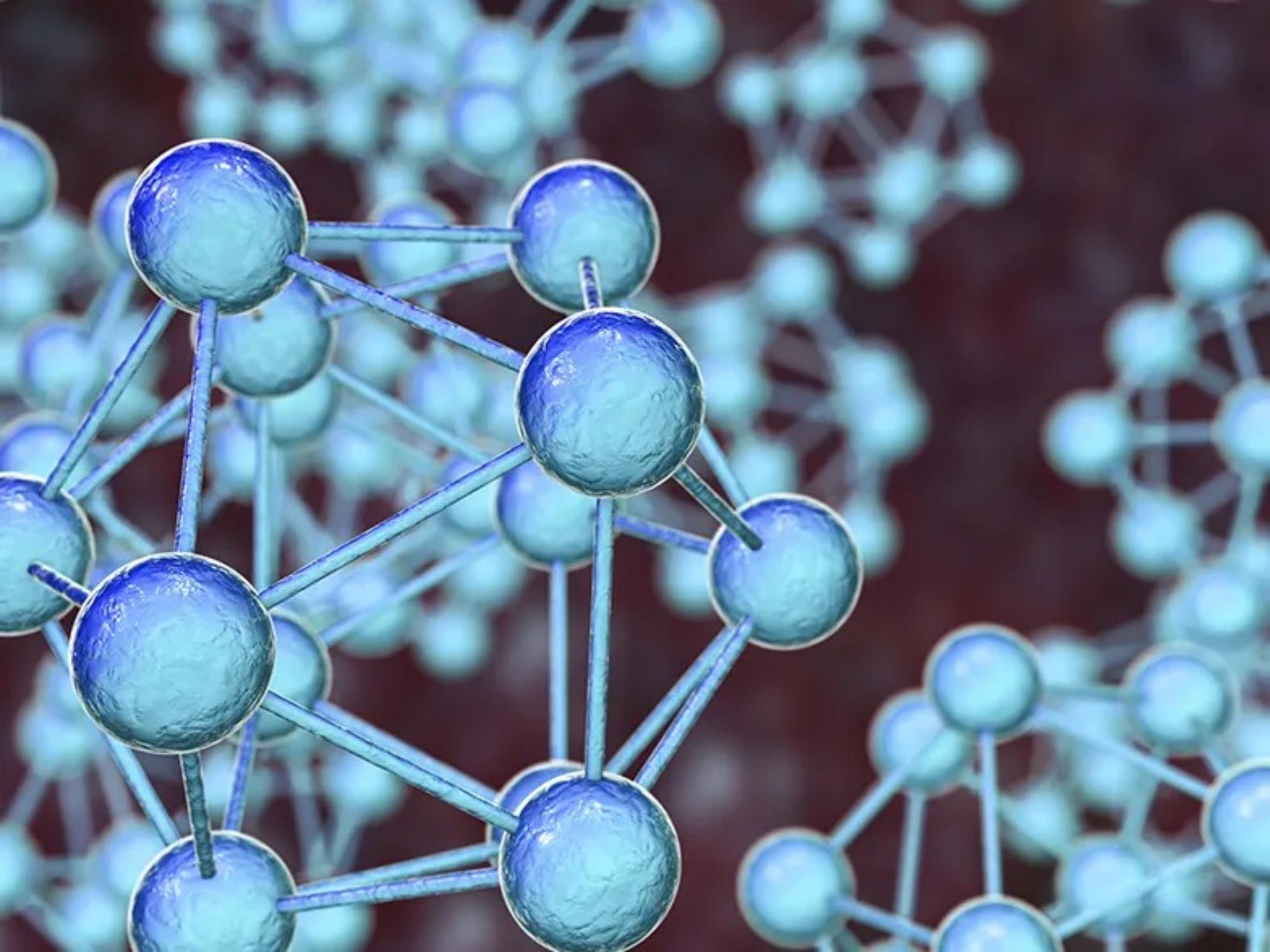


On the Concept of Term *Difficulty*

Complexity vs Difficulty

- <Complexity>: intrinsic property of an object and is given by its internal structure. It's not dependent on an individual's skills or abilities.
- <Difficulty>: level of challenge or effort (skills, knowledge, experience, and cognitive abilities) required by an individual to deal with that complexity.

Pelánek, R., Effenberger, T., & Čechák, J. (2022). Complexity and difficulty of items in learning systems. International Journal of Artificial Intelligence in Education, 32(1), 196-232.



On the Concept of *Difficulty*

- ▶ In the texts we find terms (not concepts)
 - ▶ What are the criteria for defining term difficulty?
 - ▶ What characteristics can I quantify in a term and qualify as difficult?
 - ▶ How can we measure term difficulty?



Morphology



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Federica Vezzani | SimpleText@CLEF-2023, September 18th 2023, Thessaloniki, Greece



Morphology

- ▶ Morphologically difficult terms



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Federica Vezzani | SimpleText@CLEF-2023, September 18th 2023, Thessaloniki, Greece



Morphology

- ▶ Morphologically difficult terms
 - ▶ high number of morphemes:



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Federica Vezzani | SimpleText@CLEF-2023, September 18th 2023, Thessaloniki, Greece



Morphology

- ▶ Morphologically difficult terms
 - ▶ high number of morphemes:
 - ▶ “nephrotoxicity”: nephro- (kidney) + toxic- (poison) + -ity (condition)



Morphology

- ▶ Morphologically difficult terms
 - ▶ high number of morphemes:
 - ▶ “nephrotoxicity”: nephro- (kidney) + toxic- (poison) + -ity (condition)
 - ▶ “cardiomyopathy”: cardio- (heart) + myo- (muscle) + -pathy (disease)



Morphology

- ▶ Morphologically difficult terms
 - ▶ high number of morphemes:
 - ▶ “nephrotoxicity”: nephro- (kidney) + toxic- (poison) + -ity (condition)
 - ▶ “cardiomyopathy”: cardio- (heart) + myo- (muscle) + -pathy (disease)
 - ▶ Difficulty calculated on the basis of readability indices



Morphology

- ▶ Morphologically difficult terms
 - ▶ high number of morphemes:
 - ▶ “nephrotoxicity”: nephro- (kidney) + toxic- (poison) + -ity (condition)
 - ▶ “cardiomyopathy”: cardio- (heart) + myo- (muscle) + -pathy (disease)
- ▶ Difficulty calculated on the basis of readability indices
 - ▶ Depends on each language



Semantics



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Federica Vezzani | SimpleText@CLEF-2023, September 18th 2023, Thessaloniki, Greece



Semantics

- ▶ Sense/concept disambiguation



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Federica Vezzani | SimpleText@CLEF-2023, September 18th 2023, Thessaloniki, Greece



Semantics

- ▶ Sense/concept disambiguation
 - ▶ “The patient's *sensitivity* to light was a concern for the doctor”



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Federica Vezzani | SimpleText@CLEF-2023, September 18th 2023, Thessaloniki, Greece



Semantics

- ▶ Sense/concept disambiguation
 - ▶ “The patient's *sensitivity* to light was a concern for the doctor”
 - ▶ Allergic Reaction



Semantics

- ▶ Sense/concept disambiguation
 - ▶ “The patient's *sensitivity* to light was a concern for the doctor”
 - ▶ Allergic Reaction
 - ▶ Photophobia



Semantics

- ▶ Sense/concept disambiguation
 - ▶ “The patient's *sensitivity* to light was a concern for the doctor”
 - ▶ Allergic Reaction
 - ▶ Photophobia
 - ▶ Individual knowledge/educational level



Semantics

- ▶ Sense/concept disambiguation
 - ▶ “The patient's *sensitivity* to light was a concern for the doctor”
 - ▶ Allergic Reaction
 - ▶ Photophobia
 - ▶ Individual knowledge/educational level
 - ▶ Vocabolario di base



Towards Text Simplification

Strategies

- Terminological Variation
- Terminological Definitions
- Definition vs Explanation



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Federica Vezzani | SimpleText@CLEF-2023, September 18th 2023, Thessaloniki, Greece



Terminological Variation

Terminological Variation

<Concept A>

“ganser syndrome”

“nonsense syndrome”

“prison psychosis”



Terminological Variation

<Concept A>

“hysteria”



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Federica Vezzani | SimpleText@CLEF-2023, September 18th 2023, Thessaloniki, Greece



Terminological Variation

<Concept A>

<Concept B>

“hysteria”



Terminological Variation

<Concept A>

“dysmorphophobia”



Terminological Variation

<Concept A>

<Concept B>

“dysmorphophobia”



Terminological Variation

<Concept A>

<Concept B>

“dysmorphophobia”

“body dysmorphic disorder”



Dimensions of Variation

- ▶ Diatopic variation
- ▶ Diachronic variation
- ▶ Diamesic variation
- ▶ Diaphasic variation
- ▶ Diastratic variation



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Federica Vezzani | SimpleText@CLEF-2023, September 18th 2023, Thessaloniki, Greece



Terminological Variation (diastratic)

- ▶ Replace scientific terms with popular terms
 - ▶ “hypercholesterolemia” - “high cholesterol”
 - ▶ “myocardial infarction” - “heart attack”
 - ▶ “hypertension” - “high blood pressure”



Terminological Definitions

Types of Definitions

- ▶ Lexicographic definition: “a metalinguistic discourse locked in the structures of a language and a culture” (Alain Rey, 1982)
- ▶ Terminological definition: “representation of a concept by an expression that describes it and differentiates it from related concepts” (ISO 1087: 2019)



Types of Definitions (ISO 1087: 2019)

3.3.2

intensional definition

*definition (3.3.1) that conveys the *intension* (3.2.6) of a *concept* (3.2.7) by stating the immediate *generic concept* (3.2.19) and the *delimiting characteristic(s)* (3.2.5)*

EXAMPLE 1 optical mouse: computer mouse in which movements are detected by light sensors.

EXAMPLE 2 mechanical mouse: computer mouse in which movements are detected by rollers and a ball.

Note 1 to entry: Intensional definitions are preferable to other types of *definitions* (3.3.1) because they clearly reveal the *characteristics* (3.2.1) of a *concept* (3.2.7) within a *concept system* (3.2.28): they should be used whenever possible.

3.3.3

extensional definition

*definition (3.3.1) that enumerates all the *subordinate concepts* (3.2.16) of a *superordinate concept* (3.2.15) under one *criterion of subdivision* (3.2.17)*



Natural Language Definition vs Explanation

- ▶ Definition: to provide a clear representation of the concept and set boundaries between related concepts.
- ▶ Explanation: to make a concept more understandable by providing context, background information, details, or reasons.



Natural Language Definition vs Explanation



Certainly! Let's take the concept of "gravity" as an example and provide both a definition and an explanation for it:



Definition of Gravity:

- Gravity is the force of attraction that exists between all objects with mass or energy. It is responsible for the phenomenon of objects falling toward the Earth and is described by the law of universal gravitation formulated by Sir Isaac Newton.

Explanation of Gravity:

- Gravity is a fundamental force of nature that acts between all objects that have mass. This force is what makes objects fall when dropped and keeps planets in orbit around the sun. It is described by the law of universal gravitation, which states that every object with mass attracts every other object with mass. The strength of this attraction depends on the mass of the objects and the distance between them. For example, the reason we stay on the Earth's surface is because of the gravitational pull between us and the Earth. Without gravity, we would float away into space. Gravity is a fundamental concept in physics and plays a crucial role in understanding the motion of celestial bodies and everyday phenomena like dropping a ball or jumping.

Regenerate



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

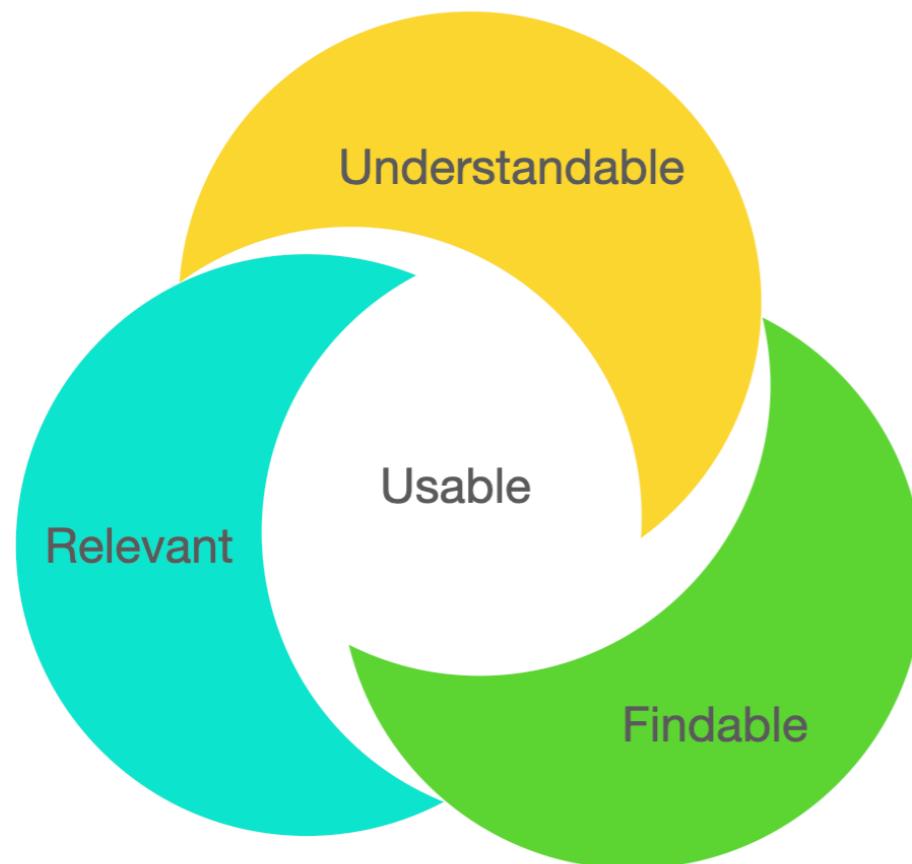
Federica Vezzani | SimpleText@CLEF-2023, September 18th 2023, Thessaloniki, Greece



Plain Language and CNLs

ISO TC/37

Language and Terminology



INTERNATIONAL
STANDARD

ISO
24495-1

First edition
2023-06

Plain language — Part 1: Governing principles and guidelines

*Langage clair et simple —
Partie 1: Principes directeurs et lignes directrices*

Overview of the Principles and Guidelines

Understandable:
Readers can easily understand what they find

Choose familiar words

Write clear and concise sentences

Write clear and concise paragraphs

Consider including images and multimedia

Project a respectful tone

Ensure that the document is cohesive

Controlled Natural Languages (CNLs)

- ▶ A language is called a controlled natural language if and only if it has all of the following four properties:
 - ▶ It is based on exactly one natural language
 - ▶ It is more restrictive concerning lexicon, syntax, and/or semantics
 - ▶ It preserves most of the natural properties of its base language
 - ▶ It is explicitly and consciously defined

Kuhn, Tobias. "A survey and classification of controlled natural languages." *Computational linguistics* 40, no. 1 (2014): 121-170.



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Federica Vezzani | SimpleText@CLEF-2023, September 18th 2023, Thessaloniki, Greece



ASD Simplified Technical English (ASD-STE)

- ▶ Controlled Natural Language for the Aerospace Industry (1986)
- ▶ Today, the language is maintained by the Simplified Technical English Maintenance Group (STEMG)
- ▶ ASD-STE is based on English with restrictions expressed in about 60 general rules + a dictionary

<https://www.asd-ste100.org>



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Federica Vezzani | SimpleText@CLEF-2023, September 18th 2023, Thessaloniki, Greece



ASD Simplified Technical English (ASD-STE)



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Federica Vezzani | SimpleText@CLEF-2023, September 18th 2023, Thessaloniki, Greece



ASD Simplified Technical English (ASD-STE)

- ▶ These rules restrict the language on



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Federica Vezzani | SimpleText@CLEF-2023, September 18th 2023, Thessaloniki, Greece



ASD Simplified Technical English (ASD-STE)

- ▶ These rules restrict the language on
 - ▶ the lexical level (rule 1.2 “use approved words from the dictionary only as the part of speech given”)



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Federica Vezzani | SimpleText@CLEF-2023, September 18th 2023, Thessaloniki, Greece



ASD Simplified Technical English (ASD-STE)

- ▶ These rules restrict the language on
 - ▶ the lexical level (rule 1.2 “use approved words from the dictionary only as the part of speech given”)
 - ▶ “test” noun not as verb



ASD Simplified Technical English (ASD-STE)

- ▶ These rules restrict the language on
 - ▶ the lexical level (rule 1.2 “use approved words from the dictionary only as the part of speech given”)
 - ▶ “test” noun not as verb
 - ▶ the syntactic level (rule 2.1 “do not make noun clusters of more than three nouns”)



ASD Simplified Technical English (ASD-STE)

- ▶ These rules restrict the language on
 - ▶ the lexical level (rule 1.2 “use approved words from the dictionary only as the part of speech given”)
 - ▶ “test” noun not as verb
 - ▶ the syntactic level (rule 2.1 “do not make noun clusters of more than three nouns”)
 - ▶ Non-STE: Runway light connection resistance calibration



ASD Simplified Technical English (ASD-STE)

- ▶ These rules restrict the language on
 - ▶ the lexical level (rule 1.2 “use approved words from the dictionary only as the part of speech given”)
 - ▶ “test” noun not as verb
 - ▶ the syntactic level (rule 2.1 “do not make noun clusters of more than three nouns”)
 - ▶ Non-STE: Runway light connection resistance calibration
 - ▶ STE: Calibration of the resistance of the runway light connection



ASD Simplified Technical English (ASD-STE)

- ▶ These rules restrict the language on
 - ▶ the lexical level (rule 1.2 “use approved words from the dictionary only as the part of speech given”)
 - ▶ “test” noun not as verb
 - ▶ the syntactic level (rule 2.1 “do not make noun clusters of more than three nouns”)
 - ▶ Non-STE: Runway light connection resistance calibration
 - ▶ STE: Calibration of the resistance of the runway light connection
 - ▶ the semantic level (rule 1.3 “use approved words only with their approved meanings”)



ASD Simplified Technical English (ASD-STE)

- ▶ These rules restrict the language on
 - ▶ the lexical level (rule 1.2 “use approved words from the dictionary only as the part of speech given”)
 - ▶ “test” noun not as verb
 - ▶ the syntactic level (rule 2.1 “do not make noun clusters of more than three nouns”)
 - ▶ Non-STE: Runway light connection resistance calibration
 - ▶ STE: Calibration of the resistance of the runway light connection
 - ▶ the semantic level (rule 1.3 “use approved words only with their approved meanings”)
 - ▶ Non-STE: Follow the safety instruction



ASD Simplified Technical English (ASD-STE)

- ▶ These rules restrict the language on
 - ▶ the lexical level (rule 1.2 “use approved words from the dictionary only as the part of speech given”)
 - ▶ “test” noun not as verb
 - ▶ the syntactic level (rule 2.1 “do not make noun clusters of more than three nouns”)
 - ▶ Non-STE: Runway light connection resistance calibration
 - ▶ STE: Calibration of the resistance of the runway light connection
 - ▶ the semantic level (rule 1.3 “use approved words only with their approved meanings”)
 - ▶ Non-STE: Follow the safety instruction
 - ▶ STE: Obey the safety instruction



Français Rationalisé

- GIFAS - Groupement des Industries Françaises Aéronautiques et Spatiales
 - Start of works: 1985
 - Guide du français rationalisé published in 1998
- The objectives of FR were twofold:
 - Enable authors to write in a French that is easily translatable into SE
 - Increase clarity, even for native French-speaking readers



Italiano Tecnico Semplificato (ITS)



ITS è un progetto

COM&TEC
Associazione Italiana per la
Comunicazione Tecnica

Cos'è?

L'ITS è il Linguaggio Naturale Controllato della lingua italiana.

"Naturale" perché usa la lingua che naturalmente si è affermata e sviluppata nella cultura Italiana. **"Controllato"** perché fornisce semplici istruzioni per scrivere e redigere testi chiari, sicuri e facilmente comprensibili.

L'ITS è uno strumento innovativo per:

- Comunicare in maniera Sicura, Efficace e a Norma nell'era digitale
- Scrivere e far capire subito e senza fraintendimenti
- Elaborare messaggi univoci, semplici e di facile assimilazione attraverso sintesi e standardizzazione linguistica
- Scrivere un manuale, una guida o un qualsiasi testo Tecnico versatile, utilissimo anche per la redazione e la revisione di qualsiasi altra tipologia di informazione e comunicazione destinate all'uso interno ed esterno
- Comunicare in maniera globale.

<https://www.italianotecnicosemplificato.it>



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Federica Vezzani | SimpleText@CLEF-2023, September 18th 2023, Thessaloniki, Greece



T4TEC Project

Terminology for Technical Efficient Communication

T4TEC

The screenshot displays the T4TEC interface, which consists of two main sections: a left sidebar with various input fields and a right panel for managing a local termbase.

Source: Contains the text "The volume control can be adjusted."

CNLs Rules: Contains the rule "Rule 3.4: 'Do not use helping verbs to make complex verb structures.'"

Simplified: Contains the simplified text "You can adjust the volume control".

Check Existing Termbases: A placeholder section.

Local Termbase: A search bar labeled "type here to search..." and a "Term" input field labeled "type here a new term...". Below these are "Create" and "Update" buttons.

Data Categories: Tabs for Semantics, Formal features (which is selected), Variation, and Usage.

Definition: An empty text area for defining terms.

External cross-ref (definition): An empty text area for external cross-references.

Di Nunzio, G. M., Vezzani, F., e Sicilia, T. (2022). T4TEC: A Prototype for Simplified Multilingual Technical Communication.

In O. Alonso, R. Baeza-Yates, T. H. King, & G. Silvello (Eds.), Proceedings of the Third International Conference on Design of Experimental Search & Information REtrieval Systems, San Jose, California, USA, August 30- 31, 2022. CEUR-WS.org.

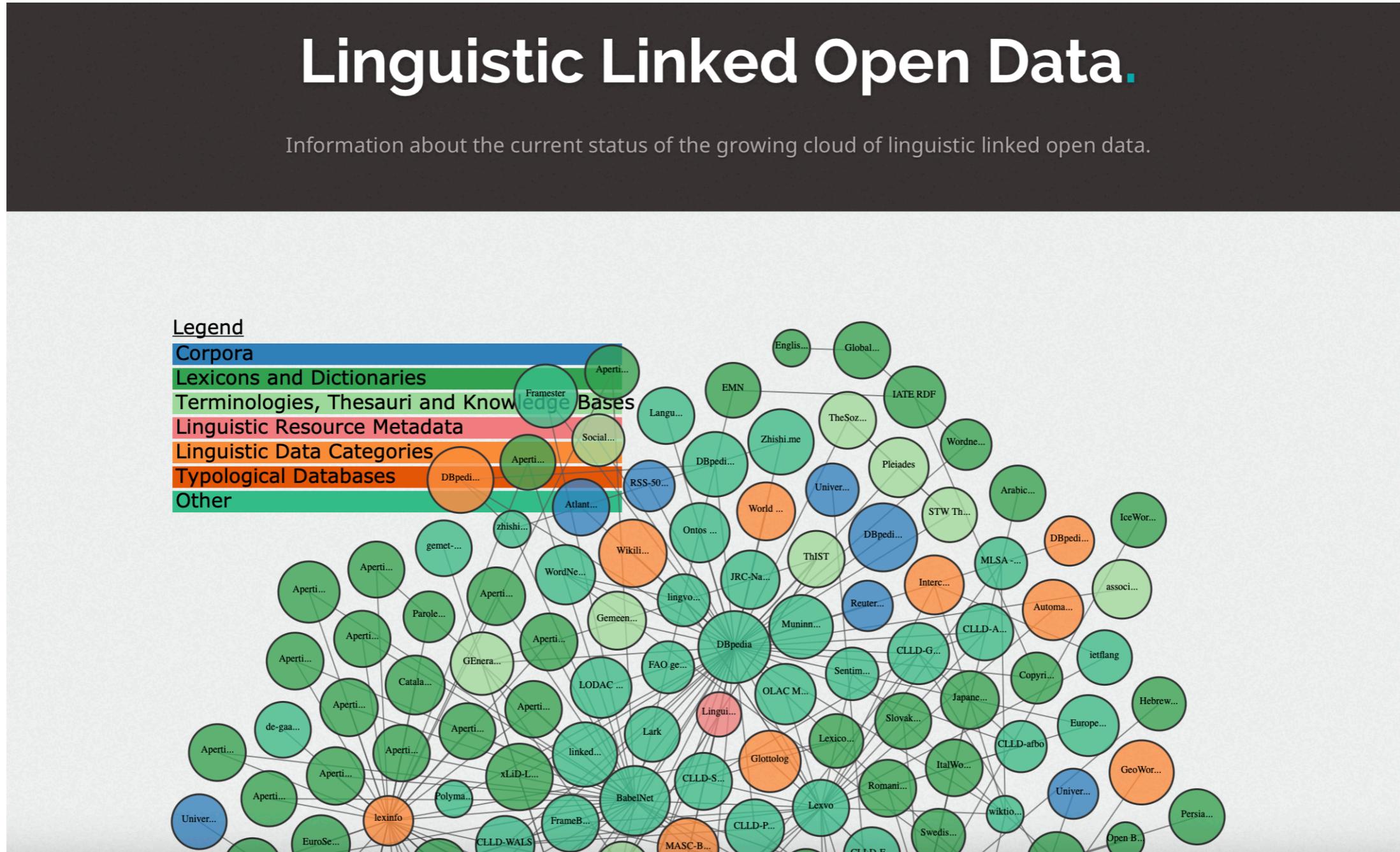
Terminology Management

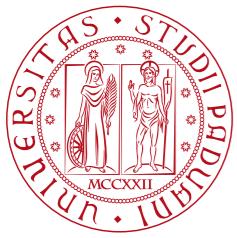
How to manage multilingual terminology?

- ▶ ISO/TC 37 “Language and terminology”
- ▶ ISO/TC 37/SC 3 “Management of terminology resources”
 - ▶ ISO 16642: 2017 “Computer applications in terminology – Terminological markup framework”
 - ▶ ISO 30042: 2019 “Management of terminology resources – **TermBase eXchange (TBX)**”



How to manage multilingual terminology?





UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Thank you for your attention!

Federica Vezzani
federica.vezzani@unipd.it

Department of Linguistic and Literary Studies
Università degli Studi di Padova

SimpleText@CLEF-2023 - Automatic Simplification of Scientific Texts
September 18th 2023, Thessaloniki, Greece

