



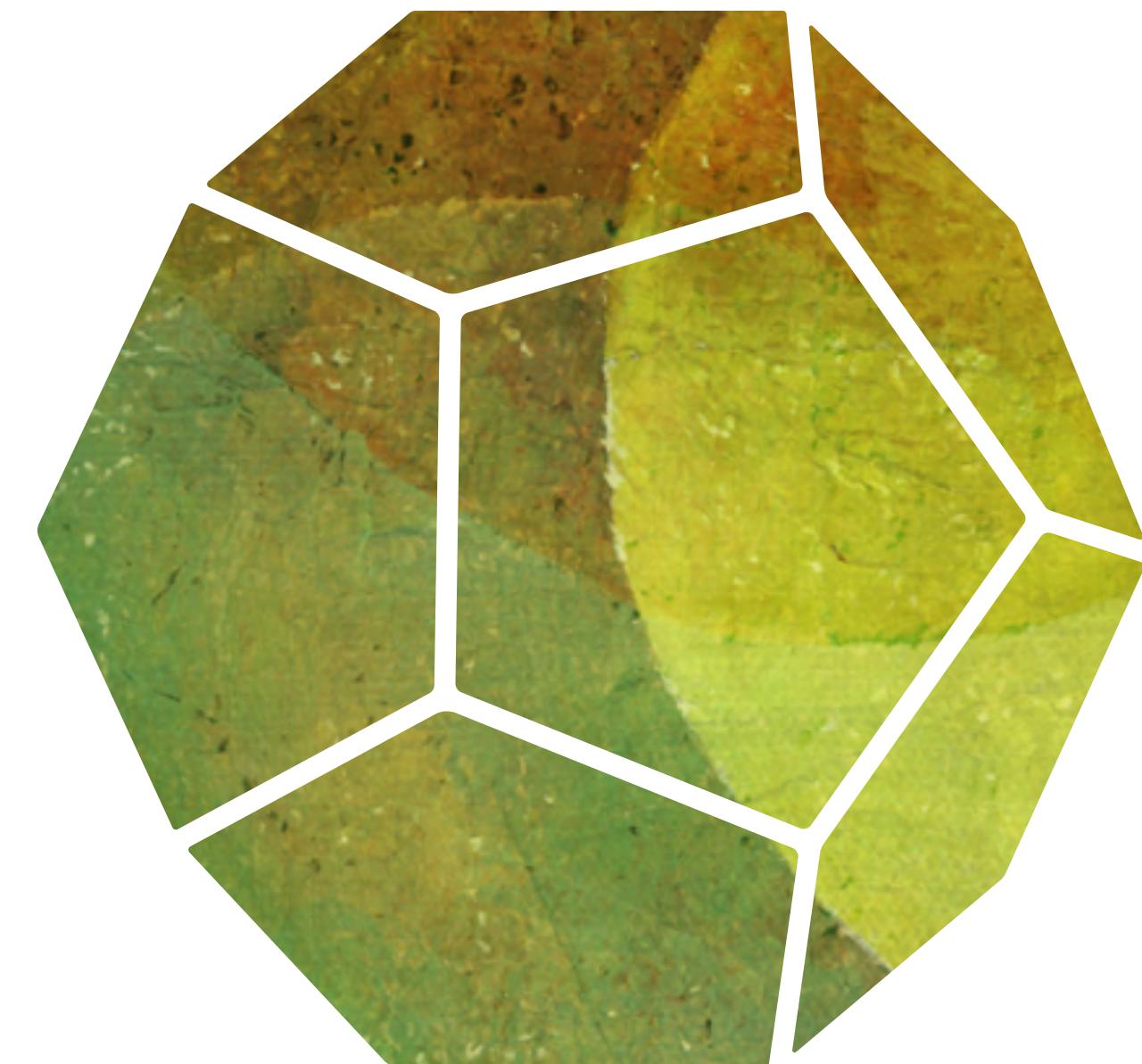
COMPUTATIONAL LINGUISTICS * DIGITAL HUMANITIES

Gemma Bel Enguix
UNAM

Summer School on Digital Humanities
UNAM, September 13-14, 2023

OUTLINE AND PRESENTATION

- LANGUAGE IS A COMPLEX OBJECT
- LANGUAGE NEEDS INTERDISCIPLINARY RESEARCH
- ROUND TRIP: TECHNOLOGY LANGUAGE TECHNOLOGY
- COMPUTATIONAL LINGUISTICS * DIGITAL HUMANITIES



LANGUAGE IS A COMPLEX OBJECT

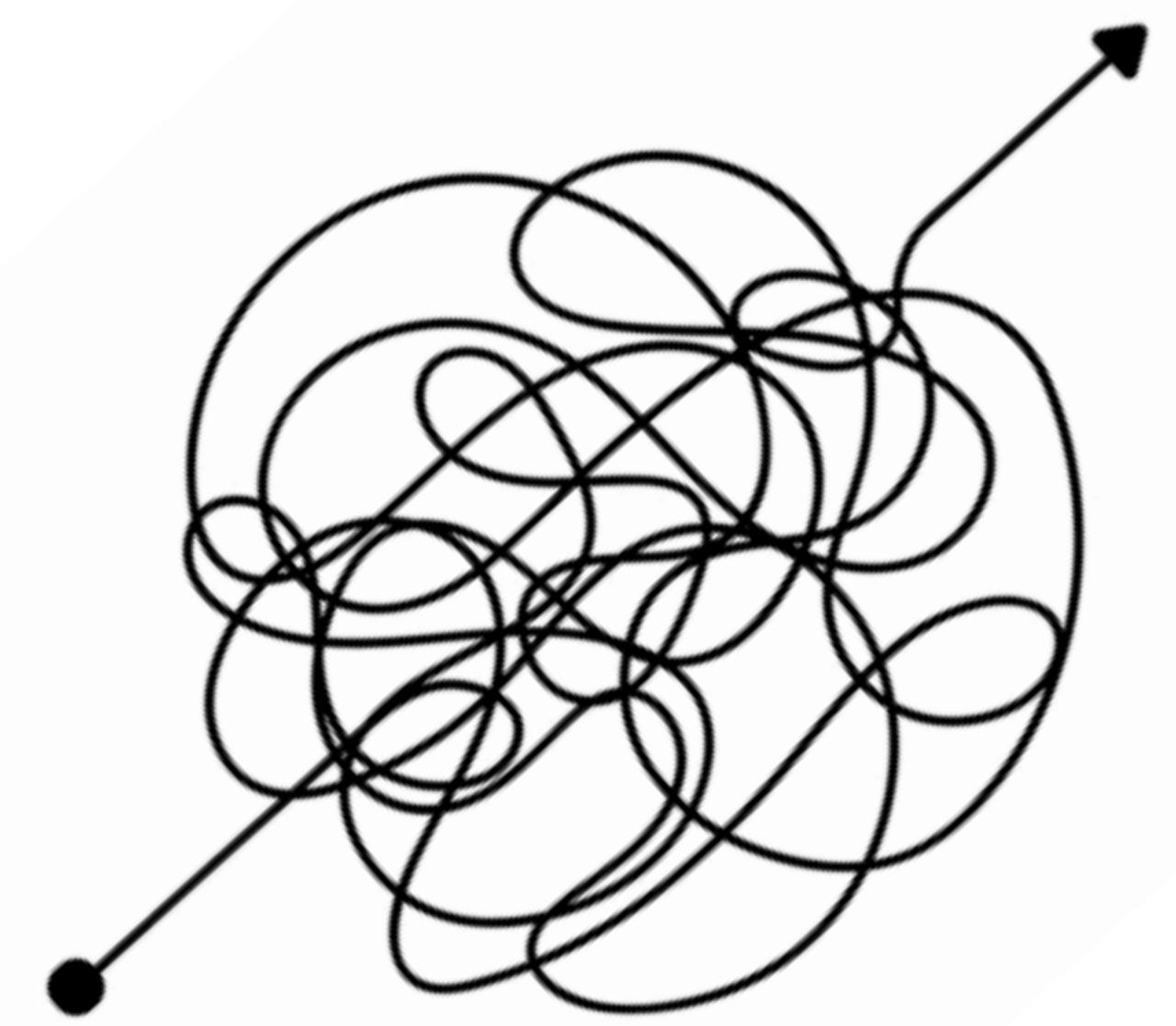


A system is complex if it displays some or all of the following properties: it is open, situated, boundary-free and replete with unstable individual; “infinitively” cardinal, uncontrollable and uncertain; dynamic, metastable and path dependent; non-linear, sensitive to initial conditions, exponentially amplifiable and in regions chaotic; emergent, non-additive, non-modularizable, irreducible and organizationally intricate. It is also self-organizing and adaptive.

Andrason, 2014

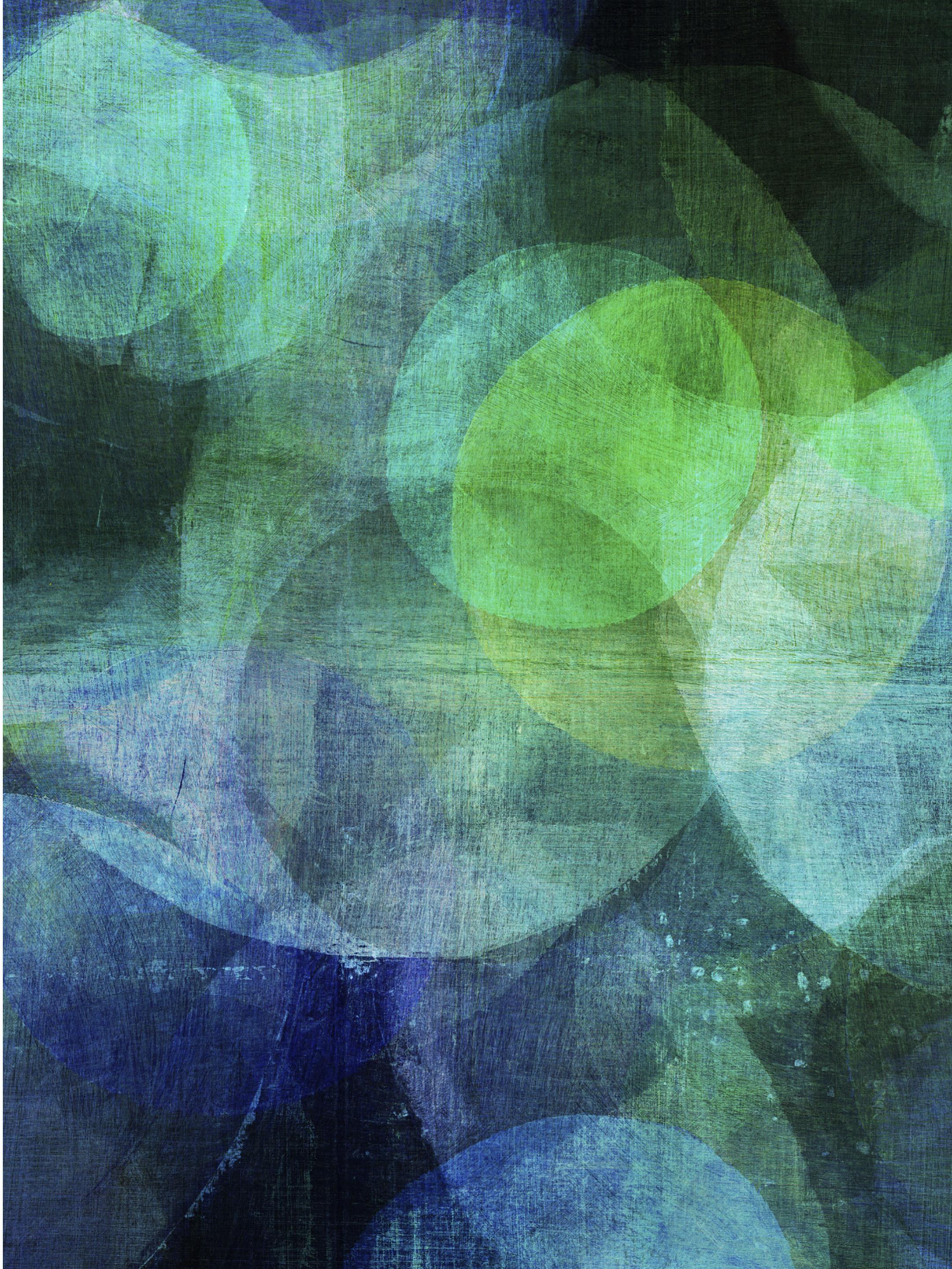
LANGUAGE IS A COMPLEX OBJECT

- LANGUAGE consists of many components that interact each other.
- Complex systems have emergent behaviour. Their behavior depends not only on the parts of which they are composed but also on the whole in which they are embedded.
- Systems that are emergent are non-resultant, non-additive and non-modularized: they cannot be explained by their microanalysis into independent parts because they are not mere superposed computations of their isolated components.



LANGUAGE IS A COMPLEX OBJECT

- The **line between** the **system** and its **environment** becomes **fluid** and the concept of boundary highly problematic.
- The **environment participates** in the system's **behaviour** and **regulates** it, being in turn simultaneously **influenced** by the system.
- When **reduced** to subsystems, they suffer a **loss of the information** that makes these systems a system.



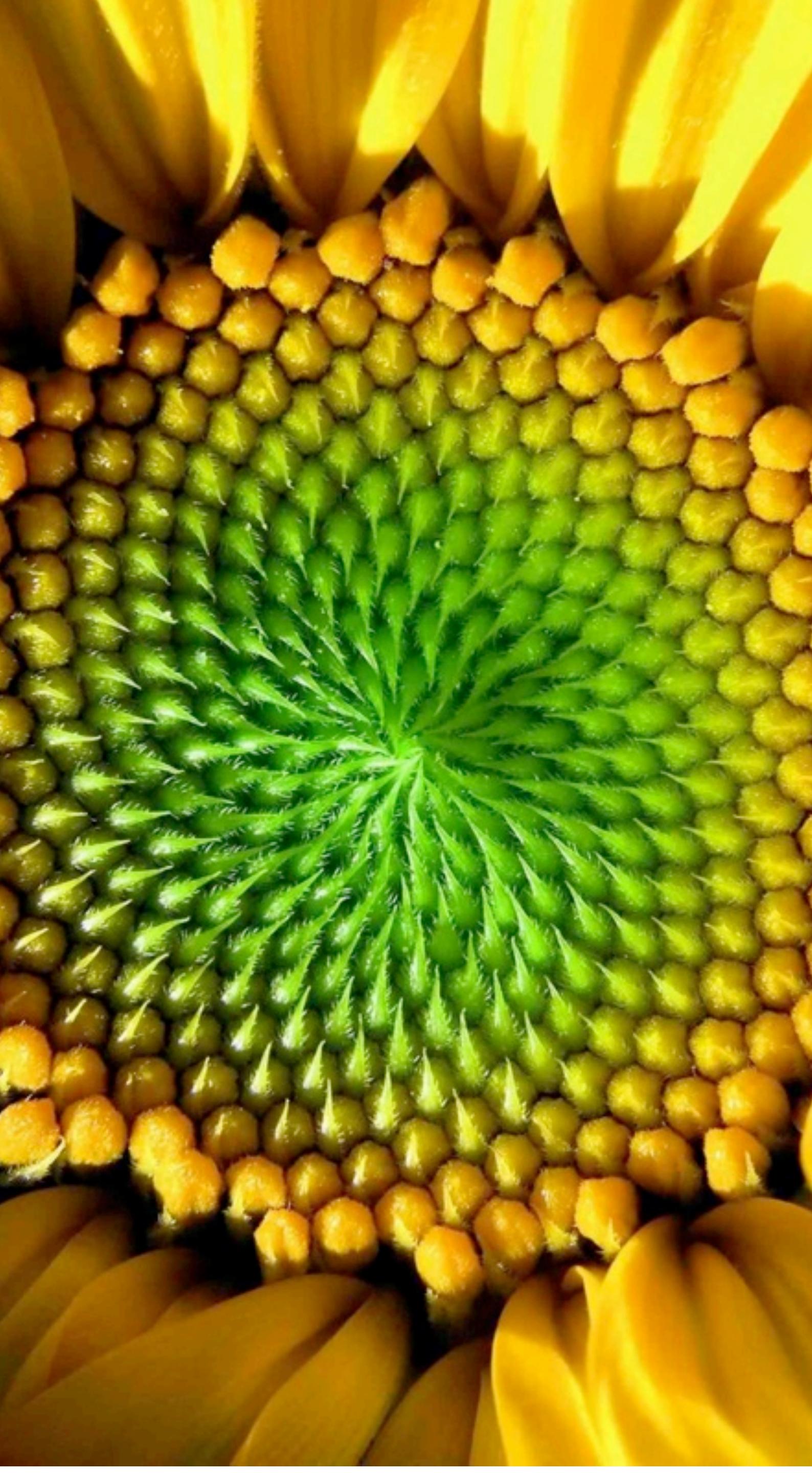
LANGUAGE IS A COMPLEX OBJECT



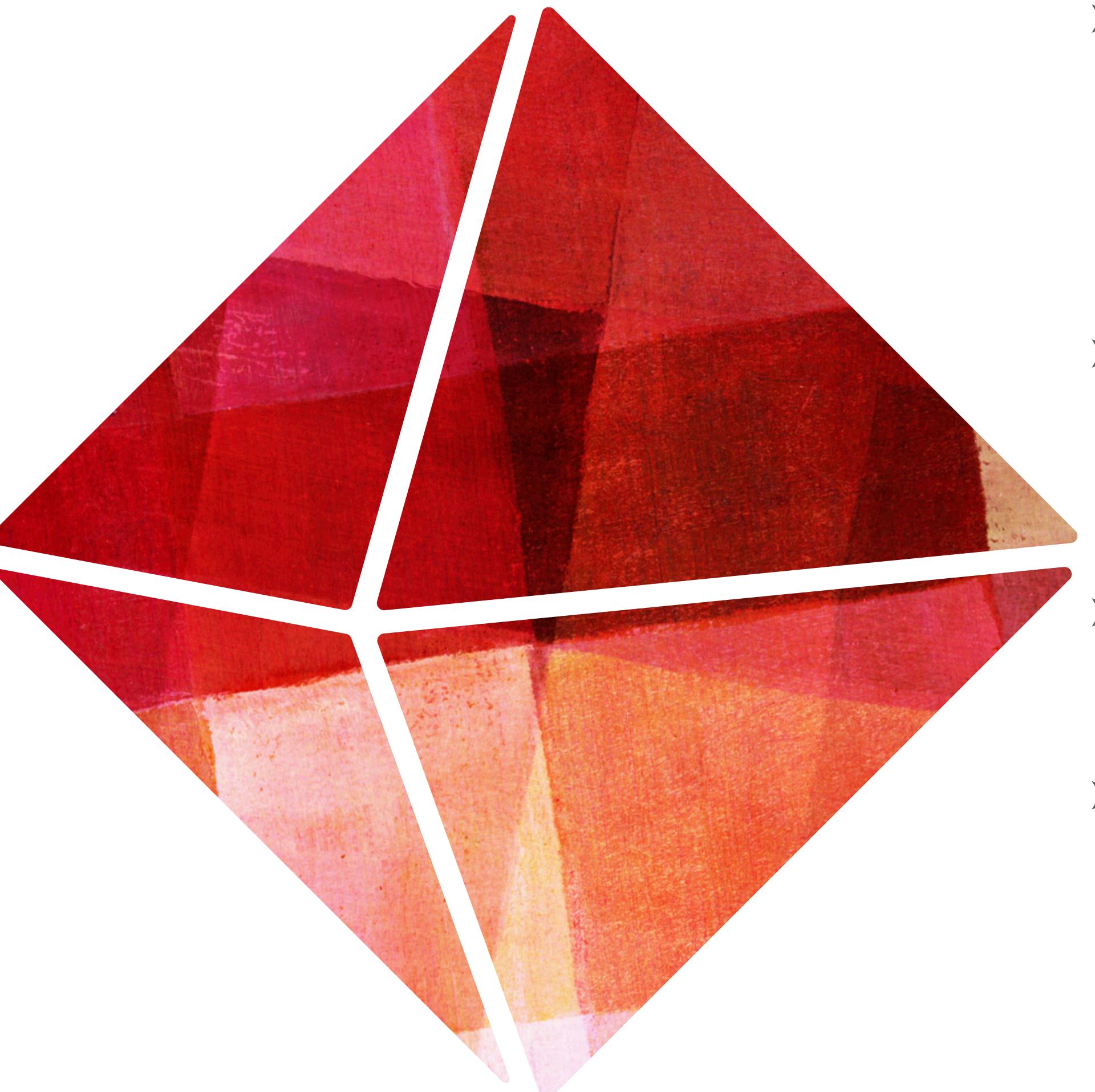
- The dynamic understanding of components of a system and the system itself as processes implies that **the system strongly depends on its history**.
- **Path dependence** implies that the momentum of the system is regulated by the precise and dynamic conditions where the first “step” was made.
- **Non-linearity**: system does not satisfy the superposition principle: its **outputs** are **not directly proportional** to the **inputs**.

LANGUAGE IS A COMPLEX OBJECT

- Non-linearity makes systems **uncontrollable**, both **synchronously** and **historically**.
- Complex systems are **incompressible** and **irreducible**. Their models are always incomplete.
- **Language evolves**. Language is an inherently dynamic phenomenon corresponding to a network of incessant fluctuations and modifications.



LANGUAGE IS A COMPLEX OBJECT



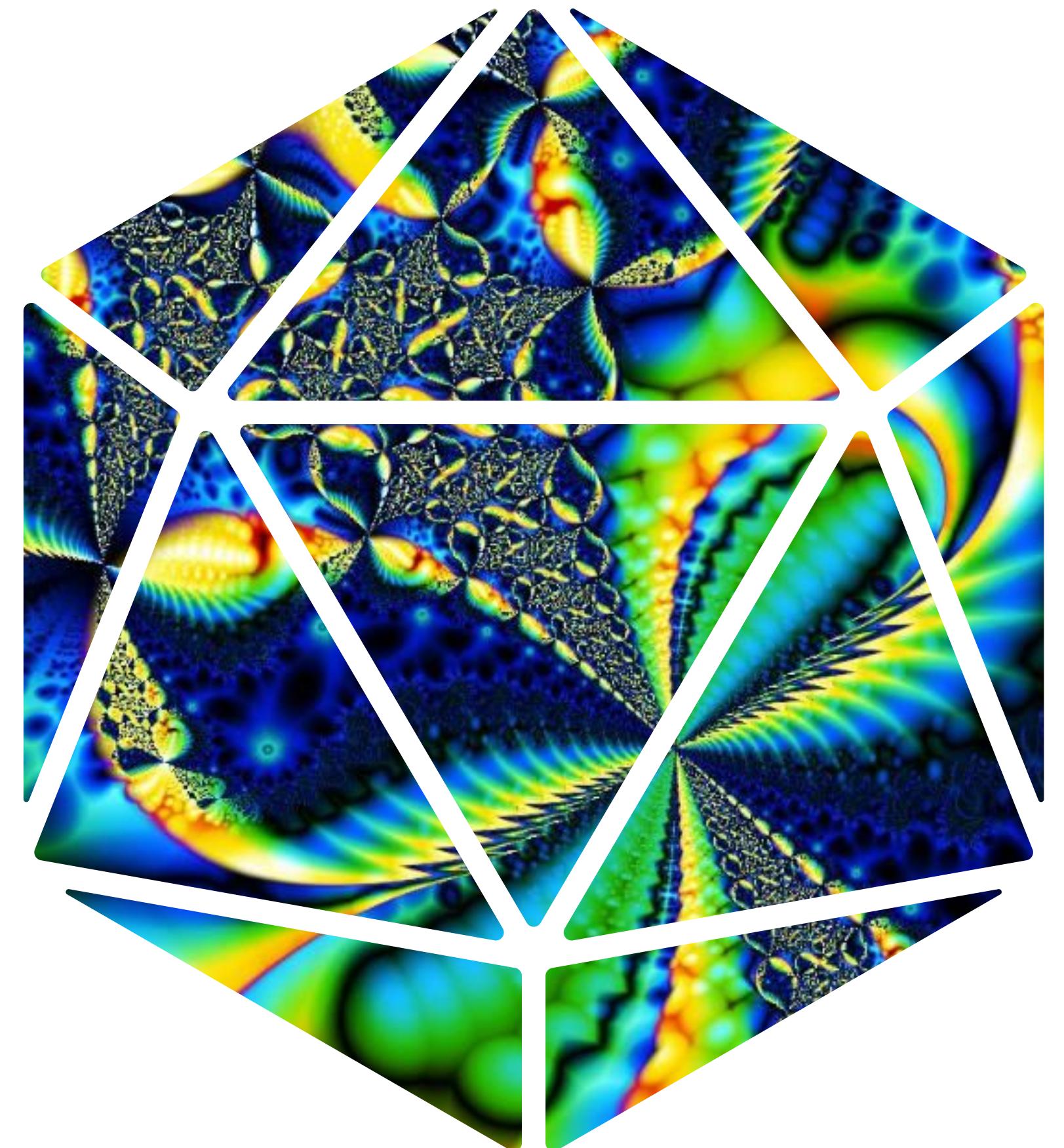
- Language is a prototypical open system that constantly exchanges material and energy with the environment. It influences our reality and perception, being, at the same time, affected by the external world.
- Language is always a situated phenomenon: it is invariably embedded in a culture, social organization and higher bodies on which it depends and to which it simultaneously contributes.
- Any fragment of a language is embedded in a larger system so that no components can be viewed as isolated with clear boundaries.
- In general, boundaries at which language comes into contact with the physical world, biology of human mind or socio-cultural institutions, are vital parts of the linguistic system itself, so that, under certain approximations, physics, biology, sociology and culture can be considered important spheres of languages.

LANGUAGE IS A COMPLEX OBJECT

- High cardinality (of components and relations or rules) constitutes only one of the features that make language an exemplary complex system. The cardinality of the components of a language is extreme.
- The amount of combinations between components that belong to various levels in clauses, phrases and sentences is absolutely unmeasurable.
- Macro - micro: Various levels exist embedded in one another and influence one another: each one with its specific emergent properties that differentiate it from the lower and higher planes.

LANGUAGE IS A COMPLEX OBJECT

- LANGUAGE is typically incompressible, model-specific and model-plural.
- Therefore:
 - There is Plurality of Models
 - And a continuum of models, because all of them represent the system just partially.



LANGUAGE IS A COMPLEX OBJECT

Taken in its entirety, language is multiform and heteroclite; It belongs to several domains, simultaneously: physical, physiological and psychological, it also belongs to the individual sphere and the social sphere. It cannot be classified into any category of human events, because it is not known how to extract its unity.

Ferdinand de Saussure

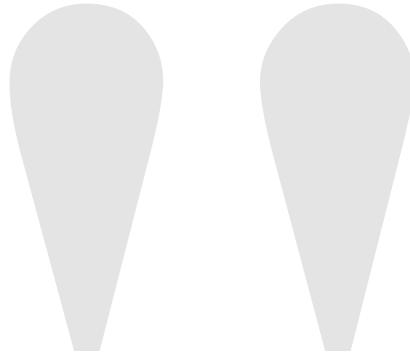


LANGUAGE IS A COMPLEX OBJECT



You can't study the language itself, because the language is a hell of a mess. But don't put that on your midterm exam.
Say..., I don't know..., say it is multiform and heteroclite

LANGUAGE NEEDS INTERDISCIPLINARY RESEARCH

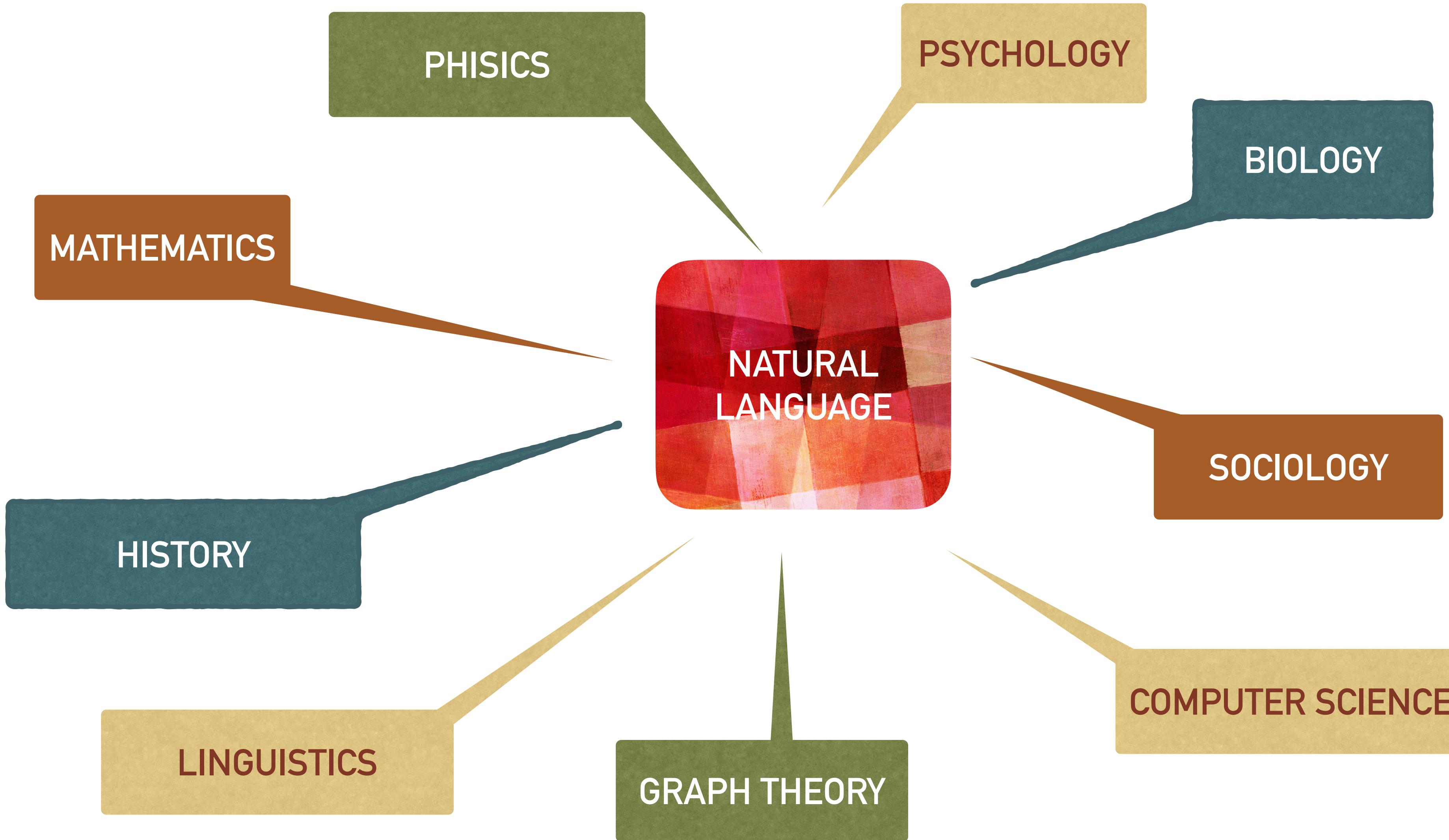


In interdisciplinary thinking and working, you try to integrate perspectives or insights from different perspectives through interaction, in order to better understand a complex phenomenon. In this, you can do more together than alone ('it goes beyond a simple sum of its parts'). Integration can occur at the level of methods, tools, concepts, theories, or insights.

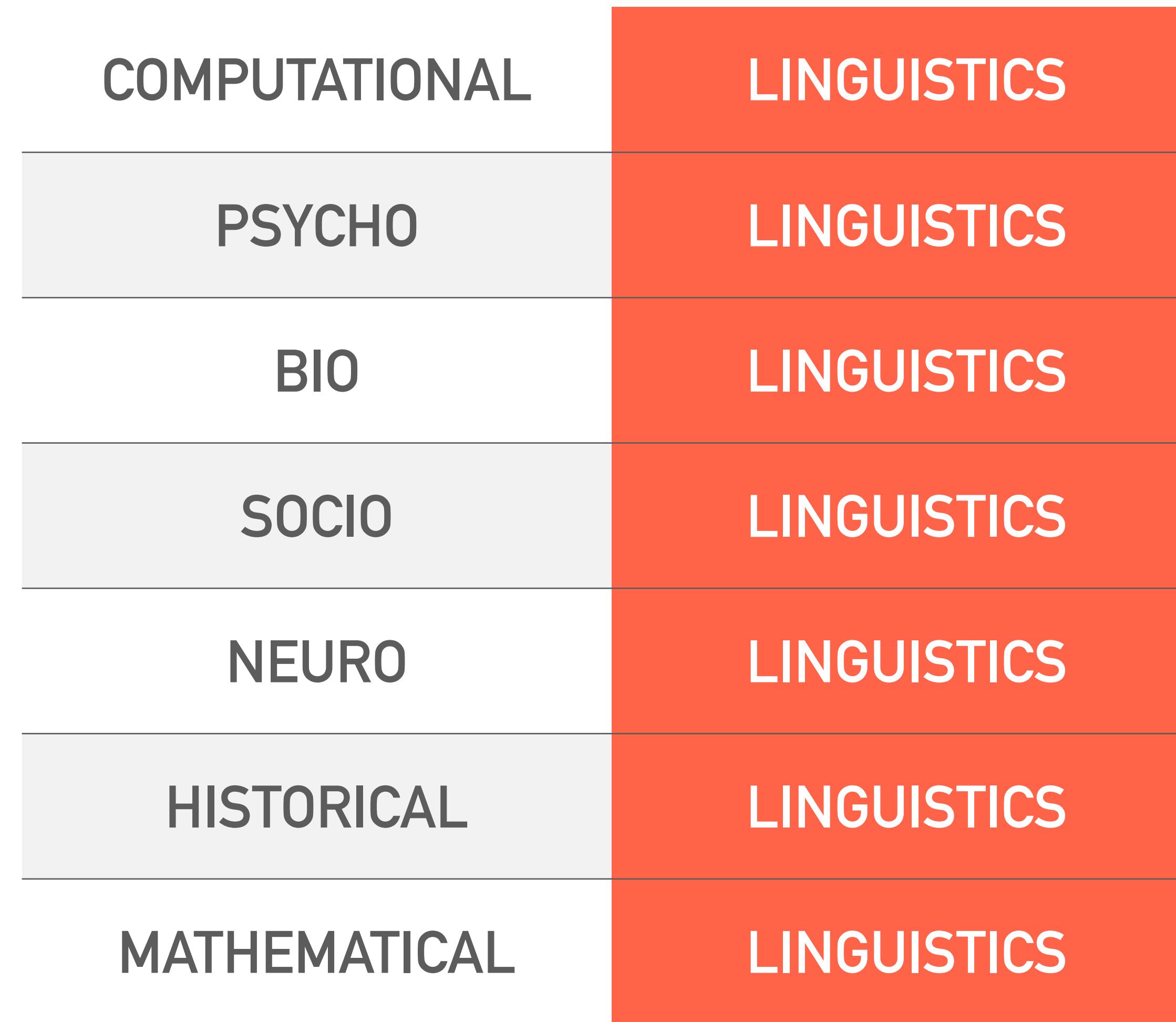
dr. Esther Slot

<https://www.uu.nl/en/education/>

LANGUAGE NEEDS INTERDISCIPLINARY RESEARCH



INTERDISCIPLINES – HYBRID DISCIPLINES



There is Plurality of Models

There is a continuum of models, because all of them represent the system just partially.

TRANSVERSAL METHODOLOGIES

GRAPH THEORY

STATISTICS



COMPUTATION

MACHINE LEARNING

PYTHON

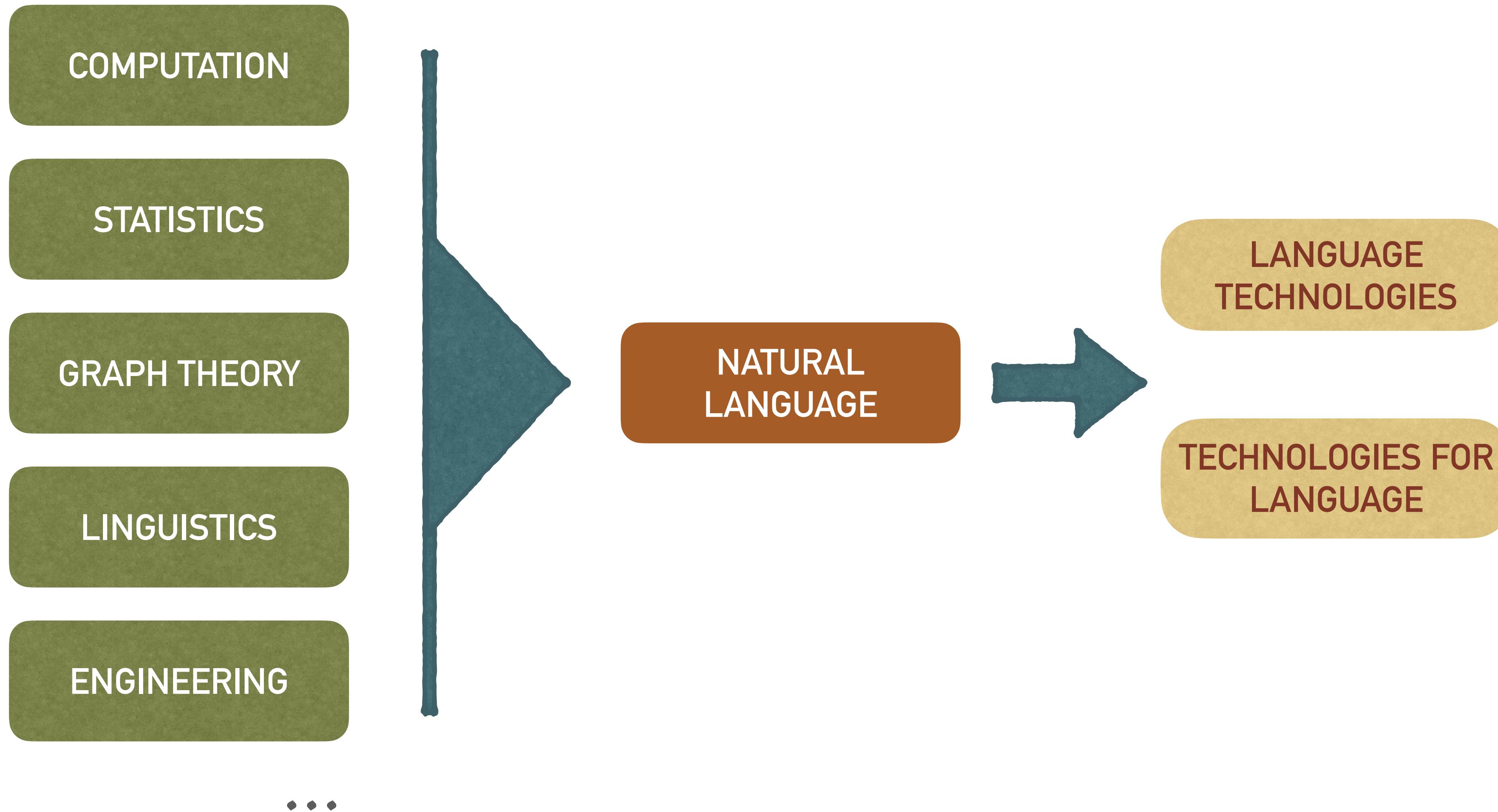
NEURAL NETWORKS

SOME IMPLICATIONS IN LANGUAGE RESEARCH

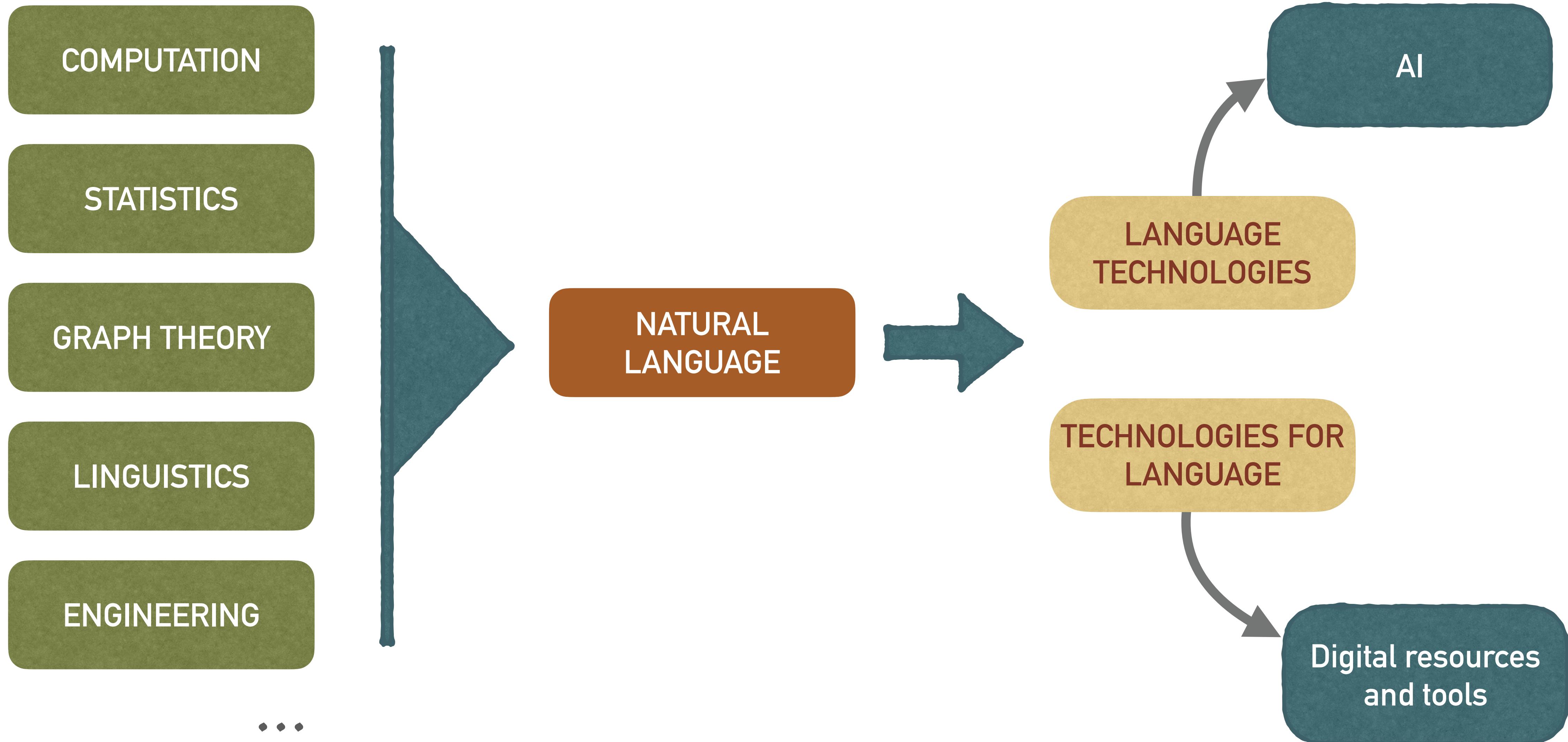
- There is no distinction between **SYNCHRONY AND DIACHRONY**
 - Dependency of the path
 - Language is a system that evolves by definition
 - There is adaptation to the environment
- No **MODULARITY**
 - It's not possible isolating a problem postulating it is not connected to the rest
- Language is not a **QUALITATIVE** discipline any more
 - Scientific methods imply the necessity of **QUANTIFICATION**
 - Data collection, analysis and interpretation
- However, an F1-score is not enough for approaching language.
 - Mixed-Methods research

**TECHNOLOGY
LANGUAGE
TECHNOLOGY (AI)**

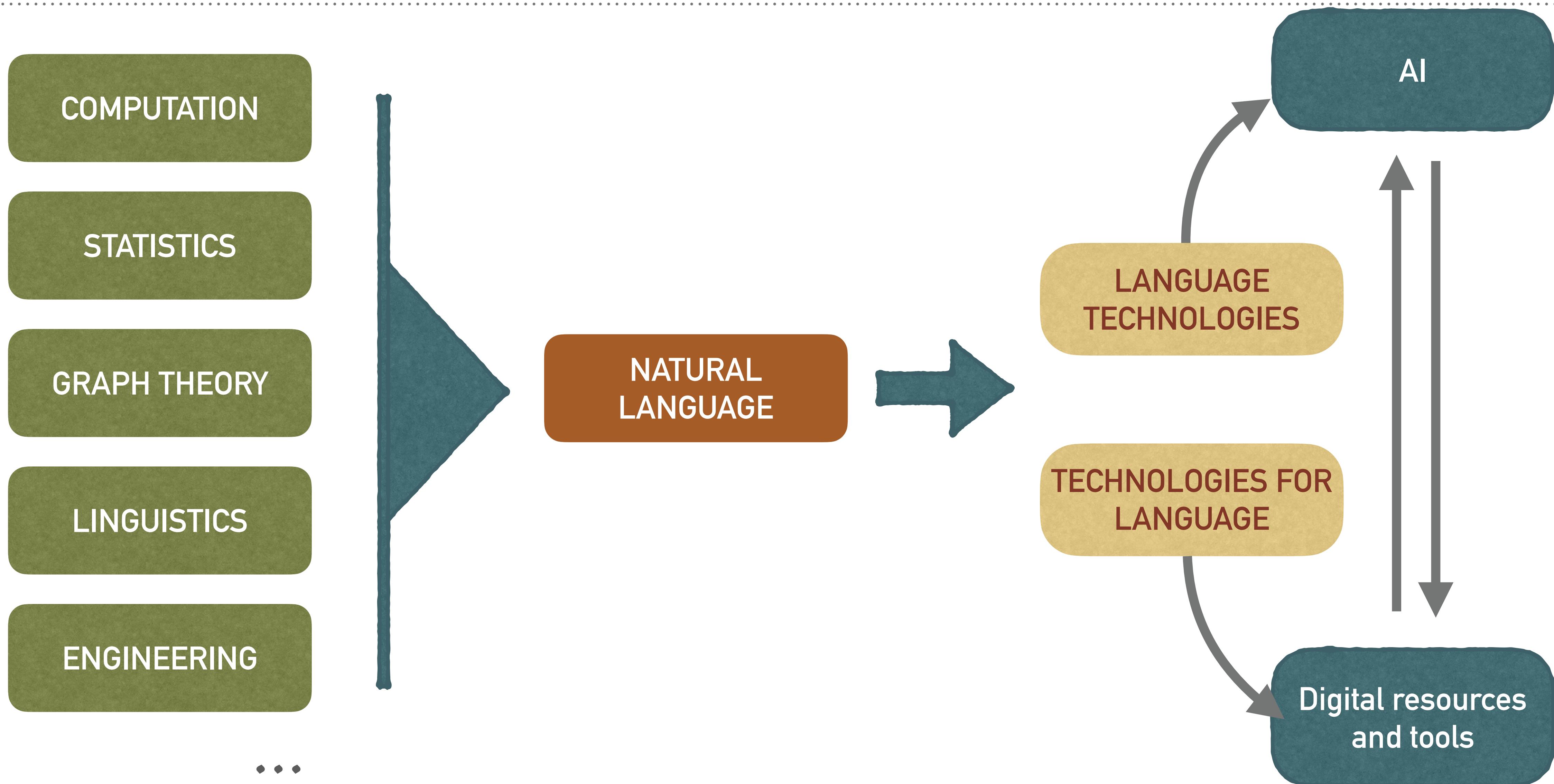
LANGUAGE UNDERSTANDING AND TECHNOLOGIES



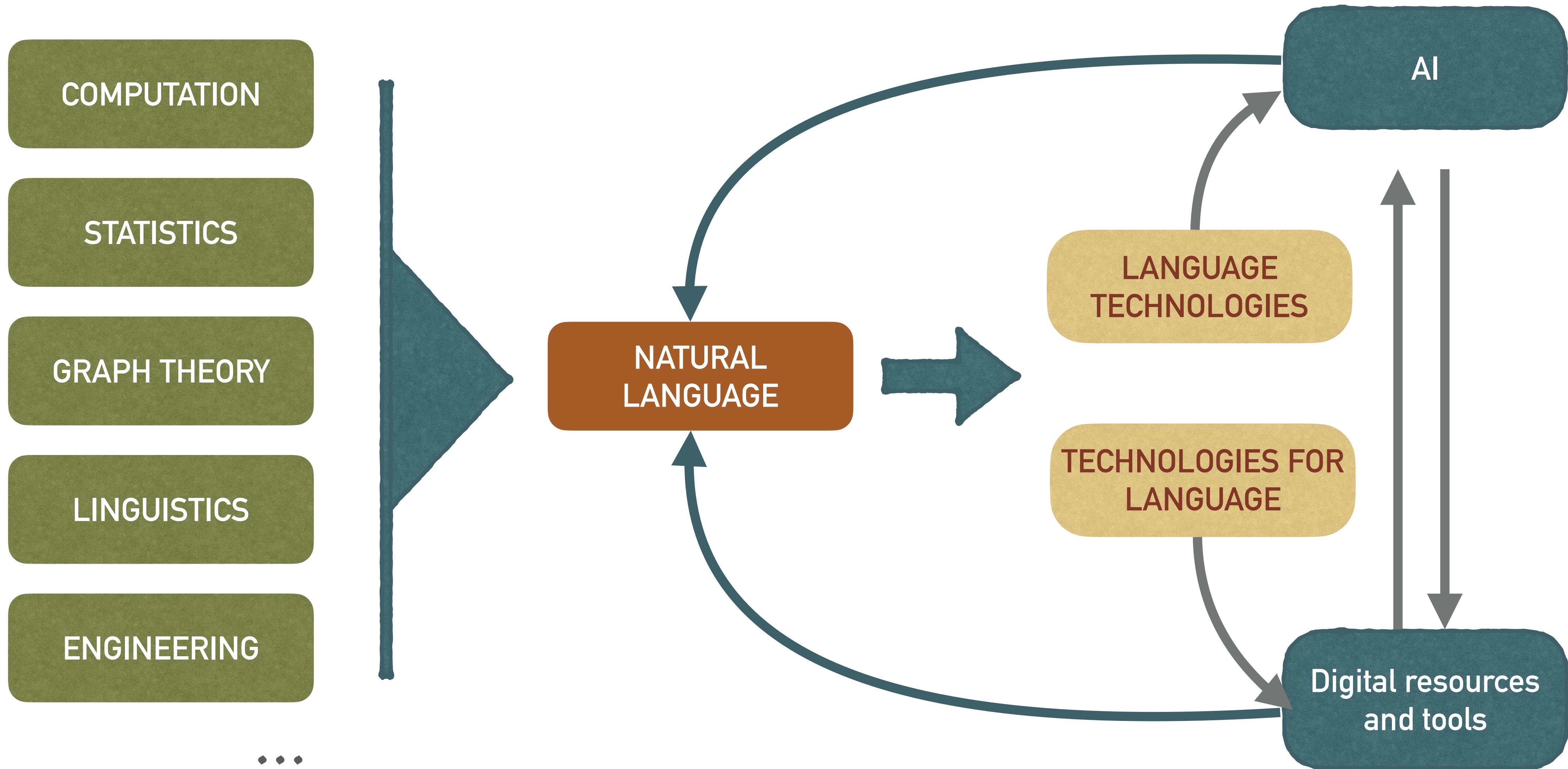
AI AND DIGITAL HUMANITIES



AI AND DIGITAL HUMANITIES



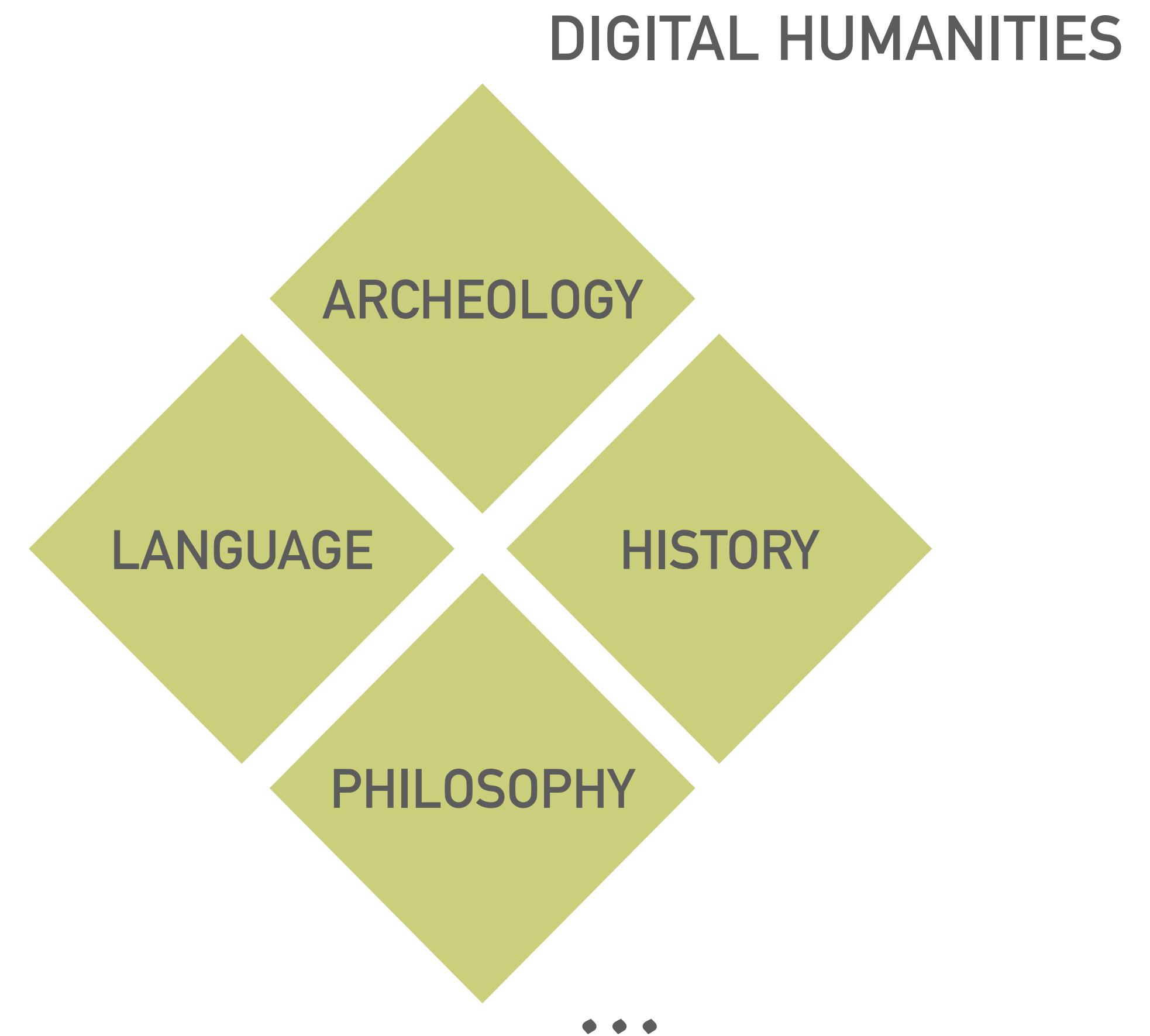
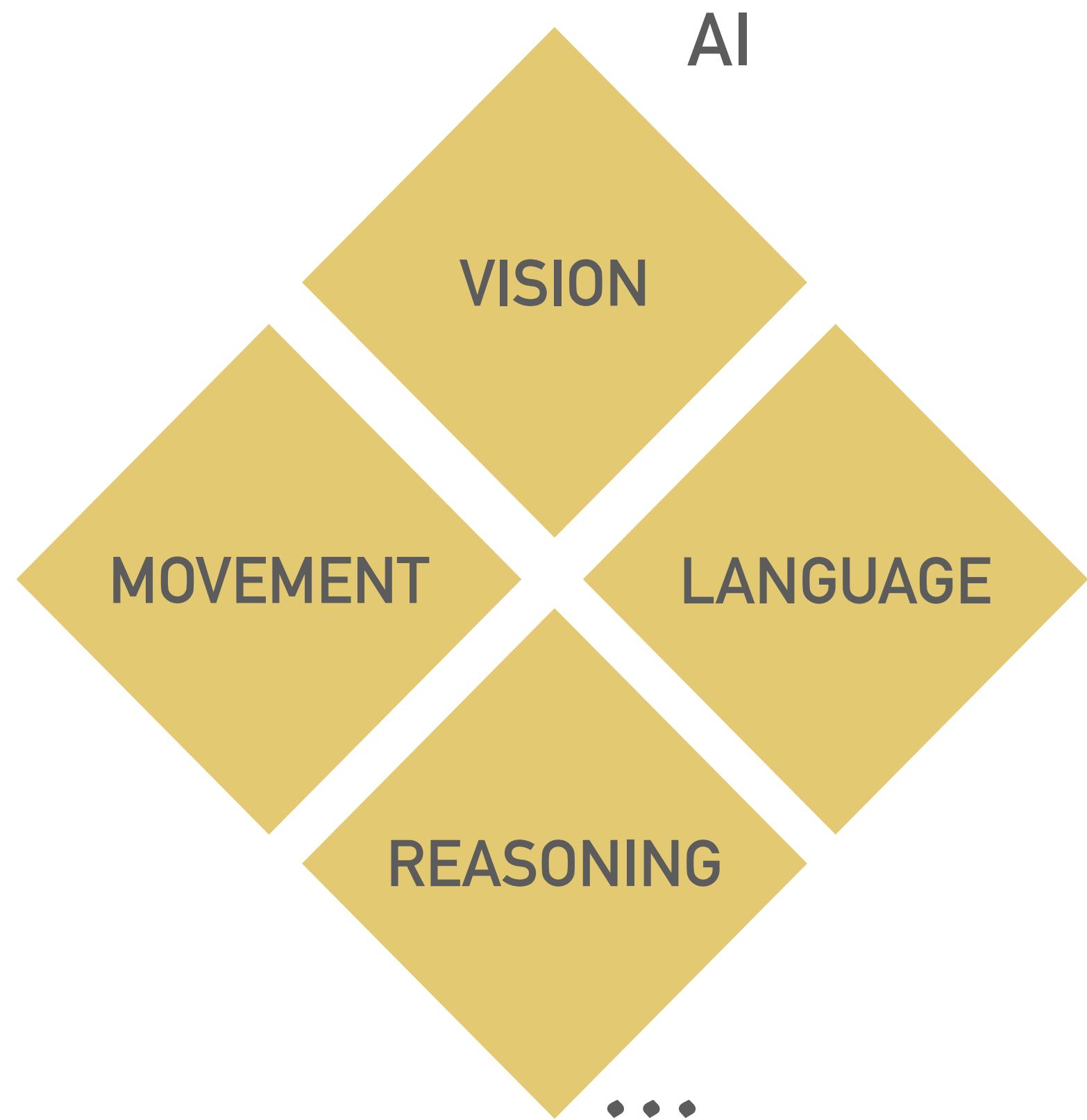
AI AND DIGITAL HUMANITIES



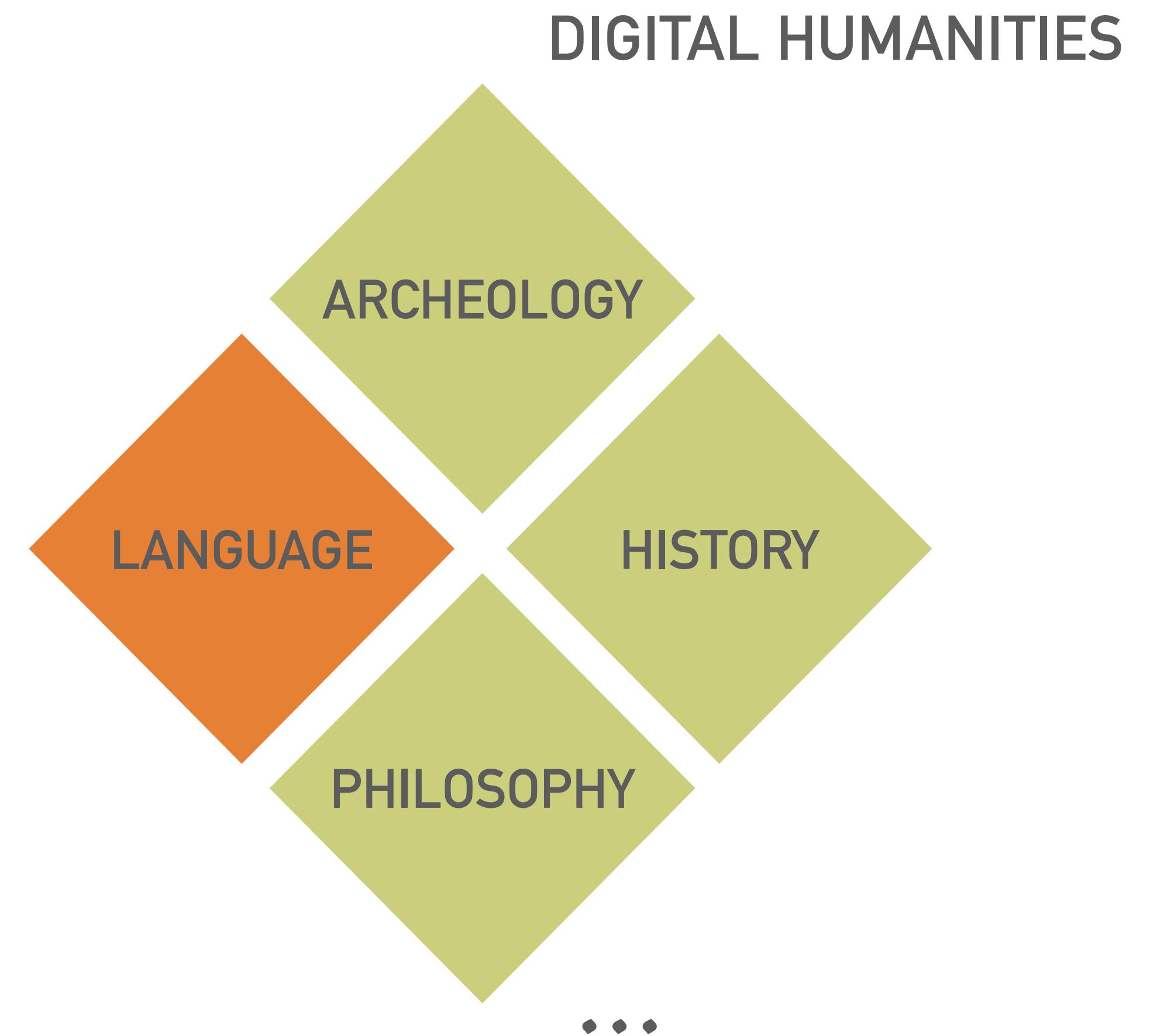
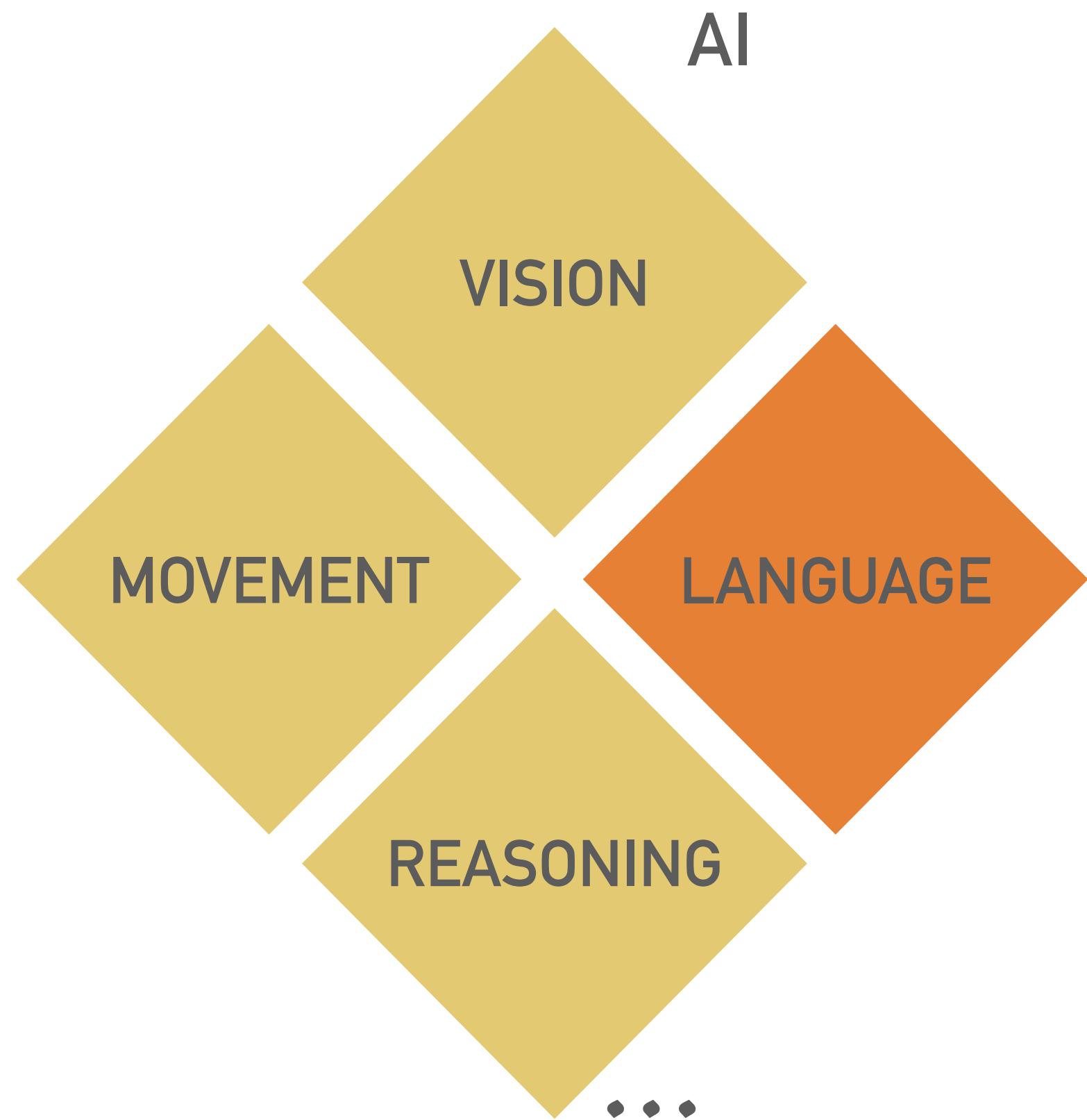
COMPUTATIONAL LINGUISTICS

DIGITAL HUMANITIES

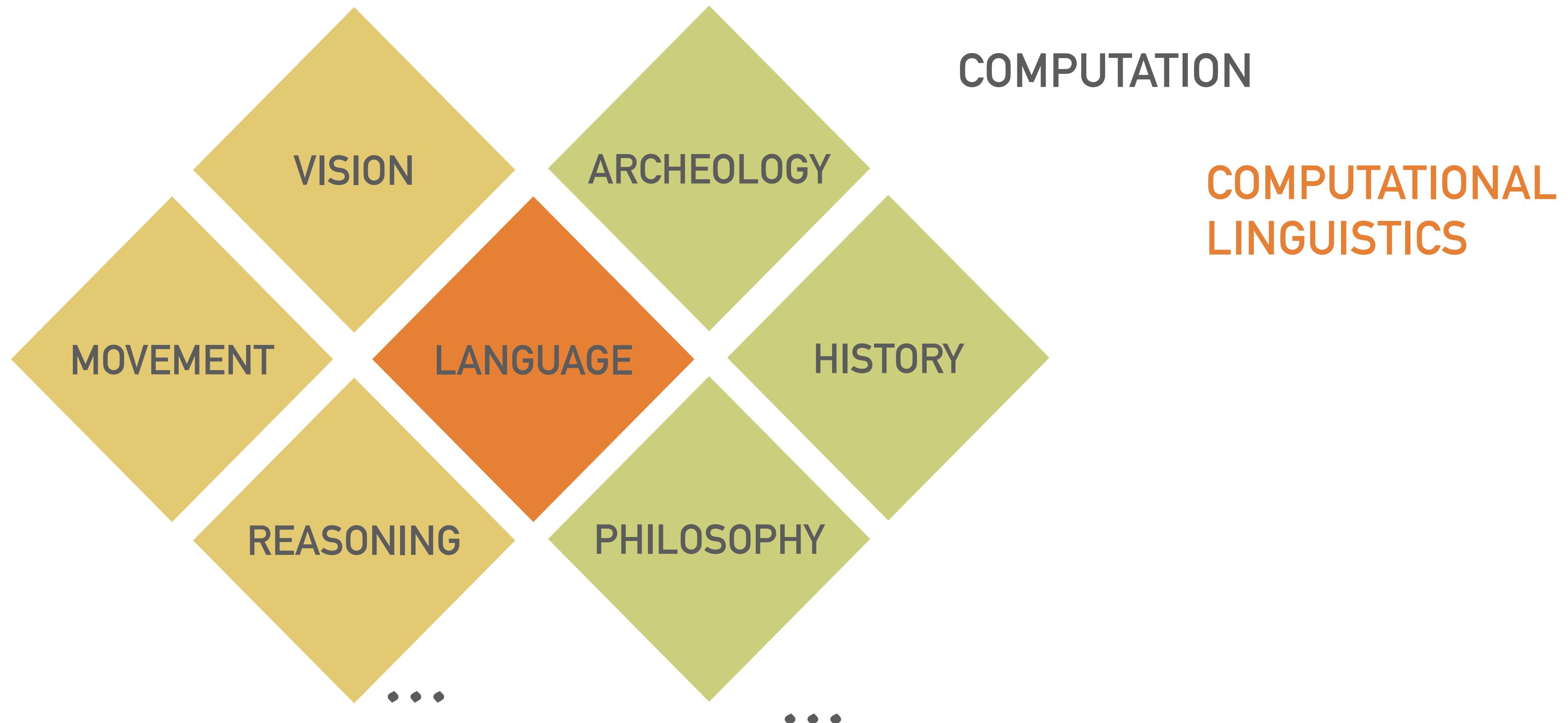
AI AND DIGITAL HUMANITIES



AI AND DIGITAL HUMANITIES



AI AND DIGITAL HUMANITIES



COMPUTATIONAL LINGUISTICS * DIGITAL HUMANITIES

- This school deals with Natural Language Processing, the most applied area of computational linguistics
- We are introducing some new and classical methods and techniques that are the state of the art for natural language research: Markov models, Transformers...
- Also, the applications of these techniques: text classification, medical entity detection, language generation, emotions...
- Moreover, we are presenting some tools for language management and analysis: Geco, Pandora, SAUTEÉ, TEXT2GRAPH_API...
- Combining the design of applications for language analysis with theoretical approaches and novel techniques,
- All of them on natural language, its understanding and applications WITH COMPUTERS

-
- Alexander Andrason. Language Complexity: An Insight from Complex-System Theory, International Journal of Language and Linguistics. Volume 2, Issue 2, March 2014 , pp. 74-89. doi: 10.11648/j.ijll.20140202.15
 - Beckner, Clay, Nick C. Ellis, Richard Blythe, John Holland, Joan Bybee, Jinyun Ke, Morten H. Christiansen, Diane Larsen-Freeman, William Croft & Tom Schoenemann. 2009. Language is a complex adaptive system: Position paper. *Language Learning* 59(1). 1–26. DOI: 10.1111/j.1467-9922.2009.00533.x