

## **Assegno di ricerca**

### **Word Specificity and other Lexical Features**

### **in Assessing Text Complexity and Performing Text Simplification**

**ABSTRACTION project (ERC-2021-STG-101039777)**

**PI: Marianna Marcella Bolognesi.**

**[ITA]**

#### **Descrizione:**

Il progetto ABSTRACTION, finanziato da una borsa europea ERC, si concentra sullo studio della rappresentazione cognitiva e computazionale di parole che variano in concretezza e specificità. L'abstract del progetto ABSTRACTION è riportato in fondo al bando.

L'assegno di ricerca sarà incentrato sul capire principalmente come la concretezza e la specificità delle parole (variabili spesso escluse dalle analisi linguistiche) influenzino la percezione di complessità di diversi tipi di testo, e di conseguenza la possibilità di semplificazione lessicale del testo stesso. La ricerca si svolgerà sia sulla lingua italiana che su quella inglese. I candidati e le candidate dovranno dunque avere conoscenze linguistiche buone di entrambe le lingue.

L'assegnista formulerà e testerà ipotesi sperimentali relative al come concretezza e specificità lessicali influenzino la percepita complessità di testi appartenenti a diversi generi, in collaborazione con la PI e con gli altri membri del gruppo di ricerca Abstraction.

L'assegnista potrà utilizzare risorse disponibili (benchmarks e formule di leggibilità basate su diverse variabili linguistiche), ma dovrà anche costruire un corpus che comprenda generi testuali diversi, su cui si testeranno gli effetti di concretezza, specificità ed altre variabili lessicali sulla percepita complessità di testi pensati per diversi tipi di lettore.

#### **Timeline e tempistiche di implementazione:**

##### **Giugno-Settembre 2025**

Studio della letteratura scientifica su come concretezza, specificità ed altre variabili lessicali influenzino la comprensibilità di un testo. Lo studio considera sia la variabilità dei generi testuali che quella sociodemografica dei lettori.

##### **Ottobre 2025 - Febbraio 2026**

Costruzione del corpus, annotazione della sua complessità in base a variabili lessicali; valutazione dei testi da parte di parlanti (crowdsourcing) relativamente alla percepita complessità dei testi stessi; analisi quantitative e qualitative volte a investigare il potenziale effetto delle variabili lessicali sulla percepita complessità dei testi inclusi nel corpus.

##### **Marzo-Maggio 2026**

Stesura degli articoli scientifici e disseminazione dei risultati.

[ENG]

**Description:**

The ABSTRACTION project, funded by a European ERC grant, focuses on the cognitive and computational representation of words that vary in concreteness and specificity. The abstract of the ABSTRACTION project is provided at the end of the call for applications.

The research fellowship will primarily focus on understanding how the concreteness and specificity of words influence the perceived complexity of different types of texts, and consequently, the possibility of lexical simplification of the texts themselves. The research will be conducted in both Italian and English, so applicants must have a good command of both languages.

The research fellow will formulate working hypotheses on how concreteness and specificity influence the perceived complexity of texts from different genres, in collaboration with the Principal Investigator (PI) and other members of the ABSTRACTION research group.

Using available resources related to text complexity (benchmarks and readability formulas based on different linguistic variables) and by building a corpus that includes different textual genres, the research fellow will investigate how concreteness, specificity, and other lexical variables affect the complexity of texts designed for different types of readers.

**Timeline and implementation schedule:**

**June-September 2025**

Review of the scientific literature on how concreteness, specificity, and other lexical variables influence text comprehensibility. The study will consider the variability of textual genres and the sociodemographic variability of readers.

**October 2025 - February 2026**

Corpus construction, annotation of its complexity based on lexical variables; evaluation of texts by speakers (crowdsourcing) regarding the perceived complexity of the texts themselves; quantitative and qualitative analyses aimed at investigating the potential effect of lexical variables on the perceived complexity of the texts included in the corpus.

**March-May 2026**

Writing of scientific articles and dissemination of results.

## **ABSTRACTION project (ERC-2021-STG-101039777)**

Abstract del progetto europeo all'interno del quale è inserito l'assegno:

[ITA]

Le parole, elementi costitutivi del linguaggio, sono etichette che definiscono diversi tipi di categorie. Alcune parole definiscono categorie di entità concrete (gatti, tavoli) mentre altre definiscono entità astratte (eredità, empatia). Alcune parole definiscono categorie generiche che comprendono molte entità diverse (veicoli, arte) mentre altre definiscono categorie più specifiche (auto sportive, impressionismo).

Per ricavare significato dall'esperienza, costruiamo diversi tipi di categorie attraverso meccanismi di astrazione. Concretezza e specificità sono le due variabili coinvolte nei processi di astrazione.

Tuttavia, quando si indagano i meccanismi e gli effetti dell'astrazione, studiosi di diversi campi si concentrano tipicamente solo sulla specificità o solo sulla concretezza. Focalizzandosi su definizioni parziali di astrazione, il dibattito tra le comunità scientifiche è ostacolato e lo sviluppo teorico è compromesso. Questo è anche dovuto al fatto che non esistono risorse lessicali generate dai parlanti per misurare la specificità delle parole.

Il team di ABSTRACTION raccoglierà dati di specificità per migliaia di parole in 2 lingue (inglese e italiano) attraverso una tecnica innovativa di gamification. Utilizzando questi dati e altre risorse lessicali, eseguiremo analisi statistiche al fine di spiegare come la specificità interagisca con la concretezza nei seguenti domini:

- Pensiero: per spiegare i risultati contrastanti che sono stati precedentemente attribuiti alla sola concretezza
- Linguaggio: per costruire testi che siano ottimalmente chiari e informativi per i lettori
- Creatività, per costruire metafore efficaci in diversi contesti

ABSTRACTION spiegherà come la specificità e la concretezza delle parole ci permettono di costruire significato a partire dall'esperienza e raggiungere generalizzazioni di ordine superiore su cui si basa gran parte del nostro pensiero e del nostro parlare. Questo è un argomento particolarmente dibattuto nelle scienze cognitive, relativamente al "grounding" dei concetti astratti, e nella ricerca sull'intelligenza artificiale, dove è ancora poco chiaro come le macchine possano costruire e usare concetti in un modo che rispecchi il funzionamento della mente umana.

[ENG]

Words, language's building blocks, are labels that define different types of categories. Some words define categories of concrete entities (cats, tables) while others define abstract entities (legacy, empathy). Some words define generic categories that encompass many different entities (vehicles, art) while others define more specific ones (sport cars, Impressionism).

To unlock meaning from experience, we construct different types of categories through mechanisms of abstraction. Concreteness and specificity are the two variables that support abstractions.

However, when investigating the mechanisms and effects of abstraction, scholars from different fields typically focus only on specificity or only on concreteness. Relying on different and partial definitions of abstraction, the debate across scientific communities is impaired and the theoretical development is jeopardized. This is also due to the fact that human-generated resources to measure specificity do not exist.

The ABSTRACTION team will collect specificity data for thousands of words in 2 languages (English and Italian) through an innovative gamification technique. Using this data and other lexical resources, we will run extensive statistical analyses aimed at explaining how specificity interacts with concreteness in:

- Thought, to explain contrasting findings that have been previously attributed to concreteness alone
- Language, to construct texts that are optimally clear and informative for the target readerships
- Creativity, to construct effective metaphors in different contexts

ABSTRACTION will explain how word specificity and concreteness enable us to unlock meaning from experience and achieve the higher-order generalizations on which much of our thinking and talking relies. This is a hot topic in cognitive science, where the grounding of abstract concepts is an open question, and in AI research, where it is still unknown how machines may construct and use concepts in the way humans do.