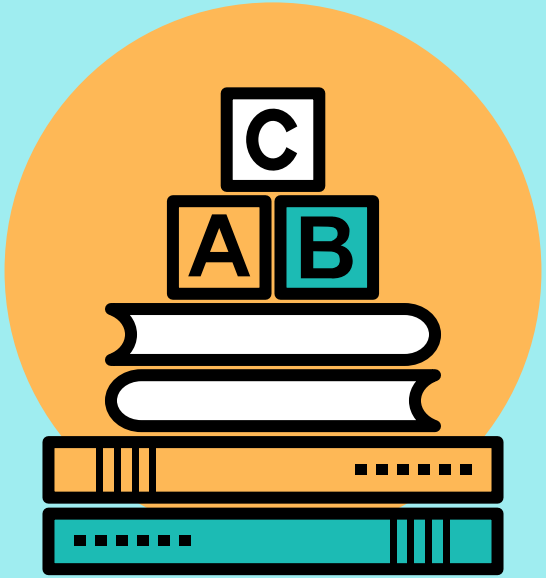



# Ressources interoperability

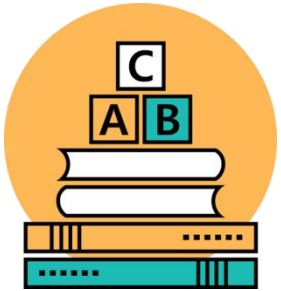
## EXPLOITING LEXICOGRAPHIC DATA TO AUTOMATICALLY GENERATE DICTIONARY EXAMPLES

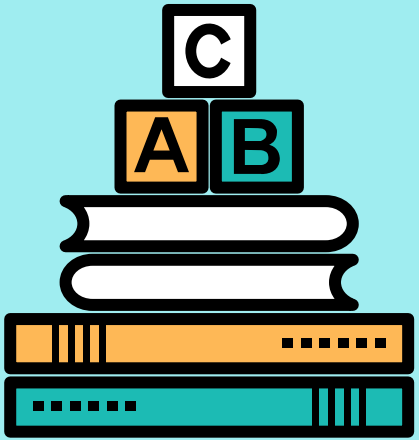
María José Domínguez / Miguel Anxo Solla / Carlos Valcárcel

GOBIERNO DE ESPAÑA MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD
MultiComb
MultiGenera
Fundación BBVA

# Content



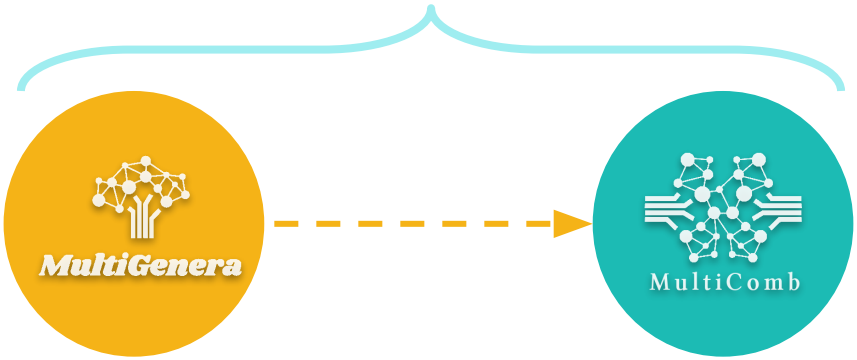





# 1. Core principles

The MultiGenera and MultiComb projects

# Two linked projects


**MultiGenera**

Nominal phrase generation

**MultiComb**

Sentence context generation

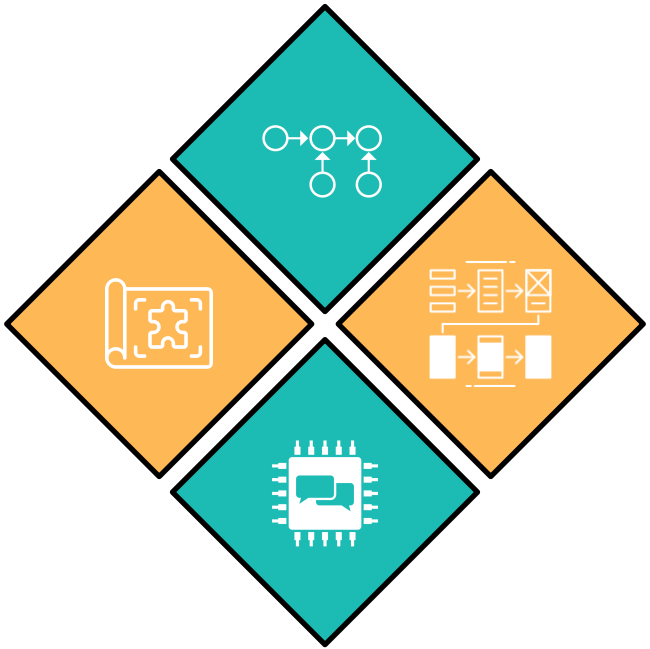
# Approaches applied

**Valency Grammar**

Arguments
Surface realizations
Semantic roles
Combinatory patterns

**Meaning to Text Theory**

Lexical functions
Semantic representation
Government patterns



**Prototypes Theory**

Lexical prototype
Semantic prototype

**Natural Language Processing**

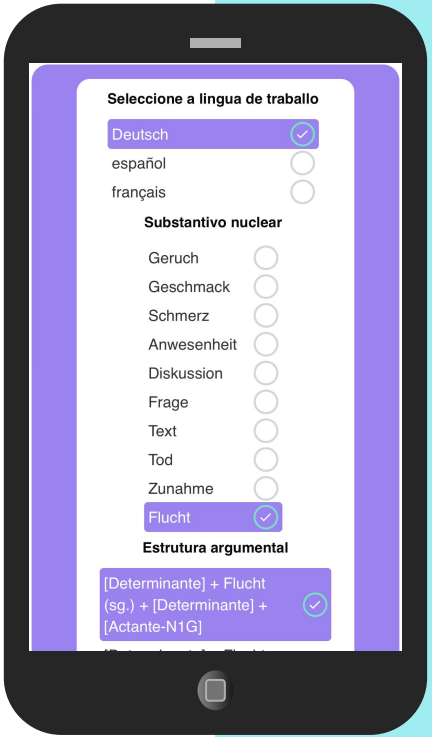
Natural Language Generation
Information extraction
Lexical mining

# Combined method

**Analysis**

**Extraction**

**Evaluation**




# Ressource interoperability

**External**

PORTLEX
WordNet
FreeLing

**Internal**

APIs
Lematiza
Combina

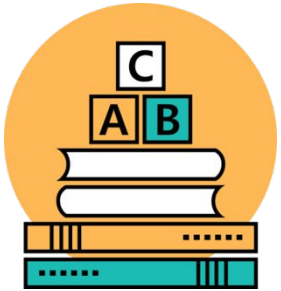
# What can it be useful for?

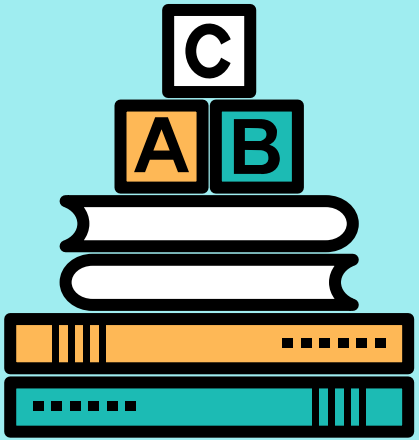
**1.** For building customized examples.

**3.** For building lexical extraction tools.

**2.** For creating reusable lexical packages.

**4.** For studying restrictions within the noun phrase.

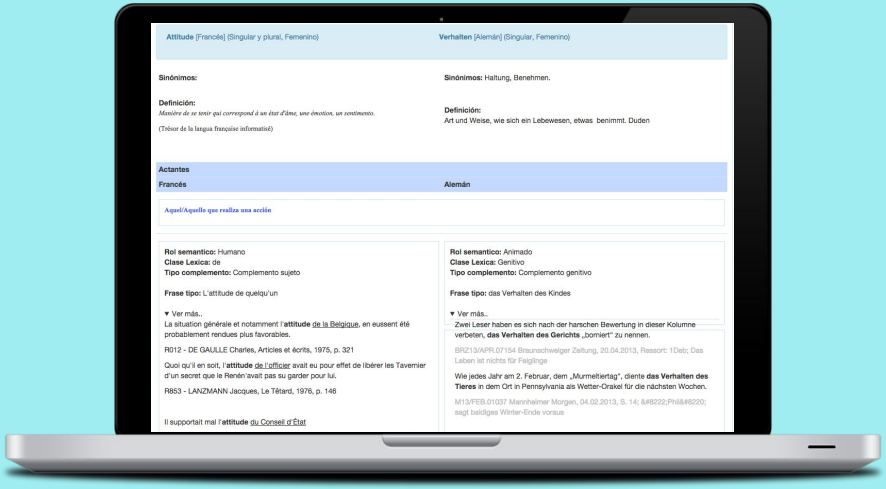





# 2. PORTLEX

Our main lexicographic source

# A valency dictionary of the noun phrase




**Valency**

**Collaborative**

**Cross-lingual**

**Modular**

CONSUMO 1
*Lexeme*

CONSOMMATION
*FR Equivalent*

COMBINATION 2+1
*Adjective*
*de + human*
La consommation énergétique des familles

COMBINATION 2+1
*Adjective*
*Adjective*
La consommation viticole mondiale

COMBINATIONS
*of actants*

ACTANTS

ACTANT 1
That who/what does the action

ACTANT 2
That who/what is affected

REALIZATION 1
*de + human*
La consommation des familles

REALIZATION 2
*Adjective*
La consommation parentale

REALIZATION 1
*de + object*
La consommation d'aliments

REALIZATION 2
*en + object*
La consommation en énergie

REALIZATION 3
*Adjective*
La consommation énergétique

REALIZATION 4
*Noun*
La consommation alcool

## Followers

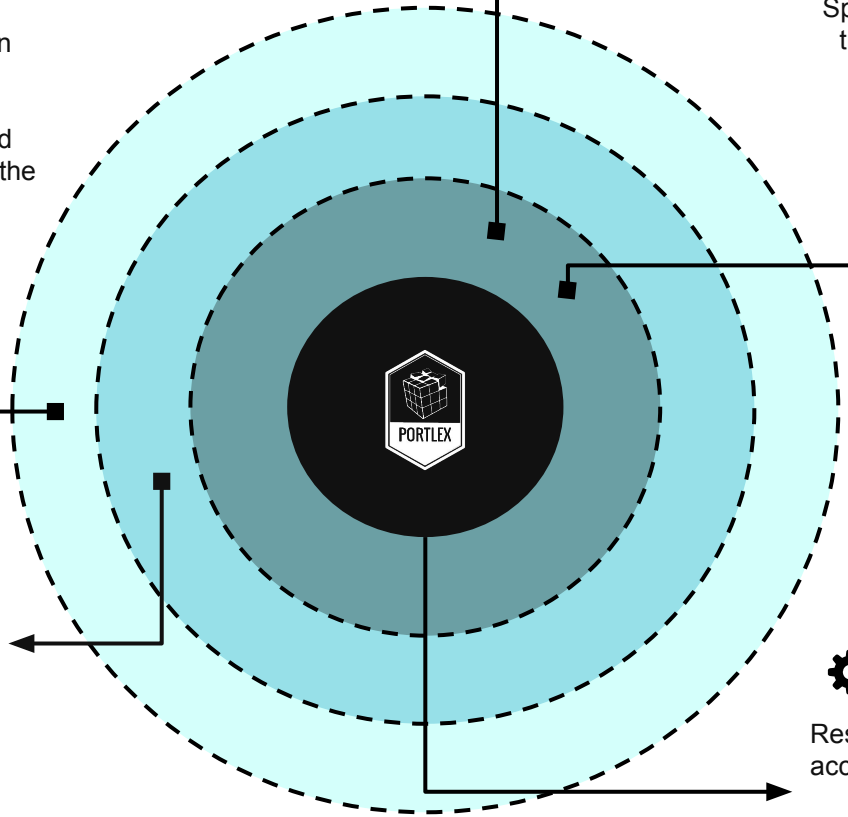
They follow PORTLEX updates on the web and/or on social media. They can interact with users and editors, but they are not registered as users since they don't consult the dictionary.

## Editors

Specialists in a particular language with training in valency grammar. They edit entries in the dictionary database..

## Reviewers

Specialists in a particular language with training in valency grammar. They validate the data entered by the editors.



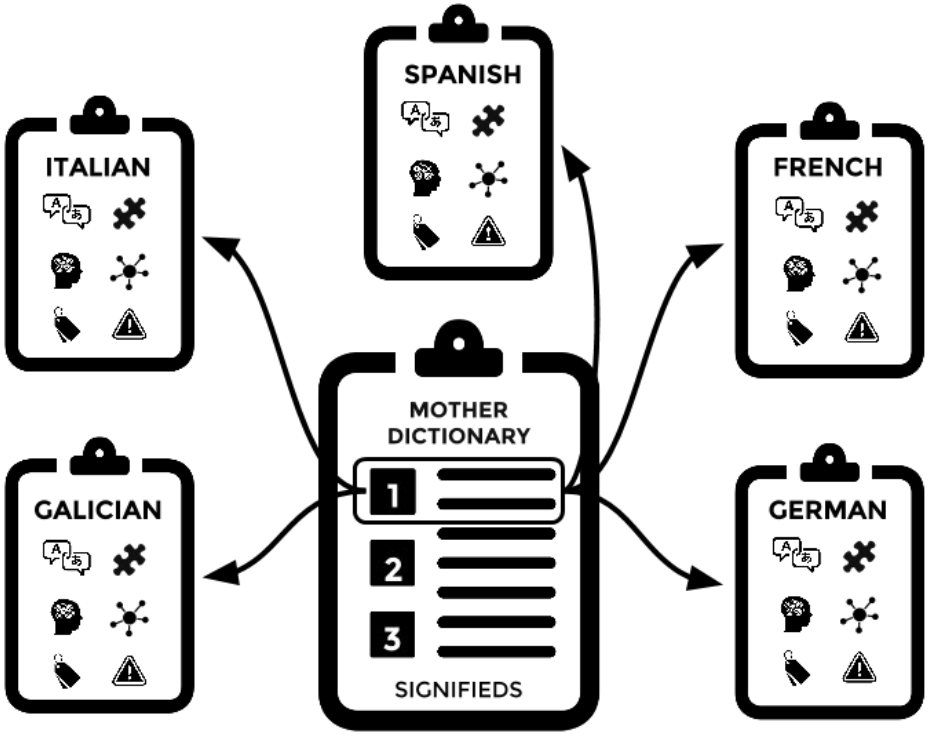

## Users

Experts or general users who usually do searches in the dictionary. They can interact with other users and with the editors. They can also become editors after a training process.

## Administrators

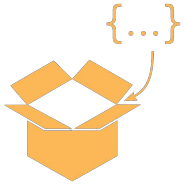
Researchers who manage the database, access and user roles.

# Modular & cross-lingual



# Difficulties encountered

when working with  corpora



**The compilation of surface realizations**

It's a very time demanding process.

**The description of argument patterns**

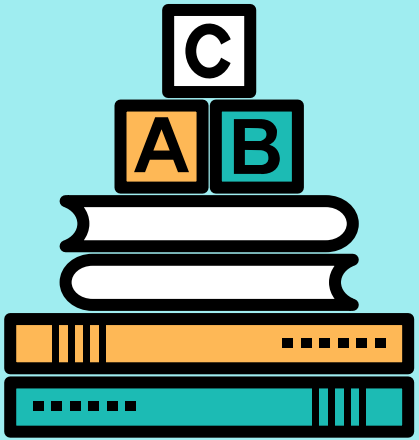
Dozens of patterns by noun to be described in detail.

**Corpus-extracted data adequacy**

Human review is also necessary.

**Semantic prototypes.**

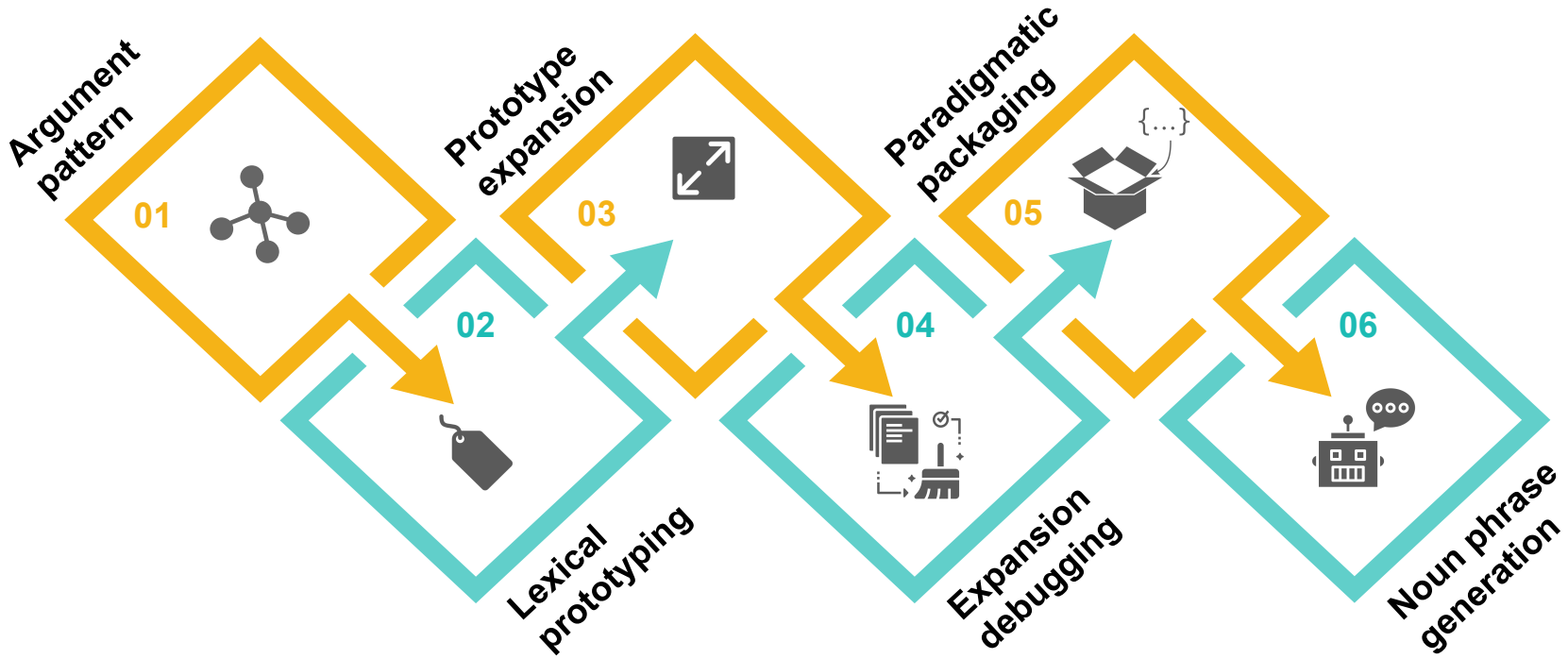
**Could they be one way out?**




# 3. Methodology & tools

APIs, Lematiza, Combina and Xera.

# Combined method phases



# Lexico-semantic prototyping

**Concordances**

Pattern description
CQL queries
Slot filling lemmata

**Frequencies**

Frequent lemmata
Lexical prototypes

**Cleansing**

Inadequacies
Repetitions

**Classification**

Semantic annotation
Semantic prototyping

# An exemple

**Det. + Flucht + aus + Noun**

*Die Flucht aus Sintra*

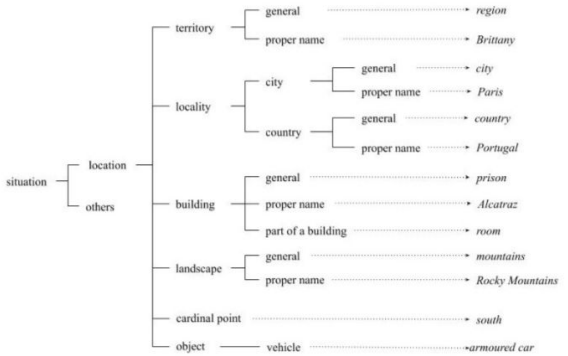
**CQL query**

DDR
Ghetto
Troja
Haus
Frankreich
Ost-Berlin
etc.

**Lexical prototypes**

FLUCHT
aus+dative +

| Lexical prototypes | 1st Level | 2nd Level | 3rd Level | 4th Level |
|---|---|---|---|---|
| Warschauer Ghetto | situation | location | territory | |
| Haus | situation | location | building | |
| Kriegsgefangenenlager | situation | location | building | |
| Wohnung | situation | location | building | |
| Troja | situation | location | locality | proper name |
| Ost-Berlin | situation | location | locality | proper name |
| Venedig | situation | location | locality | proper name |
| Frankreich | situation | location | territory | proper name |
| Deutschland | situation | location | territory | proper name |
| Italien | situation | location | territory | proper name |




**Semantic annotation**

**Semantic classes**

# Expansions: resorting to Wordnet

**SUMO**

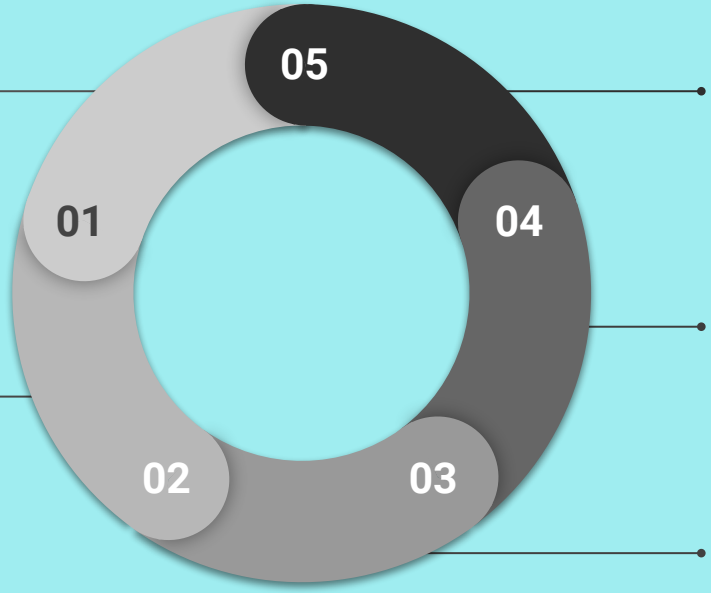
Suggested Upper Merged Ontology.
Nouns, verbs and adjectives.

**Top**

Top Concept Ontology.
Verbs and nouns.

**Hyponymes**

Linked on WordNet.




**Domains**

WordNet Domains.

**File**

Lexicographic files.
Semantic primes.

**Epinonyms**

By Guinovart & Solla.
Noun based

# Prototypes expansion

**Ontologies**

Semantic matching
LEMATIZA

**API queries**

Lexical extraction queries
APIs + COMBINA

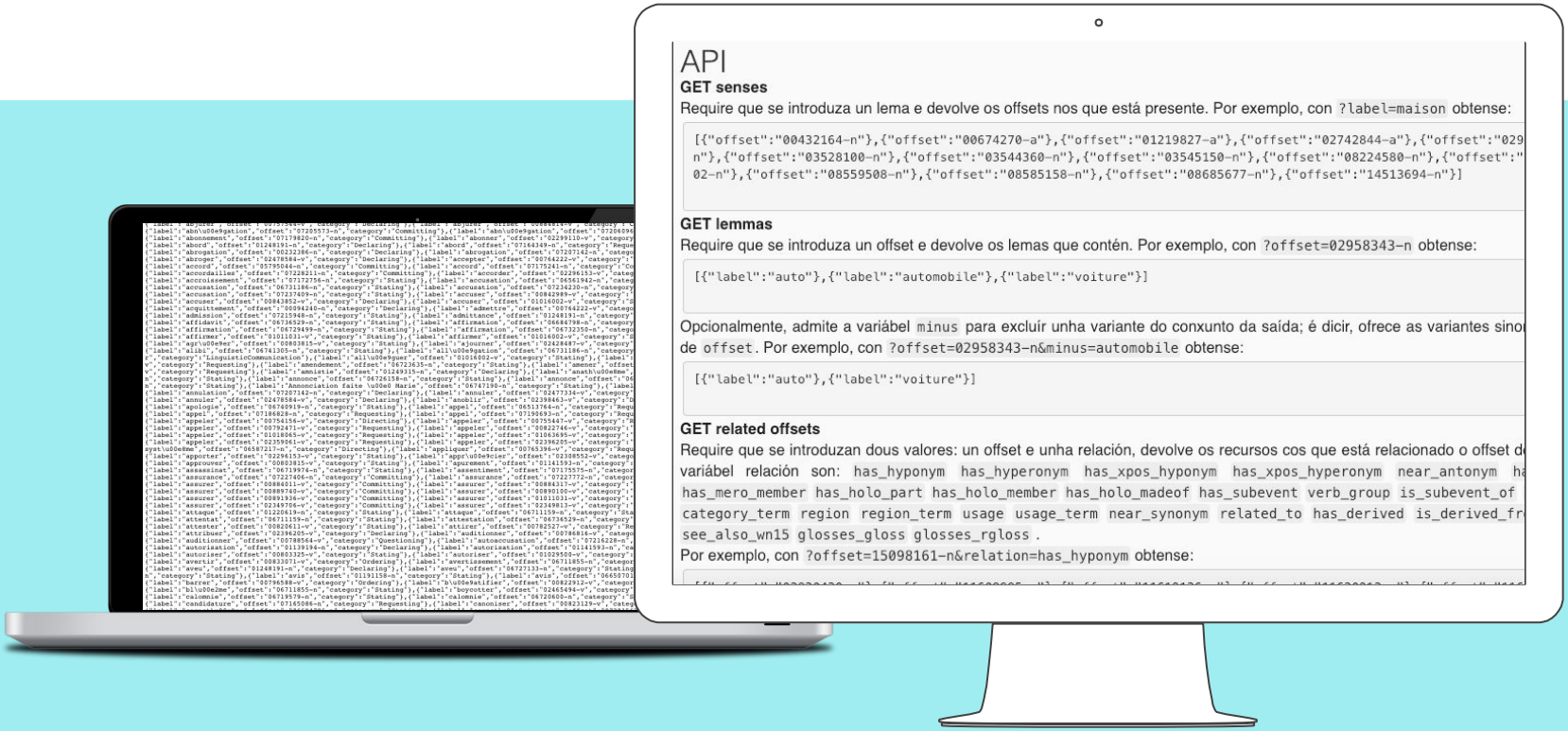
**Cleansing**

Inadequacies
Repetitions

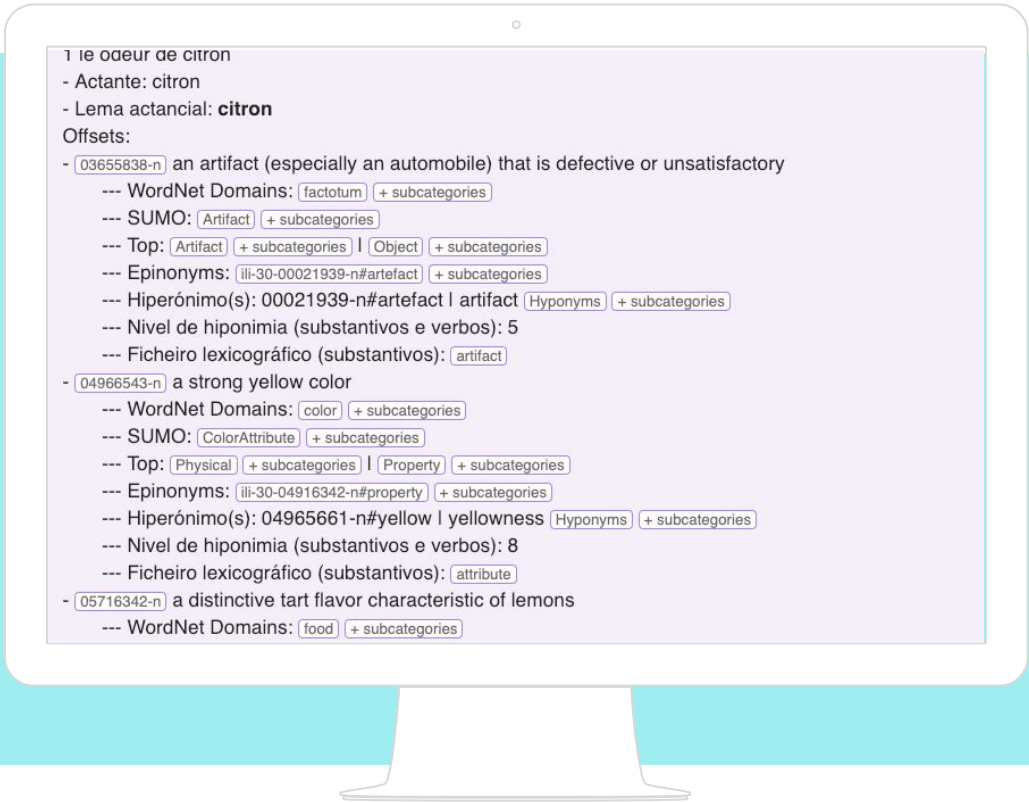
**Packaging**

Morphological annotation
Semantic tagging
FLEXIONA

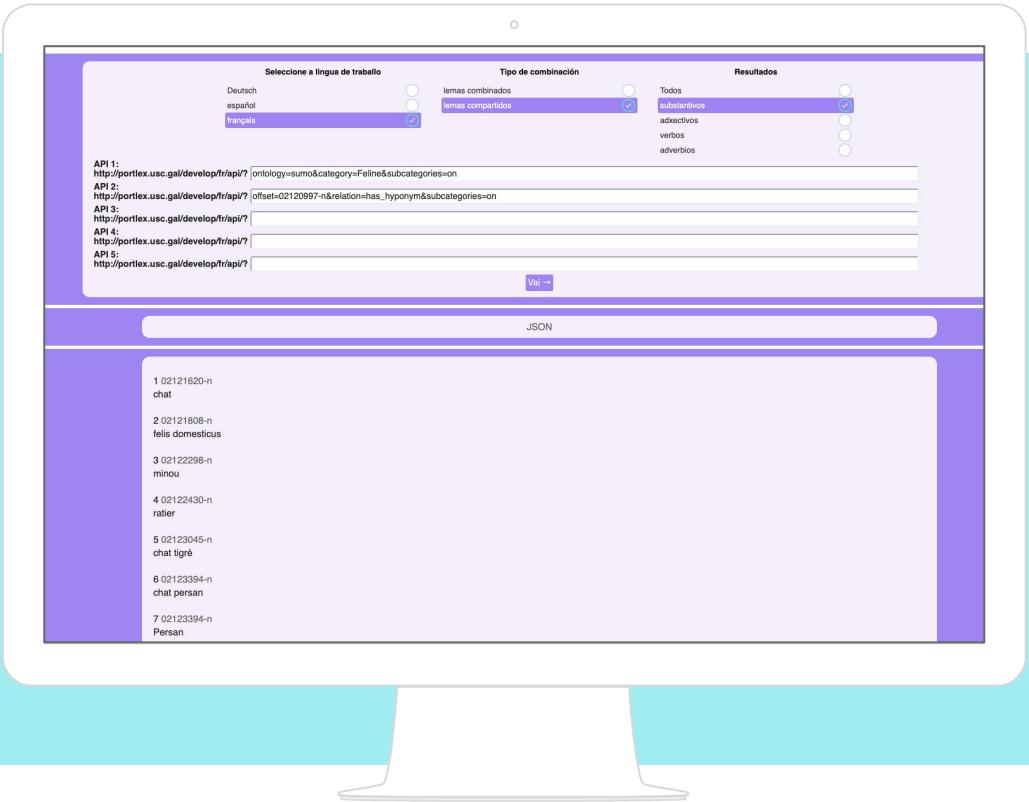
# APIs bit.ly/multigenera-tools



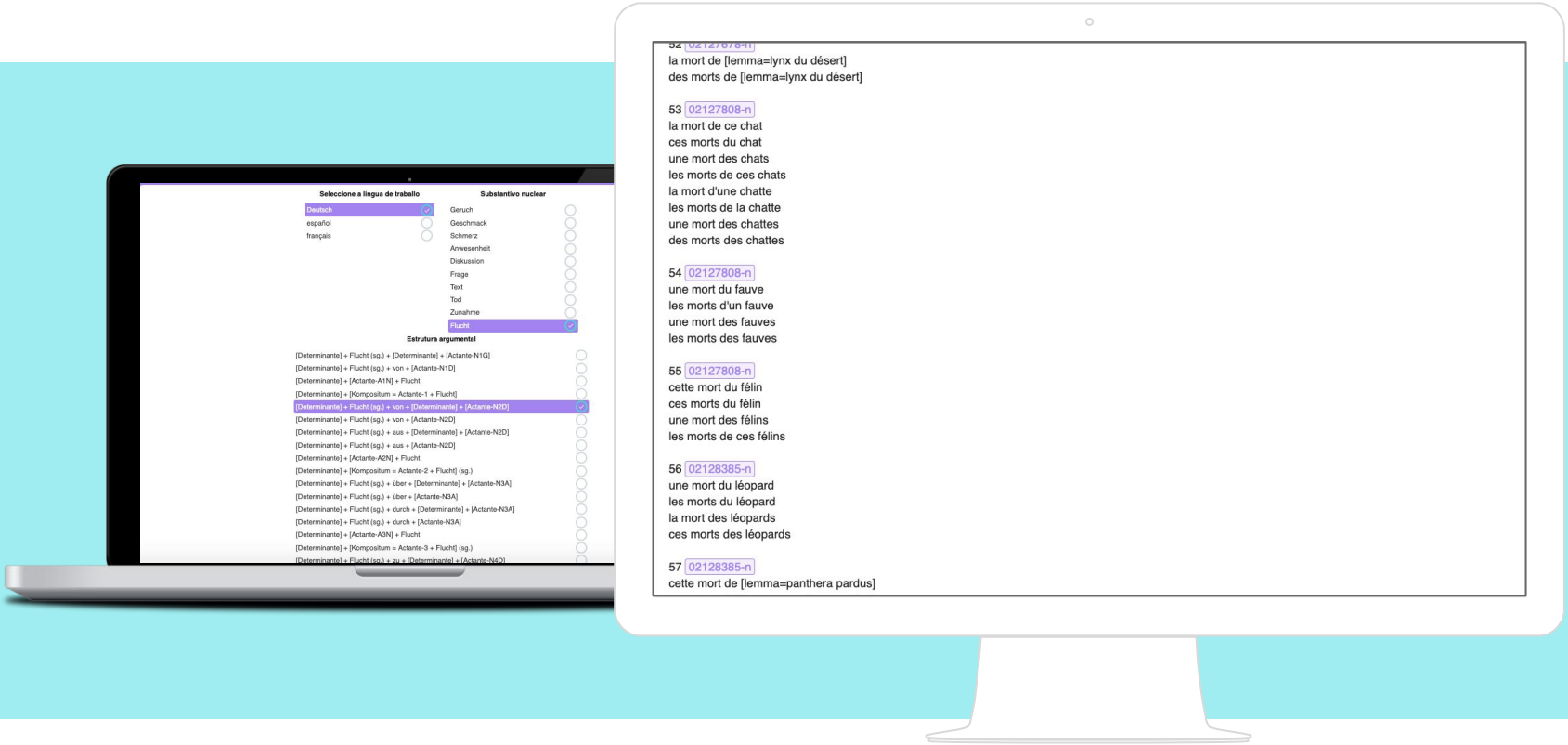
# LEMATIZA bit.ly/multigenera-tools

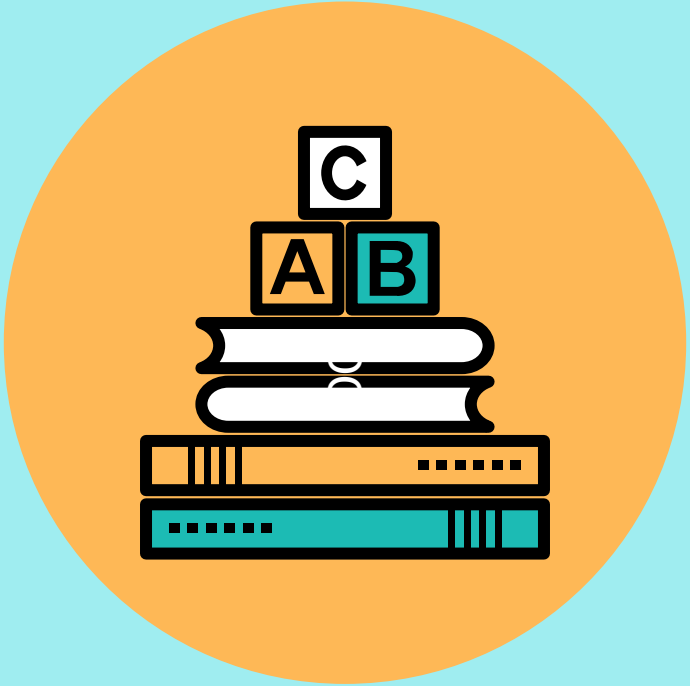


# COMBINA bit.ly/multigenera-tools



# XERA bit.ly/multigenera-tools






# Thank you! Any questions?

majo.dominguez@usc.es
miguelsolla@uvigo.es
carlos.valcarcel@uvigo.es