The REGLA database. Towards making it compatible with LiLa

Berta González Saavedra, Carlos Curiá and Esperanza Torrego

Universidad Autónoma de Madrid

June, the 4th

REGLA

Research projects

- ► FFI2009-13402-C04: 2009-2013
- ► FFI2013-47357-C4: 2013-2017

- ► COMREGLA. Compatibilizando REGLA con otros recursos digitales- AYUDAS FUNDACIÓN BBVA A EQUIPOS DE INVESTIGACIÓN CIENTÍFICA 2018

What is REGLA?

- REGLA stands for REcción y complementación en Griego Antiguo y Latín. It is a database for verbs and their Valency Frames.

- REGLA was built up as tool for a research project, started in 2009 and aimed to investigate Valency Frames of Greek and Latin verbs from a corpus of verbal forms in use as big as possible.

- It contains VFs of around 200 Latin and Greek verbs extracted from more than 25,000 texts evidences

- Theoretical approach: Functional Grammar (S. Dik 1997, *The Theory of Functional Grammar*, Berlín-New York, Mouton de Gruyter).

REGLA

Fields of analysis

- for Verbs
  - Morphology of the verbal form (lemma, person, number, aspect and tense, mood, voice)
  - Main or subordiate clause
  - Polarity
  - Complementation (Arguments and 1st level satellites)
- for Complementation
  - Presence (explicit, generic or context retrievable)
  - Integration level (Arg.1, Arg.2, Arg.3, Arg.4, Satellite)
  - Syntactic function (Subject, Direct Object...)
  - Formal codification (case, preposition, Sub. Clause ...)
  - Semantic function (Agent, Experiencer, Instrument...)
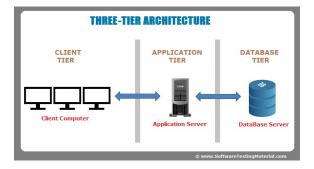 - Lexical characterisation (human being, event, concrete, place, ...)

REGLA

Results

- ► Valency Frames based on all verbal forms in use (empirical evidence)
- ► Repository of Valency Frames of the Greek and Latin verbs
- ■ *castra ponit* → [A1(Agent)(Hum)][A2(Afect)(Concr)]
- ■ *in conspectu Caesaris atque omnis exercitus res gerebatur* → [A2(Afect)(Even)][S(Ubic)(Lug)]
- ■ *corporis uero forma constabit exiguo capite, nigris oculis, naribus apertis*→ [A1(Afect)(Prop)][A2(Orig)(Concr)+A2(Orig)(Concr)]

First steps of the REGLA db, 2010-2013

- Latin Corpus:
  - Files extracted from LASLA for concrete verbs
- Greek corpus:
  - Pieces of texts copied "by hand" doing "copy and paste"
- Flex Application + MySQL database

REGLA

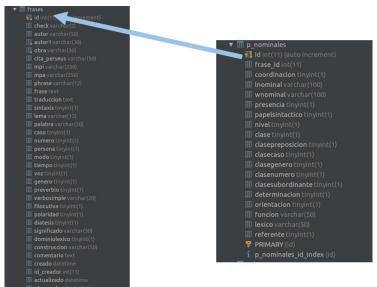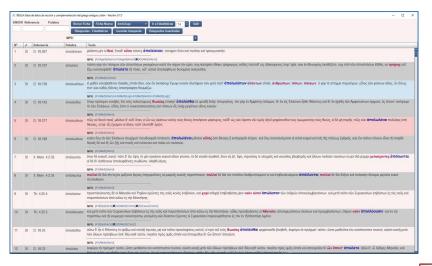# 2014, texts extracted from Perseus with web scraping for Greek and Latin

Database scheme

Corpus Text and Verbal Features → Argument Features

frases
- id int(11) (auto increment)
- check varchar(2)
- autor varchar(50)
- autor1 varchar(30)
- obra varchar(30)
- cita_perseus varchar(50)
- mpi varchar(256)
- mpa varchar(256)
- phrase varchar(12)
- frase text
- traduccion text
- sintaxis tinyint(1)
- lema varchar(15)
- palabra varchar(20)
- caso tinyint(1)
- numero tinyint(1)
- persona tinyint(1)
- modo tinyint(1)
- tiempo tinyint(1)
- voz tinyint(1)
- genero tinyint(1)
- preverbio tinyint(1)
- verbosimple varchar(20)
- filocutiva tinyint(1)
- polaridad tinyint(1)
- diatesis tinyint(1)
- significado varchar(50)
- dominiolexico tinyint(1)
- construccion varchar(50)
- comentario text
- creado datetime
- id_creador int(11)
- actualizado datetime
- id_actualizador int(11)
- PRIMARY (id)

p_nominales
- id int(11) (auto increment)
- frase_id int(11)
- coordinacion tinyint(1)
- lnominal varchar(100)
- wnominal varchar(100)
- presencia tinyint(1)
- papelsintactico tinyint(1)
- nivel tinyint(1)
- clase tinyint(1)
- clasepreposicion tinyint(1)
- clasecaso tinyint(1)
- clasegenero tinyint(1)
- clasenumero tinyint(1)
- clasesubordinante tinyint(1)
- determinacion tinyint(1)
- orientacion tinyint(1)
- funcion varchar(50)
- lexico varchar(50)
- referente tinyint(1)
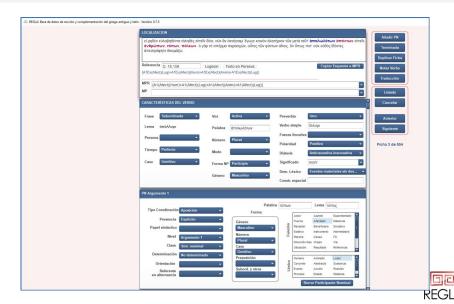- PRIMARY (id)
- i p_nominales_id_index (id)

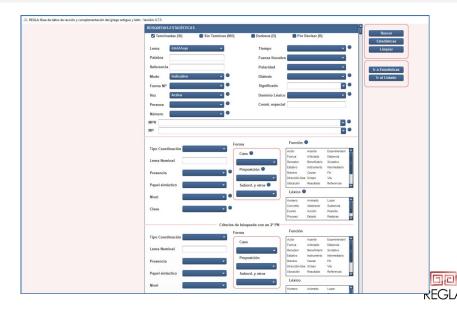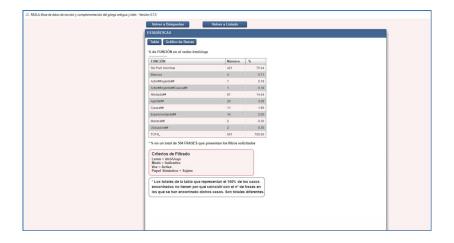REGLA

REGLA Flex Application - General View of the Data

REGLA Flex Application - Annotation Interface for Verbs

REGLA Flex Application - Annotation Interface for NPs

REGLA Flex Application - Statistics

Technical Implementation: Shortcomings (2014) and REGLA new Directions

- Database have pieces of unrelated texts
- Queries are limited to a single verb
- Queries are limited to a bunch of features
- Flex and Flash has an uncertain future

▶ Improve queries and statistics for the whole database

▶ Work with complete works, not with unconnected pieces of texts

▶ Annotated systems able to support semantic tags

▶ Leave Flex and adopt the browser as technology for the client

REGLA

The REGLA Browser