ISMLA Multilingual Session 6: Building a Spanish Verb Trainer in GWT

Johannes Dellert

Tübingen University

December 13, 2017

Plan

1. Spanish: Finite Verb Forms
2. Idea: Web-Based Verb Trainer
3. Exercise 05: A Spanish Verb Trainer in GWT

Spanish Finite Verb Forms: An Overview

Spanish verbs inflect for the following categories:

- subject agreement: `1SG`, `2SG`, `3SG`, `1PL`, `2PL`, `3PL`
- politeness distinction in `2SG` and `2PL`:
  - informal forms (`2SGfam` and `2PLfam`)
  - polite forms (`2SGpol` and `2PLpol`)
- tense: present (`PRS`), imperfect (`IPF`), preterite (`PRT`), future (`FUT`)
- mood: indicative (`IND`) and subjunctive (`SJV`) combine with tense, whereas imperative (`IMP`) and conditional (`CND`) do not (or only periphrastically)

Additional complications we will not be dealing with:

- non-finite forms (participles, infinitive, gerund)
- pronoun clitics in imperative, infinitive, and gerund
- two alternative forms of the imperfect subjunctive

Spanish: Paradigm Example *asar* "to roast"

		1SG	2SGfam	3SG 2SGpol	1PL	2PLfam	3PL 2PLpol
IND	PRS	as**o**	as**as**	as**a**	as**amos**	as**áis**	as**an**
	IPF	as**aba**	as**abas**	as**aba**	as**ábamos**	as**abais**	as**aban**
	PRT	as**é**	as**aste**	as**ó**	as**amos**	as**asteis**	as**aron**
	FUT	as**aré**	as**arás**	as**ará**	as**aremos**	as**aréis**	as**arán**
SBJ	PRS	as**e**	as**es**	as**e**	as**emos**	as**aséis**	as**en**
	IPF	as**ara**	as**aras**	as**ara**	as**áramos**	as**arais**	as**aran**
		as**ase**	as**ases**	as**ase**	as**ásemos**	as**aseis**	as**asen**
	FUT	as**are**	as**ares**	as**are**	as**áremos**	as**areis**	as**aren**
CND		as**aría**	as**arías**	as**aría**	as**aríamos**	as**aríais**	as**arían**
IMP			as**a**	as**e**	as**emos**	as**ad**	as**en**

Verb Trainer

- in order to teach such paradigms, textbooks tend to include a lot of rote exercises in the chapter where the respective form is introduced
- this leads to very easy exercises, because you only need to keep e.g. one row of the table in your head at a time
- better for long-term learning: mixed exercises tuned towards actual usage (natural focus on irregular forms, because they are frequent)
- IDEA: build a web tool which allows you to paste any Spanish text, masks the verb forms, and asks the user to reconstruct the missing forms based on the infinitive and a list of features

Exercise 05: Basic Architecture

In GWT, we need the following components to implement a verb trainer:

- a frontend (client-side code) implementing a webpage which
  - allows the user to paste some Spanish text (e.g. a Wiki snippet)
  - uses an RPC to the backend to turn the text into an exercise (partially masked text), and displays the exercise as a mixture of text boxes and text, with the hints displayed next to each gap
  - allows the user to fill the gaps, then click a button to check the answers
  - provides color-coded feedback (e.g. green for correct answers, red for incorrect ones) and the corrected version
- a backend (server-side code) which
  - maintains a database of Spanish verb forms and their features
  - implements a servlet which can process text and returns an object which represents a partially masked version of that text

Exercise 05: The Data

In `es-verb-forms.tsv`, you find a catalogue of Spanish verb forms:

- extracted from a Wiktionary dump using JWKTL, with a lot of post-processing to unify the format
- result: almost 452,000 Spanish verb forms with specifications
  - `ababillabais  ababillarse: 2PLfam IPF IND`
    ababillarse according to Wiktionary: to be sick with the stifle (of horses and other quadrupeds), *(veterinary medicine, Chile, Mexico)*.
- left column is what you will look up (and mask), right column provides the hints to display to the users next to the gaps
- missing: some very basic irregular verbs like *ser* and *hacer* (where the Wiktionary entries adhered even less to a common standard than for the more regular forms you will find)
- problem 1: many forms are ambiguous; we will just provide some description, as we lack the means to disambiguate forms in context
- problem 2: many common words like *como* and *una* will be analysed (and masked) as verb forms

Exercise 05: Necessary Steps

1. write server-side code for reading in and representing the contents of `es-verb-forms.tsv` as a simple lookup table
2. design and write a serializable data structure for representing a masked text with hints and solutions
3. implement the possibility to turn any Spanish text into an instance of that class by an RPC, masking all the verb forms
4. design a user interface with the required elements on the client side
5. implement event handling code on the client side which uses an RPC to create an exercise from text, and displays it
6. implement event handling code for checking the user answers against the exercise object, and provides color-coded feedback

Exercise 05: Implementation Hints

- data file should be in the WAR folder, and read in via `getServletContext().getResourceAsStream()`
- the text can be tokenized by `split(" ")`, although you will use some forms at the beginning and end of sentences in this way
- your serializable classes need a zero-argument constructor!
- best widgets for text: `HTMLPanel` filled with `HTML` and `TextBox`
- it is best to rely on CSS for layout in the frontend; CSS classes can be assigned to GWT widgets via `getElement().addClassName()`
- important CSS properties: `display` and `white-space`
- more about CSS: `https://www.w3schools.com/css/`, and text technology slides (attached to exercise)
- for checking the results, you will need to maintain a client-side data structure to map each `TextBox` into the respective correct answer
- a reference implementation is available at: `http://kos.sfs.uni-tuebingen.de:8080/SpanishTrainer/`

Exercise 05: Questions

Questions?