

# A spell checker for language learners of French and a learner corpus

Anne Vandeventer Faltin, Sébastien L'haire, Mar Ndiaye (University of Geneva, Switzerland)

Eurocall 2005, Cracow

## Motivations:

- Differences between native speakers' and learners' spelling errors
  - types of errors that must be treated
  - the kind of feedback

Specific error types targeted:

- phonetic spelling
- diacritics
- morphology
- agglutination
- missing punctuation

## Corpus gathering:

- Targets
  - Classes
  - Independent users
- Users must log in
- Teachers have access to all sentences from their classes

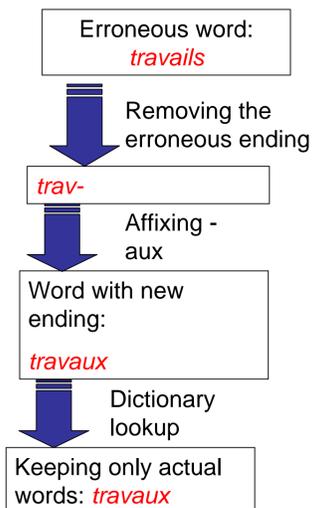
### Ad hoc method

Errors of morphology: incorrect plural of words in -ail/-al

The incorrect ending -ails/-als is removed from the stem.

The correct ending in -aux is attached to the stem.

The new word is checked in the dictionary to make sure it exists. Unknown words are discarded.

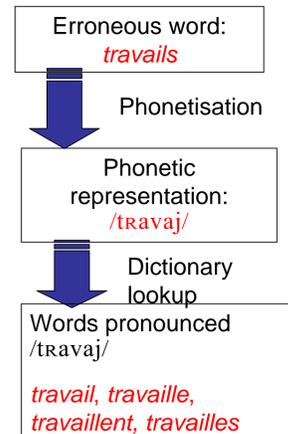


### Phonetic reinterpretation method

Errors involving phonetic spelling

A phonetic representation of the erroneous word is obtained by a phonetizer

The dictionary is accessed through a phonetic access and all words sharing this phonetic representation are retrieved



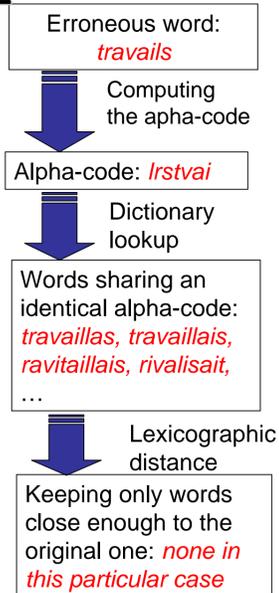
### Alpha-code method

Errors involving typos, diacritics, some omissions, etc.

An alpha-code is a synthetic representation of a word. Diacritics are removed, each letter of the word appears only once. Letters are ordered alphabetically into two groups: consonants first, followed by vowels.

The dictionary is accessed through the alpha-codes (which were computed once and for all for the dictionary) and all words sharing the same alpha-code are retrieved.

A selection is made to keep only the words close enough to the original word



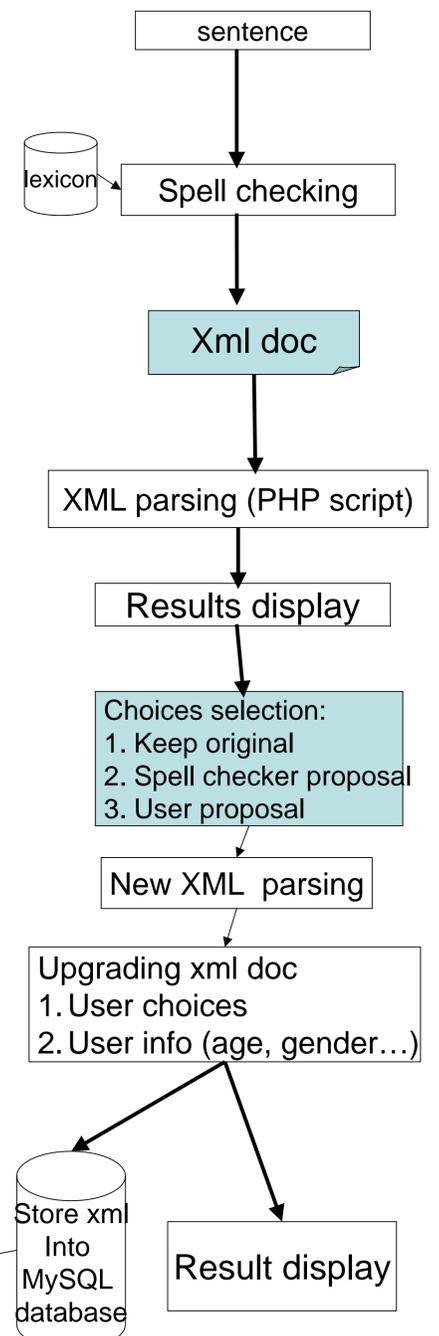
### Ordering the proposals

Number and type of method(s) used to retrieve a single proposal

Matching lexical category, gender, number, and person with the expected one (from sentence analysis)

- travaux
- travail
- travaillent
- travailles
- travaille

```
LATLCORR xml:lang="fr" mothertongue="fr" age="33"
gender="m" time="19/08/2005 15:50">
-<SENTENCE sentenceId="1">
-<ITEM index="i00001" pos="1">
<ORIGINAL itemTag="i00001">les</ORIGINAL>
-<PROPS itemTag="i00001">
<PROPOSAL index="p00001" itemTag="i00001"
method="maj" selected="yes">Les</PROPOSAL>...
```



### FipsOrtho: Correction orthographique

Votre nom d'utilisateur: seb

Votre texte:

les travaux sont terminés.

Corrections:

Passez la souris sur les mots en bleu pour mettre en évidence leur emp

Mot inconnu:	Propositions du correcteur OU	vosre suggestion
les	les (laisser tel quel)	<input type="text"/>
travaux	travaux (laisser tel quel)	<input type="text"/>
	<input type="button" value="Effacer les champs"/>	<input type="button" value="Enregistrer les corrections"/>

### FipsOrtho: Correction orthographique

Votre nom d'utilisateur: seb

Votre texte:

les travaux sont terminés.

Corrections:

Passez la souris sur les mots en bleu pour mettre en évidence leur emplacement dans la phrase.

Mot inconnu:	Propositions du correcteur OU	vosre suggestion
les	Les	<input type="text"/>
travaux	travaux (laisser tel quel)	<input type="text"/>
	travaux (laisser tel quel)	<input type="text"/>
	travaux	<input type="text"/>
	travail	<input type="text"/>
	travaillent	<input type="text"/>
	travailles	<input type="text"/>
	travaille	<input type="text"/>
	<input type="button" value="Enregistrer les corrections"/>	

### FipsOrtho: Correction orthographique: validation

Votre nom d'utilisateur: seb

Votre original: les travaux sont terminés.

Votre correction: Les travaux sont terminés.

Entrez une nouvelle phrase en cliquant sur "Effacer" ou soumettez la phrase à nouveau au correcteur

Les travaux sont terminés.

### Future developments:

- Improvements of spell checker from first results of data gathering
- User access to mon/bilingual dictionary(-ies) to improve choice selection
- User access to morphological analysis of words

**Call for participation:** The spell checker will be available soon at: <http://latlcui.unige.ch/spellchecker/> . You are welcome to use it, with or without your class.