Towards the effective integration of NLP in CALL

Piet Desmet, Dirk De Hertog

0. Overview

1. Introduction

2. Technological Readiness of NLP

    1. Native Language
    2. Learner Language
    3. Measure linguistic complexity and proficiency

3. Applications: 7 use cases

4. Future Possibilities

1. Introduction

1.1 Current Situation

Most of the CALL environments do not use NLP

-> use of NLP in classrooms is hardly mainstream practice

> "The development of systems using NLP technology is not on the agenda of most CALL experts, and interdisciplinary research projects integrating computational linguists and foreign language teachers remain very rare" (Amaral & Meurers 2011: 6).

cf. also report by the Dutch Language Union:
Onderzoek taal- en spraaktechnologie en onderwijs (Van den Heuvel, T'Sas & Verberne 2012)

1.2 Concerns

Why not? Skepticism -> technological and pedagogical concerns

(1) Native language :

Can NLP account for the full **complexity** of natural human languages?

> "I am pessimistic about the possibility of ICALL" (Nyns 1989: 46)

But is this still the case?

(2) Learner language

a) (Erroneous) application of NLP developed for native language
-> But: can NLP be developed to work for learner language?

b) Current binary scoring of errors and automatic correction
-> How do we handle the notion of error and correction?

(3) Tutor: “One size fits all” due to technological restrictions

But: can NLP help us create language and learner models

1.2 Reasonable Targets

Yet there is a high reward if NLP were successfully integrated

(1) Automatization of a number of tedious tasks such as resource creation

Exploit and enrich Native language

(2) Automatic Error Correction and feedback generation of open items

Analyse learner language

(3) Personalised teaching

Linguistic Complexity and Intelligent tutors

2. Technological readiness

Technological readiness (written language) depends on the field of application

(1) Native language

NLP is (partially) usable, requires integration

(2) Learner language

NLP requires development
should rely on insights of L2-acquisition and FLTL

(3) Model complexity (-> Intelligent tutors)

Requires an improvement in (2)
Work on integration of (1) and (2)

2.1 Native Language

NLP has evolved and improved dramatically the last 20 years
Some technologies are ready to use

Better results due to successful paradigmatic shift

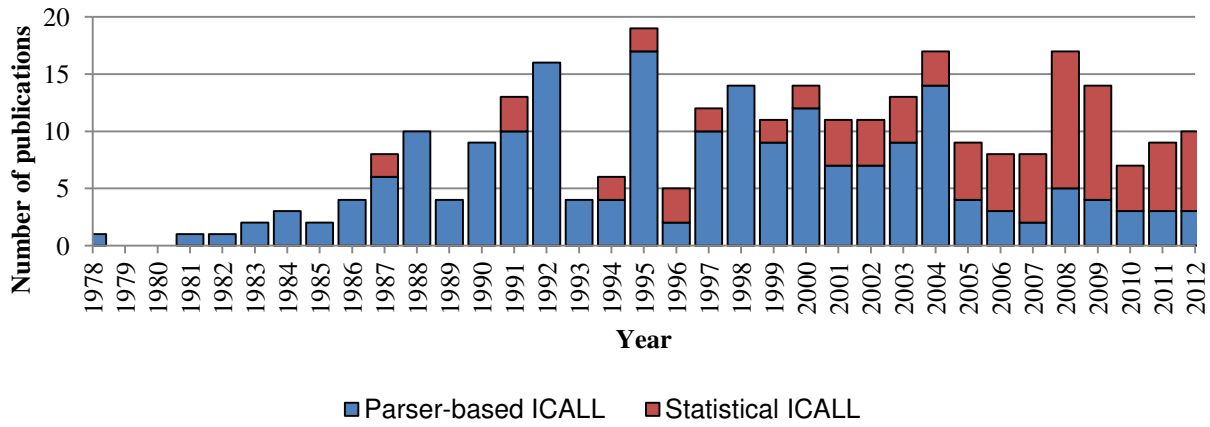
From Rule-based NLP

Manual engineering of grammar rules

To Statistical NLP

Learn from examples
use supervised machine learning techniques
to solve certain problems

Looking at the number of publications shown in this figure, parser-based CALL has dominated research up until 2005, but there seems to have been a shift in research interest and statistical error detection systems have since then taken the upper hand.

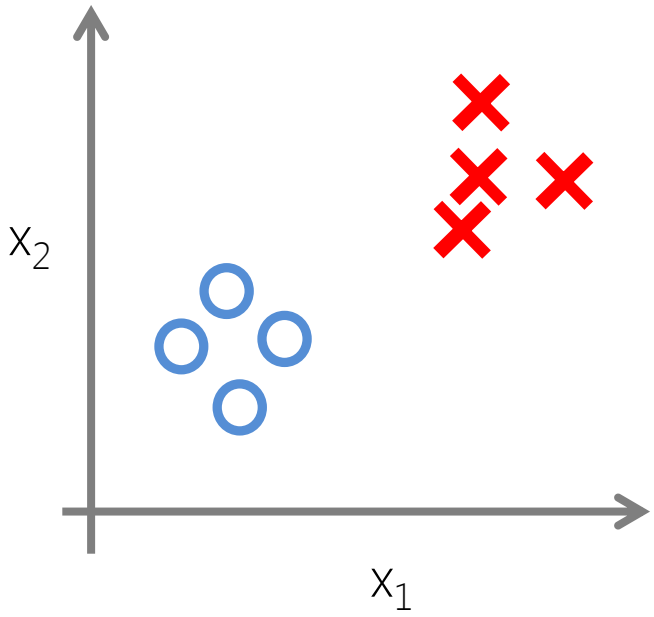


© PhD Ruben Lagatie

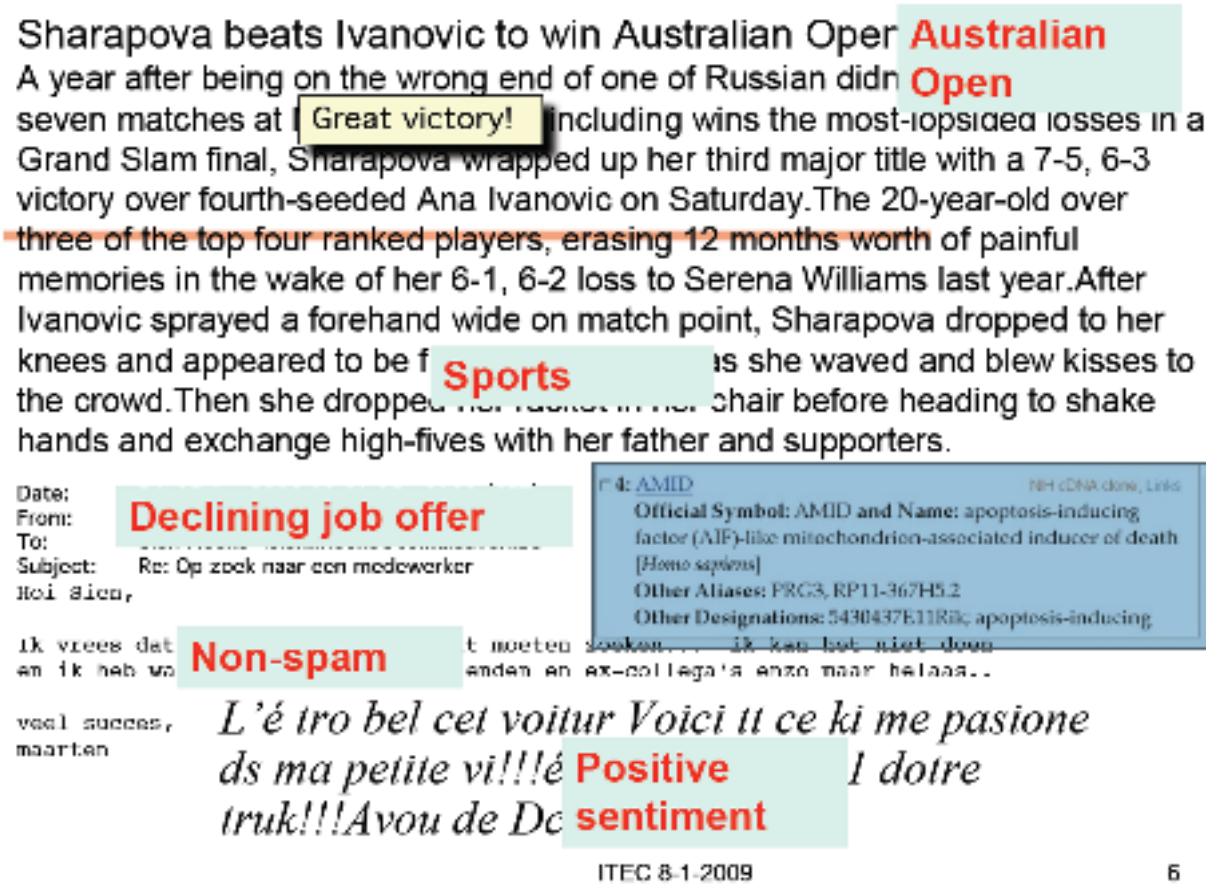
2.1.1 Statistical NLP

Supervised Learning

Learning from examples how to label them
Predict on the basis of this expert **model** how to label new examples



Examples of supervised learning



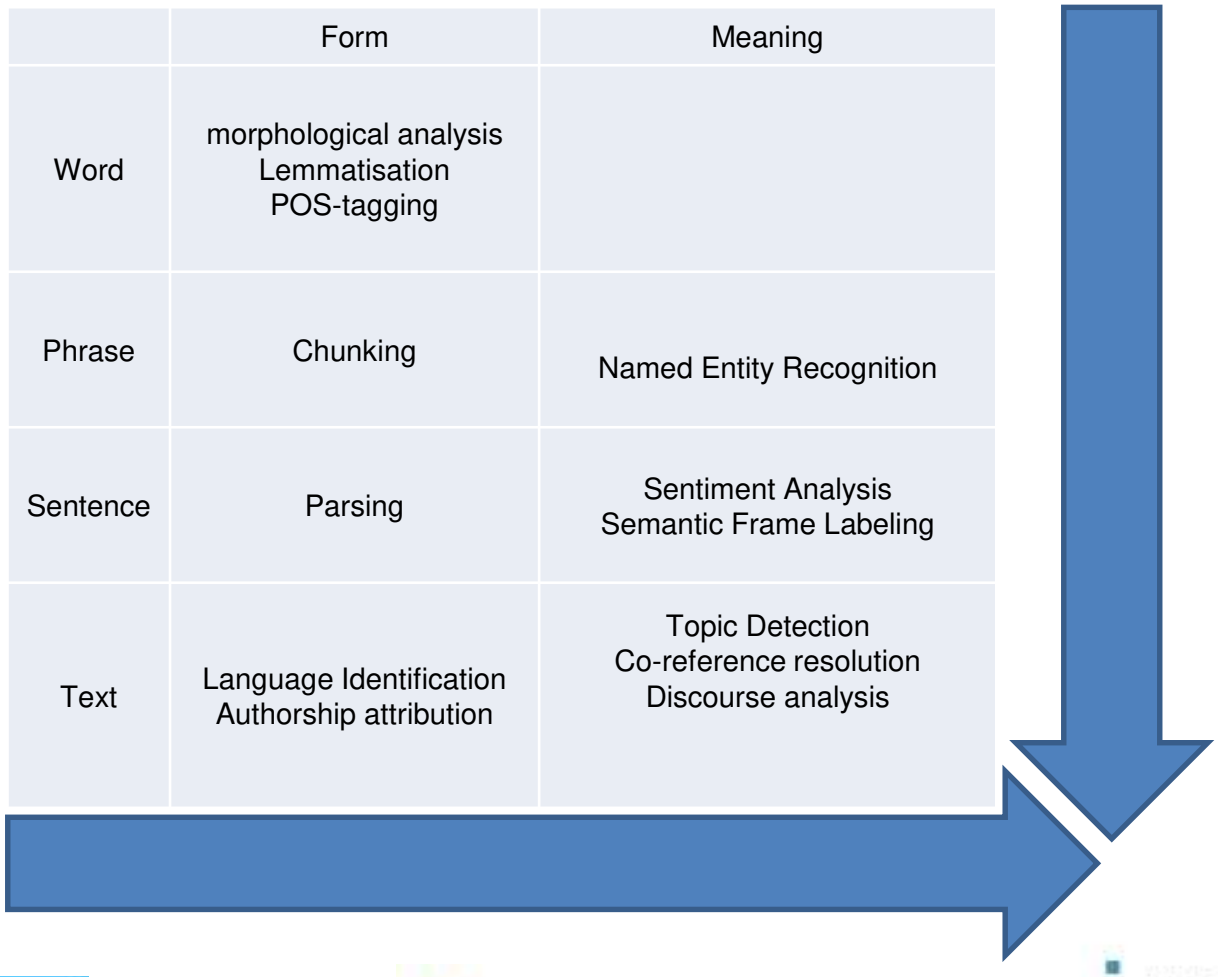
© Sien Moens (2009)

2.1.2 Native Language

	Form	Meaning
Word	morphological analysis Lemmatisation POS-tagging	
Phrase	Chunking	Named Entity Recognition
Sentence	Parsing	Sentiment Analysis Semantic Frame Labeling
Text	Language Identification Authorship attribution	Topic Detection Co-reference resolution Discourse analysis

Growing Complexity

	Form	Meaning
Word	morphological analysis Lemmatisation POS-tagging	
Phrase	Chunking	Named Entity Recognition
Sentence	Parsing	Sentiment Analysis Semantic Frame Labeling
Text	Language Identification Authorship attribution	Topic Detection Co-reference resolution Discourse analysis



Ongoing develpments and improvements

- Combination of statistical and rule-based approaches

- Technically: the use of semantic representations
  is becoming mainstay and used in more complex NLP-tasks
  => a broader generalisation possible

- Discursive component (semantic text) is gaining importance

- *Unsupervised methods stand at the sideline in applications but are ready to make their entry (cf. semantic vector spaces – infra)*

2.2. Learner Language

Using predictive models trained on native language
will only work on advanced learners' output

Solution:
NLP: For beginning learners rely on a hypothesis
of what it is they meant to communicate
= Disambiguation of what the learner is trying to say

Goal is to detect errors: the difference between
hypothesis and produced utterance
= What is wrong?

Ambiguity of target hypothesis

e.g. “He hear you tell the history.”
-> He heard you speaking about the past.

- Many different hypotheses could be generated based on the provided output
  - **I** hear you tell your history
  - He heard you **told** a **story**
  - He hears you **telling** a story
  - …

Where is the mistake?
Multiple and simultaneous mistakes take place at all linguistic levels

Constraining open-ended output

Fully open-ended learner output hard to manage
-> constrain learner output (reduce possible answers)
by reducing the search space of the processing tools

Important role for design and use of CALL tools
The tools help to disambiguate the hypothesis

Task design should be structured towards specific type of output

Constraining learner language

Constraints...	on form	on meaning
Explicit	• multiple choice • gap-filling • list of words to reuse/reorder	• set content words • translation • reading comprehension questions

E-Tutor (Heift, 2010)

Tagarela (Amaral *et al*, 2011)
CoSeC (Hahn & Meurers, 2012)





Tagarela

Constraints...	on form	on meaning	
Explicit	• multiple choice • gap-filling • list of words to reuse/reorder	• set content words • translation • reading comprehension questions	E-Tutor (Heift, 2010) Tagarela (Amaral *et al*, 2011) CoSeC (Hahn & Meurers, 2012)
Implicit → *still too much variation to handle*		• task & context • dialogic interaction • open-ended conversations	SPELL (Morton *et al*, 2012) (prototype) Reactive chatbots (e.g. *Tutor Mike*)





Dialogue-based CALL: *Dialangue* project (PhD Serge Bibauw)

Task-oriented dialogue:
Job interviews

Constraints:
Answer questions according to given profile and resume



Progression:
Task complexity
Amount of content
Linguistic complexity



System **initiative**

Immediate corrective **feedback** and writing support (**scaffolding**)

Task-oriented dialogue:
Non-linguistic outcome.
Implicit time pressure.
Performance evaluation.

2.3. Model Linguistic Complexity and Learner Proficiency

Tutor combines a **language model** to measure linguistic complexity

&

a learner model to measure proficiency

Determine **complexity** at which linguistic level?

Lexical level: frequency in general corpora
Syntactic level: sentence length, parse structure

Measure complexity in what type of language?

1) in native language
2) in learner language : requires working NLP!

3. Applications

Focus on written language input & output

Native Language : Enrich and Exploit

1. *Resource generator*
2. *Reading companion*
3. *Exercise and test generator*

Learner Language : Finding Errors

4. *Error detector, feedback generator and automatic scoring tool*
5. *Writing aid*

Model Learner Behaviour : measure Complexity

6. *Input provider*
7. *Adaptive item sequencer*

-> Conceptual outline + applications from academic R&D

3.1 Resource Generator

What? Providing learners with easy access to authentic language use

How? Exploiting authentic (annotated) corpora

=> (Intermediate or) advanced learners become linguistic investigators
Requires well developed linguistic awareness

Used methods :

- Concordancing on (bilingual) corpora
- Search on terms and filter on specific categories (cfr. annotation)

3.1 Resource Generator

Projects:

- Concordancing on bilingual corpora
  & more advanced search engines

  REBECA-project (ITEC)
  = Ressources électroniques bilingues extraites de corpus alignés

  DPC-project (ITEC) http://www.kuleuven-kulak.be/DPC/
  = Dutch Parallel Corpus

- Corpus-enriched learner dictionaries & grammars

  BLF-project (Serge Verlinde) http://ilt.kuleuven.be/blf/
  = Base Lexicale du Français

REBECA-project

5. Gezondheid en leefmilieu in België

Index

Nauwelijks een eeuw geleden leden duizenden mensen in ons land nog aan ziekten veroorzaakt door de slechte kwaliteit van het leef- en werkmilieu.

Die tijd is intussen voorbij. Tal van ziekten zijn onder controle. Maar vandaag heeft de overheid totaal andere gezondheidsproblemen, die zijn veroorzaakt door vervuiling door de industrie, het verkeer en door de menselijke activiteit in het algemeen.

De opkomst van nieuwe chemische producten, nieuwe productieprocessen en technologieën en de vermenging van allerlei pollutiebronnen hebben hun weerslag op het klimaat, de kwaliteit van de lucht en de bodem, de biodiversiteit en de voedselketen. Vaak is het effect ervan pas na enkele jaren of zelfs decennia later zichtbaar.

Bovendien is de verstedelijking sterk toegenomen. In 2000 leefde ongeveer 80% van de bevolking in stedelijke gebieden. Dat heeft gevolgen. In heel wat steden duiken hoe langer hoe meer stressverschijnselen op die te maken hebben met het leefmilieu: ozonpieken, zware luchtvervuiling, toenemend lawaai, stijgende afvalproductie, moeilijkere toegankelijkheid tot recreatiegebieden, files, enzovoort.

En dan is er nog de maatschappelijke ongelijkheid. Die heef
factoren. Die faktoren hebben een rechtstreekse invloed op

De strijd tegen ziekte en vervuiling kan dus maar succesvo
welzijn van de hele bevolking.

bron: *http://www.belgium.be - 19.09.2003*

http://cime-se.sry.fundp.ac.be/d...NL=overheid&terme_FR=autorit%E9

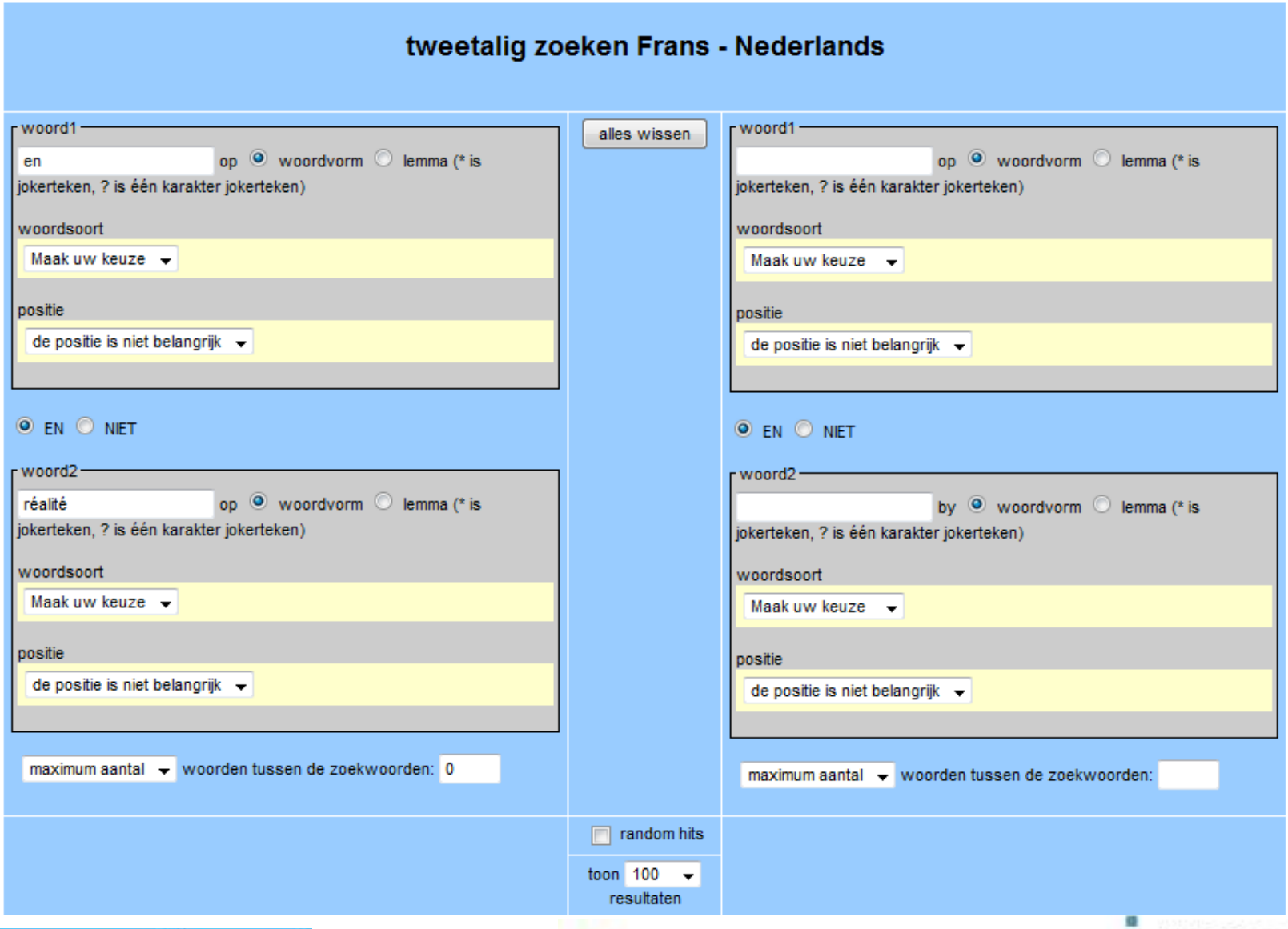
De Gemeenschappen zijn bijvoorbeeld autonoom op het vlak van onderwijs, maar de minimale vereisten voor de aflevering van diploma's blijven een aangelegenheid van de federale **overheid**, net zoals de leerplicht.	Quant aux Communautés, elles ont, par exemple, dans leurs attributions l'enseignement, mais les conditions minimales de délivrance des diplômes demeurent du ressort de l'**autorité** fédérale, tout comme l'obligation scolaire.

Extrait de : *http://www.fgov.be*

Fermer

overheid (nom, de, overheden)	
... nder toezicht van alle hogere **overheden**, in het kader van de fe ...	... nales en étant subordonnées à toutes les **autorités** supérieures.
... angelegenheid van de federale **overheid**, net zoals de leerplicht ...	... mes demeurent du ressort de l'**autorité** fédérale, tout comme l'o ...
De **overheid** heeft een nieuw reglement uitgevaardigd.	Les **autorités** ont promulgé un nouveau règlement.
... erd in welke administratie en/of **overheid** daarbij betrokken is.	... r quelle administration et/ou **pouvoir** public est impliqué dans ...
... erd in welke administratie en/of **overheid** daarbij betrokken is.	... s intéressés de savoir quelle **administration** et/ou pouvoir publ ...
... hillende administraties en/of **overheden** die hierbij betrokken z ...	... r quelle administration et/ou **pouvoir** public est impliqué dans ...
... hillende administraties en/of **overheden** die hierbij betrokken z ...	... s intéressés de savoir quelle **administration** et/ou pouvoir publ ...
Administraties en **overheden** zullen elkaars gegevens zoveel mog ...	Les **administrations** et les autorités doivent partager et utili ...

DPC-project


tweetalig zoeken Frans - Nederlands
woord1
en
op
woordvorm
lemma (* is jokerteken, ? is één karakter jokerteken)
woordsoort
Maak uw keuze
positie
de positie is niet belangrijk
EN
NIET
woord2
réalité
op
woordvorm
lemma (* is jokerteken, ? is één karakter jokerteken)
woordsoort
Maak uw keuze
positie
de positie is niet belangrijk
maximum aantal
woorden tussen de zoekwoorden:
0
alles wissen
random hits
toon
100
resultaten
woord1
op
woordvorm
lemma (* is jokerteken, ? is één karakter jokerteken)
woordsoort
Maak uw keuze
positie
de positie is niet belangrijk
EN
NIET
woord2
by
woordvorm
lemma (* is jokerteken, ? is één karakter jokerteken)
woordsoort
Maak uw keuze
positie
de positie is niet belangrijk
maximum aantal
woorden tussen de zoekwoorden:

Example: “en réalité” French - Dutch

Aantal resultaten: 114 waarvan er 114 getoond worden.

Exporteer deze resultaten naar Excel.

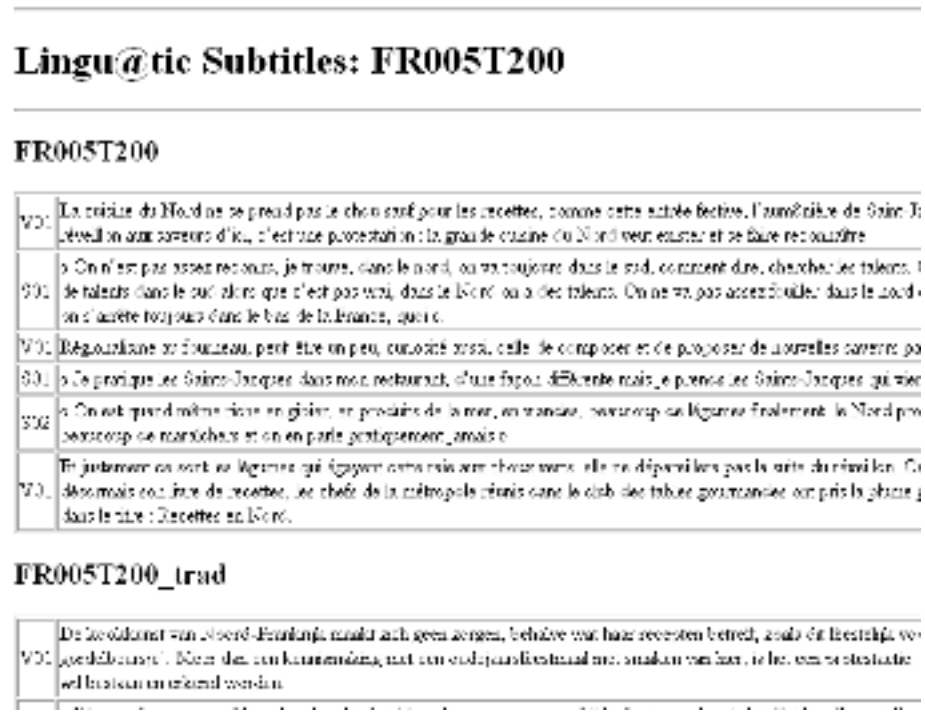
En réalité, il existe deux types de bokashi.	Er zij eigenlijk wee soorten bokashi.
Il s'agit **en réalité** d'une entourloupe : le magazine contient des articles emphatiques mais creux, dans lesquels Boorman change tout simplement le nom de l'entreprise et du produit dans chaque numéro.	In werkelijkheid s het een zeepbel: het tijdschrift bevat hoogdravende, maar artikels waarin Boorman voor elk nummer gewoon de namen van bedrijf en product vervangt.
Donc, pour nous, l'essentiel était qu'il parle, tant il racontait des choses énormes, du plus amusant au point de vue politique au préjudiciable, par exemple quand, évoquant "les poitrines de poulet et les dos bossus", il parlait **en réalité** des anciens combattants...	Voor ons kwam het er dus op aan dat hij veel aan het woord kwam, want hij vertelde veel enormiteiten, soms amusant maar ook vaak schadelijk voor de politiek. ch is zijn uitspraak over de kiekenborsten en de bochels als hij h in feite d over oud-strijders.
En réalité, la défense de la ville est constituée dans un premier temps de remparts en terre. Les murs en pierre seront dressés dans ces remparts, sur des arcades.	In werkelijkheid werden eerst aarden wallen opgetrokken en werd de stenen muur daarin gefundeerd, op bogen.
En réalité, au cours de la nuit, T'Serclaes escalade sournoisement le mur du jardin de son propre palais.	In werkelijkheid is T'Serclaes in het geniep en bij nacht en ontij over de tuinmuur van zijn eigen paleis geklommen.
« Cette inscription est gratuite : on donne ainsi l'impression que le bar n'est pas accessible à tout le monde, alors qu'**en réalité** ce n'est pas le cas.»	"Die inschrijving is gratis: je geeft zo het gevoel dat dit niet voor iedereen bereikbaar is, terwijl d wel is."
En réalité, le bruissement perceptible est dû à la combinaison de diverses fréquences renforcées par la forme du coquillage.	In werkelijkheid horen we in zo'n schelp niets anders dan een ruistoon, die een mengeling is van allerlei frequenties die door de vorm van de schelp versterkt worden.

Extension: video search

www.franel.eu

Exploitation of XML-based captions (transcriptions and translations) to render video selection more versatile and attractive:
lemmas are time tags indexes which form reference points to video selections





Lingu@tic Subtitles: FR005T200

FR005T200

V01	La cuisine du Nord ne se prend pas le chou sauf pour les recettes, comme cette entrée festive, l'aumônière de Saint-J… réveillon aux saveurs d'ici, c'est une protestation : la grande cuisine du Nord veut exister et se faire reconnaître
S01	» On n'est pas assez reconnu, je trouve, dans le nord, on va toujours dans le sud, comment dire, chercher les talents. … de talents dans le sud alors que c'est pas vrai, dans le Nord on a des talents. On ne va pas assez fouiller dans le nord … on s'arrête toujours dans le bas de la France, quoi ».
V01	Régionalisme au fourneau, peut-être un peu, curiosité aussi, celle de composer et de proposer de nouvelles saveurs pa…
S01	» Je pratique les Saints-Jacques dans mon restaurant, d'une façon différente mais je prends les Saints-Jacques qui vien…
S02	» On est quand même riche en gibier, en produits de la mer, en viandes, beaucoup de légumes finalement, le Nord pro… beaucoup de maraîchers et on en parle pratiquement jamais ».
V01	Et justement ce sont les légumes qui égayent cette raie aux choux verts, elle ne dépareillera pas la suite du réveillon. C… désormais conviviale de recettes, les chefs de la métropole réunis dans le club des tables gourmandes ont pris la plume … dans le titre : Recettes en Nord.

FR005T200_trad

V01	De kooldienst van Noord-Frankrijk maakt zich geen zorgen, behalve wat haar recepten betreft, zoals dit feestelijk ve… gedeelteerd. Meer dan een kennismaking, met een oudejaarsfeestmaal met smaken van hier, is het een protestactie … wil bestaan en erkend worden.

```
à 00:00:09.6800
a 00:00:24.8900|00:00:28.8400|00:00:59.9000|00:01:12.7000
alors 00:00:24.8900
assez 00:00:18.0300|00:00:31.9000
au 00:00:38.3000
aussi 00:00:38.3000
aux 00:00:09.6800|00:01:05.8100
bas 00:00:35.6500
beaucoup 00:00:24.8900|00:00:55.0000|00:00:59.9000
bonnes 00:00:31.9000
calais 00:00:51.1200
ce 00:01:05.8100
celle 00:00:42.0800
c'est 00:00:09.6800|00:00:24.8900
cette 00:00:06.6400|00:01:05.8100|00:01:12.7000
chefs 00:01:16.0000
chercher 00:00:18.0300|00:00:31.9000
chou 00:00:04.0000
choux 00:01:05.8100
club 00:01:16.0000
comme 00:00:06.6400
comment 00:00:18.0300
composer 00:00:42.0800
cuisine 00:00:04.0000|00:00:14.1800|00:01:12.7000
curiosité 00:00:38.3000
dans 00:00:18.0300|00:00:18.0300|00:00:24.8900|00:00:28.8400|
00:00:31.9000|00:00:35.6500|00:00:47.2800|00:01:16.0000|00:01:19.9500
```

Viewing video extracts based on lexical search in transcription and/or translation (challenge: search on semantically related words)

3.2 Reading Companion

What?

Helping learners understand foreign-language input
By providing a different and enhanced perspective

How?

Enrich language input through annotation layers

- add information
- provide alternative ways of exploration
- serve as a bridge to connect more information

Formal linguistic information (Lemma) > Connect to dictionary information
Semantic information (NER) > Connect to encyclopedic information

Projects:

GLOSSER-project (John Nerbonne)

http://www.let.rug.nl/glosser/Glosser

lemmatization of the inflected forms
dictionary entry (cf. Van Dale)
examples of the word from corpora

iRead+ project (ITEC)

Lemmatization & POS-tagging
Named entity recognition (persons, organisations, locations)

GLOSSER-project





HOME

ABOUT

API

CONTACT US



Standard de Liège

Une odeur de viande grillée flotte dans l'air. On croi
sosie de Johnny Hallyday. Des grappes de jeunes
taquinent gentiment, écharpe rouge et blanche aut
cou, vareuse floquée Van Damme. Dix mètres plus
trois quadras urinent de concert, le gobelet rempli de bière
entre les dents. Un couple s'enlace, à l'ombre du stade
rougeoyant de Sclessin. De l'autre côté de la Meuse, le
haut-fourneau d'Ougrée se dresse dans un ciel entre chien

anesthésiera (anesthésier)
Verbe

Rendre insensible à la douleur par l'anesthésie.

Example sentence: *"On employait autrefois, pour anesthésier les malades, un appareil relativement compliqué."*

Read more on Wiktionary

Gilles De Bilde, Rodrigo Beenkens, Pierre-Yves Jeholet, les Wathelet père et fils... Tous sont passés par la loge du groupe Circus. Pour faire monter la température, Philippe Fayray a sa spécialité : les « orgasmes », un cocktail de Baileys, Jet 27 et glace frappée.

Retour sur le terrain, pour un autre Standard-Genk, le 24 février. Un dimanche glacial. De la neige fondante tombe sur Liège, drue. Les supporters attendent le coup d'envoi engoncés dans leur manteau, bonnet sur la tête. Au même
apéritif, blottis
ntreprise
un médecin du
en téléphonant à
natch depuis une
p froid ? Toi et
trouves pas ?

Le froid anesthésiera aussi la partie. Après nonante minutes de jeu au ralenti, les joueurs se sépareront sur un score bloqué à zéro de chaque côté. Qu'importe, côté VIP, la bonne humeur reste de mise. On croise un homme


iRead+
the intelligent reading companion
HOME
ABOUT
API
CONTACT
Johnny Hallyday (Confidence: 1)
Jean-Philippe Smet, dit Johnny Hallyday, est un chanteur, compositeur et acteur français né le 15 juin 1943 à Paris. Avec plus de cinquante ans de carrière, il est l'un des plus célèbres chanteurs francophones et l'une des personnalités les plus présentes dans le paysage médiatique français, où plus de couvertures de magazines lui ont été consacrées.
http://www.johnnyhallyday.com/
Read more on Wikipedia
n croise un sosie de Johnny Hallyday. Des grappes de jeunes se taquinent gentiment, écharpe rouge et blanche autour du cou, vareuse floquée Van Damme. Dix mètres plus loin, trois quadras urinent de concert, le gobelet rempli de bière entre les dents. Un couple s'enlace, à l'ombre du stade rougeoyant de Sclessin. De l'autre côté de la Meuse, le haut-fourneau d'Ougrée se dresse dans un ciel entre chien
Gilles De Bilde, Rodrigo Beenkens, Pierre-Yves Jeholet, les Wathelet père et fils... Tous sont passés par la loge du groupe Circus. Pour faire monter la température, Philippe Fayray a sa spécialité : les « orgasmes », un cocktail de Baileys, Jet 27 et glace frappée.
Retour sur le terrain, pour un autre Standard-Genk, le 24 février. Un dimanche glacial. De la neige fondante tombe sur Liège, drue. Les supporters attendent le coup d'envoi engoncés dans leur manteau, bonnet sur la tête. Au même moment, une brochette d'invités prennent l'apéritif, blottis bien au chaud dans la loge d'une grande entreprise liégeoise. Coupe de champagne en main, un médecin du CHU scrute le stade à travers la vitre, tout en téléphonant à l'un de ses amis, qui s'apprête à suivre le match depuis une tribune 2 ouverte à tous les vents. « Pas trop froid ? Toi et moi, là, cela fait un peu lutte des classes, tu trouves pas ? Le froid anesthésiera aussi la partie. Après nonante minutes de jeu au ralenti, les joueurs se sépareront sur un score bloqué à zéro de chaque côté. Qu'importe, côté VIP, la bonne humeur reste de mise. On croise un homme
Persons (64):
Mario
Louis Smal
Josly Piette
Frisaye
Mercedes CLS
Bob Claes
Marc Wilmots
Justine Henin
Bernard Tapie
Mpoku
Roland Duchâtelet
Jean-Michel Javaux
M. Frisaye
Benoît Poelvoorde
Jean-Michel Zecca
Wathelet
Meusinwest
Van Damme
Philippe Fayray
Johnny Hallyday

Waarom wachtte Anderlecht zo lang? Net zoals in Genk wachtte ook Anderlecht lang om van trainer te wisselen. In wezen valt het toe te juichen dat clubs kiezen voor sportieve continuïteit. Maar soms voel je dat doorgaan met een trainer geen enkele zin heeft. Het was een opmerkelijke passage in de geschiedenis van Anderlecht, de intrede van John van den Brom. Een Nederlander met lef en flair, open en eerlijk, communicatief en toegankelijk, een adept van het offensief voetbal. De mensen stonden op de banken bij het bij momenten sprankelende spel dat Anderlecht toen bracht, de term revolutionair lag in de mond, zoiets hadden ze in het Astridpark al lang niet meer gezien. Zou John van den Brom er maandag aan hebben gedacht, op het moment dat hij gebrek aan duidelijkheid dat uiteindelijk ook het huwelijk tussen Mario Been en RC Genk deed stranden? Net zoals in Genk wachtte ook Anderlecht lang om van trainer te wisselen. In wezen valt het toe te juichen dat clubs kiezen voor sportieve continuïteit. Maar soms voel je dat doorgaan met een trainer geen enkele zin heeft. Dat was bij Genk zo en het was nu bij Anderlecht niet anders. De vraag blijft waarom de bestuurlijke top van de kampioen zolang wachtte, terwijl er wedstrijden waren dat bestuurslui voortijdig naar huis gingen. Omdat ze hoopten op een ommekeer? Omdat ze op het einde van het seizoen in schoonheid afscheid wilden nemen? Maar welke basis was er nog om op die kentering te rekenen? De wanvertoning van afgelopen zaterdag op OH Leuven was nog maar zelden vertoond, het

Selectie woordsoort (85)

Woordsoort

Werkwoord

Features

WVORM

PVTIJD

PVAGR

POSITIE

BUIGING

3.3 Exercise and test generator

What? Assist the teacher with the creation and selection of exercises
Based on authentic material

How? Create (mostly) closed-form exercises on:

- morpho-syntactic correctness
- lexical restrictions
- semantic appropriateness

based on the analysis of L1 text materials
and/or on the analysis of learner errors

For? From beginning to advanced learner

Projects :

Morpho-syntactic

iRead+ project (ITEC)

VIEW-project (Detmar Meurers) http://sifnos.sfs.uni-tuebingen.de/VIEW
= Visual Input Enhancement of the Web

Lexical restrictions:

Alfalex-project (Serge Verlinde) http://ilt.kuleuven.be/publicaties/tools.php

iRead+ Exercise generator: Three step model

- Explore
  1. Application highlights target items
  2. Learner recognizes target items and clicks on target items

- Practice
  - Fill gaps or multiple choice activities
  - Feedback

- Play
  - Target items, drill & practice
  - Time constraint





Visual Input Enhancement of the Web (VIEW)

- Home
- About VIEW
- Getting Started
- Topics and Activities
  - Articles
  - Determiners
  - Gerunds vs. Infinitives
  - Noun Countability
  - Phrasal Verbs
  - Prepositions
  - See vs. Enter
  - Wh-Questions
- FireFox Extension
- Changelog
- Feedback

Articles

The **articles** topic just highlights the use of the words *a, an, the*. If you'd like to practice with more types of articles and determiners, try the **determiners** topic.

- Color

  In the text, VIEW shows you the articles *a, an, the* in **orange**.

- Click

  Click on each article you find in the text and VIEW will show you whether you guessed **right** or **wrong**.

- Multiple Choice

  Choose an article from the drop-down list and VIEW will show you whether you guessed **right** or **wrong**. You can get help by clicking on [?].

- Practice

  Fill in the blanks with the **correct** article. If you guessed **wrong** or you find it too difficult you can get help by clicking on [?].

Try it out:

This form lets you get an idea of what VIEW can do, but it may have trouble accessing or displaying some pages. For the best results, use the VIEW firefox extension.

URL:

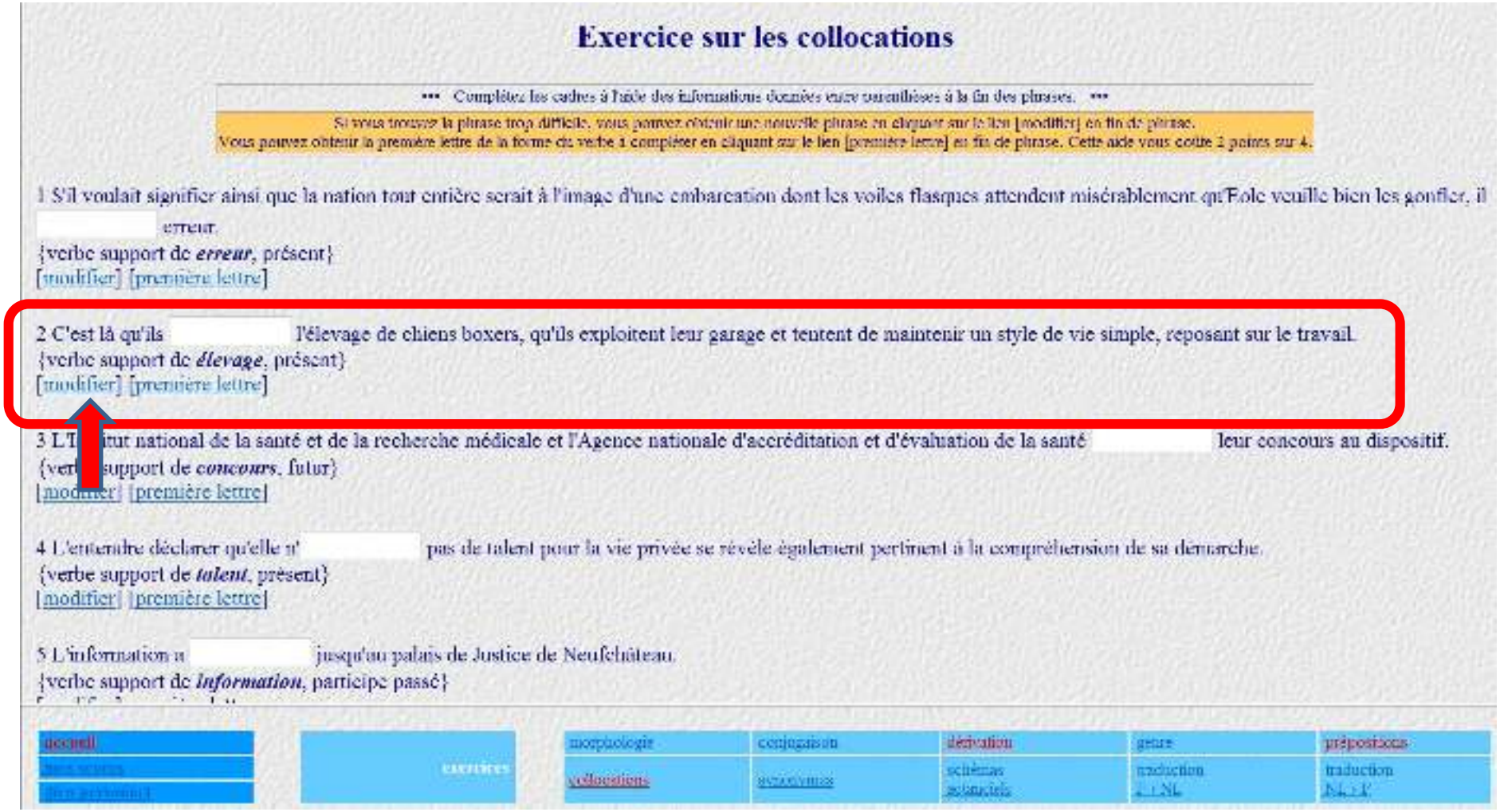
Activity type: ◉Colorize ○Click ○Multiple Choice ○Practice

Go! (opens in a new window)

Some example sites with colorizing:

- Wikipedia (computer-assisted language learning)
- The Guardian (car free cities)
- TechCrunch (Wool vs. AP)
- The Onion (on This American Life)
- Public Literature (Emma by Jane Austen)

Alfalex (KU Leuven – Serge Verlinde)



Exercice sur les collocations

*** Complétez les cadres à l'aide des informations données entre parenthèses à la fin des phrases. ***

Si vous trouvez la phrase trop difficile, vous pouvez obtenir une nouvelle phrase en cliquant sur le lien [modifier] en fin de phrase.
Vous pouvez obtenir la première lettre de la forme du verbe à compléter en cliquant sur le lien [première lettre] en fin de phrase. Cette aide vous coûte 2 points sur 4.

1 S'il voulait signifier ainsi que la nation tout entière serait à l'image d'une embarcation dont les voiles flasques attendent misérablement qu'Eole veuille bien les gonfler, il [redacted] erreur.
{verbe support de ***erreur***, présent}
[modifier] [première lettre]

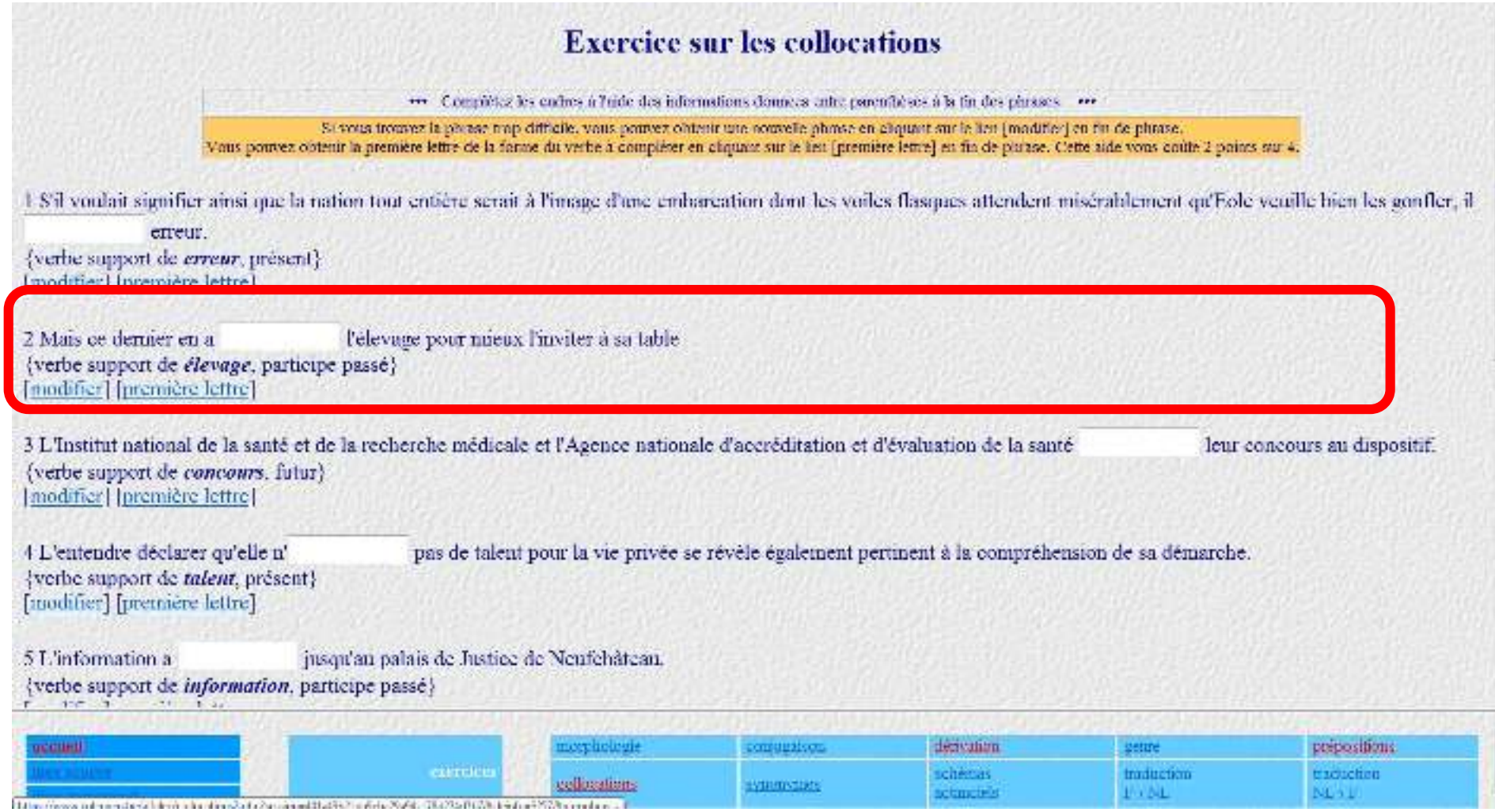
2 C'est là qu'ils [redacted] l'élevage de chiens boxers, qu'ils exploitent leur garage et tentent de maintenir un style de vie simple, reposant sur le travail.
{verbe support de ***élevage***, présent}
[modifier] [première lettre]

3 L'Institut national de la santé et de la recherche médicale et l'Agence nationale d'accréditation et d'évaluation de la santé [redacted] leur concours au dispositif.
{verbe support de ***concours***, futur}
[modifier] [première lettre]

4 L'entendre déclarer qu'elle n' [redacted] pas de talent pour la vie privée se révèle également pertinent à la compréhension de sa démarche.
{verbe support de ***talent***, présent}
[modifier] [première lettre]

5 L'information a [redacted] jusqu'au palais de Justice de Neufchâteau.
{verbe support de ***information***, participe passé}

accueil	exercices	morphologie	conjugaison	dérivation	genre	prépositions
[illegible]		collocations	synonymes	schémas actanciels	traduction L + NL	traduction NL + L
[illegible]						


Exercice sur les collocations
*** Complétez les cadres à l'aide des informations données entre parenthèses à la fin des phrases. ***
Si vous trouvez la phrase trop difficile, vous pouvez obtenir une nouvelle phrase en cliquant sur le lien [modifier] en fin de phrase.
Vous pouvez obtenir la première lettre de la forme du verbe à compléter en cliquant sur le lien [première lettre] en fin de phrase. Cette aide vous coûte 2 points sur 4.
1 S'il voulait signifier ainsi que la nation tout entière serait à l'image d'une embarcation dont les voiles flasques attendent misérablement qu'Eole veuille bien les gonfler, il erreur.
{verbe support de erreur, présent}
[modifier] [première lettre]
2 Mais ce dernier en a l'élevage pour mieux l'inviter à sa table
{verbe support de élevage, participe passé}
[modifier] [première lettre]
3 L'Institut national de la santé et de la recherche médicale et l'Agence nationale d'accréditation et d'évaluation de la santé leur concours au dispositif.
{verbe support de concours, futur}
[modifier] [première lettre]
4 L'entendre déclarer qu'elle n' pas de talent pour la vie privée se révèle également pertinent à la compréhension de sa démarche.
{verbe support de talent, présent}
[modifier] [première lettre]
5 L'information a jusqu'au palais de Justice de Neufchâteau.
{verbe support de information, participe passé}
accueil
exercices
morphologie
conjugaison
dérivation
genre
prépositions
collocations
synonymes
schémas
schémas
traduction
traduction

3. Applications

Focus on written language input & output

Native Language : Enrich and Exploit

1. *Reading companion*
2. *Resource generator*
3. *Exercise and test generator*

Learner Language : Finding Errors

4. *Error detector, feedback generator and automatic scoring tool*
5. *Writing aid*

Model Learner Behaviour : measure Complexity

6. *Input provider*
7. *Adaptive item sequencer*

-> Conceptual outline + applications from academic R&D

3.4 Error detector, feedback generator and automatic scoring tool

What?

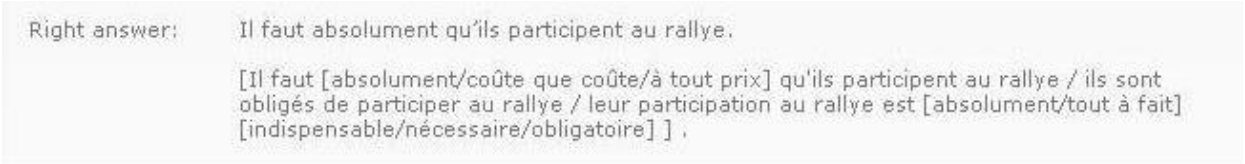
Automated analysis of learner output and generation of feedback

Not limited to closed exercises (MC, fill-in-the-blank, etc.) but also (semi-)open practice tasks (translation, correction, rephrasing, etc.)

How?

1) Approximate (or fuzzy) string matching (ASM)

-> anticipate all potential well-formed and ill-formed learner responses (with inclusion of regular expressions)



2) Parser-based (malrules, constraint relaxation, etc.) and/or statistical and machine learning methods

Towards (half-)open activities with automated analysis and feedback



Host VI Harvard University
August 11-15, 2015

HARVARD UNIVERSITY

Edumatic®

Bridging the gap between closed and open items
or
how to make CALL more intelligent

Piet Desmet & Bert Wylin

KU LEUVEN kulak

itec

iMinds
CONNECT. INNOVATE. CREATE

Projects:

Task design: require the learner to use certain language material

e-tutor-project (Trude Heift) http://www.e-tutor.sfu.ca
translation

Tagarela-project (Detmar Meurers) http://sifnos.sfs.uni-tuebingen.de/tagarela
answer should include certain words

Task design: offers clues towards a limited set of possible answers

Dialog Dungeon (ITEC)
gamified written dialog tasks
focus on specific set of problems
e.g. past tenses, subject-verb agreement, etc.

E-tutor project

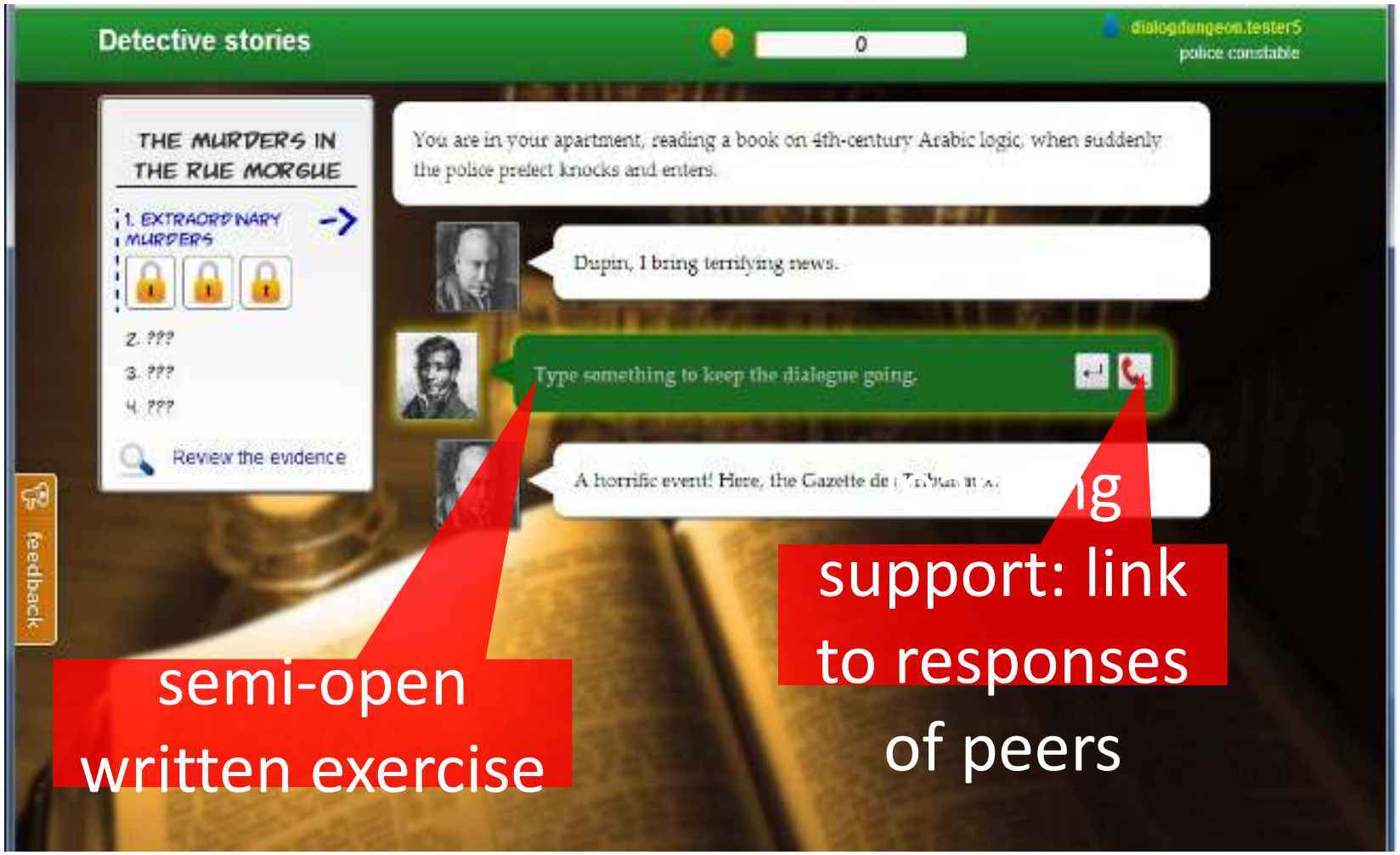


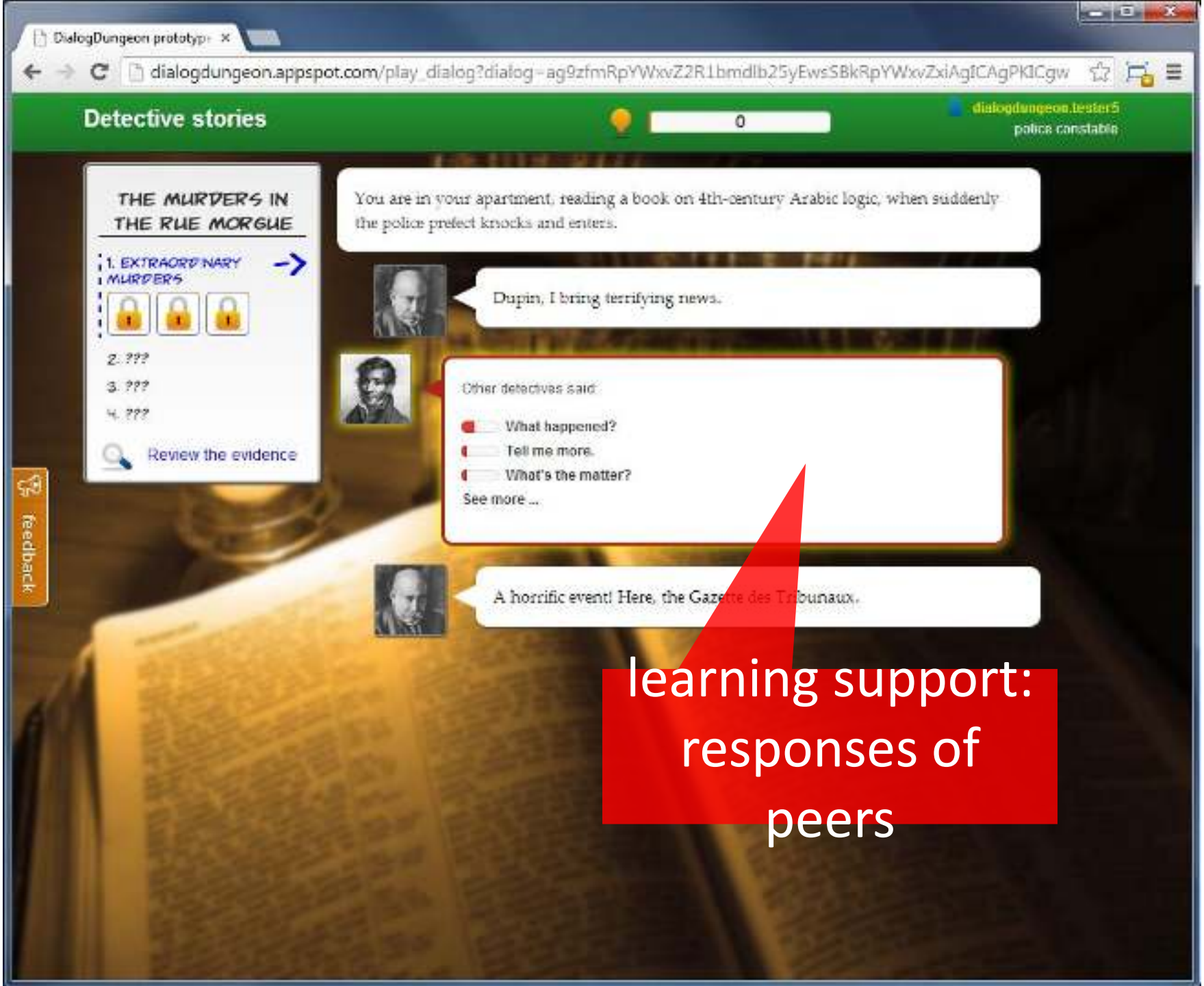
Tagarela-project

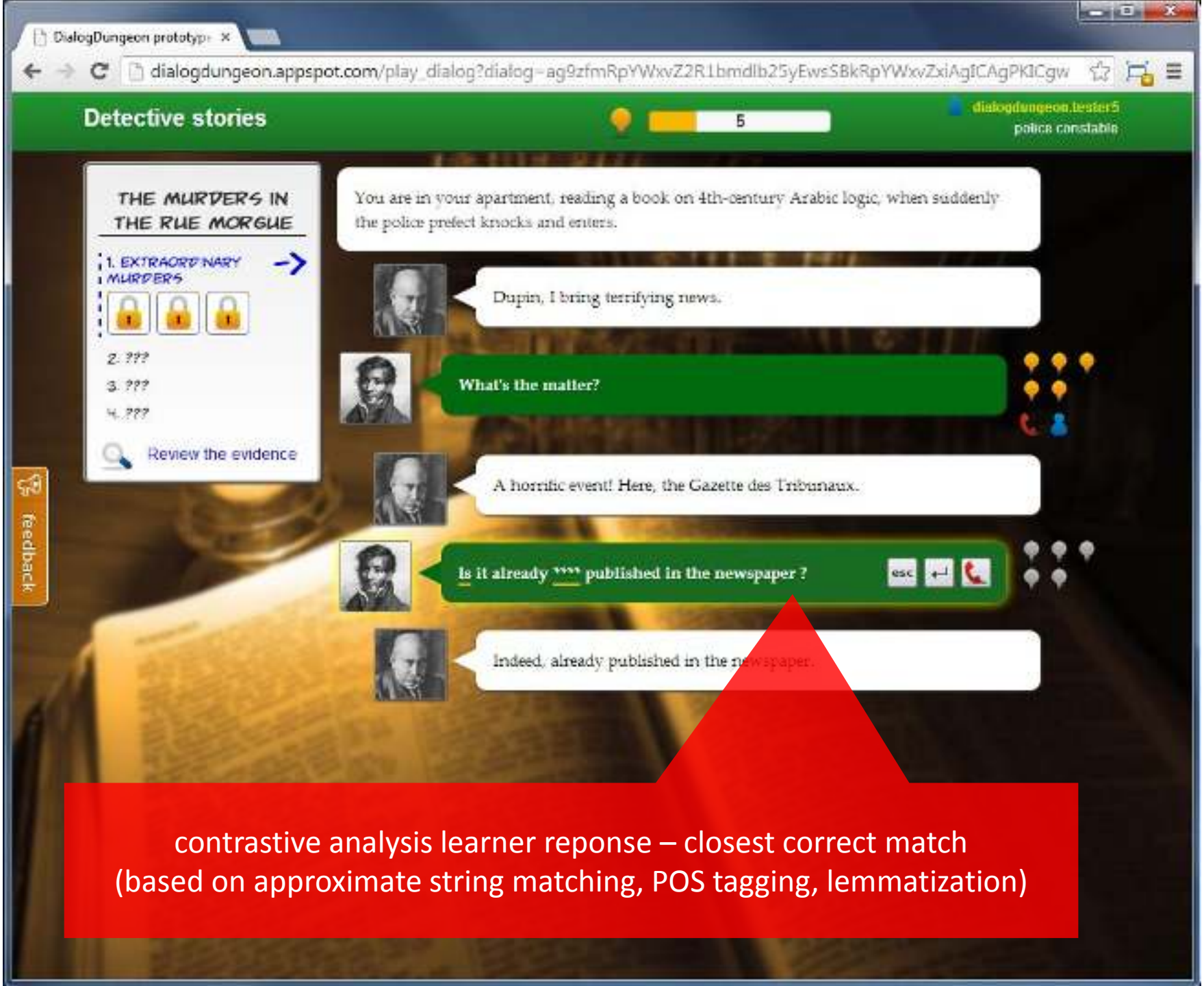


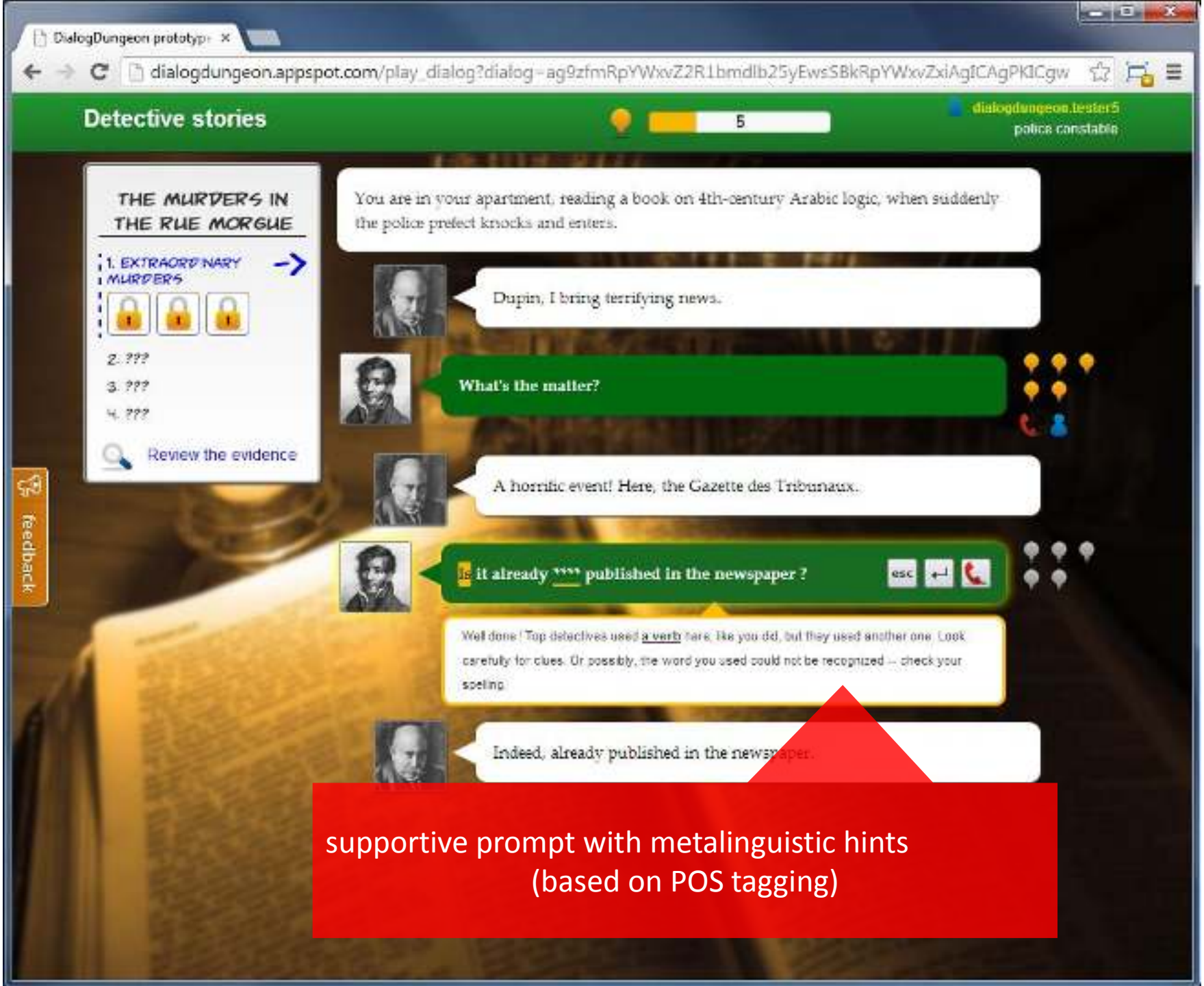
Dialog Dungeon
gamified written dialogue tasks

@Frederik Cornillie




DialogDungeon prototype
dialogdungeon.appspot.com/play_dialog?dialog=ag9zfmRpYWxvZ2R1bmdlb25yEwsSBkRpYWxvZxiAgICAgPKICgw
Detective stories
0
dialogdungeon.tester5
police constable
THE MURDERS IN THE RUE MORGUE
1. EXTRAORDINARY MURDERS
2. ???
3. ???
4. ???
Review the evidence
You are in your apartment, reading a book on 4th-century Arabic logic, when suddenly the police prefect knocks and enters.
Dupin, I bring terrifying news.
Other detectives said:
What happened?
Tell me more.
What's the matter?
See more ...
A horrific event! Here, the Gazette des Tribunaux.
feedback
learning support: responses of peers


DialogDungeon prototype
dialogdungeon.appspot.com/play_dialog?dialog=ag9zfmRpYWxvZ2R1bmdlb25yEwsSBkRpYWxvZxiAgICAgPKICgw
Detective stories
5
dialogdungeon.tester5
police constable
THE MURDERS IN THE RUE MORGUE
1. EXTRAORDINARY MURDERS
2. ???
3. ???
4. ???
Review the evidence
feedback
You are in your apartment, reading a book on 4th-century Arabic logic, when suddenly the police prefect knocks and enters.
Dupin, I bring terrifying news.
What's the matter?
A horrific event! Here, the Gazette des Tribunaux.
Is it already **** published in the newspaper ?
esc
Indeed, already published in the newspaper.
contrastive analysis learner reponse – closest correct match
(based on approximate string matching, POS tagging, lemmatization)


DialogDungeon prototype
dialogdungeon.appspot.com/play_dialog?dialog=ag9zfmRpYWxvZ2R1bmdlb25yEwsSBkRpYWxvZxiAgICAgPKICgw
Detective stories
5
dialogdungeon.tester5
police constable
THE MURDERS IN THE RUE MORGUE
1. EXTRAORDINARY MURDERS
2. ???
3. ???
4. ???
Review the evidence
feedback
You are in your apartment, reading a book on 4th-century Arabic logic, when suddenly the police prefect knocks and enters.
Dupin, I bring terrifying news.
What's the matter?
A horrific event! Here, the Gazette des Tribunaux.
Is it already **** published in the newspaper ?
esc
Well done ! Top detectives used a verb here, like you did, but they used another one. Look carefully for clues. Or possibly, the word you used could not be recognized -- check your spelling.
Indeed, already published in the newspaper.
supportive prompt with metalinguistic hints
(based on POS tagging)

3.5 Writing aid

What? Support the learner in writing a functional, well-formed text

How? No automated correction, but suggestions to help the learner

Real-time, online, assistance on
- lower-order skills (spelling) , lexical (including MWE)
- Lexico-grammatical, Grammatical skills
- Semantic, Pragmatic analysis

Through
- On-the-fly prompts
- Q&A interactive system
- Post-writing checks

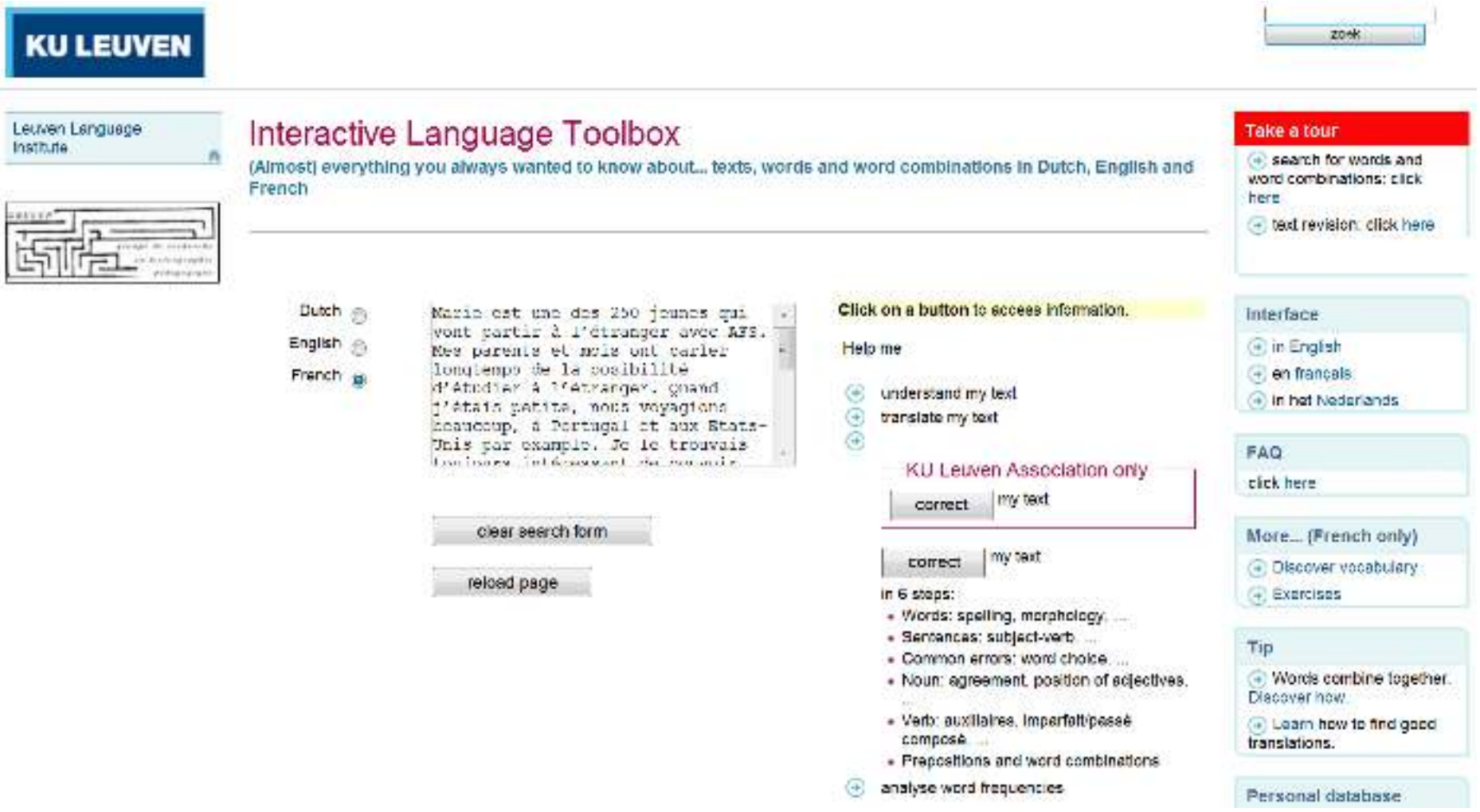
Projects:

Interactive Language Toolbox-project (Serge Verlinde)
https://ilt.kuleuven.be/inlato

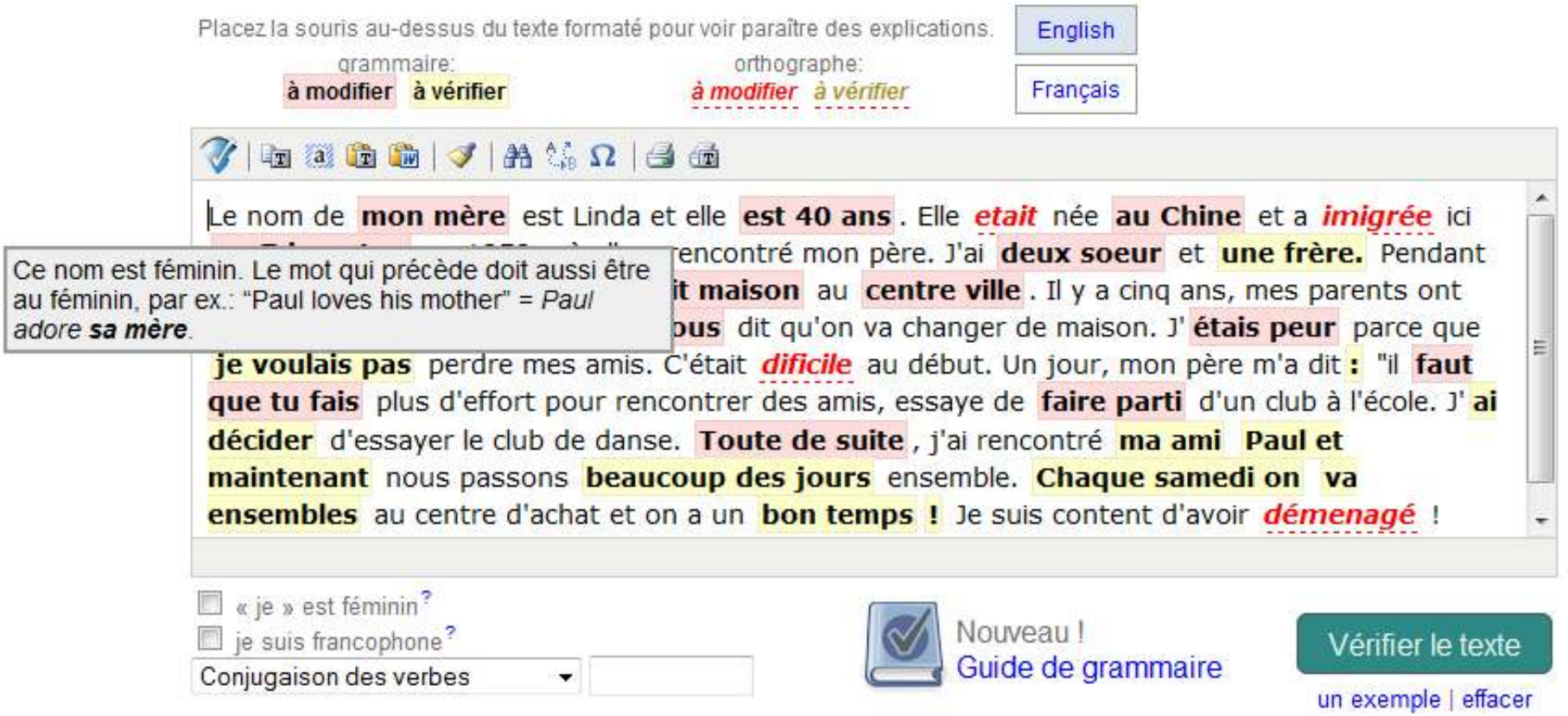
Bon patron, SpellCheckPlus & SpanishChecker (Nadaclair Language Technologies)
http://bonpatron.com; http://spellcheckplus.com; http://spanishchecker.com

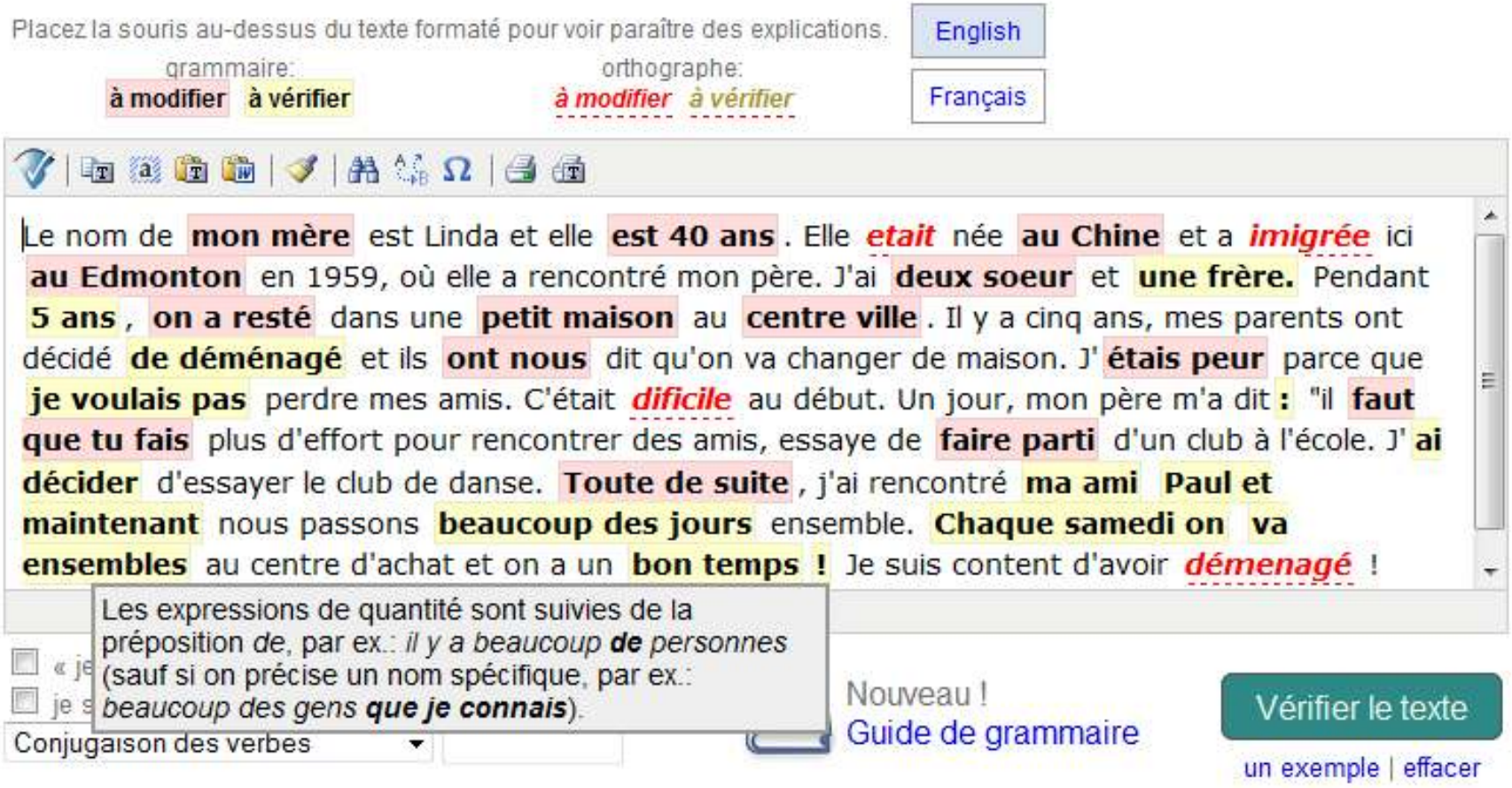
Glosser-project (Univ. of Sydney) http://www.glosserproject.org/en/

Interactive Language Toolbox

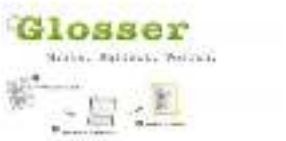



BonPatron.com
correcteur de grammaire et d'orthographe
Tapez ou collez votre texte en français ici et cliquez le bouton "Vérifier le texte".
« je » est féminin ?
je suis francophone ?
Conjugaison des verbes
Nouveau !
Guide de grammaire
Vérifier le texte
un exemple | effacer


Placez la souris au-dessus du texte formaté pour voir paraître des explications.
English
grammaire:
à modifier à vérifier
orthographe:
à modifier à vérifier
Français
Ce nom est féminin. Le mot qui précède doit aussi être au féminin, par ex.: "Paul loves his mother" = Paul adore sa mère.
Le nom de mon mère est Linda et elle est 40 ans . Elle etait née au Chine et a imigrée ici
encontré mon père. J'ai deux soeur et une frère. Pendant
it maison au centre ville . Il y a cinq ans, mes parents ont
ous dit qu'on va changer de maison. J' étais peur parce que
je voulais pas perdre mes amis. C'était dificile au début. Un jour, mon père m'a dit : "il faut
que tu fais plus d'effort pour rencontrer des amis, essaye de faire parti d'un club à l'école. J' ai
décider d'essayer le club de danse. Toute de suite , j'ai rencontré ma ami Paul et
maintenant nous passons beaucoup des jours ensemble. Chaque samedi on va
ensembles au centre d'achat et on a un bon temps ! Je suis content d'avoir démenagé !
« je » est féminin ?
je suis francophone ?
Conjugaison des verbes
Nouveau !
Guide de grammaire
Vérifier le texte
un exemple | effacer


Placez la souris au-dessus du texte formaté pour voir paraître des explications.
English
grammaire:
à modifier à vérifier
orthographe:
à modifier à vérifier
Français
Le nom de mon mère est Linda et elle est 40 ans . Elle etait née au Chine et a imigrée ici au Edmonton en 1959, où elle a rencontré mon père. J'ai deux soeur et une frère. Pendant 5 ans , on a resté dans une petit maison au centre ville . Il y a cinq ans, mes parents ont décidé de déménagé et ils ont nous dit qu'on va changer de maison. J' étais peur parce que je voulais pas perdre mes amis. C'était dificile au début. Un jour, mon père m'a dit : "il faut que tu fais plus d'effort pour rencontrer des amis, essaye de faire parti d'un club à l'école. J' ai décider d'essayer le club de danse. Toute de suite , j'ai rencontré ma ami Paul et maintenant nous passons beaucoup des jours ensemble. Chaque samedi on va ensembles au centre d'achat et on a un bon temps ! Je suis content d'avoir démenagé !
Les expressions de quantité sont suivies de la préposition de, par ex.: il y a beaucoup de personnes (sauf si on précise un nom spécifique, par ex.: beaucoup des gens que je connais).
Conjugaison des verbes
Nouveau !
Guide de grammaire
Vérifier le texte
un exemple | effacer

Glosser project (Sydney)



About

Glosser is a web-based framework for providing automated feedback. The framework allows writers to use different writing tools such as Wiki's, cloud-based Google Docs or MS Word. The documents are then processed to produce a wide range of types of feedback spanning collaborative or individual writing activities, visual and text modalities, feedback can be on surface or content features, on the writing product (the final document) or the process. This range of possibilities requires an architecture that is highly flexible and extensible.

Glosser provides different forms of feedback, each appropriate for specific writing genres. Current feedback 'tools' include:

Spelling. We have integrated ISpell so after a document is harvested the tool displays it with the misspelled words underlined. When clicking on the word the student can see a list of suggestions to fix the problem. This tool is a good example of the data flow from harvesting to user interaction so it is described in more detail later.

Flow. Two paragraphs are said to flow when they talk about similar topics. They do not flow when the semantic distance between them is too large. Glosser's tool displays warnings when the distance between two paragraphs is greater than a threshold (that can be changed by the user). This tool is a good example of how single document LSA can be used and is discussed in more detail later on.

Flow Map. This map provides a visualization of how the parts of a document relate and follow each other, producing a visual representation of a quality measures based on the average semantic distance between parts of a document (O'Rourke, et al., 2011). The visualization is not only useful to students but has also been shown to help tutors mark essays more efficiently and reliably (O'Rourke, et al., 2011).

Topics. This tool (shown in Figure 4) shows a table with the key topics developed in the document, the three most important sentences referring to that topic and the identification of the person who wrote most about it. Default trigger questions include: "Are the ideas used in the essay relevant to the question?", "Are the ideas developed correctly?".

Topic Map. The topic map tool is used to identify the structure of topics developed in the document. The tool identifies keywords using a network-based analysis then maps them to a visual representation. If two keywords are used in the same sentence, they are linked in the visualization. The structure of topics can then be subjectively inferred from the semantics of the visualization. The sentences that cover more than one of these topics can be seen when the link between them is shown on a rollover. The interactive visualization tool is implemented as a Java applet....

Concept Map. The topic map tool uses a naïve approach, where . The concept map tool incorporates an automated concept map generator discussed in more detail elsewhere (Jorge Villalon & Calvo, 2009; J. Villalon & Calvo, 2011).

Participation. This tool is aimed at supporting collaborative writing activities and/or the teaching the awareness of the writing process.

Automatic Questions. Questions of different types can be generated from a single document. We have developed an approach to automatic question generation (Liu, et al., 2010) integrated into this tool. Since the questions currently available are produced from citations found in the document, this tool is mostly useful for activities involving literature reviews.

Glosserproject is kindly supported by YourKit. YourKit, LLC is the creator of innovative and intelligent tools for profiling Java and .NET applications, and they support open source projects with licenses for their full-featured **profilers.** Take a look at YourKit's leading software products: YourKit Java Profiler and YourKit .NET Profiler.

3. Applications

Focus on written language input & output

Native Language : Enrich and Exploit

1. *Reading companion*
2. *Resource generator*
3. *Exercise and test generator*

Learner Language : Finding Errors

4. *Error detector, feedback generator and automatic scoring tool*
5. *Writing aid*

Model Learner Behavior : measure Complexity

6. *Input provider*
7. *Adaptive item sequencer*

-> Conceptual outline + applications from academic R&D

3.6 Input provider

What? Support learners in searching for texts adapted to
Their current proficiency level and interest
Optimising for reaching the next level

How? (Semi-)automated selection of comprehensible & authentic text material
For vocabulary and reading practice

- Linguistic analysis of readability and complexity
  Analysis of formal, lexical and semantic complexity
- Operationalize proficiency and link to complexity
- Perform syntactic and lexical text simplification/elaboration
- Determine topical interest of texts

Projects:

REAP-project (Carnegie Mellon) http://reap.cs.cmu.edu
"Reader-Specific Lexical Practice for Improved Reading Comprehension"

SATO-project (François Daoust) http://www.ling.uqam.ca/sato
"Système d'analyse de texte par ordinateur"

+ SATO-Calibrage
Automated formal analysis of a corpus (existing or personal)

Aim: automatically identify the lexical items in the text that are complex for specific learners

Measures for the prediction of word complexity

1. Linguistically motivated measures
2. Resource-based measures
3. Learner-based measures

© Thomas François (UCL)

1. Linguistically motivated measures

= identify word characteristics that are good predictors of word difficulty, especially ***psycholinguistic variables*** (Shardlow 2013; Gala et al. 2014)

- *lexical sophistication vs. basic vocabulary list*
- *word frequency (SUBTLEX, Lexique3, ...)*
- *word sense ambiguity (WordNet synset)*
- *complex orthographical patterns*
- *frequency of morphological affixes*
- *...*

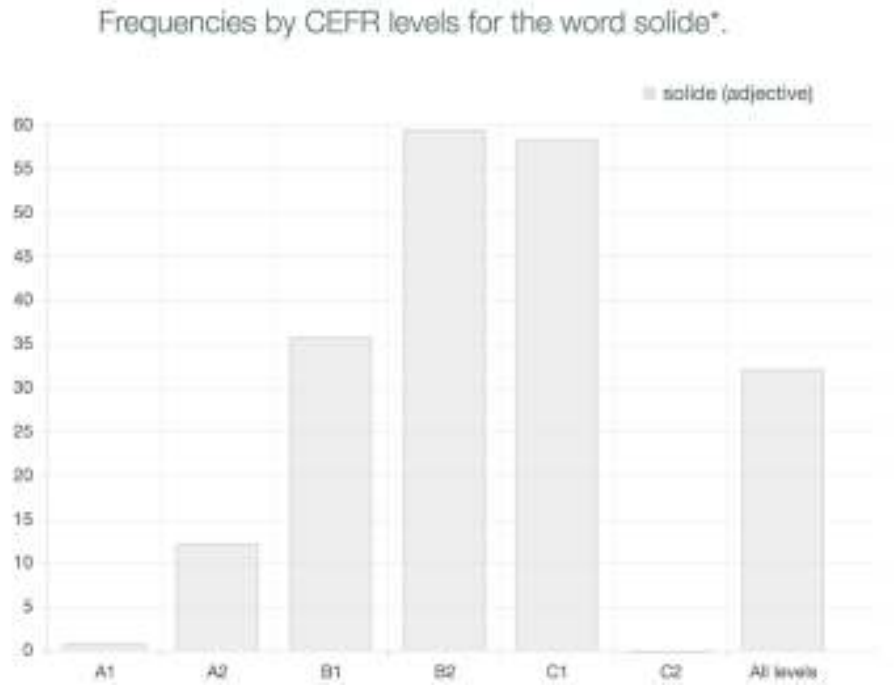
© Thomas François (UCL)

2. Resource-based measures

FLELEX: A graded lexical resource
http://cental.uclouvain.be/flelex/

- linking word frequency and word complexity
- dictionary of frequencies observed along a difficulty scale
    - L2 : CEFR-graded texts and textbooks (François et al., 2014)

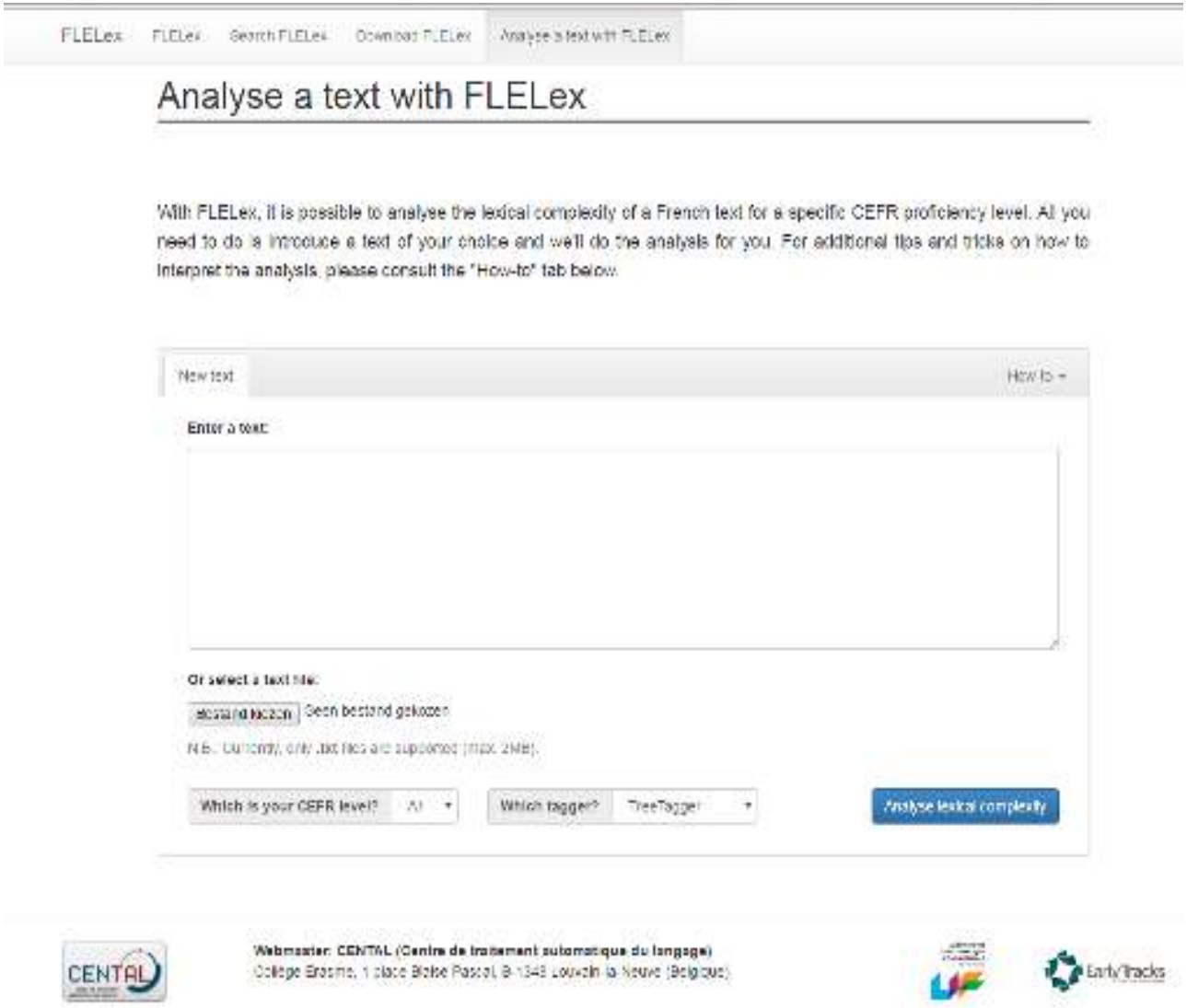
© Thomas François (UCL)



http://cental.uclouvain.be/flelex

Main issues:

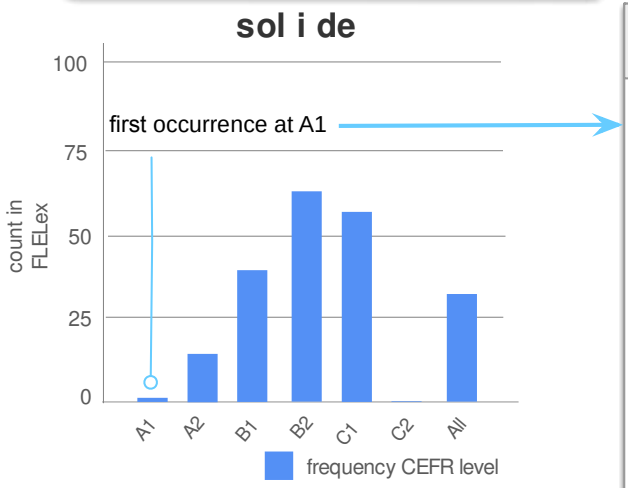
- ❑ What is the word's CEFR level? In other words, what level targets the word as « core vocabulary » to be acquired at this level ?
- ❑ How can we identify this level?
  - first occurrence
  - threshold the Xth percentile
  - ...
- ❑ How can we distinguish core and peripheral vocabulary?
  - complex words observed in A1 text might be fortuitous (i.e. peripheral vocabulary)
- ❑ How to deal with phraseology?


FLELex
FLELex
Search FLELex
Download FLELex
Analyse a text with FLELex
Analyse a text with FLELex
With FLELex, it is possible to analyse the lexical complexity of a French text for a specific CEFR proficiency level. All you need to do is introduce a text of your choice and we'll do the analysis for you. For additional tips and tricks on how to interpret the analysis, please consult the "How-to" tab below.
New text
How to
Enter a text:
Or select a text file:
Which is your CEFR level?
Which tagger?
TreeTagger
Analyse lexical complexity
Webmaster: CENTAL (Centre de traitement automatique du langage)
CENTAL
Early Tracks

Lexical competence prediction — Expert model

Predictive model based on a **graded lexical resource**

→ Predict known and unknown words in a new text *for a given level*



Archimède — Level A2

physicien, mathématicien et ingénieur. Né en 287 av. J. -C. et mort en 212 av. J. -C. à Syracuse, il a vécu en Sicile (île actuellement italienne).

unknown (B1)

Œuvre scientifique.

Archimède fut le premier à mettre en équations mathématiques les faits faciles à observer dans la nature. Ses livres démontrèrent à l'aide des figures géométriques et des proportions, comment sont fait les équilibres, ou les centres de gravité des corps solides, quelles longueurs doivent avoir des leviers pour soulever les poids.

(B1) unknown

unknown (OOV)

unknown (B1)

known (A1)

unknown (OOV)

Tack, A., François, T., Ligozat, A.-L., & Fairon, C. (2016). Evaluating Lexical Simplification and Vocabulary Knowledge for Learners of French: Possibilities of Using the FLELex Resource. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC'16)* (pp. 230–236). Portorož, Slovenia.

© Thomas François & Anaïs Tack (UCL)

3. Subjective measures

= identify words that are found difficult by a specific learner

Problem: expert model gives one-sided predictions

- identical predictions for all learners having the same level
- OOV words (out-of vocabulary) either because of (a) technicality or (b) editorial choice

Towards a more fine-grained & personalized prediction

- integrated model of word complexity features
  - psycholinguistic + resource-based + subjective measures (e.g. cognates)
- adapt predictions according to learner feedback
  - triggered by a learner clicking on a difficult word
  - incremental learning of feature weights

Lexical competence prediction — Learner model

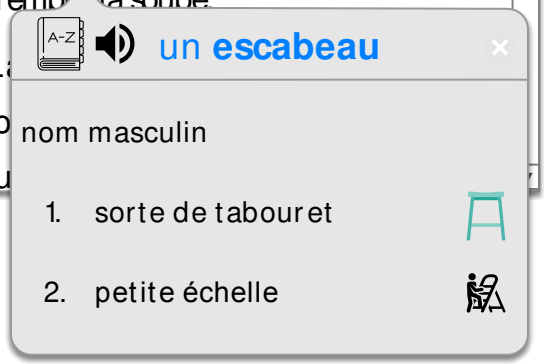
Predictive model based on **learner input**

Predict known and unknown words in a new text *for a given learner*

L'Auberge — *Guy de Maupassant*

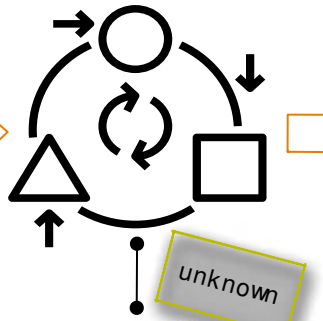
nous n'avons plus de femme maintenant, il faut préparer le dîner, tu vas éplucher les pommes de terre.

Et tous deux, s'asseyant sur des escabeaux de bois, commencèrent à tremper la soupe.



learner id A2-3

adaptive model



escabeau
cat egor y=3
l evel =7
char s=9
senses=2
. . .

learner id A2-3

Archimède

J. -C. à Syracuse, il a vécu en Sicile (île actuellement italienne).
unknown
Œuvre scientifique.
unknown
Archimède fut le premier à mettre en
unknown
équations mathématiques les faits faciles à observer dans la nature. Ses livres démontrèrent à l'aide des figures
unknown
géométriques et des proportions, comment sont fait les équilibres, ou les
unknown

Tack, A., François, T., Ligozat, A.-L., & Fairon, C. (2016). Modèles adaptatifs pour prédire automatiquement la compétence lexicale d'un apprenant de français langue étrangère. In *Actes de la 23ème Conférence sur le Traitement Automatique des Langues Naturelles (TALN'16)* (pp. 221–234). Paris, France.

3.7 Adaptive Item Sequencer

a) Adaptivity: to what?

- difficulty level of the tasks
- learner profile (prior knowledge, motivation, cognitive load, interests & preferences)
- context (time and place, device, etc.)

(b) **What to adapt?**

- adaptive form representation (form in which content is presented to the learner, e.g. dynamically generated hypermedia pages)
- adaptive content representation (e.g. with or without learner support)
- adaptive curriculum sequencing (e.g. selection of items in function of difficulty level, learner profile, etc.)

(c) **How to implement adaptivity?**

- full program control (via reasoning component 'if.. then…' rules)
- full learner control
- shared control

Wauters, K., Desmet, P., Van Den Noortgate, W. (2010). Adaptive Item-Based Learning Environments Based on the Item Response Theory: Possibilities and Challenges. *Journal of Computer Assisted Learning, 26* (6), 549-562.

3.7 Adaptive Item Sequencer

What about AI?

a) Adaptivity: to what?

- linguistic complexity -> language agent (expert system)
- item difficulty -> IRT
- learner profile (prior knowledge, motivation, cognitive load, interests & preferences) -> student agent

(b) **What to adapt?**

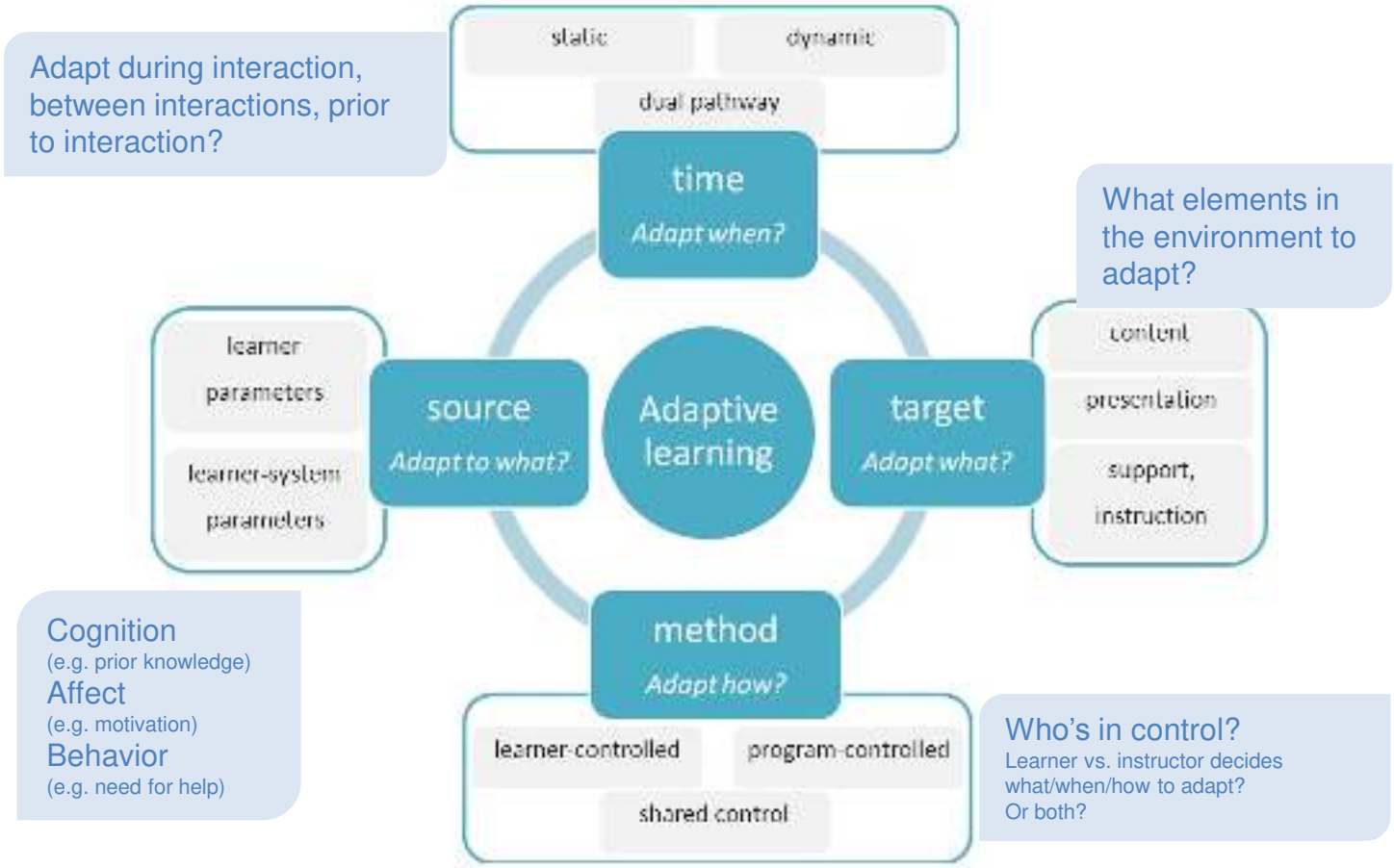
- adaptive curriculum sequencing (e.g. selection of items in function of difficulty level, learner profile, etc.) -> tutor agent

(c) **How to implement adaptivity?**

full or shared control -> tutor agent

Beuls, Katrien (2013) *Processing, learning and tutoring of Spanish verb morphology*. Brussels: VUB. (PhD)

4D-model of adaptive instruction



Vandewaetere, Desmet & Clarebout 2011 / Vandewaetere & Clarebout, 2012



http://www.slideshare.net/piet_desmet/2015-0522-presentatiecalicodesmet-vandewaetere-def

Adaptive learning

"Harnessing collective intelligence in order to make e-learning environments adaptive"
(IOF knowledge platform KU Leuven, 2007-2011)



ADAPTIVITY ON ITEM LEVEL

IRT FOR ADAPTIVE LEARNING
DYNAMIC LEARNING CURVES

DR. KELLY WAUTERS
(2012)



ADAPTIVITY ON LEARNER LEVEL

LEARNER CONTROL AND OPEN LEARNER MODELS
TOOL USE AND HELP-SEEKING

DR. MIEKE VANDEWAETERE
(2011)



ADAPTIVITY ON CONTEXT LEVEL

ONTOLOGIES FOR ADAPTIVE LEARNING
PERVASIVE, CONTEXT-AWARE SYSTEMS

DR. AHMET SOYLU
(2012)

4. Future Developments

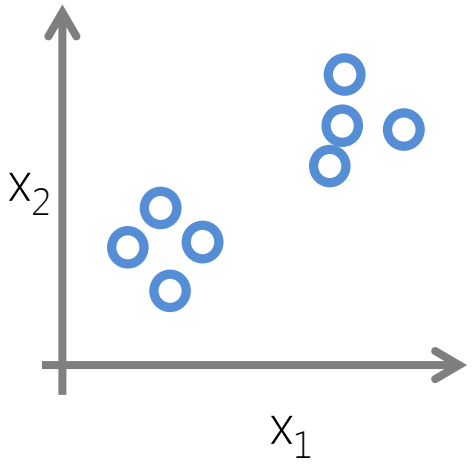
4.1. Native Language
- Unsupervised learning: Understand examples
- Learn structure, explore

4.2. Learner Language
- Towards a different model of interlanguage
- Using CALL applications for gradual error detection

4.3. Measuring linguistic complexity
- Individualized Lexical complexity
- Link known with unknown vocabulary items

4.1 Native Language

Unsupervised Learning on Native Language



Description of examples
through more finegrained clustering
without labels
Interpretation by the learner himself
~ Concordancers
~ Similar words
~ Similar constructions

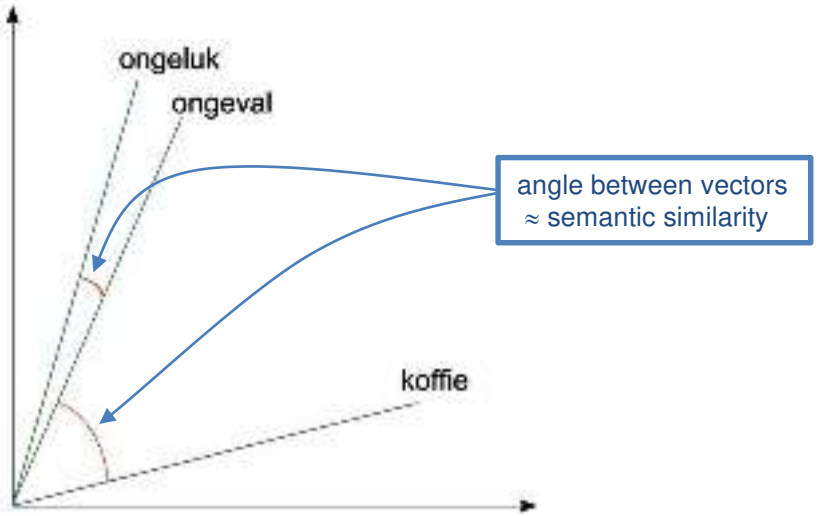
Shifts the responsibility of interpretation to the learner himself
But based on a much more informative clustering of examples

Semantic Vector Spaces

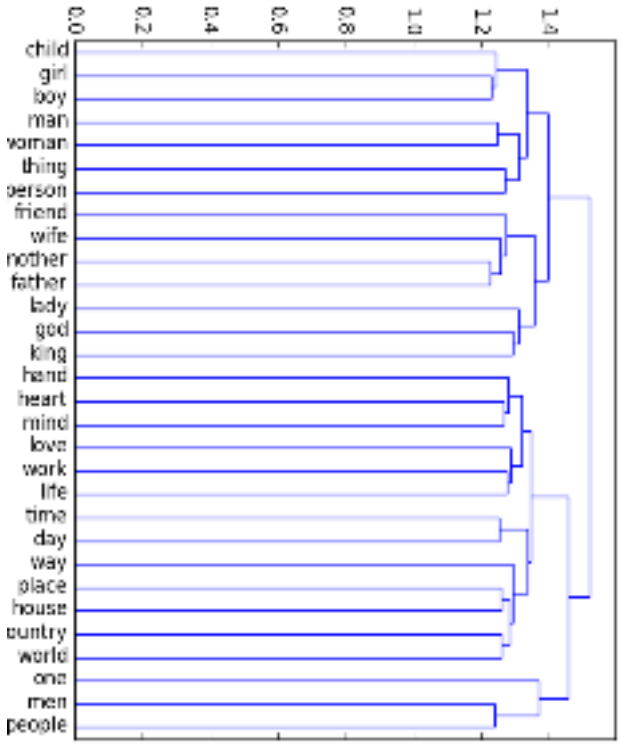
At the word level: Semantic Vector Spaces
"You shall know a word by the company it keeps" (Firth, 1957)

1) Paradigmatic characteristics of words can be retrieved on the basis of their syntagmatic behaviour.
2) Map the usage in context of a word in a high dimensional space (e.g. with contexts as dimensions, and #occurrences as value)
3) Proximity in space signals **semantic relatedness**
   Even relations can be represented as vectors in space

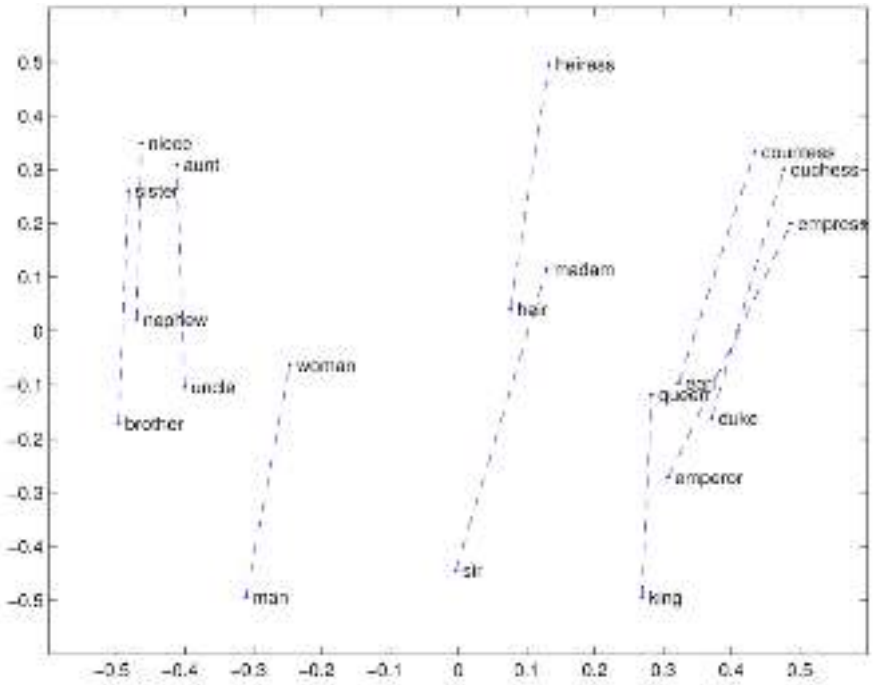
Unsupervised semantic comparisons



As the basis of Resource Creation



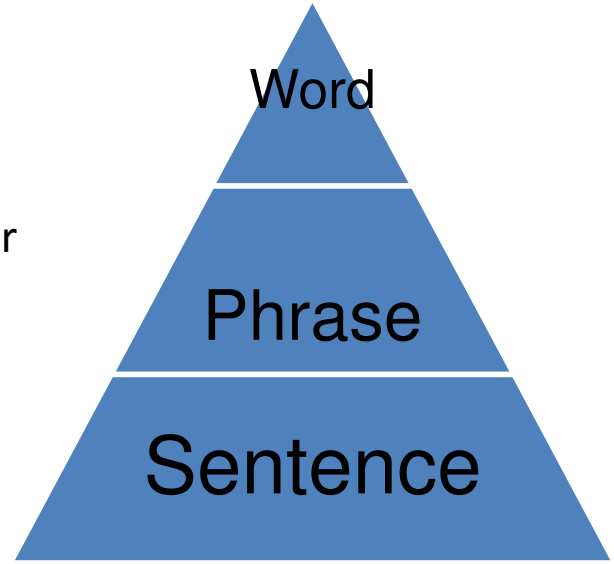
For visualising relationships



Semantic Vector Spaces

A promising future beyond the word level

The **same representation** for more complex linguistic structures using the principles of **compositionality**.



4.2. Modeling Learner Language

Now starts from native language model
= basis of error analysis

Native language is a fully developed linguistic **system**

E.g. POS-tags: What defines a POS-category knowing the full system?

- Distributional restrictions : N usually follows Det/Adj
- Morphological restrictions: N + plural s
- Lexical restrictions:

⇒Resolves to ONE POS-tag
Thus possible to train a pos-tagger

4.2.1. Towards a different model of interlanguage

However: POS might be a non-existing construct for a learner (Diaz et al., 2010)

“He is joy” (He is happy)
⇒ ‘joy’ is not permitted in the slot.

Central : learner does **not understand/know** part(s) of the native system

Solution: **no full POS resolution**
But model each of the **system-components** separately

- Distributional restrictions
- Morphological restrictions
- Lexical restrictions

⇒Does not necessarily converge to a single POS-tag
⇒Partial information can then be used for corrections/feedback

4.2.2. Gradual error detection

CALL tools provide unseen (and underused) possibilities to

- Capture L2 production
- Advance towards fully open output

How? Ask for immediate feedback from the learner

Have the tool disambiguate the output at production times
A constant closed feedback loop

Progressive Feedback

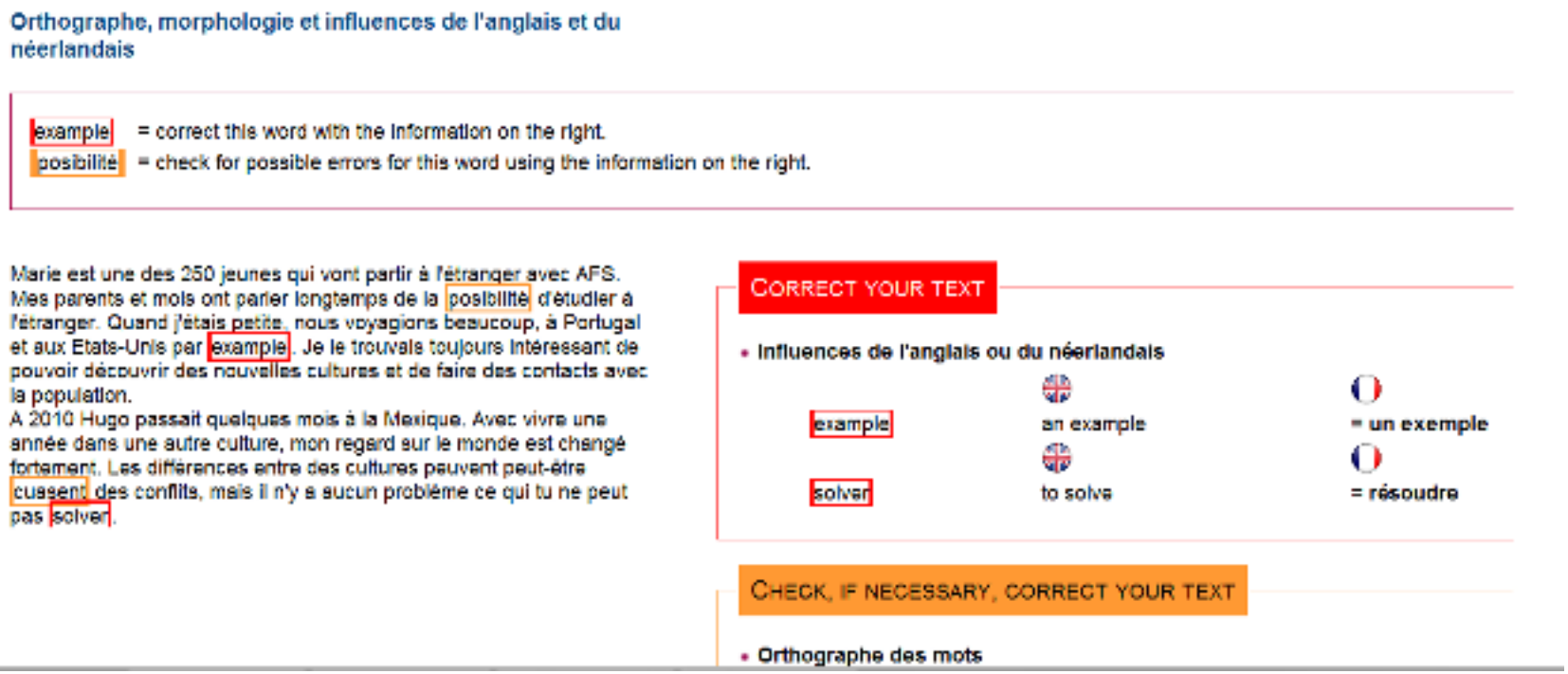
In a progressive manner: the algorithm does not propagate wrong assumptions

1) orthography
2) POS-disambiguation
3) Morphology
4) Argument structure
5) ...

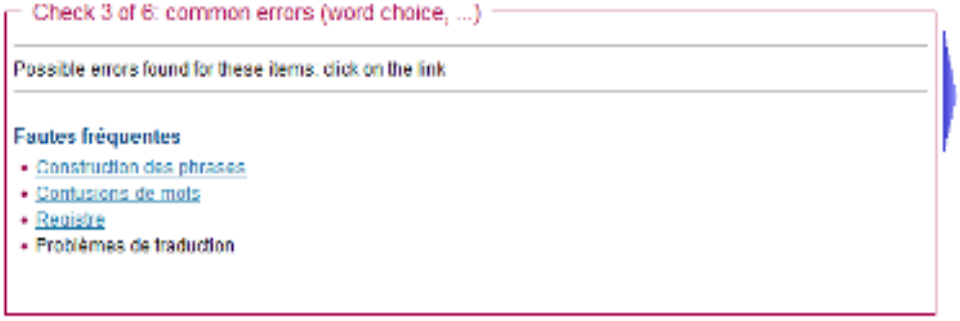
At each step construct and adjust the hypothesis
Have the system come as close as possible to verification

- Helping the user with analogous examples from corpora
- Through explanation of linguistic regularities
- …

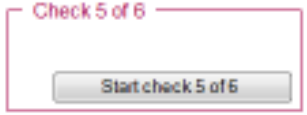

Check 1 of 6: words (spelling, morphology, ...)
Possible errors found for these items: click on the link(s)
Les mots
• Orthographe (spelling), morphologie et influences de l'anglais et du néerlandais
Check 2 of 6
Start check 2 of 6


Orthographe, morphologie et influences de l'anglais et du néerlandais
example = correct this word with the information on the right.
posibilité = check for possible errors for this word using the information on the right.
Marie est une des 250 jeunes qui vont partir à l'étranger avec AFS. Mes parents et mois ont parler longtemps de la posibilité d'étudier à l'étranger. Quand j'étais petite, nous voyagions beaucoup, à Portugal et aux Etats-Unis par example. Je le trouvais toujours intéressant de pouvoir découvrir des nouvelles cultures et de faire des contacts avec la population.
A 2010 Hugo passait quelques mois à la Mexique. Avec vivre une année dans une autre culture, mon regard sur le monde est changé fortement. Les différences entre des cultures peuvent peut-être cuasent des conflits, mais il n'y a aucun problème ce qui tu ne peut pas solver.
CORRECT YOUR TEXT
• Influences de l'anglais ou du néerlandais
example
an example
= un exemple
solver
to solve
= résoudre
CHECK, IF NECESSARY, CORRECT YOUR TEXT
• Orthographe des mots

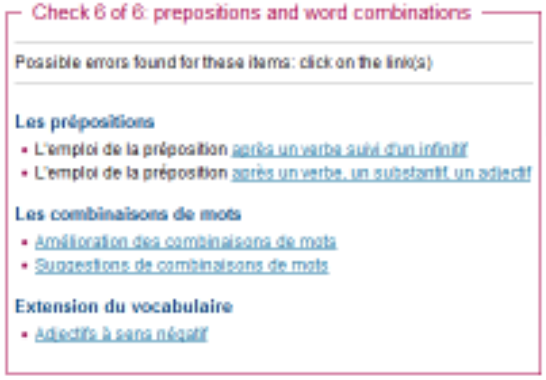

Check 2 of 6: sentences (subject-verb, ...)
Check these items for the whole text: click on the links
Les phrases
• Accord sujet-verbe: contrôle général
• Place des compléments et emploi des prépositions


Check 3 of 6: common errors (word choice, ...)
Possible errors found for these items: click on the link
Fautes fréquentes
• Construction des phrases
• Confusions de mots
• Registre
• Problèmes de traduction


Check 4 of 6: noun (agreement, position of adjectives, ...)
Possible errors found for these items: click on the link(s)
Les noms/substantifs
• Le genre: masculin et féminin
• La place de l'adjectif
Les noms de pays
• Le genre et emploi des prépositions


Check 5 of 6
Start check 5 of 6


Check 5 of 6: verb (auxiliaires, imparfait/passé composé, ...)
Les verbes
• L'accord sujet/verbe
• L'emploi du gérondif
• L'emploi du subjonctif
• L'emploi de l'imparfait et du passé composé
• L'emploi des auxiliaires

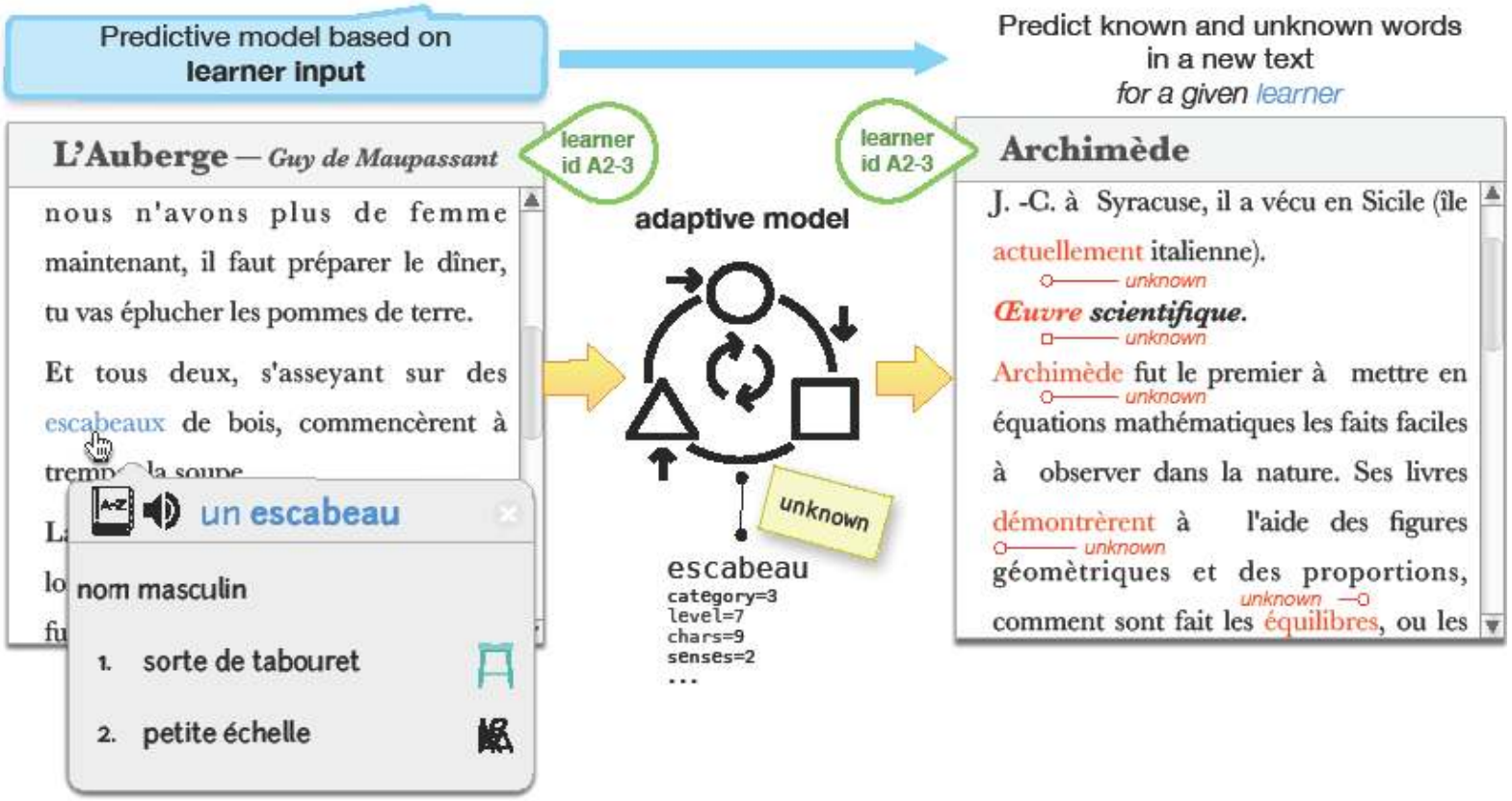

Check 6 of 6: prepositions and word combinations
Possible errors found for these items: click on the link(s)
Les prépositions
• L'emploi de la préposition après un verbe suivi d'un infinitif
• L'emploi de la préposition après un verbe, un substantif, un adjectif
Les combinaisons de mots
• Amélioration des combinaisons de mots
• Suggestions de combinaisons de mots
Extension du vocabulaire
• Adjectifs à sens négatif

4.3 Intelligent tutors and linguistic complexity

Steps away from procedures that rely on “one-size fits all”
= A truly individualized learning program

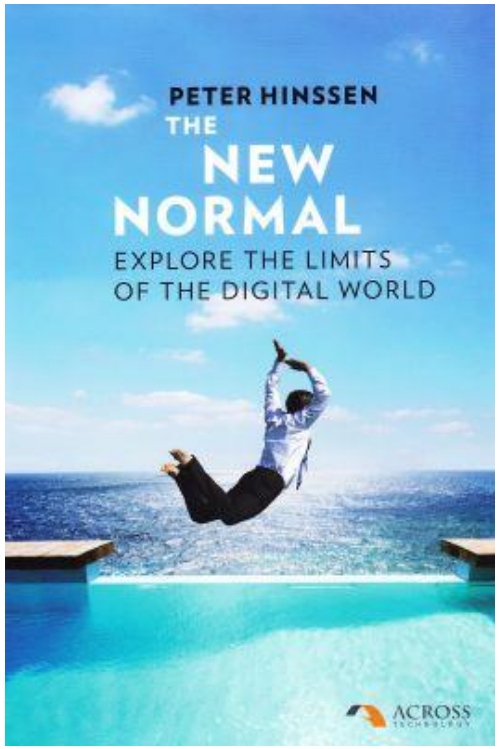
- Capture and store L2 input and production per learner
- Determine known and unknown vocabulary items

Lexical competence prediction — Learner model



Steps away from procedures that rely on “one-size fits all”
= A truly individualized learning program

Additional possibilities:
Link unknown **vocabulary items** to
known vocabulary items:
Thesauri, dictionaries
Semantic Vector Spaces



Conclusion…

NLP and CALL

Provide a world of possibilities

Analyse Native language	- Resource creation
Analyse Learner language	- Exercise Correction
Measure complexity	- Intelligent tutoring

Provide the technological means to create ready to use applications

Future possibilities include visualization to enhance understanding, immediate feedback systems, individualized learning systems

=> Warranted qualified optimism about NLP-enhanced CALL

Contact

Piet Desmet

Piet.Desmet@kuleuven.be

be.linkedin.com/in/pietdesmet

@PietDesmet

ITEC

www.kuleuven.be/itec

