TAGARELA: A web based intelligent workbook for Portuguese

Detmar Meurers and Ramon Ziai

based on joint research with
Luiz Amaral (UMass Amherst)

Berlin. October 15, 2009

TAGARELA: An Intelligent Tutoring System

- TAGARELA: Teaching Aid for Grammatical Awareness, Recognition and Enhancement of Linguistic Abilities (Amaral & Meurers 2005, 2006, 2007a,b, 2008, 2009; Amaral 2007; Ziai 2009)
- an intelligent web-based workbook for beginning learners of Portuguese
- designed to satisfy real-life FLT needs identified at OSU
- provide opportunities for students to practice their listening, reading, and writing skills
- offers individual feedback on learner input to system
- foster learner awareness of language forms and categories (Long 1991, 1996; Ellis 1994; Schmidt 1995; Lyster 1998; Lightbown & Spada 1999; Norris & Ortega 2000; Schulz 2002)

TAGARELA
Detmar Meurers & Ramon Ziai
Introduction
Feedback
System Architecture
The three models
Expert model: NLP
Annotation-based setup
Activity model
Relevance for processing
Analyzing learner language
On Tokenization
Interpretation
Portuguese Properties
Mismatches in the identification of tokens
Solution
On interpreting accented characters
Interpretation
Portuguese Properties
Mismatches in the interpretation of tokens
Solution
Wrapping up
Conclusion
Appendix
Screenshots
Eberhard Karls Universität Tübingen

System role, Activity types, Interface

- What role does the system play in teaching?
  → Self-guided activities accompanying teaching
- What type of activities are appropriate and useful for fostering awareness (and fit into the FLT approach)?
  → Activities ideally involve both form and meaning, such as listening/reading comprehension questions.
  - TAGARELA offers six types of activities:
    - listening comprehension
    - reading comprehension
    - picture description
    - fill-in-the-blank
    - rephrasing
    - vocabulary

    Similar to traditional workbook exercises, plus audio.
- What should the system interfaces look like?
  → Use L2 as far as possible (needs careful interface design).

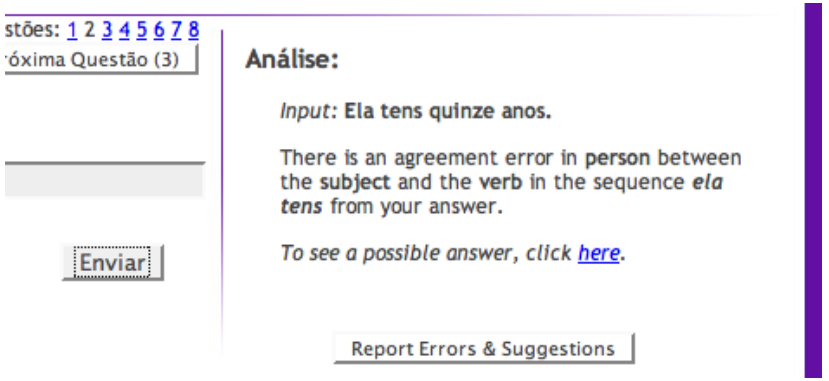
TAGARELA
Detmar Meurers & Ramon Ziai
Introduction
Feedback
System Architecture
The three models
Expert model: NLP
Annotation-based setup
Activity model
Relevance for processing
Analyzing learner language
On Tokenization
Interpretation
Portuguese Properties
Mismatches in the identification of tokens
Solution
On interpreting accented characters
Interpretation
Portuguese Properties
Mismatches in the interpretation of tokens
Solution
Wrapping up
Conclusion
Appendix
Screenshots
Eberhard Karls Universität Tübingen

Providing Feedback

- TAGARELA provides on-the-spot feedback on
  - orthographic errors (non-words, spacing, capitalization, punctuation)
  - syntactic errors (nominal and verbal agreement)
  - semantic errors (missing or extra concepts, word choice)
- Providing feedback on meaning becomes crucial for activities such as reading and listening comprehension.

TAGARELA
Detmar Meurers & Ramon Ziai
Introduction
Feedback
System Architecture
The three models
Expert model: NLP
Annotation-based setup
Activity model
Relevance for processing
Analyzing learner language
On Tokenization
Interpretation
Portuguese Properties
Mismatches in the identification of tokens
Solution
On interpreting accented characters
Interpretation
Portuguese Properties
Mismatches in the interpretation of tokens
Solution
Wrapping up
Conclusion
Appendix
Screenshots
Eberhard Karls Universität Tübingen

Feedback on Agreement



TAGARELA
Detmar Meurers & Ramon Ziai
Introduction
Feedback
System Architecture
The three models
Expert model: NLP
Annotation-based setup
Activity model
Relevance for processing
Analyzing learner language
On Tokenization
Interpretation
Portuguese Properties
Mismatches in the identification of tokens
Solution
On interpreting accented characters
Interpretation
Portuguese Properties
Mismatches in the interpretation of tokens
Solution
Wrapping up
Conclusion
Appendix
Screenshots

Feedback on Word Choice

TAGARELA
Detmar Meurers & Ramon Ziai
Introduction
Feedback
System Architecture
The three models
Expert model: NLP
Annotation-based setup
Activity model
Relevance for processing
Analyzing learner language
On Tokenization
Interpretation
Portuguese Properties
Mismatches in the identification of tokens
Solution
On interpreting accented characters
Interpretation
Portuguese Properties
Mismatches in the interpretation of tokens
Solution
Wrapping up
Conclusion
Appendix
Screenshots

TAGARELA
Detmar Meurers & Ramon Ziai
Introduction
Feedback
System Architecture
The three models
Expert model: NLP
Annotation-based setup
Activity model
Relevance for processing
Analyzing learner language
On Tokenization
Interpretation
Portuguese Properties
Mismatches in the identification of tokens
Solution
On interpreting accented characters
Interpretation
Portuguese Properties
Mismatches in the interpretation of tokens
Solution
Wrapping up
Conclusion
Appendix
Screenshots

Feedback on Wrong Word

Questões: 1 2 3 4
ma Questão (2)

e Beatriz

Análise:

Input: **Elas se chamam Ana e Maria.**

I think there is a problem with the proper noun you have chosen.

Are you sure you want to use **Maria** instead of **Beatriz**?

To see a possible answer, click here.

TAGARELA

Detmar Meurers & Ramon Ziai

Introduction
Feedback
System Architecture
The three models
Expert model: NLP
Annotation-based setup
Activity model
Relevance for processing
Analyzing learner language
On Tokenization
Interpretation
Portuguese Properties
Mismatches in the identification of tokens
Solution
On interpreting accented characters
Interpretation
Portuguese Properties
Mismatches in the interpretation of tokens
Solution
Wrapping up
Conclusion
Appendix
Screenshots

Feedback on Missing Verb

Questões: 1 2 3 4 5
óxima Questão (4)

Enviar

Análise:

Input: **Qual o seu nome?**

Your answer is close, but there is a **verb** missing in your sentence.

To see a possible answer, click here.

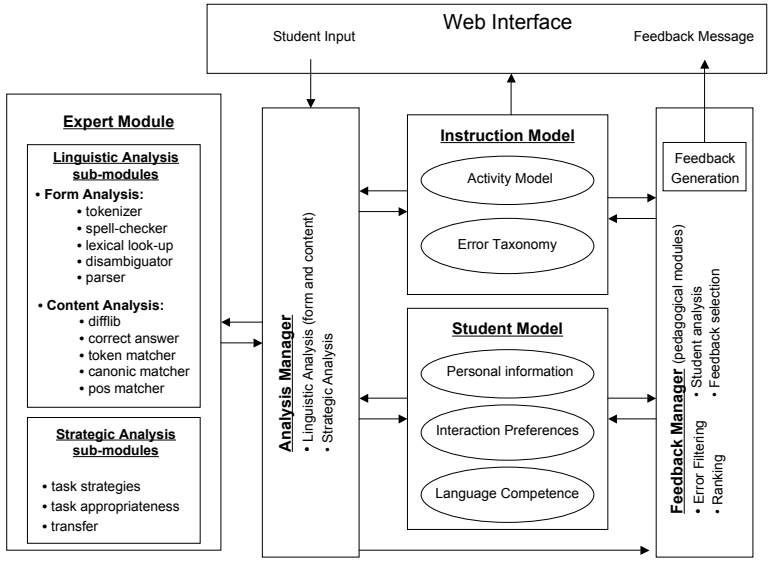
Report Errors & Suggestions

TAGARELA

Detmar Meurers & Ramon Ziai

Introduction
Feedback
System Architecture
The three models
Expert model: NLP
Annotation-based setup
Activity model
Relevance for processing
Analyzing learner language
On Tokenization
Interpretation
Portuguese Properties
Mismatches in the identification of tokens
Solution
On interpreting accented characters
Interpretation
Portuguese Properties
Mismatches in the interpretation of tokens
Solution
Wrapping up
Conclusion
Appendix
Screenshots

General Architecture of TAGARELA



TAGARELA

Detmar Meurers & Ramon Ziai

Introduction
Feedback
System Architecture
The three models
Expert model: NLP
Annotation-based setup
Activity model
Relevance for processing
Analyzing learner language
On Tokenization
Interpretation
Portuguese Properties
Mismatches in the identification of tokens
Solution
On interpreting accented characters
Interpretation
Portuguese Properties
Mismatches in the interpretation of tokens
Solution
Wrapping up
Conclusion
Appendix
Screenshots

The three models

- The TAGARELA architecture includes
  - model of domain knowledge (linguistic knowledge)
  - learner model
  - instruction/activity model
- What is the point of learner and activity models?

⇒ Providing feedback involves
  - identifying linguistic properties of the learner input and
  - interpreting them in terms of likely (mis)conceptions of the learner
    - This interpretation goes beyond linguistic form as such.
    - It needs to model the learner's use of language for a specific task in a specific context.
      → (Amaral & Meurers 2007a)

NLP analysis modules in TAGARELA

- Form Analysis:
  - tokenizer: takes into account specifics of Portuguese (cliticization, contractions, abbreviations)
  - lexical/morphological lookup: returns multiple analyses based on CURUPIRA lexicon (Martins et al. 2006)
  - disambiguator: finite state disambiguation rules narrow down lexical information, in the spirit of Constraint Grammar (Karlsson et al. 1995; Bick 2000, 2004)
  - parser: bottom-up chart parser establishes relations to check agreement, case and global well-formedness
- Content Analysis:
  - shallow semantic matching strategies between student answer and target, cf. Content Assessment Module (Bailey & Meurers 2006, 2008)

Annotation-based processing

- Allow the analysis manager to flexibly employ NLP modules relevant to a particular activity.
- To support a flexible control structure, the data structures serving as input and as output for the analysis modules need to be uniform and explicit.
- NLP analysis = a process of enriching the learner input with annotations (parallel to XML-based corpus annotation)
- The same data structure, the learner input annotated with information, is accessed throughout.
  - Closely related idea: Common Analysis System (CAS, Götz & Suhre 2004) of the Unstructured Information Management Architecture (UIMA).
  - UIMA-based reimplementation of TAGARELA's NLP (Ziai 2009)
- In addition to the information obtained by analyzing the input, we need information about the activity.

General Characteristics of Activities

Activities can be characterized and differ in:

- task specification
  - e.g.: listen, read, write, comment, complete
- level
  - e.g.: basic, intermediate, advanced
- expected input
  - e.g.: word, phrase, sentence
- nature and availability of target responses and type of variation from target that is permitted
- required skills and abilities, e.g.:
  - strategies needed (e.g., scanning, summarizing, grouping)
  - amount of content manipulation required
  - required awareness of linguistic categories and rules
- pedagogical goals behind activity and feedback provided:
  - generally: improve the required skills and abilities

Where it matters for processing

- General claim: The NLP analysis and feedback generation depend on the specific activity (type).
- The information from the activity model has an impact on
  - **Property Identification**:
    - Which linguistic properties (incl. errors) of the learner input can actually be observed in a given activity?
  - **Property Selection**: Which of the observed properties to select as likely error cause (or other relevant aspect)?
    - Which of the identified properties is most likely to provide a reliable assessment?
    - Which of the identified errors should be the focus of the feedback given activity and its specific pedagogical goals?
  - **Feedback Strategy**: Which strategy does it choose? E.g.:
    - explicit feedback on form for FIBs
    - scaffolding for reading comprehension (i.e., encouraging the use of required strategies)

Property identification in TAGARELA

- In TAGARELA, different activity types require different linguistic information to analyze student's input:
  - FIB: spell-checking, lexical information
  - Rephrasing: as above + syntactic processing and basic content assessment (correct answer, token matcher)
  - Reading: as above + all content analysis modules
- Why not always run everything?
  - "Don't guess what you know."
  - The more we know about the linguistic properties, the types of variation, and the potential errors NLP needs to detect,
    - the more specific information we can diagnose
    - with higher reliability

TAGARELA meets real life language learners

- The system was used by beginning Portuguese students at The Ohio State University.
- Studying the system logs, we identified two aspects where feedback based on the linguistically correct analysis did not seem to be helpful for learners:
  - interpretation of tokens with accented characters
  - tokenization of compounds

Identifying tokens (I)

Regiões do Brasil

O Brasil está política e geograficamente dividido em cinco regiões. Os limites de cada região (Norte, Nordeste, Sudeste, Sul e Centro-Oeste) coincidem sempre com as fronteiras dos estados que as compõem.

A região Norte ocupa a maior parte do território brasileiro, com uma área que corresponde a 45,27% da área total do País. Formada por sete Estados, tem sua área quase totalmente dominada pela bacia do Rio Amazonas.

A região Nordeste pode ser considerada a mais heterogênea do País. Dividida em quatro grandes zonas - meio-norte, zona da mata, agreste e sertão -, ocupa 18,26% do território nacional e tem nove estados.

O Sudeste é formado por quatro Estados. Esta é a região de maior importância econômica do País, onde está concentrado também o maior índice populacional - 42,63% dos brasileiros.

Já o Sul, região mais fria do País, com ocorrências de geadas e neve, é a que apresenta menor área, ocupando 6,75% do território brasileiro e com apenas três Estados. Os rios que cortam sua área formam a bacia do Paraná em quase toda sua totalidade e são de grande importância para o País, principalmente pelo seu potencial hidrelétrico.

Finalmente, a região Centro-Oeste tem sua área dominada basicamente pelo Planalto Central Brasileiro e pode ser dividida em três porções: maciço goiano-mato-grossense, bacia de sedimentação do Paraná e as depressões. Ela é formada por quatro Estados e nela está a capital do Brasil.

Questões: 1 2 3 4 5 6 7
Próxima Questão (3)

Questão 2

Em que região fica o Rio Amazonas?

O Amazonas fica na região norte.

à á â ã é ê í ó ô õ ú ü ç
À Á Â Ã É Ê Í Ó Ô Õ Ú Ü Ç

Enviar

Análise:

Input: O Amazonas fica na região norte.

Excellent!

Identifying tokens (II)

Regiões do Brasil

O Brasil está política e geograficamente dividido em cinco regiões. Os limites de cada região (Norte, Nordeste, Sudeste, Sul e Centro-Oeste) coincidem sempre com as fronteiras dos estados que as compõem.

A região Norte ocupa a maior parte do território brasileiro, com uma área que corresponde a 45,27% da área total do País. Formada por sete Estados, tem sua área quase totalmente dominada pela bacia do Rio Amazonas.

A região Nordeste pode ser considerada a mais heterogênea do País. Dividida em quatro grandes zonas - meio-norte, zona da mata, agreste e sertão -, ocupa 18,26% do território nacional e tem nove estados.

O Sudeste é formado por quatro Estados. Esta é a região de maior importância econômica do País, onde está concentrado também o maior índice populacional - 42,63% dos brasileiros.

Já o Sul, região mais fria do País, com ocorrências de geadas e neve, é a que apresenta menor área, ocupando 6,75% do território brasileiro e com apenas três Estados. Os rios que cortam sua área formam a bacia do Paraná em quase toda sua totalidade e são de grande importância para o País, principalmente pelo seu potencial hidrelétrico.

Finalmente, a região Centro-Oeste tem sua área dominada basicamente pelo Planalto Central Brasileiro e pode ser dividida em três porções: maciço goiano-mato-grossense, bacia de sedimentação do Paraná e as depressões. Ela é formada por quatro Estados e nela está a capital do Brasil.

Questões: 1 2 3 4 5 6 7

Questão 2 Próxima Questão (3)

Em que região fica o Rio Amazonas?

O Amazonas fica no região norte.

à á ã ã é ê í ó ô õ ú ü ç
À Á Ã Ã É Ê Í Ó Ô Ô Ú Ü Ç

Enviar

Análise:

Input: O Amazonas fica no região norte.

There is an agreement error in **gender** between the **determiner** and the **noun** in the sequence ***o região norte*** from your answer.

To see a possible answer, click here.

Properties of Portuguese

Tokenization

- Certain Portuguese words are syntactically complex.
- Contraction: preposition + determiner/pronoun
  - *no* = *em* (in) + *o* (the)
  - *nela* = *em* (in) + *ela* (it)
  - *destes* = *de* (of) + *estes* (these)
  - *às* = *a* (to) + *as* (the)
- Encliticization:
  - *comprá-lo* = *comprar* (to buy) + *o* (it)
  - *compram-nas* = *compram* (buy) + *as* (them)
  - *comprei-a* = *comprei* (bought) + *a* (it)

TAGARELA
Detmar Meurers & Ramon Ziai
Introduction
Feedback
System Architecture
The three models
Expert model: NLP
Annotation-based setup
Activity model
Relevance for processing
Analyzing learner language
On Tokenization
Interpretation
Portuguese Properties
Mismatches in the identification of tokens
Solution
On interpreting accented characters
Interpretation
Portuguese Properties
Mismatches in the interpretation of tokens
Solution
Wrapping up
Conclusion
Appendix
Screenshots

Mismatches in the identification of tokens

- Learner input: *O Amazonas fica no região norte.*
- System's interpretation: *no* = *em* + *o*
  - tokenized input: [em, o, região, norte]
  - syntactically analyzed: [$_{PP}$ em [$_{NP}$ o$_{masc}$, região$_{fem}$, norte]]
  - ⇒ Agreement error between *o* and *região*.
- Student's interpretation:
  - There is no *o região norte* in the sentence I wrote.
  - I used the 'preposition' *no*.
  - ⇒ So *no* seems to be the wrong preposition?

TAGARELA
Detmar Meurers & Ramon Ziai
Introduction
Feedback
System Architecture
The three models
Expert model: NLP
Annotation-based setup
Activity model
Relevance for processing
Analyzing learner language
On Tokenization
Interpretation
Portuguese Properties
Mismatches in the identification of tokens
Solution
On interpreting accented characters
Interpretation
Portuguese Properties
Mismatches in the interpretation of tokens
Solution
Wrapping up
Conclusion
Appendix
Screenshots

Addressing the Identification of Tokens

- The system needs to connect the surface form provided by the student with the system analysis of this input.
- An annotation-based NLP architecture (→ UIMA) readily supports this with multiple parallel layers of annotation for the learner input.
- The tokenization mismatch can be addressed by representing both surface and deep tokenizations of the learner input, and the mapping between the two.
  - Refer to surface form when generating the feedback.

TAGARELA
Detmar Meurers & Ramon Ziai
Introduction
Feedback
System Architecture
The three models
Expert model: NLP
Annotation-based setup
Activity model
Relevance for processing
Analyzing learner language
On Tokenization
Interpretation
Portuguese Properties
Mismatches in the identification of tokens
Solution
On interpreting accented characters
Interpretation
Portuguese Properties
Mismatches in the interpretation of tokens
Solution
Wrapping up
Conclusion
Appendix
Screenshots

Example Token Representation

```
Token
BEGIN        0
END          2
TOKENSTRING  'no'
CATEGORY     'prep'
LEXDEF       { LexiconInfo
               POS        'prep'
               CANONIC    'no'
               FREQUENCY  31
               SOURCE     'lexicon' }
DEEPFORM     { Token
               BEGIN        0
               END          2
               TOKENSTRING  'em'
               CATEGORY     'prep'
               LEXDEF       { LexiconInfo
                              POS     'prep'
                              SOURCE  'tokn' }
             , Token
               BEGIN        0
               END          2
               TOKENSTRING  'o'
               CATEGORY     'det'
               LEXDEF       { LexiconInfo
                              POS     'det'
                              SOURCE  'tokn'
                              GENDER  'm'
                              NUMBER  's' }
             }
```


Interpreting tokens: Accents (I)

TAGARELA
Detmar Meurers & Ramon Ziai

Introduction
Feedback
System Architecture
The three models
Expert model: NLP
Annotation-based setup
Activity model
Relevance for processing
Analyzing learner language
On Tokenization
Interpretation
Portuguese Properties
Mismatches in the identification of tokens
Solution
On interpreting accented characters
Interpretation
Portuguese Properties
Mismatches in the interpretation of tokens
Solution
Wrapping up
Conclusion
Appendix
Screenshots

Interpreting tokens: Accents (II)

TAGARELA
Detmar Meurers & Ramon Ziai

Introduction
Feedback
System Architecture
The three models
Expert model: NLP
Annotation-based setup
Activity model
Relevance for processing
Analyzing learner language
On Tokenization
Interpretation
Portuguese Properties
Mismatches in the identification of tokens
Solution
On interpreting accented characters
Interpretation
Portuguese Properties
Mismatches in the interpretation of tokens
Solution
Wrapping up
Conclusion
Appendix
Screenshots

Properties of Portuguese

Accents and their importance for lexical distinctions

- Accents in Portuguese encode important linguistic distinctions.
- Part-of-speech differences:
  - pronoun vs. verb
    - *esta* (this) – *está* (is)
  - conjunction vs. verb
    - *e* (and) – *é* (is)
  - verb vs. noun
    - *para* (stop) – *Pará* (state's name)
- Other differences:
  - gender
    - *avô* (grandfather) – *avó* (grandmother)
  - meaning
    - *coco* (coconut) – *cocô* (poop)

TAGARELA
Detmar Meurers & Ramon Ziai

Introduction
Feedback
System Architecture
The three models
Expert model: NLP
Annotation-based setup
Activity model
Relevance for processing
Analyzing learner language
On Tokenization
Interpretation
Portuguese Properties
Mismatches in the identification of tokens
Solution
On interpreting accented characters
Interpretation
Portuguese Properties
Mismatches in the interpretation of tokens
Solution
Wrapping up
Conclusion
Appendix
Screenshots

Mismatches in the interpretation of accents

- Learner Input: *O vaso esta em cima de mesa.*
- System's interpretation:
  - The word *esta* in the learner input is a determiner.
  - There is no form of the verb (*estar*) in the answer.
  - ⇒ The student did not include the main verb.
- Student's interpretation:
  - I included *esta* as a form of the verb *estar*.
    - (The correct spelling is *está*.)
  - There is a verb in the sentence.
  - ⇒ The lack of an accent is a spelling error.

Addressing the Interpretation of Accents

- Learners perceive the unaccented and accented versions of a character as orthographically similar and in consequence confuse linguistically unrelated forms.
- The system needs to capture the confusability of accented with unaccented characters.
  - Treat accented and unaccented characters parallel to common L1-transfer phonological confusions.
    - *está* and *esta* are confused just like
    - *liver* and *river* are by Japanese learners of English

⇒ Develop a module that compares whether different (un)accentuated variants of input words are more likely.
  - Where this is the case, provide dedicated feedback alerting learner of this confusion.

Wrapping up: Token Identification & Interpretation

- Problems for an ITS can arise from mismatches between
  - the system's interpretation of the learner input
  - how the learner perceives and conceptualize the input
- Where such mismatches arise, the feedback produced by the system is inadequate.
- We discussed two such mismatches for Portuguese tokens in TAGARELA:
  - identification of tokens: contraction, encliticization
  - interpretation of tokens: accented characters
- We argued that these problems can be addressed
  - by treating accented and unaccented characters parallel to common L1-transfer phonological confusions.
  - using an annotation-based NLP processing architecture supporting a rich representation of the learner input, including surface and deep tokenizations.

Conclusion

- Integration of computational, linguistic, and FLT/SLA expertise opens up opportunities for ICALL research
- An ITS such as TAGARELA can address specific needs in real-life FLT:
  - provide opportunities for students to practice their listening, reading, and writing skills
  - provide individualized feedback to learner
  - foster learner awareness of language forms and categories
  - provide contextualized activities integrating meaning and form
- The explicit activity design in ITS opens up unique opportunities for the collection of learner language produced in a range of controlled but meaningful activities.
  - Explicit activity design (constraining the potential learner input) makes it possible to include target answers (i.e., a premeditated set of potential target hypotheses!)

References

Amaral, L. (2007). Designing Intelligent Language Tutoring Systems: integrating Natural Language Processing technology into foreign language teaching. Ph.D. thesis, The Ohio State University.

Amaral, L. & D. Meurers (2005). Towards Bridging the Gap between the Needs of Foreign Language Teaching and NLP in ICALL. In A. Pedros-Gascon (ed.), *Proceedings of the 8th Annual Symposium on Hispanic and Luso-Brazilian Literatures, Linguistics, and Cultures*.

Amaral, L. & D. Meurers (2006). Where does ICALL Fit into Foreign Language Teaching? URL http://purl.org/net/icall/handouts/calico06-amaral-meurers.pdf. 23rd Annual Conference of the Computer Assisted Language Instruction Consortium (CALICO), May 19, 2006. University of Hawaii.

Amaral, L. & D. Meurers (2007a). Conceptualizing Student Models for ICALL. In C. Conati & K. F. McCoy (eds.), *User Modeling 2007: Proceedings of the Eleventh International Conference*. Wien, New York, Berlin: Springer, Lecture Notes in Computer Science. URL http://purl.org/dm/papers/amaral-meurers-um07.html.

Amaral, L. & D. Meurers (2007b). Putting activity models in the driver's seat: Towards a demand-driven NLP architecture for ICALL. URL http://www.ling.ohio-state.edu/icall/handouts/eurocall07-amaral-meurers.pdf. EUROCALL. September 7, 2007. University of Ulster, Coleraine Campus.

Amaral, L. & D. Meurers (2008). From Recording Linguistic Competence to Supporting Inferences about Language Acquisition in Context: Extending the Conceptualization of Student Models for Intelligent Computer-Assisted

Language Learning. *Computer-Assisted Language Learning* 21(4), 323–338. URL http://purl.org/dm/papers/amaral-meurers-call08.html.

Amaral, L. & D. Meurers (2009). Little Things With Big Effects: On the Identification and Interpretation of Tokens for Error Diagnosis in ICALL. *CALICO Journal* 27(1).

Bailey, S. & D. Meurers (2006). Exercise-driven selection of content matching methodologies. Peer reviewed conference presentation. EUROCALL'06. September 6, 2006. University of Granada.

Bailey, S. & D. Meurers (2008). Diagnosing meaning errors in short answers to reading comprehension questions. In J. Tetreault, J. Burstein & R. D. Felice (eds.), *Proceedings of the 3rd Workshop on Innovative Use of NLP for Building Educational Applications, held at ACL 2008*. Columbus, Ohio: Association for Computational Linguistics, pp. 107–115. URL http://aclweb.org/anthology-new/W/W08/W08-0913.pdf.

Bick, E. (2000). *The Parsing System "Palavras": Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. Aarhus University Press.

Bick, E. (2004). PaNoLa: Integrating Constraint Grammar and CALL. In H. Holmboe (ed.), *Nordic Language Technology, Arbog for Nordisk Sprogteknologisk Forskningsprogram 2000-2004 (Yearbook 2003)*, Copenhagen: Museum Tusculanum, pp. 183–190.

Ellis, N. (1994). Implicit and Explicit Language Learning - An Overview. In *Implicit and Explicit Learning of Languages*, San Diego, CA: Academic Press, pp. 1–31.

Götz, T. & O. Suhre (2004). Design and implementation of the UIMA Common Analysis System. *IBM Systems Journal* 43(3), 476–489.

TAGARELA

Detmar Meurers & Ramon Ziai

Introduction
Feedback
System Architecture
The three models
Expert model: NLP
Annotation-based setup
Activity model
Relevance for processing
Analyzing learner language
On Tokenization
Interpretation
Portuguese Properties
Mismatches in the identification of tokens
Solution
On interpreting accented characters
Interpretation
Portuguese Properties
Mismatches in the interpretation of tokens
Solution
Wrapping up
Conclusion
Appendix
Screenshots

Eberhard Karls Universität Tübingen

32 / 39

Karlsson, F., A. Voutilainen, J. Heikkilä & A. Anttila (eds.) (1995). *Constraint Grammar: A Language-Independent System for Parsing Unrestricted Text*. No. 4 in Natural Language Processing. Berlin and New York: Mouton de Gruyter.

Lightbown, P. M. & N. Spada (1999). *How languages are learned*. Oxford: Oxford University Press.

Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. D. Bot, C. Kramsch & R. Ginsberg (eds.), *Foreign language research in cross-cultural perspective*, Amsterdam: John Benjamins, pp. 39–52.

Long, M. H. (1996). The role of linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (eds.), *Handbook of second language acquisition*, New York: Academic Press, pp. 413–468.

Lyster, R. (1998). Negotiation of form, recasts, and explicit correction in relation to error types and learner repair in immersion classroom. *Language Learning* 48, 183–218.

Martins, R., R. Hasegawa & M. das Graças Nunes (2006). Curupira: a functional parser for Brazilian Portuguese. In *Computational Processing of the Portuguese Language, 6th International Workshop, PROPOR. Lecture Notes in Computer Science 2721*. Faro, Portugal: Springer. URL http://www.springerlink.com/content/b48vjft1l88yvrj0/fulltext.pdf.

Norris, J. & L. Ortega (2000). Effectiveness of L2 Instruction: A Research Synthesis and Quantitative Meta-Analysis. *Language Learning* 50(3), 417–528.

Schmidt, R. (1995). Consciousness and foreign language: A tutorial on the role of attention and awareness in learning. In R. Schmidt (ed.), *Attention and*

TAGARELA

Detmar Meurers & Ramon Ziai

Introduction
Feedback
System Architecture
The three models
Expert model: NLP
Annotation-based setup
Activity model
Relevance for processing
Analyzing learner language
On Tokenization
Interpretation
Portuguese Properties
Mismatches in the identification of tokens
Solution
On interpreting accented characters
Interpretation
Portuguese Properties
Mismatches in the interpretation of tokens
Solution
Wrapping up
Conclusion
Appendix
Screenshots

Eberhard Karls Universität Tübingen

32 / 39

awareness in foreign language learning, Honolulu: University of Hawaii Press, pp. 1–63.

Schulz, R. A. (2002). Hilft es die Regel zu wissen um sie anzuwenden? Das Verhältnis von metalinguistischem Bewusstsein und grammatischer Kompetenz in DaF. *Die Unterrichtspraxis—Teaching German* 35(1), 15–24. URL http://www.jstor.org/stable/pdfplus/3531951.pdf.

Ziai, R. (2009). A Flexible Annotation-Based Architecture for Intelligent Language Tutoring Systems. Master's thesis, Universität Tübingen, Seminar für Sprachwissenschaft.

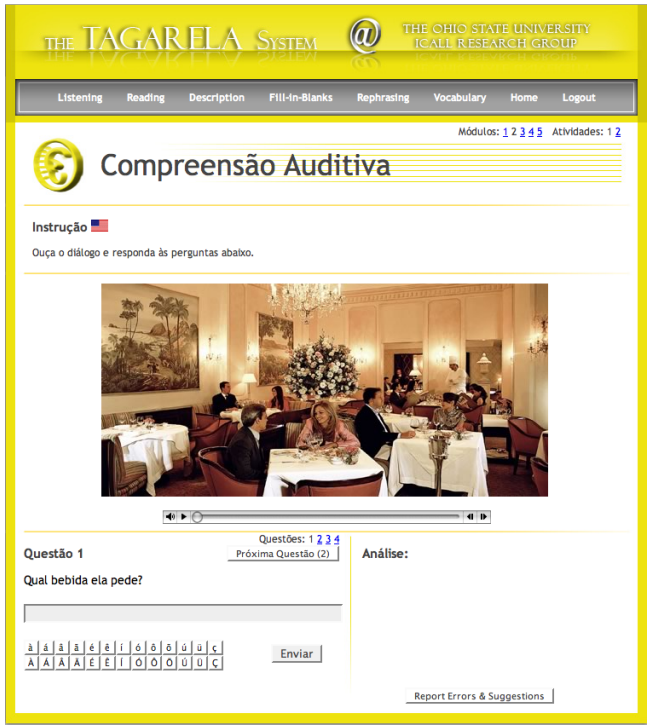
TAGARELA

Detmar Meurers & Ramon Ziai

Introduction
Feedback
System Architecture
The three models
Expert model: NLP
Annotation-based setup
Activity model
Relevance for processing
Analyzing learner language
On Tokenization
Interpretation
Portuguese Properties
Mismatches in the identification of tokens
Solution
On interpreting accented characters
Interpretation
Portuguese Properties
Mismatches in the interpretation of tokens
Solution
Wrapping up
Conclusion
Appendix
Screenshots

Eberhard Karls Universität Tübingen

32 / 39



THE TAGARELA SYSTEM @ THE OHIO STATE UNIVERSITY ICALL RESEARCH GROUP

Listening | Reading | Description | Fill-in-Blanks | Rephrasing | Vocabulary | Home | Logout

Módulos: 1 2 3 4 5 Atividades: 1

Preencha as Lacunas

Instrução

Complete as lacunas com os verbos listados abaixo. Não repita o mesmo verbo mais de uma vez. Conjugue os verbos no pretérito perfeito do indicativo.

Questões: 1 2 3 4 5 6 7

Questão 1 Próxima Questão (2)

comprar - falar - mostrar

Semana passada eu _____ com meu vizinho que queria vender meu carro. Eu _____ o carro pra ele. Hoje de manhã ele _____ o carro.

à á â ã é ê í ó ô õ ú ü ç
À Á Â Ã É Ê Í Ó Ô Õ Ú Ü Ç

Enviar

Análise:

Report Errors & Suggestions

THE TAGARELA SYSTEM @ THE OHIO STATE UNIVERSITY ICALL RESEARCH GROUP

Listening | Reading | Description | Fill-in-Blanks | Rephrasing | Vocabulary | Home | Logout

Módulos: 1 2 3 4 5 Atividades: 1

Reescreva

Instrução

Escreva uma frase comparando os dois elementos apresentados na tabela. Siga o exemplo abaixo.

ensolarada
cozinha 1: sol todo o dia
cozinha 2: sol só pela manhã

Resposta: *A cozinha 1 é mais ensolarada que a cozinha 2.*

Questões: 1 2 3 4 5 6 7 8

Questão 1 Próxima Questão (2)

caro
apartamento 1: R$150.000
apartamento 2: R$230.000

à á â ã é ê í ó ô õ ú ü ç
À Á Â Ã É Ê Í Ó Ô Õ Ú Ü Ç

Enviar

Análise:

Report Errors & Suggestions

TAGARELA

Detmar Meurers & Ramon Ziai

Introduction
- Feedback
- System Architecture
  - The three models
  - Expert model: NLP
  - Annotation-based setup
  - Activity model
  - Relevance for processing

Analyzing learner language
- On Tokenization
  - Interpretation
  - Portuguese Properties
  - Mismatches in the identification of tokens
  - Solution
- On interpreting accented characters
  - Interpretation
  - Portuguese Properties
  - Mismatches in the interpretation of tokens
  - Solution
- Wrapping up

Conclusion

Appendix
- Screenshots

THE TAGARELA SYSTEM @ THE OHIO STATE UNIVERSITY ICALL RESEARCH GROUP

Listening | Reading | Description | Fill-in-Blanks | Rephrasing | Vocabulary | Home | Logout

Módulos: 1 2 3 4 5 Atividades: 1

Vocabulário

Instrução

Observe a figura e complete a descrição com as palavras que estão faltando.

Questões: 1 2 3 4 5 6 7 8 9

Questão 1 Próxima Questão (2)

No banheiro tem _____.

à á â ã é ê í ó ô õ ú ü ç
À Á Â Ã É Ê Í Ó Ô Õ Ú Ü Ç

Enviar

Análise:

Report Errors & Suggestions

TAGARELA

Detmar Meurers & Ramon Ziai

Introduction
- Feedback
- System Architecture
  - The three models
  - Expert model: NLP
  - Annotation-based setup
  - Activity model
  - Relevance for processing

Analyzing learner language
- On Tokenization
  - Interpretation
  - Portuguese Properties
  - Mismatches in the identification of tokens
  - Solution
- On interpreting accented characters
  - Interpretation
  - Portuguese Properties
  - Mismatches in the interpretation of tokens
  - Solution
- Wrapping up

Conclusion

Appendix
- Screenshots