

Little Things With Big Effects: On the identification and interpretation of tokens for error diagnosis in ICALL

Luiz Amaral Detmar Meurers
University of Victoria The Ohio State University

CALICO 2008
Automatic Analysis of Learner Language Workshop
March 19, 2008

On the identification
and interpretation of
tokens in ICALL

Luiz Amaral & Detmar Meurers

Introduction

Research Context

The TAGARELA system
Feedback Types
Feedback Examples
System Architecture
NLP Analysis Module
Annotation-based NLP

Little Things with Big Effects

Accounts
Interpretation
Portuguese Properties
Mismatches in the
interpretation of tokens
Solution
Tabulation
Interpretation
Portuguese Properties
Mismatches in the
identification of tokens
Solution

Conclusion



1 / 23

Introduction

- ▶ ICALL systems often focus on providing automatic feedback on learner input.
 - ▶ Providing feedback involves identifying linguistic properties of the input and interpreting them in terms of likely (mis)conceptions of the learner.
- ▶ For written learner input, this includes orthographic, morphological, syntactic, and semantic properties, all of which require a first step:
 - ▶ the identification and interpretation of the words (tokens).
- ▶ Based on our experience with TAGARELA, we discuss
 - ▶ the identification and interpretation of tokens in the system
 - ▶ how students identify and interpret foreign language tokens
 - ▶ problems arising from mismatches between the two
 - ▶ our solutions to these problems

On the identification
and interpretation of
tokens in ICALL

Luiz Amaral & Detmar Meurers

Introduction

Research Context

The TAGARELA system
Feedback Types
Feedback Examples
System Architecture
NLP Analysis Module
Annotation-based NLP

Little Things with Big Effects

Accounts
Interpretation
Portuguese Properties
Mismatches in the
interpretation of tokens
Solution
Tabulation
Interpretation
Portuguese Properties
Mismatches in the
identification of tokens
Solution

Conclusion



2 / 23

Research Context: The TAGARELA system

- ▶ TAGARELA is an intelligent web-based workbook for learners of Portuguese (Amaral and Meurers 2005, 2006).
- ▶ It addresses real-life needs identified in interviews with OSU foreign language instructors (Amaral 2004, 2007).
 - ▶ It provides opportunities for students to practice their listening, reading, and writing skills.
 - ▶ Focus: on-the-spot feedback for activity types considered useful but typically assigned as homework.
- ▶ It offers six types of activities:
 - ▶ listening comprehension
 - ▶ reading comprehension
 - ▶ picture description
 - ▶ fill-in-the-blank
 - ▶ rephrasing
 - ▶ vocabulary

On the identification
and interpretation of
tokens in ICALL

Luiz Amaral & Detmar Meurers

Introduction

Research Context

The TAGARELA system
Feedback Types
Feedback Examples
System Architecture
NLP Analysis Module
Annotation-based NLP

Little Things with Big Effects

Accounts
Interpretation
Portuguese Properties
Mismatches in the
interpretation of tokens
Solution
Tabulation
Interpretation
Portuguese Properties
Mismatches in the
identification of tokens
Solution

Conclusion



3 / 23

Feedback Types

- ▶ TAGARELA provides on-the-spot feedback on
 - ▶ orthographic errors
 - ▶ non-words
 - ▶ spacing
 - ▶ capitalization
 - ▶ punctuation
 - ▶ semantic errors
 - ▶ missing concepts
 - ▶ extra concepts
 - ▶ word choice
 - ▶ syntactic errors
 - ▶ nominal agreement
 - ▶ verbal agreement

On the identification
and interpretation of
tokens in ICALL

Luiz Amaral & Detmar Meurers

Introduction

Research Context

The TAGARELA system
Feedback Types
Feedback Examples
System Architecture
NLP Analysis Module
Annotation-based NLP

Little Things with Big Effects

Accounts
Interpretation
Portuguese Properties
Mismatches in the
interpretation of tokens
Solution
Tabulation
Interpretation
Portuguese Properties
Mismatches in the
identification of tokens
Solution

Conclusion



4 / 23

Feedback on Spelling Error

Instrução

Ouçã o diálogo e responda às perguntas abaixo.



Questões: 1 2 3 4

Próxima Questão (3)

Questão 2

Onde eles vão?

Eles vão ao cinema.

A A A A A C C I C I O O O O O O C

Enviar

Análise:

Input: Eles vão ao cinema.

I cannot find the word cinema in my dictionary.

Are you sure you don't mean cínema?

On the identification and interpretation of tokens in ICALL

Luz Amaral & Detmar Meurers

Introduction

Research Context

The "MAGRELA" system

Feedback Type

Feedback Examples

System Architecture

NLP Analysis Module

Annotation-based NLP

Little Things with Big Effects

Accents

Integration

Portuguese Properties

Mismatches in the

interpretation of tokens

Solution

Transcription

Integration

Portuguese Properties

Mismatches in the

identification of tokens

Solution

Conclusion



5 / 23

Feedback on Capitalization Error



Vocabulário

Módulos: 1 2 3 4 5 Atividades: 1

Instrução

Faça uma frase completa sobre os objetos que se encontram na sala. Siga o modelo abaixo:



Na sala tem uma mesa de centro.

Questões: 1 2 3 4 5 6 7 8

Próxima Questão (3)

Questão 2



Na sala tem uma Televisão.

Análise:

Input: na sala tem uma televisão.

Your answer is correct, but review it for capitalization errors.

To see a possible answer, click [here](#).

On the identification and interpretation of tokens in ICALL

Luz Amaral & Detmar Meurers

Introduction

Research Context

The "MAGRELA" system

Feedback Type

Feedback Examples

System Architecture

NLP Analysis Module

Annotation-based NLP

Little Things with Big Effects

Accents

Integration

Portuguese Properties

Mismatches in the

interpretation of tokens

Solution

Transcription

Integration

Portuguese Properties

Mismatches in the

identification of tokens

Solution

Conclusion



6 / 23

Feedback on Missing Word



Reescreva

Módulos: 1 2 3 4 5 Atividades: 1

Instrução

Reescreva a frase abaixo usando a expressão entre parênteses.

Questões: 1 2 3 4 5

Próxima Questão (4)

Questão 3

Como você se chama? (nome)

Qual o seu nome?

A A A A A I E E I I O O O O O C
A A A A A E E I I O O O O O C

Enviar

Análise:

Input: Qual o seu nome?

Your answer is close, but there is a verb mising in the sentence.

To see a possible answer, click [here](#).

[Report Errors & Suggestions](#)

On the identification and interpretation of tokens in ICALL

Luz Amaral & Detmar Meurers

Introduction

Research Context

The "MAGRELA" system

Feedback Type

Feedback Examples

System Architecture

NLP Analysis Module

Annotation-based NLP

Little Things with Big Effects

Accents

Integration

Portuguese Properties

Mismatches in the

interpretation of tokens

Solution

Transcription

Integration

Portuguese Properties

Mismatches in the

identification of tokens

Solution

Conclusion



7 / 23

Feedback on Word Choice



Descrição

Módulos: 1 2 3 4 5 Atividades: 1 2

Instrução

Complete a frase sobre a pessoa na foto usando um dos verbos da tabela abaixo.

levantar - morar - ser - ter - dar

Questões: 1 2 3 4 5

Próxima Questão (3)

Questão 2



Ele - em Londres

Ele vive em Londres.

Análise:

Input: Ele vive em Londres.

I am not expecting the verb **viver** for this answer. Try using **morar** instead.

To see a possible answer, click [here](#).

On the identification and interpretation of tokens in ICALL

Luz Amaral & Detmar Meurers

Introduction

Research Context

The "MAGRELA" system

Feedback Type

Feedback Examples

System Architecture

NLP Analysis Module

Annotation-based NLP

Little Things with Big Effects

Accents

Integration

Portuguese Properties

Mismatches in the

interpretation of tokens

Solution

Transcription

Integration

Portuguese Properties

Mismatches in the

identification of tokens

Solution

Conclusion



8 / 23

Feedback on Agreement Error

Módulo: 1 2 3 4 5 ADMAdes: 1 2

Leitura

Instrução

Leia o texto e responda às questões usando frases completas e o vocabulário apresentado no texto.



Quem é você?

Eu me chamo Patricia Mattos, tenho quinze anos e moro em São Paulo. Eu estudo em uma escola pública e tenho muitos amigos.

Eu moro com minha mãe. Seu nome é Marta. Ela tem quarenta anos e é cozinheira em um restaurante de luxo.

Eu tenho um irmão. O nome dele é Claudio. Ele mora nos Estados Unidos e é músico. Ele toca Jazz e Blues. Ele é um excelente guitarrista.

Questões: 1 2 3 4 5 7 9
Próxima Questão (4)

Questão 3

Qual é a sua ocupação?

Ela é estudante.

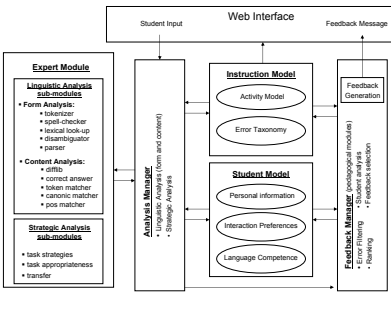
A A A A A E E I O O O O O S C
A A A A A E E I O O O O O S C

Enviar

Report Errors & Suggestions

On the identification and interpretation of tokens in ICALL
Luz Antares & Demner Meurers
Introduction
Research Context
The TAGARELA system
Feedback Types
Feedback Examples
System Architecture
NLP Analysis Module
Annotation-based NLP
Little Things with Big Effects
Accents
Interpretation
Portuguese Properties
Matches in the interpretation of tokens
Solution
Interpretation
Portuguese Properties
Matches in the identification of tokens
Solution
Conclusion
University of Victoria
9 / 23

Overall system architecture of TAGARELA



Linguistic Analysis Modules in TAGARELA

- ▶ **Tokenizer:** takes into account specifics of Portuguese (cliticization, contractions, abbreviations)
 - ▶ **Lexical/Morphological Lookup:** returns multiple analyses based on CURUPIRA lexicon (Martins et al. 2006)
 - ▶ **Finite State Disambiguation:** narrows down lexical information, in the spirit of Constraint Grammar (Karlsson et al. 1995; Bick 2000, 2004)
 - ▶ **Bottom-Up Chart Parser:** establishes relations to check agreement, case and global well-formedness
 - ▶ **Content Assessment** using shallow semantic matching between answer and target (Bailey and Meurers 2006)
- On the identification and interpretation of tokens in ICALL
Luz Antares & Demner Meurers
Introduction
Research Context
The TAGARELA system
Feedback Types
Feedback Examples
System Architecture
NLP Analysis Module
Annotation-based NLP
Little Things with Big Effects
Accents
Interpretation
Portuguese Properties
Matches in the interpretation of tokens
Solution
Interpretation
Portuguese Properties
Matches in the identification of tokens
Solution
Conclusion
University of Victoria
11 / 23

Annotation-based Processing

- ▶ NLP analysis = a process of enriching the learner input with annotations (parallel to corpus annotation).
 - ▶ The same data structure is used throughout:
 - ▶ the learner input annotated with information.
- On the identification and interpretation of tokens in ICALL
Luz Antares & Demner Meurers
Introduction
Research Context
The TAGARELA system
Feedback Types
Feedback Examples
System Architecture
NLP Analysis Module
Annotation-based NLP
Little Things with Big Effects
Accents
Interpretation
Portuguese Properties
Matches in the interpretation of tokens
Solution
Interpretation
Portuguese Properties
Matches in the identification of tokens
Solution
Conclusion
University of Victoria
12 / 23

Interpreting tokens: Accents (I)

Modulos: 1 2 3 4 5 Atividades: 1

Descrição

Instrução 

Descreva a foto usando as palavras apresentadas no exercício e uma das preposições abaixo.

em cima de - entre - embaixo de - ao lado de

Questões: 1 2 3 4
Próxima Questão (2)

Questão 1



Análise:

Input: O vaso está em cima da mesa.
Very Good! Keep going!

vaso - mesa

O vaso está em cima da mesa.

A A A A E E I I O O O O U U C C

Enviar

Report Errors & Suggestions

On the identification and interpretation of tokens in ICALL

Luiz Amaral & Demian Meurers

Introduction

Research Context
The "ICALL" system
Feedback Types
Feedback Examples
System Architecture
NLP Analysis Module
Annotation-based NLP

Little Things with Big Effects

Accents

Interpretation
Portuguese Properties
Matches in the interpretation of tokens
Solution
Transcription
Interpretation
Portuguese Properties
Matches in the identification of tokens
Solution

Conclusion

 OHIO STATE UNIVERSITY
University of Victoria 13 / 23

Interpreting tokens: Accents (II)

Modulos: 1 2 3 4 5 Atividades: 1

Descrição

Instrução 

Descreva a foto usando as palavras apresentadas no exercício e uma das preposições abaixo.

em cima de - entre - embaixo de - ao lado de

Questões: 1 2 3 4
Próxima Questão (2)

Questão 1



Análise:

Input: O vaso está em cima da mesa.
There is an important verb missing in your sentence.
Also review it for unnecessary words.
To see a possible answer, click [here](#).

vaso - mesa

O vaso está em cima da mesa.

A A A A E E I I O O O O U U C C

Enviar

Report Errors & Suggestions

On the identification and interpretation of tokens in ICALL

Luiz Amaral & Demian Meurers

Introduction

Research Context
The "ICALL" system
Feedback Types
Feedback Examples
System Architecture
NLP Analysis Module
Annotation-based NLP

Little Things with Big Effects

Accents

Interpretation
Portuguese Properties
Matches in the interpretation of tokens
Solution
Transcription
Interpretation
Portuguese Properties
Matches in the identification of tokens
Solution

Conclusion

 OHIO STATE UNIVERSITY
University of Victoria 14 / 23

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
 - ▶ *avó* (grandfather) – *avó* (grandmother)
 - meaning
 - ▶ *coco* (coconut) – *côco* (poop)

On the identification and interpretation of tokens in ICALL

Luiz Amaral & Demian Meurers

Introduction

Research Context
The "ICALL" system
Feedback Types
Feedback Examples
System Architecture
NLP Analysis Module
Annotation-based NLP

Little Things with Big Effects

Accents

Interpretation
Portuguese Properties
Matches in the interpretation of tokens
Solution
Transcription
Interpretation
Portuguese Properties
Matches in the identification of tokens
Solution

Conclusion

 OHIO STATE UNIVERSITY
University of Victoria 15 / 23

Mismatches in the interpretation of accents

- ▶ Learner Input: *O vaso está em cima de mesa.*
- ▶ System's interpretation:
 - The word *está* 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.

On the identification and interpretation of tokens in ICALL

Luiz Amaral & Demian Meurers

Introduction

Research Context
The "ICALL" system
Feedback Types
Feedback Examples
System Architecture
NLP Analysis Module
Annotation-based NLP

Little Things with Big Effects

Accents

Interpretation
Portuguese Properties
Matches in the interpretation of tokens
Solution
Transcription
Interpretation
Portuguese Properties
Matches in the identification of tokens
Solution

Conclusion

 OHIO STATE UNIVERSITY
University of Victoria 16 / 23

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 (un)accentuated variants of input words are more likely.

- ▶ Where this is the case, provide dedicated feedback alerting learner of this confusion.

On the identification and interpretation of tokens in ICALL

Luiz Amaral & Detmar Meurers

Introduction

Research Context

The *INACALL* system
Feedback Types
Feedback Examples
System Architecture
NLP Analysis Module
Annotation-based NLP

Little Things with Big Effects

Accents
Interpretation
Portuguese Properties
Mismatches in the interpretation of tokens

Solution

Tokenization
Interpretation
Portuguese Properties
Mismatches in the identification of tokens
Solution

Conclusion



17/23

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 6,79% 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.



Análise:

Input: O Amazonas fica na região norte.

Excellent!

On the identification and interpretation of tokens in ICALL

Luiz Amaral & Detmar Meurers

Introduction

Research Context
The *INACALL* system
Feedback Types
Feedback Examples
System Architecture
NLP Analysis Module
Annotation-based NLP

Little Things with Big Effects

Accents
Interpretation
Portuguese Properties
Mismatches in the interpretation of tokens

Solution

Tokenization
Interpretation
Portuguese Properties
Mismatches in the identification of tokens
Solution

Conclusion



18/23

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 6,79% 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.

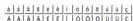
Questões: 1 2 3 4 5 6 7

Próxima Questão (3)

Questão 2

Em que região fica o Rio Amazonas?

O Amazonas fica no região norte.



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](#).

On the identification and interpretation of tokens in ICALL

Luiz Amaral & Detmar Meurers

Introduction

Research Context

The *INACALL* system
Feedback Types
Feedback Examples
System Architecture
NLP Analysis Module
Annotation-based NLP

Little Things with Big Effects

Accents
Interpretation
Portuguese Properties
Mismatches in the interpretation of tokens

Solution

Tokenization
Interpretation
Portuguese Properties
Mismatches in the identification of tokens
Solution

Conclusion



19/23

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)*

On the identification and interpretation of tokens in ICALL

Luiz Amaral & Detmar Meurers

Introduction

Research Context
The *INACALL* system
Feedback Types
Feedback Examples
System Architecture
NLP Analysis Module
Annotation-based NLP

Little Things with Big Effects

Accents
Interpretation
Portuguese Properties
Mismatches in the interpretation of tokens

Solution

Tokenization
Interpretation
Portuguese Properties
Mismatches in the identification of tokens
Solution

Conclusion



20/23

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?

On the identification and interpretation of tokens in ICALL

Luiz Amaral & Detmar Meurers

Introduction

Research Context

The TAGARELA system
Feedback Types
Feedback Examples
System Architecture
NLP Analysis Module
Annotation-based NLP

Little Things with Big Effects

Accents

Interpretation

Portuguese Properties

Mismatches in the interpretation of tokens

Solution

Transcription

Interpretation

Portuguese Properties

Mismatches in the identification of tokens

Solution

Conclusion

University of Victoria

21 / 23

OHIO STATE UNIVERSITY

University of Victoria

21 / 23

OHIO STATE UNIVERSITY

University of Victoria

21 / 23

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 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.

On the identification and interpretation of tokens in ICALL

Luiz Amaral & Detmar Meurers

Introduction

Research Context

The TAGARELA system
Feedback Types
Feedback Examples
System Architecture
NLP Analysis Module
Annotation-based NLP

Little Things with Big Effects

Accents

Interpretation

Portuguese Properties

Mismatches in the interpretation of tokens

Solution

Transcription

Interpretation

Portuguese Properties

Mismatches in the identification of tokens

Solution

Conclusion

University of Victoria

22 / 23

OHIO STATE UNIVERSITY

University of Victoria

22 / 23

OHIO STATE UNIVERSITY

University of Victoria

22 / 23

Conclusion

- ▶ In an ICALL system, problems can arise from mismatches between:
 - ▶ the identification and interpretation of the learner input by the system
 - ▶ 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:
 - ▶ interpretation of tokens: accented characters
 - ▶ identification of tokens: contraction, encliticization
- ▶ 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.

On the identification and interpretation of tokens in ICALL

Luiz Amaral & Detmar Meurers

Introduction

Research Context

The TAGARELA system
Feedback Types
Feedback Examples
System Architecture
NLP Analysis Module
Annotation-based NLP

Little Things with Big Effects

Accents

Interpretation

Portuguese Properties

Mismatches in the interpretation of tokens

Solution

Transcription

Interpretation

Portuguese Properties

Mismatches in the identification of tokens

Solution

Conclusion

University of Victoria

23 / 23

OHIO STATE UNIVERSITY

University of Victoria

23 / 23

OHIO STATE UNIVERSITY

University of Victoria

23 / 23

OHIO STATE UNIVERSITY

References

- Amaral, Luiz (2007). Designing Intelligent Language Tutoring Systems: integrating Natural Language Processing technology into foreign language teaching. Ph.D. thesis, The Ohio State University.
- Amaral, Luiz and Detmar Meurers (2006). Where does ICALL Fit into Foreign Language Teaching? CALICO Conference. May 19, 2006. University of Hawaii.
- Amaral, Luiz Alexandre (2004). ICALL Specifications. Ms. Department of Linguistics & Department of Spanish and Portuguese. Ohio State University.
- Amaral, Luiz Alexandre and Walt Detmar Meurers (2005). Towards Bridging the Gap between the Needs of Foreign Language Teaching and NLP in ICALL. In Ant6nio Pedros-Gascon (ed.), *Electronic Proceedings of the 8th Annual Symposium on Hispanic and Luso-Brazilian Literatures, Linguistics, and Cultures*.
- Bailey, Stacey and Detmar Meurers (2006). Exercise-driven selection of content matching methodologies. EUROCALL Conference. September 6, 2006. University of Granada.
- Bick, Eckhard (2000). *The Parsing System "Palavras": Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. Aarhus University Press.
- Bick, Eckhard (2004). PaNoLa: Integrating Constraint Grammar and CALL. In Henrik Holmboe (ed.), *Nordic Language Technology, Arborg for Nordisk Sprogteknologisk Forskningsprogram 2000-2004 (Yearbook 2003)*, Copenhagen: Museum Tusulanum, pp. 183-190.

On the identification and interpretation of tokens in ICALL

Luiz Amaral & Detmar Meurers

Introduction

Research Context

The TAGARELA system
Feedback Types
Feedback Examples
System Architecture
NLP Analysis Module
Annotation-based NLP

Little Things with Big Effects

Accents

Interpretation

Portuguese Properties

Mismatches in the interpretation of tokens

Solution

Transcription

Interpretation

Portuguese Properties

Mismatches in the identification of tokens

Solution

Conclusion

University of Victoria

23 / 23

OHIO STATE UNIVERSITY

University of Victoria

23 / 23

OHIO STATE UNIVERSITY

University of Victoria

23 / 23

OHIO STATE UNIVERSITY

Karlsson, Fred, Atro Voutilainen, Juha Heikkilä and Arto 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.

Martins, Ronaldo, Ricardo Hasegawa and Maria 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.

On the identification and interpretation of tokens in ICALL

Luiz Amaral & Detmar Meurers

Introduction

Research Context

The ICALL system

Feedback Types

Feedback Examples

System Architecture

NLP Analysis Module

Annotation-based NLP

Little Things with Big Effects

Accounts

Interpretation

Portuguese Properties

Mismatches in the

interpretation of tokens

Solution

Identification

Interpretation

Portuguese Properties

Mismatches in the

identification of tokens

Solution

Conclusion

