	Tapping into the synergy between SLA, FLT and NLP in ICALL	Introduction	Tapping into the synergy between SLA, FLT and NLP in ICALL
Tapping into the synergy between SLA, Foreign Language Instruction and Natural Language Processing in ICALL	Introduction Instruction FLT and ICALL VIGARELA Activity & feeback types Exercise pages VIGARELA & General Picture Example feedback pages	 The idea of this talk is to make explicit where ICALL research requires interaction between FLT practice SLA research NLP research and development. 	Introduction Instruction FLT and ICALL YEARSELA Activity & Rectoack types Exercise pages YEARSELA & General Picture Example Rectack pages
Detmar Meurers Luiz Amaral Universität Tübingen UMass Amherst	Modeling System Architecture The three models Interpreting learner language Mematiches in interpretation From diagnosting errors to over-funderates Enhanced learner corpora Summarry	 Following the workshop topic, we focus on instruction: relating ICALL and FLT practice modeling: motivating the need for explicit activity models and learner models 	Modeling System Architecture The three models Interpreting learner language Mismatches in interpretation From diagnosting errors to over-fundament Enhanced learner corpora Summarry
Symposium on Instruction and Modeling in ICALL AlLA: 15th World Congress of Applied Linguistics August 25, 2008	Next steps References	 Background: Development of the TAGARELA system, an intelligent web-based workbook for learners of Portuguese – and our current research directions there. 	Next steps References
	UNITED STATE TO THE PROPERTY TH		UNIVERSITÄT THEISOERY AMHERST 2/31
Instruction and ICALL	Tapping into the synergy between SLA, FLT and NLP in ICALL	Instruction and TAGARELA	Tapping into the synergy between SLA, FLT and NLP in ICALL
► What is the connection between ICALL research and real-life instruction, i.e., Foreign Language Teaching? ► Why should FLT practitioners and ICALL researchers care about establishing such a connection? ➡ FLT perspective: Find new ways to address currently unmet pedagogical needs (to improve learning outcomes) ➡ ICALL perspective: Obtain real-life data to ► make learner input analyzable: training of NLP tools ► make learner input interpretable: disambiguation using explicit learner and activity models ► evaluate algorithms and test systems	Introduction Instruction Instr	We conducted interviews with OSU foreign language instructors (Amaral 2004, 2007), who reported: Students often fail to achieve the communicative goal of a class activity if they cannot use the appropriate language forms and patterns. But class activities that focus on grammar patterns are perceived as problematic (reduce pace of lesson). Individual differences make it impossible to have all students do the same tasks in exactly the same time — a general problem for some activities, e.g., listening. There is a real-life FLT need for tools which provide opportunities for students to practice their listening, reading, and writing skills provide individual feedback on learner input to system	Introduction Instruction Tri and CALL Tri Tri and CALL Tri Tri and CALL Tri
	UMASS AMHERST 3/31	 foster learner awareness of language forms and categories 	UMASS AMHERST 4/31

Instruction and TAGARFI A

- ► TAGARELA currently offers six types of activities:
 - listening comprehension

Current activity and feedback types

- reading comprehension
- picture description
- ▶ fill-in-the-hlank
- rephrasing
- vocabulary Similar to traditional workbook exercises, plus audio.
- ► The system provides on-the-spot feedback on

 - · orthographic, syntactic, and
 - semantic errors.

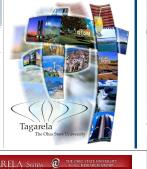


THE TAGARELA SYSTEM

Tapping into the

synergy between SLA, FLT and NLP

in ICALL



Report Errors & Suggestions



Tapping into the

synergy between SLA. FLT and NLP

in ICALL

Introduction





synergy between









Tapping into the

Modeling

Summary

synergy between

SLA, FLT and NLP

Activity & feedback types

Example feedback pages

Interpreting learner language

Mismatches in interpretation

From diagnosing errors to

Enhanced learner corpora

in ICALL

Introduction

Instruction

TAGARELA

Modeling

Summary

Next steps

References

System Architecture







Tapping into the

synergy between

SLA FIT and NIP

Activity & feedback types

Example feedback pages

Interpreting learner language

Mismatches in interpretatio

From diagnosing errors to

in ICALL

Introduction

FLT and ICALI

TAGARELA

Modeling

Summary

Next steps

References

Tapping into the

synergy between SLA. FLT and NLP

Instruction and TAGARFI A Tapping into the Instruction and TAGARFI A Tapping into the synergy between SLA, FLT and NLP synergy between SLA, FLT and NLP General issues, conclusions so far, avenues for research in ICALL General questions, conclusions so far, avenues for research in ICALL What should the system interfaces look like? Instruction Which forms of feedback are (most) successful in → Use L2 as far as possible (needs careful interface design). fostering awareness of forms/categories - and. What type of activities are appropriate and useful for ultimately, in influencing learning outcomes? fostering awareness (and fit into the FLT approach)? Modelina Modelina → Meta-linguistic feedback, highlighting (cf. Heift 2004) → Activities ideally involve both form and meaning, such more research needed into range of feedback types as listening/reading comprehension questions. what is appropriate for human-computer interaction/CMC From diagnosing errors to investigate activity types satisfying FLT needs and (cf., e.g., Sachs & Suh 2007; Petersen in preparation) providing sufficient constraints for NLP analysis and Enhanced learner corpora Enhanced learner corpora including evaluation using Summary Summary interpretation (activity models) Next stens learning outcomes Next steps online measures of noticing, e.g., using eve tracking. References Deferences What role does an ICALL system play in teaching? since no learning without noticing (Schmidt 1995) → Self-quided activities accompanying teaching investigate more guidance by system (learner models) Tapping into the Instruction and TAGARELA THE TAGARELA SYSTEM synergy between synergy between SLA FIT and NI P SLA FIT and NIP General questions, conclusions so far, avenues for research in ICALL in ICALL Módulos: 1 2 3 4 5 Atividades: 1 2 Instruction Leitura What can/should feedback be provided on? FLT and ICALL FLT and ICALL TAGARELA provides on-the-spot feedback on Activity & feedback types Activity & feedback type Exercise pages orthographic errors (non-words, spacing, capitalization, punctuation) Lela o texto e responda às questões usando frases completas e o vocabulário apresentado no texto Modelina Modelina syntactic errors (nominal and verbal agreement) Quem é você? semantic errors (missing or extra concepts, word choice) Eu me chamo Patricia Mattos, tenho quinze anos e moro em São Paulo. Eu estudo em uma escola pública e tenho muitos amigos Providing feedback on meaning becomes crucial for From diagnosing errors to activities such as reading and listening comprehension. Eu moro com minha mãe. Seu nome é Marta. Ela tem quarenta anos e é cozinheira em um Enhanced learner corpora → automatic meaning comparison can be effective Summary Summary Eu tenho um irmão. O nome dele é Claudio. Ele mora nos Estados Unidos e é músico. Ele Next stens toca Jazz e Blues. Ele é um excelente guitarrista. Next steps (Bailey & Meurers 2008) more research needed on when and how this is the References References Questão 2 Análisas Próxima Questão (3) case and what constitutes good diagnosis categories Ouantos anos ela tem Input: Ela tens guinze anos. Ela tens quinze anos the subject and the verb in the sequence ela tens from your answer. 111111111111 To see a possible answer, click here, Enviar

Tapping into the Feedback on Agreement THE TAGARELA SYSTEM @ Tapping into the synergy between SLA, FLT and NLP synergy between SLA. FLT and NLP in ICALL in ICALL Detmar Meurers & Luiz Amaral Listening Reading Description Fill-in-Blanks Rephrasing Vocabulary Home Logor Introduction Introduction Módulos: 1 2 3 4 5 Atividades: 1 2 Instruction Instruction stões: 1 2 3 4 5 6 7 8 FLT and ICALL Leitura óxima Questão (3) Análise: Activity & feedback types Exercise pages AGARELA & General Pict GARELA & General Pict Input: Ela tens guinze anos. Instrução 🎫 Leía o texto e responda ás questões usando frases completas e o vocabulário apresentado no texto. Modeling Modeling There is an agreement error in person between the subject and the verb in the sequence ela System Architecture Quem é você? tens from your answer. Interpreting learner language Eu me chamo Patricia Mattos, tenho quinze anos e moro em São Paulo, Eu estudo em uma Mismatches in interpretation escola pública e tenho muitos amigos. To see a possible answer, click here. From diagnosing errors to Enviar Enhanced learner corpora Fu moro com minha mãe. Seu nome é Marta. Ela tem quarenta anos e é cozinheira em um Enhanced learner corpora restaurante de luxo. Summary Summary Eu tenho um irmão. O nome dele é Claudio. Ele mora nos Estados Unidos e é músico. Ele Report Errors & Suggestions Next steps toca Jazz e Blues. Ele é um excelente guitarrista. Next steps References References Questão 2 Análise: Próxima Questão (3) Quantos anos ela tem? Input: Ela és guinze anos. I am not expecting the verb ser for this answer. Try using ter instead. Universität To see a possible answer, click berg. Enviar Tapping into the Tapping into the Feedback on Word Choice THE TAGARELA SYSTEM @ synergy between synergy between SLA FIT and NI P SLA FIT and NIP in ICALL in ICALL ing Reading Description Fill-in-Blanks Rephrasing Vocabulary Home Logout Introduction Módulos: 12345 Atividades: 12345 Instruction stões: 1 2 3 4 5 6 7 8 FLT and ICALL FLT and ICALL Descrição óxima Questão (3) Análise: TAGARELA TAGARELA Activity & feedback types Activity & feedback types Exercise pages Input: Ela és quinze anos. Instrução 🎫 Modeling Modeling I am not expecting the verb ser for this answer. Excreva uma frase completa usando a informação apresentada pela gravura e as palavras entre parênteses Try using ter instead. System Architecture Questões: 1234 Análise: interpreting learner language Questão 1 Próxima Questão (2) Interpreting learner language To see a possible answer, click here, Mismatches in interpretation Mismatches in interpretatio (nout: Elas se chamam Ana e Maria. From diagnosing errors to chamar-se - Ana e Beatriz From diagnosing errors to Enviar I think there is a problem with the proper noun Enhanced learner corpora Summary Are you sure you want to use Maria instead of Summary Report Errors & Suggestions Next steps Next steps To see a possible answer, click bere References Elas se chamam Ana e Maria. Enviar Report Errors & Suggestions

Tapping into the Feedback on Wrong Word Tapping into the synergy between SLA, FLT and NLP synergy between SLA. FLT and NLP THE TAGARELA SYSTEM @ in ICALL in ICALL Detmar Meurers & Luiz Amaral Introduction Introduction Instruction Instruction Questões: 1 2 3 4 FLT and ICALL Módulos: 1 2 3 4 5 Athyldades: 1 ma Questão (2) Análise: Activity & feedback types Reescreva Exercise pages ARELA & General Pict GARELA & General Pict Input: Elas se chamam Ana e Maria. e Beatriz Modeling Modeling I think there is a problem with the proper noun Instrução === System Architecture you have chosen. Reescreva a frase ababo usando a expressão entre parênteses, Interpreting learner language Interpreting learner language Are you sure you want to use Maria instead of Questões: 12345 From diagnosing errors to Ouestão 3 Análise: Beatriz? Próxima Questão (4) Enhanced learner corpora Enhanced learner corpora Como você se chama? (nome) Input: Qual o seu nome? To see a possible answer, click here. Summary Summary Your answer is close, but there is a verb missing in Qual o seu nome? your sentence. Next stens Next steps To see a possible answer, click bere Enviar Report Errors & Suggestions Tapping into the Tapping into the Feedback on Missing Verb General Architecture of TAGARELA synergy between synergy between SLA FIT and NI P SLA, FLT and NLP in ICALL in ICALL Introduction Web Interface Student Input Feedback Message Instruction FLT and ICALL FLT and ICALL Questões: 1 2 3 4 5 TAGARELA TAGARELA Análise: róxima Questão (4) Activity & feedback types Expert Module Activity & feedback types Exercise pages Instruction Model Exercise pages TAGARELA & General Pictur Input: Qual o seu nome? Linguistic Analysis Feedback sub-modules Example feedback pages Activity Model Generation Modeling Form Analysis: Modelina Your answer is close, but there is a verb missing in System Architecture spell-checker your sentence. levical look-un Error Tayonomy interpreting learner language disambiguator To see a possible answer, click here. • parser Mismatches in interpretation Mismatches in interpretatio From diagnosing errors to From diagnosing errors to · Content Analysis: Enviar · dmb Student Model Manage Analysis (* malysis Enhanced learner corpora correct answer token matcher Summary Summary · canonic matches Personal information · pos matcher Next steps Next steps Report Errors & Suggestions Analysis References Interaction Preferences References Stratenic Analysis · Error Filberi · Banking task strategies Language Competeno task appropriateness UNIVERSITÄT • transfer

The three models	Tapping into the synergy between SLA, FLT and NLP in ICALL	Interpreting learner language: an example	Tapping into the synergy between SLA, FLT and NLP in ICALL
The TAGARELA architecture includes model of domain knowledge (linguistic knowledge) student 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 2007).	Demon Mexico & List Answell Introduction Instituction Ins	(1) He sat with looking at the liver. (From HELC, Miura 1998) Doctors reporting about work in operating room? Cooks discussing in kitchen? Chinese learner of English translating sentence on rivers? activity and learner information needed to interpret data The learner model relevant for analysis even for interpreting apparently well-formed expressions. Activities in current ICALL systems are tightly controlled, e.g., build-a-sentence or translation, so that processing can implicitly rely on knowing which words should occur. For a wider range of meaning-based activities with more variation in learner input, research into constructing and using explicit learner and activity models is needed.	Centro Manura & List Answell Interdoduction Instituction
Mismatches in the interpretation of learner input Learner Input: O vaso esta em cima de mesa. TAGARELA'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: Included esta as a form of the verb estar. (The correct spelling is esta). There is a verb in the sentence. The lack of an accent is a spelling error. Linguistic analysis ≠ student intention. So what can the analysis rely on? Learner model needed for analysis/interpretation Information from the learner model should be taken into account by the NLP (cf., ICICLE, Michaud & McCoy 2004). more research needed on how learner models can define the space to be checked by the NLP	Tapping into the synergy between SAL Park Park Park Park Park Park Park Park	From diagnosing errors to over/-underuse Phrasal verbs in EFL English learners make errors in particle placement: (2) a. so they give up it b. food which will build up him Examples from the Chinese Learner English Corpus (CLEC 2004) Learners also avoid using phrasal verbs: Liao & Fukuya (2002) show that Chinese learners of English avoid phrasal verbs; similar research for other L1. We confirmed patterns of avoidance in the CLEC (2004): heavy use of pattern that is always grammatical (vorb + particle - full IPP) ittle use of patterns restricted to certain verb & object types Diagnosis of over- and underuse of language patterns only possible based on production history for a learner analyzing a corpus (cf., e.g., Borin & Prütz 2003) In ICALL system: interpreting extended learner model	Tripping into the system, believed the state of the system, believed to state of the system, believed to system, belie

Using ICALL to enhance learner corpora	Tapping into the synergy between SLA, FLT and NLP in ICALL	Summary	Tapping into the synergy between SLA, FLT and NLP in ICALL
Nost learner language corpora consist of essays. Yet learners produce language in a wide range of environments, and errors often are correlated with specific language tasks performed by the learner, e.g., answering reading or listening comprehension questions asking questions in information gap activities specific strategies needed to complete specific tasks, e.g., skimming, scanning, To obtain corpora representative of learner language, they should include language produced in a variety of tasks. CALL systems provide an ideal setup for collecting task-specific learner language. support explicit meta-information on learner and task particularly helpful for annotating learner corpora easier to infer what learner wanted to say if one knows the text they are answering questions about	Christon Manuma & Lisz Ansonal Introduction Instituction	 A strong connection between FLT, ICALL, and SLA is mutually beneficial, to obtain better tools for FLT more varied learner corpora and a valuable experimental testbed for SLA research practically relevant ICALL systems, including a wide range of activities and effective NLP ICALL systems need more extensive learner models to interpret the language as intended by the learner make processing feasible by constraining it to the relevant interlanguage detect patterns of over-/underuse, an important feature of interlanguage as studied in SLA. 	Commo Marcon & Lazi Annual Introduction Introduction Introduction Institution
Our next steps Fully modularize interface (apache2), activity specifications (xml), and NLP (UIMA) to support exploring range of activity types and feedback strategies NLP integrating learner and activity model information reuse for languages other than Portuguese (IU Bloomington: Russian, Hebrew, UMass: Spanish) Extend learner model to take into account specification provided in Amaral & Meurers (2009). Integrate content assessment approach of Bailey & Meurers (2008) and extend it: Which type of answers to what type of question given which context support reliable content assessment? What are reliable and useful diagnosis categories? Extend logging facilities so that the system can produce XML-based learner corpora from the log, including meta information on activity and learner model. Use the nicely controlled environment for experimental SLA research (e.g., addressing language transfer, models of second language acquisition).	Tagging into the systems, below the systems, belowed the systems, belowed the systems, below the systems, below the systems of	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 (2007). Conceptualizing Student Models for ICALL. In C. Conal & K.F. McCoy (eds.), Duer Modeling 2007: Proceedings of the Eleventh international Conference. Wien, New York, Berlin: Springer, Lecture Notes in Computer Science. URL http://purl.org/cim/papers/amaral-meures-um07. html. Amaral, L & D. Meurers (2009). From Recording Linguistic Competence to Supporting Inferences about Language Acquisition in Context: Extending the Conceptualization of Student Models for Intelligent Computer-Assisted Language Learning URL http://purl.org/cim/papers/amaral-meures-called). html. Amaral, L A (2004). ROAL Specifications. Mb. Department of Linguistics & Department of Spanish and Portuguese. Online of the International Content of Spanish and Portuguese. Online of the International Content of Spanish and Portuguese. Online of the International Content of Spanish and Workshop on Inconview Use of MP for Building Educational Applications. held at ACL 2006. Columbus, Ohio: Association for Computational Linguistics, pp. 107-115. URL http://achveb.org/anthology-new/WW08W08-0913.pdf. 18 Jennis L. P. Christ (2003). Beyon wine in old shirts? A corpus investigation of L1 syntactic transfer in learner language. In Tacching and language corpora (Tal.C.) 2002. Amsterdam: Rodopi.	Transporting to the system, planes on the system of the syst

Heift, T. (2004). Corrective Feedback and Learner Uptake in CALL. ReCALL 16(2). 416-431 Liao, Y. D. & Y. J. Fukuya (2002). Avoidance of phrasal verbs: The case of Chinese

learners of English. Second Language Studies 20(2), 71-106.

Michaud, L. N. & K. F. McCov (2004). Empirical Derivation of a Sequence of User Stereotypes for Language Learning. User Modeling and User-Adapted

Interaction 14, 317-350. Miura, S. (1998). Hiroshima English Learners' Corpus: English learner No. 2

CLEC (2004). Chinese Learner English Corpus. Web interface to Corpus.

(English I & English II). URL http://home.hiroshima-u.ac.jp/d052121/eigo2.html. Last Modified 14 May,

1998.

Petersen, K. (in preparation). Implicit Corrective Feedback in Computer-Guied Interaction: Does Mode Matter? Ph.D. thesis, Georgetown University. Sachs, R. & B.-R. Suh (2007). Textually enhanced recasts, learner awareness, and

L2 outcomes in synchronous computer-mediated interaction. In Oxford applied linguistics. Oxford: Oxford University Press. Schmidt, R. (1995). Consciousness and foreign language: A tutorial on the role of attention and awareness in learning. In R. Schmidt (ed.), Attention and awareness in foreign language learning, Honolulu; University of Hawaii Press, pp. 1-63.

Tapping into the synergy between SLA, FLT and NLP in ICALL

Detriar Meurens & Luiz Amaral

Introduction

Instruction

FLT and ICALL

Activity & feedback types

Exercise pages TAGARELA & General Pictur

Example feedback pages

Modeling

System Architecture The three models Interpreting learner language Mismatches in interpretation

From diagnosing errors to Enhanced learner corpora

Summary Next steps

References



