

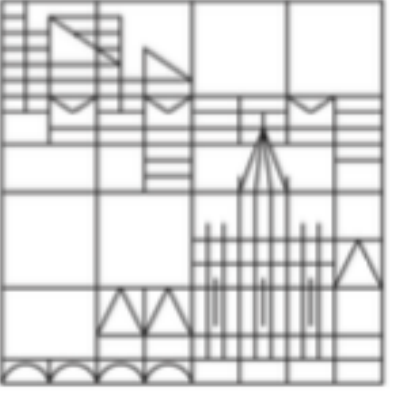
A Deep Linguistic Computer-Assisted Language Learning System for Italian

Jessica Zipf

University of Konstanz

jessica.zipf@uni-konstanz.de

Universität
Konstanz



OBJECTIVE

- Computer-Assisted Language Learning (CALL) tool for Italian
- Deep linguistic parser in tool's back-end
- Aimed at German L2 learners of Italian

INTRODUCTION

- Based on architecture proposed by Khader, Butt & King (2004)
- System parses ungrammatical input & provides user with grammatical alternative
- Crucial components:
 - ▶ Concepts from Optimality Theory (OT) (Frank et al. 1998)
 - ▶ Generation component of the Xerox Linguistic Environment (XLE)
 - ▶ A feedback system that provides user with linguistic explanations on certain phenomena (Seiss 2013)
- Frequent errors among German L2 learners of Italian (Leonini & Belletti 2003)
 - ▶ SV instead of VS order (1)
 - ▶ Clitic production (2)
 - additionally, both phenomena interact

EXAMPLES

- (1) a. (CONTEXT)
Chi è arrivato?
who AUX.3SG arrive.PST.PTCP
'Who has arrived?'
- b. (EXPECTED)
È arrivato Francesco.
AUX.3SG. arrive.PST.PTCP Francesco.
'Francesco has arrived.'
- c. (COMMON ERROR)
*Francesco è arrivato.
Francesco AUX.3SG arrive.PST.PTCP
'Francesco has arrived.'

• Subject realized postverbally to express new information

- (2) a. (CONTEXT)
Chi ha portato i fiori?
who AUX.3SG bring.PST.PTCP the flowers
'Who brought the flowers?'
- b. (EXPECTED - OBJECT CLITIC)
Li ha portati Silvia.
CL.3PL.M AUX.3SG bring.PST.PTCP Silvia
'Silvia brought them.'
- c. (COMMON ERRORS - FULL NP OR OMISSION)
*Silvia ha portato (i fuori).
Silvia AUX.3SG bring.PST.PTCP the flowers
'Silvia brought (the flowers).'

• Subject realized postverbally
• Transitive verb: direct object realized as a clitic pronoun
• Tensed verb form: proclitic

RELATED WORK

An LFG Grammar Checker for CALL (Fortmann & Forst 2004)

- ▶ Grammar checker based on an existing large scale LFG Grammar for German
- ▶ Malrules & OT-marks to parse ungrammatical sentences
- ▶ Additional f-structural annotations to identify marked word orders
 - Strictly ungrammatical constructions can be differentiated from marked ones

Arboretum: Using a precision grammar for grammar checking in CALL (Bender et al. 2004)

- ▶ Tutorial system for English based on the *English Resource Grammar* (Flickinger 2000)
- ▶ Malrules relate erroneous input to well-formed semantic representations
 - Diagnose error type
 - Generated from well-formed semantic representations to produce corrected forms

THE SYSTEM AND ITS COMPONENTS

(I) OT-MARKS

- Allow statement of preferences and dispreferences
- Ordered according to their relative importance
- Added to rules or lexical entries
- **Here:** allow to deal with ambiguous and, more importantly, ungrammatical input
 - ▶ *Ungrammatical* OT-marks to mark error rules that parse ungrammatical structures
 - ▶ O-structure & f-structure contain information on the error type → input for generator

(II) XLE GENERATOR (Crouch et al. 2011)

- Reverse of parsing
 - ▶ Same grammar to parse & generate
 - ▶ Analysis to string
- Input for generation: f-structure analysis
- **Here:** "Main engine" for feedback-system
 - ▶ Used to provide grammatical sentences given an ungrammatical input by user
 - ▶ Generate feedback on error type
 - ▶ Generate paradigms, e.g. verb paradigms in case of errors in subject-verb agreement
 - ▶ Provide linguistic information on why certain structures must be used in a given context, e.g. (1b) vs (1c)

(III) *morph-it!* (Zanchetta & Baroni 2005)

- Lexicon of inflected forms with their lemma & morphological features
- Contains 34,968 lemmas & 504,906 entries
- **Here:** compiled for the Xerox Finite State Tool (XFST) (Beesley & Karttunen 2003) and integrated in grammar's lexicon

ARCHITECTURE

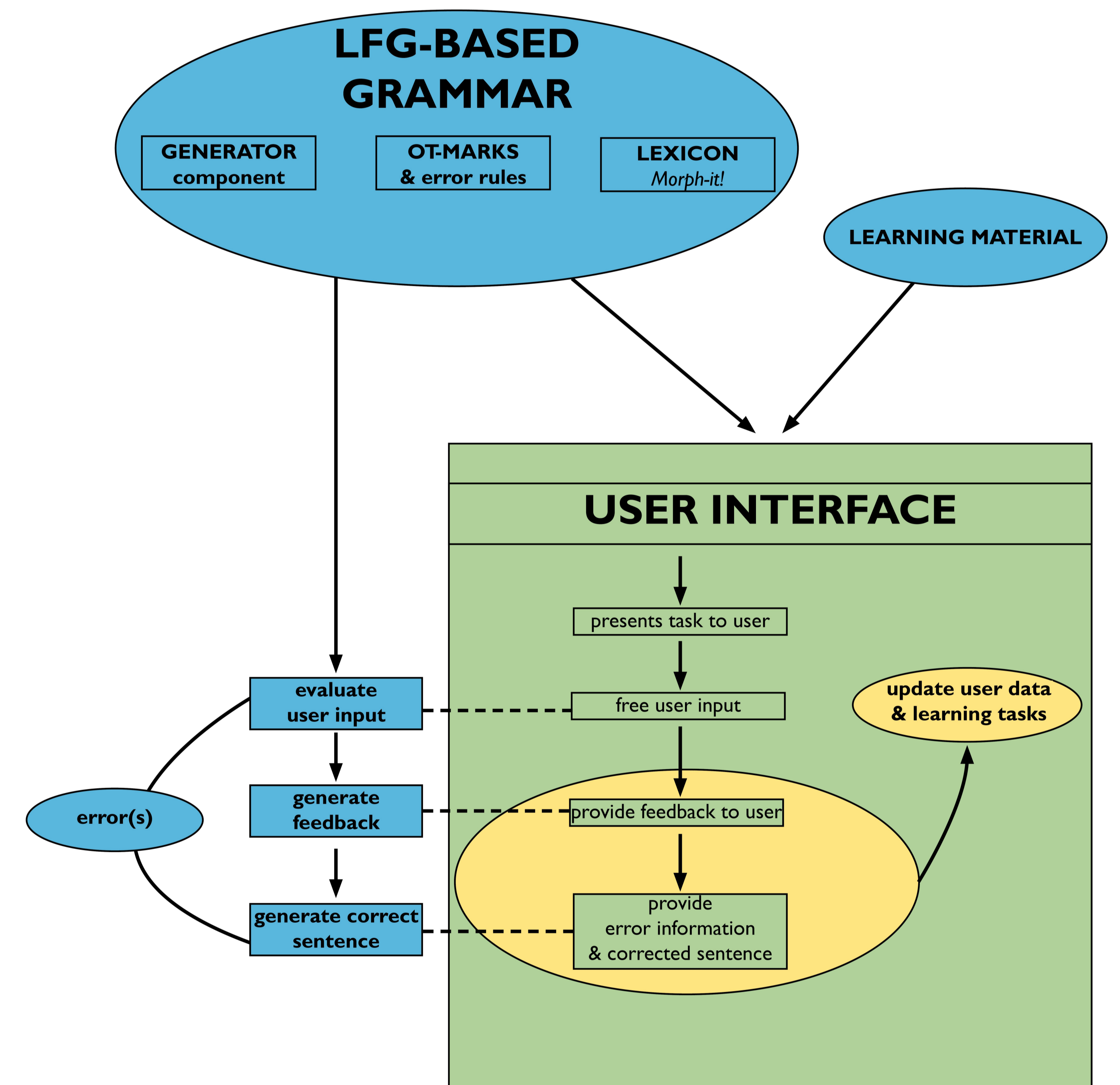


Figure 1: CALL tool - Architecture

FUTURE WORK

- Evaluate the grammar with a learner corpus:
 - ▶ How efficient is the grammar?
 - ▶ What are other common mistakes that should be detected?
- Develop learning material
- Implement user interface & sophisticated feedback system
- Integrate user model as a further component
- Evaluate the system with actual learners of Italian

References

- [1] L. Amaral and D. Meurers. On Using Intelligent Computer-Assisted Language Learning in Real-Life Foreign Language Teaching and Learning. *ReCALL*, 23(1):4-24, 2011.
- [2] K. R. Beesley and L. Karttunen. *Finite State Morphology*. CSLI Publications, Stanford, 2003.
- [3] E. M. Bender, D. Flickinger, S. Oepen, A. Walsh, and T. Baldwin. Arboretum: Using a precision grammar for grammar checking in CALL. *Proceedings of the InSTLICAL2004 Symposium on Computer-Assisted Language Learning*, 17:19, 2004.
- [4] D. Crouch, M. Dalrymple, R. M. Kaplan, T. H. King, J.T. Maxwell III, and P. Newman. *XLE Documentation*. Palo Alto Research Center, 2011.
- [5] D. Flickinger. On building a more efficient grammar by exploiting types. *Natural Language Engineering*, 6(1):15-28, 2000.
- [6] C. Fortmann and M. Forst. An LFG Grammar Checker for CALL. *Proceedings of InSTLICAL2004 - NLP and Speech Technologies in Advanced Language Learning Systems*, (1):2-5, 2004.
- [7] A. Frank, T.H. King, J. Kuhn, and J.T. Maxwell III. Optimality theory style constraint ranking in large-scale LFG grammars. In *Proceedings of the LFG98 Conference*, Stanford, 1998. CSLI Publications.
- [8] R. A. Khader, T. H. King, and M. Butt. Deep CALL grammars: The LFG-OT Experiment. [Talk at DGfS 2004; Online [https://pdfs.semanticscholar.org/presentation/8cb2/f402eab55e9dc75debbe136d5f01e1f0c12e.pdf?_a=2.249700870.144176098.1531230038-1861174644.1528199618], accessed 10-July-2018], 2004.
- [9] C. Leonini and A. Belletti. Adult L2 Acquisition of Italian Clitic Pronouns and 'Subject Inversion' / VS Structures. In *Proceedings of GALA2003*, 2003.
- [10] M. Seiss. *Murrinh-Patha Complex Verbs: Syntactic Theory and Computational Implementation*. PhD thesis, University of Konstanz, Konstanz, 2013.
- [11] E. Zanchetta and M. Baroni. Morph-it! a free corpus-based morphological resource for the Italian language. *Corpus Linguistics* 2005, 1(1), 2005.