A Deep Linguistic Computer-Assisted Language Learning System for Italian

Jessica Zipf, Universität Konstanz

In this paper we report on the development of a Computer Assisted Language Learning (CALL) tool for learners of Italian. We show how the formal power of LFG and XLE can be harnessed to design a system that can parse ungrammatical input, flag it and then provide the user with the grammatical alternative. In this, the generation component of XLE as well as the use of concepts from Optimality Theory (OT) (e.g. Prince and Smolensky 1993; Frank et al. 1998) are crucial. The basis for the tool is a self-written Grammar for Italian that integrates the morphological analyzer Morph-it! (Zanchetta and Baroni, 2005). To account for influencing factors of learners' L1, the tool is aimed at native German speakers. We demonstrate the system with respect to Italian clitic pronouns and 'subject inversion' (VS) structures, phenomena that have been shown to be frequent errors among German learners of Italian (e.g. Leonini and Belletti 2003):

(1) L(o)= ha buttata ne-l cestino.
CL.3SG.M.ACC= AUX.3SG throw.PST.PTCP in-the basket.
'He has thrown it away in the basket.'

(2) È arrivato Gianni.
AUX.3SG arrive.PST.PTCP John.
'John arrived.'

The theoretical framework of LFG is well suited to be used in a CALL environment (see e.g. Reuer 2003). The formal architecture of LFG successfully describes morpho-syntactic phenomena for a variety of languages. One important aspect captured by LFG is the difference between configurational languages with strict word order and languages with a more or less free word order. LFG and its key projections c- and f-structure resemble this difference. In Italian, word order is fairly free and strongly interacts with information structural content. For example, the subject can be placed in various different positions in a sentence, all of which are associated with different information structural roles (e.g. Antinucci and Cinque 1991; Rizzi 1997; Rizzi and Bocci 2017). The modular architecture of LFG allows to deal with this phenomenon in an appropriate manner by integrating c-, f- and i-structure projections. Hubbard (1994:59-60) argues for LFG to have properties that are useful for the pedagogical explanation of syntactic structures in a learning environment. He regards the theory's treatment of grammatical functions as primitives as a strong argument for using LFG for language teaching as it proves reasonable to state pedagogical rules using them. Aside from theoretical benefits, LFG implemented grammars have proven to be fast, robust and running efficiently (e.g. Butt et al 1999).

Another formalism used in existing CALL tools is Head-driven Phrase Structure Grammar (HPSG): for example a tutoring system that evaluates and responds to student performance in German (Heift 1998), or the grammar checker *Arboretum* (Bender et al. 2004), based on the *English Resource Grammar* (Copestake and Flickinger 2000; Flickinger 2002, 2011), among others (e.g. Amaral and Meurers 2011).

In our project, concepts from OT are incorporated to deal with ambiguous and, most importantly, ungrammatical user input. Optimality marks are ordered according to their relative importance. Additionally, OT-marks can be declared as preference marks. Given that parsing often results in multiple possible analyses, the OT-marks determine the "winner" among competing analyses by preferring the parse associated with the most preference marks. Here, we use "ungrammatical" marks to mark error rules that parse ungrammatical constructions. For example, we parse sentences with errors in subject-verb agreement. Whenever that kind of error occurs, the grammar returns to the user that subject-verb agreement is not satisfied and provides the grammatical version of the sentence by generating it from f-structure (Khader 2003).

The Italian grammar parses learner input, flags ungrammatical sentences by error type and produces the corresponding grammatical sentence on demand.

Besides expanding the Grammar and its lexicon, future work includes designing learning material and implementing a user interface. In a final step, the system needs to be evaluated and tested by real-life learners of Italian.

References

Amaral, L., Meurers, D. (2011). *On Using Intelligent Computer-Assisted Language Learning in Real-Life Foreign Language Teaching and Learning*. ReCALL, 23(1):4-24.

Antinucci, F., Cinque, G. (1991). Sull'ordine delle parole in italiano: l'emarginazione. In G. Cinque (ed.), *Teoria linguistica e sintassi italiana*, 163–196. Bologna: Il Mulino.

Bender, E.M., Flickinger, D., Oepen, S., and A. Walsh (2004). Arboretum: Using a precision grammar for grammar checking in CALL. *In Proceedings of the InSTIL/ICALL Symposium 2004*, Venice, Italy.

Butt, M., Dipper, S., Frank, A., and T.H. King (1999). Writing large-scale parallel grammars for English, French and German. In M. Butt & T.H. King (eds), *Proceedings of the LFG99 Conference.* Stanford: CSLI Publications.

Crouch, D., M. Dalrymple, R.M. Kaplan, J. Maxwell, T.H. King and P. Newman (2011). *XLE Documentation*. Palo Alto Research Center.

Frank, A., T.H. King, J. Kuhn and J. Maxwell (1998). Optimality Theory style constraint ranking in large-scale LFG grammars. In M. Butt & T.H. King (eds), *Proceedings of the LFG98 Conference*. Stanford: CSLI Publications.

Flickinger, D. (2011). Accuracy vs. Robustness in Grammar Engineering. in E.M. Bender and J.E. Arnold, eds. *Language from a Cognitive Perspective: Grammar, Usage, and Processing*, 31-50. Stanford: CSLI Publications.

Flickinger, D. (2002). On building a more efficient grammar by exploiting types. In S. Oepen, D. Flickinger, J. Tsujii and H. Uszkoreit (eds.), *Collaborative Language Engineering*, 1-17. Stanford: CSLI Publications.

Flickinger, D. (2000). An open-source grammar development environment and broad-coverage English grammar using HPSG. In *Proceedings of the Second conference on Language Resources and Evaluation (LREC-2000)*.

Heift, T. (1998). *Designed Intelligence: A language teacher model.* Ph.D. thesis, Simon Fraser University.

Hubbard, P.L. (1994). Non-transformational theories of grammar: Implications for language teaching. In T. Odlin (ed.), *Perspectives on pedagogical grammar*, 49-71. Cambridge: Cambridge University Press.

Khader, I. (2003). *Evaluation of an English LFG-based grammar as error checker*. Master's thesis, University of Manchester Institute of Science and Technology.

Leonini, C. and A. Belletti (2003). Adult L2 Acquisition of Italian Clitic Pronouns and 'Subject Inversion'/ VS Structures. *Proceedings of GALA 2003*.

Prince, A., and P. Smolensky (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*. Technical Report 2, Rutgers University Center for Cognitive Science.

Reuer, V. (2003). *Error recognition and feedback with Lexical-Functional Grammar*. CALICO Journal, 20(3):497-512.

Riezler, S., King, T.H., Kaplan, R., Crouch, D., Maxwell, J., and Johnson, M. (2002) *Parsing the Wall Street Journal using a Lexical-Functional Grammar and Discriminative Estimation Techniques*. Proceedings of the Annual Meeting of the Association for Computational Linguistics, University of Pennsylvania.

Rizzi, L., Bocci, G. (2017). The Left Periphery of the Clause . In M. Everaert & H. van Riemsdijk (eds.), *The Companion to Syntax, 2nd Edition*, 1–33. John Wiley & Sons.

Rizzi, L. (1997). The Fine Structure of the Left Periphery. In L. Haegeman (ed.), *Elements of Grammar: A Handbook of Generative Syntax*, 281–337. Dordrecht: Kluwer Academic Press.

Zanchetta, E. and M. Baroni (2005). Morph-it! A free corpus-based morphological resource for the Italian Language. In *Proceedings of Corpus Linguistics 2005*.