

5 A deep linguistic computer-assisted language learning game for Italian

Jessica Zipf¹

Abstract

This paper describes the initial stages of a project which seeks to develop a language learning game for Italian running a deep linguistic grammar at its backend for fine-grained error detection. The grammar is designed within the grammatical framework of Lexical Functional Grammar (LFG). The project aims to bring together work from different fields by combining strategies from computational linguistics with theoretical insights from Second Language Acquisition (SLA) and components from computer gaming.

Keywords: Italian, SLA, lexical-functional grammar, gaming.

1. Introduction

In recent years, many tools have been developed to facilitate language learning. Aside from commercial products such as Duolingo (Teske, 2017) or HelloTalk (Rivera, 2017), grammar checkers based on linguistic frameworks have been implemented as well. These include a grammar checker for German (Fortmann & Forst, 2004) based on a large-scale LFG grammar, and *Arboretum* (Bender et al., 2004), a tutorial system for English using Flickinger's (2000) English Resource Grammar at its backend.

1. University of Konstanz, Konstanz, Germany; jessica.zipf@uni-konstanz.de; <https://orcid.org/0000-0002-5264-0769>

How to cite this chapter: Zipf, J. (2019). A deep linguistic computer-assisted language learning game for Italian. In A. Plutino, K. Borthwick & E. Corradini (Eds), *New educational landscapes: innovative perspectives in language learning and technology* (pp. 33-40). Research-publishing.net. <https://doi.org/10.14705/rpnet.2019.36.953>

While most commercially available systems rely on pattern-matching algorithms, grammar-based tools allow a more fine-grained error detection. Instead of comparing an input string to pre-programmed answers, a deep linguistic grammar analyses an input string morphologically and syntactically. The system developed here combines an LFG based grammar for Italian with a large-scale lexicon covering a wide range of Italian vocabulary. The grammar tool allows for the detection of errors and forms the building block for generating feedback to the learner. The actual learning process is guided by insights from processability theory (Bettoni & Di Biase, 2015; Pienemann, 2005), a theory of SLA that focusses on language development over time by analysing which forms from a second language are processable at which developmental stage (Pienemann, 2005). A recent publication discussing the development of Italian as a second language (Bettoni & Di Biase, 2015) is hereby fundamental for the exercise design and the order in which they are presented to the learners to ensure a successful language learning experience.

This paper aims to present the current state of the project by introducing the components that form the backbone of the computer-assisted language learning tool: the LFG based grammar and how its architecture and components are of great benefit.

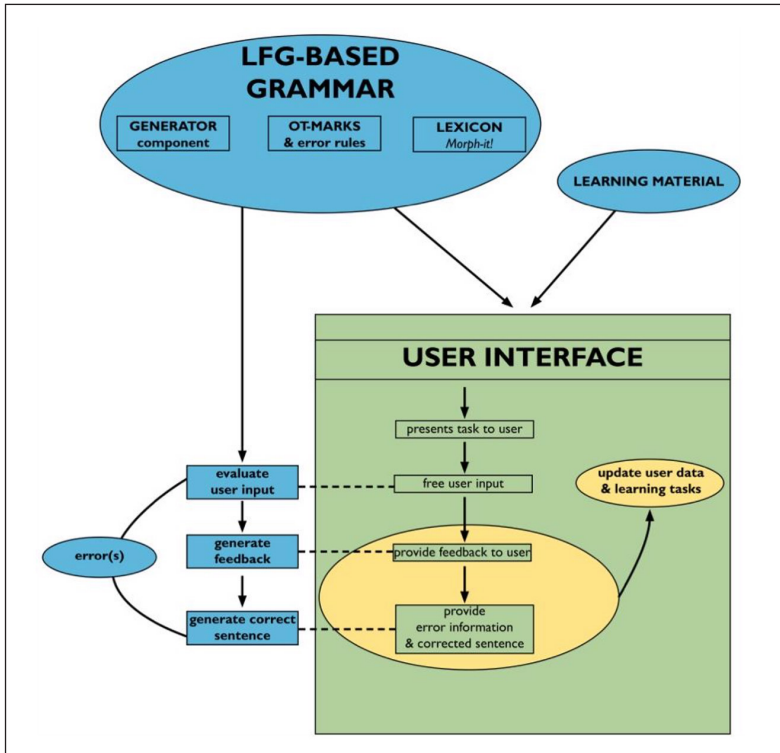
2. Method

Crucial to the tool is the combination of the following components: (1) concepts from Optimality Theory (OT) (Frank, King, Kuhn, & Maxwell, 1998) combined with error rules, (2) the generation component of the Xerox Linguistic Environment (XLE) (Crouch et al., 2008), and (3) the lexicon. Figure 1 illustrates the system's architecture and how these components feed into the language learning software.

The user interface is being fed by the LFG based grammar, alongside learning material. The grammar is responsible for evaluating free user input, generating feedback, and a corrected sentence or structure in case of erroneous input. The

user is thus presented with an exercise, inputs their answer, and receives feedback. This information is then passed on to update the user data and subsequent tasks are adapted accordingly.

Figure 1. Architecture



3. System components

3.1. Lexical Functional Grammar

LFG is a lexicalist, non-derivational theory of grammar (Dalrymple, 2001; Kaplan & Bresnan, 1982). In contrast to other generative grammars, LFG

assumes parallel representations for sentences, each with its own structure and vocabulary, and adhering to its own constraints. Constituent structure and functional structure (c- and f-structure) are the two main representations for sentences. While a c-structure depicts hierarchical relations, constituency, and linearity in the shape of a syntactic tree, an f-structure captures grammatical functions and semantic notions such as tense and aspect in an attribute-value matrix. Having a strong mathematical architecture, LFG is not only implementable but also efficient and fast in analysing input. The Italian grammar in this project is implemented with XLE, a platform commonly used to implement LFG based grammars.

3.2. OT marks

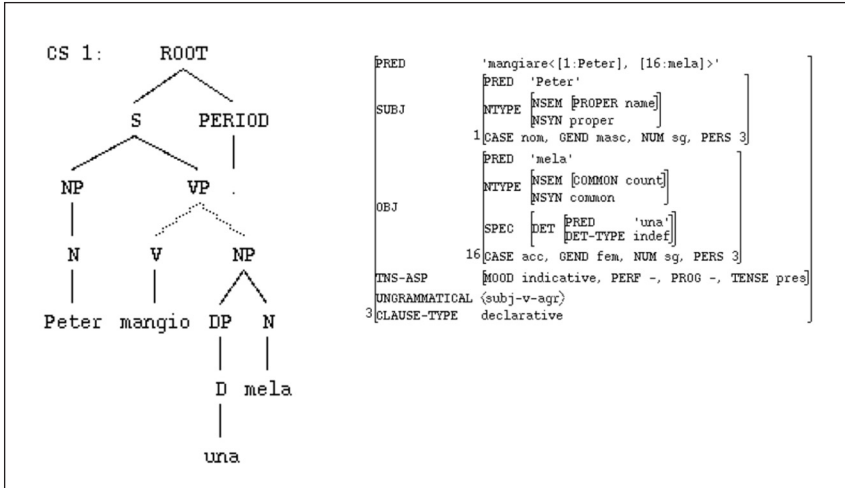
OT marks allow the statement of preferences and dis-preferences in sentence analysis and can be ordered according to their relative importance. As a result, these marks enable the system to deal with ambiguous and ungrammatical input. The mark *ungrammatical* is added to error rules in the grammar, allowing the parser to analyse ungrammatical sentences. Additionally, information on the error type can be added, passing this information on to an f-structure. Consider the subsequent sentence:

- Peter mangi-**o*** una mela.
- Peter.**3PSG** eat-**1PSG** an apple.
- ‘Peter eats an apple’.

This is an example of erroneous subject-verb agreement. Parsing the sentence with the LFG grammar yields the following output (Figure 2).

While the c-structure on the left illustrates the constituents and the hierarchical structure, the f-structure contains semantic information on the main predicate and its arguments. Additionally, it returns the information that subject-verb agreement is ungrammatical in this example. This information is passed on to the f-structure by adding certain annotations to the OT mark *ungrammatical*.

Figure 2. C- and f-structure of Example 1



3.3. XLE generator

The generator component of XLE enables the system to create a sentence given an f-structure as input. Generation being the reverse of parsing, the same grammar produces a string based on a certain f-structure analysis. Here, the generator is used to provide a grammatical sentence given an ungrammatical input. Going back to Example 1, the XLE generator takes as input the f-structure in Figure 2 and produces the grammatical alternative depicted in Figure 3.

Figure 3. Grammatical alternative to Example 1 as generated by XLE

```

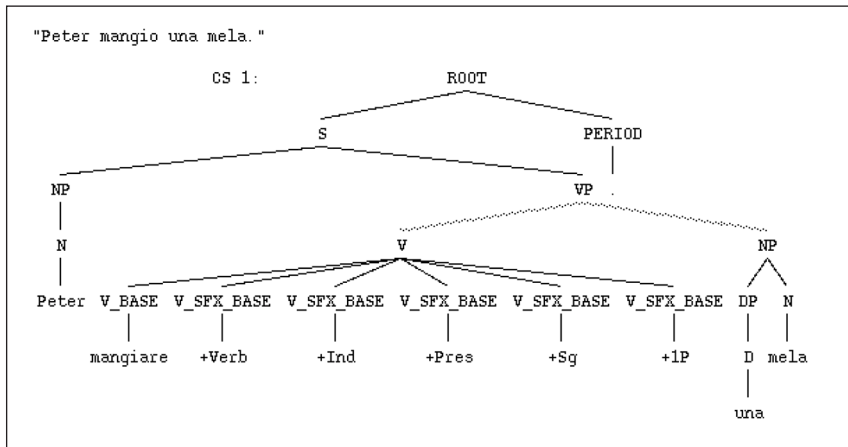
% parse "Peter mangio una mela."
parsing {Peter mangio una mela.}
1 solutions, 0.003 CPU seconds, 0.170MB max mem, 30 subtrees unified
1
%
Peter mangia una mela.
  
```

The first line illustrates the ungrammatical sentence that was passed on to the system, while the last line shows the grammatical alternative with subject-verb agreement satisfied.

3.4. Lexicon

The lexicon constitutes the third big building block of the tool. It was created by converting the *Morph-it!* lexicon (Zanchetta & Baroni, 2005) into a finite state morphological analyser using the Xerox Finite State Tool (Beesley & Karttunen, 2003). The lexicon contains 34,968 lemmas and an overall count of 504,906 entries. Not only does it provide the grammar with a large Italian lexicon, but additionally with morphological analyses that expand from the c-structure, as depicted in Figure 4.

Figure 4. Integrating a finite state morphological analyser



4. Conclusions

This paper outlines an on-going project on the development of a language learning game for Italian. The tool is based on a deep grammar within the framework

of LFG and integrates an Italian lexicon. While the grammar component is responsible for analysing the user input and returning adequate feedback, the learning process is to be guided by insights from the processability theory. At this initial stage of development, the grammar component can detect error types and return the grammatical alternatives. Next steps include further expansion of the grammar and its syntactic structures, an evaluation of the grammar using learner corpora, and the development of learning material and learning exercises. The final stages of the project are then to incorporate the grammar and the learning material into an attractive and user-friendly environment, including gaming components to enhance the learning experience.

Acknowledgements

I want to thank my supervisor Miriam Butt for the possibility to work on this project and her constant support and input, as well as Bruno Di Biase and Barbara Hinger for their input and discussions on SLA and processability theory.

References

- Beesley, K. R., & Karttunen, L. (2003). *Finite state morphology*. CSLI Publications. <http://www.fsmbook.com>
- Bender, E. M., Flickinger, D., Oepen, S., Walsh, A., & Baldwin, T. (2004). Arboretum: using a precision grammar for grammar checking in CALL. *Proceedings of the InSTILICALL2004 Symposium on Computer-Assisted Language Learning, 17*, 19. http://www.isca-speech.org/archive_open/archive_papers/icall2004/iic4_020.pdf
- Bettoni, C., & Di Biase, B. (2015). The development of Italian as a second language. In C. Bettoni & B. Di Biase (Eds), *Grammatical development in second languages: exploring the boundaries of processability theory* (pp. 117-147). EuroSLA.
- Crouch, D., Dalrymple, M., Kaplan, R. M., King, T. H., Maxwell, J., & Newman, P. (2008). XLE documentation. *Palo Alto Research Center*.
- Dalrymple, M. (2001). *Lexical functional grammar*. Academic Press.

- Flickinger, D. (2000). On building a more efficient grammar by exploiting types. *Natural Language Engineering*, 6(1), 15-28. <https://doi.org/10.1017/S1351324900002370>
- Fortmann, C., & Forst, M. (2004). An LFG grammar checker for CALL. *Proceedings of InSTIL/ICALL2004 - NLP and Speech Technologies in Advanced Language Learning Systems*, (1), 2-5.
- Frank, A., King, T., Kuhn, J., & Maxwell, J. (1998). Optimality theory style constraint ranking in large-scale LFG grammars. In *Proceedings of the LFG98 Conference*. Stanford: CSLI Publications. <http://elib.uni-stuttgart.de/opus/volltexte/1999/356/>
- Kaplan, R. M., & Bresnan, J. (1982). Lexical-functional grammar: a formal system for grammatical representation. *Issues in Lexical-Functional Grammar*, 8(5), 29-130.
- Pienemann, M. (Ed.). (2005). *Cross-linguistic aspects of processability theory*. John Benjamins.
- Rivera, A. V. (2017). HelloTalk. *CALICO Journal*, 34(3), 384-392.
- Teske, K. (2017). Duolingo. *CALICO Journal*, 34(3), 393-401.
- Zanchetta, E., & Baroni, M. (2005). Morph-it! a free corpus-based morphological resource for the Italian language. *Corpus Linguistics 2005*, 1(1).



Published by Research-publishing.net, a not-for-profit association
Voillans, France, info@research-publishing.net

© 2019 by Editors (collective work)
© 2019 by Authors (individual work)

New educational landscapes: innovative perspectives in language learning and technology
Edited by Alessia Plutino, Kate Borthwick, and Erika Corradini

Publication date: 2019/07/10

Rights: the whole volume is published under the Attribution-NonCommercial-NoDerivatives International (CC BY-NC-ND) licence; **individual articles may have a different licence.** Under the CC BY-NC-ND licence, the volume is freely available online (<https://doi.org/10.14705/rpnet.2019.36.9782490057481>) for anybody to read, download, copy, and redistribute provided that the author(s), editorial team, and publisher are properly cited. Commercial use and derivative works are, however, not permitted.

Disclaimer: Research-publishing.net does not take any responsibility for the content of the pages written by the authors of this book. The authors have recognised that the work described was not published before, or that it was not under consideration for publication elsewhere. While the information in this book is believed to be true and accurate on the date of its going to press, neither the editorial team nor the publisher can accept any legal responsibility for any errors or omissions. The publisher makes no warranty, expressed or implied, with respect to the material contained herein. While Research-publishing.net is committed to publishing works of integrity, the words are the authors' alone.

Trademark notice: product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Copyrighted material: every effort has been made by the editorial team to trace copyright holders and to obtain their permission for the use of copyrighted material in this book. In the event of errors or omissions, please notify the publisher of any corrections that will need to be incorporated in future editions of this book.

Typeset by Research-publishing.net
Cover design by © Raphaël Savina (raphael@savina.net)

ISBN13: 978-2-490057-48-1 (Ebook, PDF, colour)
ISBN13: 978-2-490057-49-8 (Ebook, EPUB, colour)
ISBN13: 978-2-490057-47-4 (Paperback - Print on demand, black and white)
Print on demand technology is a high-quality, innovative and ecological printing method; with which the book is never 'out of stock' or 'out of print'.

British Library Cataloguing-in-Publication Data.
A cataloguing record for this book is available from the British Library.

Legal deposit, UK: British Library.
Legal deposit, France: Bibliothèque Nationale de France - Dépôt légal: juillet 2019.
