

1222 · 2022
800
ANNI



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

DISL DIPARTIMENTO DI STUDI
LINGUISTICI E LETTERARI

**Centro
Linguistico
di Ateneo**



6th LEARNER CORPUS RESEARCH CONFERENCE
Padua, 22-24 September 2022

Book of Abstracts

Table of Contents

PLENARY TALKS

Bernardini, Silvia	9
A marriage of two minds? Learner translation corpora in learner corpus research	
Lüdeling, Anke	10
Explorations of variability: Evidence from L1 and L2 corpora of German	
Nesi, Hilary	12
Learner corpus research: Some problems, some questions, and some possible answers	

FULL PAPERS

Balakina, Ksenia	13
Splitting and joining sentences in Italian-Russian inverse translation	
Bear, Elizabeth, Xiaobin Chen, and Detmar Meurers	15
Linguistic style in a second language: Exploring cross-task individual differences in complexity in a large-scale corpus	
Biber, Doug, Tove Larsson, Gregory Hancock, Bethany Gray and Randi Reppen	17
Dimensions of grammatical complexity in L1/L2 writing: A comparative analysis of theory-based models	
Bienati, Arianna and Jennifer Carmen Frey	19
Development of explicit causal connectives in Italian L1 and L2 student writing: A comparison of argumentative texts from lower and upper secondary school	
Brocca, Nicola, Maria K. Rudigier, and Valentin A. Spielthener	21
A corpus-based approach in foreign-language teacher education: A case study on politeness in instant messages in Italian L2 by Germanophone learners	
Callies, Marcus	23
Challenges in the annotation and analysis of learner corpora	
Charles, Maggie, Ahmed Halil, Michael Jenkins, and Karin Whiteside	25
What gets funded? A learner corpus study of grant proposal summaries by L1 Arabic-Syrian academics	
Crawford, William J.	27
Disfluencies in L2 peer interaction: A corpus analysis of cognitive fluency	
Dagbandan, Sepideh	29
A comparison between colloquial Persian used by English-speaking learners of Persian and Iranian speakers of Persian: Insights from a learner corpus-based study	
Derkach, Kateryna and Dora Alexopoulou	31
The differential effect of specificity on definite and indefinite article accuracy in learner English	
Deshors, Sandra C. and Steven Gagnon	33
“ <i>The store is wanting staff</i> ”: A multifactorial approach to progressive marking in Korean English	
Di Nuovo, Elisa Bianca, Maria De Paolis, Cristina Bosco, and Elisa Corino	35
Error identification, normalization and tagging: Three inter-annotator agreement experiments in a picture-elicited learner corpus	

Dupont, Maité and Sylviane Granger	37
Connector placement in EFL learner writing: Focus on <i>however</i>	
Gagnon, Steven and Sandra C. Deshors	39
‘ <i>Growing up students</i> ’: A collostructional analysis approach to phrasal verbs in Korean learner English	
Gaillat, Thomas	41
Exploring the operationalisation of L2 microsystems as functional complexity metrics for proficiency assessment	
Gee, Roger W., M. Karen Jogan, and Kathleen S. Jogan	43
Developmental use of pronominal noun modifiers by Spanish L1 EFL teachers	
Gesuato, Sara and Elisabetta Pavan	45
Students’ requestive emails to faculty-pragmatic proficiency in elicited and spontaneous Italian L1 and English L2	
Glaznieks, Aivars and Jennifer-Carmen Frey	47
Syntactic variation in German <i>weil</i> -clauses: A Comparison between immersed and non-immersed learners of German	
Gries, Stefan Th.	49
Most dispersion measures do not measure dispersion, and the implications of that for LCR	
Gries, Stefan Th. and Magali Paquot	50
Association measures in learner corpus research: Problems and pointers for improvement	
Guziurová, Tereza	52
Code glosses in L2 learner writing: Reformulation and exemplification in master’s theses by Czech university students	
Hartle, Sharon, Giorgia Andreolli, and Emanuela Tenca	54
Visual Thinking Strategies (VTS) in online EFL learner discussions: Creating a micro-corpus of spoken learner discourse for qualitative analysis	
Hasselgård, Hilde	56
Young writers' use of adverbial intensification in English L1 and L2	
Hasund, Ingrid Kristine	58
Genres in young learner EFL writing: A genre typology for the TRAWL (tracking written learner language) corpus	
Iurato, Alessia	60
Compiling a corpus of written and spoken L2 Chinese: Combining pragmatic and error annotation to study the Chinese <i>shi</i> 是... <i>de</i> 的 cleft construction	
Ivaska, Ilmari	62
Register effects and morphosyntactic complexity affecting the use of the preterite construction in advanced L2 Finnish	
Izquierdo, Marlén and Naroa Zubillaga	64
Empirical translation studies: Contrasting learner translations in a diglossic environment	
Kaatari, Henrik, Tove Larsson, Ying Wang, Seda Acikara Eickhoff, and Pia Sundquist	66
Exploring the effect of target-language extramural activities on students’ written production	
Kavalir, Monika and Gašper Ilc	68
Use of English negation in the Slovene subcorpus of ICLE	

Kia, Elnaz and Fernando Rubio	70
Lexical bundles and L2 Spanish writing development: A case of dual language immersion	
Kim, Sangeun	72
Multidimensional analysis of syntactic complexity development in L2 learner writing in an American university EAP programme	
Kircili, Kathrin	74
Non-canonical syntax in learner language: Between language transfer, language universals and idiosyncrasies	
Kisselev, Olesya, Rossina Soyán, Dmitrii Pastushenkov, and Jason Merrill	76
Lexical and syntactic complexity development in L2 Russian texts and correlations with curricular levels and raters' scores	
König, Alexander, Jennifer-Carmen Frey, Egon W. Stemle, Aivars Glaznieks, and Magali Paquot	78
Towards standardizing LCR metadata	
Larsson, Tove, Tony Berber-Sardinha, Bethany Gray, and Doug Biber	80
Exploring early L2 writing development: A register-functional approach to grammatical complexity	
Le Foll, Elen	82
Teaching pre-service teachers to create corpus-informed materials: The effectiveness of different types of tasks in an e-learning setting	
Lee, Joseph J. and Robert Bern	84
Changing patterns of linking adverbials in L2 university student writing	
Leńko-Szymańska, Agnieszka, Piotr Pęzik, and Michał Adamczyk	86
Phraseology in the assessment of L2 writing	
López-Sako, Nobuo Ignacio and Cristóbal Lozano	87
Redundancy in subject anaphora resolution: A corpus-based study of L1 Japanese learners of L2 Spanish	
Montaño, Jorge and Ana Díaz-Negrillo	89
Does mode affect referring expression selection? A corpus-based study of advanced L1 Spanish-L2 English narratives	
Oksuz, Dogus Can, Dora Alexopoulou, Kate Derkach, and Ianthi Maria	91
The influence of L1 typology on the acquisition of the L2 English articles: A large-scale learner corpus study	
Paquot, Magali, Rachel Rubin, and Nathan Vandeweerd	93
Introducing the CLAP project: Adaptive comparative judgment as a community-based solution for enriching learner corpora with crowdsourced L2 proficiency assessment	
Poli, Francesca	95
"Let's say maybe it's our Italian culture": Expressions of uncertainty in Italian learners of English	
Puga, Karin	97
F0 range in L2 discourse as evidence for the existence of a prosody interlanguage system	
Quesada, Teresa and Cristóbal Lozano	99
Using two comparable learner corpora to investigate the production of referring expressions bidirectionally: L1 Spanish-L2 English vs. L1 English-L2 Spanish	
Reppen, Randi and Doug Biber	101
Studying individual longitudinal development in a corpus of 'natural' disciplinary writing	

Rudebeck, Lisa and Gunlög Sundberg	103
On the other side of the error tag: The nature and functions of the corrected texts	
Spina, Stefania	105
Task effects on phraseological complexity in learners' written and oral production: A structural equation modeling study	
Sun, Qiuyi	107
Modality in Chinese EFL learners' academic writing: From semantic meaning to disciplinary variation	
Tayeh Chamoun, Jessica and Nicolas Ballier	108
Automatic classification of Arabic learners of English based on complexity metrics	
Tomson, Annely	110
Acquisition of Norwegian as a second language: What are the differences between the written and spoken language of the learners?	
Vandeweerd, Nathan	111
The effect of phraseological complexity on ratings of oral versus written French proficiency	
Weiss, Zarah and Detmar Meurers	113
How do tasks impact the different domains of L2 linguistic complexity?	
Weiss, Zarah, Nina Selina Hicks, Detmar Meurers, and Thomas Studer	115
Using linguistic complexity to probe into genre differences? Insights from the multilingual SWIKO learner corpus	
Wuttisrisiriporn, Niwat	117
Investigating effects of L1 and discipline on syntactic complexity in master's theses and research articles	
 WORK-IN-PROGRESS REPORTS	
Ahmed, Abdelhamid, Lameya Rezk, and Xiao Zhang	119
A corpus-based contrastive analysis of transition markers in L1 Arabic and L2 English argumentative writing	
Bottini, Raffaella	122
Lexical complexity in L2 English speech: Exploring monologic and dialogic tasks in the Trinity Lancaster corpus	
Bulantová, Barbora	123
Measuring syntactic complexity in L2 speech at advanced proficiency levels	
Burton, Graham and Maria Cristina Gatti	125
English in a bilingual German-Italian community: Collecting data and investigating learner variables in creating the EdiCoMC corpus	
De Cock, Sylvie	127
<i>Do you love me</i> : Interrogatives in learner speech in LINDSEI and in the Trinity Lancaster corpus	
De Kuthy, Kordula and Detmar Meurers	129
Extending experimental research on the effectiveness of an intelligent tutoring system: A corpus study systematically identifying targeted language means in authentic ESL student essays	
Dusturia, Nida	131
The use of connectors in spoken and written argumentative texts of Indonesian EFL learners: A corpus-based study	

Klavan, Jane	133
A multifactorial learner corpus approach to genitive alternation in non-native English	
Li, Jen-Yu, Thomas Gaillat, and Elisabeth Richard	135
Exploring the use of dependency parsing in automatic erroneous collocation extraction in learner English	
Lopopolo, Olga	137
The acquisition and use of the progressive aspect by multilingual learners of English as L3: Preliminary results from a longitudinal learner corpus-based study	
Lorenz, Eliane	139
“So I’ll need English like good English” – Functions and use of discourse marker <i>like</i> in UAE English	
Murakami, Akira	141
Towards more appropriate modeling of (and with) linguistic complexity indices	
Quinci, Carla	143
“Today’s lesson was really interesting”: Improving second-language learning and obtaining feedback through students’ reflective Padlet posts	
Shadrova, Anna	145
Lexical similarity in L1 and L2 German as evidence for the structure and dynamics of the lexicon	
Sugiura, Masatoshi, Akiko Eguchi, Mariko Abe, Remi Murao, Takashi Koizumi, Daisuke Abe	147
Using IPSyn to measure early L2 syntactic development	
Thomas, Anita and France Rousset	149
Corpora as input and output: A fragile link in classroom research	
Wedig, Helena, Carola Strobl, and Jim Ureel	151
Investigating connective use in L2 German: A corpus study	
 POSTERS	
Alameer, Sadeem Ibn, Dagmar Divjak, and Paul Thompson	153
An exploratory corpus-based study of Arab learners’ usage of English phrasal verbs	
Bear, Elizabeth, Bronson Hui, Haemant Santhi Ponnusamy, Björn Rudzewitz, Xiaobin Chen, and Detmar Meurers	154
Using ICALL to collect spoken learner data in real-life conversation tasks	
El Ayari, Sarra	156
Sarramanka: An online tool for learner corpora analysis	
Flores Hernández, Ana Abigahil and Pauline Moore	158
Mexican learner corpus: Designing and collecting a longitudinal spoken corpus of Mexican university learners of English	
Forti, Luciana, Irene Fioravanti, and Fabio Zanda	160
Lexical complexity across proficiency levels in L2 Italian: Some preliminary findings	
García-Guerrero, Elena and Cristóbal Lozano	162
Is planning time beneficial for L2 production? A corpus-based study of anaphora resolution in L1 Spanish – L2 English learners	
Hammond, Thomas A.	164

From production short-cuts to syntactic development? Analysing the production of fixed expressions (FEs) with the development of the L2 computational component	
Holmquist, Kristoffer and Therese Lindström Tiedemann	165
A corpus-based study of derivational morphology in written L2 Swedish	
Lopopolo, Olga and Fabio Zanda	167
The relevance of inter and intra-rater reliability in multi-layer annotation procedures	
Maso, Sara	169
Tracking the development of written language competence in L2 Italian: A NLP-based approach	
Migliorelli, Alice	171
Variants and varieties of learning preserved in the historical archives of the University for Foreigners of Perugia: Toward the building of a digital learning corpus	
Wedig, Helena, Carola Strobl, and Jim Ureel	173
The Beldeko corpus: A new resource for investigating L2 German texts written by L1 Dutch students	
Weiss, Zarah, Moritz Sahlender, Inga ten Hagen, Anastasia Knaus, and Stefanie Helbig	175
Investigating spoken classroom interactions in linguistically heterogeneous learning groups – An interdisciplinary approach to compile multi-modal corpora in second language classrooms	
Zasina, Adrian Jan and Elżbieta Kaczmarek	177
Czech errors in writings based on the Polish learner corpus PoLKO: A pilot study	
 SOFTWARE DEMOS	
Ballier, Nicolas and Helen Yannakoudakis	179
Towards crowdsourcing research for learner keylogging data	
Chitez, Mădălina, Cosmin Strilechi, and Karla Csűrös	181
Meeting ROGER: An open-access bilingual corpus search platform	
Glaznieks, Aivars, Jennifer-Carmen Frey, Maria Stopfner, Lorenzo Zanasi, Lionel Nicolas	183
LEONIDE: A longitudinal trilingual corpus of young learners of Italian, German and English	
Lozano, Cristóbal and Nobuo Ignacio López-Sako	185
Demonstration of the CEDEL2 (version 2) interface: A multi-L1 corpus of L2 Spanish	
Spina, Stefania, Irene Fioravanti, Luciana Forti, Francesca Malagnini, Angela Scerra, Valentino Santucci, and Fabio Zanda	187
The CELI corpus: A new resource to analyse Italian L2	
Volodina, Elena Therese Lindström Tiedemann, and Yousuf Ali Mohammed	189
Swedish L2 profile – A tool for exploring L2 data	



Sponsor of the Benjamins Poster Prize

Development of explicit causal connectives in Italian L1 and L2 student writing: A comparison of argumentative texts from lower and upper secondary school

Arianna Bienati, Jennifer Carmen Frey
Institute for Applied Linguistics, Eurac Research
arianna.bienati@eurac.edu, jennifercarmen.frey@eurac.edu

When writing, discourse or coherence relations (Mann & Thompson 1988; Kehler 2002; Asher et al. 2003; Miltsakaki et al. 2004) are a paramount strategy to logically connect semantically related stretches of text. Formally, languages provide extensive sets of connectives that encode these semantic relations explicitly (Pander Maat & Sanders 2006). Although the use of such explicit cohesive devices is not necessarily correlated with coherence or text quality judgments (Crossley et al. 2016), its acquisition is an important steppingstone in text competence development. Thus, an in-depth analysis of the types and variety of connectives used at different stages of a writer's school education could provide important empirical data for training of textuality features and writing assessment in L2 and L1 teaching practice of a particular language.

In our contribution, we analyzed the quantity and repertoire of explicit connectives found in argumentative texts of L1 and L2 speakers of Italian in the 3rd year of lower secondary school, and after four years of training, i.e., in the 4th year of upper secondary school. In our analysis, we focus on explicit causal connectives as one important means for constructing coherence in argumentative texts, in that they explicitly point out supporting reasons and anticipated consequences, to convince an audience of a statement.

Our research questions are:

- Are there any common trends in the use of explicit connectives employed by students through time, regarding quantity and repertoire of uses?
- Are there any significant differences in the use of explicit connectives by L1 and L2 speakers at the same developmental stage?

To answer these questions, we automatically annotated the explicit causal connectives in a sample of 200 texts, evenly distributed between the four conditions, namely first/second language and lower/upper secondary school. All texts were gathered in the multilingual province of Bolzano/Bozen in Italy and originated from three different learner corpora. Argumentative texts of lower secondary school writers were randomly sampled from the L2 and L1 writers in the Italian sub-corpus of LEONIDE (Glaznieks et al. 2022). The texts of upper secondary school writers were drawn as a random sample from the Italian Kolipsi-2 corpus (L2 data, Glaznieks et al. 2021) and from data collected in the ITACA project (L1 data, <https://itaca.eurac.edu/>). The automatic annotation follows a dictionary-based approach aided by the Lexicon for Italian COnnectives (LICO) (Feltracco et al. 2016), a repository of Italian connectives aligned with the PDTB 3.0 (Webber et al. 2019). To analyze quantity, we observed both the number of causal connectives per text (normalized per 100 words to account for text length differences) and the ratio of causal connectives of all connectives. Furthermore, we investigated the students' repertoire of causal connectives qualitatively and quantitatively, extracting frequencies from a reference corpus (CORIS, Rossini Favaretti et al. 2002) to understand which kind of connectives (if low or high frequency) were present in the four groups. We calculated both the mean and the standard deviation of the frequencies of connectives used in each group, to measure differences in the repertoires.

Our analysis, aided by linear regression models, shows that the number of causal connectives decreases significantly in the upper grades, independently of L1/L2 variable. However, the category is not internally homogeneous: causal connectives of the result type (e.g., *quindi*, *di conseguenza*) display a remarkable relative growth in upper secondary school, suggesting that result relations are more complex and therefore learned later on. Regarding the kind of connective used, older students of both groups use significantly less common connectives than younger students. Changes in the variety over time exist only in the L1 group, in which new, rarer connectives (e.g., *per via*, *siccome*, *cosicché*), may emerge aside from the high-frequency ones typical of lower grades (e.g., *per*, *perché*, *così*, *quindi*). L2 students, instead, seem to use a narrower range of connectives with similar frequency (higher for lower grades and lower for upper grades). In general, the change in the use of causal connectives over time was similar for both L1 and L2 students, with the only significant difference between L1 and L2 students visible only in the variety and average frequency of connectives used in the upper grades. Results suggest that for both L1 and L2 writers, quantity and variety of connectives employed are in a tradeoff: while through time students may learn other strategies to express coherence relations – determining the decrease

in the quantity of connectives –, they learn to use also rarer connectives, supposedly the ones present in formal, academic language.

References

- Asher, N. M., & Lascarides A. (2003). *Logics of Conversation*. Cambridge: CUP.
- Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The development and use of cohesive devices in L2 writing and their relations to judgments of essay quality. *Journal of Second Language Writing*, 32, 1-16.
- Feltracco, A., Jezek, E., Magnini, B., & Stede, M. (2016). LICO: A Lexicon of Italian Connectives. In A. Corazza, S. Montemagni, & G. Semeraro (Eds.). *Proceedings of the Third Italian Conference on Computational Linguistics CLiC-it 2016*. Torino: Accademia University Press, 141-145.
- Ferrari, A. (2014). *Linguistica del testo. Principi, fenomeni, strutture*. Roma: Carocci.
- Glaznieks, A., Frey, J.-C., Nicolas, L., Abel, A. & Vettori, C. (2021). Kolipsi-2 Corpus v1.0, Eurac Research CLARIN Centre, <http://hdl.handle.net/20.500.12124/30>
- Glaznieks, A., Frey, J.-C., Stopfner, M., Zanasi, L., & Nicolas, L. (2022). LEONIDE: A longitudinal trilingual corpus of young learners of Italian, German and English. *International Journal of Learner Corpus Research*, 8(1), 97-120.
- Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. Stanford: CSLI Publications.
- Mann, W., & Thompson, S. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8, 243–281.
- Miltsakaki, E., Prasad, R., Joshi, A., & Webber, B. (2004). The Penn Discourse Treebank. In M. T. Lino, M. F. Xavier, F. Ferreira, R. Costa, & R. Silva (Eds.). *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*. Lisbon: European Language Resources Association, 2237-2240.
- Pander Maat, H., & Sanders, T. (2006). Connectives in Text. In K. Brown (ed.). *Encyclopedia of Language & Linguistics*. Amsterdam: Elsevier, 33-41.
- Rossini Favaretti, R., Tamburini, F., & De Santis, C. (2002). CORIS/CODIS: A corpus of written Italian based on a defined and a dynamic model. In A. Wilson, P. Rayson, & T. McEnery (Eds.). *A Rainbow of Corpora: Corpus Linguistics and the Languages of the World*. München: Lincom-Europa, 27-38.
- Webber, B., Prasad, R., Lee, A., & Joshi, A. (2019). The Penn Discourse Treebank 3.0 Annotation Manual.