



Università
per Stranieri
di Perugia



INSTITUT
FRANÇAIS
ITALIA

DISCOVERING DATA-DRIVEN LEARNING HOW CORPORA CAN HELP TRANSLATORS AND LEARNERS OF A SECOND LANGUAGE

Book of abstracts

Libro degli abstract



**VENERDÌ
29 MARZO 2019
PALAZZO GALLENGA
SALA GOLDONI**

Program

Programma

14:00-14:15
Opening greetings



14:15-14:45
***Data-driven learning in a nutshell:
what, how, why... and does it work?***
Alex Boulton (ATILF, Université de Lorraine & CNRS)

14:45-15:15
***Corpora for translators:
affordances and limitations
of an under-explored resource***
Silvia Bernardini (University of Bologna, Forlì)

15:15-15:45
***Corpus-based approaches
for the analysis and description
of learner languages: ways and means***
Andrea Abel (Eurac Research, Bolzano)

15:45-16:15
Coffee break



16:15-16:45
***Corpus use and Data-driven
learning in L2 French:
from research issues to commonsense
applications for the modern learner***
Henry Tyne (Université de Perpignan Via Domitia)

16:45-17:15
***From Learner Corpus Research
to Data-driven learning activities:
some ready-to-use examples
for the Italian L2 classroom***
Luciana Forti & Stefania Spina
(University for Foreigners of Perugia)



17:15-18:15
***The future of corpora in translation
and second language learning***
Round table with
**Andrea Abel, Silvia Bernardini,
Alex Boulton, Luciana Forti,
Stefania Spina, Henry Tyne**
and the public



18:15-18:30
Closing greetings

Alex Boulton, (ATILF - Université de Lorraine & CNRS)

Data-driven learning in a nutshell: what, how, why... and does it work?

Despite a substantial history since scholarly papers first started appearing in the 1980s, the use of language corpora for learning and teaching (aka data-driven learning or DDL) still has the reputation as a niche activity. This paper aims to provide a broad introduction to set the theme for the day, addressing the main questions arising. What is DDL (the corpora and tools)? How it can be used (the types of activities for learners, teachers and other users)? And why (the pedagogical arguments for and against)? We close with a look at existing empirical research based on a recent meta-analysis of the field.



Questioni chiave del Data-driven learning: cosa, come, perché e... funziona?

Nonostante una storia considerevole a partire da primi articoli pubblicati negli anni '80, l'uso dei corpora per l'apprendimento e l'insegnamento (noto come data-driven learning o DDL) continua ad avere la reputazione di essere un approccio di nicchia. Questa presentazione vuole fornire un'ampia introduzione al tema della giornata, prendendo in considerazione le principali questioni chiave. Che cos'è il DDL (i corpora e le risorse)? Come può essere utilizzato (tipologie di attività per apprendenti, insegnanti e altri utenti)? E perché (argomenti pedagogici pro e contro)? Si concluderà con uno sguardo alla ricerca empirica attualmente disponibile basata su una recente meta-analisi.

Silvia Bernardini (Università di Bologna, Forlì)

Corpora for translators: affordances and limitations of an under-explored resource

A corpus is a collection of texts in electronic format that can be consulted through a dedicated software. In its various forms (specialist, reference, parallel, learner – and so on – corpora), it has become a fundamental resource for the study of languages, for the creation of dictionaries, and for the development of textbooks and teaching syllabi, especially in English-speaking contexts. In the world of translation, on the other hand, it is still an under-explored resource, despite the fact that documentation work ("what does x mean/ how do you say y in a specialised language variety or genre?") is one of the most time-consuming and complex tasks in the translation process. In this paper, I will highlight the value and the potential, but also the limitations, of these resources compared to other sources of information, and I will offer an overview of the variety of corpora and software programs that are currently available, focusing in particular on those that are free to use, and relevant for a translator.

possèdent un rôle dans l'apprentissage des langues qui, tout en se focalisant sur le
Les logiciels actuels d'apprentissage des langues sont, le plus souvent, peu pertinents pour le
contemporain portrait of second language learning and teaching, to identify major trends
can self-access facilitate play in language learning. These are two areas which the
two key variables affecting language learning styles, i.e., the general
language learning and language learning is available.

Il corpus: pro e contro di una risorsa ancora poco conosciuta (ed amata) dai traduttori

Un corpus è una raccolta di testi in formato elettronico consultabile attraverso software dedicato. Nelle sue varie incarnazioni (corpus di specialità, di riferimento, parallelo, di testi di apprendenti ecc.), negli anni è diventato una risorsa essenziale per lo studio delle lingue, per la creazione di dizionari, per la stesura di libri di testo e di programmi didattici, soprattutto in contesti anglosassoni. Nel mondo della traduzione invece è ancora una risorsa poco utilizzata, nonostante l'attività di documentazione ("cosa significa x/come si dice y in una lingua di specialità o genere particolare?") sia probabilmente una delle più dispendiose e complesse nel processo traduttivo. In questo contributo metterò in evidenza il valore aggiunto e le possibilità ma anche i limiti di queste risorse rispetto ad altre fonti di informazione e offrirò una panoramica dell'ampia varietà di corpora e software attualmente disponibili, soffermandomi in particolare su quelli liberi, o comunque gratuiti, maggiormente rilevanti per un traduttore.

Henry Tyne (Université de Perpignan Via Domitia)

Corpus use and data-driven learning in L2 French: from research issues to commonsense applications for the modern learner

This paper will present studies based on French L2 (and francophone contexts) which plead for the use of DDL. It will also give an overview of existing DDL work in/on French, providing examples and general recommendations. The use of corpus techniques and associated tasks in "commonsense" everyday DDL practices will be discussed, as will possible pitfalls and problems concerning French and francophone contexts.

possèdes autour de l'apprentissage des langues qui, tout en se raccrochant sur le
Les logiciels actuels d'apprentissage des langues sont, le plus souvent, peu pertinents pour le contexte contemporain d'enseignement et d'apprentissage. Ils sont conçus pour aider les apprenants à acquérir une langue étrangère, mais ne tiennent pas compte de la réalité de l'apprentissage des langues dans les contextes d'enseignement et d'apprentissage. Ils sont également moins efficaces que les méthodes traditionnelles d'enseignement et d'apprentissage.
Le présent article vise à démontrer que les logiciels d'apprentissage des langues peuvent être utilisés de manière efficace pour l'apprentissage des langues étrangères, en utilisant des méthodes et des stratégies d'enseignement et d'apprentissage qui sont pertinentes pour le contexte contemporain d'enseignement et d'apprentissage. Il propose également quelques recommandations pour l'utilisation des logiciels d'apprentissage des langues dans les contextes d'enseignement et d'apprentissage.

L'uso dei corpora e del data-driven learning nel francese L2: da questioni di ricerca ad applicazioni ragionevoli per l'apprendente moderno

Questa presentazione descriverà una serie di studi basati sul francese L2 (e contesti francofoni) che supportano l'uso del DDL. Fornirà anche una panoramica dei lavori esistenti sul DDL in e sul francese, presentando esempi e linee guida generali. Sarà discusso l'uso di tecniche basate su corpora e task “ragionevoli” per pratiche DDL quotidiane, così come saranno evidenziate possibili falle e problemi relativi al francese e a contesti francofoni.

Andrea Abel (Eurac Research, Bolzano)

Corpus-based Approaches for the Analysis and Description of Learner Languages: Ways and means

Those who work with second or foreign languages are also always dealing with learner languages. In recent years learner corpora have become an important resource for an empirically based analysis of learner languages.

The aim of the talk is to give an overview over possible strategies for a corpus-based description of learner languages. Particular attention will be given to the analysis and annotation of errors and other characteristics of interlanguages. The corpus-based work will be illustrated by means of concrete examples taken from the trilingual learner corpus MERLIN. MERLIN addresses the need for a concrete picture of the levels of the Common European Framework of Reference for Languages (CEFR) for Czech, German and Italian.

possèdent autour de l'apprentissage des langues qui, tout en se focalisant sur le
Les logiciels actuels d'apprentissage des langues sont, le plus souvent, peu performants pour donner un portrait de second langage learning and teaching, to identify major traits which the
can self-access facilitate play in language learning. These are two areas which the
two key variables affecting language learning styles, i.e., the general
variables for language teaching and language learning is available.

Approcci corpus-based per l'analisi e la descrizione delle lingue di apprendenti: modalità e strumenti

Chi lavora con le lingue seconde o straniere ha sempre a che fare con varietà linguistiche di apprendenti. In anni recenti, i corpora sono diventati una risorsa inestimabile per l'analisi empirica delle varietà linguistiche di apprendenti.

L'obiettivo di questa presentazione è quello di fornire una panoramica sulle possibili strategie per una descrizione basata su corpora delle varietà linguistiche degli apprendenti. Un'attenzione particolare sarà rivolta all'analisi e all'annotazione degli errori e di altre caratteristiche relative alle interlingue. Il lavoro basato su corpora sarà illustrato attraverso esempi concreti tratti dal corpus trilingue MERLIN. MERLIN risponde al bisogno di avere un quadro concreto dei livelli del Quadro comune europeo di riferimento per le lingue (QCER) per il ceco, il tedesco e l'italiano.

Luciana Forti & Stefania Spina

From Learner Corpus Research to Data-driven learning activities: some ready-to-use examples for the Italian L2 classroom

Corpora provide researchers with a wealth of new avenues into the observation of language use, and this has determined significant shifts in the analysis of both native and learner varieties. Learner Corpus Research (LCR), in particular, provides second language learning specialists with corpus-based evidence on how a language is learned. On the other hand, corpora provide learners with tools and resources that enable them to explore patterns of language use through what has become known as Data-driven learning (DDL). By coming into direct contact with corpus data, learners explore real language use through the guided-discovery activities prepared by the teacher, or in a freer and more exploratory manner.

This paper seeks to provide an overview of the most recent work conducted in the fields of LCR and DDL within the Italian L2 learning and teaching context. It will highlight the possible links between the two fields of research and practice, through the presentation of practical DDL activities for the Italian L2 classrooms, based on the empirical evidence related to learner needs.

possènes autour de l'apprentissage des langues qui, tout en se focalisant sur le
Les logiciels actuels d'apprentissage des langues sont, le plus souvent, peu performants pour l'enseignement des langues étrangères. Ils sont cependant très utiles pour aider les apprenants à identifier les principales caractéristiques de leur apprentissage. Ces logiciels peuvent également aider les apprenants à développer leur compétence en langage en leur fournissant des exercices et des activités adaptées à leur niveau. Ils peuvent également aider les enseignants à mieux comprendre les besoins et les intérêts des apprenants.
can self-access facilitate language learning? These are the questions which the
two key variables affecting language learning styles, i.e., the general
language learning and language learning environment.
language learning and language learning environment.

Dalla Learner Corpus Research alle attività di Data-driven learning: alcuni esempi pratici per la classe di italiano L2

I corpora forniscono ai ricercatori numerosissimi percorsi osservativi degli usi linguistici reali, e questo ha determinato innovazioni significative nell'analisi delle varietà linguistiche sia native che non native. La Learner Corpus Research (LCR), in particolar modo, fornisce agli specialisti di apprendimento delle lingue seconde evidenze basate su corpora in relazione a come venga appresa una lingua.

D'altra parte, i corpora forniscono agli apprendenti gli strumenti e le risorse che consentono loro di esplorare le regolarità negli usi linguistici attraverso l'approccio noto come Data-driven learning (DDL). Venendo in contatto diretto con i dati tratti dai corpora, gli apprendenti esplorano usi linguistici autentici attraverso attività di scoperta guidata preparate dall'insegnante, o in modalità più libera e autonoma.

Questa presentazione mira a fornire una panoramica dei lavori più recenti svolti nei campi della LCR e del DDL per contesti di apprendimento e insegnamenti relativi all'italiano L2. Metterà in luce le connessioni possibili tra i due campi, attraverso la presentazione di attività DDL da poter integrare nella didattica in classe, fondate sull'evidenza empirica relativa ai bisogni degli apprendenti.

Andrea Abel, Silvia Bernardini, Alex Boulton, Luciana Forti, Stefania Spina, Henry Tyne and the public.

Round table: The future of corpora in translation and second language learning

The aim of the round table is to bring speakers and public together in discussion. It will focus on some of the main challenges characterising the integration of DDL into translation and language learning practices today, by gathering insights and comparing experiences from those participating in the workshop.

possènes autour de l'apprentissage des langues qui, tour en se recouvrant sur ie
Les logiciels actuels d'apprentissage des langues sont, le plus souvent, peu perf
contemporain portrait of second language learning and teaching, to identify major tr
can self-access facilities play in language learning? These are questions which the
language style is play of language learning and secondly, it looks at the cont
two key variables affecting language learning styles, ie, the general
language to language teaching and language learning is available

Il futuro dei corpora nella traduzione e nell'apprendimento delle seconde lingue

L'obiettivo della tavola rotonda è di nutrire una discussione tra i relatori e il pubblico. Si focalizzerà su alcune delle principali sfide che caratterizzano oggi l'integrazione del DDL nelle pratiche traduttive e in quelle legate all'apprendimento linguistico, raccogliendo spunti e confrontando le esperienze dei partecipanti alla giornata di studi.

Selected bibliography

- Bernardini, S. (2000). Systematising serendipity: Proposals for concordancing large corpora with language learners. In L. Burnard & T. McEnery (Eds.), *Rethinking Language Pedagogy from a Corpus Perspective* (pp. 225–234). Frankfurt: Peter Lang.
- Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348–393.
- Boulton, A. & Tyne, H. (2014). Des documents authentiques aux corpus: démarches pour l'apprentissage des langues. Paris: Didier.
- Bybee, J. L. (2006). From Usage to Grammar: The Mind's Response to Repetition. *Language*, 82(4), 711–733.
- Chambers, A. (2007). Popularising corpus consultation by language learners and teachers. *Language and Computers*, 61(1), 3–16.
- Chen, M., & Flowerdew, J. (2018). Introducing data-driven learning to PhD students for research writing purposes: A territory-wide project in Hong Kong. *English for Specific Purposes*, 50, 97–112.
- Chiari, I. (2011). Teaching language variation using Italian corpora. In N. Kübler (Ed.), *Corpora, Language, Teaching, and Resources: From Theory to Practice* (pp. 301–323). Bern: Peter Lang.
- Corino, E., & Buschini, C. (2018). “*There’s a woman having her hair cut*”. A case study of data driven learning in a vocational school for hairdressers. Presented at the Teaching and Language Corpora Conference, University of Cambridge.
- Corpas Pastor, G. & Seghiri Domínguez, M. (eds). (2016). *Corpus-based Approaches to Translation and Interpreting: From Theory to Applications*. (Studien zur romanischen Sprachwissenschaft und interkulturellen Kommunikation, 106). Frankfurt: Peter Lang.
- Granger, S., Lerot, J. & Petch-Tyson, S. (eds). (2003). *Corpus Based Approaches to Contrastive Linguistics and Translation Studies*. Amsterdam: Rodopi.
- Johns, T. (1991). Should you be persuaded - Two examples of data driven learning materials. *Classroom Concordancing, English Language Research Journal* 4, 1–13.
- Kennedy, C., & Miceli, T. (2017). Cultivating effective corpus use by language learners. *Computer Assisted Language Learning*, 30(1–2), 91–114.
- Laviosa, S. (2002). *Corpus-based Translation Studies. Theory, Findings, Applications*. Amsterdam & New York: Rodopi.
- Lee, H., Warschauer, M., & Lee, J. H. (2018). The Effects of Corpus Use on Second Language Vocabulary Learning: A Multilevel Meta-analysis. *Applied Linguistics*.
- Mizumoto, A., & Chujo, K. (2015). A Meta-analysis of Data-driven Learning Approach in the Japanese EFL Classroom. *English Corpus Studies*, 22, 1–18.
- Polezzi, L. (1993). Concordancing and the teaching of ab initio Italian language for specific purposes. *ReCALL*, 5(09), 14–18.
- Sinclair, J. M. (2003). *Reading Concordances*. London: Pearson.
- Spina, S. (tbp). The development of phraseological errors in Chinese learner Italian: a longitudinal study.
- Tognini Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam-Philadelphia: Benjamins.
- Viganò, P. B. (2011). I corpora e il loro sfruttamento in didattica. *Italiano LinguaDue*, 3(2), 115–128.
- Zanettin, F., Bernardini, S., Stewart, D. (eds). (2003). *Corpora in Translator Education*. Manchester: St Jerome.