

Corpora and Applied Linguistics (IT-FR)

Marie–Pierre Escoubas Benveniste

Academic Researcher and Lecturer

Language and Translation studies–French language L–LIN/04

DIPARTIMENTO DI METODI E MODELLI
PER L'ECONOMIA IL TERRITORIO E LA FINANZA
MEMOTEF



SAPIENZA
UNIVERSITÀ DI ROMA

Corpora and Applied Linguistics (IT-FR)

1. A brief foreword
2. Defining corpus and applied linguistics
3. What corpus linguistics is applied to?
4. Some research areas of interest
5. Open questions and suggestions

Corpora and Applied Linguistics (IT-FR)

1. A brief foreword

“the actual enjoyment of fundamental rights (expression, education, participation in cultural life, benefiting from scientific progress) is highly conditioned by linguistic factors. Languages thus provide an **essential medium** for exercising those rights.”

(Unesco, 2007)

“ beaucoup de **langues européennes** pourraient être coupées de la communication numérique et devenir sans importance dans une **société en réseau**”

(*Livre blanc*, 2012, CNRS)

“...research has revealed positive impacts of multilingualism, including **greater cognitive flexibility and language control**, more **complex understanding of words**, [...] From an economic perspective, several studies [...] reveal that knowledge of more than one national language can have an **economic advantage** [...]”

(Unesco, 2019)

As an academic institution, Dipartimento Memotef promotes multilingualism :

4 european languages : Italiano, Español, Deutsch, Français + English

Corpora and Applied linguistics (IT-FR)

2. Defining Corpus and Applied linguistics

- **Language “data” in digital form.** Usually large amounts but **not necessarily “big (amount of) data”**.
- **Corpus** : a set of (subsets of) **selected items** (texts, text + sound, text + image, text + video) displayed in digital format and structure.
=> WWWeb ≠ Corpus
- **Applied linguistics** : when social needs encounter AL => acquiring new insights (language process vs language system)

=> virtuous loop : applied and theoretical ling. “feed” each other

Corpora and Applied Linguistics (IT-FR)

3. What is Corpus linguistics applied to?

- LSP empirical description
- Terminology Studies (monolingual/bilingual/multilingual)
 - Terminology
 - Term Variation (over time, text-genre, domain, users)
- Translation Studies (i.e. *Translationese* features)
- Teaching/Learning how to translate
- Language Teaching/Learning (L1, L2, L3)
- Understanding Language Risk (Health, Industry, ...)
- Fine tuning of resources and Computer Assisted Tools
- ...

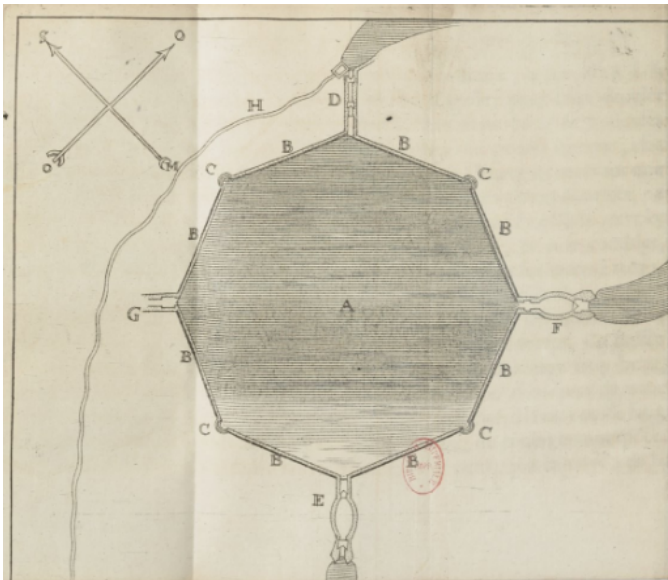
Corpora and Applied Linguistics (IT-FR)

4. Some research areas and projects

- **Monolingual corpus based term analysis**

2022: L' "épanchoir" of the Two Sees Channel (Occitany, FR), a term of high cultural content.

1681 (H: *L'épanchoir de la rigole*)



L'épanchoir des Patiasses © J.F. Benne, 2022



Corpora and Applied Linguistics (IT-FR)

4. Some research areas and former projects

- **Monolingual LSP corpus**

2009: “La définition dans le discours économique de vulgarisation savante”

- **Textual data analysis (JADT 2010 Roma, JADT 2012 Liegi)**

2010: Le “Grenelle de l’Environnement” (Le Monde)

Corpus based semantic analysis of a new term (cf. Sarkozy’s mandate)

2012: with O. Floquet (Dip. SEAI) and S. Bolasco (Dip. Me.Mo.T.E.F.)

Corpus based comparative quantitative and qualitative analysis in written contemporary French (gerund / present participle)

- **Comparable Corpora**

2013 Spoken and written French corpora for differential analysis (TIPA)

2016-2018- LSP : Comunicazione aziendale (IT FR) ongoing research

Corpora and Applied Linguistics (IT-FR)

4. Some research areas and projects

- **Bilingual corpora (FR-IT)**

2008 A bilingual corpus for LSP teaching (=> train-the-trainer)

2012 Contrastive analysis of a bilingual corpus of judgments of the European Court of Justice (=> train-the translator)

2018-2022: Studying how a bilingual corpus (FR-IT) may contribute to terminology network building (CA method)

Corpora and Applied Linguistics (IT-FR)

4. Some research areas

- **Learner Corpora and error analysis (=> train-the-trainer)**

2017 « La dictée de français et les étudiants italophones. Analyse préliminaire des erreurs de morphologie verbale ».

⇒ **2019 Corpus design**

⇒ **2020 NoVerFra Project** (prot. RP120172B88FC1FA)

Ongoing: **Building a digital resource for researchers and trainers interested in a better understanding of the acquisition of bi-modal French L2 verbo-nominal morphosyntax**, with Paolo G. Bottoni (Dip. Informatica)

Corpora and Applied Linguistics (IT-FR)

5. Open questions & suggestions

- **Research enhancement and orientation: suggestions**

- from **bi-** to **multilingual** applied linguistic research ?

(cf. CorMusT Project, with Di Vito)

- **Computer linguists skills** needed for data extraction & processing

- **Statistical analysis** : expertise needed in methodology

- **Domain experts collaboration needed** for validation/invalidation of textual data interpretations provided by corpus linguists

=> building reliable multilingual terminologies for translation

- **Research product valorisation : where and how?**

- Product = **resources** (i.e. train-the-trainer platform, data sets, ...)

- **Hosting on academic server** (where?)

- **Resource access** policy (of what sort?)

Corpora and Applied Linguistics (IT-FR)

Related bibliography

(**submitted**) coauthor with Condamines & Federzoni "Apport d'un corpus de presse spécialisé aligné français/italien à l'analyse des marqueurs et de la relation de méronymie" Actes du congrès ONELA Numéro spécial de ELA (Etudes de Linguistique appliquée) Amélie Josselin-Leray & Denyze Toffoli (eds.)

2022 "L' "épanchoir" du Canal des Deux Mers, un terme de haute teneur" in F. Attruia, A. Naccarato, A. Orlandi, C.Preite (éds), « *Nous plongeons nos mains dans le langage* ». Hommage à Paola Paissa, Coll. Au coeur des textes n°42, Editions Academia, Louvain-la-Neuve. pp. ISBN: 978-2-8061-0661-2

2021 coauthor with Condamines & Federzoni. Apport de la traduction dans l'étude des marqueurs de relations conceptuelles. Une étude en corpus aligné français-italien" In Cécile Frérot et Mojca Pecman (coord.) Des corpus numériques à la modélisation linguistique en langues de spécialité, `Collection « Langues, Gestes, Paroles » co-dirigée par Jean-Marc Colletta et Anne Vilain aux Éditions UGA (Université Grenoble Alpes).

2020 La fraseologia nella traduzione francese del Visconte dimezzato (1952) di Italo Calvino: verso un'analisi bidirezionale / Escoubas, Marie Pierre. - (2020), pp. 573-618. in Koesters at laii (a cura di) Si dice in molti modi - STUDI E RICERCHE STUDI UMANISTICI - ANTICHIstica.Sapienza - Editrice

2019 « L'apprentissage de la morphologie du nom et du verbe français par des italophones : un corpus de contextes pour l'expérimentation », *Repères DoRiF*, n. 20 – Modèles linguistiques et cognitifs et didactique des langues, DoRiF Università, Roma, dicembre 2019

2018 with Condamines & Federzoni. (2018) Usare i corpora allineati per migliorare la costituzione di "reti terminologiche" e l'insegnamento della traduzione. *International Conference of the Italian Association of Applied Linguistics*, Feb 2018, Università degli studi di Roma Tre, Rome, Italy

Corpora and Applied Linguistics (IT-FR)

Related Bibliography

2017 with S. Di Domenico « La dictée de français et les étudiants italophones. Analyse préliminaire des erreurs de morphologie verbale ». In Gerolimich et al. (éds.), *Comment les apprenants étrangers s'approprient-ils la morphologie verbale du français ?*, Bulletin suisse de linguistique appliquée, N° 105, 2017, pp. 77-94, ISSN 1023-2044**2013**

2012: Prédicats juridiques et schémas d'arguments dans les textes des arrêts de la Cour : approche bilingue français-italien“

2013 « Usages du gérondif et du participe présent en français parlé et écrit : étude comparée basée sur corpus », *TIPA. Travaux interdisciplinaires sur la parole et le langage* [En ligne], [29 | 2013](#). URL : <http://tipa.revues.org/908>

2010 “Repérer et s'approprier le lexique spécialisé grâce aux TIC: défense et illustration “, in *Multimedialità e linguaggi specialistici: riflessioni teoriche e iniziative didattiche*, In Rita Salvi (a cura di) RILA, p. 247-262.

2010 “Le Monde et le « Grenelle de l'Environnement » : pistes pour l'analyse sémantique assistée par ordinateur d'un corpus de presse“, In *Proceedings of 10th International Conference Journées d'Analyse statistique des Données Textuelles*, JADT 2010, Sapienza, University of Rome. Roma, 9-11 giugno 2010, a cura di Sergio Bolasco et al., Milano: Edizioni universitarie di Lettere Economia Diritto, vol. 1/2010, p. 409-422.

2009 “La définition dans le discours économique de vulgarisation savante”, In Micaela Rossi (coord.) *Autour de la définition 6*, Publif@rum, Università degli Studi di Genova. ISSN 1824-7482, codice: E186699, Rivista: PUBLIF@RUM
http://publiforum.farum.it/ezone_articles.php?id=146

2008 “Le corpus français-italien LSSEE (Langue Spécialisée du domaine “Systèmes éducatifs européens”)”, Marie Hédiard (a cura di), *Corpus et logiciels pour la linguistique*, [CD-Rom], Centro editoriale di Ateneo, Università degli Studi di Cassino.

2003 “(-)mener ou (-)porter ? Avantages d'une approche lexicale de la traduction de portare, verbe de déplacement à sens concret”, *RILA-Rassegna Italiana di Linguistica Applicata XXXV*, 1-2, Milano: Bulzoni Editore, 2003, pp.127-149. ISSN 0033-9725.

2002 “An ou année ? Un exemple d'analyse contextuelle guidée par corpus comme aide à la sélection lexicale pour les apprenants italophones”, in Marie Hédiard (coord.) *Le nuove tecnologie nella ricerca in linguistica e glottodidattica*, Pubblicazioni dell'Università degli Studi di Cassino, pp.41-74.