

Press release



L'AVENIR EST FAIT D'AUDACE

Erganeo and Emvista sign a license for the use of a unique language resource.

In March 2022, Erganeo signed a technology license transfer agreement with Emvista, a software company, concerning a linguistic resource: the French Tree Bank.

The analysis and reproduction of natural language requires a clear understanding and structuring of the meaning of the sentence. The *French Tree Bank* corpus (1996-2016) is a unique lexical and syntactic resource, offering 664,500 tokens (words) and 21,500 sentences richly annotated (categories, groups and syntactic functions) by linguists. In particular, it addresses problems related to Automatic Natural Language Processing (NLP), the semantic web, the integration of language in a human-machine dialogue search engine (chatbots) or even spell checking. The corpus is used in several hundred laboratories around the world for research or automatic processing of French.

The *French Tree Bank* corpus was created from articles taken from the French daily newspaper, *Le Monde*, and was registered by Université Paris Cité, which entrusted Erganeo with the development of this linguistic resource. The technology, developed by researchers Anne Abeillé (professor, UMR 7110 CNRS Université Paris Cité) and Lionel Clément (doctoral student and lecturer at the University de Bordeaux), led to a patent application in 2008.

The *French Tree Bank* has now been transferred to Emvista, a Montpellier-based software company, which is developing a state-of-the-art technology for analysing all textual content and structuring the information conveyed in it. This information structuring technology is based on numerous linguistic indicators, including those annotated in the *French Tree Bank* corpus. These indicators will then be used within learning and knowledge-based systems. The *French Tree Bank* is also a reference resource in the academic world and is therefore indispensable to the Emvista R&D team when it wishes to compare the performance of the systems being tested.

Currently, the corpus is distributed free of charge for research purposes, and a trial version is available to all on the *French Tree Bank* website.

(<http://ftb.linguist.univ-paris-diderot.fr/>)

About Erganeo – www.erganeo.com

Erganeo is a French investment fund that specialises in breakthrough innovations (deep tech) with a major societal impact. We invest early in the game to secure researchers' newest inventions before they are transferred to companies or before start-ups are founded, across a wide range of scientific fields: Biotech, Infotech (Telecom, connected objects, big data, AI) and Eneritech (new energy, chemistry, materials), among others. Erganeo aims to accelerate and simplify links between research and industry for the benefit of society. To do so, we finance and accompany the new generation of French researchers and entrepreneurs on the road to international recognition and success. As a member of the SATT Network, Erganeo draws on the Ile-de-France network in building the foundations of a better tomorrow, tapping into a talent pool of over 20,000 researchers across 350 cutting-edge, leading laboratories. Since it was founded, Erganeo has invested over €41m, thus contributing to the signing of 88 licensing agreements with companies of all sizes and setting up 31 start-ups

Press contact : Caroline Pontifice
Communication and Marketing Manager
caroline.pontifice@erganeo.com

About Emvista - www.emvista.com

Emvista is a software publisher. Founded in Montpellier in 2018, and with a strong R&D team in Automatic Natural Language Processing, the company offers innovative products based on state-of-the-art technologies. Its flagship product, the email management assistant, enables its users to be more efficient in managing their emails by locating all relevant information in the mass of information received (e.g. requests for information and actions that require action to be taken). Emvista markets a number of other text analysis services such as opinion and emotion analysis, anonymisation of sensitive texts and entity extraction.

Press contact: Cédric Lopez - Research Director
cedric.lopez@emvista.com