Sujet de Thèse
Contrat Doctoral de la SFR Agorantic
"Culture, Patrimoines, Sociétés numériques"

Directeur de thèse Rachid El-Azouzi, PR, Laboratoire d'Informatique d'Avignon
Co-directeur: Gonzalez Madelena, PR au Laboratoire Identité Culturelle, Textes et Théâtralité (ICTT); Eitan Altman, DR1, INRIA Sophia Antipolis.

Intitulé du projet de thèse : Nouveaux outils pour l'analyse de l'évolution des styles dans la littérature anglaise

Titre en anglais : New linguistic tools for the analysis of the evolution of style in English literature

Mot-Clés : Evolutionary games, English literature, evolution of writing styles, Information theory, Data analysis, Evolutionary dynamics.

Co-tutelle : Non

New linguistic tools for the analysis of the evolution of style in English literature Artistic styles in general, and in particular styles in literature are part of the cultural heritage of humanity. An artistic style can be viewed as a language common to the artist and understood by the art consumers, which has its own evolutionary rules. Yet, while there has been a lot of work within different communities (e.g. linguists, computer scientists and biologists) on the evolution of spoken and of written languages using these tools **[R1-R5]** and on the evolution of culture **[R6]**, there has been little work **[R8,R9]** on the evolution of styles and none using the tools that we propose.

The objective of this thesis is to develop novel tools to identify and to understand the evolution of styles in literature as well as the coevolution of various styles that coexist. These will include tools from different fields such as computer science, information theory and evolutionary game theory. The thesis will further use these tools to classify English prose and poetry within existing styles and will develop tests to identify style innovations, i.e. creation within a new literary style. The main objective is to create mathematical and computational models grounded in sociology, linguistics and evolutionary dynamics in order to model and understand the dynamics of evolution of style in English literature at a lexical and linguistic levels

An important challenge would be to develop tools for postmodern literature style, in which styles evolve much faster and where the author is expected to create his own style.

The work will be highly interdisciplinary and will therefore require the participation of an interdisciplinary team of advisors: Rachid El-Azouzi (expert in computer science and evolutionary game theory) Eitan Altman (expert in information theory, evolutionary games theory and in styles in music) and Madelena Gonzalez (expert in analysis of English literature).

Why and how to use evolutionary game theory in this thesis. Evolutionary games have been introduced by biologists to explain evolution of species. In his book "the selfish gene", the biologist

Richard Dawkins advocates the use of this branch of game theory for explaining the evolution of cultural phenomena as well. It is thus natural that classical evolutionary game theory has been used in evolution of languages so as to answer what motivates the introduction of new words, and what are the features of a word that favors its adoption in a language. Yet the authors of **[R7]** have recently proposed an alternative theory for evolution among societies which is much more adapted to cultural phenomena. The thesis will therefore attempt to use not only classical evolutionary game theory but also the novel approach of **[R7].**

Why and how to use information theory concepts. In **[R3]** the authors show that there are some mathematical objects (the entropy, which is a central object in information theory) that are invariant over different languages (in fact different linguistic families). In a similar way, the thesis will try to identify invariants within and across styles.

Use of data sets in the thesis. The thesis will make heavy use of data sets available at google N-gram that provides digital data on occurrence and co-occurrence of words from millions of books.

References

[R1] Nowak, M. (2000). Evolutionary biology of language. *Philosophical Transactions Royal Society B: Biological Sciences,* 355, 1615-1622.

[R2] Nowak, M. and D. Krakauer (1999). The evolution of language. In: *Proceedings of the National Academy of Sciences of the United States of America,* 96,8028-8033.

[R3] Montemurro MA, Zanette DH (2011) Universal Entropy of Word Ordering Across Linguistic Families. PLoS ONE 6(5): e19875. doi:10.1371/ journal.pone.0019875

[R4] Proceedings of the Workshop on Language Variation and Change and Cultural Evolution Feb. 2015, Univ of York, Department of language and linguistic Science

[R5] Twitter and the Geo-linguistic fingerprint, Eitan Altman and Yonathan Portilla, Proceedings of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), August, Istanbul, August 2012

[R6] Pietarinen, A.-V. (2005). Evolutionary game-theoretic semantics and its foundational status. In: *Evolutionary Epistemology, Language and Culture: A Nonadaptationist Systems-Theoretical Approach* (N. Gontier, *J.* P. Van Bendegem and D. Aerts, eds.), pp. 429-452. Springer, Dordrecht.

[R7] **Altruism in Groups: An Evolutionary Games Approach,** Ilaria Brunetti, Eitan Altman, Rachid El-Azouzi, *7th International Conference on NETwork Games COntrol and OPtimization (NETGCOOP 2014)*, Oct 2014, Trento, Italy. 2014

[R8] Formula follows the evolution of writing style, The New Scientist, see https://www.newscientist.com/article/dn21767-formula-follows-the-evolution-of-writing-styles/

[R9] Quantitative patterns of stylistic influence in the evolution of literature, James M. Hughesa, Nicholas J. Fotia, David C. Krakauerb,c, and Daniel N. Rockmore, PNAS vol. 109 no. 20 > James M. Hughes, 7682–7686, doi: 10.1073/pnas.1115407109