

ABSTRACT: Reading Darwin between the Lines: a Computer-Assisted Analysis of the Concept of Evolution in *The Origins of Species*.

SAINTE-MARIE, Maxime

Doctorate, Cognitive Computer Science

Laboratoire d'Analyse Cognitive de l'Information (*Cognitive Information Analysis Laboratory*)

Université du Québec à Montréal

MEUNIER, Jean-Guy

Department of Philosophy

Laboratoire d'Analyse Cognitive de l'Information (*Cognitive Information Analysis Laboratory*)

Université du Québec à Montréal

Simultaneously praised and condemned by both clergymen and scientists, the Darwinian theory of modified descent by means of natural selection marks the birth of a radically new and modern conception of nature, life, science and man, built on the revitalization and reworking of an old biological concept: *evolution*. While Darwin's first major work, *On the origin of species by means of natural selection or the preservation of favoured races in the struggle for life*, is generally considered as the birth document of this theoretical revolution, studies on and around this book often overlook the fact that the word *evolution* itself is rarely used by Darwin.

In the first (November 24th 1859), second (January 7th 1860), third (March 1861) and fourth (June 1866) editions, there is only one occurrence related to the term evolution: it is the last word of the conclusion of the work, *evolved* (1st & 2nd: p. 490; 3rd: p. 525; 4th: p. 577;). In the fifth edition, the same term *evolved* appears a second time, the first occurrence appearing in the fourteenth chapter (p. 573) and the second at the same last spot as in the earlier editions (p. 579). Surprisingly, only in the sixth and last edition are the term *evolution* and its derivatives more extensively and systematically employed: *evolutionists* (Chapter VII: pp. 189, 201), *evolve* (Chapter VII: p. 191), *evolved* (Chapter VII: pp.191, 202 (2); Chapter VIII: pp. 425,429), *evolution* (Chapter VIII: pp. 201(2), 202; Chapter VIII, p. 215; Chapter X, p. 282; Chapter XV, pp. 424, 424, 424).

This lexical outbreak of the word *evolution* in the last edition, while contrasting with its lexical scarcity in the earlier editions, doesn't necessarily mean however that the concept of evolution wasn't present in the first five printings of the book. Said differently, the word *evolution* used by Darwin in the sixth edition of *The Origins of Species* refers to a vast and complex conceptual network – including notions and themes such as descent with modification, natural selection, adaptation, variation, reproduction, heredity, fitness, survival, etc.– whose definitive configuration may already have been present and systematically referred to in the first five editions, independently of any proper denomination.

Such conceptual and lexical considerations, while emphasizing the distinction between the semantic associations of specific concepts and their embodiment in natural language, also seems to imply the possibility of “reading between the lines”, that is, of identifying concepts on the sole basis of their relations with other concepts, without any appeal to their proper linguistic designations. This specific issue will be addressed by means of a computer-assisted analysis of the concept of evolution in *The Origin of Species*: by means of automatic classification, categorization and annotation strategies and applications used in text mining, a new and automated approach for “reading Darwin between the lines” will be proposed, aiming to identify where the author “conceptually” refers to evolution, regardless of the presence or the absence of the word itself. This technologically-assisted attempt, while bringing new insights in the understanding of the concept of evolution, might also shed new light onto the conceptual analysis of the Darwinian theory itself and on digital philology, hermeneutics and text interpretation in general.