COMPUTATIONAL LEXICOLOGY IN THE PUBLIC EYE

CĂTĂLIN DEHELEAN

Abstract

This article is all about purpose and means. The wide public of Linguistics may be interested only in finding out the latest trends. Therefore more than a few words about the Computational Linguistics Terminology may not be absolutely necessary. They can be found in short online free non-expertly written articles. Parts of the public disenchanted with the amount and nature of such information may want to re-orient and read general content books on Computational Linguistics. However, if the latter do not offer the required information, one can always study specific content books on Computational Lexicology.

Keywords: *Computational Lexicology, Free Resources, Specialised Resources*

1. Introduction

It goes without saying that the public interested in the current and thus future direction in Lexicology should attempt an introduction into Computational Lexicology. To achieve just that one must access resources which bear an immediate relationship to this topic. There are both the free and specialised resources one may access, which tend to offer a rather general perspective of Computational Lexicology.

2. Free Resources

For the purposes of this paper the free resources are understood to mean online material which is available in full to the general public. Free resources offer a partial view on the topic which might

be perceived as useful as means of introducing any reader to the topic. The two most obvious free resources are English Wikipedia Project and ACLWiki Project.

The English Wikipedia Project article on Computational Lexicology[1] and the ACLWiki article on Computational Lexicology[2] seem to be, for the most part, identical.

They both debut with a definition. According to the English Wikipedia Project "Computational lexicology is that branch of computational linguistics, which is concerned with the use of computers in the study of lexicon"; the ACLWiki on the other hand is slightly more specific and states that "Computational Lexicology is the use of computers in the study of the lexicon". This definition stems according to both articles from a Doctoral dissertation by R.A. Amsler[3].

This definition is then elaborated upon. The said Wikipedia article explains that "It is distinguished from *computational lexicography*, which more properly would be the use of computers in the construction of dictionaries, though some researchers have used computational lexicography as synonymous." The ACL Wiki article provides a virtually identical piece of information about Computational Lexicology: "It is distinguished from Computational Lexicography, which more properly would be the use of computers in the construction of dictionaries, though some researchers have used Computational Lexicography as synonymous with Computational Lexicology."

Henceforth, the two wiki articles move on and state that they will present the history of Computational Lexicology. The degree of success in the matter is rather questionable. The information provided by the English Wikipedia Project article on the history of Computational Lexicology is limited to a short paragraph where only a few landmarks are mentioned, namely the situation in the 1960s

[1] http://en.wikipedia.org/wiki/Computational_lexicology
[2] http://www.aclweb.org/aclwiki/index.php?title=Computational_Lexicology
[3] Amsler, R.A. (1980). *The Structure of the Merriam-Webster Pocket Dictionary*, Doctoral Dissertation, TR-164, University of Texas, Austin.

and the present-day situation. The ACLWiki article, on the other hand, sums up the history of Computational Lexicology somewhat more comprehensively mentioning, besides the achievements of the pioneering years and the contemporary developments, the research of the 1980s.

But under the idea of history, both articles present a topic which is related to the development of Computational Lexicology, but not which is not a description of its development. This topic is related to the characteristics of print dictionaries. The Wikipedia article is straight forward on the matter. The ACLWiki article is not so straightforward in an attempt to explain its relevancy to the history of Computational Lexicology. Both articles then end the part on the characteristics of print dictionaries with paragraphs which mention the need for further development. The two paragraphs are identical: "Many computational linguists were disenchanted with the print dictionaries as a resource for computational linguistics because they lacked sufficient syntactic and semantic information for computer programs. The work on computational lexicology quickly led to efforts in two additional directions."

The sequel thus set, one is going to read in both articles about a questionable topic called "Successors to Computational Lexicology". Each article elaborates on the topic in four identical paragraphs. The paragraphs are built on two basic ideas. The first idea is the "role that corpora played in creating dictionaries". The second idea is "the creation of Lexical Knowledge Bases".

The slight differences between the two articles provide, however, the public with welcome new pieces of information. Such is the case of the English Wikipedia Project article on Computational Lexicology. Its text, unlike that of the ACLWiki article, does not end with the history of Computational Lexicology. It features instead a third section on standardization of lexicons. It briefly mentions the standard ISO/TC37 and its sequel.

3. Specialised Resources

While the nature of information in free resources was relatively easy to grasp specialised resources are somewhat more problematic. They are rarely, if ever, available freely or completely online. One is bound to get accustomed to their printed versions, that is to say, with books. Specialised resources may offer a broader view on Computational Lexicology but this requires the perspective reader to go deeper into the topic.

There are two types of specialised resources. The first type is comprised of general purpose books where one is supposed to be told as much as possible about Computational Linguistics. The second type refers to the specific content books which are supposed to narrow down the perspective on a certain part of Computational Linguistics, in this case on Computational Lexicology.

3.1. General Content Books

The first attempt of any dedicated reader to find information on Computational Lexicology means coming into contact with information provided by general content books on Computational Linguistics. The attraction of the public to general content books lies in the almost encyclopaedic nature of such works, which are, not without reason, seen as landmark books on Computational Linguistics. It may be thus assumed that readers taking this path do so since they believe that the nature general content books of may provide an easy-to-understand description of what they are looking for.

The interested public is said to expect a lot from these landmark books, but such an expectation should be realistically limited since they often overlook the necessity of defining Computational Lexicology, focusing instead on resources and/or Computational Lexicography. One is bound to encounter such limitations even before the act of reading the text of a landmark book, when one is in the process of getting acquainted with the table of contents. Having this in mind, it is about time to see how general content / landmark books on Computational Linguistics present Computational

Lexicology. Two such books are "The Oxford Handbook of Computational Linguistics"[4] and "Computerlinguistik und Sprachtechnologie"[5][6].

In "The Oxford Handbook of Computational Linguistics" Computational Lexicology is a term which is not mentioned as such. However there is a chapter called "Lexicography" made up of several parts. This chapter mostly refers to Computational Lexicography, where one is offered a view on constructing dictionaries. The first part is an "Introduction" which defines Computational Lexicography from two perspectives. In the next part, it presents a "Historical Background" which is used to speak about "Deficiencies", the "Usability of lexicon", "Machine-readable dictionaries" and, "Corpus-based dictionaries". Then it speaks about "Restructuring and Exploiting Human Dictionaries for Computational Reasons", "Dictionary structure", "Using Computational techniques to compile new dictionaries (with "Challenges of corpus-based lexicology", "Corpus-based revision", "WordNet"), "Linking meaning and use", "Exploring the future". However, as Computational Lexicology and Computational Lexicography exhibit a mutual affinity, the reference to lexicological resources is unavoidable and actually occurs in the last part of the chapter called "Further reading and relevant resource".

A partially similar perspective on Computational Lexicology is presented in "Computerlinguistik und Sprachtechnologie". The third chapter of this book is called "Methoden", i.e Methods. It describes in several parts the branches of Linguistics: Phonology, Morphology, Syntax, Semantics, Pragmatics, and, Text Generation. But it has no part dedicated to explaining the theoretical foundations of Computational Lexicology.

---

[4] Mitkov, R. (2004). *The Oxford Handbook of Computational Linguistics*. Oxford University Press: Oxford, UK

[5] "Computerlinguistik und Sprachtechnologie" roughly translates "Computational Linguistics and Speech Technology"

[6] Carstensen, K.-U., Ebert, Ch., Endriss, C., Jekat, S., Klabunde, R., Langer H. (2004). *Computerlinguistik und Sprachtechnologie*. Spektrum Academisher Verlag: München

"Computerlinguistik und Sprachtechnologie" does however contain a chapter which presents lexical resources. On is bound to find definitions, classifications and other pieces of information on Text Corpora, Tree Banks, Lexical-Semantic Word Networks, Lexicons for Multimodal Systems, Speech Databases, Non-Linguistic Databases, The World Wide Web. This chapter is indeed helpful for any person interested in where Computational Lexicologists look their words up, feed them in to, and / or provide feedback to, thus helping the development of these resources.

On reading these otherwise comprehensive books, one is under the impression that Computational Lexicology either does not exist or it is simply a matter of writing dictionaries. But one needs to remember that these books mention lexicological resources without placing them in the immediate field of writing dictionaries. It goes without saying that the study of lexemes, regardless the means and perspective, is the part of the field of Lexicology. Therefore the views these books offer on Computational Lexicology are highly implicit, which makes it very difficult for interested parties to understand them.

In the end one is left with the paradox of general content books, which is that, even though they offer preciously little in the sense of the disambiguation of the term Computational Lexicology, since they are perceived as landmark books, they are more likely to be read.

3.2. Specific Content Books

They offer a more precise if rather academic outlook on Computational Lexicology, its resources and applications and of course. Specific Content Books are the one to go to in order to gain a better appreciation of the relationship between Computational Lexicology and Computational Lexicography.

An example of a specific content book is "Computational Approaches to the Lexicon"[7]. While this title suggests the study of

[7] Atkins, B.T.S., Zampoli A. (1994). *Computational Approaches to the Lexicon*. Claredon Press: Oxford, UK

Computational Lexicography rather than that of Computational Lexicology, the book is almost immediately offering an interesting perspective.

Upon reading the introduction of this book one is presented with definitions which parallel descriptive and computational approaches to Lexicology and Lexicography while relating the ideas of Lexicology and Lexicography.

As such one is bound to find out that Lexicology is considered a part of Linguistics which mainly deals with studying the lexicon, Computational Lexicology being mostly about using computers in achieving the same purpose, although the fact that it has larger resources complicates matters.

At the same time Lexicography is said to be about collecting lexical material while Computational Lexicography is said to gather lexical material with the help of computers, which of course means more comprehensive collections.

Just as noteworthy is the observation that both Computational Lexicology and Computational Lexicography are experiencing a genuine boom.

While the introduction to this book is truly revealing, much like the rest of the book which is not discussed here, there is a shortcoming. The drawback of specific content books is that by their specific nature and academic style, they are not commonly consulted by the public.

Conclusions

The advantages and disadvantages at every step of the way are connected to the amount of time and effort one is able or willing to invest in this endeavour. Realistically speaking the broadest part of the public is not going to go beyond the free sources. Most of those who are willing to go beyond are going to settle for the information in general content books on Computational Linguistics. Only a few are going to read specific books on Computational Lexicology.

To avoid the spread of imprecise definitions and highly subjective views on Computational Lexicology this article has two

suggestions to make. The first suggestion addresses the public. It is advised to strive against the common practice of reading what is freely available online and find any necessary information in specific content books. The second suggestion is more of an appeal to linguists. Their help is instrumental in the expert checking and re-writing of free sources as well as in suggesting ways of reconsidering the position of Computational Lexicology in general content books on Computational Linguistics.

Bibliography

Amsler, R.A. (1980). *The Structure of the Merriam-Webster Pocket Dictionary*, Doctoral Dissertation, TR-164, University of Texas, Austin.

Atkins, B.T.S., Zampoli A. (1994). Computational Approaches to the Lexicon. Claredon Press: Oxford, UK

Carstensen, K.-U., Ebert, Ch., Endriss, C., Jekat, S., Klabunde, R., Langer H. (2004). Computerlinguistik und Sprachtechnologie. Spektrum Academisher Verlag: München

Hausser, R. (2000). Grundlagen der ComputerLinguistik. Springer: Heidelberg

Mitkov, R. (2004). The Oxford Handbook of Computational Linguistics. Oxford University Press: Oxford, UK

http://en.wikipedia.org/wiki/Computational_lexicology

http://www.aclweb.org/aclwiki/index.php?title=Computational_Lexicology

LEXICOLOGIA COMPUTAŢIONALĂ ÎN CONTACT CU PUBLICUL LARG
(Rezumat)

Acest articol dezbate ideea de atingere a unui scop. Publicul larg care intră în contact cu Lingvistica este adesea interesat de a afla ultimele realizări în domeniu. Când vine vorba despre Lingvistica Computaţională, terminologia este uşor de aflat. Ea este expusă în articole online neverificate de specialişti. Dacă publicul nu este mulţumit de aceste informaţii se poate re-orienta spre cărţi care tratează Lingvistica Computaţională în general. Dacă informaţia căutată nu se găseşte nici în aceste surse, publicul este invitat să apeleze la cărţi specializate pe Lexicologie Computaţională.