Universitatea „Babeş-Bolyai”Cluj-Napoca
Facultatea de Litere
Caterdra de Limba şi Literatura Engleză

The Human Algorithms and the Computer Language

Summary

Coordonator Ştiinţific:
Prof. Univ. Dr. Mihai M. Zdrenghea

Doctorand:
Cătălin Dehelean

Cluj-Napoca 2010

The Human Algorithms and the Computer Language

(PhD Thesis Summary)

Table of Contents

Introductory arguments

Linguistics is a field of great opportunity to anyone interested, as different approaches can and often do bring to light new ways understanding the language phenomena. Structural linguistics has been doing us a great service by trying to accurately describe these phenomena with impressive results.

However, these descriptions rely on the assumption that language is a human only occurrence, and thus it is used only for inter-human communication. While it is true that language came to exist inside the human mind and it has been used to exchange content for the better part of the human history, things have been slowly turning in a new direction since the second half of the twentieth century.

With the advent of powerful computation machines, the new challenge of exchanging content between man and machine has increasingly gained weight. Here we can easily see the conceptual mutation from describing the exclusively human language phenomena to describing machine language phenomena as well, and the process of exchanging content between the two. The description of this newly established/newly accepted relationship would however stretch structural linguistics to its limits and beyond.

A new orientation in linguistics has slowly come into being to bridge this gap. Computational linguistics is basically meant to describe human – machine interaction in a rather clever way. Instead of describing human language and the artificial language of the machines separately, it simply describes human language using logical patterns which are machine friendly. Machines are then required to replicate and use these patterns when engaging in a dialogue with a human interlocutor.

The success of this language acquisition is measured in degree of human understanding. Failures to render meaning are analysed and the physical limitations of the human interfaces pointed out, so that future versions may contain extra features thus improving dialogue. But computational linguistics steers clear of computer programming.

It is therefore important to highlight that Computational Linguistics is intrinsically human language oriented. The following parts chapters will be an attempt to demonstrate this last statement by showing that Computational Linguistics has its own history, pursues language levels in its unique way and it is able to observe human-machine interaction independently.

Once the need for an introduction is satisfied the lecturer of this work will be met with information in a manner which allows for a guided tour into the depths of Computational Linguistics and its application. Before reading this work, one should expect a traditional structure of three chapters followed by conclusions.

After reading the three chapters with all their parts one will be met by some conclusions. This part will recap the information presented in the three chapters. Then it will remind the reader of the claim of this work, namely the questioned identity of Computational Linguistics, and it will invite the reader to consider it according to the aforementioned information.

Chapter 1: A Brief Overview of Computational Linguistics

The purpose of this chapter is to try to present the evolution of computational linguistics objectively, by trying to balance several views related to this topic. This chapter is therefore be a work of synthesis.

This chapter tries to focus on the development of the science called Computational Linguistics, although the journal called Computational Linguistics is undoubtedly an important part of the latter, as it was and still is one of the most important means of spreading the latest achievements in this field.

There are two chronological approaches to the evolution of Computational Linguistics. These approaches try to answer the question when had Computational Linguistics

appeared and when it ended if it did so at all. They both start from the idea that the science called Computational Linguistics is ultimately a product of the general effort to develop automated translation systems in the 1950s. But then they diverge as to when or if Computational Linguistics ceased to be an active field of research. This chapter suggests that Computational Linguistics has indeed come into being in the 1950s and has been progressing ever since.

This proposed history is meant to be ordered in a logical-chronological manner as it is fairly easy to see it as a sequence of related theories and their development. This history starts with a description of the ideas immediately leading to the emergence of Computational Linguistics and end with contemporary tendencies.

This first chapter was meant to introduce the reader to Computational Linguistics. It literally presents basic notions and ideas related to this field. The birth and development of Computational Linguistics was shown to be inextricably linked to the development of Speech Technology. But while Speech Technology was advanced by means of Natural Language Processing, a field pertaining to Information Technology and thus to Computer Programming, Computational Linguistics was developed in the spirit of Linguistics. The first chapter thus showed that, throughout its development, Computational Linguistics acquired an identity of its own and was marked by an association and even a periodical, but what is more important by a unique methodology.

Having in mind everything that has been written so far one can go one step further. One can try and tackle questions which have, so far, been left unanswered. In the next chapter the reader is invited to get familiarised with the said methodology.

Chapter 2: Computational Approaches to Language Levels

This chapter constitutes the theoretical backbone of the present work, but not an island in itself. The overview of Computational Linguistics is by its very nature scanty and requires further development as far as specialised information is concerned. It may seem that Logics dictates that the following part is meant to fill in just such a gap.

Unsurprisingly however, the story doesn't end here. Theory in itself is not necessarily the only other argument when proving a higher point, especially when that point is the acceptance of Computational Linguistics as a field of Linguistics. As a consequence of this matter, it is necessary that the readers enjoy a follow-up, that is a third chapter, which should present the details of at least one application of Computational Linguistics, which derives to a certain extent from the second one.

Returning however to the current issues it seems only fair to warn the reader of the extraordinary structure and, thus, size and nature, of the second chapter when compared to the other two. Instead of the usual sections which are common throughout the other two chapters one is bound to find larger parts which in turn consist of sections. This particular arrangement was deemed necessary in order to achieve gain in understanding at the price of breaking the fluency of the text.

In this light, one can say that the second chapter is made up of five contiguous parts. The criterion for this arrangement was the attempt to order language levels. Although in the course of the research for this work one has come across several ways of ordering language levels, for the sake of a better comprehension and to rule out the risk of omissions, it was deemed fit to use a rather traditional order. The five contiguous parts therefore present issues on Computational Phonology, Computational Lexicology, Computational Morphology, Computational Syntax and Computational Semantics.

These parts however, exciting as they may sound, are themselves but introductions. They contain both elements which pertain to Descriptive Linguistics and to Computational Linguistics proper.

2.1 On Computational Phonology

The actual debut of chapter two means an introduction into the first language level, phonology. It may not necessarily seem a novelty of deep importance. But given the

computational perspective and the thesis of this work its presence and placing is actually vital.

This first part has presented the defining properties of different types of phonologies. The approach has been situated with respect to current and past developments in phonology and computational linguistics.

In the next part of this chapter the focus will switch to the next level. It is interesting to see how the same computational perspective ushers in new exploits for lexicology.

2.2 On Computational Lexicology

This part is meant to open a discourse on Computational Lexicology. It offers general widely-available information. Although structured, this introductory section does not penetrate deep into the subject matter, rather presenting a few of the concepts further elaborated on. It tries to give an outlook of the discipline; it suggests a definition and it attempts to mark the relationship between Computational Lexicology and Lexicography.

Reading the second part of the second chapter, may seem a bit difficult but with a bit of patience one will discover that the already generous level of lexicology may include even more resources.

The age old issue of finding ones words in an organized, well-described fashion is being discussed. Different criteria of putting together lexical material such as texts in Text corpora, dependency trees in Treebanks, semantically related lemma in Lexical-Semantic Word Networks, and the online material of the World Wide Web are being discussed.

The part on Computational Lexicology is however connected to the next ones. Lexicology is offers a perspective on words, yet it uses methods loaned from the fields studying the upper language levels. In the next part, our attention span shall be broadened by Computational Morphology.

2.3 On Computational Morphology

The purpose of this chapter is to get an insight into computational morphology and its applications. It is therefore desired and recommended that a strong theoretical background is to be laid down for a better understanding of the description of applications.

The first part of this chapter is dedicated to the basics. It explains terminology and lists the main problems which are likely to be met throughout the whole chapter. All these issues are arranged inclusively, that is to say from the general to the specialized.

Chapter two part three was meant to show the public the correlation between the descriptive and the computational view on morphology on one hand, and the computational methodology in the approach of morphology on the other. The longest and most interesting parts were those describing Models from Generative Linguistics and the Default Root Networks: DATR.

However, in the end of this third part, we are left with an interesting but rather fragmentary image of the language levels. All this is about to change for the next part is providing us with the extra information we need on how Computational Linguistics views syntax.

2.4 On Computational Syntax

The purpose of chapter is to create an image of computational syntax. The main issues will of course be the distinctions, methods and history.

The task of the first part is to find the borderline between the computational morphology and computational syntax. The second task of this work is to present a picture of computational syntax both from a historical and from a contemporary perspective. The

third task is to take one more step and speak about the connection between syntax and parsing.

The part on Computational Syntax began in honest by linking and somewhat distancing itself from the part on Computational Morphology. Then it dealt with all sorts of problems such as Syntactic Structures, grammatical models and parsing. They revealed, if only in brief, the fact that Computational Linguistics is able to describe even the most complicated structures in its own way.

For almost all means and purposes the second chapter could have ended here, but it didn't. The reason is that it would have been simply unfair to ignore the search for the meanings of the words. That is the prime task of the next part.

2.5 On Computational Semantics

The last part of chapter two deals with the matter of the units of meaning. The Computational perspective in semantics is advocated as it opens up new possibilities.

This part presents basic concepts and issues of computational semantics. It speaks about the syntax-semantics relationship, semantic representations and approaches to syntax.

Having considered all that it is high time to end the theoretical procedures of the second chapter and move on to a more down to Earth third chapter which is meant to do nothing more and nothing less than describe an application of Computational Linguistics.

Chapter 3: On Machine Translation

The third and final chapter is meant to present an application of Computational Linguistics. It was not an easy choice between the various applications all with their degree of importance. On might have presented Dialog Systems for their astonishing complexity, where one would have had to take into account all the problems concerning

every single language level, Grammar and Spelling Checkers, where comprehensive grammar and orthography knowledge is required, or electronic dictionaries wich allows for a plethora of lexicographical issues to be discussed, and the list could go on. But, interesting as they undoubtedly are none of these made it to the final. Instead, it was deemed fit for the final chapter to choose the application which actually gave us the computational perspective on language.

Born out of the ambitions of the 1950s, machine translation proved more difficult than it looked like during the first attempts. Failure to produce advanced machine translation prompted linguists to search for new insights of the human or natural language, on the grounds that, if it could be understood it might be possible to replicate it. However, given the sheer volume of information, classical descriptive methods seemed unlikely to achieve that goal alone. The solution to this problem was to employ computational methods. It all seemed to make perfect sense because if the logics of the language were described in a machine readable manner then, they could be fed into a computational device which would then be able to generate new contents which would be both machine and user-friendly.

That didn't quite happen although obvious progress has been made. The reason for this lag is that even with the use of computational methods the understanding of such a complex matter as the human language is advancing at a very slow step. Amazing as the development of Computational Linguistics might have been since the late 1950s, it had to deal with the same problems as all the other types of linguistics and it just couldn't be expected to describe everything. So, even if computation machines have evolved to a point where it can process information fast enough for any needs in this case it just doesn't have enough information to process. The shortcomings of machine translation therefore reflect the shortcomings of computational linguistics. The following parts are meant to highlight exactly this statement.

Concluding Arguments

Before reading this paper we were aware of the existence of fields such as History, Linguistics and Computer Science. They seemed as different as night and day, each with little importance for the development of the others. It was therefore not expected to gain a meaningful inside into one of the fields by looking from the other's perspective.

Moreover, while history was expected to study everything, that is including linguistics, albeit from a superficial perspective, there was little hope of finding any connection between Linguistics and Computer Science. The study of human language, machine language and their interaction was deemed to be the exclusive task of computer scientists which deal with Natural Language Processing.

Computational Linguistics was often seen as a term interchangeable with Natural Language Processing and the achievements of Computational Linguistics were attributed to Natural Language Processing. In this light, most linguists tended to be circumspect towards Computational Linguists, regarding it as an experimental science, at best.

To dispel such pre-existing ideas it was necessary to provide compelling arguments. By putting these arguments together it is hoped that a new perspective regarding the identity of Computational Linguistics will eventually prevail. Therefore this work was conceived in such a manner as to address the three issues: the history, the theoretical workframe and applications of Computational Linguistics.

The first chapter of this work did nothing else but to show that there is a rather well defined notion of Computational Linguistics, despite debates on the matter. That it has a moment of inception and a constant evolution.

The second Chapter, which had to be, by its very nature, the lengthiest, was meant to prove that Computational Linguistics, far from being an experimental approach to language, is in fact set on solid theoretical foundations. In the spirit of this statement, this chapter was divided into five parts, one for each language level.

The Third Chapter is perhaps the most exciting as it deals with applications of Computational Linguistics. Machine Translation is the order of the day. They are defined, classified and explained. In this process new inside is gained on the language issues.

In the end, having presented with the three arguments regarding Computational Linguistics, it is safe to say that, while one must not necessarily embrace Computational Linguistics with all its theories and applications, one should recognize it as a fully-fledged part of Linguistics.

Partial Bibliography

1. 1994. Head-Driven Phrase Structure Grammar. Chicago: CSLI Publications and University of Chicago Press.
2. A.V. Aho, R. Sethi, J.D. Ullman: Compilers: principles, techniques, and tools. Boston/MA: Addison-Wesley Longman Publishing Co., Inc., 1986
3. Alan W. Black, Perfect synthesis for all of the people all of the time. IEEE TTS Workshop 2002.
4. Albro, D. M. (1994). `AMAR: A Computational Model of Autosegmental Phonology'. B.S. Thesis, Dept. of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.
5. Allen, J. F. (1983). Maintaining knowledge about temporal intervals. Communications of the ACM, 26,832-43.
6. Allen, J., Hunnicutt, S., and Klatt, D. (1987). From Text to Speech: The MITalk System. Cambridge University Press.
7. Amason, K. (1986). The segmental and suprasegmental status of preaspiration in Modern Icelandic. Nordic Journal of Linguistics, 9,1-23.
8. Anderson, J. M. and Durand, J. (eds.) (1987). Explorations in Dependency Phonology. Dordrecht: Foris.
9. Anderson, S. R. (1985). Phonology in the Twentieth Century: Theories of Rules and Theories of Representations. The University of Chicago Press.

10. Anderson, S. R. (1989). Johnson-Laird: The computer and the mind. Language, 65,800-11.
11. Antworth, F. (1990). PC-KIMMO: A Two-Level Processor for Morphological Analysis. Dallas: SIL.
12. Archangeli, D. (1988). Aspects of underspecification theory. Phonology, 5, 183-207.
13. Archangeli, D. and Pulleyblank, D. (1987). Maximal and minimal rules: effects of tier scansion. In McDonough, J. and Plunkett, B. (eds.), Proceedings of the Seventeenth Meeting of the North East Linguistic Society, pages 16-35. Graduate Linguistic Student Association, University of Massachusetts at Amherst.
14. Archangeli, D. and Pulleyblank, D. (1989). Yoruba vowel harmony. Linguistic Inquiry, 20,173-217.
15. Avery, P. and Rice, K. (1989). Segment structure and coronal underspecification. Phonology, 6,179-200.
16. Bach, E. (1976). An extension of classical transformational grammar. In Problems in Linguistic Metatheory, Proceedings of the 1976 Conference at Michigan State University, pages 183-224. Michigan State University.
17. Bach, E. (1983). On the relationship between word-grammar and phrase-grammar. Natural Language and Linguistic Theory, 1, 65-89.
18. Bach, E. and Wheeler, D. W. (1981). Montague phonology: a first approximation. In Chao, W. and Wheeler, D. W. (eds.), University of Massachusetts Occasional Papers in Linguistics, volume 7 pages 27-45. Graduate Linguistics Student Association, University of Massachusetts.
19. Barwise, J. and J. Perry. 1983. Situations and Attitudes. Cambridge, Mass.: MIT Press.
20. Batóg, T. (1967). The Axiomatic Method in Phonology. London: Routledge Kegan Paul.
21. Bear, J. (1990). Backwards phonology. In Karlgren, H. (ed.), Proceedings of the Thirteenth International Conference on Computational Linguistics, volume 3, pages 13-20. International Committee on Computational Linguistics.

22. Beesley, Kenneth and Lauri Karttunen. 2003. Finite State Morphology. Stanford: CSLI Publications.
23. Belew, Richard K.: Corpus-Based Linguistics and WordNet. Finding Out About. September 21st 2000. - URL http: //www.cs.ucsd.edu/~rik/foa/l2h/foa-6-3-2.html
24. Bellgard, M. I. (1993). 'Machine Learning of Temporal Sequences: Applications of the Effective Boltzmann Machine'. PhD thesis, University of Western Australia.
25. Belnap, N. D. (1977). A useful four-valued logic. In Dunn, J. M. and Epstein, G. (eds.), Modern Uses of Multiple-Valued Logic, volume 2 of Episteme pages 8-37. D. Reidel Publishing Company.
26. Bennett, M. and Partee, B. H. (1972). Toward the Logic of Tense and Aspect in English. Indiana University Linguistics Club.
27. Berners-Lee, Tim: The World Wide Web: A Very Short Personal History. W3C World Wide Web Consortium. July 1998. - URL http://www.w3.org/People/Berners-Lee/ShortHistory.html
28. Bird, S. (1991 b). Focus and phrasing in Unification Categori al Grammar. In Bird, S. (ed.), Declarative Perspectives on Phonology, pages 139-66. University of Edinburgh.
29. Bird, S. (1991a). Feature structures and indices. Phonology, 8,137-44.
30. Bird, S. (1992). Finite-state phonology in HPSG. In Proceedings of the Fifteenth International Conference on Computational Linguistics, pages 74-80. International Committee on Computational Linguistics.
31. Bird, S. (1994a). Automated tone transcription. In Proceedings of the First Meeting of the ACL Special Interest Group in Computational Phonology, pages 1-12. Association for Computational Linguistics.
32. Bird, S. (1994b). Introduction to computational phonology. Computational Linguistics, 20(3), iii—ix.
33. Bird, S. and Blackburn, P. (1991). A logical approach to Arabic phonology. In Proceedings of the Fifth Meeting of the European Chapter of the Association for Computational Linguistics, pages 89-94. Association for Computational Linguistics.

34. Bird, S. and Ellison, T. M. (1994). One level phonology: autosegmental representations and rules as finite automata. Computational Linguistics, 20,55-90. Bird, S. and Klein, E. (1990). Phonological events. Journal of Linguistics, 26, 33-56.
35. Bird, S. and Klein, E. (1994). Phonological analysis in typed feature systems. Computational Linguistics, 20,455-91.
36. Bird, S. and Ladd, D. R. (1991). Presenting autosegmental phonology. Journal of Linguistics, 27,193-210.
37. Bird, S., Bow, B. & Hughes, B. 2003. Towards a General Model of Interlinear Text Proceedings of EMELD 2003.
38. Bird, S., Coleman, J. S., Pierrehumbert, J. B., and Scobbie, J. M. (1992). Declarative phonology. In Crochetičre, A., Boulanger, J.-C., and Ouellon, C. (eds.), Proceedings of the Fifteenth International Conference of Linguists. Quebec: Presses de l'Université Laval.
39. Blackburn, P. (1990). 'Nominal Tense Logic and other Sorted Intensional Frameworks'. PhD thesis, University of Edinburgh.
40. Blackburn, P. (1993). Nominal Tense Logic. Notre Dame Journal of Formal Logic, 34,56-83.
41. Blackburn, P. and J. Bos. 1999. Representation and inference for natural language. MS. www. comsem.org/.
42. Bledsoe, W. W., and Browning, I. 1959. Pattern Recognition and Reading by Machine. In Proceedings of the Eastern Joint Computer Conference, 225-232.
43. Bloch, B. (1948). A set of postulates for phonemic analysis. Language, 24,3-46.
44. Bloomfield, L. (1926). A set of postulates for the science of language. Language, 2,153-64. Reprinted in M. Joos (ed.) (1957), Readings in Linguistics I: The Development of Descriptive Linguistics in America, 1925-56, pages 26-31. The University of Chicago Press
45. Bloomfield, L. (1933). Language. London: Allen & Unwin. (Reprinted 1976). Bobrow, D. 0. and Fraser, J. B. (1968). A phonological rule tester. Communications of the ACM, 11,766-72.

46. BO, J. (1989). A prosodic theory of epenthesis. Natural Language and Linguistic Theory, 7,217-59.
47. Bobrow, R. and Webber, B. (1980). Knowledge representation for syntactic/semantic processing. In Proceedings of AAAI-80, pages 316-23. American Association for Artificial Intelligence.
48. Boley, H. (1991). Declarative and procedural paradigms — do they really compete? In Boley, H. and Richter, M. M. (eds.), International Workshop on Processing Declarative Knowledge, pages 383-5. Springer Verlag.
49. Boolos, G. S. and Jeffrey, R. C. (1989). Computability and Logic (third ed.). Cambridge University Press.
50. Bouma, C., R. Malouf, and I. A. Sag. 2001.`Satisfying constraints on extraction and adjunction. Natural Language and Linguistic Theory,19,1-65.