A NATURAL LOGIC FOR ARTIFICIAL INTELLIGENCE, AND ITS RISKS AND BENEFITS

Gyula Klima

Department of Philosophy, Fordham University, New York, USA

ABSTRACT

This paper is a multidisciplinary project proposal, submitted in the hopes that it may garner enough interest to launch it with members of the AI research community along with linguists and philosophers of mind and language interested in constructing a semantics for a natural logic for AI. The paper outlines some of the major hurdles in the way of "semantics-driven" natural language processing based on standard predicate logic and sketches out the steps to be taken toward a "natural logic", a semantic system explicitly defined on a well-regimented (but indefinitely expandable) fragment of a natural language that can, therefore, be "intelligently" processed by computers, using the semantic representations of the phrases of the fragment.

KEYWORDS

Natural logic, natural vs. artificial intelligence, semantics-driven language processing.

1. INTRODUCTION

The purpose of this proposal is to launch a research project facilitating *intelligent natural language processing by computers*. In order to do so, we intend to pool the resources of scholars working in various fields, ranging from linguistics, philosophy, and history of logic, as well as model theoretical semantics and formal ontology, to computer science and artificial intelligence, cognitive psychology, and ethics. The need for such wide-ranging collaboration becomes apparent once we consider the enormity of the task.

2. BACKGROUND AND MOTIVATION

To make computers "intelligent", at least to the extent of being capable of processing bits and pieces of a human language with some semblance of understanding, we need to understand what the intelligent use of language consists in. For of course linguistic symbols can be used, processed, and manipulated without any understanding whatsoever, as it happens in simple word processors. Indeed, through the interactions of an intelligent user, even such primitive symbol manipulators can produce useful results, such as neat formatting, spell-checking, or answering simple questions about a text (or even an enormous number of documents) through string-searches and other mere syntax-oriented queries, without any representation of the meanings of these symbols in the machine doing the language processing itself.

By contrast, the Holy Grail of intelligent language processing would be the ability for the machine itself to process linguistic symbols with a semblance of understanding to do the processing regarding the meanings of the symbols being processed. This is what we can refer to as *semantics-driven language processing*, that we, humans do, when we use language with understanding. The proposed project will address this intuitive idea of intelligent language use

David C. Wyld et al. (Eds): CCSIT, SIPP, PDCTA, AISC, NLPCL, BIGML, NCWMC - 2021
pp. 117-123, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.110710

through tackling the conceptual and practical issues involved in understanding how natural language processing works for us, humans, the natural users of our natural languages as the medium of human thought, understanding and reasoning, and the applicability of the lessons learned from this study to artificial intelligence in computer science. Therefore, the two tasks indicated in the title of this project do require such broad-based collaboration: dealing with *natural logic* and *artificial intelligence* necessitates the recruitment of scholars from the wide range of fields indicated above.

Logic, as taught and practiced today as an academic discipline in the Frege-Russell tradition, is still primarily a formal mathematical study of certain fundamental forms of reasoning whose validity hinges on the fixed meanings of so-called “logical constants” or “logical connectives”, such as those expressed by our languages’ syncategorematic terms, like ‘and’, ‘or’, ‘if … then’, ‘not’, ‘some’, or ‘every’, which provide the formal structure of various forms of valid reasoning about any type of material expressed by our categorematic terms, namely, the terms that function as the subject or predicate of our categorical propositions.

However, classical formal logic (that is, standard predicate logic) is known to diverge from natural languages on various levels.

2.1. Mismatch of syntax

First, there is a known mismatch between the syntax of predicate logic and natural languages, based on predicate logic’s treatment of all common categorematic terms as predicates (hence the name, “predicate logic”) of singular referring expressions (individual names and variables, meant to represent proper nouns and pronouns of natural languages, respectively). Accordingly, predicate logic does not acknowledge the role of common terms in their referring function, forcing a reinterpretation of simple categorical propositions as conditionals or conjunctions, or leading to mere bewilderment over certain (so-called “pleonotetic”) phrases. For example, on this approach, ‘Some/All/Most Greeks are mortal’ would turn into ‘Some x is a Greek and x is mortal’, and ‘Every x is such that if x is Greek, then x is mortal’, and just a source of embarrassment in the last case (and so also a motivation for generalized quantification theory), respectively. For more on this, see [1] and [3].

2.2. The divergence between formal and material validity

Besides this well-known mismatch in syntax, there is also the known fact that there are various valid forms of reasoning not captured by the notion of logical validity defined for the formal language of predicate logic, namely, those forms of reasoning whose validity is based on the information content of the categorematic terms of our propositions, which is precisely what is disregarded by the formal language. (For instance, “the page you are reading is in front of you; therefore, it is not behind you” or “it is white; therefore, it is not black” is a perfectly valid inference, the validity of which, however, is not captured by standard predicate logic.) Furthermore, there are obviously invalid forms of reasoning, which, however, based on their syntactical form alone, would appear to be instances of formally valid patterns of reasoning. (For example, “whatever is healthy is alive, but the food in this health food store is healthy; therefore, it is alive”, which is invalid because of the equivocation of ‘healthy’ in the premises.)

This divergence between the notion of formal validity of a formal system and validity of actual pieces of natural language reasoning is nothing new. In fact, Aristotle (“the father of logic”) recognized that his formal system of syllogistic reasoning did not capture all valid forms of reasoning (which is why he wrote the *Topics*, not surprisingly, in connection with the *Categories*, to deal with valid non-syllogistic reasoning), and that there were many invalid forms of reasoning appearing to fit into valid syllogistic patterns (which is why he wrote his *Sophistical Refutations*, cataloguing various forms of fallacious reasoning).

It was also the realization of this divergence that motivated scholastic logicians' sophisticated discussions of the notion of logical consequence, striving to provide a unified account of valid reasoning in a "regimented" (explicitly regulated) version of academic Latin. But over the course of history's twists and turns, the scholastics' achievements were nearly completely forgotten, and were in modern times superseded by the theory and practice of formal logic as we know it. [8]

2.3. The need for a "natural logic"

Considerations of this sort recently more and more often prompted the expressed need for "a natural logic", both among philosophers of language and among computer scientists. ([6], [11]) A "natural logic" in the requisite sense would be a formal semantic system for a well-regimented fragment of a natural language with explicit phrase structure rules for its syntax and a recursively defined model for its semantics. (For an early attempt along these lines, see [5].)

What would make this approach "natural" in the first place would be cutting out the intermediary of a formal language along with its translation-rules (*à la* [10]) from the well-regimented fragment of natural language. In the second place, its semantics would allow the model-theoretical definition of a categorial structure, much in the vein of the Aristotelian theory of C*ategories*, licensing formal inferences based on the formal relations among the semantic contents of its categorematic terms, much in the vein of the *loci* of the Aristotelian *Topics*. Indeed, the compositional semantics for its propositions would enable the system to construct the *semantic content* of its propositions, thereby allowing for a content-sensitive definition of valid inference, based on the idea of the containment of the semantic content of the conclusion in the semantic content of the premises, along the lines of the conception of the *via antiqua* tradition of scholastic logic, yielding a relevant logic without the so-called paradoxes of entailment. [8] The recursive formulation of the formal semantics will allow the computability of the semantic values of any complex phrases for any arbitrarily chosen ontology; hence the system should easily offer itself for AI, enabling the computer to "see" the implications of all well-formed sets of sentences, which can actually be simply grammatical sentences of a natural language, thereby getting really close to "intelligent" natural language processing. (For the reason for the quotes, see, however [12]).

Approaching the natural logic in question from the starting point of standard predicate logic, the following steps need to be taken:

1. Represent a noun-phrase with a restricted variable, the values of which come out of the extension of the noun-phrase, provided it is not empty, otherwise its value is a zero-entity outside the universe of discourse. (Example: 'All humans are mortal' will become '(x.Hx)(Mx.)', where the values of 'x.Hx' will be elements of the extension of 'Hx', provided it is not empty, otherwise it is 0, an arbitrary item outside the universe of discourse.) The advantages of this approach include overcoming the mismatch of syntax mentioned above, the restoration of the full traditional Square of Opposition and syllogistic, and the immediate access to generalized quantification without the ontological extravaganzas of generalized quantification theory. A game-theoretical model for this approach as well as its perfect match with scholastic logic has been presented in [4].
2. Provide a tensed-modalized version of the previous system, in which the tense and modal operators in the matrices of restricted variables can perfectly model what scholastic logicians called the *ampliation* of terms: the phenomenon that in intensional contexts the

range of reference of our common terms becomes extended beyond actual existents, allowing quantification over and reference to non-existents. (Example: ‘Every horse is alive; no dead horse is alive; therefore, no dead horse is a horse; however, whatever is a dead F was an F and was alive; therefore, a dead horse was a horse and was alive.) The advantages of this approach should be obvious to anyone who considers how ampliation in natural languages works.

3. Define a signification-function (inspired by Geach in [2]; see, however, [7] as well) as the semantic function of predicates in a model that contains at least actual and non-actual elements in the domain of discourse, thus: SGT(P)(u) is an element of the domain, and ‘Px’ is true (relative to f), just in case SGT(P)(f(x)) is an element of the actual part of the domain, where f(x) is the value of x in an evaluation f. [Example: ‘Socrates is wise’ will be true, just in case SGT(‘wise’)(f(‘Socrates’)) is an element of the actual part of the domain, i.e., just in case Socrates’ wisdom is actual.] This is what medievalists refer to as the scholastics’ “inherence theory of predication”, the idea that the truth-maker of a simple predication is the actuality of the individualized property signified by the predicate in the individual(s) referred to by the subject. One immediate advantage of this approach is that it yields “fine grained intentions” even for logically equivalent predicates (SGT(‘triangular’) does not have to be the same as ‘SGT(‘trilateral’) despite their equivalence), which then can be cashed out in intentional (psychological) contexts (for example: “John knows that a circle is a circle, yet he doesn’t know that a circle is the locus of points equidistant from a given point”, since he has no concept of a geometrical locus).
4. Choose a manageable “regimented” fragment of a natural language, and apply the semantic ideas listed above, including a categorization of its categorematic terms, allowing for “topical” inferences based on the significations the categorematic terms as well as on the syncategorematic structure of its propositions. To that end, define propositional significations compositionally, and allow mapping all semantic values of all your phrases onto a categorially structured ontology, as outlined in [9].

The rest is just a matter of deft programming, and you can have a machine that will “understand” your regimented linguistic fragment, insofar as it will have a semantic representation of a potential infinity of phrases generable in your fragment, and will, therefore, “intelligently” converse about the issues expressible in your fragment. The fragment can of course grow, and “cannibalize” ever larger portions of a natural langue; indeed, several natural languages, allowing for more intelligent translations than any syntax-driven systems can produce.

3. PRELIMINARY METHODOLOGY

To be able to test whether we are moving in the right direction, we need to start small. We should start out with a small vocabulary and a very restricted set of construction and interpretation rules, so we can see with our finite human intuitions that the rules we input into the machines do indeed produce the intuitively correct results. Still, we must do this with a view to the further ends. As should be clear from the foregoing, the natural logic to be taught to computers requires a well-regulated, regimented language, in which the terms themselves “bear their contents on their sleeves”, namely, they carry essential information about the things they name. That is precisely the scenario, for instance, in organic chemistry, where the strictly regulated nomenclature serves exactly this purpose. So, taking some basic samples of texts in this field, we should first see if the results produced by the machine would fit our intuitive expectations, and whenever we find some anomalies, we need to chisel our rules accordingly. However, this “trial-and-error” period can be significantly shortened, and the whole enterprise can be substantially broadened by the input of our colleagues working in the fields listed above. So, the entire project needs to be given a more definite shape through a launch event, where our collaborators can pitch in with their ideas from their fields of expertise.

4. THE RISKS AND THE BENEFITS

So, what if the production of genuinely intelligent, language using robots (as opposed to the glorified word processors of today) becomes a reality?

On the one hand, the exponential growth of the processing and storage capacities of today's computers and the similarly exponential growth of scientific data make it virtually impossible for our individual human intellects to keep pace with the explosion of scientific information. So, intelligent computers may soon surpass humans in many fields, especially those involving long, and boring tasks that humans would not and/or could not tackle, while robots would handle with ease and without complaints. So much for the benefits.

On the other hand, the potential for humans' being surpassed by robots in areas requiring intelligence immediately raises the spectre of the well-worn staple of Sci-Fi: the "killer bots". The usual Sci-Fi way to face these potential risks is to require the implementation of some "safety measures" (think Asimov's "laws of robotics"), and the plot usually unravels to show the ways in which those safety measures can be overridden by evil humans or can go awry through misinterpretation of their intent by merely "robotically intelligent" robots. The novel approach of this project would be based on the insight that intelligence and benevolence are not incompatible; indeed, on the contrary, ideally, they would go together. But if artificial intelligence is the artificial implementation of the ideals of human intelligence, then it should be the implementation not only of the theoretical, intellectual ideals (flawless calculation and reasoning), but also of the moral ideals (acting for the perfection of humanity in each individual and all of mankind). Thus, one of the tasks of this project will also be the articulation of these ideals, along with the means of their implementation.

5. LAUNCH EVENT

Accordingly, the launch event should be a workshop of invited participants who can provide their input concerning the requisite tasks and sub-tasks of the project. Who will oversee the fragment of a natural language (English) to be processed? Who will work on its semantics? What type of semantics should it be? What sort of ontology would it require? What sorts of categories will it include? What will be the basis of the categorization? What types of inference rules will the semantics license besides those licensed by syncategorematic structure? What are the "topical" rules of inference we can use from scholastic logic? How shall we treat fallacious forms of reasoning (that scientific texts are obviously not immune to)? What are the expectable risks and benefits even in the early stages of the project? In what ways will the resulting artificial "intelligence" be different from human intelligence, and what are the potential security and ethical issues emerging from them? This is just a somewhat random sampling of the theoretical and practical questions such a workshop will have to address.

6. TOPIC LIST FOR THE WORKSHOP

- natural logics, their semantics, proof systems and computational feasibility
- the development of natural logics: the Aristotelian and medieval logic tradition, and the later connections to the relational and quantifier logics of De Morgan, Frege, Russell and Peirce.
- natural logic as extended syllogistic logics

- natural reasoning and domain specific reasoning, e.g., handling of plurals and partonomies.
- natural logics and diagrammatic logics
- relationships between natural logics and natural languages (in the plural)
- semantic issues such as intensionality, and collective vs. distributive readings in natural logics
- natural logic for knowledge bases and AI systems, e.g., computational natural language processing
- formal ontology and category theory
- categories and topical reasoning in natural logic
- connections between natural logics and description logics
- security and ethical issues related to the emergence of "intelligent robots".

ACKNOWLEDGEMENTS

The idea of this workshop came up in collaboration with a Danish group of computer scientists working with Jørgen Fischer Nilsson,jfni@dtu.dk, who generously invited me for a discussion of the plans outlined above in 2018. For various practical reasons (including Jørgen's retirement and my getting married across two continents) the project eventually failed to materialize in the year we planned it for. This is just my attempt to revive what still appears to me a worthy idea.

REFERENCES

[1] G. Boolos: "To Be Is to Be a Value of a Variable (or To Be Some Value of Some Variable)", *The Journal of Philosophy*, 8(1984), pp.430-431.

[2] Geach, P. T. "Form and Existence," *Proceedings of the Aristotelian Society* 55(1954–1955); reprinted in his *God and the Soul* (London: Routledge & Kegan Paul, 1969) 42–64.

[3] Klima, G. (1990) "Approaching Natural Language via Medieval Logic", in: J. Bernard-J. Kelemen: *Zeichen, Denken, Praxis,* Institut für Sozio-Semiotische Studien: Vienna, pp. 249-267.

[4] Klima, G. and Sandu, G. (1990) with "Numerical Quantifiers in Game-Theoretical Semantics", *Theoria*, 56, pp. 173-192.

[5] Klima, G. (1991) "Latin as a Formal Language: Outlines of a Buridanian Semantics", *Cahiers de l'Institut du Moyen-Âge Grec et Latin*, Copenhagen, 61, pp. 78-106.

[6] Klima, G. (2010), "Natural Logic, Medieval Logic and Formal Semantics", *Magyar Filozófiai Szemle*, 54(2010), pp. 58-75.

[7] Klima, G. (2015) "Geach's Three Most Inspiring Errors Concerning Medieval Logic", *Philosophical Investigations*, 38(2015), pp. 34-51. Online "early view" DOI: 10.1111/phin.12075

[8] Klima, G. (2016) "Consequence", in Read, S.L.-Dutilh-Novaes, C., *The Cambridge Companion to Medieval Logic*, CUP: Cambridge, UK, pp. 316–341. doi:10.1017/CBO9781107449862.014

[9] Klima, G. (2021) "Form, Intention, Information: from Scholastic Logic to Artificial Intelligence", in Ludger Jansen & Petter Sandstad (eds.) *Neo-Aristotelian Perspectives on Formal Causation,* Routledge, 2020. ISBN: 9780367341206; https://www.routledge.com/Neo-Aristotelian-Perspectives-on-Formal-Causation/Jansen-Sandstad/p/book/978036734Bene1206

[10] Montague, R. (1974) "English as a Formal Language", in: R Montague: *Formal Philosophy,* Yale University Press, New Haven-London.

[11] Nilsson, J. F. (2015). "In Pursuit of Natural Logics for Ontology-Structured Knowledge Bases". In *Proceedings of the 7th International Conference on Advanced Cognitive Technologies and Applications* (COGNITIVE 2015) (pp.42-46). IARIA

[12] Stephan, K. and Klima, G. (2020) with "Artificial Intelligence and Its Natural Limits", AI & Society, 36(2020), pp. 1-10; Springer online first: DOI: https://doi.org/10.1007/s00146-020-00995-z; read online: https://rdcu.be/b4wmM

AUTHOR

Gyula Klima is a Professor of Philosophy at Fordham University in New York City and the Director of the Research Centre for the History of Ideas at the Institute for Hungarian Research in Budapest. He is specialized in ontology and logical semantics and works on comparative analyses of the relevant scholastic and contemporary theories**.**