How Limited is the Limit?

Prakash Mondal
International Institute of Information
Technology, Gachibowli
Hyderabad 500032
prakash.mondal@research.iiit.ac.in

Abstract

In this paper an exploratory map of what intelligent natural language processing systems can achieve will be drawn up, given the advances that have been made in recent years as revealed in the latest developments in practical applications of natural language technology in areas as diverse as natural language generation, natural language understanding, machine translation, dialog system etc. Here a mathematical exploration of the issue in question will lay out the constraints on what they can achieve in their goal of automatizing language processing that humans do. It will be shown that these constraints together constitute a fundamental limit which these systems seem to fail to cross.

Keywords

Intelligent natural language processing systems; constraints; mathematical exploration; language processing.

1. Introduction

In recent years we have encountered a massive change in our conception of what natural language processing systems have achieved [1]. We have also gained a broader understanding of conceptual and empirical challenges that we face today in designing and implementing better systems that can be robust without incurring heavy computational costs [2]. With all this we are perhaps moving more towards the goal of automatization of human language processing in machines. But in spite of what has been gained in terms of theoretical and conceptual understanding of the problems, challenges, prospects in building natural language processing systems, there still seems to be an enormous gap between the level of performance these systems have come up to and how human language processing occurs [3] [4]. Is there any fundamental reason why the gap cannot be bridged? If there is any reason, why cannot we overcome it? And what is it about us that makes us do effortlessly all that these systems are designed to do, but are still far behind fully being capable of doing? It would be argued that a fundamental answer to all these questions perhaps exists. And the fundamental answer underlies a fundamental nature of human language processing mechanism.

2. What has been achieved

In recent years we have seen a spurt in the growth of computational models and practical systems in natural language processing. In the domain of natural language generation systems we have seen massive developments in phrase-based and feature-based systems of natural language generation [5], [6]. In the case of natural language understanding we have gone through ELIZA, PARRY, SIR etc. and we have viewed a range of rule-based and statistical natural language understanding systems for tutoring, medical advising etc. [7], [8] [9]. At the same time developments in parsing technology have moved on from rule-based models to data-driven statistical models and now future progress is moving toward hybrid models [10]. In addition, parallel progress has been made in machine translation systems and spoken dialog systems as well [11], [12], [13]. However, even if a lot has been achieved so far, a whole lot more is still to be achieved, which is upon the future generation high-computing technology [14].

3. What next?

3.1 The seemingly unbridgeable gap

Against this background, it seems that there is something missing. And such concerns are also recently found in a few recent works [15], [16], [17]. The missing link underpins the disparity between the level of performance of these systems as enumerated above and the seemingly effortless capabilities of human language processing. Can we really hope to scout out this missing link? Can there be a possible role of some hidden constraints, variables which operate in cognitive processing of human language, but are missing in inert machines? Even if the answers to these questions may not be easy to be spelled out, here it will be proposed that there is some deeply hidden fundamental law or law-like contingency that reflects the missing link. Before that a general architecture of cognitive processing of language will be drawn up to show that such a model of general cognitive processing is viable for placing linguistic processing in its ecological niche which has perhaps been ignored by the community [4].

3.2 Cognitive topography of human language processing

It is now necessary to present a model of the cognitive substrate of language processing by fitting the functional system of language into an architecture incorporating other cognitive domains so that how this architecture of language interacts with other cognitive domains can be linked to neural processing at the lower level. The broader picture in Figure 1 below is symmetric with certain constraints on mutual interaction of the cognitive domains/systems such that not all information passes from one domain/system to

another. Such constraints- computational, algorithmic and implementational in Marr's [18] sense- are specified by our genome.
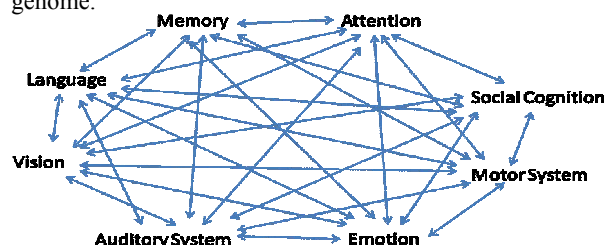
"Figure 1. Interconnection of Language to other Cognitive Domains/Systems."

Here the architecture is much more general in having a broader interconnected network of connections spanning through a plexus of cognitive domains. Such a network resembles Hopfield network more [19]. And research in a number of fields like biology, dynamical systems theory, behavioral genetics is moving more toward such interconnected models because of overlapping and shared neural mechanisms between many of them. [20], [21].

4. The hidden fundamental
4.1 Two fundamental variables
4.1.1 Meaning
It has generally been realized by philosophers, linguists, computer scientists, and perhaps neuroscientists what meaning in language is not, rather than what meaning in language is. However, despite the progress made in understanding meaning in language we are no nearer to having a complete grasp of what it is [22], [23]. And because of this the problem persists in fields which heavily rely on linguistic meaning, such as the field of natural language processing or AI.

Here the notion of meaning in language as we know it will be taken in a natural and general sense, even if no attempt will be made to define it given that there exists a serious danger in defining what meaning is. Roughly it can be said that meaning in language is what we understand in given circumstances from a linguistic structure- whether lexical or phrasal or sentential or discoursal. And it can correspond to a concept as is customary in *conceptual semantics* [24] or to a formal specification in *formal semantics* [25], or to a neurally connected network of activation patterns [26]. So meaning is here being treated as a fundamental variable in language in that language is impossible without this entity. Meaning is inherent in language as a system or language processing. It is so fundamental that it is perhaps least vulnerable in language disorders [27], [28].

4.1.2 Structure
Now the focus can be thrown on the other fundamental variable of language. The concept of structure has also a long history of varying implications. Right from the start linguistic structure has had a privileged place in the history of linguistic studies. Structural linguistics was entirely based on the conception of linguistic structure. Then with the advent of Generative grammar more stress has been given on syntactic structure as it is thought to be the computational systems with semantics and phonology as interpretive systems [29]. Here (linguistic) structure will be used in a general sense as it was done for meaning. It will be used to mean lexical structure, or phrasal structure or sentential or discoursal structure and phonological structure. The form of each type of structure is what has been characterized in the literature in terms of hierarchical organization [22].

Structure in language is also fundamental in terms of the scaffolding it provides to language. It is uncontroversial that natural language in any modality- auditory or visual (sign language) – must have structures which are constituted by phonological, lexical and syntactic structures. The exact representation of these kinds of structures may vary from one theoretical framework to another, but the fact that they form the skeletal architecture of language is fairly established.

4.2 The fundamental principle
Here the entire system of language has been reduced to two fundamental variables of language, namely, structure and meaning both of which can be either stable or variable; even if it is true that language interfaces with other cognitive domains/systems, which falls out of the architecture in Figure 1. above. However, soon this notion of language will be interwoven with the cognitive architecture shown above. Before that let us proceed to the fundamental principle that meaning and structure in language give rise to. Meaning and structure in language show a fundamental duality, in that what language is is manifested at varying levels either as meaning or as structure. For example, in the well-known case of tip-of-the-tongue phenomenon, what exists in mind is meaning, not the structure. Similar things happen in cases where we feel that a meaning is so intricate that it cannot be expressed in linguistic structure. Or for example in the case of patients with Broca's aphasia what is missing in them is structure at some level, but not meaning. The reverse is also found in a range of cases, though it is relatively rarer. Consider the case of *semantic satiation* where the structure of a word or phrase is repeated to the extent that it is bleached of its meaning [19]. Or for example, take the case of reading a poem which has linguistic structures in it, but often readers do not understand the meaning, even if they can decode the structures. In addition, in language itself we find specific words or phrases that have no meaning despite having structure; for example, light verbs like 'keep', the word 'of' in English are of such a nature.

Structure-meaning duality is actually at the heart of language and is perhaps more pervasive than has been

realized. In particular, it can be emphasized that this duality does not in itself diminish the complexity of language as we know it, rather it adds to it. It is because in the absence of structure it is meaning that fills up the vacuum and perhaps vice versa. We understand it more deeply when we see a complementarity between the two in our everyday affairs. Pre-linguistic children understand the meaning, but are not capable of producing structure [19]. We so often fail to understand the convoluted structures in language, though we sense the meaning. In fact, in our normal day-to-day life we transform meaning into structure and structure into meaning. And this can be simply expressed in the following manner-

$$f(M) = S \qquad \ldots (1)$$

and

$$f(S) = M \qquad \ldots (2)$$

By treating them equivalent we get

$$f(M) = f(S) \qquad \ldots (3)$$

But if we become more refined, the duality can be represented in the following way-

$$M = S \cdot k \qquad \ldots (4)$$

Here M denotes meaning as has been characterized above; S denotes structure and k is a constant specifying the constraints that operate on structure. What this ultimately amounts to is that meaning is actually equivalent to structure and vice versa. On the face of it, it may seem to be counterintuitive, since there is a general feeling that sometimes more meaning comes out of a given structure, so meaning cannot be equivalent to structure. But let us stop for a moment to understand what we mean by it. What is being claimed is that at a given instant, a given structure must be equivalent to meaning or vice versa, hence for example no one can compute two meanings from a single pattern of structure within a temporal window of , say, 100-200 milliseconds, which falls within the neural threshold for action potentials [30]. In the same way, no human being can compute two structures from a single meaning within the same temporal window. But it is of course possible to have

$$M > S \qquad \ldots (5)$$

or

$$S > M \qquad \ldots (6)$$

In that case we must have the following scenario as can be represented here as

$$C(t) = \sum_{i=1}^{n} A(t_1 \ldots t_n) \cdot \delta^{(h)} \qquad \ldots (7)$$

where $C(t)$ refers to total conceptual emergence, A is a mapping from S to M or vice versa with different instants from $t_1 \ldots t_n$ and $\delta^{(h)}$ is the distribution of them in a linear (or real or complex) space Ś.

Now this total conceptual emergence $C(t)$ defined over the mapping A can be related to processing function $P^{(d)}$ relativized to a cognitive domain d. First we have as $P^{(d)}$

$$P^{(d)} = \sum_{i=1}^{N} P_i \int d_{p1, \ldots} d_{pN} \delta_{(p1 \ldots pN)} \cdot \Delta\psi \qquad \ldots (8)$$

Here $p1 \ldots pN$ are differential probability functions coming out of the interaction dynamics of language with other cognitive domains (Fig. 1); $\Delta\psi$ is a temporal continuum distribution. Relating $C(t)$ to $P^{(d)}$ we get

$$\frac{\partial C(t)}{\partial t} = \frac{P^{(d)} \rho_i (C(t))}{P^{(d)} \rho_j (C(t))} \qquad \ldots (9)$$

What this boils down to is that C(t) if at time t, C(t) exists with the processing function P(d) with the probability density ρj, then it is true at any other time as well when there exists a current form of probability function ρi.

4.3 Layers of recursive emergence

It should now be clarified at this moment that the equation (1) derives from A which is an emergent reality that is embedded in another emergent reality at a higher level characterized by $P^{(d)}$. So it appears that here we have got a case of recursive emergence with one being nested inside a larger one. However, it may be noted that reading meaning-structure equivalence into language is itself an emergent reality which is nested inside another layer of emergent level representing $P^{(d)}$. And this $P^{(d)}$ is a non-algorithmic process as being mediated by dynamical non-linearity [19], [21], [30]. Interestingly, being an emergent level of fundamentality, this duality can never be broken down into either meaning or structure separately. In other words, if one tries to treat meaning representations or specifications in a way thinking that they are actually representing meaning which, to them, is a phenomenologically different entity altogether even if it stands for the structure in question, they are misled into a wrong direction. The reason is quite simple. Let us suppose that in a given scenario a meaning M' stands for or is symbolized by a structure S'; and assume that M'≠ S'. In a certain case let us say that S' is non-existent, but M' is not, since M' exists. In that scenario (for example, tip-of-the-tongue situation) quite often we get another S'' that is a sort of surrogate of S' when the speaker tries to substitute another structure for the one missing. What if we get another surrogate S''' for S'' itself? It will of course lead us into an infinite regress, which does not of course happen in normal situations. The point to be made is that if meaning had not been equivalent to structure, this would have never happened. Meaning is not tied to any specific structure, rather we can say meaning is structure or vice versa.

4.4 Meaning-structure equivalence and multiple grammars

Where does meaning-structure equivalence come from? Thus the concept of meaning-structure as specified above can be more generally linked to the concept of multiple grammars. Recently the concept of multiple grammars has been proposed to account for the existence of a range of hypotheses about grammar in language acquisition, the existence of overlapping but diverging grammars in language change [31] etc. And this can be generalized to the extent that it may well be said that at any stage in language processing a number of grammars operate in a hyperspace of potentiality from where the best one(s) that help(s) interpret the structure or convert the meaning into structure are chosen or utilized. Here the word 'grammar' is being used in a more general sense. It is of such nature that it may be used in interpreting structure(s) or for converting meaning into structure; so in this sense it is not just syntax that constitutes grammar, it may include semantic/pragmatic, morphological, phonological rules as well. In fact, meaning or structure is an actualization or realization of a selection form a wave of probability density of multiple grammars. It can be represented as

$$\sum_{i=1}^{n} h\left(P_1(G_1), \ldots, P_n(G_n)\right). P_1(G_1)., \ldots, . P_n(G_n) \qquad \ldots (10)$$

At any time there exist grammars $G_1, \ldots, G_n$ with the probabilities $P_1, \ldots, P_n$, where m is an arbitrary number and $n \leq m$, $n \geq 2$. But it must be the case that $m \geq 2$, since $n \geq 2$. This picture becomes much clearer in cases of ambiguity in natural language. Even in cases where there does not apparently exist any ambiguity; a selection must occur from that potential probability waving through the hyperspace. What is realized as structure or meaning at any time from that potential probability is an actualization at an emergent level. It may well be the case that this emergent actualization is unique in most instances [32].

5. Natural language processing, machine intelligence and meaning-structure equivalence

Below are some reasons for why machines cannot manifest any signs of meaning-structure equivalence.

First, natural language processing or even AI in general is still mostly based on linear, logical, rule-governed, algorithmic, and perhaps deterministic view of language processing, whereas the meaning-structure equivalence is a an emergent property of natural intelligence belonging to humans. We still find algorithms in natural language processing that are of such nature and that is why in the absence of meaning, machines cannot provide for structure in substitution of the missing meaning [33].

Second, it is indeed true that language computations are bound below by NP-hardness and are NP complete as has been mathematically proved by Ristad [34]. Where does this complexity come from? It can be said that this is the essence of natural language which reflects emergent non-linearity of meaning-structure equivalence.

Third, computational modeling of meaning-structure equivalence fails for the following reason. Let us suppose that there is a mapping function W from meaning-structure equivalence M-S^E to an algorithm K. Then we get W (M-S^E) $\rightarrow K$. Now if one has to computationally model this mapping, that mapping has to be embedded in another mapping function, say, Q which maps the above mapping to another potential algorithm for the purpose of modeling nested layers of emergence. Given that the goal of natural language processing is to build a robust system without incurring greater computational costs, such a computational modeling as depicted above will certainly involve greater computational costs.

6. What is the limit?

So what is the limit then? The limit lies in what the duality of meaning and structure shows machines to be confined to. It seems that in this way computation of natural language is bound from below by the constraints posed by meaning-structure equivalence in humans. We can put it here in the following fashion

$$\varphi \leq W\,(\text{M-S}^E) \qquad \ldots (11)$$

Here the functional mapping W (M-S^E) is bound below by the threshold limit φ which is not crossed by computers. This is what makes the problem of computational tractability. Indeed, it leads to a different view of constraints as operating on natural language processing. They allow us to refine our understanding of what it means to process language not just in a robust way which is a perfect epiphenomenon of this duality, but also in a broader view. The duality is like a self-referential loop as one gets the same thing in dealing with either.

7. Conclusion

This is a preliminary sketch of what it is that is behind the limitations today's natural language processing systems are facing. The paper has proposed that there is a natural reason behind all that. And it is this very fundamental of meaning-structure equivalence that is so elusive that it escapes computational tractability. There is of course no logical reason why the future cannot forge a different picture of natural language processing. Perhaps the future lies in a hybrid model of neural fuzzy systems combined with evolutionary computation subsumed under a dynamical non-linear approach. This has the potential of handing over to the research community the necessary self-organizing dynamics in language processing that we urgently require. But whether that will really place us in a

position to claim that we have succeeded in building natural or seemingly natural intelligence into machines will depend on our emerging views about language processing, human intelligence and perhaps cognition in general.

References

[1] Robert Dale, H. L. Somers and Hermann Moisl.(Eds.), Handbook of Natural Language Processing. Marcel Dekker, New York, 2000.

[2] Shalom Lappin. A Sequenced Model of Anaphora and Ellipsis Resolution. In António Branco, Tony McEnery and Ruslan Mitkov (Eds.), Anaphora Processing: Linguistic, Cognitive and Computational Modeling. John Benjamins, Amsterdam, 2005.

[3] J. A. Feldman. Computational Cognitive Linguistics. In Proceedings of 20th International Conference on Computational Linguistics. Association for Computational Linguistics, Morristown, New Jersey, 2004.

[4] Roger K. Moore. Towards a Unified Theory of Spoken Language Processing. In Proceedings of 4th IEEE International Conference on Cognitive Informatics, Irvine, USA, 2005.

[5] Eduard Hove. Language Generation. In Ronald Cole, Joseph Mariani, Hans Uszkoreit, Givanni Battista Virile, Annie Zaenen, and Antonio Zampoli. (Eds.), Survey of the State of the Art in Human Language Technology. Cambridge University Press, New York, 1998.

[6] C. L. Paris, W. R. Swartout, and W. C. Mann. (Eds.). Natural Language Generation in Artificial Intelligence and Computational Linguistics. Kluwer Academic, Boston, 1990.

[7] James Allen. Natural Language Understanding. CA: Benjamin/Cummings, Redwood City, 1994.

[8] Nicholas L. Cassimatis, Arthi Murugesan and Magdalena D. Bugajska A Cognitive Substrate for Natural Language Understanding. In Proceedings of the First Conference on Artificial General Intelligence, University of Memphis, 2008.

[9] Mark G. Core and Johanna D. Moore. Robustness versus Fidelity in Natural Language Understanding. In Proceedings of the Human Language Technology-North American Chapter of the Association for Computational Linguistics, Boston, USA, 2004.

[10] A. Murugesan, and N.L. Cassimatis. A Model of Syntactic Parsing Based on Domain-General Cognitive Mechanisms. In Proceedings of the 28th Annual Conference of the Cognitive Science Society, Canada, Vancouver, 2006.

[11] S. Nirenburg, H. L Somers and York Wilks. (Eds.), Readings in Machine Translation. MIT Press, Cambridge, Mass., 2003.

[12] Yorick Wilks: Machine Translation: Its Scope and Limits. Springer, New York, 2009.

[13] Egidio Giachin. Spoken Language Dialogue. In Ronald Cole, Joseph Mariani, Hans Uszkoreit, Givanni Battista Virile, Annie Zaenen, and Antonio Zampoli. (Eds.), Survey of the State of the Art in Human Language Technology. Cambridge University Press, New York, 1998.

[14] Jeremy Pekham. (Ed.). Recent Developments and Applications of Natural Language Processing. Kogan Page, London, 1989.

[15] S. Nireburg and Victor Raskin. Ontological Semantics. MIT Press, Cambridge, Mass., 2004.

[16] N.L. Cassimatis Cognitive Science and Artificial Intelligence Have the Same Problem. In Proceedings of the 2006 AAAI Spring Symposium on Between a Rock and a Hard Place: Cognitive Science Principles Meet AI-Hard Problems, 2000.

[17] Martin Kay. A Life of Language. Computational Linguistics 31(4): 425-438, 2005.

[18] David Marr. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. W. H. Freeman, San Francisco, 1982.

[19] Michael Spivey. The Continuity of Mind. Oxford University Press, New York, 2007.

[20] Klaus Mainzer. Thinking in Complexity: The Computational Dynamics of Matter, Mind and Mankind. Springer, New York, 1994.

[21] Myrna Estep. Self-Organizing Natural Intelligence: Issues of Knowing, Meaning and Complexity. Springer, Dordrecht, 2006.

[22] Ray Jackendoff. Foundations of Language. Oxford University Press, Oxford, 2002.

[23] P. Reccah. What is an Empirical Theory of Linguistic Meaning a Theory of? In Z. Frajzyngier, A. Hodges, and D. S. Rood. (Eds.), Linguistic Diversity and Language Theories. John Benjamins, Amsterdam, 2007.

[24] Ray Jackendoff. Semantic Structures. MIT Press, Cambridge, Mass., 1990.

[25] G. Chierchia and S. McConnell-Ginet. Meaning and Grammar: An Introduction to Semantics. MIT Press, Cambridge, Mass., 2000.

[26] A. M. Collins and M. R. Quillian. Retrieval Time from Semantic Memory. Journal of Verbal Learning and Verbal Memory. 8, 240-247, 1969.

[27] P. Fletcher and J. F. Miller. (Eds.), Developmental Theory and Language Disorders. John Benjamins, Amsterdam, 2005.

[28] Ewa Dąbrowska. Language, Mind and Brain. Edinburgh University Press, Edinburgh, 2004.

[29] Noam Chomsky. The Minimalist Program. MIT Press Cambridge, Mass., 1995.

[30] Eugene M. Izhikevich. Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting. MIT Press, Cambridge, Mass., 2007.

[31] Charles Yang. Knowledge and Learning in Natural Language. Oxford University Press, New York, 2002.

[32] Evelyne Andreewsky. Complexity of the Basic Unit of Language: Some Parallels in Physics and Biology. In M. Mugur-Schachetr, and Alwyn van der Merwe. (Eds.), Quantum Mechanics, Mathematics, Cognition and Action. Springer, Amsterdam, 2002.

[33] C. Havasi, R. Speer and J. Alonso. ConceptNet3: A Flexible Multilingual Semantic Network for Common Sense Knowledge. In Proceedings of the Recent Advances in Natural Language Processing, Borovets, Bulgaria, 2007.

[34] Eric Sven Ristad. The Language Complexity Game. MIT Press, Cambridge, Mass., 1993.