

Contextual Knowledge for Language-Endowed Intelligent Agents

Sergei Nirenburg, Marjorie McShane

Department of Cognitive Science
Rensselaer Polytechnic Institute

Abstract

That context plays an important – possibly, crucial – role in language processing is a truism. But what counts as context is different for different people. Automatic language understanding and generation has not yet attained the levels of quality habitually demonstrated by people. It is therefore not surprising that the notion of context has been gradually extended to include previously untapped phenomena, knowledge and data in efforts to improve system performance. This paper briefly discusses textual, knowledge, situational and interpersonal facets of a comprehensive view of context when characterizing the processes of understanding and generating natural language texts by artificial intelligent agents.

The role of context in eliminating ambiguity in text has always been recognized in AI and computational linguistics (CL). However, there have been different views on what counts as the context of a natural language utterance or text and what contextual features are necessary for language-related applications ranging from machine translation to question answering to building language-endowed intelligent agents. This paper discusses the contextual needs of this latter application.

Overt Textual Context

Already in 1949, in his famous Memorandum on machine translation (Weaver 1949), Warren Weaver offered this vision of the mechanization of the translation process: "... if ... one can see not only the central work in question, but also say N words on either side, then, if N is large enough one can unambiguously decide the meaning of the central word. The formal truth of this statement becomes clear when one mentions that the middle word of a whole article or a whole book is unambiguous if one has read the whole article or book, providing of course that the article or book is sufficiently well written to communicate at all." Much of

the subsequent work on language processing in AI and CL has been till this day devoted to finding ways to implement this "unambiguous decision" process on the basis of word meaning, using a variety of approaches to what constitutes word meaning.

One method of encoding lexical meaning was to use the artificial metalanguage of semantic primitives (such as "semantic markers" first popularized by Katz and Fodor 1963) that would in combination interpret the meanings of words in natural languages. Many systems of such primitives and ways of using them for lexical disambiguation and combining them to generate meanings of texts were proposed over the years by philosophers, linguists and AI researchers and used in a variety of proof-of-concept systems. Quillian 1968, Wilks 1975, Schank 1975, Hirst 1988 are among the most influential ones. Both the metalanguages and the systems using them in this research were small-scale, and the computational dictionaries that used the metalanguages for describing lexical meanings had to be compiled by hand.

When human-oriented dictionaries and thesauri became available in machine-readable form, large-scale efforts were mounted to automatically transform information they contained into machine-tractable metalanguage statements to support disambiguation. Alas, years of relatively large-scale efforts to make the vast amount of knowledge in these resources machine-tractable essentially failed (see Ide and Veronis 1993 for a discussion). Machine-readable dictionaries were also used by Lesk (1986) and others to carry out word sense disambiguation directly, without conversion into machine-tractable form. Navigli and Lapata (2010) used WordNet (Fellbaum 1998), a hand-built thesaurus, for similar purposes.

Once work has been done on tagging online corpora with word senses from a machine-readable dictionary or thesaurus, it became possible to apply a variety of machine learning techniques to word sense disambiguation. A very large number of systems of this kind have been developed over the past 15 years or so (see Navigli 2009 for a survey).

Hidden Context: World Knowledge

Well before the flowering of this word sense disambiguation work, it became clear that knowing the dictionary meanings of words in word *W*'s context is routinely not enough to disambiguate *W*. Thus, Bar Hillel (1960) argued that fully automatic high quality machine translation was impossible because the capability that Weaver postulates as a requirement can be attained only by "intelligent readers." His example was *The box was in the pen*. This sentence is easier to understand in the context of a short story: *Little John was looking for his toy box. Finally he found it. The box was in the pen. John was very happy*. What allows people to disambiguate *pen* is their knowledge that a typical writing implement is too small to contain a typical (even small!) box. This information is nowhere to be found in the context understood as "N words on either side," even if the meanings of all these words are available in a machine-readable dictionary. It was clear to Bar Hillel that for a machine to be able to disambiguate it requires, in addition to a dictionary, a "universal encyclopedia." So, he promptly concluded that this requirement "is utterly chimerical and hardly deserves any further discussion."

Today nontrivial subsets of Bar Hillel's universal encyclopedias exist and are known as ontologies. There has been a flowering of efforts on building ontologies since the early 1980s, CYC (Lenat and Guha 1990), SUMO (e.g., Niles and Pease 2001) and DOLCE (e.g., Gangemi et al. 2002) being good examples of upper-level ontological systems. Ontological knowledge bases were developed to support general automatic reasoning, not specifically for the purposes of eliminating ambiguity in language processing. One ontology that was originally developed to support language processing, specifically interlingual machine translation, is the OntoSem ontology (e.g., Nirenburg and Raskin 2004). One motivation for creating this ontology was to allow automatic language processors to generate unambiguous meaning representations for sentences like Bar Hillel's example. That is why, among many other knowledge elements describing concepts in this ontology included both the knowledge-level counterparts of selectional restrictions¹ and properties that are not connected to text processing, such as typical sizes of physical objects.

Another capability licensed by the OntoSem approach is using detailed descriptions of complex events, with their concomitant participants (agents and objects), forming

¹ Selectional restrictions are constraints on co-occurrence of word senses – e.g., that the meaning of any direct object of the English word *address*, as in *He addressed the letter in longhand*, is a reference to a physical object that is a kind of missive; this sense of *address* will thus be distinguishable from *address* in *He addressed the students* and in *They addressed the problem in time*.

causal and typical temporal chains. The use of this kind of knowledge has been first popularized in AI by Schank and his co-authors under the name of scripts (e.g., Schank and Abelson 1977). Scripts provide a wealth of hidden context that is not available overtly in texts and help resolve both lexical and referential ambiguity. In *John made Bill carry two heavy suitcases from campus to his house. No wonder that he was exhausted, he refers to Bill*. This referential ambiguity cannot be resolved without the knowledge that carrying heavy loads makes people tired.

Hidden Context 2: Concept Instances

Note that it is impossible to determine from the above text whether *his* refers to John or Bill. Nothing in the text or in the knowledge about the world helps to resolve this referential ambiguity. However, if the first sentence of this text were something like *John's house is very far from campus*, then humans won't have any problem resolving it. In order for artificial intelligent agents to do so, they need to keep track of the history of the narrative in the text (or statements in a dialog). This history is best viewed as a set of statements in a metalanguage representing meanings of text sentences or dialog turns. These representations consist of instances of ontological concepts that interpret the meanings of words and phrases used in the text or dialog, relations among these concept instances and values of some general parameters, such as time and location as well as modality. In the terminology of cognitive science this context-providing information is part of the agent's short-term memory (STM).

The knowledge about the distance between John's house and the campus may be available to the agent before it starts reading the text or participate in a dialog. This means that knowledge about the world (stored in its long-term memory, LTM) must include not only an ontological model of types of objects and events in the world but also remembered instances of such objects and events. In OntoSem, this context-providing knowledge base is called fact and belief repository. Its content includes all of the agent's STM and a significant subset of its LTM.

Hidden Context 3: Mindreading

An agent's fact and belief repository contains the agent's beliefs about other agents. This special type of knowledge includes facts like where their houses are located. But it also includes additional types of important context-providing beliefs: beliefs about other agents' view of the world, their social roles, their inventories of goals and plans, their personality traits and even their physical and mental (emotional) states. These beliefs provide indispensable context for language understanding. For example, suppose on a November morning A and B are in a room with an open window and A utters "It's too cold here." B recognizes that B is uncomfortable and has on its goal

agenda an active instance of the goal MaintainPhysical-Comfort. If A has a higher social or organizational standing, B may also interpret this statement as an indirect speech act requesting that the window is closed. However, if the relative social standing between the two agents is reversed, B may interpret this statement as a different speech act – that of asking permission to close the window.

Similar examples can be evoked to illustrate the importance of other parameters comprising the profiles of other agents within an agent’s belief repository. What this and other similar examples demonstrate is that systems that aim to understand the intended meaning of a text or a dialog turn should extend the notion of context well beyond “N words on each side of a given word.” Note that extracting and representing meaning is not a necessary condition for success of many popular applications. Modern statistics-based machine translation and information extraction systems do not deal with meaning and so can make do with the narrow notion of context that is more in line with Weaver’s original statement.

Still, human-level intelligent agents of the future must be able to extract and represent the meaning of language inputs. This means that they must be able to deal with a variety of issues: multiple kinds of ambiguity – lexical, referential, pragmatic (as in the case of indirect speech acts) and others; vagueness; ellipsis; unexpected – ungrammatical, semantically novel (e.g., metaphorical) – input; etc.

The extended notion of context described above defines a set of parameters that can serve as arguments for decision functions or algorithms devoted to each of the tasks in meaning extraction and representation. Maintaining and determining time-sensitive values of these parameters is an ongoing background operation in an artificial intelligent agent model that facilitates high-quality decision making (see, e.g., McShane et al 2011 for a discussion of a subset of agent memory management operations needed for maintaining referential coherence of its belief repository). It consists of updating the agent’s belief repository as a side effect of the agent’s functioning. This includes, among other actions, introducing new conceptual beliefs (“A thinks it is cold here”) and modifications to the set of profiles of other agents (“A seems to be in great discomfort and is irritated.”)

Context in Language Production

Context is important not only for understanding but also for generating dialog turns. For this purpose, the agent’s belief repository must include beliefs not only about other agents but also about self (yes, we do recognize that these beliefs can be erroneous). In the MVP system (McShane et al. 2009), we modeled a broad variety of artificial intelligent agents playing the role of virtual patient (VP) in a physician training environment. In this system, VPs were supplied with their individualized versions of ontology and

semantic lexicon, individualized belief repositories, goal and plan inventories, profiles of other agents and profiles of themselves that included individualized personality profiles and biases (see McShane et al. 2013 for a discussion). They were also equipped with a simulated dynamic model of their physiological and pathological processes, which allowed them to obtain input through simulated interoception – so that in their physical state profile they could record pain, difficulty swallowing and other physical symptoms. All the above constituted sources of contextual information – values of parameters used in decision making during processing perceptual input, reasoning and decision-making with respect to action.

As an example of this kind of functioning, consider an experiment that we ran with several different VPs. Each of them was given (by manipulating the physiological model) the primary complaint of difficulty swallowing and put in the situation of a diagnostic dialogue with a user playing the role of a medical professional. When asked the question “Have you been traveling recently?” they responded in different ways depending on their general knowledge, character traits, more or less successful mindreading of the interlocutor and other contextual parameter values. Among the answers were the following:

VP1: No, I haven’t been anywhere that might have made me sick.

VP2: No, I have not traveled recently.

VP3: Yes, I drove to Washington to visit my sister.

VP4: Yes, I have been on a crowded plane last week.

VP5: No. Why are you asking?

VP1 and VP4 correctly “read” the goal pursued by the user (diagnosing) and had in their ontologies the knowledge that their symptom might be due to an air-borne infection whose propagation is facilitated in particular situations, such as when many people are bunched in a confined space. It so happened that the belief repository of VP1 did not contain any event in which the agent was in a confined space while the fact repository of VP4 did. In choosing their responses VP1 and VP4 undertook mindreading of beliefs they ascribed to their interlocutor. VP2 decided to bypass any mindreading and to respond to the question factually. Nothing can be concluded about the content of its ontology or belief repository. It is not clear whether VP3 mindread the user’s goal. What is clear though is that, unlike VP1 and VP4, it did not have the necessary knowledge about the world relating to the provenance of its symptoms. Finally, VP5 decided not to mindread the user and did not have the requisite world knowledge but attempted to learn what goal the user pursued in asking this question. If given an answer, VP5 might then have opted to augment its ontology. The user could also conclude that VP5 has a heightened level of the personality trait of suspiciousness – but it could not be ascertained in the brief exchange.

Until recently, ambiguity resolution was firmly associated with language understanding but not with language production. Indeed, text or dialog turn producers start from unambiguous representations of the meaning of what they want to convey, and their main tasks include selection of the lexical and grammatical elements that most faithfully render this meaning. A deeper analysis of this process with a view of applications in cognitive robotics and development of artificial intelligent agents shows that in fact ambiguity does play a role in language production. Non-textual context can dictate particular lexical and grammatical choices in language production. To give a simple example, consider how you would answer the question “What is a violin?” depending on who is asking. The Merriam-Webster Learner’s Dictionary gives the following definition of *violin* that is suitable for children or people whose knowledge of English is limited:

a musical instrument that has four strings and that you usually hold against your shoulder under your chin and play with a bow

The regular Merriam-Webster’s definition uses unconstrained vocabulary and assumes knowledge of musical terminology (and, presumably, a special interest in things musical – which could be considered an instance of generalized mindreading, that is, assessing the group of readers that would use this definition):

a bowed stringed instrument having four strings tuned at intervals of a fifth and a usual range from G below middle C upward for more than 4½ octaves and having a shallow body, shoulders at right angles to the neck, a fingerboard without frets, and a curved bridge.

Another set of choices for the language producer in a dialog involves the contextual assessment of how brief a text to be produced can with a view of minimizing the overall effort of the speaker and a hearer in a dialog. Generating a brief utterance takes less immediate effort but can lead to extra effort down the line if the utterance is too elliptical for the interlocutor to understand. In such a situation, the interlocutor is likely to ask for clarifications, and this would mean extra effort would need to be expended. Even if the interlocutor is able to understand the brief utterance, to do so might cause him or her to expend too much effort. So, a cooperative dialog participant would strive to minimize the joint effort, not his or her individual one (see Piantadosi et al. 2012 for a discussion of communicative ambiguities and the principle of least effort).

It’s Not All About Language

In the example of the open window, Agent B may choose to respond to A’s statement verbally. But it may also decide to respond in other ways – for example, by closing the

window or by doing nothing at all. The choice space for action transcends a particular action mode. Often this mode could be mixed – Agent B could say “OK, let’s make you more comfortable” while closing the window. But the decision to choose a particular action or combination of actions still uses the same contextual parameters as those that would be used by a purely conversational agent.

In general, the reasoning and decision making apparatus of an artificial intelligent agent does not necessarily need to be different for each perception, reasoning and action modality (see Nirenburg and McShane 2015 for a discussion).

Final Thoughts

In this paper, we attempted to motivate the need for including several kinds of hidden context -- essentially, elements of general and situational world knowledge – to support language-endowed intelligent agents. We did not touch upon the many issues arising from the need to put this knowledge to use and, importantly, how to take into account and function in the very realistic eventualities of the knowledge being unreliable or insufficient, how to make it possible for the agent to recover from failures and learn (from experience, from being told, from reading). For a glimpse into our recent experience in addressing such issues in language-endowed intelligent agent (LEIA) systems see, e.g., McShane and Nirenburg 2012, 2015; McShane et al. 2013a, b, 2015, 2016; Nirenburg and McShane 2015, 2016.

References

- Bar Hillel. Y. 1960. A Demonstration of the Nonfeasibility of Fully Automatic High Quality Translation. Appendix III of 'The present status of automatic translation of languages', *Advances in Computers*, vol.1 (1960), p.158-163.
- Fellbaum, C. (Ed.). (1998). *WordNet: An Electronic Lexical Database*. MIT Press.
- Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., and Schneider, L. 2002. Sweetening ontologies with DOLCE. *Proceedings of EKAW 2002*, Sigüenza, Spain.
- Hirst, G. 1988. Semantic Interpretation and Ambiguity. *Artificial Intelligence*, 34, 131-177.
- Ide, N., & J. Véronis, 1993. Extracting knowledge bases from machine-readable dictionaries: Have we wasted our time? In *Proceedings of KB&KS'93 Workshop* (pp. 257–266). Tokyo.
- Katz, J. and J. Fodor. 1963. The Structure of a Semantic Theory. *Language*, 39, 170-210.
- Lenat, D. and R. Guha. 1990. *Building Large Knowledge-Based Systems*. Reading, MA: Addison-Wesley.
- Lesk, M. (1986). Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. In *SIGDOC '86: Proceedings of the 5th annual international conference on systems documentation*.
- McShane, M. and Nirenburg, S. 2012. A knowledge representation language for natural language processing, simulation and

- reasoning. *International Journal of Semantic Computing*, 6(1): 3-23.
- McShane, M. and Nirenburg, S. 2015. OntoAgents gauge their confidence in language understanding. In Ahmed, N., Cummings, M., and Mille, C. (Eds.), *Self-Confidence in Autonomous Systems: Papers from the AAAI Fall Symposium*. Technical report FS-15-05, pp. 22-29. Menlo Park, CA: AAAI Press.
- McShane, M., S. Nirenburg, B. Jarrell, S. Beale, G. Fantry. 2009. Maryland Virtual Patient: A knowledge-based, language-enabled simulation and training system. *Bio-Algorithms and Med-Systems*: 5(9): 57-63
- McShane, M., Nirenburg, S., and Jarrell, B. 2013a. Modeling decision-making biases. *Biologically-Inspired Cognitive Architectures (BICA) Journal*, 3: 39-50.
- McShane, M., Nirenburg, S., Beale, S., Jarrell, B., Fantry, G. and Mallott, D. 2013b. Mind-, body- and emotion-reading. *Proceedings of the Annual Meeting of the International Association for Computing and Philosophy (IACAP 2013)*
- McShane, M., Nirenburg, S., Jarrell, B., and Fantry, G. 2015. Learning components of computational models from texts. In Finlayson, M. A., Miller, B., Lieto, A. and Ronfard, R. (Eds.), *Proceedings of the 6th Workshop on Computational Models of Narrative (CMN'15)*, pp. 108-123.
- McShane, M., Nirenburg, S. and Beale, S. 2016. Language understanding with Ontological Semantics. *Advances in Cognitive Systems* 4:35-55.
- Navigli, R. (2009). Word sense disambiguation: A survey. *ACM Computing Surveys (CSUR)*, 41(2), 10.
- Navigli, R. and M. Lapata (2010). An experimental study of graph connectivity for unsupervised word sense disambiguation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(4), 678-692.
- Niles, R. and A. Pease, Towards a standard upper ontology, in *Proceedings of Formal Ontology in Information Systems (FOIS 2001)*, Ogunquit, Maine, USA, 2001.
- Nirenburg, S. and McShane, M. 2015. The interplay of language processing, reasoning and decision-making in cognitive computing. In Biemann, C., Handschuh, S., Freitas, A., Mezziane, F. and Métais, E. (Eds.), *Natural Language Processing and Information Systems, Proceedings of the 20th International Conference on Applications of Natural Language to Information Systems (NLDB 2015)*, pp. 167-179. Springer Lecture Notes in Computer Science 9103.
- Nirenburg, S. and McShane, M. 2016. Slashing metaphor with Occam's Razor. *Proceedings of The Fourth Annual Conference on Advances in Cognitive Systems*.
- Nirenburg, S. and V. Raskin 2004. *Ontological Semantics*. Cambridge, MA: MIT Press.
- Piantadosi, S. T., Tily, H. & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition* 122, 280-291.
- Quillian, M. R. (1968). Semantic memory. In M. Minsky (Ed.), *Semantic information processing*. Cambridge, MA: MIT Press.
- Schank, R.C. 1975 *Conceptual Information Processing*. North Holland, Amsterdam, Holland.
- Shank, R.C. and H. Abelson. 1977. *Scripts, Plans, Goals and Understanding*. Englewood Cliffs, NJ: Erlbaum.
- Weaver, W. 1949. Translation. Reprinted in: W.N. Locke and D.A. Booth (eds.) *Machine Translation of Languages*. Cambridge, MA: MIT Press, 15-23.
- Wilks, Y. (1975). A preferential pattern-matching semantics for natural language. *Artificial Intelligence*, 6, 53-74.