What's in a symbol: ontology, representation and language

Sergei Nirenburg and Yorick Wilks

June 24, 1999

Abstract

This paper is in a form unconventional in modern journals but traditional for the discussion of foundational questions: a dialogue. It is a form that makes it possible to contrast two deeply held but incompatible views, each with its standard forms of defense, in order to seek common ground and make the differences more precise. In artificial intelligence, or at least in the major part of it still committed to symbolic representations, there is a long history of discussion of the origin and nature of the symbols we use in representations, symbols which normally look like words, English words in fact, but which most researchers deny are such words, since to concede that would put in question the abstract nature of the representation. In what follows, we examine our common ground and then diverge over five specific questions on the issue of representations. The discussion focuses on symbol use in representations of language, because there the similarity is most acute — between the representation and the represented — but the issues are general and apply to symbolic AI as such.

1 Introduction

That language and cognition are closely related in humans is undisputed, but in machines the connection is more problematical and disputed. A dialogue is a form that can sharpen difficult questions, at least that is our hope, but not one that lends itself to historical or scholarly introductions, though it should be said at the outset that the central question in what follows, the nature, origin, acquisition and use of the symbols in AI representations, is one that has been discussed in print many times, by AI researchers (e.g. [12]; [6]) among others) as well as by many spectators of AI in philosophy and other cognitive sciences. After the discussion created by some of these papers, the issues raised then disappear again, unresolved, perhaps because they are unresolvable. This paper starts with common-ground between the two opposed, and apparently irreconcilable positions, and then attempts to refine the differences between them piecemeal: these positions are that the symbols we use in AI representations are (or, conversely, are utterly different from) the English words they plainly resemble. Much AI research is only comprehensible on the negative assumption (in parentheses above), yet an important question, rarely asked, is whether it makes any real difference to the course of, or results from, AI research which of the above positions is true or false.

This paper, as will become clear, is in a very conservative tradition: core representationalist AI. It retains Newell's assumptions [14] about semantics as information processing as much as it does McCarthy's vision of AI as the method of heuristics [11], as opposed to continuing attempts to make properties of representational structures provable in some strong sense. Indeed some version of this dialogue gets published as a paper every five years or so, as in the citations above, and readers will have to judge for themselves whether any progress is being made. We, above all, want to clarify the questions and seek clear differences among the answers.

Within the representationalist camp, we wish to separate ourselves from those aspects of the formalist movement, whether within linguistics or mainstream AI, who believe the solution, to whatever problem there is here, is to continue to seek formalisms with a logical semantics. We have discussed elsewhere [25], [16] the claims of the formal approach to natural language processing (NLP) and will not repeat them here: in a nutshell, our view is that there is no reason to believe that systems for which notions like deductive closure are important have any demonstrable relationship to NLP, either as an empirical, engineering task or as a model of human processing.

The central issues for us are: first, whether or not one believes the symbols in representations (whether of language itself or some other part of the world) are fundamentally language-like in nature, and, secondly, whether or not the answer to this question affects our expectations concerning the development of large-scale application systems and largely automatic acquisition of representational resources (lexicons, knowledge-bases etc.).

If there were no relationship between our enquiry here into the nature of symbols and the processes within which we intend (as AI researchers) to use them, then our enterprise would be purely philosophical. The second issue above is currently of great practical importance in NLP. But we will argue here that the former, more apparently philosophical, question may influence the outcome of any research program.

Our discussion will be organized round the following five questions:

1. Are representation languages (RLs) natural languages (NLs) in any respect?

2. Are languages (natural or representational) necessarily extensible as to sense?

3. Are language acquisition and extensibility linked?

4. If automatic acquisition is possible, what are the consequences for any RL/NL difference?

5. What are the consequences of all of this, if any, for representations for humans (versus for machines).

2 A dialogue on representation: prolegomenon

SN: Of course, even our agreement might be more of the kind illustrated by the old Soviet-era joke from the "Radio Yerevan" series:

Question: Have you heard that Academician Ambartsumian has just won a Volga car in state lottery?

Answer: Of course I have. Only he's no academician. He's a night watchman. And his name is not Ambartsumian. It's Rabinovich. And it was not a car. It was a hundred rubles. And he played poker, not the state lottery. Oh, and by the way: he didn't win!

Some form of representationalism remains the mainstream AI position, even after much pressure from connectionism and, in the particular case of NLP, the recent and much publicized successes of statistical methods. For example, purely statistical machine translation (MT) has risen to a level of success, in terms of the percentage of sentences correctly translated, of roughly the level of code redundancy in natural languages, i.e. 50-60%.

Our examination of the basis of meaning representation will return constantly to the following thesis: are representations in NLP, KR etc. coded in terms close to those of natural language itself, and what are the consequences of this fact, if it is one.

Of course, the crucial issue is to define what is meant by "close", or, in other words, what is natural language and where the boundaries between natural and artificial languages are to be drawn.

YW: What then are the salient features of the methodology that underwrite this extraordinary ability of NLP and KR researchers to go on writing down knowledge structures, linguistic or otherwise, in formalisms whose abstract structure is defined, more or less, but whose content — predicates, primitives, classifiers etc., — is never set out in any formal terms but the most trivial?

SN: Here one has to agree first on the semantics of "formal"? Are there gradations of formality? Or do we all have to subscribe to logic and algebraic definitions of theories in order to be considered formal? Maybe the ability to be directly computable is what makes a structure formal?

YW: Remember that this question can only be asked of those who remain, more or less, within the representational paradigm: it cannot be asked of a fully distributed connectionist, nor of one, like Ken Church or Peter Brown who adopts a view of NLP in which words are simply symbol strings without significance.

Yet, in our day jobs we subscribe to the methodology of "NLP-as-Language-Engineering" and much discussed matter, and our position can be expressed very succinctly: formalists do sometimes change formalisms but they never go through a process of rejecting a complex set of hypotheses in the face of large scale, statistically assessable evidence. That process, the distinctively scientific process, the lifeblood of science in all its aspects (except perhaps cosmology), never occurs in purely formal theories.

SN: We cheerfully admit, of course, that we are unlikely to solve any such problems in this discussion, but at least we hope to tease out additional issues and maybe to learn to formulate the five following questions a bit better.

2.1 The first question: are representation languages natural languages?

YW: This dialogue, though not a philosophical or psychological one, has overlap with one of the main aspects of Fodor's Language of Thought [4] claims: that the basis of mental representation is language-like in nature. Fodor presents a set of claims concerning the language-like properties of his putative LOT: in particular, the hierarchical, or tree-like, nature of its structures and the non-compositionality of the meanings of its predicates. The former has involved Fodor in extended disputes with connectionists about whether or not tree structures can be replicated by connectionist learning techniques. The latter runs in the face of standard compositionality- oriented accounts of text (or sentence) meaning. We differ from Fodor on the crucial issue of what it means to be "language-like".

SN: Fodor's is only one set of criteria for what features make a language "NL-like". We can suggest additional ones. Two basic features, will probably include the "live", "unconstructed" character of NL and the functional criterion of being designed to support human communication, that is, relying on the human apparatus of understanding.

YW: The first feature of language that should concern us in this discussion is as follows: can the predicates of a formal representational language avoid ending up ambiguous as to sense? A negative answer to this question would make RLs NL-like. It will also mean that understanding a representation involves knowing what sense a symbol is being used in. If NLs are necessarily extensible as to sense — and words get new senses all the time – then can RLs that use NL symbols avoid this fate?

SN: The predicates of a representational language are consciously CONSTRUCTED. They do not exist except through the will of the acquirer. We can argue about the process of construction and how the elements of a representational language get realized in practice. But the crucial difference is that NLs HAPPEN, RLs are MADE. You presuppose somehow that an RL is not constructed but rather EXISTS. And if, indeed, RL symbols are allowed to be ambiguous, then having to know in what sense an RL symbol is being used simply sends the task of disambiguation one step further: either to yet another, this time, unambiguous, RL.

YW: Here is where we start to disagree strongly: for me, RLs are not made, or rather they are made up of existing NL bits, all too often English. And I can give no sense to the claim that we make symbols ambiguous or not; we have no such control of NLs or RLs.

SN: In some NLP applications, texts in an RL (such as the Mikrokosmos TMR, see, e.g. [15]) are typically used to represent the meaning of an NL text. Stating that RL elements are ambiguous is equivalent to saying that NL meaning cannot be truly extracted and represented. Another complication is the difficulty in using ambiguous RL statements as inputs to generation.

One can reconstruct the impetus for the above question in a deep pessimism about whether one can create an unambiguous RL. This is an important issue, though different from the original one. Briefly, there are two ways in which a representation language can be ambiguous. First, when one and the same RL representation can correspond to two or more non-synonymous NL texts and, secondly, when one and the same NL text can be represented with two or more distinct representations. In the latter case, an added issue is how to establish whether these distinct representations are synonymous or maybe differ only in the grain size of description (which might be considered allowable variation for the purposes of NLP applications). I believe that occurrences of ambiguity of the former kind are to be avoided if possible in RLs.

YW: Yes, but that is not what I mean by ambiguity in RLs. I see no difference between what you describe and the question of whether a passage of, say, Italian and of French are synonymous. I have no problem allowing that they are, modulo Quinean doubts. My worry is about the symbols that comprise them, be they RLs or NLs. Fodor faces this problem no more than do formalists who write "runs(John)" and appear to simply know which of the may senses of "run" the symbol bears in that context.

SN: One reason is that formalists do not usually work with an RL which has an interpreted vocabulary (they sometimes refer to such an entity when they talk about models in model-theoretic semantics). Of course, runs(John) is not an expression in a natural language even if "run" is taken as an unspecified sense of the English word. The reason is almost trivial: the parentheses and the intended assignment of a function-like character to "runs" and argument-like character to "John". Next, the artificial language used in this notation assumes an interpreter with human semantic capacity because a particular sense of the English word "run" must be selected, as well as a particular sense of john. Of course, the most blatant abuse of the similarity with NL and the most overt proof that such notations expect human interpreters is the ending -s on the predicate. Logicians would not think twice of writing "run(John&Mary)" fully expecting the interpreter to understand that "runs" and "run" are identical! It is only in this sense we can say that this representation is like English. The formal system in which such language is used never bothers to explain formally this reliance on the human processor, concentrating instead on studying the formal manipulations of symbols.

YW: Your run/runs point is a nice one and tells on my side, I feel, at least as to how formalists and would-be formalists actually use formulae in a casual, self-deceptive, way. Shall we now ask, what are the essential properties of being language-like and does a representational language have any of those properties, accidentally or necessarily?

SN: Suppose I set these out in a table as follows, where a plus against one of our initials means agreement with the purported feature to the left (as applied to NL or RL symbols), a minus means disagreement, a query means uncertainty, and +/- means both yes and no.

	NL	RL
ambiguous as to sense	SN:+ YW:+	SN:- YW:+
extensible	SN:+ YW:+	SN:+/- YW:+
constructed (vs. accepted)	SN:- YW:-	SN:+ YW:+
structures hierarchical (JF)		
non-compositional		
presupposes human processor:		
– at processing time	SN:+ YW:+	SN:- YW:-
– at acquisition time	SN:+ YW:+	SN:+ YW:- (?)
primitives are:		
NL words/phrases	SN:+ YW:+	SN:- YW:+
NL word-senses	SN:- YW:?	SN:+/- YW:-

YW: Let us put this matter very crudely: you seem to believe that the classificatory hierarchy of, say, WordNet [13] consists of English words while that of Mikrokosmos does not. To me they seem both to consist of an ascending thesaurus of similar terms up to some notional top nodes, but I can see no difference between them IN PRINCIPLE, only in richness of structure. This point is merely about the interpretation of symbols in a verbal/ontology hierarchy and how they are normally interpreted.

SN: Normally interpreted by WHOM? By people? Or by computer programs?

One of the differences between us may simply relate to how information is stored in the static knowledge sources of an NLP system, and used as a basis for inferences on representations. If the representation is ambiguous, as I believe it would if RL were an NL, then the inference system would have to disambiguate RL symbols. The decision to retain ambiguity in an RL leads to the situation where disambiguation occurs AFTER a representation is obtained.

YW: I can accept of course that life would be simpler, if duller, if NLs consisted of unambiguous symbols, and the same goes for RLs. What I cannot grasp here is that it is, as you say, a matter of decision, to let a representation be ambiguous or not. I cannot understand that. Let me ask again at its simplest: how can you believe the elements in Mikrokosmos, or Schank's CD [21], or anything else like that, are other than English words, with their own sense ambiguity?

SN: This question may be understood in at least two different ways. Let me first comment on the one for which I feel I have a better repartee: surely RL elements are not just additional senses of NL words with which they share the ASCII codes. You wouldn't say that the Spanish MAYOR is another sense of the English MAYOR, would you? As to the second interpretation of your statement, let me just say that having a separate language for primitives helps to explain paradigmatic relations among word- and phrase-senses, such as synonymy, antonymy etc. Of course, as in WordNet, one can bypass this explicitness about relations through the use of devices like synsets, but then one ends up with a knowledge source which does not support all the operations necessary for automatic meaning processing.

YW: OK, so we are moving into Quinean territory now, as when he used the impossibility of veridical paraphrase and translation to attack the folk-content notion of meaning [19]. It would seem natural if we want to defend the latter (what shall we call it–Commonsense semantics?) that we also believe in some provable/demonstrable notion of paraphrase. I would argue that even the present poor quality Information Extraction (IE) and MT are steps towards a practical notion of paraphrase–but that's not a philosophical defense.

SN: Do you mean that in order to defend the feasibility of meaning representation one needs to defend the feasibility of paraphrasing and translation? Clearly, even paraphrases and translations made by humans do not always convey EXACTLY the same meaning. The simple defense might be that philosophers habitually operate with an ideal RL and do not take into account the notion of the grain size of description, to say nothing about the possibility of a slip of judgment or an outright error on the part of acquirers.

YW: There is a venerable tradition of describing meaning through translation and paraphrase without representing it separately. It was Frege who seems to have wanted a functional notion of word meaning representation, or sense, (Sinn) that related surface entities (referents, or Bedeutungen, for Frege) but which did not ITSELF YIELD OR POSSESS CONTENT. For Frege [5] sense does not contain a coded meaning: it is just a function that allows you to specify or locate plausible referents in the world: a black box, or a sort of recognizer if you like!

SN: And that brings us back to the issue of whether a representation of text meaning is required or can simply be pointed at, and compared with the meaning of another text OR connected with a particular denotatum (which is in the real world, not in another set of symbols). In reality, we MODEL the world of denotata using a set of symbols because we simply cannot avoid it if we want to develop computer applications which require meaning representations.

YW: Well, of course, that is exactly what connectionists deny—they think they can give some sense to non-symbolic models–but I don't suppose we need bother with them here. And I admit that my obsessive questions about the exact status of primitives in a KR (and whether they are NL words or not) ignores Frege's best known injunction which was not to consider the sense of the symbol OUTSIDE AN EXPRESSION. Will we get any closer to resolution here by considering a possible scale of NL likeness for an RL:

English or Bulgarian
Esperanto
Predicate calculus
Some Interlingua

Are these all equally expressive and if so how could we know or prove it? If they are then that is one NL/RL link: anything one can say in one one can say in the other. Certainly many users of Predicate calculus and Interlingual formalisms have held this position of equal expressivity.

SN: An RL must support automatic inferencing operations. One might just consider the difficulty (or otherwise) of adapting any of the above kinds of RLs to this task. The major consideration is, again, whether the language is intended for people or for machines. The answer is easy in the case of English, Bulgarian and Esperanto; more problematic for predicate calculus; and impossible for interlinguas, which can be constructed. Any two independently constructed grammars of a language will be different though may well have the same weak or even strong generative capacity. The ideal interlingua would be good for both computers and people: it would support inferencing in a broad domain, thus permitting high-quality meaning representation for texts; and it would also be easy to repair and expand (which, for the foreseeable future will remain largely a task for people).

YW: I am not sure this tells one way or the other on the language-likeness issue, though it does make one ask if we have a good notion of "equivalent coverage"" for representational languages in the way we do for grammars.

A key phrase that may help clarify our difference is that you say NLs are comprised of words and RLs of word-senses. But language research is different from other AI areas because, in all areas but language, we can imagine a computer system being better than us: better than physicians or grandmaster chess players. We CANNOT imagine the system understanding language better than people, and this point is not often appreciated in some NLP areas.

SN: At the risk of sounding like a broken record, I'd like to insist that the purpose of a representation is to get the symbol ambiguity out, which is exactly what you think cannot be done. But does that point require an objective measure of symbol ambiguity, anywhere in our discussion or outside it?

YW: Somewhere in his discussion of what he calls "The Concrete Lexicon versus The Abstract Dictionary", Martin Kay [9] seemed to be arguing that the brain MUST subscript symbols to separate the senses of (brain) RL primitives within the Concrete Lexicon, i.e. the head. I have never been sure quite what he meant, but he was clearly discussing the same issue as us, as many have before, and he seemed to me to be roughly on my side here: conceding that RL atoms could be ambiguous and this would have to be resolved by the processor that used them, that is, the brain itself. Indeed, this is what you said earlier would be needed, and is not the same as the RL expressions being ambiguous, which you take me to mean here and I do not.

SN: If we must talk about the brain, I am agnostic. I don't know an awful lot about what is going on inside that device.

YW: No, neither do I, but people like us who talk about the nature of RLs for human knowledge must, like all AI workers, be making at least potential claims about the brain, whether we admit it or not.

SN: I wonder whether we indeed do. Maybe if we concentrate on representation by computers and for computers, we will be off the hook.

2.2 The second question: are languages necessarily extensible?

YW: How can anything that is a language be other than extensible? If that is obvious, one can then ask how can such extended information about a language be acquired. This could be seen as a traditional Chomskian question [3] about language and the child's learning of L1, its first language, but we intend it in the more accessible sense of an enquiry about how a computer can come to acquire new information about language, and whether that could ever be equated with the mastery of a merely finite, static, resource.

SN: Of course, language is extensible. However, any sublanguage used in an application, has, up till now, been finite and static. In AI applications, acquisition of knowledge typically precedes its use. When a new word must be entered in the lexicon of an MT system, it has been done by people or, at least, sanctioned by people. One can argue that the associated representation language was static and was used on any new input text as such, until the need for further extensions arose.

YW: Another of our key questions here is whether this feature of language is universal and, if so, must it be also be possessed by RLs, too?

SN: It is well known that people have difficulty recognizing ambiguities; they immediately choose the contextually appropriate sense for each word or phrase. This seems to suggest that, if indeed meanings are represented, the elements of the representation are not ambiguous, as the operation of retrieving the other senses of an input language element is so expensive.

YW: Ah, yes, this is Wittgenstein's [28] famous point that "the senses of a word do not pass in front of my mind". But your point does not, to me, prove anything about the nature of the representation: it is only a point about our lack of ACCESS to our processes. And in any case I am not claiming that representations are ambiguous: only that the items in them can be ambiguous (out of context presumably) in just the way NL items can. A difference of emphasis between us reflects our intellectuial upbringings, You, I fear, focus on the whole representation (in RL), I on the RLs constituents!

Do we therefore need to discuss the issue of what it is to know, or assess, objectively, in some sense, that a symbol in a representational system/language is ambiguous (within or out of context). It is clear from the variation of lexicographic intuitions (10 senses for a given word versus 2, in different dictionaries) that mere intuition is not enough. Remember, too, Wierzbicka's argument [24] that polysemy is mostly an illusion.

SN: Surely, lexicographic intuitions are about NL, not necessarily RL. That lexicographers disagree may simply mean that there does not exist some "correct" number of senses. I intuitively dislike the suggestion (that there is such a number), but maybe in some system-operational approach, one could define word senses cross-linguistically. This latter point connects with the idea of using an almost Hjelmslevian [7] view of the semes across languages as an impetus for humans to select senses for representation even in an internalized RL.

YW: Yes, the translation-as-representation case, between NLs, has had a new lease of life recently hasn't it, and it is a strange reprise of the Fodorian comedy of the LOT as the translation one can't get at. I used to suffer the temptation at meetings to ask Fodor how he KNEW the LOT wasn't, say, Latin, but I fortunately never gave way to it, since I know he doesn't know.

More seriously, and given that LISP was considered almost a Language-of-Thought by AIers in the seventies: consider NIL in LISP, now usually thought of as 3-ways ambiguous (an empty list, an atom, and a Boolean value). Was there an objective test of that? Did it matter until it was noticed, in terms of the usefulness or otherwise of LISP? Was there a formal criterion for spotting it: i.e. is "giving a formal semantics of a representation" a revelatory mechanism for exposing "ambiguity"? I suspect not.

SN: The fact that a lexical ambiguity in a representation language can be contextually "benign" does not necessarily prove that ambiguity can be introduced with similar impunity into RLs designed for the purpose of representing meanings of texts.

YW: Agreed, and as we know, NLs, unlike RLs, can be metalanguages for themselves, and this is probably a point on your side showing a clear NL/RL difference. Though I still do not need to concede, what you insist on, that we can allow or prevent ambiguity in RLs. These matters are under no one's control: in RLs like CYC [10] no one was able to control the coders' use of the predicates effectively. There is no RL/NL distinction there, where you seem to want it for RL coding, and this, for me, rebuts your earlier claim that applications are static and finite.

The case of corpus statistics may be interesting here because its users (e.g. [1]) generally have no use for terms like "word sense" which they find unbearably intuitive; for them, symbols simply occur in environments which may or may not be usefully separable into classes of occurrence.

I am not sure there is any objective demonstration of the ambiguity of a symbol, which would require showing the Reality of Word Senses? I have always used the Schvaneveldt Pathfinder nets [22] as a justification; they can show "bank" having separable subgraphs with an algorithm that requires no seeding or stimulation to do that. The other well-known statistical methods usually do not show ambiguity unless you assume it to start with.

SN: Your position here is similar to that of the "lexical-rule"-oriented lexical semanticists (e.g. [18] who prefer to propose few (usually, one) word sense for recording in the dictionary, and then to add rules for accommodating meanings that do not directly conform to statically defined constraints. This single word sense is, indeed, ambiguous. Unfortunately, the generative lexicon approach does not discuss the vocabulary of the representation language in any detail.

YW: No, I am not assuming single senses for words, nor lexical rules for creating dictionaries. Look at this a slightly different way: the relation between an expression in NL and in its corresponding RL may be either a relationship like that between a language and its metalanguage or one of (presumably mutual) translation. If the former case holds, then the languages need not really differ in type; they simply have an asymmetric relationship and might differ in expressiveness, but, as is well known, a meta-RL is as much in need of its meta-language as the object-NL. There is an agreement in the formal world to stop worrying about this, and probably rightly, but, if the relationship is of that sort, there is no reason to believe the two levels differ over, say, polysemousness or extensibility of meanings.

Alternatively, if the relationship is one of translation, then, almost by definition, TRANSLATE (X, Y) if X and Y are both symbolic, requires that X and Y be of the same TYPE, that is, both are NL-like, in this case!

SN: Of course, we cannot tolerate an infinite regression of metalanguages. The relation between NL and RL is, to me, asymmetrical, though there will be both many to one relations between elements of NL and RL (e.g., synonymy) and one-to-many ones (most notably, polysemy). Internal consistency is achieved for RL through maintaining the complex cross-relationships in an ontology (the RL vocabulary). The issue of meaning grounding is more difficult and we might want to state, cautiously, that it is achieved through the multiple connections of elements of an RL with multiple NLs, through human judgement of quality of translation correspondence.

Your argument about the relation of translation hinges centrally on how one defines TYPE. It may be that we do not disagree, but you elect to stress similarities between NL and RL while I persist in looking for differences. Let them be of the same type, but RLs must support machine inferencing while this cannot be asked of NLs. The case of the Dutch company BSO working with Esperanto as its interlingua [27] for MT clearly showed how much a human-oriented (though invented) language had to be modified in order to serve as a kind of RL. Even the developers themselves, Esperanto enthusiasts all, had to call the new language somewhat differently: BCE or "binary-coded Esperanto."

2.3 The third question: are acquisition and extensibility linked?

YW: Acquisition in our sense is linked to the necessity or otherwise of symbol ambiguity, because much acquisition (especially automatic acquisition, i.e. machine learning) is of new ambiguities or senses of symbols.

SN: The extent to which automatic acquisition of content is possible may indeed be a major practical undercurrent of this paper. A question for you: does explaining the meaning of an ambiguous symbol in terms of another ambiguous symbol actually constitute disambiguation?

YW: This a practical question, too, of course. We are seeking, in our everyday research, and outside dialogues like this, practical, robust, NL processors, not necessarily wedded to one particular theory, but ones that tackle areas of NL and KR representations. I am, in a sense, rather neutral about particular representations but strong on assessment and large systems and data. On your question: again, I accept that an (ambiguous) symbol can be defined, more or less, by a string that is not, as a whole, ambiguous.

An assumption about communication behind all this is that the trivial diagram we are all familiar with of humans communicating with their separate representations (in head balloons) via the very narrow linear language stream from their mouths, is wrong in one crucial respect. It is normally shown with the SAME structure in the two heads. But there is no reason at all to believe that human communication requires identical logics, lexicons, grammars, parsers etc. in both heads, any more than it does identical beliefs.

I suggest the most striking feature of communication is that humans who differ about these structures can communicate, just as can individuals with different dialects, or those writing to others at later historical periods.

SN: Yes, there should be no presupposition of a similarity between the knowledge and processing resources of various people, modulo the hardware (wetware?) and possibly some other, perhaps genetic, constraints. The difference is clear in the case of conversations between people who are native speakers of different languages or belonging to different professional and social strata, people of different ages, etc. It is indeed amazing how adaptable people are when viewed as information processors. At the same time, on the surface, what this shows is only that there may be as many "proprietary" devices for processing language as there are people.

YW: The commonsense fact is that communication can take place within a bandwidth of difference, and human-computer communication in a way explores the limits of this bandwidth and how far it can be extended in special cases by tuning lexicon structures and beliefs to each other in the course of communication itself. But this issue cannot be separated from the problem of language representation itself, for we cannot understand the nature of the representation of meaning in lexicons, say, unless we can see how to extend lexicons in the presence of incoming data that does not fit the lexicon we started with. Extension of representation is part of an adequate theory of representation.

SN: I think I understand your intended meaning: first, no set of static knowledge sources will have complete coverage; therefore, representations need to be extensible; therefore there must be a mechanism of adding elements to representations, preferably, on the fly.

Further, many of such representation elements are lexical. And the easiest way of naming these new elements would be through the natural language strings that refer to them in the input and which triggered the representation augmentation process in the first place.

This, of course, presupposes automatic acquisition, because if a human is involved in acquisition other suggestions could become quite palatable. In short, the argument for allowing natural language into a representation becomes thus also practical: we need it because otherwise we will have problems naming new atoms.

YW: Supppose we write

$$\text{I: structure1 X corpus} \rightarrow \text{structure2}$$

as a basic model of acquisition of a representational structure, be it an ontology or a lexicon, to indicate that a state of the structure itself plays a role in the acquisition, of which structure2 is then a proper extension (capturing new concepts, senses etc). This is a different model from the wholly automatic model of lexicon aquisition in, say TIPSTER related work (e.g. 20), which can be written:

$$\text{II: corpus} \rightarrow \text{structure}$$

This case is one which does not update or "tune" an existing lexicon but derives one directly and automatically from a corpus. We are arguing the essential role of representational structure in this process, and hence the first process, which we may also take to involve some essential human intervention as well. But whatever is the case about that, we are not discussing the ab initio / tabula rasa case. Interestingly perhaps, neither of these is an analogue to the Chomskian approach to (first) language acquisition [3], which might be written:

$$\text{III: Universal-Constraints X corpus} \rightarrow \text{structure}$$

If the constraints here are of the same format as a lexicon structure then this third form is closer to I above, especially in Fodor's work, where the constraints become a sort of primitive-ontology or -lexicon.

SN: This classification seems to skirt the issue of human involvement. In reality, fully automatic acquisition of lexical information does not, at this time, go anywhere deep enough to yield material of use in solving hard problems such as full-text lexical disambiguation or even syntactic analysis. In TIPSTER, for instance, as far as I know, the automatic acquisition of subcategorization patterns for some English verbs was accompanied by massive manual acquisition. Personally, I would choose to use a combination of all three of the above methods of acquisition, depending on the quality of the input data and availability of good-quality constraints and structures.

2.4 The fourth question: if automatic acquisition is possible, what are the consequences for or against RLs as NLs?

YW: If automatic acquisition of content is possible to any degree, from a Machine-Readable Dictionary or corpus then, since those are plainly in NL, does this suggest that in some form NL is a representation language for information about language, and that settles the issue discussed raised earlier.

SN: First of all, I think that this premise is a moot point at the moment, because automatic acquisition of content can be considered possible only if content is plainly trivial. Any success in the automatic acquisition of content is predicated on the ability of the developers to model (in the weak sense, with no claims of similarity of the model to the modeled other than at output!) the disambiguation and other meaning assignment processes of humans. More concretely, this modeling involves overt, human-directed, formulation, at the time of acquisition, of the background knowledge and processes which support the automatic assignment of meaning at processing time.

But even if the premise of your argument is given, the argument itself still seems to be a bit of a sleight of hand. It is rather similar methodologically to the use by our colleagues at USC ISI (e.g.[8]) of the fact that the ontology in the Pangloss DARPA-funded MT project used English as its metalanguage: the Spanish lexicon in that project explained the meanings of Spanish words in terms of an ontology whose atoms were homographs of English words and expressions.

YW: Well, if they can do it, I might want to say it is not a sleight of hand but proof of my NL-RL point. I also want to use the metaphor of a dictionary as containing a lexicographer's "conscious, explicit, knowledge", which is what we might extract by these processes–but other computations over the result could yield meaning connections no lexicographer had actually seen (and which might be said to model his unconscious).

SN: A representation needs to be reformulated and fleshed out for machines. Lexicographers in writing (printed) dictionary entries heavily (if subconsciously!) rely on the fact that their representations, such as definitions, will be processed by a high-quality language processor, namely, the human! This may be the very crux of our disagreement. The task of NLP knowledge acquirers is to use their language processing capacity to state information as overtly as possible given a desired grain-size of description AND in a format which facilitates access by machine (e.g., frames). The latter condition is, of course, of secondary importance: it is a convenience consideration only. The former condition is contentful in that it presumes that the definition is not complete by itself but only together with the human understander of that definition. This can be proven wrong, incidentally, if it is shown that dictionary entries, in fact, do not rely on extraneous human knowledge in specifying definitions. But if that were so, why do lexicographers say that if you don't know some meaning, you won't understand it from the dictionary? Is this just frivolity?

YW: It is frivolity and, if true in general, would make their products useless. I still think it is an open question whether structures derived pretty much automatically from MRDs can be useful for NLP [26]. If they are your position weakens. Is our difference really one of bottom-up versus top-down approaches to the same information? You believe that the acquisition of the core of these knowledge resources can be done only semi-automatically, but under human supervision: for instance, in automatic production of lexicon entries through lexical rules.

2.5 The fifth question: the relationship of these issues to representations for humans and for machines.

SN: Representations for humans assume the presence of an extremely powerful analysis system and a huge amount of background knowledge. One has to specify things at a much finer grain of description for machines than for humans even if the purposes of the two descriptions are compatible.

YW: This is an excellent question and not as much discussed as it should be. A difference in machine versus human handling of representations used to be called the Gensym issue (e.g. [12]): a machine can handle English expressed as arbitrary Gensyms substituted for words, but a human native speaker cannot without vast retraining, if then. We can both accept the difference between the comprehensible representations that humans need and the fact that they have no meaning for machines, and use it to prove opposite views as regards NL and RL. Your observation proves to me that, for that very reason, RLs must be accessible to humans (as well as machines) and THEREFORE must be NL like in certain respects.

SN: This is a weaker form of your original argument about NLs as RLs, and I would fully agree with the premise: just like the computer programs, which are written in part to be read by people (an estimated 80% of the time of software engineers is spent on maintenance: that is, reading and improving other peoples' code), so should the knowledge structures in an RL. That in, say, Mikrokosmos names of atoms are words or phrases in English is due exactly to this fact. It is, on my view, a conceptual fallacy to read more into this state of affairs: for instance, to claim that there is an intrinsic necessity for RL elements to be also elements of an NL.

YW: Charniak's final argument [2] against connectionism was that you couldnt understand the structures such systems acquired; and they were therefore not acceptable representations, regardless of whether machines could use them or not. How much of our underlying disagreement is over whether structure must be comprehensible?

SN: Well, to comprehend anything which is non-trivial, one must learn. One can, in fact, learn to read meaning representations. It has been proved in practice. Of course, it is very desirable to avoid having people read unadorned RL structures, but this might be a premature hope.

YW: Would we be helped by thinking about how actual coders use RLs? An example that has interested me is that of how some Japanese researchers use interlinguas for, say, MT or in the Tokyo EDR dictionary project [17], but with English symbols. It has been argued that it may be an advantage for them because they do not, in many cases, see more than the main sense for any primitive and this makes its use easier and less confusing than for a native speaker of the interlingua, if you will allow that term. The question might then be: has that fact any analogies with how you see a machine as handling a representation: the difference between human and machine handling of representations being, I think, crucial for your position, though not for mine?

SN: The analogy with understanding by machines is clear: they operate with fewer word and phrase senses (to say nothing about connotations) than people. However, I do not see any bearing that this observation has on the differences between RLs and NLs. If the Japanese researchers you mention do not know English well enough, this does not impinge in any way on the issue of whether RLs for computers should be either bad or good English or any other NL, or an artificial language (with either narrow or broad coverage of meanings in an NL).

YW: Maybe when we model understanding we aim at too high a target: in ordinary situations people may understand just a fraction of what is said by a speaker but they ask clarification questions only when it matters. In reality, there are few penalties for failures such as miscommunication or misunderstanding: contrast medical counselling dialogue, legal searches, patents, and philosophical discussions, in all of which misunderstanding is thought disastrous and maybe carries real costs.

I have a feeling we may have swapped sides here a bit. Part of our difference may arise from what one could call my Wittgensteinian prejudices [28 from bad early training perhaps, which cause me to think language central and unreplaceable in thought and representations so that there will never be any alternative to doing what we do now –whatever happens to AI or computational linguistics—–because we are self-defined by language and we can't expunge it from representations.

SN: If the issue here is that, however people may try, they will not be able to produce RLs which are not ambiguous, it is, or will be, a verifiable matter. Possibly, this will be happening asymptotically. But it is surely not plausible that people are somehow constitutionally unable to come up with an unambiguous RL, not because of the size and complexity of the problem (which can be ameliorated through tools, partial automation etc.) but rather by definition! We would need to go through a much more detailed discussion of the influence of the fallibility of human acquirers on the nature of the RL: Sapir-Whorfian [23] influences of native tongues, difficulties with listing all senses of a lexeme or all synonyms of a word, as opposed to the human faculty of judgment as to whether any two words are synonymous.

YW: Ah, so at the end we really do differ. I think it is beyond human ability to design an RL without the features they now have, and for the reason we touched on: they must remain comprehensible to us, and if they do they will be like NL, where I quite accept that "like" inevitably remains a bit fuzzy. It is as if comprehensibility will carry a price; which will be loss of control of the sort you think we can retain.

SN: But it is exactly your understanding of the meaning of "like" which is the crux of the matter here! As long as it is fuzzy, one cannot very well argue about it. Further, I assume that by "control" you mean whether people can be taught deliberately to produce representations that, as they or their project managers believe, would be processable by machines. As I understand it, you think this implausible because vestiges of human language will remain and corrupt the RL representations. I think it necessary, unless we can teach machines to reason using knowledge bases which are inconsistent or ambiguous. Mind you, I do not have any illusions about the practical attainability of knowledge bases which are fully consistent and unambiguous. The methodological choice is to carry on pretending that they are until special mechanisms are developed for dealing with such inconsistencies and ambiguities.

YW: Our misunderstandings persist to the end: vestiges of NL in an RL does not mean for me that a machine cannot "understand" it in a particular application. Of course not, it is happening all the time in millions of working programs.

SN: Right. So long as we agree on what constitutes understanding, but that would need another conversation!

References

[1] Brown, P., V. Pietra, P. deSouza, J. Lai, and R. Mercer, Class-based n-gram models of natural language, Computational Linguistics, 18(4) (1992).
[2] Charniak, E. Connectionism and Explanation, in: Y. Wilks, ed., Theoretical Issues in Natural Language Processing. (Erlbaum, Hillsdale, NJ, 1991).
[3] Chomsky, N. Aspects of the Theory of Syntax. (MIT Press, Cambridge, MA, 1965).
[4] Fodor, J.A. Psychosemantics. (MIT Press, Cambridge, MA, 1987).
[5] Frege, G. On sense and reference, in: P. Black, ed., Translations from the Philosophical Works of Gottlob Frege. (Blackwell, Oxford, 1960).
[6] Hirst, G. Existence assumptions in knowledge representantation, Artificial Intelligence, 49 (1991).
[7] Hjelmslev, L. Structural Analysis of Language, in: B. Malmberg, ed., Readings in Modern Linguistics. (Mouton, The Hague, 1972).
[8] Hovy, E.H. and K. Knight, Motivation for Shared Ontologies: An Example from the Pangloss Collaboration, Proceedings of the Workshop on Knowledge Sharing and Information Interchange, IJCAI-93. (Chambery, France 1993).
[9] Kay, M. The Concrete Lexicon and the Abstract Dictionary, Proceedings of the 5th Annual Conference of the UW Centre for the New Oxford English Dictionary. (Oxford, England, 1989).
[10] Lenat, D., M. Prakash, and M. Shepherd, CYC: Using common sense knowledge to overcome brittleness and knowledge acquisition bottlenecks, The AI Magazine, 6(4) (1986).
[11] McCarthy, J. Epistemological Problems of Artificial Intelligence, Proceedings of the 5th International Joint Conference on Artificial Intelligence (1977).
[12] McDermott, D. Artificial Intelligence meets Natural Stupidity, in: J. Haugheland, ed., Mind Design. (MIT Press, Cambridge, MA, 1976).
[13] Miller, G., R. Beckwith, C. Fellbaum, D. Gross and K. Miller, WordNet: on on-line lexical data-base, International Journal of Lexicography 3(4) (1990).
[14] Newell, A. Artificial Intelligence and the Concept of Mind, in: R. Schank and K. Colby, eds., Computer Models of Thought and Language. (Freeman, San Francisco, 1973).
[15] Nirenburg, S. and C. Defrise, Practical Computational Linguistics, in: R. Johnson and M. Rosner, eds., Computational Linguistics and Formal Semantics. (Cambridge University Press, Cambridge, 1991).
[16] Nirenburg, S. and V. Raskin, Ten Choices for Lexical Semantics. (MCCS-96-304, CRL, NMSU, 1996).
[17] Nomura, N. Functions of the set of concept explications in an MTD and methodology for developing bilingual concept explications, Proceedings of the EDR Workshop. (University of Pennsylvania, 1993).
[18] Pustejovsky, J. The Generative Lexicon. (MIT Press, Cambridge, MA, 1995).
[19] Quine, W. V. O. The problem of meaning in linguistics, in: From a Logical Point of View. (Harper and Row, New York, 1963).

[20] Riloff, E. Automatically constructing a dictionary for information extraction, Proceedings of the 11th National Conference on Artificial Intelligence (1993).
[21] Schank, R., ed., Conceptual Information Processing. (Amsterdam, North Holland, 1972).
[22] Schvaneveldt, R., ed., Pathfinder Networks: Theory and Applications. (Ablex, Norwood, NJ, 1990).
[23] Whorf, B.L. Language, Thought and Reality, ed. J. Carroll, (Wiley, New York, 1956).
[24] Wierzbicka, A. Semantic Primitives. (Athanaeum, Frankfurt, 1972)
[25] Wilks, Y. Form and Content in Semantics, in: R. Johnson and M. Rosner, eds., Computational Linguistics and Formal Semantics. (Cambridge University Press, Cambridge, 1992).
[26] Wilks, Y., L. Guthrie and B. Slator, Electric Words: Dictionaries, Computers and Meanings. (MIT Press, Cambridge, MA, 1996)
[27] Witkam, A.P.M. Distributed Language Translation: Feasibility Study of a Multilingual Facility for Videotex Information Networks. (BSO, Utrecht, 1983)
[28] Wittgenstein, L. Philosophical Investigations. (Blackwell, Oxford, 1958)