From NLP to NLU

Robert D. Van Valin, Jr.
Heinrich Heine University Düsseldorf
University at Buffalo, The State University of New York

A commonsensical idea about the difficult task of getting computers to understand human language would be that the job would involve crucial input from linguistics. Linguistics is, after all, the scientific study of language, and linguists have developed theories of the relation between form (words, sentences, etc.) and meaning. The investigation of form falls primarily into the domains of morphology (concerning the form of words) and syntax (concerning the structure of phrases and sentences), while the analysis of meaning belongs to the domains of semantics (roughly, the meanings of words, phrases and sentences) and pragmatics (roughly, inferred meanings based on the use of language). The best known theories of the structure of language have been proposed by Noam Chomsky of the Massachusetts Institute of Technology. Chomsky's theories have dominated linguistics since the 1960's and been highly influential throughout the field of cognitive science, which Chomsky helped to create.

It might, therefore, come as a surprise that linguistics does not play a dominant role in the field of natural language processing [NLP] by computer. Rather, statistical models are preeminent in NLP. Generalizing from the co-occurrence patterns of words in numerous pre-analyzed texts, statistical models induce the rules governing the patterns and learn to recognize similar patterns in unanalyzed texts. This is termed 'machine learning'. It underlies most applications involving natural language, from Watson to Siri. Why did computer scientists turn away from linguistic models?

The answer to this question can be found in changes in linguistic theory initiated by Chomsky in the 1970's and 1980's. In his earliest models, from the 1950's and 1960's, syntactic rules were procedural in nature, e.g. 'move a question word like *who* to the beginning of the sentence in a question', and the approach to the creation of a sentence was a derivation consisting of a series of procedural operations.[1] Such a sequence of procedures could readily be modeled by a computer algorithm. However, the goals for linguistic theory that Chomsky championed did not include modeling how sentences are actually constructed or interpreted by speakers in real time. He proposed a distinction between linguistic competence, a native speaker's knowledge of the grammar of his/her native language, and linguistic performance, the application of this competence in actual situations. Linguistic theory, Chomsky maintained, concerns itself exclusively with the investigation of linguistic competence and its acquisition by children; only when competence had been fully understood and described could performance be studied. The primary data used in the study of competence are a native speaker's judgements as to whether a particular sentence is grammatical or not. Given this focus on grammatical competence, the system of procedural rules was shown to be extremely problematic as a means for modeling it, and Chomsky turned to what has come to be known

---

[1] Noam Chomsky, *Syntactic Structures* (Mouton, 1957) and *Aspects of the Theory of Syntax* (MIT Press, 1965).

as the 'principles and parameters' approach.[2] In it a grammar does not consist of a set of procedural rules but rather a set of abstract principles which show some parametric variation reflecting cross-linguistic differences. (For example, there are general principles which govern the structure of phrases, e.g. verb + object [VP] or preposition + object [PP], which are parameterized to reflect the fact that some languages have the object after the V or P (English) and other have it before the V or P (Japanese).) Whether a sentence is grammatical or not is determined by its compliance with all of the parameterized principles that make up the grammar of the language, and there is no set of procedural rules as before. This new approach proved very difficult, if not impossible, to implement algorithmically, and computational linguists and computer scientists were faced with a situation in the early 1980's in which the dominant, most widely accepted linguistic theory was not suitable for computational NLP work. So they began to explore other options, and statistical approaches came to be paramount in the field.[3]

These statistical approaches have been very successful to date, but they have their limitations, as any one who has had problems communicating with a digital assistant or gotten weird translations from a translation program knows. Watson, arguably one of the most impressive systems, famously defeated a human champion on the TV game show 'Jeopardy'. But at a lecture at Stanford University in 2012, one of the leaders of the Watson team made the following remark as he ended his talk: "The only advantage the human contestant had over Watson was that he understood the questions."[4] This leads us to a crucial point: statistical systems can accomplish natural language *processing* to a considerable degree, but they cannot achieve natural language *understanding* [NLU], which necessarily involves meaning, something which purely statistical approaches cannot capture. Something more is needed. To start down the path from NLP to NLU we have to go back to linguistics.

Human language has often been characterized as a systematic relationship between form and meaning, and the many linguistic theories which have been proposed over the years since the advent of Chomsky's theory represent different formulations of this relationship and different ways of capturing it formally. In order to make the transition to NLU possible, a theory must have a rich semantic representation which can be related to the form of sentences by a set of computationally tractable rules or principles, and just such a theory is found in Role and Reference Grammar [RRG], a theory whose development I have been involved in since the late 1970's.[5] RRG is well-suited to the task, for the reasons to be detailed below, and it fits well with Patom Theory, as John Ball's initial implementation work has shown.[6]

---

[2] Noam Chomsky, *Lectures on Government and Binding* (Reidel, 1981).

[3] See Brian Roark & Richard Sproat, *Computational Approaches to Morphology and Syntax* (Oxford UP, 2007).

[4] See Rob High, *IBM Watson*, available at http://www.redbooks.ibm.com/redpapers/pdfs/redp4955.pdf. On p. 4 there is a section heading "IBM Watson understands language" followed on the next page by "However, Watson does not really understand the individual words in the language." Clearly, the verb *understand* is not being used in the same sense in these two passages, and thus, crucially, Watson's so-called "understanding" is not based on meaning the way a human being's understanding of language is.

[5] William Foley & Robert Van Valin, Jr., *Functional Syntax and Universal Grammar* (Cambridge UP, 1984), Robert Van Valin, Jr. & Randy LaPolla, *Syntax: Structure, Meaning and Function* (Cambridge UP, 1997), Robert Van Valin, Jr., *Exploring the Syntax-Semantics Interface* (Cambridge UP, 2005). See also the RRG homepage, http://linguistics.buffalo.edu/people/faculty/vanvalin/rrg.html for my 'Brief Overview of Role and Reference Grammar' for a concise introduction to the theory.

[6] John Ball, 'Patom Theory', Pat, Inc. White Paper.

The basic architecture of RRG is given in (1).

(1)   Semantic representation <——Linking algorithm——> Syntactic representation
        [meaning]   [form]

The linking system takes a semantic representation and maps it into possible sentence forms which can express that meaning; this is semantics-to-syntax (meaning-to-form) linking. The system, however, is bi-directional, which is why the arrows point in both directions: it also takes a sentence form and maps it into the meaning it expresses; this is syntax-to-semantics (form-to-meaning) linking. Why is this bi-directionality important? Because it represents what speakers and hearers do. Speakers formulate a message and map it into a sentence form and utter it (meaning-to-form linking), whereas hearers perceive a sentence and give it an interpretation (form-to-meaning linking). NLU is about modeling what hearers do, and RRG's architecture includes the linking from form to meaning that is at the heart of NLU. One might ask, doesn't every theory have this structure? The answer is 'no'. Most theories claim to be direction 'neutral', but in practice their rules and principles are formulated primarily in terms of what is essentially meaning-to-form mapping. Thus, RRG's structure makes it particularly well-suited for the requirements of a system capable of NLU.

A second important aspect of RRG is its use of enriched semantic representations to capture the core meaning of verbs. This is done by means of what is called *lexical decomposition*. This can be illustrated with the verb *shatter* in English. There are two main uses of this verb: (i) as a change of state predicate, e.g. *The window shattered*, and (ii) as an induced change of state predicate, e.g. *The man shattered the window*. In the first case, the window goes from being whole to being in pieces, and no external cause is attributed to the event. In the second case, however, some action by the man is asserted to be the cause of the window undergoing a change of state. These two meanings of *shatter* can be represented as in (2).

(2) a. *The window shattered*
     a′. BECOME **be.in.pieces′** (window)
     b. *The man shattered the window*
     b′. [**do′** (man, Ø)] CAUSE [BECOME **be.in.pieces′** (window)]

The phrases *the man* and *the window* are called 'arguments' of the verb in (2). They are arguments from both syntactic and semantic perspectives: syntactically they function as subject or direct object in the sentence, and semantically they occupy positions in the semantic representation. The meaning of *the window shattered* is a part of the meaning of *the man shattered the window*, and this is directly represented in (2b′). The formula in (2b′) can be understood as 'the man does something unspecified which causes the window to come to be in pieces'. The usefulness of lexical decomposition as a way to represent meaning can be seen when we look at other languages. In Lakhota, a Native American language spoken in the Dakotas,[7] these two uses of *shatter* would be coded differently. The basic element is *bléča*, which means 'a flat, brittle object is in pieces', and to it is added a prefix which signals how the action which causes this state of affairs is done. For the meaning in (2a), where there is no outside force acting on the window, the prefix would be *na-*, which means 'by internal force',

---

[7] See Jan Ullrich, *New Lakota Dictionary,* 2nd. Ed. (Lakota Language Consortium, 2014)

yielding (3a). If there is an outside force involved, as in (2b), then it is necessary to specify it; for example, *ka-* signals 'to cause something to happen by striking', or *wo-* 'to cause something to happen by action from a distance', yielding (3b) and (3c). Rough semantic representations of (3b) and (3c) are given in (4). (In technical RRG work, the Lakhota words for 'man' and 'window' would occur in the representations.)

(3) a. *Ožáŋžaŋglepi kiŋ **nab**léče.*   'The window shattered.'
      window   the shatter

    b. *Wičháša kiŋ ožáŋžaŋglepi kiŋ **kab**léče.*   'The man shattered the window [by hitting
      man   the window   the shatter   it].'

    c. *Wičháša kiŋ ožáŋžaŋglepi kiŋ **wob**léče.*   'The man shattered the window [by throw-
      man   the window   the shatter   ing something through it or shooting it].'

(4) a. [**do´** (man, [**strike´** (man, window)])] CAUSE [BECOME **be.in.pieces´** (window)]
      = (3b)

    b. [**do´** (man, [**affect.from.distance´** (man, window)])] CAUSE [BECOME **be.in. pieces´** (window)] = (3c)

Thus, the structure of the Lakhota verb more directly reflects the semantic contrast between (2a) and (2b) than its English counterpart, which has the same form for both meanings, and lexical decomposition allows one to see the semantic similarities and differences between the English verb *shatter* and the Lakhota verbs *nabléča, kabléča* and *wobléča.*

    We are now in a position to illustrate how the linking system in (1) works for these English and Lakhota sentences. The exposition is oversimplified in many respects, but the point is to give a sense of how the system works, rather than to get bogged down in the full range of complexities. In the semantic representations, *the man* is the doer of the action, which is indicated by '**do´** (man, ...', and *the window* is the undergoer of a change of state from not being in pieces to being in pieces, the change being represented by 'BECOME' and the resulting state by '**be.in.pieces´**'. In the mapping from semantics to syntax, the essential issue is to select the subject and direct object, because they will be mapped into specific positions in the clause in these two languages. In (2b), (3b) and (3c), the doer of the action will function as the subject of the sentence and the undergoer of the change of state will function as the direct object. This is true for many verbs in the two languages. In both languages the subject normally occurs first in the sentence, but direct object does not occur in the same place in them: in English it follows the verb, whereas in Lakhota it comes before the verb. With respect to (2a) and (3a), there is only one candidate for the subject in the semantic representation, and it will occur as subject. Aside from the position of the direct object, there is another important difference between the two languages: in English, but not in Lakhota, it is possible to have the undergoer serve as the subject, the doer of the action occurs as a non-subject with special marking (in English, marked by the preposition *by*), and the verb occurs with special marking (in English, in a special form together with the auxiliary *be*, as in (5).

(5) The window was shattered by the man.

This option is found in some but not all languages. The semantic representation of (5) is the same as that of (2b), namely (2b´), because they both mean the same thing. The sentence in

(5) is a coding option for the meaning expressed in (2b´). The result of the linking from semantics to syntax is the sentences in (2), (3) and (5).

For the linking from syntax to semantics, which is a crucial part of NLU, we start from the sentences in (2), (3) and (5) and map them into the semantic representations given in (2) and (4). When there is only one syntactic argument of the verb, as in (2a) and (3a), then it must be the subject. There are more than one semantic representations associated with the verb in both languages, and because there is only one syntactic argument in the sentence, the semantic representation with only one argument must be selected, in order for the linking to be successful. Thus the single syntactic argument is linked to the single argument position in the semantic representation.

Things can be more complex with two argument verbs, as in (2b), (3b), (3c) and (5). In these Lakhota sentences the situation is relatively straightforward: the subject is the doer of the action, and the direct object is the undergoer. Because English allows both (2b) and (5), the situation is more interesting, as the subject is not always the doer of the action, as (5) shows. In languages like English, the interpretation of the subject is determined by the form of the verb: when the form is 'plain', e.g. *shattered, has shattered, can shatter, will shatter*, etc., then the subject is the doer of the action. On the other hand, if it is a special form, namely *be* plus *shattered*, e.g. *was shattered, has been shattered, can be shattered, will be shattered*, then the subject is the undergoer, not the doer, and the doer occurs as the object of the preposition *by*, as in (5). The system associates the sentences in (2b) and (5) with the same semantic representation. The endpoint of the syntax-to-semantics linking is the same as the starting point for the semantics-to-syntax linking, namely the semantic representations in (2) and (4), and the linking algorithm in (1) which does the mapping between form and meaning models central aspects of speech production and, more important for our concerns, crucial aspects of speech comprehension, NLU.

This brief discussion of the linking system highlights another important feature of RRG in contrast with statistical systems: multilinguality is an inherent feature of RRG. No linguistic theory is designed to deal with only one language, but RRG is especially strong in this regard, since from its inception it was designed to deal with so-called 'exotic' languages, e.g. Philippine, Papuan, Native American and Australian Aboriginal languages. Statistical systems are by definition monolingual, as they are trained on data sets from a single language, yielding statistical patterns that apply to that language only. In order to deal with more than one language, an NLP system is typically trained on parallel corpora in the two languages under consideration, yielding statistical correlations that can be compared across them. The result is not true multilingual processing but rather 'multi-parallel-monolingual' processing. Contrast this with the RRG system. First, the linking algorithm for each language is worked out, and there are a lot of cross-linguistically valid principles and generalizations in the theory, which means that one is not starting from scratch with each new language. Second, going from one language to another is mediated by the meanings expressed; this is summarized in the following figure ('LA' stands for 'linking algorithm').
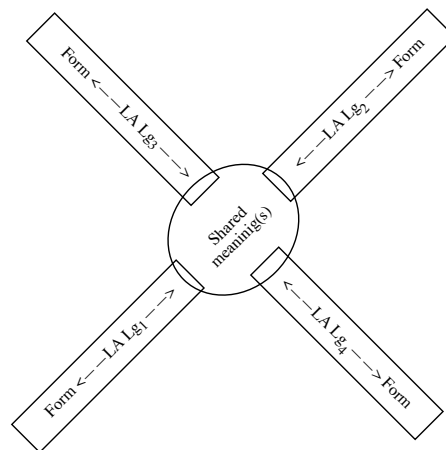
The transition between languages is mediated by the semantic representations that are common to them, as illustrated in the simple example from English and Lakhota above. The English sentences in (2b) or (5) would be linked to the semantic representation in (2b′), which is then compared with the Lakhota semantic representations in (4). They are more specific with respect to the causing activity than the English representation and therefore are both compatible with it, which is not specific about the nature of the causing action. The Lakhota linking algorithm would then map the semantic representations into the possible sentences in (3b) and (3c), which are the Lakhota translation equivalents of the English sentences we started with.[8]

A further example of the importance of having a representation of meaning can be seen in the contrast between the sentences in (6).

(6) a. Sam started a new novel.
    b. Sam started a ham sandwich.

The first sentence can mean either that Sam started reading a new novel or that he started writing one. The second, on the other hand, can mean that Sam started eating a ham sandwich or that the started fixing himself one. The first can't mean that Sam began eating or fixing himself a new novel, nor can the second mean that he began reading or writing a ham sandwich. Given that the verb is the same in the two sentences, where does the interpretive difference come from? These examples were first discussed by James Pustejovsky, a semanticist and computational linguist, who argued that the interpretations derive from properties of the direct objects, which he termed 'qualia'.[9] Two of these properties are important here: the

---

[8] As a further example of this, see the RRG-based machine translation system from Arabic to English presented in Brian Nolan & Yasser Salem 'UniArab: RRG Arabic-to-English Machine Translation' (in W. Nakamura, ed., *New Perspectives in Role and Reference Grammar*, 312-46, Cambridge Scholars Publishing, 2011). This program does not employ Patom Theory and so is rather different from the system developed by John Ball.

[9] James Pustejovsky, *The Generative Lexicon* (MIT Press, 1995).

telic property, which concerns what an object is used for, and the agentive property, which concerns how an object comes into being. By means of the notation for semantic representation introduced earlier, the relevant qualia properties can be expressed as in (7).

(7) a. *Novel*
   Telic: **do´** (x, [**read´** (x, novel)])
   Agentive: **do´** (x, [**write´** (x, novel)])
   b. *Ham sandwich*
   Telic: **do´** (x, [**eat´** (x, ham sandwich)])
   Agentive: **do´** (x, [**make´** (x, ham sandwich)])

The sentences in (6) lack a verb specifying what Sam started to do, and this information is plugged in from the direct object's qualia: he either started reading or writing a novel, or he started eating or making a ham sandwich. These interpretations are inferences derived from aspects of the meaning of the noun phrases in the sentence, and it is difficult to see how a purely statistical analysis of co-occurrence patterns would be able to arrive at them.

Thus, if we are to make the transition from natural language processing to natural language *understanding*, meaning must be involved, as understanding, unlike processing, necessarily involves grasping the meaning of a linguistic expression. To do that we need a representation of meaning and a means for linking it to its syntactic form, and the theory of Role and Reference Grammar provides strongly cross-linguistically motivated means to achieve this. RRG, together with Patom Theory, takes us beyond statistically-driven natural language processing toward meaning-based natural language understanding.