Natural Language Semantics and Computability

Richard Moot, Christian Retoré

▶ **To cite this version:**

Richard Moot, Christian Retoré. Natural Language Semantics and Computability. Journal of Logic, Language and Information, 2019, 28 (2), pp.287-307. 10.1007/s10849-019-09290-7 . hal-01315316

HAL Id: hal-01315316
https://inria.hal.science/hal-01315316

Submitted on 13 May 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Natural Language Semantics and Computability

Richard Moot[1] and Christian Retoré[2]

[1] CNRS LaBRI
[2] Université de Montpellier & LIRMM

Abstract. This paper is a reflexion on the computability of natural language semantics. It does not contain a new model or new results in the formal semantics of natural language: it is rather a computational analysis of the logical models and algorithms currently used in natural language semantics, defined as the mapping of a statement to logical formulas — formulas, because a statement can be ambiguous. We argue that as long as possible world semantics is left out, one can compute the semantic representation(s) of a given statement, including aspects of lexical meaning. We also discuss the algorithmic complexity of this process.

Introduction

In the well-known Turing test for artificial intelligence, a human interrogator needs to decide, via a question answering session with two terminals, which of his two interlocutors is a man and which is a machine (Turing 1950). Although early systems like Eliza based on matching word patterns may seem clever at first sight, they clearly do not pass the test. One often forgets that, in addition to reasoning and access to knowledge representation, passing the Turing test presupposes automated natural language analysis and generation which, despite significant progress in the field, has not yet been fully achieved. These natural language processing components of the Turing test are of independent interest and used in computer programs for question answering and translation (however, since both of these tasks are generally assumed to be AI-complete it is unlikely that a full solution for these problems would be simpler than a solution for the Turing test itself).

If we define the semantics of a (sequence of) sentence(s) σ as the mapping to a representation $\phi(\sigma)$ that can be used by a machine for natural language processing tasks, two very different ideas of semantics come to mind.

1. One notion of semantics describes what the sentence(s) speaks about. The dominant model for this type of semantics represents meaning using word vectors (only involving referential/full words nouns, adjectives, verbs, adverbs, ... and not grammatical words) which represent what σ speaks about. This is clearly computable. One must fix a thesaurus of n words that acts as a vector basis. Usually words not in the thesaurus or basis are expanded into their definition with words in the thesaurus. By counting occurrences of words from the thesaurus in the text (substituting words not in the thesaurus with their definition) and turning this into a n-dimensional vector reduced to be of euclidian norm 1, we obtain word meanings in the form of

n-dimensional vectors. This notion of semantics provides a useful measure of semantic similarity between words and texts; typical applications include exploring Big Data and finding relevant pages on the internet. This kind of semantics models what a word (or a text) speaks about.

2. The other notion of semantics, the one this paper is about, is of a logical nature. It models what is asserted, refuted, ... assumed by the sentences. According to this view, computational semantics is the mapping of sentence(s) to logical formula(s). This is usually done compositionally, according to Frege's principle *"the meaning of a compound expression is a function of the meaning of its components"* to which Montague added *"and of its syntactic structure"*. This paper focuses on this logical and compositional notion of semantics and its extension (by us and others) to lexical semantics; these extensions allow us to conclude from a sentence like "I started a book" that the speaker started *reading* (or, depending on the context, writing) a book.

We should comment that, in our view, semantics is a (computable) function from sentence(s) to logical formulae, since this viewpoint is not so common in linguistics.

- Cognitive sciences also consider the language faculty as a computational device and insist on the computations involved in language analysis and production. Actually there are two different views of this cognitive and computational view: one view, promoted by authors such as Pinker (1994), claims that there is a specific cognitive function for language, a "language module" in the mind, while others, like Langacker (2008), think that our language faculty is just our general cognitive abilities applied to language.
- In linguistics and above all in philosophy of language many people think that sentences cannot have any meaning without a context, such a context involving both linguistic and extra-linguistic information. Thus, according to this view, the input of our algorithm should include context. Our answer is firstly that linguistic context is partly taken into account since we are able to produce, in addition to formulae, discourse structures. Regarding the part of context that we cannot take into account, be it linguistic or not, our answer is that it is not part of semantics, but rather an aspect of pragmatics. And, as argued by Corblin (2013), if someone is given a few sentences on a sheet of paper without any further information, he starts imagining situations, may infer other statements from what he reads, ..., and such thoughts are the semantics of the sentence.
- The linguistic tradition initiated by Montague (1974) lacks some coherence regarding computability. On the one hand, Montague gives an algorithm for parsing sentences and for computing their meaning as a logical formula. On the other hand, he asserts that the meaning of a sentence is the interpretation of the formula in possible worlds, but these models are clearly uncomputable! Furthermore, according to him, each intermediate step, including the intensional/modal formulae should be forgotten, and the semantics is defined as the set of possible worlds in which the semantic formula is true: this cannot even be finitely described, except by these intermediate formulas; a fortiori it cannot be computed. Our view is different, for at least three reasons, from the weakest to the strongest:

- Models for higher order logic, as in Montague, are not as simple as is sometimes assumed, and they do not quite match the formulas: completeness fails. This means that a model and even all models at once contains less information than the formula itself.
- We do not want to be committed to any particular interpretation. Indeed, there are alternative relevant interpretations of formulas, as the following non exhaustive list shows: dialogical interpretations (that are the sets of proofs and/or refutations), game theoretic semantics and ludics (related to the former style of interpretation), set of consequences of the formula, structures inhabited by their normal proofs as in intuitionistic logic,...
- Interpreting the formula(s) is no longer related to linguistics, although some interpretations might useful for some applications. Indeed, once you have a a formula, interpreting it in your favourite way is a purely logical question. Deciding whether it is true or not in a model, computing all its proofs or all its refutations, defining game strategies, computing its consequences or the corresponding structure has nothing to do with the particular natural language statement you started with.

1 Computational semantics à la Montague

We shall first present the general algorithm that maps sentences to logical formulae, returning to lexical semantics in Section 2. The first step is to compute a syntactic analysis that is rich and detailed enough to enable the computation of the semantics (in the form of logical formulae). The second step is to incorporate the lexical lambda terms and to reduce the obtained lambda term — this step possibly includes the choice of some lambda terms from the lexicon that fix the type mismatches.

1.1 Categorial syntax

In order to express the process that maps a sentence to its semantic interpretation(s) in the form of logical formulae, we shall start with a categorial grammar. This is not strictly necessary: Montague (1974) used a context free grammar (augmented with a mechanism for quantifier scope), but if one reads between the lines, at some points he converts the phrase structure into a categorial derivation, so we shall, following Moot & Retoré (2012), directly use a categorial analysis. Although richer variants of categorial grammars are possible, and used in practice, we give here an example with Lambek grammars, and briefly comment on variants later.

Categories are freely generated from a set of base categories, for example np (noun phrase), n (common noun), S (sentence), by two binary operators: $\backslash$ and $/$: $A \backslash B$ and B / A are categories whenever A and B are categories. A category $A \backslash B$ intuitively looks for a category A to its left in order to form a B. Similarly, a category B / A combines with an A to its right to form a B. The full natural deduction rules are shown in Figure 1.

A lexicon provides, for each word w of the language, a finite set of categories $lex(w)$. We say a sequence of words $w_1, \ldots, w_n$ is of type C whenever $\forall i \exists c_i \in lex(w_i)\ c_1, \ldots, c_n \vdash C$. Figure 2 shows an example lexicon (top) and a derivation of a sentence (bottom).

$$\frac{\Gamma \vdash A \quad \Delta \vdash A \backslash B}{\Gamma, \Delta \vdash B} \backslash_e \qquad \frac{A, \Gamma \vdash B}{\Gamma \vdash A \backslash B} \backslash_i$$

$$\frac{\Delta \vdash B / A \quad \Gamma \vdash A}{\Gamma, \Delta \vdash B} \backslash_e \qquad \frac{\Gamma, A \vdash B}{\Gamma \vdash B / A} \backslash_i$$

Fig. 1. Natural deduction proof rules for the Lambek calculus

Word	*Syntactic Type*
kid	n
cartoon	n
watched	$(np \backslash S) / np$
every	$(S / (np \backslash S)) / n$
a	$((S / np) \backslash S) / n$

$$\dfrac{\dfrac{\dfrac{\dfrac{\textit{every} \quad \textit{kid}}{(S/(np\backslash S))/n \quad n}/_e \quad \dfrac{\textit{watched}}{(np\backslash S)/np \quad [np]^1}/_e}{(S/(np\backslash S))}\backslash_e}{\dfrac{S}{S/np}/_i(1)} \qquad \dfrac{\dfrac{\textit{a} \quad \textit{cartoon}}{((S/np)\backslash S)/n \quad n}/_e}{(S/np)\backslash S}\backslash_e}{S}$$

Fig. 2. Lexicon and example derivation

1.2 From syntactic derivation to typed linear lambda terms

Categorial derivations, being a proper subset of derivations in multiplicative intuitionistic linear logic, correspond to (simply typed) linear lambda terms. This makes the connection to Montague grammar particularly transparent.

Denoting by e the set of entities (or individuals) and by t the type for propositions (these can be either true or false, hence the name t) one has the following mapping from syntactic categories to semantic/logical types.

$(\text{Syntactic type})^* = \text{Semantic type}$	
$S^* = t$	a sentence is a proposition
$np^* = e$	a noun phrase is an entity
$n^* = e \to t$	a noun is a subset of the set of entities (maps entities to propositions)
$(A \setminus B)^* = (B / A)^* = A^* \to B^*$	extends easily to all syntactic categories

Using this translation of categories into types which forgets the non commutativity, the Lambek calculus proof of Figure 2 is translated to the linear intuitionistic proof shown in Figure 3; we have kept the order of the premisses unchanged to highlight the similarity with the previous proof. Such a proof can be viewed as a simply typed lambda term with the two base types e and t.

$$(a^{(e\to t)\to((e\to t)\to t)}\ cartoon^{e\to t})(\lambda y^e(every^{(e\to t)\to((e\to t)\to t)}\ kid^{e\to t})(watched^{e\to e\to t}\ y))$$

$$\dfrac{\dfrac{\dfrac{\dfrac{every}{(e\to t)\to(e\to t)\to t}\quad\dfrac{kid}{(e\to t)}}{(e\to t)\to t}\to_e\quad\dfrac{\dfrac{watched}{e\to e\to t}\quad\dfrac{y}{[e]^1}}{e\to t}\to_e}{t}\to_e}{\dfrac{t}{e\to t}\to_{i(1)}}\qquad\dfrac{\dfrac{a}{(e\to t)\to(e\to t)\to t}\quad\dfrac{cartoon}{(e\to t)}}{(e\to t)\to t}\to_e$$

$$t\ \to_e$$

Fig. 3. The multiplicative linear logic proof corresponding to Figure 2

As observed by Church (1940), the simply typed lambda calculus with two types e and t is enough to express higher order logic, provided one introduces *constants* for the logical connectives and quantifiers, that is a constants "$\exists$" and "$\forall$" of type $(e \to t) \to t$, and constants "$\wedge$", "$\vee$" et "$\Rightarrow$" of type $t \to (t \to t)$.

In addition to the syntactic lexicon, there is a semantic lexicon that maps any word to a simply typed lambda term with atomic types e and t and whose type is the translation of its syntactic formula. Figure 4 presents such a lexicon for our current example. For example, the word "every" is assigned formula $(S/(np\setminus S))/n$. According to the translation function above, we know the corresponding semantic term must be of type $(e \to t) \to ((e \to t) \to t)$, as it is in Figure 3. The term we assign in in the semantic lexicon is the following (both the type and the term are standard in a Montagovian setting).

$$\lambda P^{e\to t}\ \lambda Q^{e\to t}\ (\forall^{(e\to t)\to t}\ (\lambda x^e(\Rightarrow^{t\to(t\to t)}\ (P\ x)(Q\ x))))$$

Unlike the lambda terms computed for proof, the lexical entries in the semantic lexicon need not be linear: the lexical entry above is not a linear lambda term since the single abstraction binds two occurrences of x.

Similarly, the syntactic type of "a", the formula $((S/np)\backslash S)/n$ has corresponding semantic type $(e \to t) \to ((e \to t) \to t)$ (though syntactically different, a subject and an object generalized quantifier have the same *semantic* type), and the following lexical meaning recipe.

$$\lambda P^{e\to t}\ \lambda Q^{e\to t}\ (\exists^{(e\to t)\to t}\ (\lambda x^{e}(\wedge^{t\to(t\to t)}(P\,x)(Q\,x))))$$

Finally, "kid", "cartoon" and "watched" are assigned the constants $\texttt{kid}^{e\to t}$, $\texttt{cartoon}^{e\to t}$ and $\texttt{watched}^{e\to(e\to t)}$ respectively.

word	***syntactic type*** u ***semantic type*** u^* ***semantics:*** λ***-term of type*** u^*
every	$(S/(np\backslash S))/n$ (subject) $(e\to t)\to((e\to t)\to t)$ $\lambda P^{e\to t}\ \lambda Q^{e\to t}\ (\forall^{(e\to t)\to t}\ (\lambda x^{e}(\Rightarrow^{t\to(t\to t)}\ (P\,x)(Q\,x))))$
a	$((S/np)\backslash S)/n$ (object) $(e\to t)\to((e\to t)\to t)$ $\lambda P^{e\to t}\ \lambda Q^{e\to t}\ (\exists^{(e\to t)\to t}\ (\lambda x^{e}(\wedge^{t\to(t\to t)}(P\,x)(Q\,x))))$
kid	n $e\to t$ $\lambda x^{e}(\texttt{kid}^{e\to t}\ x)$
cartoon	n $e\to t$ $\lambda x^{e}(\texttt{cartoon}^{e\to t}\ x)$
watched	$(np\backslash S)/np$ $e\to(e\to t)$ $\lambda y^{e}\ \lambda x^{e}\ ((\texttt{watched}^{e\to(e\to t)}\ x)\ y)$

Fig. 4. Semantic lexicon for our example grammar

Because the types of these lambda terms are the same as those of the words in the initial lambda term, we can take the linear lambda term associated with the sentence and substitute, for each word its corresponding lexical meaning, transforming the derivational semantics, in our case the following[3]

$$(a^{(e\to t)\to((e\to t)\to t)}\ cartoon^{e\to t})(\lambda y^{e}(every^{(e\to t)\to((e\to t)\to t)}\ kid^{e\to t})(watched^{e\to e\to t}\ y))$$

into an (unreduced) representation of the meaning of the sentence.

$$((\lambda P^{e\to t}\ \lambda Q^{e\to t}(\exists^{(e\to t)\to t}\ (\lambda x^{e}(\wedge^{t\to(t\to t)}(P\,x)(Q\,x)))))\,\texttt{cartoon}^{e\to t})$$
$$((\lambda y^{e}(((\lambda P^{e\to t}\ \lambda Q^{e\to t}\ (\forall^{(e\to t)\to t}\ (\lambda x^{e}(\Rightarrow^{t\to(t\to t)}\ (P\,x)(Q\,x)))))\,\texttt{kid}^{e\to t}\ x)))(\texttt{watched}^{e\to e\to t}y))$$

[3] There are *exactly* two (non-equivalent) proofs of this sentence. The second proof using the same premisses corresponds to the second, more prominent reading of the sentence whose lambda term is: $(every\ kid)(\lambda x^{e}.(a\ cartoon)(\lambda y^{e}((watched\ y)\ x))$

The above term reduces to

$$(\exists^{(e\to t)\to t}\ \lambda x^{e}(\wedge^{t\to(t\to t)}(\texttt{cartoon}\,x)(\forall^{(e\to t)\to t}\ (\lambda z^{e}(\Rightarrow^{t\to(t\to t)}\ (\texttt{kid}\,z)((\texttt{watched}\,x)\,z))))))$$

that is[4]: $\exists x.\texttt{cartoon}(x)\wedge\forall z.\texttt{kid}(z)\Rightarrow\texttt{watched}(z,x)$

The full algorithm to compute the semantics of a sentence as a logical formula is shown in Figure 5.

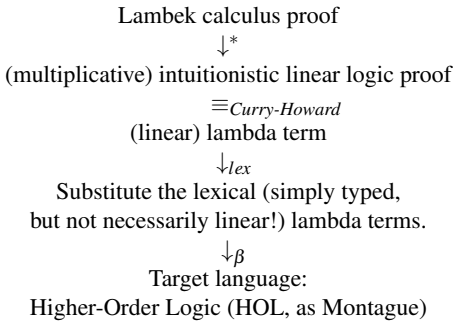

Fig. 5. The standard categorial grammar method for computing meaning

2 Adding sorts, coercions, and uniform operations

Montague (as Frege) only used a single type for entities: e. But it is much better to have many sorts in order to block the interpretation of some sentences:

(1) * The table barked.

(2) The dog barked.

(3) ?The sergeant barked.

As dictionaries say "*barked*" can be said from animals, usually dogs. The first one is correctly rejected: one gets $\texttt{bark}^{dog\to t}(\texttt{the table})^{artifact}$ and $dog\neq artifact$.

However we need to enable the last example $\texttt{bark}^{dog\to t}(\texttt{the sergeant})^{human}$ and in this case we use coercions (Bassac et al. 2010, Retoré 2014): the lexical entry for the verb "*barked*" which only applies to the sort of "*dogs*" provides a coercion $c : human\to dog$ from "*human*" to "*dog*". The revised lexicon provides each word with the lambda term that we saw earlier (typed using some of the several sorts / base type) and some optional lambda terms that can be used if needed to solve type mismatches.

Such coercions are needed to understand sentences like:

[4] We use the standard convention to translate a term $((p\,y)\,x)$ into a predicate $p(x,y)$.

(4) This book is heavy.

(5) This book is interesting.

(6) This book is heavy and interesting.

(7) Washington borders the Potomac.

(8) Washington attacked Iraq.

(9) * Washington borders the Potomac and attacked Iraq.

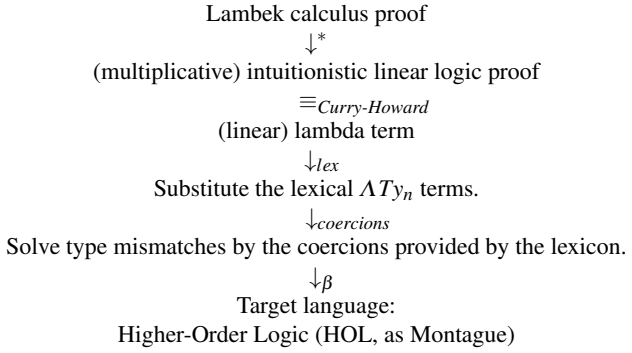

Fig. 6. Computing meaning in a framework with coercion

The first two sentences will respectively use a coercions from book to physical object and a coercion from books to information. Any time an object has several related meanings, one can consider the conjunction of properties referring to those particular aspects. For these operations (and others acting uniformly on types) we exploit polymorphically typed lambda terms (system F). When the related meanings of a word are incompatible (this is usually the case) the corresponding coercions are declared to be incompatible in the lexicon (one is declared as rigid). This extended process is described in Figure 6. Some remarks on our use of system F:

- We use it for the syntax of semantics (a.k.a. metalogic, glue logic)
- The formulae of semantics are the usual ones (many sorted as in Ty_n)
- We have a single constant for operations that act uniformly on types, like quantifiers or conjunction over predicates that apply to different facets of a given word.

3 Complexity of the syntax

As we remarked before, when computing the formal semantics of a sentence in the Montague tradition, we (at least implicitly) construct a categorial grammar proof. Therefore, we need to study the complexity of parsing/theorem proving in categorial

grammar first. The complexity generally studied in this context is the complexity of deciding about the existence of a proof (a parse) for a logical statement (a natural language sentence) as a function of the number of words in this sentence[5].

Perhaps surprisingly, the simple product-free version of the Lambek calculus we have used for our examples is already NP-complete (Savateev 2009). However, there is a notion of order, which measures the level of "nesting" of the implications as defined below.

$$order(p) = 0$$
$$order(A/B) = order(B\backslash A) = \max(order(A), (order(B) + 1))$$

As an example, the order of formula $(np\backslash S)/np$ is 1, whereas the order of formula $S/(np\backslash S)$ is 2. For the Lambek calculus, the maximum order of the formulas in a grammar is a good indication of its complexity. Grammars used for linguistic purposes generally have formulas of order 3 or, at most, 4. We know that once we bound the order of formulas in the lexicon of our grammars to be less than a fixed n, parsing becomes polynomial for any choice of n (Pentus 2010)[6].

The NP-completeness proof of Savateev (2009) uses a reduction from SAT, where a SAT problem with c clauses and v variables produces a Lambek grammar of order $3+4c$, with $(2c+1)(3v+1)$ atomic formulas.

The notion of order therefore provides a neat indicator of the complexity: the NP-completeness proof requires formulas of order 7 and greater, whereas the formulas used for linguistic modelling are of order 4 or less.

Even though the Lambek calculus is a nice and simple system, we know that the Lambek calculus generates only context-free languages (Pentus 1995), and there is good evidence that at least some constructions in natural language require a slightly larger class of languages (Shieber 1985). One influential proposal for such a larger class of languages are the mildly context-sensitive languages (Joshi 1985), characterised as follows.

- contains the context-free languages,
- limited cross-serial dependencies (i.e includes $a^n b^n c^n$ but maybe not $a^n b^n c^n d^n e^n$)
- semilinearity (a language is semilinear iff there exists a regular language to which it is equivalent up to permutation)
- polynomial fixed recognition[7]

[5] For many algorithms, the complexity is a function of the number of atomic subformulas of the formulas in the sentence. Empirically estimation shows the number of atomic formulas is a bit over twice the number of words in a sentence.

[6] For the algorithm of Pentus (2010), the order appears as an exponent in the worst-case complexity: for a grammar of order n there is a multiplicative factor of $2^{5(n+1)}$. So though polynomial, this algorithm is not necessarily efficient.

[7] The last two items are sometimes stated as the weaker condition "constant growth" instead of semilinearity and the stronger condition of polynomial parsing instead of polynomial fixed recognition. Since all other properties are properties of formal languages, we prefer the formal language theoretic notion of polynomial fixed recognition.

There are various extensions of the Lambek calculus which generate mildly context-sensitive languages while keeping the syntax-semantics interface essentially the same as for the Lambek calculus. Currently, little is known about upper bounds of the classes of formal languages generated by these extensions of the Lambek calculus. Though Moot (2002) shows that multimodal categorial grammars generate exactly the context-sensitive languages, Buszkowski (1997) underlines the difficulty of adapting the result of Pentus (1995) to extensions of the Lambek calculus[8].

Besides problems from the point of view of formal language theory, it should be noted that the goal we set out at the start of this paper was not just to generate the right string language but rather to generate the right string-meaning pairs. This poses additional problems. For example, a sentence with n quantified noun phrases has up to $n!$ readings. Although the standard notion of complexity for categorial grammars is the complexity deciding whether or not a proof exists, formal semanticists, at least since Montague (1974), want their formalisms to generate all and only the correct readings for a sentence: we are not only interested in whether or not a proof exists but, since different natural deduction proofs correspond to different readings, also in what the different proofs of a sentence are[9].

When we look at the example below

(10) Every representative of a company saw most samples.

it has five possible readings (instead of $3! = 6$), since the reading

$$\forall x.\texttt{representative\_of}(x,y) \Rightarrow \texttt{most}(z,\texttt{sample}(z)) \Rightarrow \exists y.\texttt{company}(y) \wedge \texttt{see}(x,z)$$

has an unbound occurrence of y (the leftmost occurrence). The Lambek calculus analysis has trouble with all readings where "a company" has wide scope over at least one of the two other quantifiers. We can, of course, remedy this by adding new, more complex types to the quantifier "a", but this would increase the order of the formulas and there is, in principle, no bound on the number of constructions where a medial quantifier has wide scope over a sentence. A simple counting argument shows that Lambek calculus grammars cannot generate the $n!$ readings required for quantifier scope of an n-quantifier sentence: the number of readings for a Lambek calculus proof is proportional to the Catalan numbers and this number is in $o(n!)$[10]; in other words, given a Lambek

[8] We can side-step the need for a Pentus-like proof by looking only at fragments of order 1, but these fragments are insufficient even for handling quantifier scope.

[9] Of course, when our goal is to generate (subsets of) $n!$ different proofs rather than a single proof (if one exists), then we are no longer in NP, though it is unknown whether an algorithm exists which produces a sort of shared representation for all such subsets such that 1) the algorithm outputs "no" when the sentence is ungrammatical 2) the algorithm has a fairly trivial algorithm (say of a low-degree polynomial at worst) for recovering all readings from the shared representation 3) the shared structure is polynomial in the size of the input.

[10] We need to be careful here: the number of readings for a sentence with n quantifiers is $\Theta(n!)$, whereas the maximum number of Lambek calculus proofs is $O(c_0^{c_2 n} C_{c_1 c_2 n})$, for constants c_0, c_1, c_2 which depend on the grammar (c_0 is the maximum number of formulas for a single word, c_1 is the maximum number of (negative) atomic subformulas for a single formula and c_2 represent the minimum number of words needed to add a generalized quantifier to a sentence,

calculus grammar, the number of readings of a sentence with n quantifiers grows much faster than the number of Lambek calculus proofs for this sentence, hence the grammar fails to generate many of the required readings.

Since the eighties, many variants and extensions of the Lambek calculus have been proposed, each with the goal of overcoming the limitations of the Lambek calculus. Extensions/variations of the Lambek calculus — which include multimodal categorial grammars (Moortgat 1997), the Displacement calculus (Morrill et al. 2011) and first-order linear logic (Moot & Piazza 2001) — solve both the problems of formal language theory and the problems of the syntax-semantics interface. For example, there are several ways of implementing quantifiers yielding exactly the five desired readings for sentence 10 without appealing to extra-grammatical mechanisms. Carpenter (1994) gives many examples of the advantages of this logical approach to scope, notably its interaction with other semantic phenomena like negation and coordination.

Though these modern calculi solve the problems with the Lambek calculus, they do so without excessively increasing the computational complexity of the formalism: multimodal categorial grammars are PSPACE complete (Moot 2002), whereas most other extensions are NP-complete, like the Lambek calculus.

Even the most basic categorial grammar account of quantifier scope requires formulas of order 2, while, in contrast to the Lambek calculus, the only known polynomial fragments of these logics are of order 1. Hence the known polynomial fragments have very limited appeal for semantics.

Is the NP-completeness of our logics in conflict with the condition of polynomial fixed recognition required of mildly context-sensitive formalisms? Not necessarily, since our goals are different: we are not only interested in the string language generated by our formalism but also in the string-meaning mappings. Though authors have worked on using mildly context-sensitive formalisms for semantics, they generally use one of the two following strategies for quantifier scope: 1) an external mechanism for computing quantifier scope (e.g. Cooper storage, (Cooper 1975)), or 2) an underspecification mechanism for representing quantifier scope (Fox & Lappin 2010).

For case 1 (Cooper 1975), a single syntactic structure is converted into up to $n!$ semantic readings, whereas for case 2, though we represent all possible readings in a single structure, even deciding whether the given sentence has a semantic reading *at all* becomes NP-complete (Fox & Lappin 2010), hence we simply shift the NP-completeness from the syntax to the syntax-semantics interface[11]. Our current understanding therefore indicates that NP-complete is the best we can do when we want to generate the semantics for a sentence. We do not believe this to be a bad thing, since pragmatic and processing constraints rule out many of the complex readings and enumerating all readings of sentences like sentence 10 above (and more complicated examples) is a difficult task. There is a trade-off between the work done in the syntax and in the syntax-semantics interface, where the categorial grammar account incorporates more

---

i.e. $c_2 n$ is the number of words required to produce an n-quantifier sentence) and $O(c_0^{c_2 n} C_{c_1 c_2 n})$ is in $o(n!)$.

[11] In addition, Ebert (2005) argues that underspecification languages are not expressive enough to capture all possible readings of a sentence in a single structure. So underspecification does not solve the combinatorial problem but, at best, reduces it.

than the traditional mildly context-sensitive formalisms. It is rather easy to set up a categorial grammar parser in such a way that it produces underspecified representations in time proportional to n^2 (Moot 2007). However, given that such an underspecified representation need not have any associated semantics, such a system would not actually qualify as a parser. We believe, following Carpenter (1994) and Jacobson (2002), that giving an integrated account of the various aspects of the syntax-semantics interface is the most promising path.

Our grammatical formalisms are not merely theoretical tools, but also form the basis of several implementations (Morrill & Valentín 2015, Moot 2015), with a rather extensive coverage of various semantic phenomena and their interactions, including quantification, gapping, ellipsis, coordination, comparative subdeletion, etc.

4 Complexity of the semantics

The complexity of the syntax discussed in the previous section only considered the complexity of computing unreduced lambda terms as the meaning of a sentence. Even in the standard, simply typed Montagovian framework, normalizing lambda terms is known to be of non-elementary complexity (Schwichtenberg 1982), essentially due to the possibility of recursive copying. In spite of this forbidding worst-time complexity, normalization does not seem to be a bottleneck in the computation of meaning for practical applications (Bos et al. 2004, Moot 2010).

Is there a deeper reason for this? We believe that natural language semantics uses a restricted fragment of the lambda calculus, soft lambda calculus. This calculus restricts recursive copying and has been shown to characterize the complexity class P exactly (Lafont 2004, Baillot & Mogbil 2004). Hence, this would explain why even naive implementations of normalization perform well in practice.

The question of whether soft linear logic suffices for our semantic parser may appear hard to answer, however, it an obvious (although tedious) result. To show that all the semantic lambda terms can be typed in soft linear logic, we only need to verify that every lambda in the lexicon is soft. There is a finite number of words, with only a finite number of lambda terms per word. Furthermore, words from open classes (nouns, verbs, adjectifs, manner adverbs,... in which speakers may introduce new words... about 200.000 inflected word forms) are the most numerous and all have soft and often even linear lambda terms. Thus only closed class words (grammatical words such as pronouns, conjunctions, auxiliary verbs,... and some complex adverbs, such as "too") may potentially need a non-soft semantic lambda term: there are less than 500 such words, so it is just a matter of patience to prove they all have soft lambda terms. Of course, finding deep reasons (cognitive, linguistic) for semantic lambda terms to be soft in *any* language would be much more difficult (and much more interesting!).

When adding coercions, as in Section 2, the process becomes a bit more complicated. However, the system of Lafont (2004) includes second-order quantifiers hence reduction stays polynomial once coercions have been chosen. Their choice (as the choice of the syntactic category) increases complexity: when there is a type mismatch $g^{A\to X}u^{B}$ one needs to chose one of the coercions of type $B \to A$ provided by the entries of the words in the analysed phrase, with the requirement that when a rigid coercion is used,

all other coercions provided by the same word are blocked (hence rigid coercions, as opposed to flexible coercions decrease the number of choices for other type mismatches).

Finally, having computed a set of formulas in higher-order logic corresponding to the meaning of a sentence, though of independent interest for formal semanticists, is only a step towards using these meaning representations for concrete applications. Typical applications such as question answering, automatic summarization, etc. require world knowledge and common sense reasoning but also a method for deciding about entailment: that is, given a set of sentences, can we conclude that another sentence is true. This question is of course undecidable, already in the first-order case. However, some recent research shows that even higher-order logic formulas of the type produced by our analysis can form the basis of effective reasoning mechanisms (Chatzikyriakidis & Luo 2014, Mineshima et al. 2015) and we leave it as an interesting open question to what extent such reasoning can be applied to natural language processing tasks.

5 Conclusion

It is somewhat surprising that, in constrast to the well-developed theory of the algorithmic complexity of parsing, little is known about semantic analysis, even though computational semantics is an active field, as the recurring conferences with the same title as well as the number of natural language processing applications show. In this paper we simply presented remarks on the computability and on the complexity of this process. The good news is that semantics (at least defined as a set of logical formula) is *computable*. This was known, but only implicitly: Montague gave a set of instruction to compute the formula (and to interpret it in a model), but he never showed that, when computing such logical formula(s):

- the process he defined stops with a normal lambda terms of type proposition (t),
- eta-long normal lambda terms with constants being either logical connectives or constants of a first (or higher order) logical language are in bijective correspondence with formulas of this logical language (this is more or less clear in the work of Church (1940) on simple type theory).
- the complexity of the whole process has a known complexity class, in particular the beta-reduction steps which was only discovered years after his death (Schwichtenberg 1982).

A point that we did not discuss is that we considered *worst case complexity* viewed as a function from the *number of words in a sentence* a logical formula. Both aspects of our point of view can be challenged: in practice, grammar size is at least as importance as sentence length and average case complexity may be more appropriate than worst case complexity. Though the high worst case complexity shows that computing the semantics of a sentence is not always efficient, we nevertheless believe, confirmed by actual practice, that statistical models of a syntactic or semantic domain improve efficiency considerably, by providing extra information (as a useful though faillible "oracle") for many of the difficult choices. Indeed, human communication and understanding are very effective in general, but, from time to time, we misunderstand eachother or

need to ask for clarifications. For computers, the situation is almost identical: most sentences are analysed quickly, while some require more time or even defeat the software. Even though it is quite difficult to obtain the actual probability distribution on sentence-meaning pairs, we can simply estimate such statistics empirically by randomly selecting manually annotated examples from a corpus. The other aspect, the sentence length, is, as opposed to what is commonly assumed in complexity theory, not a very safisfactory empirical measure of performance: indeed the average number of words per sentence is around 10 in spoken language and around 25 in written language. Sentences with more than 100 words are very rare[12]. Furthermore, lengthy sentences tend to have a simple structure, because otherwise they would quickly become incomprehensible (and hard to produce as well). Experience with parsing shows that in many cases, the grammar size is at least as important as sentence length for the empirical complexity of parsing algorithms (Joshi 1997, Sarkar 2000, Gómez-Rodríguez et al. 2006). Grammar size, though only a constant factor in the complexity, tends to be a *big* constant for realistic grammars: grammars with between 10.000 and 20.000 rules are common.

We believe that the complexity of computing the semantics and of reasoning with the semantic representations are some of the most important reasons that the Turing test is presently out of reach.

[12] To given an indication, the TLGbank contains more than 14.000 French sentences and has a median of 26 words per sentence, 99% of sentences having less than 80 words, with outliers at 190 and at 266 (the maximum sentence length in the corpus).

Bibliography

Baillot, P. & Mogbil, V. (2004), Soft lambda-calculus: a language for polynomial time computation, *in* 'Foundations of software science and computation structures', Springer, pp. 27–41.

Bassac, C., Mery, B. & Retoré, C. (2010), 'Towards a type-theoretical account of lexical semantics', *Journal of Logic, Language and Information* **19**(2), 229–245.

Bos, J., Clark, S., Steedman, M., Curran, J. R. & Hockenmaier, J. (2004), Wide-coverage semantic representation from a CCG parser, *in* 'Proceedings of COLING-2004', pp. 1240–1246.

Buszkowski, W. (1997), Mathematical linguistics and proof theory, *in* J. van Benthem & A. ter Meulen, eds, 'Handbook of Logic and Language', Elsevier, chapter 12, pp. 683–736.

Carpenter, B. (1994), Quantification and scoping: A deductive account, *in* 'The Proceedings of the 13th West Coast Conference on Formal Linguistics'.

Chatzikyriakidis, S. & Luo, Z. (2014), 'Natural language inference in Coq', *Journal of Logic, Language and Information* **23**(4), 441–480.

Church, A. (1940), 'A formulation of the simple theory of types', *Journal of Symbolic Logic* **5**(2), 56–68.

Cooper, R. (1975), Montague's Semantic Theory and Transformational Grammar, PhD thesis, University of Massachusetts.

Corblin, F. (2013), *Cours de sémantique: Introduction*, Armand Colin.

Ebert, C. (2005), Formal Investigations of Underspecified Representations, PhD thesis, King's College, University of London.

Fox, C. & Lappin, S. (2010), 'Expressiveness and complexity in underspecified semantics', *Linguistic Analysis* **36**(1–4), 385–417.

Gómez-Rodríguez, C., Alonso, M. A. & Vilares, M. (2006), On the theoretical and practical complexity of TAG parsers, *in* 'Proceedings of Formal Grammar (FG 2006)', pp. 87–101.

Jacobson, P. (2002), 'The (dis)organization of the grammar: 25 years', *Linguistics and Philosophy* **25**(5-6), 601–626.

Joshi, A. (1985), Tree adjoining grammars: How much context-sensitivity is required to provide reasonable structural descriptions?, *in* 'Natural Language Parsing', Cambridge University Press, pp. 206–250.

Joshi, A. (1997), Parsing techniques, *in* R. A. Cole, J. Mariani, H. Uszkoreit, A. Zaenen & V. Zue, eds, 'Survey of the State of the Art in Human Language Technology', Cambridge University Press and Giardini, chapter 11.4, pp. 351–356.

Lafont, Y. (2004), 'Soft linear logic and polynomial time', *Theoretical Computer Science* **318**(1), 163–180.

Langacker, R. (2008), *Cognitive Grammar — A Basic Introduction.*, Oxford University Press.

Mineshima, K., Martınez-Gómez, P., Miyao, Y. & Bekki, D. (2015), Higher-order logical inference with compositional semantics, *in* 'Proceedings of EMNLP', pp. 2055–2061.

Montague, R. (1974), The proper treatment of quantification in ordinary English, *in* R. Thomason, ed., 'Formal Philosophy. Selected Papers of Richard Montague', Yale University Press.

Moortgat, M. (1997), Categorial type logics, *in* J. van Benthem & A. ter Meulen, eds, 'Handbook of Logic and Language', Elsevier/MIT Press, chapter 2, pp. 93–177.

Moot, R. (2002), Proof Nets for Linguistic Analysis, PhD thesis, Utrecht Institute of Linguistics OTS, Utrecht University.

Moot, R. (2007), Filtering axiom links for proof nets, *in* L. Kallmeyer, P. Monachesi, G. Penn & G. Satta, eds, 'Proccedings of Formal Grammar 2007'.

Moot, R. (2010), Wide-coverage French syntax and semantics using Grail, *in* 'Proceedings of Traitement Automatique des Langues Naturelles (TALN)', Montreal. System Demo.

Moot, R. (2015), 'Linear one: A theorem prover for first-order linear logic', `https://github.com/RichardMoot/LinearOne`.

Moot, R. & Piazza, M. (2001), 'Linguistic applications of first order multiplicative linear logic', *Journal of Logic, Language and Information* **10**(2), 211–232.

Moot, R. & Retoré, C. (2012), *The Logic of Categorial Grammars: A Deductive Account of Natural Language Syntax and Semantics*, Springer.

Morrill, G. & Valentín, O. (2015), Computational coverage of TLG: The Montague test, *in* 'Proceedings CSSP 2015 Le onzième Colloque de Syntaxe et Sémantique à Paris', pp. 63–68.

Morrill, G., Valentín, O. & Fadda, M. (2011), 'The displacement calculus', *Journal of Logic, Language and Information* **20**(1), 1–48.

Pentus, M. (1995), Lambek grammars are context free, *in* 'Proceedings of Logic in Computer Science', pp. 429–433.

Pentus, M. (2010), 'A polynomial-time algorithm for Lambek grammars of bounded order', *Linguistic Analysis* **36**(1–4), 441–471.

Pinker, S. (1994), *The Language Instinct*, Penguin Science.

Retoré, C. (2014), The Montagovian Generative Lexicon ΛTy_n: a Type Theoretical Framework for Natural Language Semantics, *in* 'Proceedings of TYPES', pp. 202–229.

Sarkar, A. (2000), Practical experiments in parsing using tree adjoining grammars, *in* 'Proceeding of TAG+ 5'.

Savateev, Y. (2009), Product-free Lambek calculus is NP-complete, *in* 'Symposium on Logical Foundations of Computer Science (LFCS) 2009', pp. 380–394.

Schwichtenberg, H. (1982), Complexity of normalization in the pure typed lambda-calculus, *in* 'The L. E. J. Brouwer Centenary Symposium', North-Holland, pp. 453–457.

Shieber, S. (1985), 'Evidence against the context-freeness of natural language', *Linguistics & Philosophy* **8**, 333–343.

Turing, A. (1950), 'Computing machinery and intelligence', *Mind* **49**, 433–460.