www.mind-consciousness-language.com, (2005)

A device in order to improve the quality of machine translation, based on the correlational theory of thought

Giulio Benedetti [1]

(Translation from Italian is my own; it has been kindly revised by Mr Nicholas White)

[1] I am grateful to Giorgio Marchetti for his assistance in formulating this article.

Abstract

In this article the author describes a device in order to improve the quality of machine translation, conceived by Silvio Ceccato and based on his correlational theory of thought. Basing themselves on this theory Ceccato and his collaborators worked on the problem of machine translation from 1959 to 1966, carrying out almost exclusively theoretical researches. The project was interrupted when MT research in USA stopped because of the well-known ALPAC report[1], before it was possible to evaluate the results this device could give.

The author thinks that this device and the theory it is based on, although very little known for various reasons, are still extremely valid today. Therefore they could permit, in the field of MT, better results than the ones we can achieve with the programs available today, especially when the source language is a language with very little morphology and a lot of ambiguities, such as English, and the target language is a language with rich or very rich morphology. This is essentially because: a) the correlational theory of thought is (also) *a linguistic theory which is completely new and very different from all others*; b) the program based on the aforesaid device *partially reproduces what the human being really does by using his intelligence* when translating. Besides, this MT system is probably the only one based completely and exclusively on a single linguistic theory and conceived by the person who proposed the linguistic theory itself. The only problem it presents, which is the fact that the human work that has to precede the program implementation increases exponentially when the number of the dictionary entries increases, is not theoretical but only practical.

Warning

This article cannot be understood without having previously acquired some knowledge of Operational Methodology. A sufficient knowledge can be quickly acquired reading my article "A presentation of Operational Methodology ". The present article has in fact been conceived as having the other article as its indispensable preamble. Reading another article of mine, "Basic mental operations which make up mental categories", is not strictly necessary but it can be useful in order to have a more complete idea of the subject.

This article has been conceived to be read also by persons lacking any knowledge of Machine Translation, but who are interested in knowing how Operational Methodology and the correlational theory of thought can have an application in this field.

Introduction

In this article I intend to describe a device designed to implement a machine translation (MT) program, which I mentioned in my article "A presentation of Operational Methodology". As I said in this article, the theories making up Operational Methodology (OM) began to be developed in the '50s and Ceccato's thought reached its full maturity in the '60s and the '70s. The project of Ceccato and his collaborators was carried out between 1959 and 1966, the year when the ALPAC report made the funding stop[1]. Therefore Ceccato's project dates back to the first phases of MT history. Therefore it may seem strange that in a modern article such an old project, even if with some modifications, is presented again. In my opinion there are at least four good reasons why this project should be presented again. Firstly, this project and the theories it is based on, even if so old, are very little known. Secondly, these theories (therefore also the project itself) have been really understood only by very few people. Thirdly, the project was interrupted after a few years, therefore it has not been possible to see what results it would have been able to give. Finally, progress in computer hardware and software technology over the last 40 years have been enormous and they could allow us to achieve results, which were not achievable at the time this project was conceived.
Among these four reasons the most important is surely the second. It is natural to ask why OM is such a little diffused and so rarely well understood theory. In an article of this kind it is not possible to speak in-depth about this problem, nevertheless it is necessary to point it out at least, widening at least a little what I have said in the brief article "A presentation of Operational Methodology". The main reasons for this fact, in my opinion, are at least four.

1) The first reason is that it can actually be difficult to understand OM theories, both because of their nature itself and because they are radically innovative. Nevertheless this is surely the less important among the four aforesaid reasons. In fact, using an adequate foundation of the exposition, I am sure that this difficulty can be overcome. The didactic experience I have personally gained up until now, even if it is not much, has always confirmed this conviction.

2) The second reason is just that in the exposition of these theories a foundation, which in my opinion is completely wrong, has been used. OM has been always presented as a radical criticism of traditional philosophical theories, but just because of this it has been viewed as a new philosophical theory, that is something that few people are interested in. The frequent references to philosophy have then required the use of an often difficult language, and furthermore exposition is not always clear.

[1] In 1966, the report of the government committee formed in the USA for the evaluation of the results and prospects of MT (Automatic Language Processing Advisory Committee, ALPAC), saw no need for further investment in MT research, bringing a virtual end to it in the USA for over ten years.

3) The third reason is that Ceccato's theories, in my opinion, contain serious errors, especially in their most general part (above all, the part concerning perception). Since this part is the one Ceccato emphasised more and the one he always exposes first, it is very probable that a lot of the people who approached these theories quickly stopped considering them before really understanding them in depth.

4) Finally, as far as I know, Ceccato often had attitudes and positions, which turned a lot of people against him, especially in the academic environment.

My work has been a careful, systematic and in-depth critical revision of the whole thought of Ceccato and *Scuola Operativa Italiana* (Italian Operational School, S.O.I.), trying to keep what is valid and to eliminate errors. This work has led to very deep modifications and a lot of developments, so that the *theories I propose are considerably different from the original ones*. Besides, I have used *completely different general lines of exposition*, which should allow a much easier understanding of them.

Operational Methodology and Machine Translation

OM was not born as a linguistic theory, even less as a linguistic theory conceived in order to tackle the problem of MT. OM was born as a theory concerning human mental activity (that is perception, representation, memory, thought, attitudes etc.) studied with a strictly introspective method. Thus it can be considered as a theory in the field of Cognitive Psychology. Nevertheless, since language is nothing else but the means by which we express our thought, OM is *also* a linguistic theory.

Because of this OM can be important in the field of MT. In fact, as Hutchins and Somers affirm, "the major obstacles to translating by computer are, as they have always been, not computational but linguistic"[2]. Nevertheless relations between MT research and linguistic theories have always been problematic. The same authors have drawn a clear picture of the situation:

"...'state-of-the-art' MT systems are not in general based on one single linguistic theory. [...] a number of approaches to the linguistic description of natural languages have been influential in the design of MT systems. However, there have been very few systems based on a single linguistic model, and these have been predominantly experimental projects undertaking basic research using MT as a test-bed for computational linguistic theories. The great majority of MT systems are amalgams of different approaches and models, or even occasionally (particularly in the early years of MT research [...]) with no discernible theoretical basis at all. Most commonly, systems are vaguely based on a general theory, such as transformational grammar or dependency theory, greatly modified by borrowings from other theories and by the events of computational implementation.

MT research has often been criticised for ignoring developments in linguistic theory. There would appear to be a wide communication gap between theoretical linguistics and practical MT research. Some observers believe that there are good reasons for this situation: until recently, linguistic theories had not provided adequate accounts of all aspects of language use [...]. In the past, and unfortunately it is still generally true today, much of a linguistic theory was based on phenomena observed in English, the language of the majority of the theoretical linguists. This neglect of other languages has been a further reason why linguistic theory has had less impact on MT than some observers might have expected.

[...] MT research is sometimes regarded as an 'engineering' task, a search for computational methods that work with the facts of language. The aims of many theoretical linguists are more abstract, concerned with investigations of human faculties in general, the nature of language itself, and the psychological foundations of language acquisition and use.

The results has been somewhat cynically described by Yorick Wilks (1989; 59):

"… the history of MT shows, to me at least, the truth of two (barely compatible) principles that could be put crudely as 'Virtually any theory, no matter how silly, can be the basis of some effective MT' and 'Successful MT systems rarely work with the theory they claim to."

The search of solutions that work, whatever their theoretical status and whether or not they fit the alleged principles of the project, has meant that MT systems inevitably present a confusing picture of disparate methodologies and that researchers have been obliged to take much more pragmatic attitudes to theoretical issues than their colleagues in computational linguistic and in linguistic theory."[3]

[2] Hutchins W. J. (with Somers H. L.), *An introduction to machine translation*, p. 2.
[3] Ibidem, p. 82.

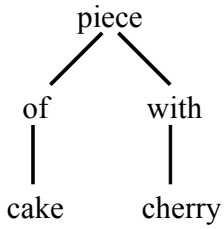
In a situation like this, it is clear that a linguistic theory radically new and different from the ones already existing could have an impact in the field of MT much greater than the one that other linguistic theories have had.

Differences between S.O.I.'s linguistic theories and the other linguistic theories. Implications for Machine Translation

S.O.I.'s theories are (*also*, as we saw) linguistic theories *deeply different* from the others. The similarities are often only superficial or apparent. There are three reasons for this deep diversity.

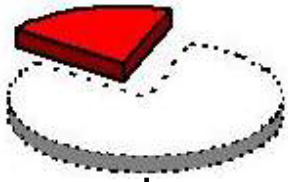
1) The first reason, still being the most important in an absolute sense, is the one that, at least in the short or medium period, has less importance for MT. In the far future, instead, it could have a decisive importance. This reason is that OM, as I tried to stress in the brief aforesaid article, has given some answers to the fundamental problem of the meaning of words, particularly the words grammar is concerned with (conjunctions, prepositions, pronouns, fundamental verbs such as "to be", "to have" etc., main adverbs etc.). It is possible that the S.O.I.'s theories concerning this still have to be partially corrected, modified and developed (this is what I have tried and I am trying to do), but substantially they are, in my opinion, hugely more satisfying than the ones any other kind of approach has produced. OM asserted that the meanings of these words are made up by groups of mental operations where the ones of attention have a leading role and it has provided analysis of these words in terms of these operations. Since these operations, at least theoretically, seem reproducible, OM, in my opinion, perhaps opened the way to the construction of real Artificial Intelligence. I know very well that, even if S.O.I. views and my opinion are correct, this is all the same an extremely difficult task and there could also be insuperable obstacles (for example, certain functions supported by nervous structures, that is biological structures, could not be reproducible by means of electronic components). But if it is not like this, and in the far future we shall be somewhat successful in the construction of a real Artificial Intelligence, that is a machine able to understand a text, this would surely be extremely important in order to produce a high quality MT. However, this is an aim which we are not sure of reaching and anyway is very far away, while in this article I am interested in proposing what can be done at present.

2) The second reason is the one of fundamental importance for MT. The correlational theory of thought, independently from the fact that it is just a theory concerning the human activity of thought and not only a linguistic theory, is deeply different from all other linguistic theories because it starts from a presupposition which is the opposite of the one they start from. Other linguistic theories and traditional grammar start from the presupposition that in sentences words are referred to each other in a direct way, that is a word is referred or tied to another in such a way that the ties among the words can be represented by lines. The sentence structure is therefore represented by the well-known "trees" or with other kinds of representation that are anyway substantially equivalent. The presupposition OM starts from is completely opposite. According to this theory, words belong to two quite distinct categories, the one of correlators (some dozen of words), that is the words having the function of *tying* other words, and the one of correlata (all other words), that is the words that *must be tied* by something else. That is, two correlata are *never* tied in a direct way, but there is always a third element that ties them. This element can be either explicit, that is it can be another word or a marker of the two correlated words, or it can be implicit. In the latter case we can indicate what the words are that it ties by putting them one after the other, but this is not always possible, and in this case this indication is lacking completely. . This is exactly, as we shall see, one of the biggest obstacles for MT program implementation.
I can show this difference with a very simple example. Let's take the expression "piece of cake with cherry". According to traditional grammar and linguistic theories different from correlational theory of thought (from now on, for the sake of simplicity, we shall call them "conventional linguistics") the structure corresponding to this expression would be the following:



That is, a substantive ("piece") governs two prepositional phrases, "of cake" and "with cherry". We have to notice that this structure is fundamentally hierarchical. This is reflected in the terminology typical of grammar and of conventional linguistics ("to depend on", "to govern", "to be governed", "governator", "head", "mother", "daughter", "sister" etc.). That is there are words that would be more "important" (generally the noun which is the subject and the verb in the personal form are considered the most important words, because the sentences we produce almost always contain at least these two elements) to which the other words of the sentence would be directly tied with a subordinated relationship. But, at the level of thought, is it really like this? Let's try to understand what happens in our mind, that is what we represent, when we hear this expression. I should not say we represent the abstract notion of "piece", to which we subordinate the prepositional phrase "of cake" and then we do the same thing with the other prepositional phrase "with cherry". I should say we tie, by means of a mental category indicated by the preposition "of", another mental category, the one indicated by the word "piece", to the representation of a physical object, that is a "cake". That is we imagine a cake, then we restrict our attention to a part of it (generally having a triangular form, as the parts we cut cakes into are), discarding the rest but mentally keeping present that the part we selected by attention comes from the whole cake (category of "piece" applied to "cake"). As every other correlator, also the one designated by the preposition "of" indicates that the two elements which it ties are *both present in our mind, together with the correlator itself*, and they are on the same level (the only thing that can be considered a "hierarchy" is the fact that "piece" precedes "cake" in time, therefore we have to refer "cake" to "piece" and not vice versa). If we wish to give a graphic representation, what we have in our mind is not this:

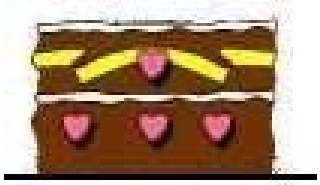
First hierarchical level:



(abstract notion of "piece")

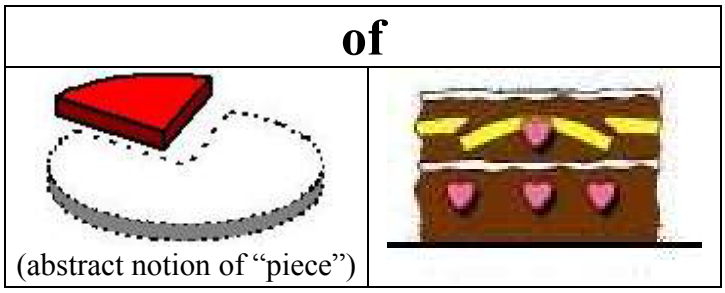
Second hierarchical level: preposition "of"

Third hierarchical level:

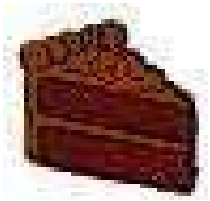


(representation of "cake")

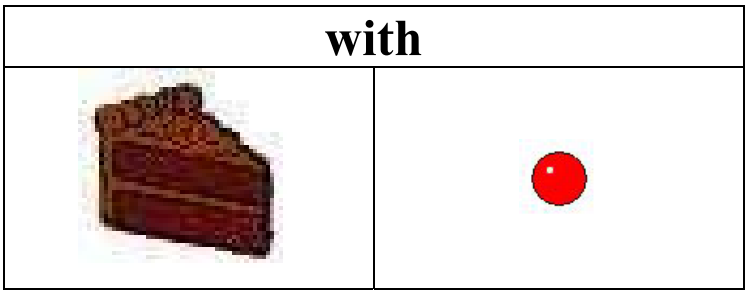
but this:



Its meaning is this:



Then we correlate *the whole correlation* "piece of cake" (*not only the notion of "piece"*) with the representation of the "cherry". The correlator "with" indicates that two distinct objects are in such a relationship that attention is induced to focus on them together[4]. Also in this case both the two correlata and the correlator are present together in our mind and they make up a single whole, that is this:

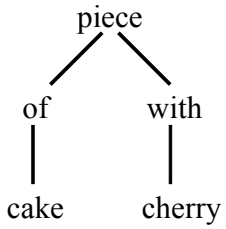


which means this:

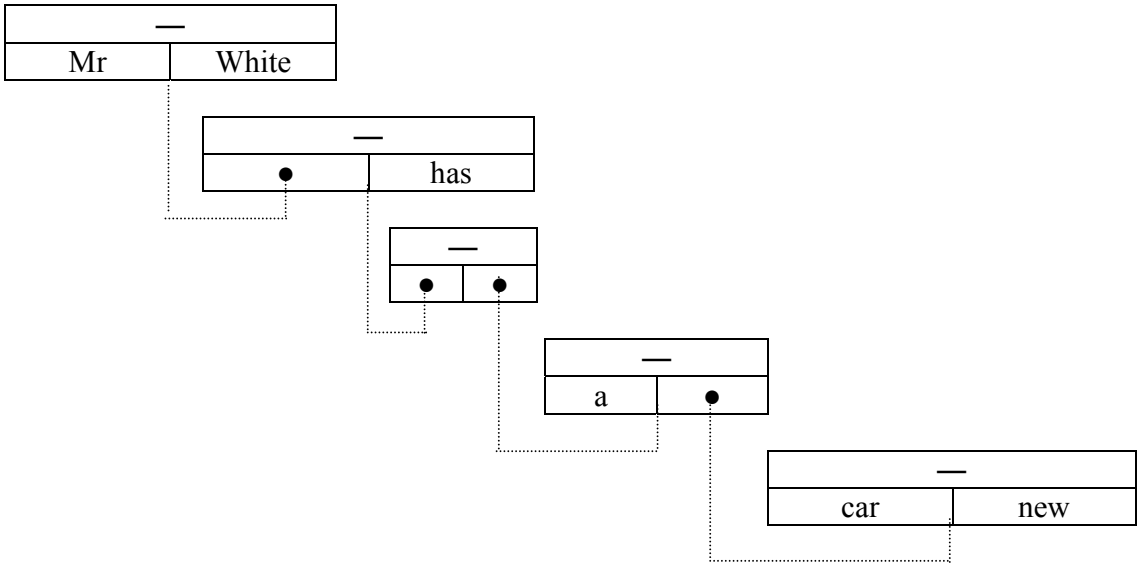
Such relationships cannot be graphically represented by means of dependency trees. In fact, the tree

[4] This analysis is my own modification of the original Ceccato's analysis (which instead is "two things are focused together by attention, then they are divided by it").

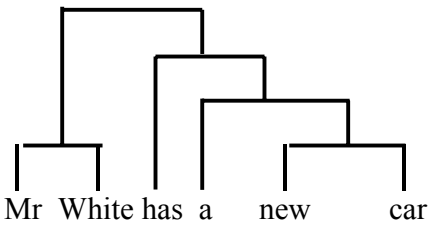
that we have above drawn



does not (and cannot) indicate the fact that, as we have just seen, *the whole correlation* "piece of cake" (*not only the notion of "piece"*) is correlated with the representation of the "cherry".
The most suitable representation is the one we saw, that is graphic elements made up by an upper box, the one of the correlator, which, so to say, "embraces" the two lower boxes, the ones of the correlata.
The correlational triad made up like this can be the correlatum of another correlational triad. This is indicated by a dotted line that starts from the middle of the graphic element representing the first correlational triad (so as to symbolise that the first triad has to be kept as a whole) and stops with a dot in the middle of one of the two correlata boxes of the second triad. Therefore the graphic representation of the structure of a sentence looks like a network, a tissue, and not the ramifications of a tree.
In the cases where in the sentence there are no prepositions or conjunctions, we can have the false impression that the correlational network is the same thing as a dependency tree. For example, the correlational network corresponding to the sentence "Mr White has a new car", that is this:

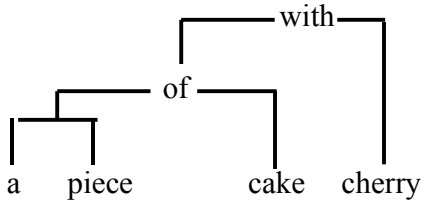


can seem the same thing as the corresponding dependency tree:



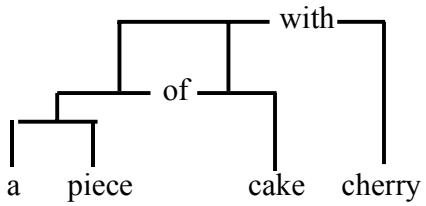
But this is only due to the fact that in this sentence there are no prepositions or conjunctions so we have not had to put *some words* of the sentence in such a way as to show that they tie other words.
If we try to represent the correlational network of thought by a structure in some way similar to a tree, the result is not good because this kind of representation is quite unnatural for the correlational network of thought. In fact, for all the sentences containing prepositions or conjunctions (as, for example, “a piece of cake with cherry”), we should build structures either as this:

[1]

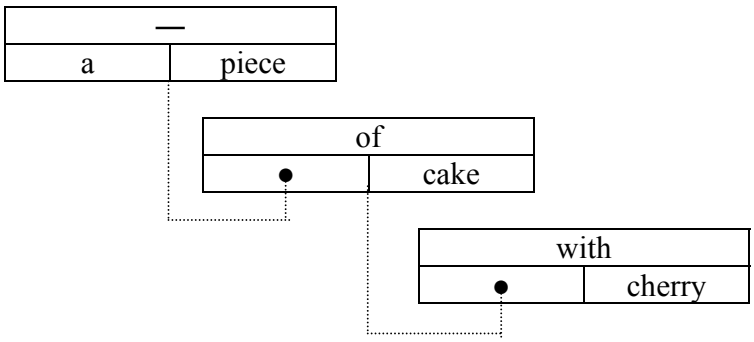


which is completely unsatisfactory because the preposition "of" seems to be tied with the preposition "with"; or as this:

[2]



which is unsatisfactory too because in it the fact that the correlation "a-piece" (having as correlator the implicit correlator) is the first correlatum of the correlation which has as correlator the preposition "of" and as second correlatum "cake" and that *all this structure as a whole* is the first correlatum of another correlation having as correlatum the preposition "with" and as a second correlatum "cherry", which is very clear in this kind of representation:



is not clear at all.
Apart from the fact that tree structures such as [1] and [2] are unnatural and unsatisfactory as representations of correlational network of thought, they are anyway deeply different, as we can see very well, from the traditional dependency trees, because some words of the sentence are *inside* the branches of the tree, and not *attached* to them.
If we want to make a comparison, a rough one but one which can be useful to give the idea in a visual way, we can compare the correlational network with a structure like a metal frame or latticework,

where there are parts (bars, beams etc.) which are tied together by other parts having the specific function of tying together (bolts, nuts etc.), while sentence structure, according to traditional linguistics, is like a tree, with its trunk, its branches and its leaves attached to them.
This comparison is useful also to highlight a very important fact which has to be underlined, that is that, according to the correlational theory of thought, unlike conventional linguistics, sentence structure *is not hierarchical*, that is there are not some words which are more "important" than the others, governing the latter (as instead happens in a tree, where the trunk is more important than the branches and the branches are more important than the leaves), but all elements are equally essential (as in a frame or latticework where the bolts are as essential as the bars they tie). The fact that the noun that is the subject and the verb in the personal form are considered the fundamental elements of the sentence derives, as we pointed out, simply from the fact that most sentences we produce contain these two elements at least. This happens because a sentence generally is used to talk about something that can exist independently (that is something indicated by a noun[5]) and what happens to that thing in time (which is designated by a verb in the personal form because it is the second correlatum of a noun, which therefore becomes its subject). But it is not always like this. Sometimes, especially in spoken language, we produce sentences lacking the subject-personal verb structure, like for example "Happy birthday and many happy returns of the day" or "What a kind person!"[6]. Therefore, according to the correlational theory of thought, some words are more "important" than others only from a practical point of view, while at the level of thought structure, that is at the level of the "deep structure"[7] corresponding to a sentence there are no hierarchies but only couples of correlata tied by an explicit or implicit correlator, so making up triads, in their turn tied by correlators. Surely it is true that in the sentence certain words are subordinated to certain others (for example in a sentence like "John has blue eyes" the verb "has" depends on the subject "John" and the adjective "blue" depends on the substantive "eyes", and both depend on the verb). But, according to OM, this verb "to depend" only means, in the example just given, that we first focus our attention on "John" and, keeping him present, we make up the mental category corresponding to the verb "has", then we correlate the triad "John-implicit correlator-has" with the triad "eyes-implicit correlator-blue"[8] previously made up. It is a dependency that derives from the fact that we *first* focus something by attention and *then*, keeping it present, we focus attention on something else, so making the second thing referred to the first. But the kind of relationship (the correlational triad) that ties the words of this sentence and all sentences is always the same and there are not more "important" words. The only thing that can be considered a "hierarchy" is the fact that the first correlatum precedes the second correlatum in time, therefore it is the latter which is referred to the former and not vice versa.

3) OM, as regards to the parts of speech, uses the same terminology as traditional grammar, that is it uses the terms "article", "noun", "adjective", "verb", "adverb" etc. Therefore, with regard to this, it seems to have introduced no novelty in comparison with traditional grammar. This is due to the fact that OM's linguistics theory is in no way opposed to traditional grammar. On the contrary, it recognises

[5] The adjective, instead, designates a *feature* of something, that is a thing which can not exist independently. For example, what is designed by the noun "pen" can exist independently, while what is designed by the adjective "red" can not exist independently from something being red. The only exception are the so-called abstract nouns ("lightness", "happiness", "beauty" etc.), where we consider isolatedly a feature of a thing in order to speak specifically about it . Because of this in the correlation substantive-adjective the substantive is always the first correlatum, even if in some languages, such as English, the adjective usually precedes the noun which it is referred to.

[6] We could think -as some grammar books assert- that in such cases a verb is understood. But we have to notice that in such cases the verbs which can be understood are generally more than one and also very different. For example, in the expression "What a kind person!" the verbs that could be understood are "is", "I met", "I knew" and more again. It is much more simple to think that in such cases, at the level of thought, there is no verb.

[7] The expression "deep structure" is here naturally used in the most general sense it has in linguistics, that is without any reference to Chomskj's generative grammar.

[8] See note 5.

that the distinctions and classifications of traditional grammar are perfectly valid and very useful. But between traditional grammar and the linguistic theory of OM there are two differences. The first difference, less important, is that prepositions and conjunctions are classified together in the category of the "correlators", while the category of "pronouns" is inserted in a larger category having another name and where there are also other elements (the reason for this cannot be explained in a relatively brief work like this). The second difference, much more important, is that OM does not limit itself to accepting the classifications of speech parts made by traditional grammar, but *defines* the parts of speech in a satisfying way, in my opinion, while traditional grammar and other linguistic theories were not able to do this (in the aforesaid article I quoted the definition of "noun", while in this article we shall see the one of "personal verb"). Linguists themselves are perfectly aware of this fact, as this passage, drawn from a recent linguistics treatise, clearly shows:

"Le parti del discorso possono essere perciò riconosciute in base a criteri puramente **distribuzionali**: i nomi, i verbi, ecc.. saranno definiti in base alle altre classi di parole assieme alle quali possono, oppure non possono, ricorrere. La definizione precisa delle varie parti del discorso in termini distribuzionali è un'operazione complessa, ma abbiamo visto che le definizioni tradizionali di tipo semantico sono decisamente inadeguate, e un'impostazione distribuzionale è la via più promettente per superare queste difficoltà."[9] [sottolineature mie]
["So the parts of speech can be recognised basing ourselves on purely distributional criteria: nouns, verbs etc. will be defined basing ourselves on the other classes of words they can, or they cannot, recur with. The precise definition of the parts of speech in distributional terms is a complex operation, but we have seen that the traditional definitions of a semantic kind are clearly inadequate, and a distributional foundation is the most promising way to overcome such difficulties."] (underlining is my own)

As regards the problem of MT, the fundamental difference between the linguistic theory of OM and other linguistic theories is the one exposed in 2). The overall meaning of a sentence is given by two things:

1) the meanings of the single words which made up it;
2) the way words are tied to each other, that is the structure of the corresponding thought.

The human being who understands the overall meaning of a sentence (action which always precedes translation) understands the meaning of the single words and reconstructs the corresponding structure of thought. Today building a machine which understands the meaning of words is absolutely impossible, but it is possible to try to implement a program which, starting from a sentence, tries to reconstruct the corresponding structure of thought. *The fact that there are words that can be only correlators and others that can be only correlata is a powerful tool in order to identify what is the correct structure of the correlational network corresponding to a sentence*. This happens because of two reasons:

1) the words designating correlators can be put only in the upper boxes of the triad, while the ones designating correlata can be put only in the lower boxes;
2) some correlators, as we shall see, accept as correlata only some kinds of correlata.

Also some actual MT programs try in some way to identify the structure of the sentences, that is how the aforesaid "trees" are done, but they do not have the tool we just described. To do this they have only:

1) the grammar rules of a language (for example the ones which state that there must be an agreement of genre and number between noun and adjective, of person between personal form of the verb and its subject etc., rules that also the program we shall describe uses);
2) a "notional sphere" which tries to reproduce the knowledge of the real world which human beings have, "notional sphere" which, in the case of our program, has huge dimensions, as we shall see.

[9] Giorgio Graffi, Sergio Scalise, *Le lingue e il linguaggio*, 2003, p. 118.

The main problems of Machine Translation

As everybody knows, in order to translate a text, it is necessary to substitute the words of the source language, arranged in a certain order, with suitable words of the target language, also arranged in a certain order, which is often different from the other language. If only one word of the target language always corresponded to each word of the source language and the word order were the same, MT would not be a problem. Unfortunately, it is not like this at all. A certain amount (which varies according to the two languages considered) of the words of the source language must be replaced by words of the target language that are different according to the cases. This is not a problem for the human being who translates. He, starting from the text of source language, *understands* the original text, that is he rebuilds *one and only one* meaning for it; then, he expresses this univocal meaning using the words and the rules of the target language. Naturally, the machine cannot do anything like this. Many problems derive from this fact. The main ones are the following two.

1) The first problem is that there are some words, which the dictionary lists as entries, which have more than one meaning. There can be two or more meanings not related to each other (for example "bank", which can indicate a financial institution or the side of a river or channel); or, much more frequently, one derived from the other (for example "head", which can indicate a part of our body or a person who gives orders). Furthermore, what may happen is that the flexion of noun, verb etc. leads to the formation of words which are equal to others (for example "bears", which can be the plural of a name of an animal or a part of the verb "to bear"). These cases are not a problem for a human being because he, basing himself on the context, can understand which is the meaning at stake from time to time. If, for example, we are talking about carrying a suitcase and in the text we have to translate there is the English word "light" the human being who translates understands at once that the word does not designate "not dark", but "not heavy".

A problem strictly related to this is that in languages with very little morphology, such as English for example, the same word can be a noun, a verb, an adjective etc. and the role the word plays from time to time can be established only by basing oneself on the general structure of the sentence and its meaning. In the case of a target language that has a rich morphology this is a very important problem, because, as a rule, these languages have different words for the nominal form, the verbal form etc.

2) The second big problem is that, more or less frequently according to the language, *not all* the relationships making up the correlational network of thought are indicated in the speech that is the expression of that thought. Let's consider the following examples:

- *some amount of water, even small*
- *some amount of water, even cold*
- *Scotch whisky bottle*
- *empty whisky bottle*

In the first example the end adjective "small" is referred to "amount", while in the second example the end adjective "cold" is instead referred to "water". In the third example the adjective "scotch" is referred to "whisky", while in the forth example the adjective "empty" is referred to "bottle". But, in the English expression of these thoughts, there is nothing which indicates it is so. In other words, in these cases the grammatical information in the speech (that is morphology and word order) does not allow us to rebuild one single correlational network of thought. The human being understands the meanings of the words "amount", "water", "small", "cold", "scotch", "whisky", "bottle", "empty" and, using the basic culture we all have (in the texts about MT, this basic culture is often called "real world knowledge"), understands at once which is the right correlational network. Also in this case, the problem in MT depends on what the target language is. When, for example, we translate the first two sentences from English into another language lacking cases too, such as Italian for example, there is no problem because in this case the simple word for word translation is satisfying. But it is not so when we trans-

late into a language owning cases, such as German or Russian, for example. In this case only understanding the English sentence allows us to translate the two end adjectives “small” and “cold” in the correct way, that is flexing the former to the nominative case and the latter to the genitive case. In cases like these actual programs translate either substantially at random, therefore with a high probability of error or, if they use some device, this proves insufficient, as we shall see further on.
We could think that such cases are not very frequent. It is not like this at all. The frequency of this phenomenon varies across languages, but it is surely present in a lot of languages and probably in all languages. In languages with a rich morphology, that is languages owning the flexion of noun and even of adjective according to various cases (nominative, accusative, genitive, dative etc.) and the flexion of verb according to the person it is referred to, such ambiguities are less frequent than in languages having very little morphology, as English. But even in the former the phenomenon is present. However, English is a hugely widespread language as a mother-language and even more so as a second language, thus it is surely strongly desirable to have programs translating from and into this language in a satisfying way.
Naturally, as we can understand very well, this phenomenon causes particularly serious problems when we need to translate from a language with very little morphology into a language with rich or very rich morphology, such as many Indo-European languages (for example Spanish, Italian, French, German, Russian etc.). If we think of how often such a translation is required, we understand very well that the importance this problem assumes is huge.
Therefore, let’s see some examples of what happens in the translation from a language having very little morphology into a language with a rich morphology. In English, for example, the verb is flexed very little according to the person that it is referred to: practically, only the third person of the present indicative has a termination different from the other persons. Moreover also the infinitive mood is the same as the indicative mood, and the personal forms of the past of regular verbs are the same as the past participle. Moreover, the verb is often the same as the substantive. For example, the English word “love” will have to be translated, according to the context, with one of the following Italian words: “amore”; “amo”, “ami”, “ama”, “amiamo”, “amate”, “amano”; “amiate”, “amino”. The English word “loved” will have to be translated with one of the following Italian words (or couples of words): “amai”, “amasti”, “amò”, “amammo”, “amaste”, “amarono”; “amavo”, “amavi”, “amava”, “amavamo”, “amavate”, “amavano”; “avevo amato”, “avevi amato”, “aveva amato”, “avevamo amato”, “avevate amato”, “avevano amato”; “ebbi amato”, “avesti amato”, “ebbe amato”, “avemmo amato”, “aveste amato”, “ebbero amato”; “amassi”, “amasse”, “amassimo”, “amaste”, “amassero”; “abbia amato”, “abbiamo amato”, “abbiate amato”, “abbiano amato”; “avessi amato”, “avesse amato”, “avessimo amato”, “aveste amato”, “avessero amato”; “amato”, “amata”, “amati”, “amate”. Furthermore, English has no genders, while many others languages have the masculine and the feminine gender and some languages the neuter gender too. Also in this case, a single English word, for example an adjective, must be translated in two or three different ways according to the substantive it is referred to (plus another two or three ways, if we think that, as a rule, in these languages there is a flexion according to the gender also in the plural).
Naturally, also in languages with very little morphology, not always do we need to resort to the understanding of the meanings of the words or to real world knowledge in order to rebuild the right correlational networks. This happens, on one hand, because also these languages have some morphology (for example, in English, the addition of the ending “-s” to the stem of the verb when it is referred to a third person singular and conjugated in the present tense of the indicative mood), but above all because these languages have a very rigid word order and also because some relationships are understandable taking into account the general structure of the sentence. For example, in the English sentence “John hit Bob” the fact that the subject is “ John” and the direct object is “Bob” is unequivocally indicated by the word order; and the fact that “hit” is a verb and not a noun (this could be, from the morphological point of

view) can be gathered by the almost general rule that sentences have a verb in a personal form (therefore we can make a program take into account this and use it to make the right choices).
As we can see very well, a program able to reconstruct the correct correlational network of thought starting from text would guarantee the complete absence of errors in the translation of the very frequent unflexed words which are present in languages with little morphology, words which must be converted into flexed words of languages with rich morphology.
Naturally, since the beginning of MT history it has been well understood that, in order to operate a correct machine translation, it is necessary to rebuild the sentence structure basing us on some linguistic theory. The existing programs are already provided with devices partly allowing them to do this, but the results they give are not, as we shall see, completely satisfying. Instead, the program I shall describe, *always* tries to operate an *integral* reconstruction of the sentence structure according to a *single* and *strict* linguistic theory. The program makes this basing itself on a complex device where the simulation of the real world knowledge plays a fundamental role. Therefore the program closely simulates what the human being does when translating.
Because of this, I think this device still preserves today all its theoretical and practical value and it could allow us to reach, in the field of machine translation, results decidedly better with respect to the ones we can reach using programs available today which are not based on this device and on correlational theory of thought. The fact that, even for very small dictionaries, a great amount of human work of linguistic kind is necessary before the program implementation, and that this work increases enormously when the number of the entries of the program dictionary increases, is only a practical problem and not a theoretical one. If we have at our disposal enough resources this problem could be overcome. The use of a large amount of resources would be largely repaid by the big commercial spin-off, which a program able to do a substantially better machine translation than the one the actual programs can do, surely would have. In regard to this it is important to notice that the quality of machine translation does not seem to have substantially improved during, say, the last ten years, neither do the approaches used up until today seem to promise important improvements for the future. In regards to this I quote two very meaningful passages from an authority on the matter, W. J. Hutchins.

The 'real' question is whether the *quality* of machine translation has improved. Some will say that it *must* have. Others of an older generation insist that MT has the same problems and throws up the same errors as it did in the 1960s and 1970s. Many, however, will perhaps share my intuition that while progress in quality is not evident during say the last ten years, there are discernible advances since the early 1980s and definite improvements since the ALPAC report [see note 1, editor's note] of the mid 1960s.[10]

Despite the prospects for the future, it has to be said that the new approaches of the current decade have not yet resulted in substantial improvements in the quality of the raw output from translation systems. These improvements may come in the future, but overall it has to be admitted that at present the actual translations produced do not represent major advances on those made by the MT systems of the 1970s. We still see the same errors: wrong pronouns, wrong prepositions, garbled syntax, incorrect choice of terms, plurals instead of singulars, wrong tenses, etc., etc. – errors that no human translators would ever commit.
[…]
Unfortunately, this situation will probably not change in the near future. There is little sign that basic general purpose MT engines are going to show significant advances in translation quality for many years to come.[11]

[10] Hutchins W. J., '*Has machine translation improved*?'. MT Summit IX: proceedings of the Ninth Machine Translation Summit, New Orleans, USA, September 23-27, 2003, 181-188. [East Stroudsburg, PA: AMTA.]
[11] Hutchins W. J., '*Retrospect and prospect in computer-based translation*'. In Machine Translation Summit VII, 13th-17th September 1999, Kent Ridge Digital Labs, Singapore. Proceedings of MT Summit VII "MT in the great translation era", 30-34. [Tokyo]: Asia-Pacific Association for Machine Translation, 1999.

Now let's see some practical examples of the errors actual machine translation programs make. The first example consists of the English and French versions of an extremely simple text. They are examples of tourist information found on the Internet site of a very famous museum. Naturally the editor did all is possible for the text to be extremely simple and clear. In fact, nearly all the words are commonly used words and the sentences, which are all very brief, are nearly all made up by only one clause.
Instead of quoting the whole passage and then the whole translation, I put the translation under each paragraph of the original text, so that the reader can more easily compare the original text with the translated one. The translation done by the program (one of the most famous of those available today[12]) is written in red type. The translation done by the human being is instead written in blue type. The mistakes in machine translation are highlighted in yellow (the meaning of the underlining will be further explained), while the inaccuracies are highlighted in grey. The correct translation of the parts that instead are wrong in the machine translation has been highlighted by the same two colours, so that the reader can easily identify it.
It is better if the reader reads machine translation text first (that is all the text in red types), so as to have an idea of the whole quality of the machine translation. Then he will be able to examine the translation mistakes one by one.

FRENCH ORIGINAL TEXT

Aide à la visite

Plan/Information

Un plan/information gratuit, en 9 langues (français, anglais, espagnol, allemand, italien, japonais, chinois, néerlandais, arabe), est disponible à la banque d'information. Il permet au visiteur de s'orienter dans le musée et les collections.

Aiuto alla visita

Piano/informazione

Un piano/informazione gratuito, in 9 lingue inglese (francesi, spagnola, tedesco, italiano, giapponesi, cinesi, olandesi, arabo), è disponibile alla banca d'informazione. Permette all'ospite di orientarsi nel museo e le raccolte.

Aiuto alla visita

Piano/informazione

Un piano/informazione gratuito, in 9 lingue inglese (francese, spagnolo, tedesco, italiano, giapponese, cinese, olandese, arabo), è disponibile al banco informazioni. Permette all'ospite di orientarsi nel museo e nelle raccolte.

Audioguides

Un audioguide (lecteur individuel de CD ROM) est à la disposition des visiteurs du musée; il comprend des commentaires en français, anglais et espagnol sur plus de 1000 oeuvres et lieux majeurs du musée, sans parcours préétabli. Le commentaire est disponible aussi en allemand, italien et japonais pour 350 oeuvres.

L'audioguide est proposé aux 3 entrées du musée (Sully, Denon et Richelieu, niveau entresol). Son prix de location est de 5 euros, en échange d'une pièce d'identité.

Règlement uniquement en espèces, chèque bancaire ou chèque de voyage.

Audioguides

Audioguide (lettore individuale di CD ROM) è a disposizione degli ospiti del museo; comprende commenti in francese, inglese e spagnolo su più di 1000 opere e luoghi principali del museo, senza percorso prestabilito. Il commento è disponibile così in tedesco, italiano e giapponese per 350 opere.

Audioguide è proposto alle 3 entrate del museo (Sully, Denon e Richelieu, livello entresol). Il suo prezzo d'affitto è di 5 euro, in cambio di un documento d'identità.

Regolamento soltanto in specie, assegno bancario o assegno di viaggio.

Audioguide

L'Audioguida (lettore individuale di CD ROM) è a disposizione degli ospiti del museo; comprende commenti in francese, inglese e spagnolo su più di 1000 opere e luoghi principali del museo, senza percorso prestabilito. Il commento è disponibile anche in tedesco, italiano e giapponese per 350 opere.

Audioguide è proposto alle 3 entrate del museo (Sully, Denon e Richelieu, al mezzanino). Il suo prezzo d'affitto è di 5 euro, in cambio di un documento d'identità.

---

[12] The results produced by only one program are quoted because the results produced by the best programs today available do not differ in a substantial way.

Pagamento soltanto in contanti, assegno bancario o traveller's cheque.

Les feuillets

Dans les salles du musée se trouvent des "feuillets", textes édités en plusieurs langues (français, anglais, allemand, espagnol, italien, japonais), sous forme de plaques plastifiées mobiles. Il s'agit de textes explicatifs des collections, portant sur une oeuvre en particulier, un artiste, un style, un site, une technique...

La version française en format réduit est en vente à la librairie, hall Napoléon.

Gli strati

Nelle sale del museo si trovano "strati", testi pubblicati in molte lingue (francesi, inglesi, tedesche, spagnolo, italiano, giapponesi), sotto forma di targhe plastificate mobili. Si tratta di testi esplicativi delle raccolte, riguardanti un'opera in particolare, un artista, uno stile, una località, una tecnica...

La versione francese in formato ridotto è in vendita alla libreria, hall Napoleone.

I foglietti

Nelle sale del museo si trovano dei "foglietti", testi pubblicati in molte lingue (francese, inglese, tedesco, spagnolo, italiano, giapponese), sotto forma di targhe plastificate mobili. Si tratta di testi esplicativi delle raccolte, riguardanti un'opera in particolare, un artista, uno stile, una località, una tecnica...

La versione francese in formato ridotto è in vendita alla libreria, hall Napoleone.

Public handicapé

Un guide d'orientation spécifique (en français et en anglais) pour les visiteurs à mobilité réduite indiquant les itinéraires avec ascenseurs est disponible à la banque d'information, ainsi que des fauteuils roulants.

Pour les visiteurs aveugles et malvoyants, le département des sculptures a ouvert un espace tactile regroupant une vingtaine d'œuvres que l'on peut découvrir par le toucher et à l'aide d'un audioguide.

Des visites-conférences pour les groupes d'handicapés peuvent être organisées sur demande.

Tel.: (33) 01 40 20 59 90

Pubblico minorato

Una guida d'orientamento specifico (in francese ed in inglese) per gli ospiti con mobilità ridotta che indicano gli itinerari con ascensori è disponibile alla banca d'informazione, così soltanto delle poltrone mobili.

Per gli ospiti ciechi e portatori di handicap visivo, il dipartimento delle sculture ha aperto uno spazio tattile che raccoglie una ventina di opere che si possono scoprire con il contatto ed all'aiuto di audioguide.

Visite-conferenza per i gruppi di minorati possono essere organizzate su domanda.

Tel.: (33) 01.40.20.59.90

Pubblico minorato

Una guida d'orientamento specifico (in francese ed in inglese) per gli ospiti con mobilità ridotta che indica gli itinerari con ascensori è disponibile al banco informazioni, così come delle poltrone mobili.

Per gli ospiti ciechi e portatori di handicap visivo, il dipartimento delle sculture ha aperto uno spazio tattile che raccoglie una ventina di opere che si possono scoprire con il tatto e con l'aiuto di audioguide.

Visite-conferenza per i gruppi di minorati possono essere organizzate su domanda.

Tel.: (33) 01.40.20.59.90

Règlement de visite

Pour le confort des visiteurs, l'usage des flashes est vivement déconseillé.

Il est interdit de fumer dans le Hall Napoléon et dans les salles du musée.

Ne sont pas acceptés au vestiaire, ni à la bagagerie, ni dans les collections: les aliments, les boissons, les animaux, les gros sacs et les valises.

Le livret résumant le règlement de visite est à disposition des visiteurs à la banque d'information sous la pyramide.

Regolamento di visita

Per la comodità degli ospiti, l'impiego degli flashes è vivamente sconsigliato.

È vietato fumare nell'hall Napoleone e nelle sale del museo.

Non sono accettati al guardaroba, né alla bagagerie, né nelle raccolte: i prodotti alimentari, le bevande, gli animali, le grandi borse e le borse.

L'opuscolo che riassume […] regolamento di visita è a disposizione degli ospiti alla banca d'informazione sotto la piramide.

Regolamento di visita

Per la comodità degli ospiti, l'impiego dei flashes è vivamente sconsigliato.

È vietato fumare nell'hall Napoleone e nelle sale del museo.

Non sono accettati al guardaroba, né al deposito bagagli, né nelle collezioni: i prodotti alimentari, le bevande, gli animali, le grandi borse e le valigie.

L'opuscolo che riassume [il] regolamento di visita è a disposizione degli ospiti alla banca d'informazione sotto la piramide.

Pourquoi il ne faut pas toucher les oeuvres

Les oeuvres d'art sont uniques et fragiles.

Elles ont traversé les siècles et doivent être conservées pour les générations futures.
Toucher, même très légèrement, une peinture, un objet, une sculpture, un meuble l'abîme.
Surtout lorsque ce geste est répété des milliers de fois.
Aidez-nous à protéger notre patrimoine commun.

Perché non occorre toccare le opere
Le opere d'arte sono uniche e fragili.
Hanno attraversato i secoli e devono essere conservate per le generazioni future.
Toccare, anche molto leggermente, una vernice, un oggetto, una scultura, un mobile lo danneggia.
Soprattutto quando questo gesto è ripetuto migliaia di volta.
Aiutiamo a proteggere il nostro patrimonio comune.

Perché non bisogna toccare le opere
Le opere d'arte sono uniche e fragili.
Hanno attraversato i secoli e devono essere conservate per le generazioni future.
Toccare, anche molto leggermente, un dipinto, un oggetto, una scultura, un mobile lo danneggia.
Soprattutto quando questo gesto è ripetuto migliaia di volte.
Aiutateci a proteggere il nostro patrimonio comune.

ENGLISH ORIGINAL TEXT

Visit Information

Handbook
A free handbook in 9 languages (French, English, Spanish, German, Italian, Japanese, Chinese, Dutch, Arabic) to help visitors find their way in the museum and the collections, is available from the Information Desk.

Le Informazioni Di Chiamata

Manuale
Un manuale libero in 9 lingue (francese, inglese, spagnolo, tedesco, italiano, giapponese, cinese, olandese, arabo) per aiutare gli ospiti a trovare il loro senso nel museo e nelle collezioni, è disponibile dallo scrittorio delle informazioni.

Informazioni di visita

Manuale
Un manuale gratuito in 9 lingue (francese, inglese, spagnolo, tedesco, italiano, giapponese, cinese, olandese, arabo) per aiutare gli ospiti a trovare la loro via nel museo e nelle collezioni, è disponibile al banco delle informazioni.

Audioguides
An audioguide is provided for visitors to the museum. It includes commentaries in English, French and Spanish on 1000 works and major locations within the museum, without the need to follow a specific route. Available as well in German, Italian and Japanese on 350 works. It can be obtained from the entrances to the three wings of the museum (Sully, Denon and Richelieu), on presentation of an identity card. Fee: 5 euros.
Payment only in cash, bank cheque or traveller's cheque.

Audioguides
Un audioguide è fornito per gli ospiti al museo. Include i commenti in inglese, francese e lo Spagnolo su 1000 impianti e le posizioni importanti all'interno del museo, senza la necessità di seguire un itinerario specifico. Disponibile pure in tedesco, l'italiano ed il giapponese su 350 impianti. Può essere ottenuto dalle entrate alle tre ale del museo (Sully, Denon e Richelieu), sulla presentazione di una carta di identità. Tassa: 5 euros.
Pagamento soltanto in denaro, l'assegno della serie o l'assegno del viaggiatore.

Audioguide
Una audioguida è fornita per gli ospiti al museo. Include i commenti in inglese, francese e spagnolo su 1000 opere e i luoghi importanti all'interno del museo, senza la necessità di seguire un itinerario specifico. Disponibile pure in tedesco, italiano e giapponese su 350 opere. Può essere ottenuta dalle entrate alle tre ali del museo (Sully, Denon e Richelieu), su presentazione di una carta di identità. Tassa: 5 euros.
Pagamento soltanto in denaro, assegno bancario o traveller's cheque.

Information sheets
Information sheets are available for consultation in the museum's galleries. These are hard plastic-covered pages, A3 size, available in several languages (French, English, German, Spanish, Italian, Japanese). The texts explain the collections in that room, or deal with a specific work, artist, style, location, technique etc.
The French version is on sale in reduced format at the bookshop in the Hall Napoléon.

Fogli delle informazioni
I fogli delle informazioni sono disponibili per consultazione nelle gallerie del museo. Queste sono pagine duro plastica-coperte, formato A3, disponibile in parecchie lingue (francese, inglese, tedesco, spagnolo, italiano, giapponese). I testi spiegano le collezioni in quanto stanza, o l'affare con un lavoro, un artista, uno stile, una posizione, una tecnica specifici ecc.
La versione francese è sulla vendita nella disposizione ridotta al bookshop nel Corridoio Napoléon.
Fogli delle informazioni
I fogli delle informazioni sono disponibili per consultazione nelle gallerie del museo. Queste sono pagine coperte con plastica dura, formato A3, disponibili in parecchie lingue (francese, inglese, tedesco, spagnolo, italiano, giapponese). I testi spiegano le collezioni in quella stanza, o trattano di un lavoro, un artista, uno stile, un luogo, una tecnica specifici ecc.
La versione francese è in vendita in formato ridotto alla libreria nella Hall Napoléon.
Disabled Visitors
An orientation guide (in French and English) for less mobile visitors offers a special tour, and is available at the Information Desk. Loan of wheelchairs upon request, tel.: (33) 01 40 20 53 17.
For blind and partially sighted visitors, the Department of Sculptures has opened a gallery where around twenty works can be explored by touch and with the aid of an audioguide.
Guided tours for groups of disabled persons can be organised on request, tel.: (33) 01 40 20 59 90.
Ospiti Disabled
Una guida di orientamento (in francese ed inglese) per gli ospiti meno mobili offre un giro speciale ed è disponibile allo scrittorio delle informazioni. Prestito delle sedie a rotelle su richiesta, tel.: (33) 01 40 20 53 17.
Per i ciechi e gli ospiti parzialmente avvistati, il reparto delle sculture ha aperto una galleria in cui intorno venti impianti possono essere esplorati tramite il tocco e con l'aiuto di un audioguide.
I giri guidati per i gruppi delle persone disabled possono essere organizzati a richiesta, tel.: (33) 01 40 20 59 90.
Ospiti Disabili
Una guida di orientamento (in francese ed inglese) per gli ospiti meno mobili offre un giro speciale ed è disponibile allo banco delle informazioni. Prestito delle sedie a rotelle su richiesta, tel.: (33) 01 40 20 53 17.
Per gli ospiti ciechi e con capacità visiva parziale, il reparto delle sculture ha aperto una galleria in cui circa venti opere possono essere esplorate tramite il tatto e con l'aiuto di un audioguide.
I giri guidati per i gruppi delle persone disabili possono essere organizzati a richiesta, tel.: (33) 01 40 20 59 90.
Regulations for Visitors
Please refrain from using flashes.
It is prohibited to smoke in the Hall Napoléon and in the galleries.
No food may be brought into the galleries or the cloakrooms.
Visitors are not allowed to bring cumbersome objects or animals into the galleries.
The Visit regulation book for visitors may be consulted at the Information desk.
Regolazioni per gli ospiti
Prego astensione dal usando i flash.
È proibita per fumare nel Corridoio Napoléon e nelle gallerie.
Nessun alimento non può essere introdotto nelle gallerie o nei guardaroba.
Agli ospiti non sono permessi introdurre gli oggetti o gli animali ingombranti nelle gallerie.
Visit regulation Il libro per gli ospiti può essere consultato allo scrittorio delle informazioni.
Regole per gli ospiti
Prego astenersi dall'usare i flash.
È proibito fumare nella Hall Napoléon e nelle gallerie.
Nessun alimento può essere introdotto nelle gallerie o nei guardaroba.
Agli ospiti non è permesso introdurre gli oggetti ingombranti o gli animali nelle gallerie.
Il libro del regolamento di visita per gli ospiti può essere consultato al banco delle informazioni.
Why works of art should not be touched
Works of art are unique and fragile.
They have survived centuries and must be preserved for future generations.
Touching, even lightly, a painting, object, sculpture or piece of furniture causes damage.
Especially when this gesture is repeated thousands of times.
Help us protect our common heritage.
Perchè le opere d'arte non dovrebbero essere toccate
Le opere d'arte sono uniche e fragili.
Hanno secoli superstiti e devono essere conservate per le generazioni future.
Toccando, persino leggermente, una pittura, l'oggetto, la scultura o la parte di mobilia causa danni.
Particolarmente quando questo gesture è migliaia ripetute dei periodi.

Aiutili a proteggere la nostra eredità comune.
Perché le opere d'arte non dovrebbero essere toccate
Le opere d'arte sono uniche e fragili.
Sono sopravvissute per secoli e devono essere conservate per le generazioni future.
Toccando, persino leggermente, un dipinto, un oggetto, una scultura o un mobile causa danno.
Particolarmente quando questo gesto è ripetuto migliaia di volte.
Aiutaci a proteggere la nostra eredità comune.

In regards to the Italian translation of the original French text, the general impression is, I should say, that it is good enough. Nevertheless, we have to notice that Italian and French are two closely related languages, so that an Italian can, with a little effort, understand a lot of a French text even if he does not know this language at all. Italian and French have a nearly equal morphology and because of this a lot of words that form with the flexion and conjugation of stems correspond to each other without any ambiguity. Also syntax is almost the same and because of this there is no need to modify words order. Lexicon is made up by words that very often correspond to each other perfectly in their meaning. Therefore, in the case of these two languages, even crude word-by-word translation gives good enough results.
As regards the Italian translation of the English original text, the general impression is clearly worse. Mistakes are much more frequent and there are also parts that are not understandable or that are only hardly understandable or are understandable thanks to the experience we all have about things that are usually written in information leaflets given to visit a museum.
Then it is important to notice that the source text is really elementary. It is quite unusual to write like this.
Therefore, let's try to consider a text more similar to the usual way of writing. This is information about a very famous Internet search engine. It is still a very simple text, but sentences are made up of more than one clause, as usually happens in common speaking and writing. Original texts are still French and English texts (translation from English into French is a little free, but this has no influence on our speech) and the translation into Italian, both the one done by the machine and the one done by the human being are quoted following the same conventions as in the former example.

FRENCH ORIGINAL TEXT

Introduction
Les performances et la précision des recherches Google reposent sur la qualité du matériel et des logiciels utilisés. La quasi-instantanéité des résultats est due en partie à l'efficacité de notre algorithme de recherche et en partie aux milliers (!) de PC que nous avons installés en réseau pour constituer un moteur de recherche ultrarapide.
L'élément fondamental de notre logiciel est PageRank, un système de classement des pages Web mis au point par les fondateurs de Google (Larry Page et Sergey Brin) à l'université de Stanford. Et pendant que plusieurs dizaines d'ingénieurs et de spécialistes consacrent leurs journées à améliorer les différents aspects de Google, PageRank reste la pierre angulaire de nos outils de recherche.
Introduzione
Le prestazioni e la precisione delle ricerche Google si basano sulla qualità strumentazione e software utilizzati. Il quasi-instantanéité dei risultati è dovuta in parte all'efficacia del nostro algoritmo di ricerca ed in parte alle migliaia (!) di PC che abbiamo installato in rete per costituire un motore di ricerca ultrarapida.
L'elemento fondamentale del nostro software è PageRank, un sistema di classificazione delle pagine web messo a punto dai fondatori di Google (Larry pagina e Sergey Brin) all'università di Stanford. E durante che molte decine degli ingegneri e di specialisti dedicano i loro giorni da migliorare i vari aspetti di Google, PageRank resta la pietra angolare dei nostri attrezzi di ricerca.

Introduzione

Le prestazioni e la precisione delle ricerche Google si basano sulla qualità del hardware e software utilizzati. La quasi-instantaneità dei risultati è dovuta in parte all'efficacia del nostro algoritmo di ricerca ed in parte alle migliaia (!) di PC che abbiamo installato in rete per costituire un motore di ricerca ultrarapida.

L'elemento fondamentale del nostro software è PageRank, un sistema di classificazione delle pagine web messo a punto dai fondatori di Google (Larry Page e Sergey Brin) all'università di Stanford. E mentre molte decine di ingegneri e di specialisti dedicano i loro giorni a migliorare i vari aspetti di Google, PageRank resta la pietra angolare dei nostri strumenti di ricerca.

PageRank

PageRank est un champion de la démocratie: il profite des innombrables liens du Web pour évaluer le contenu des pages Web -- et leur pertinence vis-à-vis des requêtes exprimées. Le principe de PageRank est simple : tout lien pointant de la page A à la page B est considéré comme un vote de la page A en faveur de la page B. Toutefois, Google ne limite pas son évaluation au nombre de " votes " (liens) reçus par la page; il procède également à une analyse de la page qui contient le lien. Les liens présents dans des pages jugées importantes par Google ont plus de " poids ", et contribuent ainsi à " élire " d'autres pages.

Les sites qui se distinguent par leur qualité sont affectés d'une valeur PageRank plus élevée, et Google en tient compte lors de chaque recherche. Bien entendu, les pages jugées " importantes " par Google vont vous laisser indifférent si elles ne répondent pas à vos requêtes... Aussi, pour retrouver les pages qui correspondent au mieux à votre requête, Google complète l'évaluation PageRank par des mécanismes évolués de correspondance de texte. Google ne se contente pas de compter le nombre d'occurrences d'un terme de recherche dans une page : il examine différents aspects du contenu de cette page (et du contenu des pages liées à celle-ci) afin de déterminer si elle correspond à votre requête.

PageRank

PageRank è un campione della democrazia: approfitta dei legami innumerevoli del web per valutare il contenuto delle pagine web e la loro pertinenza di fronte delle richieste espresse. Il principio di PageRank è semplice: ogni legame che indica della pagina A alla pagina B è considerato come un voto della pagina A in favore della pagina B. tuttavia, Google non limita la sua valutazione al numero di "voti" (legami) ricevuti dalla pagina; procede anche ad un'analisi della pagina che contiene il legame. I legami presenti in pagine giudicate importanti da Google hanno più "pesi", e contribuiscono così "ad eleggere" altre pagine.

Le unità che si distinguono con la loro qualità sono destinate di un valore PageRank più elevato, e Google ne tiene conto allora di ogni ricerca. Ben intese, le pagine giudicate "importanti" da Google vi lasceranno indifferente se non rispondono alle vostre richieste... Inoltre, per trovare le pagine che corrispondono al massimo alla vostra richiesta, Google completa la valutazione PageRank con meccanismi evoluti di corrispondenza di testo. Google non si accontenta di contare il numero di verificarsi di un termine di ricerca in una pagina: esamina vari aspetti del contenuto di questa pagina (e del contenuto delle pagine legate a questa) allo scopo di determinare se corrisponde alla vostra richiesta.

PageRank

PageRank è un campione di democrazia: approfitta dei legami innumerevoli del web per valutare il contenuto delle pagine web e la loro pertinenza nei confronti delle richieste espresse. Il principio di PageRank è semplice: ogni legame che indica dalla pagina A alla pagina B è considerato come un voto della pagina A in favore della pagina B. Tuttavia, Google non limita la sua valutazione al numero di "voti" (legami) ricevuti dalla pagina; procede anche ad un'analisi della pagina che contiene il legame. I legami presenti in pagine giudicate importanti da Google hanno più "pesi", e contribuiscono così "ad eleggere" altre pagine.

I siti che si distinguono per la loro qualità assumono un valore PageRank più elevato, e Google ne tiene conto al momento di ogni ricerca. Ben inteso, le pagine giudicate "importanti" da Google vi lasceranno indifferenti se non rispondono alle vostre richieste... Perciò, per trovare le pagine che corrispondono al meglio alla vostra richiesta, Google completa la valutazione PageRank con meccanismi evoluti di corrispondenza di testo. Google non si accontenta di contare il numero di volte che un termine di ricerca ricorre in una pagina: esamina vari aspetti del contenuto di questa pagina (e del contenuto delle pagine legate a questa) allo scopo di determinare se corrisponde alla vostra richiesta.

Intégrité

Les méthodes complexes et automatiques utilisées par les recherches Google rendent quasi impossible toute manipulation humaine des résultats. Comme nous l'indiquons clairement dans nos listes de résultat, certains sites peuvent être associés à une publicité " Sponsored Link ". Toutefois, Google ne pratique pas la vente des positions dans ces résultats ; autrement dit, il n'est pas possible d'acheter une valeur PageRank supérieure à la réalité du Web. Avec la recherche Google, vous disposez d'une solution simple, rapide, honnête et objective pour trouver des sites Web de la plus haute qualité et dont les informations répondent parfaitement à vos besoins.

Integrità

I metodi complessi ed automatici utilizzati dalle ricerche Google rendono quasi impossibile ogni manipolazione umana dei risultati. Come lo indichiamo chiaramente nei nostri elenchi di risultato, alcune unità possono essere associate ad una pubblicità "Sponsored Link". Tuttavia, Google non pratica la vendita delle posizioni in questi risultati; in altre parole, non è possibile comperare un valore PageRank superiore alla realtà del web. Con la ricerca Google, disponete di una soluzione

semplice, rapida, onesta ed oggettiva per trovare siti web della più alta qualità e le cui informazioni rispondono perfettamente alle vostre necessità.

Onestà

I metodi complessi ed automatici utilizzati dalle ricerche Google rendono quasi impossibile ogni manipolazione umana dei risultati. Come indichiamo chiaramente nei nostri elenchi di risultato, alcuni siti possono essere associate ad una pubblicità "Sponsored Link". Tuttavia, Google non pratica la vendita delle posizioni in questi risultati; in altre parole, non è possibile comperare un valore PageRank superiore alla realtà del web. Con la ricerca Google, disponete di una soluzione semplice, rapida, onesta ed oggettiva per trovare siti web della più alta qualità e le cui informazioni rispondono perfettamente alle vostre necessità.

ENGLISH ORIGINAL TEXT

Introduction

Google runs on a unique combination of advanced hardware and software. The speed you experience can be attributed in part to the efficiency of our search algorithm and partly to the thousands of low cost PC's we've networked together to create a superfast search engine.

The heart of our software is PageRank™, a system for ranking web pages developed by our founders Larry Page and Sergey Brin at Stanford University. And while we have dozens of engineers working to improve every aspect of Google on a daily basis, PageRank continues to provide the basis for all of our web search tools.

Introduzione

Google funziona su una combinazione unica di hardware e di software avanzati. La velocità che sperimentate può essere attribuita in parte al risparmio di temi della nostra procedura di ricerca ed alle migliaia del we've del pc di basso costo networked per creare parzialmente insieme un motore di ricerca del superfast.

Il cuore del nostro software è PageRank™, un sistema per le pagine di Web di posto sviluppate dai nostri fondatori Pagina di Larry e Sergey Brin all'università di Stanford. E mentre abbiamo dozzine degli assistenti tecnici che lavorano per migliorare ogni funzione di Google su una base quotidiana, PageRank continua a fornire la base per tutti i nostri strumenti di ricerca di Web.

Introduzione

Google funziona in base a una combinazione unica di hardware e di software avanzati. La velocità che sperimentate può essere attribuita in parte alla efficienza del nostro algoritmo di ricerca e parzialmente alle migliaia di pc di basso costo che abbiamo connesso insieme in rete per creare un motore di ricerca superveloce.

Il cuore del nostro software è PageRank™, un sistema per classificare le pagine Web sviluppato dai nostri fondatori Larry Page e Sergey Brin all'università di Stanford. E mentre abbiamo dozzine di ingegneri che lavorano quotidianamente per migliorare ogni aspetto di Google, PageRank continua a fornire la base per tutti i nostri strumenti di ricerca Web.

PageRank Explained

PageRank relies on the uniquely democratic nature of the web by using its vast link structure as an indicator of an individual page's value. In essence, Google interprets a link from page A to page B as a vote, by page A, for page B. But, Google looks at more than the sheer volume of votes, or links a page receives; it also analyzes the page that casts the vote. Votes cast by pages that are themselves "important" weigh more heavily and help to make other pages "important."

Important, high-quality sites receive a higher PageRank, which Google remembers each time it conducts a search. Of course, important pages mean nothing to you if they don't match your query. So, Google combines PageRank with sophisticated text-matching techniques to find pages that are both important and relevant to your search. Google goes far beyond the number of times a term appears on a page and examines all aspects of the page's content (and the content of the pages linking to it) to determine if it's a good match for your query.

PageRank Ha spiegato

PageRank conta sulla natura unicamente democratica del Web usando la relativa struttura ampia di collegamento come indicatore di valore della pagina specifica. Nell'essenza, Google interpreta un collegamento dalla pagina A per paginare la B come .. voto, dalla pagina A, per la pagina B. Ma, Google guarda più del volume puro dei voti, o collega una pagina riceve; egualmente analizza la pagina che lancia il voto. I voti lanciano dalle pagine che sono essi stessi "importanti" pesano più pesante e contribuiscono a rendere altre pagine "importanti."

I luoghi importanti e di alta qualità ricevono un più alto PageRank, che Google si ricorda di ogni tempo che conduce una ricerca. Naturalmente, le pagine importanti non significano niente a voi se non abbinano la vostra domanda. Così, Google unisce PageRank con le tecniche testo-abbinanti specializzate alle pagine del ritrovamento che sono sia importanti che relative alla vostra ricerca. Google va lontano oltre il numero di volte che un termine compare ad una pagina ed esamina

tutte le funzioni del soddisfare della pagina (e del contenuto delle pagine che si collegano ad esso) per determinare se esso […] una buona corrispondenza per la vostra domanda.

PageRank Ha spiegato

PageRank conta sulla natura unicamente democratica del Web usando la sua ampia struttura di collegamento come indicatore di valore della pagina specifica. In essenza, Google interpreta un collegamento dalla pagina A alla pagina B come un voto, dalla pagina A, per la pagina B. Ma, Google guarda più del volume puro dei voti, o collegamenti che una pagina riceve; esso analizza anche la pagina che dà il voto. I voti dati dalle pagine che sono esse stesse "importanti" pesano di più e contribuiscono a rendere altre pagine "importanti."

I siti importanti e di alta qualità ricevono un più alto PageRank, che Google ~~si~~ ricorda ~~di~~ ogni volta che conduce una ricerca. Naturalmente, le pagine importanti non significano niente per voi se non rispondono alla vostra domanda. Così, Google unisce PageRank con ~~le~~ sofisticate tecniche di abbinamento del testo per trovare pagine che sono sia importanti che pertinenti alla vostra ricerca. Google va molto oltre il numero di volte che un termine compare su una pagina ed esamina tutti gli aspetti del contenuto della pagina (e del contenuto delle pagine che si collegano ad essa) per determinare se esso è una buona risposta per la vostra domanda.

Integrity

Google's complex, automated methods make human tampering with our results extremely difficult. And though we do run relevant ads above and next to our results, Google does not sell placement within the results themselves (i.e., no one can buy a higher PageRank). A Google search is an easy, honest and objective way to find high-quality websites with information relevant to your search.

Integrità

Google complesso, metodi automatizzati rende l'alterazione umana i nostri risultati estremamente difficile. E benchè facciamo funzionare il ads relativo sopra e vicino ai nostri risultati, Google non vende la disposizione all'interno dei risultati essi stessi (cioè, nessuno possono comprare un più alto PageRank). Una ricerca di Google è un modo facile, onesto ed obiettivo […] trovare i Web site di alta qualità con le informazioni relative alla vostra ricerca.

Onestà

I complessi metodi automatizzati di Google rendono l'alterazione umana dei nostri risultati estremamente difficile. E benché facciamo funzionare un rilevante ads [*sistema di annunci pubblicitari, N.d.T.*] sopra e vicino ai nostri risultati, Google non vende il piazzamento all'interno dei risultati stessi (cioè, nessuno può comprare un più alto PageRank). Una ricerca di Google è un modo facile, onesto ed obiettivo per trovare siti Web di alta qualità con le informazioni rilevanti per vostra ricerca.

The more complex structure of the sentences in this text has not substantially modified the quality of machine translation in the case of translation from French into Italian, as was to be expected since morphology and syntax of these two languages are very similar. Instead, the considerable differences in syntax and especially the big differences in morphology between English and Italian leads, in the case of this more complex text, to a translation quality which is clearly worse.

I do not quote translations of other passages because, in the case of medium complex texts, that is texts more complex than this (which is still a simple text), expected translation quality is obviously worse or, at most, similar (naturally, this is true, and even more so, in the case of complex or very complex texts). Let's look carefully at the kind of errors in the machine translation. They are substantially of three kinds.

1) The first kind of errors are the ones due to the fact that the program sometimes does not "understand" which other word (or words group) a word (or words group) is referred to and/or which grammar category it belongs to. This does not cause many errors in the translation from French into Italian because of the reasons we just talked about. But in the translation from a language with very little morphology such as English into a language with a rather rich morphology as Italian (and even more so in the case of languages having an even richer morphology because they have cases) the errors are many. In the machine translation of the two passages quoted above there are examples of verbs translated in a completely wrong way in regards to time and mood, of nouns mistaken for verbs and vice versa, of gross mistakes in word order and so on. Let's notice that such errors, besides being numerous, are also the most serious. In fact, for the reader who does not know the language of the original text at all, it is sometimes impossible to understand the meaning of those parts of machine translation which are

clearly completely lacking in sense. This is true even if he has a bilingual dictionary, because it has no flexed forms. In the text of the machine translation quoted as an example the errors that are strictly of this kind have been highlighted by underlining (besides the yellow highlighter, common to all the serious mistakes).

2) Another kind of error is the one due to the fact that some words have more than one meaning (for example, the English word "light" can mean "not dark" or "not heavy") or different nuances. This kind of error, frequent too, is however less serious than the former, because they can be corrected by the reader basing himself on the general sense of the text or using a bilingual dictionary.

3) The third kind of error consists of the inaccuracies, the expressions sounding bad in a certain language etc. These errors, after all, do not compromise the understanding of the text. Thus, even if they are rather frequent, they are clearly the least serious.

The following table shows the frequency of these three kinds of errors in the text of the machine translation quoted as an example.

	1st kind errors (reference errors)	***2nd kind errors (meaning errors)***	***3rd kind errors (inaccuracies etc.)***
"museum" French passage (22 sentences, 403 words)	13 **0,59 — 3,22**	18 **0,81 — 4,46**	4 **0,18 — 0,99**
"museum" English passage (22 sentences, 379 words)	31 **1,40 — 8,17**	22 **1 — 5,80**	16 **0,72 — 4,22**
"Google" French passage (16 sentences, 449 words)	3 **0,18 — 0,66**	18 **1,16 — 4,00**	**3** **0,18 — 0,66**
"Google" English passage (15 sentences, 346 words)	30 **2 — 8,59**	16 **1,06 — 4,62**	21 **1,4 — 6,06**

The upper digit in the box indicates the absolute number of the mistakes, while the two lower digits (highlighted in bold type) indicate, the one on the left the average number of mistakes in a sentence (generally, every single word translated incorrectly has been considered as one error), the one on the right the percentage of incorrectly translated words. In the case of the translation from English into Italian, the average number in a sentence of the 1st and 2nd kind mistakes (the ones we are more interested in), even in the case of the first passage, which is extremely simple and the sentences of which are made up of only one clause, is equal or superior to the unity, for each of these two kinds of errors. As we can see, they are rather high frequencies, if we consider that even only two or three mistakes of these two kinds in one sentence can sometimes make the general sense of the whole sentence not understandable or hardly understandable.

Ceccato's device (with some modifications and additions of my own)

So, let's see what is the device proposed by Ceccato in order to solve, completely or at least largely, the problem of the errors in machine translation derived from the fact that the machine is not able to reconstruct the correlational network of thought starting from text because it lacks the ability of understanding the meaning of words neither does it have the general culture which in many cases allows the human being to understand the way words are related to each other.

The first and fundamental theoretical presupposition which is the base of this device is, naturally, the correlational theory of thought, briefly exposed in my article "A presentation of Operational Methodology" and, a little more in detail, in this article. The second presupposition is that if we in some way reconstruct the correlational network of thought of which a sentence is the expression and we base MT

on this result, we achieve a clearly better translation quality. This is easily understandable basing ourselves on what we have said so far. Coming back to our example of the two expressions

- "some amount of water, even small"
- "some amount of water, even cold"

if we first rebuild the corresponding correlational network of thought, then it will be easy to make the program translate in the correct way the two adjectives "small" and "cold" into a language which flexes the adjective according to the case of the noun it is referred to. We have seen other examples too where lack of a correct reconstruction of the correlational network of thought leads to gross translation errors and we have also seen that these cases are very frequent.

Therefore, reconstructing the correlational network of thought corresponding to speech expressing it before translating will allow a big increase in translation quality itself.

Ceccato started from the presupposition that this reconstruction had to be a task of the machine. This was because his research was financed by USA Armed Forces who, in the '50s and '60s, that is at the time of the so-called "cold war" between USA and Soviet Union, where interested in a more economic solution than human translation for the huge amount of Russian texts which was at the time translated every day. In this case those who produced these texts had no intention of making their translation easier, but it is not always like this. On the contrary, I should say that today, considering the strong tendency towards international diffusion of information, the authors of a large part of produced texts would like these texts to be easily translatable into other languages.

Thus, even if having a program that translates in a satisfying way starting from a non pre-edited text is more desirable, it would be surely much easier to implement a program that translates starting from a text where the reconstruction of the correlational network of thought has been done before by the human being. Already in such a way the human translator, that is the person who knows at least two languages, is completely eliminated. The pre-editing of text, written in a certain language, is done by a person who only needs to know that language very well (and, naturally, the correlational theory of thought). Hence this work can be done by anyone for any text written in his mother-language.

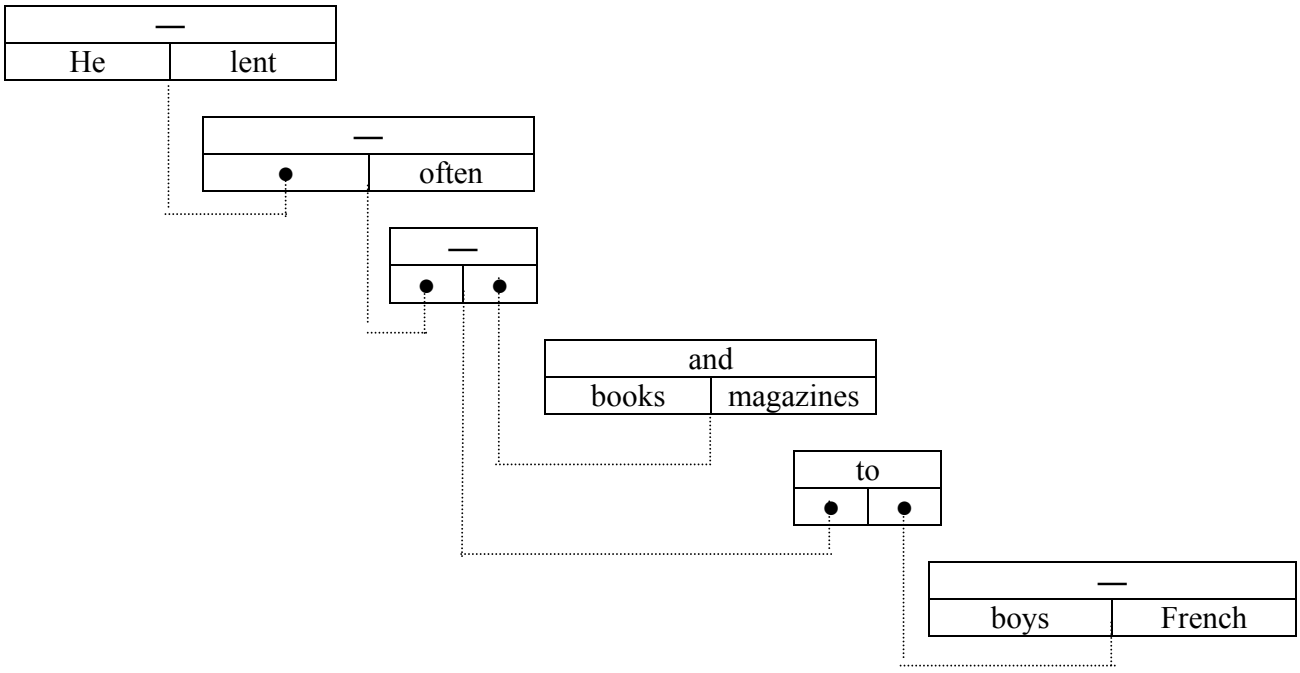
Even if in this article I shall mainly speak about Ceccato's device (conceived in order to implement a fully automatic translation program), I would like to point out a proposal of mine in order to implement a semi-automatic translation program.

Therefore my proposal is to use as a starting base for machine translation some kind of "formatting" of the text clearly showing which is the corresponding correlational network of thought. This formatting of the text naturally must be very simple to write and to read. The way of representing correlational networks proposed by Ceccato is not suitable for this purpose, first of all because it occupies a lot of space in the page, and then because the graphic representation of the network could be more simple.

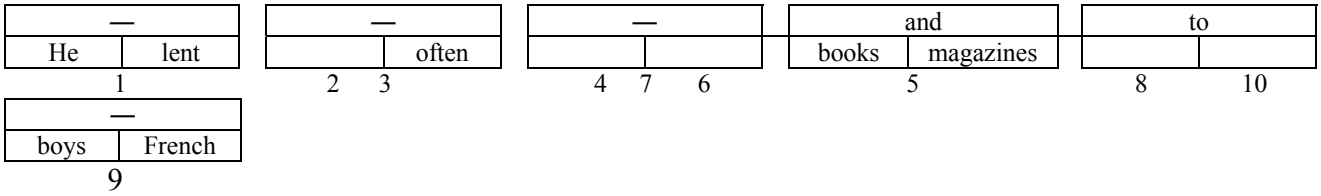
Let's take as an example a rather simple sentence, that is a sentence made up by only one clause, yet including all fundamental parts of speech (article, noun, pronoun, adjective, verb, adverb, conjunction and preposition) and fundamental syntactical elements (subject, predicate, direct object, indirect object). Many of the sentences we produce are more or less of this kind. An example of this kind could be:

He often lent books and magazines to the French boys.

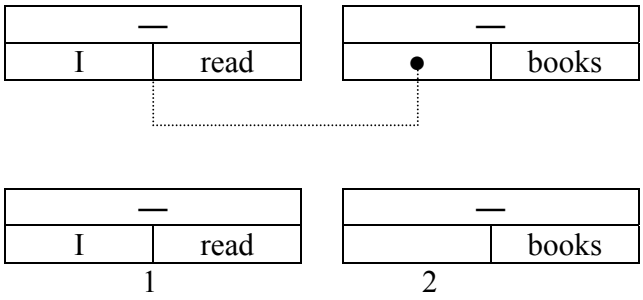
According to the way of representing correlational networks proposed by Ceccato the one corresponding to this sentence is the following (for the sake of simplicity, the correlation article-noun, which has the implicit correlator as correlator, has been indicated putting the two terms in the same box instead of using a specific correlational triad):



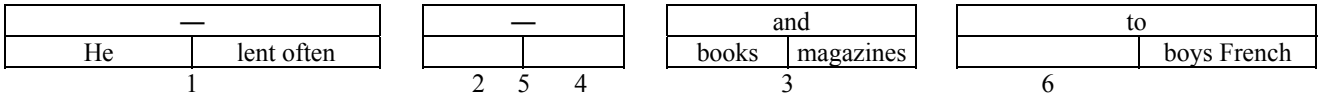
In order to solve the problem of the big amount of space occupied and the rather complex graphic representation of the network, it is sufficient to dispose the correlations in a linear way and to represent their relationships by simple numbers, in this way:



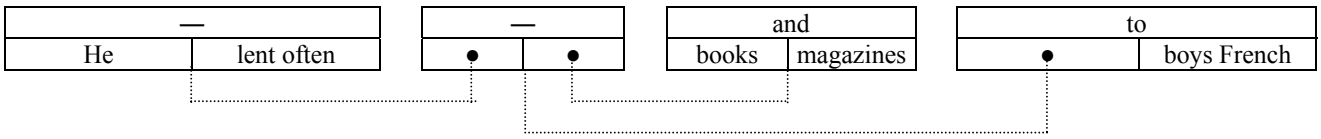
The number under the division between the two lower boxes means that the whole correlation is the correlatum of another correlation; the number under one of the two boxes of correlata, in an approximately central position, indicates which is this correlation and the first correlation being either the first or the second correlatum of the second correlation. That is the two numbers, the one immediately following the other in the series of numbers, substitute, the first the beginning of the dashed line and the second its end (represented by the big black dot); the obliged way to go from the former to the latter, following numbers order, substitutes the dashed line, in this way:



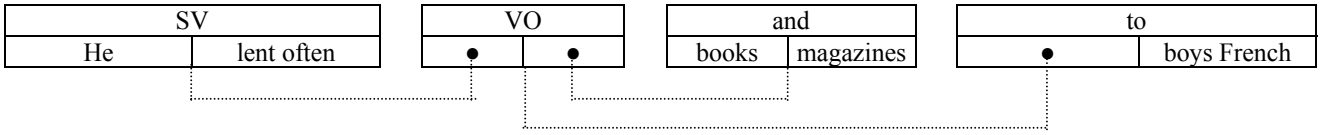
If we want to simplify even more, we can use the convention we used in the article-noun correlation, that is not representing the two terms separately, also with other correlations of which the correlator is the implicit correlator, as verb-adverb correlation, substantive-adjective correlation etc., simply putting, also in this case, the two terms in the same box. In this case the correlational network corresponding to the sentence of our example would be represented like this:



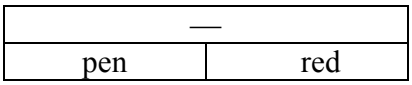
It is very easy to implement a program which allows the human being to write and eventually to quickly modify the correlational networks represented as I proposed (it is only necessary that the page is already formatted with rows made up by triads of boxes and underlying lines for the numbers; that there is a system in order to put the numbers quickly in the right place etc.). With the aid of a program like this typing in a text in the form of correlational networks of thought is not much longer than typing in it in the usual way. Also identifying the correct correlational network starting from the sentence is a task a human being learns, after some training, to do quickly.
Indicating the relationships among the correlational triads by numbers instead of dashed lines is completely lacking in ambiguities, hence it is perfectly suitable for a computer program. Instead, the human being catches these relationships more quickly and easily if they are indicated by dashed lines. It is very easy to build a program that allows us to pass automatically from one way of indicating the correlational networks to the other. The dashed lines can be used even if the correlational triads are disposed in a linear way, putting the dashed lines parallely under the correlational triads, in this way:



Finally, it is suitable, because of reasons we shall further see, to indicate some kinds of syntactic relationships between correlata of which the correlator is the implicit correlator, like the relationships subject-verb (SV) and verb-object (VO). They can be indicated in this way:



Having a machine translation program starting from a text where the correlational network has been identified by the human being is useful in itself. But such a program is interesting also for another reason. In fact, both if the task of identifying the correlational network of thought in source language is given to a human being, or if we are trying to implement a program able to do this, once we have achieved this result the following phase is the same in both cases, that is implementing a program which starts from the correlational network of thought and expresses it by the lexicon and the grammar rules of the target language. In other words, the human aided machine translation program that starts from a text "formatted" so as to indicate the correlational network of thought is the same as the second of the two phases a fully automatic translation program must consist of. Hence such a program must be created in any case. The creation of such a program should not be problematic. In any language we have to follow rigid and well-known rules when we pass from a thought to the speech that is its expression. Therefore, the implementation of a program doing this should not be difficult. For example, the correlation substantive-adjective, such as the following:



is generally expressed in English putting adjective before substantive, while in Italian the adjective follows the substantive, except in particular cases. Certainly putting these rules in the program is not difficult. Therefore, since the correlational network of thought we start from is well known, the translation will be always correct. Gross translation errors can occur, as we saw, when the correlational network of thought is unknown.
Apart from the considerations about the necessity and the utility of a machine translation program which starts from the correlational network of thought, nevertheless a program able to translate with satisfying results without needing a pre-editing of text done by the human being would be much more desirable. Therefore let's look at the device proposed by Ceccato in order to implement such a program.

As in other programs, the single words are used as "input items" and, in the case of inflectional languages, the program dictionary is a dictionary where there are all possible flexed forms of every single entry of the source language dictionary. The originality of Ceccato's proposal is the fact that the program tries to reconstruct the correlational network of thought of every single sentence. This is done essentially in two ways

1) All the words of source language are classified into two categories. The first category is the one of correlators, the second the one of correlata. Therefore, the first category is made up of conjunctions, prepositions and, in languages having cases, also by cases, while the second category is practically made up of all other words. As we saw, in the correlational triad correlata always occupy the lower boxes. Instead, correlators always occupy the upper box, except rare cases that anyway are generally marked in a particular way. An example of these rare cases is the sentence

- "and" and "or" are conjunctions

where the conjunction "and" and the conjunction "or" put between inverted commas are not correlators, but correlata tied by the other conjunction "and". Such cases, apart from the fact that they are very rare, anyway are, as we can see, clearly marked, because the two conjunctions, when performing the

role, which is anomalous for them, of correlata, when we write they are put between inverted commas and in speech are isolated by particular pauses.
There are also words indicating a correlatum and a correlator at the same time. Examples of this kind are all the verbs in a personal form in the languages, like Italian and many other languages, which flex the verb according to the person it is referred to. In fact, the personal verb is a word indicating three things at the same time:

- a correlatum
- a particular correlator, the implicit correlator
- the position of the correlatum, that is the one of second correlatum.

That is, for example, the personal form of the verb "to laugh" in the sentence "John laughs" indicates that the corresponding correlation of thought is without doubt the following:

—	
John	laughs

or, written in a more exact form, the following:

-s	
John	laugh-

because the second correlatum is not properly "laughs", but the meaning of the verb "to laugh", that is the stem "laugh-", while the termination "-s" indicates, beyond the other, that the meaning of the verb "to laugh" is correlated by the implicit correlator to a third person singular, in this case John, which is the first correlatum, while the verb "to laugh" has to be the second correlatum. Therefore, if we wanted to write this correlation in a completely exact way, we should write it like this:

-s (= implicit correlator, which correlates the second correlatum with a third singular person)	
John	**laugh-** + **-s** (= indicative present)

This is because English is sometimes, even if much less than many other languages, a "fusional" language, that is it sometimes tends just to fuse in only one termination more than one meaning. In fact, that simple termination "-s" indicates that the meaning of the verb "to laugh" is conjugated to the indicative present tense and it is correlated by the implicit correlator with a third person singular.
Another very important example of words at the same time indicating a correlatum and a correlator are, in languages having cases, all the words flexed according to a case, that is all substantives and eventually also all adjectives. In the case of the substantive flexed according to a particular case, for example the Latin substantive flexed to the genitive case *urbis* ("of the city"), the stem (*urb*-) indicates the correlatum, that is "city", while the termination (*is*-) indicates the correlator, that is the mental category "genitive", which instead in English is expressed by a preposition, the preposition "of" (in languages not having cases the corresponding mental categories are expressed by prepositions). When a language, such as Latin, flexes also the adjective giving it the same case it gives to substantive which it is referred to, the flexed form of the adjective does not indicate the correlator "case" (genitive, dative etc.), but the implicit correlator and the fact that it ties the adjective, acting as second correlatum, with a substantive flexed according to the same case.
Naturally, the cases where a word at the same time indicates a correlatum and a correlator, and eventually gives us also some information about how the other correlatum must be, make it much easier to re-

construct the correlational network of thought, both for the human being listening or reading and for the computer program we are talking about.
We said that the first presupposition on which this machine translation device is based is the distinction of all the words of a language into two categories, the one of correlators and the one of correlata. Thus, the program will have two lists of words, the words almost always (except the aforesaid rare cases) indicating correlators (that is the words grammar classifies as prepositions and conjunctions), and the list of words always indicating correlata. In the case of the languages tending to fuse into only one word a correlatum and a correlator (and eventually also other information about the other correlatum) the list of correlata will comprehend all the possible flexed forms for every entry of common dictionaries. Beside each flexed form there will be the stem (or the nominative form) and what the flexion (generally a termination) expresses, specified following grammar terminology (for example: "genitive", "dative", "implicit correlation with a third person singular" etc.).
The program will always (unless in the aforesaid rare cases) put the words of the correlators list in the upper boxes of the correlational triads, while it will always put in the lower boxes the words of the correlata list. This is the first device allowing the program to reconstruct the correlational network of thought starting from the text expressing this thought.

2) The second device is to provide the program with a so-called "notional sphere" made in such a way as to allow the program to complete the reconstruction of the correlational network of thought carrying out the same function which the basic culture, we all have, has in the human being listening or reading. As we saw, this basic culture very often allows us to reconstruct the structure of some correlations and the relationships among them even if in the speech or the text the relevant information is not present. The human being can do this because he understands the meaning of the words of a language which are correlata and he also has a basic culture which allows him to establish if there can be a relationship or not between the things designated by these words. Reproducing this function in a machine is at the moment absolutely impossible, but it is possible to approximate to human translation outcome by substituting the comprehension of the meanings of the words and the notions about the possible relationships among things designated by them with complex lists of classifications which make up the notional sphere. Essentially the lists are the following three.

a) A first classification consists of dividing all the words of the program dictionary in more or less general classes. The general criterion in order to establish what these categories must be is that inside each class all the words have features determining their possibility or impossibility to be correlated with other words. For example, all items of the class "solid food" cannot be the direct object of a verb like "to drink", while they can be the direct object of a verb like "to eat", and it is very probably like this if in a sentence one of these words is near the verb "to eat"; all the items of the class "vegetables" cannot be the subject of verbs like "to walk", "to fly" etc., while they can be the subject of verbs like "to bloom", "to dry up", "to grow" etc. Nevertheless it is not possible *a priori* to exactly establish what these classes must be. Ceccato and his collaborators did these divisions into classes basing themselves on the experience acquired making up a lot of sentences (about 1000) of the sentences that can be made up starting from a certain dictionary (they used an essential English dictionary of about 500 words). First they introduced very general classes (for example: "living beings", "not living beings", "human beings", "animals", "plants", "minerals", "solid substances", "liquid substances" etc.), then, examining each of the aforesaid 1000 sentences, they verified if the classes list they made out was sufficient to exclude the incorrect correlational networks. If it was not like this, they introduced new more specific classes. For example, if the class of words designating "animals" was not sufficient to exclude that an item of it, for example the word "dog", present in one of the aforesaid 1000 sentences, was part of a grammatically possible but incorrect correlation (for example the correlation having "dog" as subject of the verb "to fly") a new class or new classes allowing the exclusion of this incorrect correlation were introduced (for example: "air animals", "water animals" and "earth animals"). The word "dog" in our

example was hence classified again also as "earth animal" so as to exclude the possibility of it being the subject of a verb like "to fly".
Below I reproduce the list of classifications used by Ceccato and his collaborators. As we can see very well, the list comprehends more general categories and less general categories. All items of each category presumably can or cannot be correlata of certain other words. There are also even more specific classifications (for example the class "gardening tools", which, as I said, have been introduced to solve cases of ambiguity in the reconstruction of the correlational network which really occurred in the aforesaid 1000 sentences.

Table 1

List of the classifications[13]

1) living beings
2) not living beings
3) animated beings
4) not animated beings
5) human beings
6) animals
7) plants
8) minerals
9) parts of 003
10) parts of 005
11) parts of 006
12) parts of 007
13) collectives of 002, 004
14) collectives of 005
15) collectives of 006
16) collectives of 007
17) 005 + occupation (or activity carried out)
18) 005 + geographic belonging
19) 005 + political belonging
20) 005 + familiar relationships
21) 005 + social relationships
22) political communities
23) water animals
24) air animals
25) earth animals
26) slithering animals
27) herbivorous animals
28) carnivorous animals
29) dangerous animals
30) meek animals
31) wild animals
32) domestic animals
33) digging animals
34) predatory animals
35) animals for slaughter
36) animals for hunting
37) fruit trees
38) opaque things
39) transparent things
40) liquid substances
41) aeriform substances
42) solid substances
43) fluid substances
44) powders
45) fluid substances and powders
46) aeriform transparent substances
47) aeriform opaque substances
48) solid transparent substances
49) celestial bodies
50) atmospheric agents
51) atmospheric phenomena
52) meteorological conditions
53) cardinal points
54) geographic extensions
55) geographic extensions of earth
56) geographic extensions of water
57) geographic extensions characterised by the shape
58) geographic extensions characterised by the ground

[13] An item of the original list has been eliminated because I think it is too similar to another one.

59) geographic extensions characterised by the flora
60) foods
61) solid foods
62) liquid foods and drinks
63) powdered foods
64) fruit
65) vegetables
66) natural objects
67) man-made objects
68) inhabited places
69) buildings
70) parts of 069
71) inner parts of 069
72) external parts of 069
73) objects of interior decoration
74) furniture
75) textiles
76) clothes
77) personal belongings
78) parts of 074
79) parts of 076
80) instruments
81) means of transport
82) aquatic means of transport
83) aerial means of transport
84) terrestrial means of transport
85) domestic tools
86) containers
87) musical instruments
88) gardening instruments
89) toys
90) measuring instruments
91) measures
92) linear measures
93) square measures
94) cubical measures
95) weights
96) time measures
97) names of the days of the week
98) names of the months
99) indications of time coming from astronomy
100) economic objects
101) semantic objects
102) events
103) public places
104) places open to the public
105) places where economic activity is carried out
106) public services
107) covered things
108) covering things
109) things that can be opened
110) things opened or closed by removal or adding
111) things opened or closed by their position as regards other
112) things opened or closed by the position of parts of them
113) things opening or closing something by their rotation or sliding
114) instruments in order to open or to close
115) things that can be hung
116) instruments in order to hang
117) products of art
118) things that can be held in hand
119) fixed things (that can be held in hand)
120) mobile things (that can be held in hand)
121) transportable things
122) things transportable by sliding
123) things transportable by pushing them
124) things transportable by pulling them
125) pointed things
126) cutting things
127) signs and things used in order to indicate
128) numbers
129) things being in pairs
130) things being in rows
131) ....

---

b) The second fundamental classification we have to do is the one concerning the relationships among things. We need to identify what are the most common relationships between two things. The list of these relationships made out by Ceccato and his collaborators is the following.

Table 2

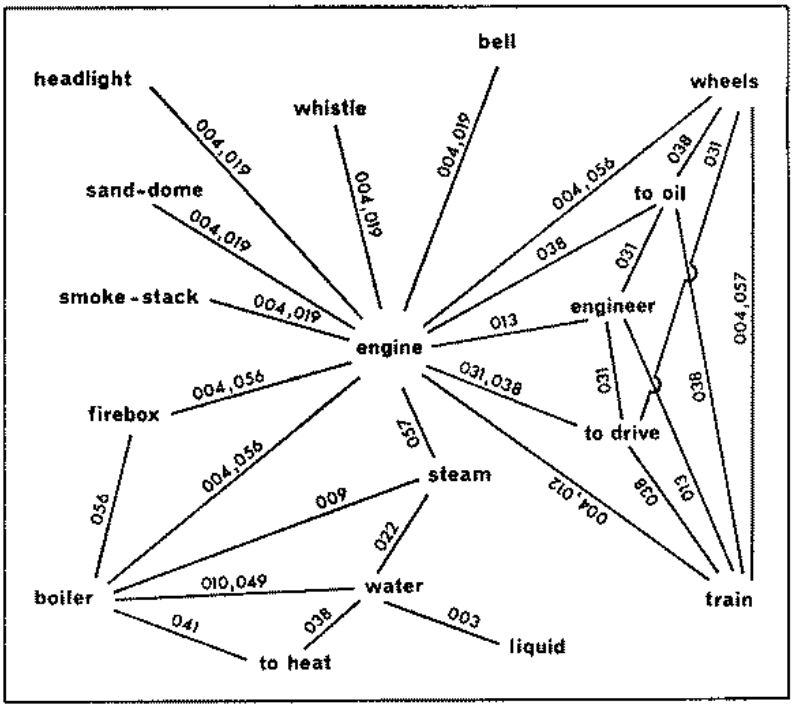
Notional sphere: relationships list[14]

1)	item	group
2)	individual	class
3)	species	genus
4)	part	whole
5)	component	compound
6)	characteristic	characterised thing
7)	product	producer
8)	product	place of production
9)	contents	container
10)	pulled thing	pulling thing
11)	directed or guided thing	directing or guiding thing
12)	thing coming from	place it comes from
13)	preceding thing	following thing
14)	covered or closed thing	covering or closing thing
15)	decorated thing	decoration
16)	pushed thing	pushing thing
17)	main thing	accessory thing
18)	mean of protection	thing we protect against
19)	material	form
20)	material	its physical state
21)	preceding stage of development	following stage of development
22)	ascendant kinship	descending kinship
23)	collateral kinship	
24)	historical association	
25)	contiguity	
26)	opposition	
27)	economic relationship	
28)	semantic relationship	
29)	subject	activity
30)	subject	object of the activity
31)	subject	result of the activity
32)	subject	material of the activity
33)	subject	instrument of the activity
34)	subject	usual place of the activity
35)	subject	usual time of the activity
36)	subject	object
37)	activity	result
38)	activity	material
39)	activity	instrument
40)	activity	usual place
41)	activity	usual time
42)	result of the activity	material of the activity

[14] Two items of the original list have been eliminated because I think they are too similar to others.

43) result of the activity		instrument of the activity
44) object of the activity		result of the activity
45) object of the activity		material of the activity
46) object of the activity		instrument of the activity
47) object		usual place
48) object		usual time
49) material of the activity		instrument of the activity
50)	complementary instruments	
51)	complementary objects	
52) thing		material
53) function		organ
54) organ		machine
55) subject		agent of the activity
56) ……..		…….

At this point we need to actually build the notional sphere relevant to the program dictionary. A piece of the notional sphere relevant to a very small dictionary (the one made up by the words of a children tale, “The little train” by L. Lenski) is shown in picture 1.



Picture 1

The lines uniting the words indicate that between the things designated by these words themselves at least one of the relationships listed in table 2 has been found; the numbers along the lines are the code numbers of the particular kind of relationship. For example, the line uniting the two words “boiler” and “engine” means that between the things designated by these two words there are the relationships of “part-whole” (004) and “organ-machine” (056).

Not all possible kinds of relationship are explicitly indicated, but some relationships are drawn by the program, according to derivation rules, starting from some basic relationships. For example, in the case of the relationship genus-species a rule allows us to transfer the relationships of the thing classified as genus to the thing classified as species (naturally, except the relationship species-genus with the opposite direction). For example, if "water" is classified as species of "liquid", and "liquid" as object of "to pour", the latter relationship will be transferred also to "water", which will automatically become an object of the activity of pouring. Instead, the relationship genus-species that "liquid" has with "whisky", for example, will not be transferred to "water".
Another rule allows us to classify as "species of the same genus" words like "water" and "whisky" through the relationship they have with "liquid".
Another rule establishes that some relationships, such as species-genus, part-whole, object-material etc. have the property of automatically transferring in the ambit of the same relationship and the same direction. If, for example, "piston" is a part of "engine", which in its turn is a part of "car", also "piston" is a part of "car".
Other rules are described by Ceccato and collaborators in the following way:

Un altro gruppo di regole riguarda i rapporti fra attività e loro complementi Il verbo "dipingere", per esempio, è stato inserito in una relazione di attività-materiale con "colore"; di attività-risultato con "quadro"; e di attività-strumento con "pennello". Da questi rapporti è possibile dedurre che fra "colore" e "quadro" si pone il rapporto di materiale-risultato, fra "quadro" e "pennello" quello di risultato-strumento, ecc. I rapporti sono stati invece espressi caso per caso quando il dizionario non possieda un nome per l'attività. Per esempio, è esplicitamente indicato il rapporto 'soggetto-risultato della sua attività' per 'ciabattino/calzolaio' e 'scarpe', in quanto le attività designate con 'fare', 'fabbricare', 'aggiustare', ecc., hanno raggio di applicabilità troppo vasto per essere connesse con particolari soggetti o con particolari risultati/prodotti.

Another group of rules concerns the relationships between the activities and their objects. The verb "to paint", for example, has been put in a relationship of activity-material with "colour"; of activity-result with "painting"; and of activity-instrument with "brush". From these relationships it is possible to deduce that between "colour" and "painting" there is the relationship of material-result, between "painting" and "brush" the one of result-instrument etc. The relationships have instead been expressed case by case when the dictionary has not a name for the activity. For example, the relationship "subject-result of its activity" is explicitly indicated for "cobbler/shoemaker" and "shoes", because the activities designated by "to make", "to produce", "to mend" etc. have an applicability which is too large to be related with particular subjects or particular results/products.

c) The third classification consists of identifying all the verbs of which the meaning already suggests which objects they can have. For these verbs individual notional spheres called "constellations" are built. Below I partially[15] reproduce, as an example, the constellation-table built by Ceccato and collaborators for the Italian verb "portare".

[15] The original table also comprehends the columns relevant to Russian and Polish.

Table 3

Contenuto semantico		Caratteristiche del soggetto	Caratteristiche dell'oggetto e suoi rapporti	Modalità	Costellazione	Francese		Inglese		Tedesco	
							Costellazione		Costellazione		Costellazione
1	Attività di trasportare una cosa (animata o inanimata) da un luogo ad un altro	Essere animato Forza naturale	Non muoventesi di moto proprio. Rispetto al soggetto è il rapporto di sostenuto-sostenente. (all'accusativo [in italiano per i pronomi])		da a, a + art. dat. (pron.) in su + art.	porter porter porter porter porter	de à, à + art. dat. (pron.) en, dans sur	to carry to carry to carry to carry to carry	from to, at dat. In on	tragen tragen tragen tragen tragen	von zu dat. in auf
2	Reggere un peso				con	porter	avec	to carry	with	tragen	mit
				dentro			dedans	to carry in		hereintragen	in
				fuori	di, da	sortir	de	to carry out	of	heraustragen	aus
				su		monter	à + art.	to carry up	at, on	herauftragen	auf
				giù	di, da	descendre		to carry down	from	hereuntertragen	von
				via	di, da	emporter		to carry away	from	wegtragen	von
				indietro	a, dat. (pron.)	reporter		to carry back	to, dat. (pron.)	zurücktragen	dat.
				direzione del soggetto, o di colui che parla	a, dat. (pron.)	apporter		to bring		bringen	zu, dat.
		Essere umano	Essere animato muoventesi di moto proprio (all'accusativo)	direzione del soggetto, o di colui che parla	a, dat. (pron.)	amener		to bring		bringen	zu, dat.
3	Attività di trasferimento dell'oggetto da un luogo ad un altro tramite un mezzo di trasporto	Essere umano	Essere animato o mezzo di trasporto (all'accusativo)		a, a + art.	porter conduire	à, à + art. à, à + art.	to drive	to	führen führen	zu, nach
		Mezzo di trasporto meccanico o animato			a, a + art.	porter	à, à + art.	to drive	to	führen	zu, nach
				Contenente l'arrivo	a, a + art.	apporter	à, à + art.	to bring	to	bringen	
				Contenente la partenza	via, da	emporter	de, de + art.	to carry away	from	wegführen	von
4	Condurre	Essere animato, strada, via			a, a + art.	conduire	à, à + art.	to lead	to	führen	zu, nach
5	Recare su sé o indosso	Essere umano	Indumenti Effetti personali (all'accusativo)			porter		to wear		tragen	

So, these are the devices, based on the correlational theory of thought, I propose in order to improve the quality of machine translation. As we can see very well, the first of the two devices (dividing all the words of a language into two categories, the one of correlata and the one of correlators, and always putting, except in the rare aforesaid cases, the words of the correlators list in the upper boxes of the correlational triads, and the words of the correlata list in the lower boxes) is very simple. The second device (the notional sphere) is instead extremely complex and the human work required in order to actually build it increases exponentially when the number of the words of the program increases. In order to have an idea of how big this work is, let's calculate the one necessary to build the part of the notional sphere concerning the relationships between things (point b. and table 2), that is the one that is easiest to calculate. Every word of the dictionary has to be related with all others (that is all the words minus the word itself) and we have to say, between the two words of each of the couples so formed, if there is or not each of the kinds of relationship listed in table 2. Forming all possible couples and asking the question whether, between the two elements of each couple, there is or not each of the kinds of relationship listed can be done quickly by a computer. But only a human being can answer these questions. The number of these questions is given by the following formula:

$$n_q = (n_w{}^2 - n_w) \times n_{rel}$$

where n_q is the number of the questions, n_w is the number of the words of the program dictionary and n_{rel} is the number of the relationship kinds listed. For example, for a number of program dictionary words equal to 1000 (that is a rather small dictionary) the number of the questions is:

$$n_q = (1000^2 - 1000) \times 55 = 54.945.000$$

while for a dictionary of only 50 words (probably the minimum for a first level experimentation) the number of the questions is:

$$n_q = (50^2 - 50) \times 55 = 134.750$$

Even if the number of the questions is high, we have to consider that the time a human being needs to answer them "yes" or "no" is very brief, less than two seconds on average (whole groups of relationship kinds can be excluded at first sight, for example the ones including an activity if the two things are not activities). Therefore the working hours required in the first case are 30.525 (3.816 working days). Instead, in the second case the time is about 75 hours (about 9 working days), therefore rather brief (see the conclusions of this article for some considerations about the amount of human work necessary to make first level experimentations of this program).
It is important to notice that this part of the notional sphere, once it has been built, is made up of a lot less items than the questions we need to answer in order to build it. In fact, if we assume, for the sake of simplicity, that between two things there is only one relationship or no relationship (actually, as we can see in picture 1, the relationships can also be more than one, sometimes two and, rarely, more than two), so to calculate the minimum theoretical number (which anyway is not very far from the actual number, as we can easily understand), the number of the items that make up this part of the notional sphere is the same as the couples of words we can form combining each word of the program dictionary with all others, that is:

$$(n_w{}^2 - n_w)$$

Each element is made up simply by the two words of the couple and a number beside them: 0 (no relationship) or the code number of one of the relationship kinds listed in table 2.
Let's look at the phases of the procedure the program follows in order to put into practice the two aforesaid devices.

The phases of the procedure

1) The program takes into consideration only one sentence at a time. This is because the sentence, that is the group of words included between two full stops, is a closed correlational network. Therefore the program tries to reconstruct the correct correlational network for each sentence, trying to correctly dispose the words making it up.
The program generates all the possible correlational structures that, according to the correlational theory of thought, can correspond to the sentence under examination, from a theoretical point of view, that is taking into account only the distinction between correlata and correlators and not considering the (possible or impossible) relationships among them. We have to notice that the number of these combinations is a finite number and that it is not particularly high. It can be mathematically calculated starting from the number of the words making up the sentence and the number of the explicit correlators present in it.
How to calculate this number and how the program generates these combinations are problems that cannot be tackled in an article such as this, which is only a general presentation. We can notice however that a lot of combinations are not even generated because they are incompatible with the word order of the text. For example, in a sentence like "there are a bottle of wine and a tray with some cakes" the rules that regulate the word order in Italian (and probably in any other language) exclude that "bottle" is correlated with "cakes" or "wine" is correlated with "tray".

2) The program, as other machine translation programs, is provided with a dictionary having also all the flexed forms of every entry: therefore, it is made up by words and not by entries. For every single word a series of information is recorded. The first and fundamental information is if the word is a correlator or a correlatum (in the program model proposed by Ceccato and collaborators also the possibility that a word that usually designates a correlator is instead a correlatum was provided for, but these cases, as we saw, are so rare that, in my opinion, in the first phases of the program implementation it is not convenient to provide for them). Since correlators can occupy only the upper boxes of the graphic elements we use in order to symbolise the correlations, while correlata can occupy only the lower boxes, this fundamental classification drastically reduces the combination possibilities among the words of the sentence.
The Italian correlators list made by Ceccato and collaborators (table 4, words in italics[16]) was long, probably complete (it comprehended 82 items, almost all prepositions and conjunctions, and a very few words grammar considers adverbs), because it comprehended also rarely used or obsolete Italian correlators, as for example *acciocché, allorché, allorquando* etc. Most of these terms have the same mean-

[16] For the reader who does not know Italian, I have added, beside each Italian correlator, between brackets, the English words or expressions used in order to translate it. We must notice that in many cases not a single English correlator corresponds to a single Italian correlator. This is because English in some cases makes distinctions or has meaning nuances that are different or absent in Italian; or, more frequently, because in some cases English uses different correlators. I have highlighted with bold type the English word (or expression) that can be considered the homologue of the Italian correlator in question (sometimes it has been necessary to indicate more than one word). The other words are the other ways the Italian correlator can or, in some cases, must be translated (because in those cases English uses another correlator). The fact that a correlator must be translated in different ways, depending on the cases, when we translate from a language into another, is, for MT, an important problem, which we shall further talk about.

ing as much more commonly used terms (for example: *allorquando* has the same meaning as *quando* ["when"]), so that probably we had better make the program first substitute the rare or obsolete term with the common one.

Table 4

1) *a* (**to**; **at**; in, on)
2) *acciocché* (**so that**)
3) *affinché* (**so that**)
4) *allorché* (**when**)
5) *allorquando* (**when**)
6) *ancorché* (**even if**, though)
7) *anzi* (**on the contrary**; **even better**; **rather**)
8) *anziché* (**rather than**, instead of)
9) *appena* (**as soon as**)
10) *appresso* (**close**; **near to**; **behind**)
11) *attraverso* (**through**; **across**)
12) *benché* (**although**, though, however)
13) *che* (**that**, no English word required, than, when, since, for, only, but)
14) *circa* (**about**, as regard, concerning)
15) *come* (**as**, as soon as)
16) *con* (**with,** by)
17) *contro* (**against**)
18) *cosicché* (**so**, so that)
19) *da* (**from**, away (from), to, at, through, by, for, with, since, (as) from, like, as, adjectival forms)
20) *dacché* (**since**)
21) *dentro* (**inside**, in, within, into)
22) *di* (**of**, possessive case, attributive adjective, adjectival or adverbial form, at, in, by, than, with, for, from)
23) *dietro* (**behind**, after)
24) *dopo* (**after**, past, since)
25) *durante* (**during**)
26) *e* (**and**, but)
27) *eccetto* (**except**, excepting, but, save)
28) *entro* (**within**, in, by, before)
29) *finché* (**till**, until, as long as)
30) *fino* (**till**, until, up to, as far as, to)
31) *fintantoché* (= *finché*)
32) *fuorché* (**except**, but, other than, apart from)
33) *fuori* (**out** of, outside)
34) *giacché* (**since**, as, now that)
35) *in* (**in**, at, inside, on, to, into, through, across, adjectival form, by)
36) *innanzi* (**before**, in front of)
37) *lungo* (**along**, during, over)
38) *ma* (**but**, still, yet, no English word required)
39) *malgrado* (**in spite of**, notwithstanding, with all, for all)
40) *mediante* (**by**, by means of, through)
41) *meno* (**except**, but)
42) *mentre* (**while**, whereas)
43) *né* (**neither**, nor, either, or)
44) *nonché* (**let alone**, still less, as well as)
45) *nonostante* (**in spite off**, despite, for all)
46) *o* (**or**, either… or…, whether… or…)
47) *oltre* (**beyond**, over, more than)
48) *onde* (**so that**, in order that)
49) *oppure* (**or**, or on the other hand, or alternatively; **or else**, otherwise)
50) *ovvero* (**that is**, in other words, or, or rather)
51) *per* (**for**, in order to, to)
52) *perché* (**because**; **so that**, in order that, so as)
53) *più* (**plus**)
54) *poiché* (**since**, as, for)
55) *presso* (**near**, **beside**, by, next to, with, in, at, among)
56) *però* (**yet**, but; **nevertheless**, though, however)
57) *purché* (**provided**, as long as; **if only**)
58) *qualora* (**if**)
59) *quando* (**when**, whenever, while, since, if)
60) *quanto* (**what**)
61) *quantunque* (**although**; **even if**)
62) *quasi* (**as if**)
63) *rasente* (no English homologue[1])
64) *salvo* (**except**, but, save, barring; apart)
65) *se* (**if**, whether, if only)
66) *sebbene* (**though**)

67) secondo (**according to**, in accordance with, depending on)
68) senza (**without**, -less, un-, in-)
69) seppure (**even if**)
70) sicché (**so**, so that)
71) siccome (**as**, since, because)
72) sopra (**on**, up, on to, onto; **over**; above; after)
73) sotto (**under**, beneath, underneath, below)
74) su (**on**, up, on to, onto; **over**; above; after; about; out of)
75) talché (= *cosicché*)
76) tra (**between**, **among**, **amid**, **within**, in)
77) tramite (**through**)
78) tranne (**except**, but, save)
79) verso (**toward**(s); near, about, to)
80) via[2] (**through**, by means of)
81) (relativo) (relative)
82) *(operazione matematica)* (mathematical operation)

[1] It is a preposition with the same meaning as the verbs "to graze", "to skim", generally translated with such verbs or another motion verb plus "close to".
[2] The Italian noun "*via*" [way] can be used as a preposition with the meaning of "through", "by means of" (for example: "*Lo so via mio fratello*" [I know it through my bother]).

Besides these correlators, the list comprehended another series of correlation kinds where the correlator is always the same, that is the implicit correlator, and what changes is the grammatical category of the correlata. That is, these correlations were listed:

Table 5

1) subject-verb
2) verb-object
3) article-……
4) adjective-……
5) …… -adjective
6) adverb-……
7) predicate-……
8) ……-predicate
9) auxiliary verb-past participle
10) auxiliary verb-present gerund
11) auxiliary verb-infinitive mood
12) "si"-verb (so called "intransitive pronominal conjugation" of verb, present in Italian but not in English)
13) middle form-verb (verbal form, absent in English, expressing actions done in advantage of the subject)
14) "ci, vi"-verb (forms of the Italian reflexive verb)
15) ……-verb in the subjunctive mood
16) ……-verb in the conditional mood
17) ……-verb in the participle mood
18) verb in the gerund mood
19) verb in the subjunctive mood-verb in the conditional mood
20) verb in the subjunctive mood-……
21) verb in the conditional mood-……
22) verb in the participle mood-……
23) verb in the gerund mood-……
24) verb in the conditional mood-verb in the subjunctive mood
25) ……-genitive
26) verb-dative
27) verb-separation (absent in English)
28) ……-relationship
29) ……-…… (repetition)
30) common noun-common noun
31) common noun-proper noun
32) proper noun-common noun
33) proper noun-proper noun

Even if the correlator of all these kinds of correlation is always the same, it is necessary to distinguish all kinds of correlations it can originate according to the grammatical nature of the two correlata. It is necessary to do this for two reasons.

– The implicit correlator is by far the most used of all correlators and in most sentences the majority of correlators are implicit correlators. Therefore, distinguishing the correlations it originates according to the grammatical nature of what the implicit correlator ties is fundamental for the reconstruction of the correlational network of thought.

– This classification is fundamental for the subsequent translation work (for example, if the correlation is "subject-verb" the first correlatum, in languages having cases, will have to be flexed to the nominative, if the correlation is "verb-object" the second correlatum will have to be flexed to the accusative etc.).

The words that can be only correlata, as we said, can occupy only the lower boxes of the correlational triad. A long and patient work has to be done on them.

First, each single word of source language (L1) has to be classified from the grammatical point of view. That is, it is necessary to specify which part of speech it can be. We have to notice that the possibilities are more than the ones that are listed in the common dictionaries for the single entries. In fact, the English word "bears", for example, can be the third person of the indicative present tense of the verb "to bear", but it can also be the plural of the noun "bear"; the Italian word *amo*, according to the dictionary, can only be a noun ("fishhook"), but in the program dictionary, where there are also the forms deriving from flexion, it can also be the first person of the indicative present tense of the verb *amare* ("to love"); etc. That is, flexion can lead to the formation of two or more words having identical spelling but completely different meaning. Moreover, there is the fact, which is much more important, that when source language is a slightly inflectional language while target language is a highly inflectional language, it frequently happens that in the target language a lot of different words correspond to the same word of the source language (above we saw how many can be the Italian outputs for words like the English words "love" and "loved"). Therefore we shall have a list of possible meanings (M1, M2, M3 etc.) for each word.

Each of these possible meanings, in its turn, has to be classified from the grammatical point of view. For example, the English word "bears" will have the following two possible meanings, which will be classified like this:

- **BEARS**
  - M1: "noun"
    - number: plural
  - M2: "verb"
    - mood: indicative
    - tense: present
    - person: 3rd

For each of the possible meanings of every single word it is then necessary to indicate which are its possibilities to relate with each correlator. That is, it is necessary to specify if each meaning can be or cannot be the correlatum of each of the aforesaid correlators and, in the case it can be, if it can be the first or the second correlatum or both. For example, the meaning M1 of the English word "bear" (the animal, a noun from the grammatical point of view) can be, for example, the first correlatum in the correlation "subject-verb" ("the bear sleeps"), but it cannot be, for example, the first correlatum of the correlation ".....-adverb", nor it can be a correlatum of the many correlation kinds where the correlator is the implicit correlator we listed in table 4 (see the table). In regard to the other correlators, the word "bear" as a noun can be (with the article) the correlatum of the prepositions "of", "at", "for", "by" etc. ("the fur of the bear", "look at the bear", "food for the bear", "a branch broken by the bear" etc.), but it

cannot be the correlatum of a conjunction like "when" which always correlates two sentences; it cannot be the first correlatum of the conjunction "during" because it must be a sentence, nor its second correlatum because it must be the name of something that has a duration; etc. (if the reader will try to match some words with the correlators we listed in table 4 he will realise that the incompatibilities are many).
The meaning M2 of the English word "bears" (the part of the verb "to bear") can practically be only the second correlatum of the correlation "subject-verb" ("he bears") among the 34 kinds of correlation we listed in table 4, because in English a personal form of the verb requires an explicit subject (differently from Italian, for example, where the word corresponding to "bears", *porta*, can also be the first correlatum of the correlation "verb-object" ("*porta pacchi*", "he bears parcels").
As we can see very well, indicating what possibilities each of the possible meanings of each word has of relating to each correlator excludes many of the correlational triads which can theoretically form combining correlators and correlata of a sentence in all the theoretically possible ways.

3) At this point the program examines the theoretically possible correlational networks which have remained, taking into account two series of rules of the source language grammar, that is:

a) the gender, number, person, case etc. concordance rules which, as we said, are many in inflectional languages. In Italian, for example, it is impossible that a feminine adjective is referred to a noun of masculine gender, the personal verb always accords in the person with the pronoun it is correlated to etc.;

b) the rules regarding the order that words must have, which, as we said, are very important in isolating languages (in English, for example, the direct object always follows the subject, the adjective generally precedes the noun which is referred to etc.).

This phase of the processing leads to a drastic reduction of the correlational networks theoretically possible. This is particularly true in highly inflectional languages such as Latin, Slavic languages etc. Probably in many cases, particularly in this kind of languages, at this point only one correlational network, the correct one, proves possible.

4) If it is not like this, the "notional sphere" comes into play. The notional sphere, as we pointed out, is composed of the two following parts.

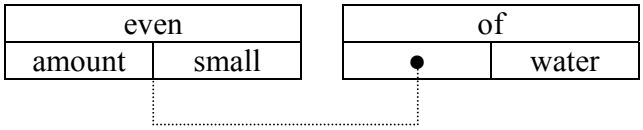
a) A list of classifications like the one in table 1. One or more of these classifications are attached to each of the words of the program dictionary. For example, the word "dog" will be classified in one of the following categories of table 1:

- living beings (1)
- animated beings (3)
- animals (6)
- earth animals (25)
- carnivorous animals (28)
- domestic animals (32)

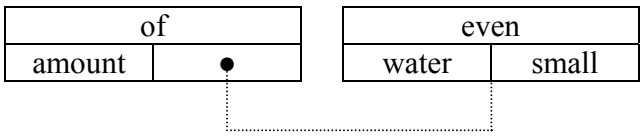
b) A network of relationships among words such as the one we partially quoted in picture 1, where a line uniting two words indicates that one or more relationships between their meanings, the ones indicated by the numbers along the line, have been found.

The correlational networks not excluded by the preceding phase are examined one by one taking into account the many classifications making up the two aforesaid parts of the notional sphere. All networks resulting incompatible with one or more of these classifications are eliminated. For example, in the aforesaid case of the two expressions "amount of water, even small" and "amount of water, even cold", the program, at the end of this phase, should have considered as theoretically possible only these two correlational networks for each expression:

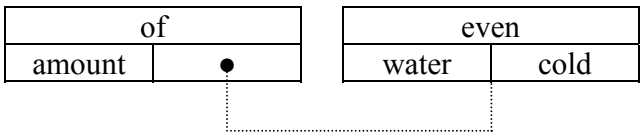
1a)



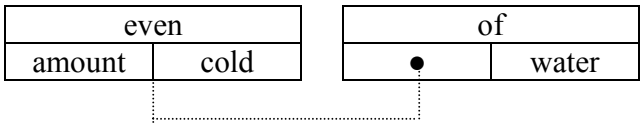
1b)



2a)



2b)



If in the notional sphere the word “amount” is tied to the word “small” by a relationship classified as “characteristic-characterised thing” (No. 6 in table 2), the correlational network 1b) will be excluded, hence only the correct correlational network will remain. If in the notional sphere the word “cold” is classified, beyond the rest, as “physical feature” while the word “amount” is classified as “abstract thing”, in the case of the sentence “amount of water, even cold “the correlational network 2b) will be excluded, hence only the correct correlational network will remain.

Another example of an expression where, after the processing carried out up to the end of the phase described in 4), two correlational networks are still possible is “red ball pen”. This is a more critical case, because the adjective “red” is perfectly compatible both with the noun “pen” and the noun “ball”. Therefore the list of classifications in table 1 does not exclude any of the two possibilities. But in this case a notional sphere, such as the one in picture 1, should indicate, if well done, that between “pen” and “ball” there are the relationships “whole-part” and “compound-component”. In this case I propose the program to proceed not by elimination but by probability. Because of the fact that the correlation between “red” and “ball” is only possible while the correlation of “pen” with “ball” is supported by a good two kinds of relationship of the notional sphere in table 2, the program will choice the latter possibility.

Naturally, the more rich and accurate the group of the classifications making up the notional sphere, the greater the probability that the program is able to consider only one correlational network possible for each sentence and that this network is the correct one. In the practical implementation of the program which I am talking about the problem is just this: the human work necessary to build the notional sphere increases in an exponential way when the number of the words of the program dictionary increases. But, as we can see very well, this is only a practical problem and not a theoretical one.

There could be another problem, we might think. The expressions or the sentences we have quoted up till now as examples of translated text or text which has to be translated are all brief and without punctuation marks. Reconstructing the correlational networks corresponding to them is very easy for a hu-

man being. But the sentences making up a text often are not so brief. They can also be very long and complex, depending on the style of the writer and, partially, also on the language used. We can ask ourselves if also in these cases the correlational theory is valid, that is if for each sentence we can reconstruct a correlational network of the same kind of the ones we have seen, even if much more complex. Basing myself on both my own experience and that of the S.O.I. up to now, the answer is affirmative. In some cases we can have some doubts, some perplexities, which it is probably best not to discuss in a very general article like this, but they are solvable problems. In tackling long sentences, at the beginning we may have some problems, but with a little experience the task is not difficult. This is true for the human being, but will it be true for the program too? It is impossible, or at least extremely difficult, to give this question a sure answer *a priori*. It is necessary to see how the program behaves in practice. Obviously, the first experiments have to be done with very small dictionaries and very simple sentences. When the length of the sentences increases, the amount of the necessary processing increases, and the increase is very big. But surely this is not a problem, given the huge power of modern processors. Nevertheless, when the sentence is made up of two clauses or more (period), a main clause with other co-ordinate and/or subordinate clauses tied to it, complex networks of related words are a correlatum of a correlational triad. It is possible that this creates problems for the program. As far I know, this problem was not seriously tackled by Ceccato and collaborators in the few years of their research program (from1959 to 1966). I have not done deep research into this problem, but at first sight it seems to me that it can be tackled in two ways.

1) The first way is to consider punctuation marks. We have seen that the full stop closes a correlational network. It is the same for the question mark and the exclamation mark. We can ask ourselves what the comma, the semicolon and the colon designate. In grammar books we find definitions of this kind:

"punctuation marks [...] are used to indicate, when we are writing, the more or less long pauses we all make when we are speaking, and to divide the clauses of a period, and also the parts of the same clause" [Bruno Migliorini, Grammatica Italiana].

Such an assertion is not wrong, but it is surely not satisfying. We have to understand what these divisions are, and why they are put in certain positions and not in others. I have still not done deep research into this problem, but basing myself on the first achieved results I can make some considerations. Let's see what the main uses of the comma seem to be, giving some examples. We have yet to state beforehand that the use of the comma is rather free, that is in several cases the comma may or may not be used.

a) The comma divides the main clause from the co-ordinate clause or the subordinate clause in some cases, but not in others:

- He could not swim, but he saved himself anyway.
- He has a car, even better two.
- I think, therefore I am.
- He told me he had gone.
- He had gone, he told me.
- I don't know how he did.
- How he did, I don't know.
- When you will arrive, you will find John in the station.
- You will find John in the station when you will arrive.
- While you were sleeping, they have phoned twice.
- They have phoned twice while you were sleeping.
- If you will come, I shall be very pleased to give you hospitality.
- I shall be very pleased to give you hospitality, if you will come.
- I don't know if I will have to start.

b) We put between two commas appositions, if they follow the substantive they refer to; sometimes we put them and sometimes we do not when they come before it.

- Christopher Columbus, the discoverer of America, was born in Genoa.
- The discoverer of America, Christopher Columbus, was born in Genoa.
- Homer, the great Greek poet, was blind.
- The great Greek poet Homer was blind.

c) We put between two commas the incidental clauses and expressions:

- John, in my opinion, was the best.
- Doctor White, they say, is very rich.
- The device, invented by engineer Brown, is very good.

while we do not put the commas if we want the expression or the clause not to have incidental value:

- The device invented by engineer Brown is very good.

d) We put the comma after vocatives:

- Dear friends, I say goodbye to you.

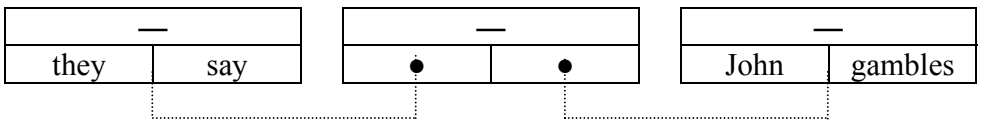
e) We put commas after the items of a list, except the last, generally, but sometimes also after it:

- I bought pears, apples and grapes.
- Artists, intellectuals, scientists joined the demonstration.

f) We put the comma when we might think that two adjacent words are related but it is not so:

- He left, the work not being ready.
- Device in order to improve the quality of machine translation, based on the correlational theory of thought.

Basing ourselves on these examples, it seems that the comma indicates, in most cases, that the words between two commas (or between the comma and the full stop before it or after it) make up together a "section" of the correlational network of the sentence, that is a part of it which is tied *as a whole* to the rest of the sentence. This part can be a correlatum of a word (or group of words) in the rest of the sentence, as in the case above in a), but it also can be apparently more independent, as in the case of the incidental clauses and expressions. In this case we could even think that a single sentence is made up of two independent correlational networks, but I think that, even if from a grammatical point of view the main clause and the incidental clause sometimes could exist also independently, at the level of thought there is an indubitable relationship between the two clauses. That is a sentence such as "John, they say, gambles" is the same as "They say that John gambles". Therefore in the case of the incidental expressions and clauses I should say that the correlational network of thought is the following:



Therefore we can say that in most cases the comma excludes that the word immediately before it and the word immediately after it are directly related. The only exception seems to be the case in f), when

the comma divides the items of a list. In this case I should say that the comma has the value of a correlator, that is the conjunction “and”, which otherwise will be repeated with an unpleasant effect.
Hence the commas divide the sentences into relatively independent parts, which can be considered by the program as sentences apart, that is word sequences of which the program has to reconstruct the correlational structure without taking into account the words making up the remaining sentence. In this way we achieve a very high reduction of the processing necessary in order to reconstruct the correlational network of long and complex sentences. In the case of pre-editing of source text before machine translation, an abundant use of commas, as long as correct, would therefore be recommendable. This way of proceeding in several cases can still cause problems, particularly when the source language is a language with little inflection, while the target language is a highly inflectional one. Let’s consider the sentence I have just written “ an abundant use of commas, as long as correct, would therefore be recommendable”. In this case, in translation for example from English into Italian, the program, which considers the expression corresponding to “as long as correct “ as an expression apart, would not know if the word “correct” has to be translated using the corresponding Italian word inflected to the masculine or the feminine genre. Therefore, in the case there are such ambiguities, the program should look, in the usual way, for the word or the correlation with which the group of words between two commas (or between the comma and a full stop) is correlated. This does not mean that commas are not a way of very much reducing the processing necessary to translate a sentence.
The particular case of lists is a problem that still has to be solved. In fact, in the case of a sentence like "I have bought apples, pears and grapes", if the program considers "pears and grapes" an independent expression this can be acceptable in the case of a language without cases, but in the case of a language which owns them the program would not know which case to use for the two words which have to translate "pears" and "grapes". Yet it should not be difficult to make the program recognise the lists because they are always made up by sequences of words (or expressions) that are divided by one or more commas and a conjunction "and" at the end. Besides, these words, or the first words of the expressions, belong to the same grammar category, that is they are either all nouns or all adjectives etc.
As regards other punctuation marks, that is the semicolon, the colon, the brackets, the dashes, the quotation marks, what we said about the comma is even more valid, because they are stronger interruptions than the comma.

2) The second way to reduce the processing necessary to reconstruct the correlational network of sentences made up by more than one clause, is to take into account the fact that the clauses coordinated or subordinated to the main clause are introduced by specific words. In the case of the coordinate clauses we speak of coordinating conjunctions (“and”, “or”, “neither”, “but”, “therefore”, “besides”, “instead” etc.); in the case of the subordinated clauses we speak of subordinating conjunctions (“that”, “as”, “when”, “while”, “if”, “though”, “until”, “since”, “because” etc.). A lot of these conjunctions, must have, or at least very often have, as correlata not singles words but whole clauses. Therefore these conjunctions can be used as markers of the beginning of clauses that in a first moment can be considered independent, that is of words sequences which have be correlated among themselves so as to form a correlation network without taking into account the rest. Only in a second moment shall we look for the word (or word group), preceding the conjunction (or, more rarely, following it), which the conjunction correlates to this network.
As we can see, using these two methods also long and complex periods reduce to a group of clauses which are decidedly more brief, and for which therefore it is more easier for the program to rebuild the correlation network. Only in a second moment, if necessary, these clauses will be correlated amongst each other.

Once the program has rebuilt the correlation network, we pass to the phase of real translation. If the correlation network has been correctly rebuild, the program cannot make translation mistakes derived from not establishing the word or word group every single word is referred to, because, as we said, the

rules which govern the passage from the correlational network of thought to its linguistic expression are very precise in all languages.
Also a lot of the translation mistakes, deriving from the fact that some dictionary entries give origin, because of the flexion, to identical words, but having completely different meaning, will be avoided. For example, if the program rebuilds the correlational network of thought corresponding to a sentence like “The expression fits and sounds clear” we can completely exclude that the word “sounds” will be considered a noun (the plural of “sound”) producing, for example, an Italian translation such as “L’espressione va bene e i suoni si schiariscono” (“The expression fits and the sounds clear up”).
Yet the problem of the homonymous words not deriving from flexion and having identical grammatical characteristics is still present. For example, the English word “bank”, as we said above, can mean the limit of a river or channel or a financial institution, still being a noun. Also in these cases the reconstruction of the correlation network of thought is extremely useful. In the luckiest cases, when such a word is in a direct correlation with something else which clearly indicates what the meaning at stake is, as happens for example in the correlation “river bank”, the notional sphere of the program, including the classification of the relationship between the two objects as “contiguity”, automatically will ensure the correct choice is made. For the other cases we could provide the program with a function of context “exploration”. That is, when there is one of these critical words, like “bank” in a sentence, the program, after having rebuilt the relevant correlational network, could examine the words making it up (and possibly also the words making up the immediately preceding or following sentences) looking for words indicating if one meaning of the word at stake is more probable than the other (for example, if in a sentence where there is the word “bank” there are also the words “money”, “cheque”, “clerk”, “robbery” etc. the meaning will very probably be the financial one. Naturally, for this function of context exploration we should use the classifications of the notional sphere we listed in table 2.
Yet another problem MT has always had is the problem of the translation of prepositions[17], therefore of correlators. As I pointed out in note 16), the cases where a certain correlator of source language cannot be translated always with the same correlator of target language, but has to be translated in different ways depending on context, are frequent. A complex notional sphere such as the one we described, containing the large number of classifications we listed in table 2, is able to solve, at least largely, this problem. In fact, let's suppose, for example, we have to translate from Italian into English the two correlations "*profumo di rosa*" (scent of rose) and "*quadro di Raffaello*" (picture of Raphael). In Italian both these correlations have as correlator the preposition "di", but in English in the second case we cannot use the preposition "of", but we must use "by". In the notional sphere, between the words "*profumo*" (scent) and "*rosa*" (rose) surely there will be the relationship n° 6 of table 2 (characteristic-characterised thing) and between the words "*quadro*" (picture) and "*Raffaello*" (Raphael) surely there will be the relationship n° 7 (product-producer). Therefore it will not be difficult to make the program translate the Italian preposition "*di*" with the English preposition "of" in the first case and the preposition "by" in the second case.
The last phase of processing is the insertion or removal or substitution of some elements, following the grammatical rules of the target language (for example, insertion or removal or substitution of articles depending on the two languages having them or not or using them in a different way; insertion of personal pronouns which in languages such as Italian can be omitted while in languages such as English they must be expressed; etc.). Also in this phase, naturally, the previous reconstruction of the correlational network of thought would be very useful.

[17] Hutchins W. J., ‘*Retrospect and prospect in computer-based translation’*, p. 2. In Machine Translation Summit VII, 13th-17th September 1999, Kent Ridge Digital Labs, Singapore. Proceedings of MT Summit VII "MT in the great translation era", 30-34. [Tokyo]: Asia-Pacific Association for Machine Translation, 1999.

We can provide for further phases too, for example identifying idiomatic expressions, which are easier to identify and to put in the correct position in their translated version when their correlational structure has been identified.

Conclusions

Therefore it seems to me that we can affirm, as a conclusion to this article, that the implementation of an MT program based on the reconstruction of the correlational network of thought, even if it needs a huge amount of human work, exponentially increasing when the number of the words of the program dictionary increases, should allow a considerable improvement in MT quality. The fundamental reason for this is that the reconstruction, by means of the described devices, of the correlation network of thought *simulates, even if does not reproduce, a part of what the human being actually does by using his intelligence when he translates*. In fact, reconstructing the correct correlational network of thought corresponding to a text means doing one of the two fundamental operations that, as we saw, make up the understanding of a text. If the program does this before translating, the serious errors, which sometimes make the output text incomprehensible, derived from the fact that the program does not "understand" which words (or group of words) a certain word is referred to, are completely eliminated. The only serious errors that can remain are fundamentally the ones deriving from the fact that the program did not choose the right meaning in the case of the words having more than one meaning. But a program such as the one we described should be able, as we saw, to solve also this problem, at least largely. Furthermore, the latter problem is less important because, if a person reading the output text understands that a certain word has nothing to do with the general sense of the sentence, he always can, if the program allows the access to its bilingual dictionary, look for the other possible meanings of that word. As an alternative, the program itself can indicate, putting them for example between square brackets, which are the possible meanings of the main homonyms. MT presents other problems, as, for example, the fact that sometimes two languages use different constructions in order to express the same concept (as for example in the case of the English expression "John is 20 years old" which has to be translated into Italian as "*Giovanni ha 20 anni*"), that is literally "John has 20 years", the fact that there are differences in style etc. But these are less important problems. Generally, as happens in the example we just gave, the human being reading the output text understands the sense of it even if there are errors derived from the fact that the program is not able to solve problems of this kind. What is really important, in my opinion, is to achieve an output text which is *understandable and without distortions*, so that the user who does not understand a certain language at all or understands it very little can understand a text written in that language[18].

Summarising, the peculiarities that make the device in order to implement a MT program, we described, different from the others, are the following.

1) The device is based on the correlational theory of thought, which is a radically new and deeply different linguistic theory from the other linguistic theories.

2) Basing itself on this theory, the program simulates, even if does not reproduce, a part of what the human being actually does by using his intelligence when he translates.

3) The device is probably the only one completely and exclusively based on a single linguistic theory.

[18] The use of MT programs in order to make translation quicker when a professional translator is available is, in my opinion, less interesting than in the past, because the excellent results achieved by speech recognition programs allow a professional translator to write at the speed he speaks (the errors are very few, so their correction is very quick). In this way it is possible to get the translation of a text in a time that is not very different from the one is needed to make a professional translator correct the MT of the same text.

4) The device is probably the only one conceived by the person who proposed the linguistic theory which the device itself is based on.

5) The number of ways in which the words making up a sentence can combine according to the correlational theory of thought is a finite number and it is not particularly high (it can be mathematically calculated starting from the number of the words making up the sentence and the number of the explicit correlators present in it). The program generates all the possible combinations (in order to examine them later), therefore also the right combination is surely produced.

If the program, at the end of the processing, is not able to choose between two (or more) theoretically possible structures, it can signal this fact to the user so that he can be the one who makes this choice (very probably, the case which makes this choice impossible, that is that, between two (or more) theoretically possible structures, there are more than one which make sense, rarely happens).

If the effective implementation of a machine translation program able to rebuild the correlational network of thought and having a wide enough dictionary needs a huge amount of human work, a first level experimentation in order to verify its efficacy is, on the contrary, rather simple. In fact it is possible to proceed in the following way.

1) At the beginning we can provide the program with a very small dictionary, a dictionary of a few dozens words. Also with so few words the number of sentences we can build is big and it is possible to verify if the program does not make the kind of errors that instead are common with programs that are currently available. We can implement only the generation module, that is the part of the program that starts from the rebuilt correlational network to generate the text in the target language. Implementing this part of the program instead of the one which rebuilds the correlational network is hugely easier and it does not depend on how wide a dictionary we use.

2) It is also possible to verify the efficacy of the program "on paper", that is before having effectively implemented the program. When we have established what classifications to use in order to build the notional sphere we spoke about and they have been applied even to a few dozens words from the dictionary, it is possible to verify, so to say "manually", how the program will behave in the translation of a certain sentence.

Select Bibliography

Benedetti G. (2004) *A presentation of Operational Methodology*, in www.mind-consciousness-language.com by G. Marchetti.

Benedetti G. (2004) *Basic mental operations which make up mental categories*, in www.mind-consciousness-language.com by G. Marchetti.

Ceccato S. (edited by) (1969) *Corso di linguistica operativa*, Milan, Longanesi.

Ceccato, S. Zonta B. (1980) *Linguaggio consapevolezza pensiero*, Milan, Feltrinelli.

Glasersfeld, E. von, Pisani, P.P. (1970), *The multistore parser for hierarchical syntactic structures*. Communications of the ACM 13(2), 74-82.

Hutchins W. J., (1986) *Machine translation: past, present, future*, (Ellis Horwood Series in Computers and their Applications.) Chichester, Ellis Horwood.

Hutchins W. J. (with Somers H. L.), (1992) *An introduction to machine translation*, Academic Press, London.

Hutchins W. J., '*Retrospect and prospect in computer-based translation*'. In Machine Translation Summit VII, 13th-17th September 1999, Kent Ridge Digital Labs, Singapore. Proceedings of MT Summit VII "MT in the great translation era", 30-34. [Tokyo]: Asia-Pacific Association for Machine Translation, 1999.

Hutchins W. J., '*Towards a new vision for MT*'. Introductory speech at the 'MT Summit VIII' conference, 18-22 September 2001, Santiago de Compostela, Galicia, Spain.

Hutchins W. J., '*Machine translation and human translation: in competition or in complementation?*'. *International Journal of Translation*, vol.13, no.1-2, Jan-Dec 2001, pp. 5-20. Special theme issue on machine translation, [editor] Michael S. Blekhman. Also in: *Machine translation theory & practice*, edited by Michael S. Blekhman. New Delhi: Bahri Publications, 2001. (BP Series in Translation Studies, 8).

Hutchins W. J., '*Machine translation today and tomorrow'*. In Computerlinguistik: was geht, was kommt? Festschrift für Winfried Lenders, hrsg. Gerd Willée, Bernhard Schröder, Hans-Christian Schmitz. Sankt Augustin: Gardez! Verlag, 2002, pp.159-162.

Hutchins W. J., '*Has machine translation improved*?'. MT Summit IX: proceedings of the Ninth Machine Translation Summit, New Orleans, USA, September 23-27, 2003, 181-188. [East Stroudsburg, PA: AMTA.]