UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

A Critical Analysis of the Performance of English - Portuguese - English MT Systems

Osvaldo Novais de Oliveira Jr
Ana Raquel Marchi
Mônica Saddy Martins
Ronaldo Teixeira Martins
Maria das Graças Volpe Nunes

Nº 47

NOTAS

São Carlos - SP

UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação
ISSN 0103-2577

A Critical Analysis of the Performance of English - Portuguese - English MT Systems

Osvaldo Novais de Oliveira Jr
Ana Raquel Marchi
Mônica Saddy Martins
Ronaldo Teixeira Martins
Maria das Graças Volpe Nunes

Nº 47

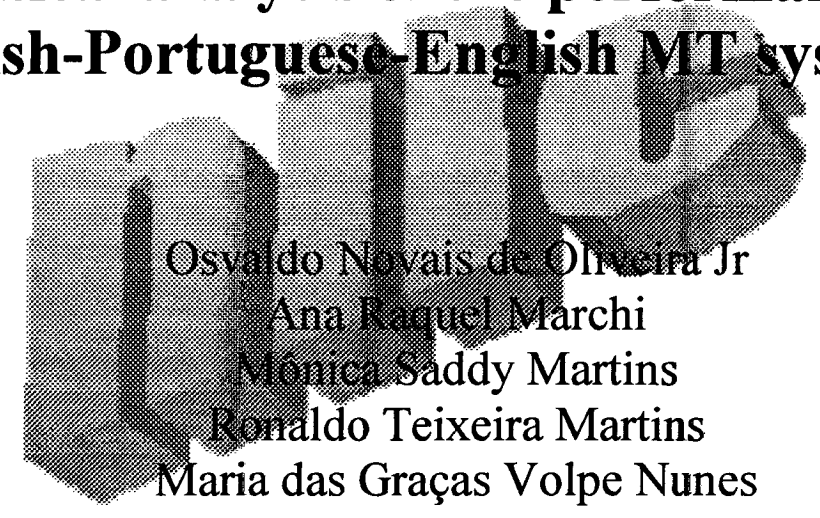
NOTAS

Série Computação

São Carlos – SP
Jun./2000

Universidade de São Paulo - USP
Universidade Federal de São Carlos - UFSCar
Universidade Estadual Paulista - UNESP

A critical analysis of the performance of English-Portuguese-English MT systems



Osvaldo Novais de Oliveira Jr
Ana Raquel Marchi
Mônica Saddy Martins
Ronaldo Teixeira Martins
Maria das Graças Volpe Nunes

NILC-TR-00-6

Junho, 2000

Série de Relatórios do Núcleo Interinstitucional de Lingüística Computacional
NILC - ICMC/USP - São Carlos, Caixa Postal 668, 13560-970 São Carlos, SP

A critical analysis of the performance of English-Portuguese-English MT systems

Osvaldo Novais de Oliveira Jr.
chu@ifsc.sc.usp.br

Ana Raquel Marchi
raquel@nilc.icmsc.sc.usp.br

Mônica Saddy Martins
monica@nilc.icmsc.sc.usp.br

Ronaldo Teixeira Martins
ronaldo@nilc.icmc.sc.usp.br

Maria das Graças Volpe Nunes
mdgvnune@icmc.sc.usp.br

Núcleo Interinstitucional de Lingüística Computacional (NILC)
http://nilc.icmc.sc.usp.br
CP 668 - ICMC-USP, 13560-970 São Carlos, SP, Brazil

Abstract

The aim of this paper is to report conclusions derived from a comparative analysis of six machine translation systems available for Portuguese/English and English/Portuguese conversion. Twenty excerpts of American and Brazilian journalistic texts were used to identify the problems related to the three proposed levels of analysis: lexical, syntactic, and semantic-pragmatic. As a conclusion, we confirmed that the output quality is not good enough yet, due mainly to three basic reasons: lack of linguistic resources in the analyzed tools; mistaken hypotheses on the functioning of the English and Portuguese languages; and the complexity inherent in the translation task itself. Except for the latter reason, we observed that the obstacles are not insurmountable at all, they could be easily overcome if there was a change in perspective when considering the idiosyncrasies of each language.

Resumo

Este artigo reporta as conclusões derivadas de uma análise comparativa de seis sistemas de tradução Português/Inglês e Inglês/Português. Vinte trechos de textos de jornais americanos e brasileiros foram usados para identificar os problemas relativos aos três níveis de análise aqui propostos: lexical, sintático e semântico-pragmático. Confirmou-se que a qualidade das conversões ainda não é boa, devido principalmente a três razões básicas: falta de recursos lingüísticos nas ferramentas analisadas; hipóteses errôneas sobre o funcionamento das línguas inglesa e portuguesa; e a complexidade inerente da tarefa de tradução propriamente dita. Exceto pela última razão, observou-se que os obstáculos não são instransponíveis, e que eles podem ser facilmente transpostos se houver uma mudança de perspectiva quando são consideradas idiossincrasias de cada uma das línguas.

Introduction

There is a widespread perception in the computational linguistics community of the low quality of the output from MT systems for Portuguese. However, there has been no systematic study in which the limitations of existing systems are scrutinized and possible ways of improving them are put forward. Given this scenario we have instituted a project aimed at evaluating the performance of MT English-Portuguese-English systems commercially available, within a computational linguistics perspective, considering not only limitations deriving from inadequate linguistic processing but also practical limitations inherent in a general-purpose natural language processing (NLP) tool.

In this paper we first analyze the performance of several systems in translating text passages extracted from newspapers from English into Portuguese and vice-versa. In addition to identifying the main limitations, we discuss the reasons why such limitations exist. From a linguistics perspective, we classify the reasons for malfunction in terms of automatic processing of lexical, syntactic, and semantic-pragmatic information.

1. Evaluation of machine translation tools

As in other areas of NLP, there are three types of evaluation generally accepted: adequacy evaluation, diagnostic evaluation, and performance evaluation[1]. Adequacy evaluation is generally used to assess MT systems adequacy to certain operational contexts. Diagnostic evaluation aims at identifying limitations, errors, and deficiencies of MT systems. Finally, performance evaluation is intended to test stages of system development. The latter two types are generally performed by researchers and developers, whereas the first is usually undergone by users and purchasers.

In this paper we report the assessment of some MT tools developed to English/Portuguese/English translation which was made according to the diagnostic evaluation, in order to achieve a broader view of the errors and deficiencies found in the outputs produced. As a benchmark for the analysis, we worked the intelligibility of non-edited (raw) translations, regardless of their accuracy, fidelity, and style appropriateness, as well as the MT systems usability of facilities to improve the outputs (creating and updating dictionaries, for instance). In this sense, our work intends to be close to the one developed by ARPA (White *et. al.* 1994), which compared the unedited output of the three experimental systems (Pangloss, Candide, and Lingstat) supported by ARPA with the output from other systems, namely: Globalink, PC Translator, Microtac, Pivot, PAHO, Metal, Socatra XLT, Systran, and Winger.

[1] Lehrberger & Bourbeau 1988; Hutchins & Sommers 1992; Arnold *et. al.* 1994.

2. Methodology for assessment

Six MT systems were employed in our study, viz. Translator Pro (from now on TP), Alta Vista (AV), Intertran (IT), GO Translator (GO), Tradunet (TN), and Enterprise Translator Server (ET). The first three provide both English/Portuguese and Portuguese/English translations, the fourth provides only Portuguese/English translation, and the latter two provide only English/Portuguese translation. Tradunet and Translator Pro are commercially available software packages, whereas the others are available on the Web.

In order to compare the tools mentioned, we randomly compiled a corpus of 20 text passages (with one or more sentences), of which 10 were extracted from the first supplement (editorials, national and international news, and reader's opinion sections) of a Brazilian daily newspaper (*Folha de S. Paulo*), and the other 10 from the corresponding sections of an American newspaper (*New York Times*). This procedure was intended to restrict the comparison to journalistic texts. However, if this restriction of corpus has the merit of allowing a more punctual comparison among the performance of various translation tools, it is necessary to recognize that several limitations derive from that, affecting the conclusions we expect to make from this paper: the number of sentences is flagrantly small to allow a real assessment of these translators' quality; and the restriction imposed on the type of texts surely implies that different results could be obtained for texts which belonged to other discourse genres. Nevertheless, as our objective here is not a hierarchy of translation tools according to their performance quality, but the discussion of the problems all of them are still tackling, we consider valid the given delimitation of data, since a significant set of problems is included in this discussion.

The outputs were analyzed in two ways, according to the source language sentence interpretation and to the target language sentence generation. Although locating the origin of problems in machine translation is not always possible, there is a pressing need to identify the difficulties either in the process these tools employ to analyze the original sentences and extract from them a meaning of interlinguistic (and even translinguistic) validity or in the process in which this meaning is materialized in the target language in a way it can be accessed by a new group of users. These problems of interpretation and the problems of generation can be among the several levels usually proposed for a linguistic analysis, and this will be the axis of the analysis undergone in the next section.

3. Results

The analysis of the outputs produced by MT tools followed three levels of linguistic analysis: lexical, syntactic, and semantic-pragmatic. For each one of them, we are going to present performance results and the behavior of tools in face of problems which traditionally affect each one of these domains.

3.1 Lexical level

At the lexical level, the MT systems performance was tested in four situations: dictionarization, homonymy, connotation and idioms.

In the first case, we can check that the translation quality is largely due to a correct and comprehensive dictionarization of the signs from the source language. The most frequent problems are proper names or words derived from them. Some MT systems did not translate into Portuguese words such as names of countries or nationalities ("*Northern Ireland*" (TN), "*Irish*" (AV), "*Hungarian*" (AV), "*Hungary*" (TN, AV, ET), "*Russian*" (AV)). The solution for this problem seems simple at first, but the consideration of expressions such as "*World Trade Center*", which point to a tendency, in Brazilian Portuguese, of preserving the English form of toponyms and anthroponyms, requires additional caution - which, besides, the tools do not seem to take: the expression was translated into "Centro de Comércio Mundial" (TP), "centro de comércio de mundo" (AV), "Mundial de Centro de Negócio" (ET) and "mundo Comércio Centro" (IT).

Another difficulty concerns the accuracy of dictionarized relations. In "*Os comerciantes da cidade pegaram carona na celebração da Páscoa e rechearam suas vitrines com todo tipo de opções de presentes*", most MT systems (TP, AV, GO) translated "*Páscoa*" as "*Passover*", when it should be translated as "*Easter*". Moreover, and maybe this is the most relevant datum here, we observed problems of dictionarization of even flexional variants and roots of words reasonably frequent in Portuguese and in English: "*weekend*" (TN), "*international*" (ET), "*trafficking*" (TN, ET, IT) , "*avantajado*" (AV, GO), "*fleuma*" (all tools), "*pelúcia*" (all tools), "*sofredores*" (all tools), "*nascituros*" (TP, IT), "*tradicionais*" (IT), "*camisetas*" (AV, IT), "*nova*" (IT), "*meses*" (IT), among others.

Homonymy problems reveal that dictionarizing is not enough. It is also necessary to develop strategies of selection among the several dictionarized options, every time it is necessary. The frequent presence of homonyms in Portuguese and English creates the need for the translation tool to use, to some extent, strategies of lexical disambiguation, because two or more linguistic forms in the target language can correspond to the same linguistic form in the source language. The quality of that choice affects the process of translation in various

degrees: if a badly made choice falls upon secondary lexical items in the source sentence (in the position of adjuncts, for example), the problem is less serious, because the sentence, although harmed in its local coherence, preserves a global coherence which, as a last resort, preserves the informative nucleus of the original sentence[2]. Nevertheless, a bad choice of lexical items in core positions (such as the subject) or in complementary positions (such as objects) largely sacrifices not only the local coherence, but also the very global coherence of the sentence, which often becomes unintelligible.

This problem was observed, for instance, in the translation of "*Each day, in the courtyard of the nation's most secure federal prison in Florence, Colo., a strange convocation takes place*", which Tradunet translated as "*Cada dia, dentro o pátio da nação tem a maior parte afiançar prisão federal em Florence, Colo., Uma convocação estranha acontece*". In this case, the interpretation of "*'s*" as the contraction of the verb "*have*", which takes the position of head of the predicate instead of a mark of the genitive case, significantly jeopardizes the intelligibility of the sentence. However, in the translation (by the same Tradunet) of "*Irish newspapers reported at the weekend that the IRA had linked an offer on disarmament to a withdrawal of British troops from Northern Ireland*", the lack of lexical correspondence for "*Northern Ireland*" and "*weekend*", which have both marginal positions in relation to the informative content of the sentence, has affected very little the quality of the output: "*Jornais irlandeses informaram no weekend que o IRA tinha associado uma oferta em desarmamento para um retirada de tropas Britânicas de Northern Ireland.*"

In a certain way, the same is valid for translations obtained for proper names and acronyms, such as "*São Paulo*" and "*F.B.I.*", which are words that resulted in absurd translations ("*are Paulo*" (IT) and "*F.B.eu.*" (ET), respectively), mainly due to the homonym "*São*" (which means "*Saint*", and not the verb "*to be*" in this case) and to the letter "*I.*" in "*F.B.I.*" (which means "*Investigation*", and not the nominative singular pronoun as it was translated). In these cases, as in the previous ones, we perceive that the translation tools do no put much effort into the process of lexical disambiguation, betting on a decision based only on the frequency of occurrences.

There are uncountable other evidence that lexical disambiguation is made in a schematic form, without considering the syntactic and semantic contexts in which the lexical items occur. For example, in "*F.B.I. reports say Mr. Mogilevich (...) has run prostitution rings (...)*", TN, TP, ET and IT translated "*has run*" as "*correu*", but the best translation here is "*dirigiu [negócios]*" ("*to manage or conduct[business]*"). Another example is the translation of "*firing*" in "*The F.B.I will also have the final say over the hiring and firing of*

[2] For further information on local and global coherence see Koch & Travaglia 1990.

the 10 hungarian agents". TN, TP, AV and IT translated the noun "*firing*" ("*demissão*") as the verb "*fazendo fogo*" or "*incendiando*" ("*to set fire*"), which is, in fact, one of its meanings, but not the one required in the passage. As the "*firing*" example above, in which tools changed the word category, there are also many other examples. One of them is the preposition "*but*" ("*exceto, menos, salvo*", in Portuguese), which was translated by TN, TP, AV and ET as the conjunction *"mas"* in the context *"(...) that good times would do anything but continue to roll*".

Nevertheless, problems concerning homonyms are not restricted to the correct grammatical classification of a word. Within the classes themselves, the meaning of linguistic forms can vary intensely, almost always as a result of the context in which they occur. This is what happened to the adverb "*ever*", whose meaning ranges from "*always*" to "*never*", and was incorrectly translated in the following sentence, *"(...) Hungary has ceded more sovereignty than many other nations – including the United States – would ever consider (...)"*. In most cases (TN, TP, AV and IT), "*ever*" was translated as "*sempre*" ("*always*"), but in this sentence the adequate translation would be "*jamais*" ("*never*"). This is a case of non-assertive form, in which the negative meaning is implicit in the comparative phrase "*more than*". A similar problem happened to the adverb "*scarcely*", which, although used with the verb in the affirmative, conveys a negative meaning, as in *"'To paraphrase a celebrated epitaph, prosperity left scarcely any of our industries untouched, and touched nothing it did not enrich'"*. The idea in this sentence is that almost none industry was left untouched by prosperity, however, in all outputs the meaning was the opposite.

The third problem observed at the lexical level concerns the connotative use of words from Portuguese (or from English) that, circumscribed to a very specific cultural context, cannot be automatically transferred from language to language, damaging the interpretability of the translated sentences. An example: in the sentence "*Os comerciantes da cidade pegaram carona na celebração da Páscoa e rechearam suas vitrines com todo tipo de opções de presentes*", the expression "*pegaram carona*" ("*to hitchhike")* is not being used in its literal meaning, but in the sense of "to take advantage", which is not supported in English by the verb *"to hitchhike"*. However, TP, AV and GO did not identify this difference. Another sentence in which some semantic knowledge is required is "*In fact one recent report from the British Heart Association analyzed the diets of European countries and found the French tops for a healthy diet*", since the fragment "*the French tops*" should be understood and translated as "*o francês lidera*" ("*the French leads*") and not as "*os topos franceses*" ("*the French peaks*": TN, TP and ET), or "os altos franceses" ("*the French heights*": AV), or even "os Franceses cimos" ("*the French summits*": IT).

Finally, we perceive serious problems related to the treatment of idioms, that is, to the situations in which the meaning of the expressions is not given in a compositional form, by the sum of the meanings of the words which compose them, but in a formulaic manner. Here also the lexical selection can affect the intelligibility of the translation to a higher or lower degree, depending on the importance the idiom has in the original sentence. Severe limitations of dictionaries are represented by cases where a single meaning is expressed in multiword lexical units, e.g. "as far as", "provided that" or "da mesma forma", "abrir mão de", etc. In this particular case, inadequate translation of English phrasal verbs leads to a number of errors. In the experiments, phrasal verbs were translated word by word, as in "*puxado sem*" (TN) or "*arrancou de*" (TP) from "*pulled out*", and "*quebrando acima*" (TN) or "*quebrar acima*" (AV, ET and IT) from "*breaking up*". Thus, the meaning of these phrasal verbs can only be understood by someone with a good knowledge of English. The translation of "*earlier this month*" as "*mais cedo este mês*" (TN, TP, AV and ET) can be simple and understandable if the person reading this fragment is an expert of the area or knows the English structure very well. But if the person does not have a good knowledge of the source language, this fragment can be understood as an event that happens every month, and which exceptionally happened earlier this month, which could cause a misunderstanding.

In general, the performance of the tools at the lexical level was as follows:

Table 1. Errors in lexical performance

Translators	Portuguese-English (495 words in total)					English-Portuguese (530 words in total)				
	A	B	C	D	total	A	B	C	D	total
TRANSLATOR PRO	11	13	31	8	63	01	11	15	7	29
Alta Vista	12	16	33	8	69	29	18	20	4	71
Intertran	47	85	60	7	199	12	33	26	4	85
GO Translator	10	16	33	8	67	-	-	-	-	-
TRADUNET	-	-	-	-	-	14	20	13	7	54
Enterprise Translator Server	-	-	-	-	-	09	15	10	4	38

A – Dictionarization (absence of translation for a certain lexical item)

B - Homonymy (incorrect translation of homonyms)

C - Idioms (incorrect translation of formulaic expressions)

D - Connotation (generalization of the figurative use of expressions which, in the target language, cannot play the same role)

The data shows that the performance in terms of output quality is approximately the same for all translators investigated, with the exception of Intertran, which presented a much larger number of errors. It is also worth mentioning that Alta Vista and GO Translator probably employ the same MT system, since their numbers are quite the same.

Concerning the errors found, we can see that most translation problems at the lexical level are associated with the treatment of formulaic expressions, that is, with situations in which a group of words acquires meanings that, in principle, are different from those manifested by the sum of their isolated meanings. Curiously, this is a problem that arises with a much higher frequency in the translation from Portuguese into English than in the opposite direction. Unfortunately, the data presented do not allow us to conclude whether it happens due to some internal characteristic of one of the two languages (the number of formulaic expressions in the daily use of Portuguese would be higher than in English) or due to the style adopted by editors in each country (American journalism would comparatively use less formulaic expressions than the Brazilian one).

In all senses, the main partial conclusion to be made from the data presented concerns, obviously, the proximity in numbers within types of errors. Save for the performance of Intertran, the outputs from Portuguese into English are practically the same In the opposite direction, that is, from English into Portuguese, the oscillation is significantly higher, what reveals that there are reasonably divergent linguistic processing strategies mobilized here.

3.2 Syntactic level

At the syntactic level, the MT systems performance was tested in either the retrieval of dependence relations in the original sentence or the establishment of such dependence relations in order to allow the interpretation of the translated sentence. The inversion of position between the subject and the object in the interpretation of the original sentence, for instance, although can produce intelligible sentences, will not produce sentences co-referential to those which served as a source for the translation process. Thus, we can observe at least two problems during the interpretation of sentences: the retrieval of possible syntactic relations between the lexical items and the selection of the most adequate relation from a repertoire of possible relations. The first problem is normally associated with parsing strategies, that is, those of reorganization of a list of words in a tree structure. The second problem derives from the very frequent possibility that a list of words correspond to more than one tree structure, what requires strategies for syntactic disambiguation.

From the perspective of the generation, the syntactic problems are of a different order, and concern obedience to grammatical rules, which characterize the good syntactic formation

of sentences in English and Portuguese (such as agreement and the use of prepositions and articles). It is a matter of linearization of a tree structure already defined, and not of the construction of the tree structure itself, as it is in a case of interpretation.

The dependence relation the syntactic level establishes with the lexical level and with the semantic level is unmistakable, and it will be described next. Among the best known strategies of lexical disambiguation surely is the consideration of minimal contexts, on the left and on the right, within which the ambiguous form appears – thus, it is a matter of syntactic information. Equally, the incorrect classification of a lexical item can lead to absurdity the syntactic analysis of a certain construction. This is what we perceived, in the previous section, when "'s" was classified as the contraction of the verb "have" instead of being classified as a particle which indicates the genitive case. The same happens between the syntactic and the semantic level. Lexical disambiguation turns out to be possible only if semantic information about the construction under analysis is available.

In the tools mentioned we observed problems with article-noun agreement ("um retirada" instead of "uma retirada", TN and IT), noun-verb agreement ("outras nações ia considerar" instead of "outras nações iam considerar", TN), use of verb tenses ("had been condemned" instead of "were condemned", AV), use of preposition ("ao fim de semana" instead of "no fim de semana", TP), use of article ("the fraternity" instead of "fraternity", TP, AV and GO), use of pronoun ("its" instead of "his", TP, AV and GO), use of comparison ("more early" instead of "earlier", AV and GO), as well as absence of preposition ("direito levar armas" instead of "direito de levar armas", TN), absence of article ("candidate" instead of "the candidate", TP), absence of reflexive pronouns ("he controlled" instead of "he controlled himself", TP) and absence of conjunctions ("dizem Senhor Mogilevich contratou..." instead of "dizem que Senhor Mogilevich contratou...", TN).

Part of these problems - agreement, among others – is associated with the process of linearization, which, as a general rule, do not include strategies of grammar checking. The lack of agreement between "um" and "retirada", for example, could be easily avoided if the application had a rule of generation that had implicitly the achievement of a article-noun agreement. However, a significant number of cases points to another direction: the difficulty of syntactic mapping between two languages. The uses of preposition, article, and conjunction are emblematic of the differences established between the Portuguese and English languages, for which the translation tools are no conveniently prepared yet. For they assume a biunique relation when, in fact, significantly different structures are involved. Although these problems become visible during the process of linearization, they are tributary to the process of

syntactic analysis, which could not retrieve the relations implicit there, when it attributed a tree structure to the list of words which composed the source sentence.

Finally, another minor problem of linearization concerns the use of specific lexical notations, which can provoke much noise in the interpretation of statements. We observed, for instance, problems with the use of capital letters, as in *"Em crise Mais cedo Este mês"* (for *"into crisis earlier this month"*), translated by TN.

In general, the comparative performance of tools in relation to the errors of a syntactic nature is as follows:

Table 2. Errors in syntactic performance

Translators	Portuguese-English (25 sentences in total)	English-Portuguese (22 sentences in total)
TRANSLATOR PRO	47	51
Alta Vista	64	44
Intertran	82	105
GO Translator	64	-
TRADUNET	-	66
Enterprise Translator Server	-	51

It is possible to consider that the table above presents just the problems of a syntactic order which were visible in the process of linearization. It does not mean that all problems of a syntactic nature are considered there, either because many of them are at the root of the problems of lexical order already pointed out or because certain contexts are less sensitive to the explanation of problems of this nature.

Impressions based on the previous table are confirmed for these data as well: except for the system Intertran, there is a significant convergence among the others, mainly if we consider the systems which go from English to Portuguese.

3.3 Semantic-pragmatic level

At this last level of analysis, two perspectives of evaluation can be considered: co-referentiality and interpretation of the sentences translated. We understand co-referentiality as the fidelity of translation to what was intended to be translated, that is, co-referentiality must be understood as the capability of target language sentences to denote the same portion of reality denoted by source language sentences. It is the ability of preserving reference across languages, in spite of the idiosyncratic way through which each language depicts reality. Co-referentiality is very close to intelligibility, but translated sentences may be intelligible even

when they are not co-referential to the original ones. In this case, we would say that translation is somewhat misleading.

As interpretability we understand the capacity of a certain list of words to stop being just a list of words and become a broader linguistic unit, which convey meaning. It presupposes in many senses a success in the process of lexical and syntactic mapping between the two languages compared, but it does not predict, except subsidiarily, the fidelity to the original.

In the analysis undergone here, our concern lied rather in interpretability than in co-referentiality. In order to assess this feature, we adopted a procedure in which a human knowledgeable of the target language could retrieve the message conveyed in the passage. Therefore, even errors related to inappropriate choice of lexical items among distinct meanings of an item could blur the message, and these cases were considered as failure in gist preservation. Within this perspective, only 40% of the passages had their gist preserved in Portuguese/English translations, whereas for the English/Portuguese counterparts the percentage is even lower (ca. 30%). In these figures, we did not consider Intertran for the reasons already commented upon. One may speculate that this asymmetry in performance is due to the difficulties in linearizing syntactic structures in Portuguese, simply because much less experience has been gathered than it is true for linearization of text in English. Interestingly, the overall number of mistakes in the translation is, in average, larger in the Portuguese/English translations, though the latter are more accurate in terms of gist preservation. This observation reinforces the need to analyze data on errors with extreme caution, for grammatical or spelling mistakes may have completely different impact on the accuracy of the translation.

Table 3a. Interpretability scores

Translators	Portuguese-English										
	01	02	03	04	05	06	07	08	09	10	%
TRANSLATOR PRO	N	N	Y	Y	N	N	Y	N	N	Y	40
Alta Vista	N	N	Y	N	N	Y	Y	N	N	Y	40
Intertran	N	N	N	N	N	N	N	N	N	N	0
GO Translator	N	N	Y	N	N	Y	Y	N	N	Y	40
TRADUNET	-	-	-	-	-	-	-	-	-	-	-
Enterprise Translator Server	-	-	-	-	-	-	-	-	-	-	-
TOTAL (%)	0	0	75	25	0	50	75	0	0	75	-

Table 3b. Interpretability scores

Translators	English-Portuguese										
	01	02	03	04	05	06	07	08	09	10	%
TRANSLATOR PRO	N	Y	N	N	N	Y	N	N	N	N	20
Alta Vista	N	Y	N	N	N	Y	Y	N	N	N	30
Intertran	N	N	N	N	N	Y	N	N	N	N	10
GO Translator	-	-	-	-	-	-	-	-	-	-	-
TRADUNET	N	Y	N	N	N	Y	N	N	N	N	20
Enterprise Translator Server	N	Y	N	N	N	Y	Y	N	N	N	30
TOTAL (%)	0	80	0	0	0	100	40	0	0	0	-

The table above points out the human evaluation of the tools performance for the set of twenty excerpts whose translation was automatically produced. The human evaluator was charged, mainly, with the task of checking if translators' output was interpretable, yes (Y) or no (N), from a neuter perspective concerning the source sentence. According to the data obtained, we can see that the sentences have, consistently, some degree of difficulty or ease which is verifiable in all translation systems. A typical example of a sentence in English easily translatable into Portuguese is the excerpt number 6, transcribed below, with the output obtained:

Table 4. Brazilian Portuguese Generation outputs for the English original sentence: "The child is at the center of an international fight between his father in Cuba, Juan Miguel González, and relatives in Miami."

Translators	Outputs
TRANSLATOR PRO	A criança está no centro de uma briga internacional entre o pai dele em Cuba, Juan Miguel González, e parentes em Miami.
Alta Vista	A criança está no centro de uma luta internacional entre seu pai em Cuba, em Juan Miguel González, e em parentes em Miami.
Intertran	A criança é no centro de um internacional batalhar entre dele pai em Cuba, Juan [Miguel] [González], e familiares em Miami.
TRADUNET	A criança está no centro de uma briga internacional entre seu pai em Cuba, Juan Miguel González, e parentes em Miami.
Enterprise Translator Server	A criança está no centro de uma luta de international entre seu pai em Cuba, González de Miguel de Juan, e parentes em Miami.

The opposite happens in the excerpt number 1, which was not conveniently translated by any of the applications employed.

Table 5. Brazilian Portuguese Generation outputs for the English original sentence: "The Northern Ireland peace process was plunged into crisis earlier this month when the IRA pulled out of talks with the British province's independent disarmament body and withdrew a proposal to put its arms 'beyond use'"

Translators	Outputs
TRANSLATOR PRO	O Irlanda paz processo Do norte foi mergulhado mais cedo em crise este mês quando o IRA arrancou de conversas com o corpo de desarmamento independente da província britânica e retirou uma proposta para pôr seus braços `` além de uso.''
Alta Vista	O processo da paz de Irlanda do Norte foi mergulhado na crise mais cedo este mês em que o IRA puxou fora das conversas com o corpo independente do disarmament da província britânica e retirou uma proposta para pôr seu uso do ``beyond dos braços.''
Intertran	O Irlanda do Norte processo de paz era mergulho para a crise anteriormente este mês quando o IRA puxado ausente de conversas com o Britânico competência independente desarmamento corpo e retirado um proposta colocar o seu armas [``beyond] use.''
TRADUNET	A paz do Northern Ireland processa Esteve mergulhado Em crise Mais cedo Este mês quando o IRA puxado sem conversas com o desarmamento independente da província Britânica corporifica E withdrew uma proposta pôr seus braços ` ` além de uso.''
Enterprise Translation Server	O processo do norte de paz de Irlanda foi mergulhado em crise mais cedo este mês quando o IRA arrancaram de conversas com a província Britânica corpo independente de desarmamento e retirou uma proposta a por seu uso de ççbeyond de braços.''

The same happens in the opposite direction: there are sentences in Portuguese which turn out to be simpler in the process of translation into English (excerpts 03, 07 e 10, for instance), and there are others for which no tool can achieve a reasonable solution (excerpts 01, 05, 08 e 09). Due to limitations in space, these sentences are not going to be transcribed here.

In general, the conclusions that can be made from the data reported point to the fact that texts written in the SVO form, as in Excerpts 2 and 6 of the second part, produced better translations than longer and more complex texts (as in Excerpts 7, 8 and 10 of the second part), which posed more problems to the MT systems employed.

In a subsidiary experiment, an expert human translator was asked to try and retrieve the original information out of the translated passages. Using her knowledge of both languages and previous experience as a translator and having all versions of a given passage

at the same time, she managed to extract the correct information from practically all passages. This is certainly a much better result than we quoted for passages taken in isolation, and without the expertise of an experienced human translator. It shows that, in spite of the problems discussed here, MT systems may be close to being of real use, even in open environments. It is interesting to note that if one could somehow combine the outputs from distinct systems, considerable improvement would be obtained. This occurs because the MT systems analyzed differ in their weak and strong points, and some degree of complementarity should be expected. In addition, in some cases, previous knowledge about the subject matter of the text translated can help to understand the message even when translation is poor.

Final Remarks

There are countless conclusions to be made from the data analyzed. The first of them involves the confirmation of what we could call an intuition of users of machine translation tools for Brazilian Portuguese: the commercially available systems do not work properly, and their outputs are seldom useful without post-editing or expertise in the source language. Less than 50% of the excerpts proposed could lead to intelligible outputs. Nevertheless, we could perceive that this deficiency has not been motivating its developers to test alternative strategies since the scores of performance are almost always the same, with the exception of Intertran.

Apparently, Intertran was based on the direct approach, in a word-by-word translation, for it is clear that fixed translations are employed for some lexical items. It does not present any agreement between words (adjectives and nouns, or verbs and nouns), employs standard translations for some words, such as "*your own*" for "*seu*" (possessive pronoun); "*in case that*" for "*se*" (reflexive pronoun); "*than it is to it*" for "*que*" (relative pronoun); "*as of*" for "*de*" (preposition); and "*the one*" for "*a*" or "*o*" (articles) in the beginning of a sentence, regardless of their function in it. Even though such information is not provided along with the systems, the other MT systems were probably developed using the transfer approach, since the mapping of syntactic structures is apparent in some translations.

The problems observed concern in general three main factors. The first is the absence of linguistic resources or a bad quality in the available ones. This point is made clear when we consider the issues related to the dictionaries used and to the strategies employed to access them. We could perceive that the dictionaries are still significantly poor in terms of coverage, and the access to the few existent entries is made in a very schematic way, that is, there is a naive belief that it is possible to assign default values to lexical entries, regardless of the

syntactic and semantic contexts in which they appear. In consequence, there is little or no effort in disambiguation procedures, which would lead to more reliable results.

The second problem concerns the mistaken hypotheses about English/Portuguese similarities and differences. Besides the idea that there is a one-to-one correspondence between these two languages lexical items, there is a widespread belief in the universality of semantics, as if a sentence meaning could be equally built in both languages. Thus, it is assumed that for every semantic structure (e.g., subject-agent/verb-action structure) of English language there is a straightforward correspondence in Portuguese, and vice-versa. It is ignored that in many cases semantic structures are context-sensitive or even culturally-sensitive.

Finally, it is worth observing that natural language translation is not a simple task at all. The expectation of users is that translation preserves the original sentence function and power, that is, retrieves faithfully not only the informational content but also the information wrapping, since form is as meaningful as content. However, as the information interactors belong to different cultures and occupy different discourse places, the task of preserving utterer's intentions in translation is almost impossible even for human beings. It is not completely unexpected, though, that machine translation fails in the reproduction of the original sentences efficacy, for this is still a challenge even for non-computational translation.

References

Arnold, D. *et al.* (1994). *Machine translation: An introductory guide.* NCC/Blackwell, Manchester, Oxford.

Hutchins, W.J. and Sommers, H.L. (1992). An introduction to machine translation. *In An introduction to machine translation.* Academic Press, London.

Koch, I.G.V. and Travaglia, L.C. (1990). *A coerência textual.* Contexto, São Paulo.

Lehrberger, J. and Bourbeau, L. (1988). *Machine translation: Linguistic characteristics of MT systems and general methodology of evaluation.* John Benjamins, Amsterdam, Philadelphia.

White, J.S. *et al.* (1994). The ARPA MT evaluation methodologies: Evolution, lessons, and future approaches. *In Technology partnerships for crossing the language barrier*: Proceedings of the First Conference of the Association for Machine Translation in the Americas, pages 193-205, Washington, DC, Association for Machine Translation in the Americas.

NOTAS DO ICMC

SÉRIE COMPUTAÇÃO

046/2000 SOUZA, P.S.L.; SANTANA, M.J.; SANTANA, R.H.C. – Escalonamento de processos: uma contribuição para a convergência da área.

045/2000 MARTINS, R.T.; RINO, L.H.M.; NUNES, M.G.V.; MONTILHA, G.; OLIVEIRA JR., O. N. – An interlingua diming at communication on the Web: how language-independent can it be?

044/99 FELTRIM, V.D.; FORTES,R.P.M. – Evaluation of a reverse engineering method through the application in a hypermedia system.

043/99 PANSANATO, L.T.E.; NUNES, M.G.V. – EHDM: Método para projeto de hiperdocumentos para ensino.

042/99 MORABITO, R.; ARENALES, M. – Optimizing the cutting of stock plates in a furniture company.

041/98 SANTOS-MEZA, E.; SANTOS, M.O.; ARENALES, M.N. – Lot sizing and scheduling in na automated foundry.

040/98 FELTRIM, V.D.; FORTES, R.P.M. – Uma modelagem do domínio de engenharia reversa de software utilizando o método OOHDM.

039/98 FERREIRA, V.G.; MELLO, O .D.; OLIVEIRA, J.N.; FORTUNA, A . O . - Tópicos teóricos e computacionais em escoamentos de fluidos.

038/98 CAVICHIA, M.C.; ARENALES, M.N. - Piecewise linear programming via interior points.

037/98 REZENDE, S. O. ; PUGLIESI, J.B. - Aquisição de conhecimento explícito ou manual - versão 1.0