

An Experiment in Spanish-Portuguese Statistical Machine Translation

Wilker Ferreira Aziz¹, Thiago Alexandre Salgueiro Pardo¹ and Ivandré Paraboni²

¹ University of São Paulo – USP / ICMC
Av. do Trabalhador São-Carlense, 400 - São Carlos, Brazil

² University of São Paulo – USP / EACH
Av. Arlindo Bettio, 1000 - São Paulo, Brazil

wilker.aziz@usp.br, taspardo@icmc.usp.br, ivandre@usp.br

Abstract. Statistical approaches to machine translation have long been successfully applied to a number of ‘distant’ language pairs such as English-Arabic and English-Chinese. In this work we describe an experiment in statistical machine translation between two ‘related’ languages: European Spanish and Brazilian Portuguese. Preliminary results suggest not only that statistical approaches are comparable to a rule-based system, but also that they are more adaptive and take considerably less effort to be developed.

1 Introduction

As in many other Natural Language Processing (NLP) tasks, statistical techniques have emerged in recent years as a highly promising approach to Machine Translation (MT), and are now part of mainstream research in the field. By automatically analysing large collections of parallel text, statistical MT is often capable of outperforming existing (e.g., rule-based) systems even for so-called ‘distant’ language pairs such as English and Arabic. Regular international contests on MT organized by NIST (*National Institute of Standards and Technology*) have shown the advances of statistical approaches. In the 2005 contest, when statistical and rule-based MT systems competed, the best-performing statistical system – the Google Translator – fared about 376% higher than the widely known rule-based system Systran for the language pair English-Arabic. For the English-Chinese pair, Google Translator outperformed Systran by 140%.

Researchers in the area argue that, since the statistical MT techniques are able to perform so well for distant languages, with all likelihood even better results should be achieved for ‘closely-related’ (e.g., Romance) languages. As an attempt to shed light on this issue, in this work we describe an experiment in statistical machine translation between two ‘related’ languages, namely, European Spanish and Brazilian Portuguese. In doing so, rather than seeking to produce a fully developed MT system for this language pair, we limit ourselves to investigate our basic statistical MT system as

compared to the shallow-transfer approach in [1]. We shall then argue that our preliminary results - although not overwhelming given the small scale of the experiment - provide remarkable evidence in favour of the statistical approach and pave the way for further research.

The rest of this paper is structured as follows. Section 2 describes the statistical translation model used in our work. Section 3 describes the training data, the sentence and word alignment procedures and the alignment revision techniques involved. Section 4 presents results of a preliminary evaluation work and a comparison with the system described in [1]. Finally, Section 5 summarizes our work so far.

2 Translation Model

Generally speaking, a statistical approach to translate from, e.g., Spanish to Portuguese involves finding the Portuguese sentence p that maximizes the probability of p given a Spanish sentence e , that is, the probability $P(p | e)$ of p being a translation of e . Given the probabilities $P(p)$ and $P(e)$ obtained from Portuguese and Spanish language models, respectively, the problem can be expressed in Bayesian terms as

$$P(p | e) = \frac{P(p)P(e | p)}{P(e)}$$

Since $P(e)$ remains constant for the input sentence e , this amounts to maximizing

$$P(p)P(e | p)$$

in which $P(e | p)$ is the *translation model*. The *language model* $P(p)$ may be computed as the probability of the sequence of n-grams in the translation p .

Given a pair $\langle e, p \rangle$ of sentences, it is said that e and p are mutual translations if there is at least one possible lexical alignment among them, i.e., correspondences among their words and/or phrases. Assuming the set of all possible alignments among e and p to be a , obtaining $P(e | p)$ can be seen as the maximization of the sum of individual contributions of every single alignment a_i , a process called *decoding*:

$$P(p | e) = P(p) \sum_i P(a_i, e | p)$$

The expression $P(a_i, e | p)$ can be obtained empirically from a number of parameters intended to capture relevant aspects of the translation process. In this work we use the basic IBM 4 model [10], in which the following parameters are defined:

- $\langle a \rangle$ the set of alignments for $\langle e, p \rangle$.
- ε the probability of a target-sentence of length m being the translation of a source-sentence of length l .
- t the probability of e being a translation of p .
- d the probability of a target word being placed at the j^{th} position in the target sentence given that it is a translation of a source word at the i^{th} position in the source sentence.
- p_l the probability of obtaining a spurious word.

- p_0 (or $1 - p_1$) the probability of not obtaining a spurious word.
- φ_0 the fertility of spurious cases¹.
- φ_i fertility rate associated with the i^{th} word.
- n the probability of a word p_i having fertility φ_i .

Fertility is defined as the number of target words generated from a given source word. A zero fertility value corresponds to a deletion operation. A spurious word appears so to speak “spontaneously” in the target sentence when there is no corresponding source word. This is modelled by the parameters p_1 and φ_0 and the NULL symbol in the word alignment representation (see Section 3.)

The above parameters are actually the same ones defined in model 3 described in [10]. However, the presently discussed model 4 also distinguishes among three word types (heads, non-heads and NULL-generated), allowing for more control over word order. For the interested reader, the complete model is defined as

$$P(a, e | p) = \binom{m - \varphi_0}{\varphi_0} p_0^{(m - 2\varphi_0)} p_1^{\varphi_0} \prod_{i=1}^l n(\varphi_i | p_i) \prod_{i=1}^l \prod_{k=1}^{\varphi_i} t(\tau_{ik} | p_i) \times$$

$$\times \prod_{i=1, \varphi_i > 0}^l d_1(\pi_{i1} - c_{\rho i} | \text{class}(e_{\rho i}), \text{class}(\tau_{i1})) \prod_{i=1}^l \prod_{k=2}^{\varphi_i} d_{>1}(\pi_{ik} - \pi_{i(k-1)} | \text{class}(\tau_{ik})) \prod_{k=1}^{\varphi_0} t(\tau_{0k} | \text{NULL})$$

For more details, we refer to [10]. In the remainder of this paper we will focus on the above model only.

3 Training

In order to build a translation model as described in the previous section, we collected 645 Portuguese-Spanish text pairs from the Environment, Science, Humanities, Politics and Technology supplements of the on-line edition of the “Revista Pesquisa FAPESP”, a Brazilian magazine on scientific news. The corpus consists of 17,681 sentence pairs comprising 908,533 words in total (being 65,050 distinct.) The Portuguese version consists of 430,383 words (being 32,324 distinct) and the Spanish version consists of 478,150 words (being 32,726 distinct.) We are aware that our data set is considerably smaller than standard training data used in statistical MT (standard training data in the area comprises about 200 million words, while the Google team have been using over 1 billion words in their experiments.) However, as we shall see in the next sections, the amount of training data that we used in this work turned out to be sufficient for our initial investigation.

Portuguese sentential segmentation was performed using SENTER [2]. The tool was also employed in the segmentation of the Spanish texts with a number of changes (e.g., in order to handle Spanish abbreviations.) Despite the similarities between the two languages, it is immediate to observe that word-to-word translation is not feasible, as Example 1 should make clear: besides the differences in word order, there are

¹ Defined as the <NULL> symbol in the word alignment – see Example 2 and 3 in Section 3.

subtle changes in meaning (e.g., “espantosa” vs. “impresionante”, analogous to “amazing” vs. “impressive”), additional words (e.g. “ubicada”) and so on.

Example 1. A Portuguese text fragment (left) and corresponding Spanish translation (right.)

Ao desencadear uma cascata de eventos físico-químicos poucos quilômetros acima da floresta, a espantosa concentração de aerossóis na Amazônia no auge da estação (...)

Esa impresionante concentración de aerosoles en la Amazonia, al desencadenar una cascada de eventos fisicoquímicos ubicada a algunos kilómetros arriba del bosque, en el auge de la estación (...)

For the purposes of this work, a *sentence alignment* a is taken to be an ordered set of $p(a)$ sentences in our Portuguese corpus and an ordered set of $s(a)$ related sentences in the Spanish corpus. Values of $p(a)$ and $s(a)$ can vary from zero to an arbitrary large number. For example, a Portuguese sentence may correspond to exactly one sentence in the Spanish translation, and such 1-to-1 relation is called a *replacement* alignment. On the other hand, if a Portuguese sentence is simply omitted from the Spanish translation then we have a 1-to-0 alignment or *deletion*. In our work we are interested in replacement alignments only. This will not only reduce the computational complexity of our next task – word alignment – but will also provide the required input format for MT tools that we use, such as GIZA++ [5].

For the sentence alignment task, we used an implementation of the *Translation Corpus Aligner* (TCA) method [3] called TCAalign [4]. The choice was based on the high precision rates reported for Portuguese-English (97.10%) and Portuguese-Spanish (93.01%) language pairs [4]. The set of alignments produced by TCAalign consists of m-to-n relations marked with XML tags. As our goal is to produce an aligned corpus as accurate as possible, the data was inspected semi-automatically for potential misalignments (which were in turn collapsed into appropriate 1-to-1 alignments.) As a result of the sentence alignment procedure, 10% of the alignments in our corpus were classified as unsafe, and their manual inspection revealed that 1,668 instances (9.43%) were indeed incorrect.

A major source of misalignment was due to segmentation errors, which caused two or more sentences to be regarded as a single unit. These cases were adjusted manually so that the resulting corpus contained a set of Portuguese sentences and their Spanish counterparts in 1-to-1 relationships. There were also cases in which n Portuguese sentences were (correctly) aligned to n Spanish sentences in a different order, and had to be split into individual 1-to-1 alignments. Some kinds of misalignment were introduced by the alignment tool itself, and yet others were simply due to different choices in translation leading to correct n-to-m alignments. Since punctuation will be removed in the generation of our translation models, in these cases it was possible - when there was no change in meaning - to manually split the Spanish sentence and create two individual 1-to-1 alignments.

Two versions of the corpus have been produced: one represents the aligned corpus in its original format (as the examples in the previous section), with capital letters, punctuation marks and alignment tags; the other represents the aligned corpus in GIZA++ [5] format, in which the entire text was converted to lower case, punctuation marks and tags were removed, and the correspondence between sentences is given

simply by their relative position within each text (i.e., with each sentence representing one line in the text file.)

We used GIZA++ to align the second version of the corpus at word level. A word alignment contains a sentence pair identification (sentence pair id, source and target sentence length and the alignment score) and the target and source sentences, in that order. The word alignment information is represented entirely in the source sentence. Each source word is followed by a reference to the corresponding target, or an empty set { } for source words that do not have a correspondence, i.e., n-0 (deletion) relationships. The source sentence contains also a NULL set representing the target words that do not occur in the source, i.e., 0-n (inclusion) relationships.

The following examples illustrate the output of the word alignment procedure for two Spanish-Portuguese sentence pairs. In the first case, the source words “de” and “los” were not aligned with any word in the target sentence, i.e., they disappeared in the Portuguese translation. On the other hand, all Portuguese words were aligned with their Spanish correspondents as shown by the empty NULL set. In the second case, the source words “la”, “de” and “el” were not aligned to any Portuguese words, but “niño” was aligned with two of them (i.e., “el niño”). In the same example the NULL set shows that the target word 4 (i.e., “do”) is not aligned with any Spanish words, i.e., it was simply included in the Portuguese translation.

Example 2. Deletion of source (Spanish) words “de” and “los”.

Sentence pair (1) source length 8 target length 6 alignment score : 0.00104896
os dentes do mais antigo orangotango
NULL ({}) hallan ({} 1) dientes ({} 2) del ({} 3) más ({} 4) antiguo ({} 5) de ({}) los ({}) orangutanes ({} 6)

Example 3. Deletion of source (Spanish) words “la”, “de” and “el”, inclusion of target (Portuguese) word “do” and 1-2 alignment of “niño” with “el niño”.

Sentence pair (14) source length 8 target length 6 alignment score : 1.28219e-08
salinidade indica chegada do el niño
NULL ({} 4) la ({}) salinidad ({} 1) indica ({} 2) la ({}) llegada ({} 3) de ({}) el ({}) niño ({} 5 6)

The word alignments were generated using the following GIZA++ parameters. The probability p_0 was left to be determined automatically; language model parameters (smoothing coefficient, minimum and maximum frequencies and smoothing constant for null probabilities) were left at their default values; the maximum sentence length was set to 182 words as seen in our corpus, and the maximum fertility parameter was set to 10 (i.e., word fertilities will range from 0 to 9) as suggested in [5]. As a result, 489,594 word alignments were produced according to the following distribution.

Table 1. Word alignment distribution.

Alignment Type	Instances	Probability
0-1	15040	3.072 %
1-0	71543	14.613 %
1-1	398175	81.328 %
1-2	3344	0.683 %
1-3	791	0.167 %
1-4	305	0.062 %
1-5	187	0.038 %
1-6	85	0.017 %
1-7	42	0.009 %
1-8	27	0.005 %
1-9	55	0.011 %

About 82% of all alignments were of the word-to-word type. Moreover, very few mappings (701 cases or 0.14% in total) involved more than three words in the target language (i.e., alignments 1-4 to 1-9.) These results may suggest a strong similarity between Portuguese and Spanish as argued in [1] for the languages spoken in Spain.

4 Decoding and Testing

The translation model described in Section 2 and associated data (e.g., word class tables, zero fertility lists etc. - see [5] for details) were generated from a small training set of about 17,000 sentence pairs using a simple trigram-based language model produced by the CMU-Cambridge Tool Kit [8]. Both Spanish-Portuguese and Portuguese-Spanish translations were then decoded using the ISI ReWrite Decoder tool [9] as follows:

$$\begin{aligned}
 \text{Spanish-Portuguese translator: } & P(p) = P(p)P(e | p) \\
 & \text{Language Model (Portuguese): } P(p) \\
 & \text{Translation Model (Portuguese-Spanish): } P(e / p) \\
 \text{Portuguese-Spanish translator: } & P(e) = P(e)P(p | e) \\
 & \text{Language Model (Spanish): } P(e) \\
 & \text{Translation Model (Spanish-Portuguese): } P(p / e)
 \end{aligned}$$

Recall from Section 2 that given a translation from Spanish to Portuguese, the decoding process intends to find the sentence p that maximizes

$$P(p | e) = P(p) \sum_i P(a_i, e | p)$$

The decoding task was based on the IBM model 4 (described in Section 2) using 5 iterations during training procedure. The selected decoding strategy was the A* search algorithm described in [5] using the translation heuristics for maximizing $P(t / s) * P(s / t)$.

For evaluation purposes, a test set of 649 previously unseen sentence pairs was taken from recent issues of “Revista Pesquisa FAPESP” and it was used as input to our translator. In the Appendix at the end of this paper we present a number of instances of Spanish-Portuguese (Table 4) and Portuguese-Spanish (Table 5) translations obtained in this way. In each example, the first line shows a test sentence in the source language; the second line shows the reference (i.e., human-made) translation, and the third line presents the output of our statistical machine translator. Occasional target words shown in the original (source) language are due to the overly small size of our training data.

Using the same test data, we compared the BLEU scores [6] obtained by our statistical approach to those obtained by the rule-based system Apertium [1]. Briefly, BLEU is an automatic evaluation metric which computes the number of n-grams shared between an automatic translation and a human (reference) translation. The higher the BLEU score, the better the translation quality. BLEU scores have been shown to be comparable to human judgements about translated texts. Accordingly, BLEU has been extensively used in MT evaluation. Table 2 summarizes our findings.

Table 2. BLEU scores for Statistical versus rule-based translations.

Approach	Portuguese-Spanish	Spanish-Portuguese
Statistical	0.5673	0.5734
Apertium	0.6282	0.5955

From the above results a number of observations can be made. Firstly, although the Apertium system slightly outperforms our statistical approach, their BLEU scores remain nevertheless remarkably close. This encouraging first trial allows us to predict that these results would most likely be the other way round had we used a more realistic amount of training data². The growth potential of our approach is still vast, and it can be achieved at a fairly low cost. This seems less obvious in the case of tailor-made translation rules. Secondly, our development efforts probably took only a fraction of the time that would be required for developing a comparable rule-based system, and that would remain the case even if we limited ourselves to one-way translation. Our system naturally produces two translation models, and hence translates both ways, whereas a rule-based system would require two sets of language-specific translation rules to be developed by language experts. Finally, it should be pointed out that as language evolves and translation rules need to be revised, a rule-based system may ultimately have to be re-written from scratch, whereas a statistical translation model merely requires additional (e.g., more recent) instances of text to adapt.

It may be argued however that BLEU does not capture translation quality accurately in the sense that it does not reflect the degree of difficulty experienced by humans in the task of post-edition. For that reason, we decided to carry out a (manual) qualitative evaluation of both statistical and rule-based MT in the Spanish-Portuguese direction. More specifically, we compute the number of lexical and syntactical errors as suggested in [11] and word error rates (WER) as follows:

² And possibly a more powerful language model e.g., [7].

$$WER = \frac{I + D + R}{S}$$

In the above, ‘I’ stands for the number of insertion steps necessary to transform the automatic translation into the reference one, ‘D’ is the number of deletions, ‘R’ is the number of replacements and ‘S’ is the number of words in a source reference set.

A random selection of 20 instances of translations (482 words in total) taken from our test data was analysed at lexical and syntactical levels. The lexical level takes into account *dictionary* errors (e.g., words not translated or wrong translations of proper names, abbreviations, etc.), *homonyms* (wrong translation choices), *idioms* (incorrect or word-to-word translation of idiomatic expressions), and *connotative* errors (literal translation of expressions with no literal meaning.) At the syntactic level we looked into sentence structure, i.e., errors stemming from verb or noun agreement, the choice of determiners and prepositions, and the appropriate use of verbal complements. The results of the analysis are show in Table 3 below, in which lower WERs are better.

Table 3. Statistical versus rule-based Spanish-Portuguese translations.

Approach	Lexical Level				Syntactic Level	WER
	Dictionary	Homonyms	Connotative	Idioms		
Statistical	19	7	0	2	32	0.3216
Apertium	40	5	0	2	24	0.2635

At first glance the results show that the statistical approach presented a higher WER, which may suggest that its output would require more effort to become identical to the reference translation if compared to Apertium. However, we notice that the WER difference was mainly due to the lack of preposition + determiner contractions as in “de” (of) + “a” (the) = “da”, causing WER to compute one replacement and one deletion for each translation. Thus, given that preposition post-processing can be trivially implemented, we shall once again interpret these results as indicative of comparable translation performances under the assumption that the present difficulties could be overcome by using a larger training data set. Due to the highly subjective nature of this evaluation method, however, we believe that more work on this issue is still required.

5 Final Remarks

In this paper we have described a first experiment in Portuguese-Spanish statistical machine translation using a small parallel corpus as training data. Preliminary results seem to confirm long held claims of the statistical MT community suggesting that this approach may be indeed comparable to rule-based MT in both performance (particularly as measured by BLEU scores) and development costs.

We now intend to expand the current training data to the (much higher) levels commonly seen in modern work in the field, and possibly make use of a more sophisticated translation model. We are also aware that the translation task between such

‘closely-related’ languages is somewhat simpler and does not reveal the full extent of translation difficulties found in other (i.e., so-called ‘distant’) language pairs. For that reason, we will also investigate the translation of Portuguese texts to and from English, having as an ultimate goal the design of a robust, state-of-art statistical translation model of practical use for Brazilian Portuguese speakers.

Acknowledgments

The authors acknowledge support from FAPESP (2006/04803-7, 2006/02887-9 and 2006/03941-7) and CNPq (484015/2007 9.)

References

1. Corbí-Bellot, A.M.; Forcada, M.L.; Ortiz-Rojas, S.; Pérez-Ortiz, J.A.; Ramírez-Sánchez, G.; Sánchez-Martínez, F.; Alegria, I.; Mayor, A.; Sarasola, K. An open-source shallow-transfer machine translation engine for the romance languages of Spain. 10th Annual Conference of the European Association for Machine Translation (2005) 79-86.
2. Pardo, T. A. S. SENTER: Um Segmentador Sentencial Automático para o Português do Brasil. NILC Technical Reports Series NILC-TR-06-01. University of São Paulo, São Carlos, Brazil (2006).
3. Hofland, K. and Johansson, S. The Translation Corpus Aligner: A program for automatic alignment of parallel texts. In: Corpora and Cross-linguistic research Theory, Method, and Case Studies. S. Johansson & S. Oksefjell (Eds.): Rodopi, Amsterdam (1998).
4. Caseli, H. M. Indução de léxicos bilíngües e regras para a tradução automática. Doctoral thesis, University of São Paulo, São Carlos, Brazil (2007).
5. Och, F.J. and Ney, H. A Systematic Comparison of Various Statistical Alignment Models. Computational Linguistics, Vol. 29, nro.1 (2003) 19-51.
6. Papineni, K.; Roukos, S.; Ward, T. and Zhu, W. BLEU: a Method for Automatic Evaluation of Machine Translation. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (2002) 311-318.
7. Chen, S.F. and J. Goodman. An empirical study of smoothing techniques for language modeling. Computer Speech and Language 13 (1999) 359-394.
8. Clarkson, P. R. and Rosenfeld, R. Statistical Language Modeling Using the CMU-Cambridge Toolkit In Proceedings of ESCA Eurospeech (1997).
9. Germann, U.; Jahr, M.; Knight, K.; Marcu, D., and Yamada, K. Fast Decoding and Optimal Decoding for Machine Translation. Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics (2001).
10. Brown, P. E.; Pietra, S. A. D.; Pietra, V. J. D. and Mercer, R. L. The Mathematics of Statistical Machine Translation: Parameter Estimation. Computational Linguistics, Vol. 16, N. 2 (1993) 79-85.
11. Oliveira Jr., O. N.; A. R. Marchi; M. S. Martins and R. T. Martins. A Critical Analysis of the Performance of English-Portuguese-English MT Systems. V Encontro para o processamento computacional da Língua Portuguesa Escrita e Falada (2000).

Appendix – Examples of Human and Machine Translation

Table 4. Spanish source sentences followed by (r) Portuguese reference, (s) statistical and (a) rule-based (Apertium) machine translations.

si fuera así probaremos con otro hasta lograr el resultado deseado

(r) mas aí testaremos outro até conseguirmos o resultado desejado

(s) se fora assim probaremos com outro até conseguir o resultado desejado

(a) se fosse assim provaremos com outro até conseguir o resultado desejado

en esa región viven cerca de 40 mil personas en nueve favelas y tres conjuntos habitacionales

(r) na região vivem cerca de 40 mil pessoas em nove favelas e três conjuntos habitacionais

(s) em essa região vivem cerca de 40 mil pessoas em nove favelas e três conjuntos habitacionales

(a) nessa região vivem cerca de 40 mil pessoas em nove favelas e três conjuntos habitacionais

sus textos escritos en lengua cuneiforme en forma de cuña grabada en tablas de barro son poesías en homenaje a una diosa llamada inanna adorada por la sacerdotisa

(r) seus textos escritos na linguagem cuneiforme em forma de cunha gravada em tábuas de barro são poesias em homenagem a uma deusa chamada inanna adorada pela sacerdotisa

(s) seus textos escritos em língua cuneiforme em forma de cuña gravada em tábuas de barro são poesias em homenagem à uma deusa chamada inanna adorada por a sacerdotisa

(a) seus textos escritos em língua cuneiforme em forma de cuña gravada em tabelas de varro são poesias em homenagem a uma deusa chamada inanna adorada pela sacerdotisa

Table 5. Portuguese source sentences followed by (r) Spanish reference, (s) statistical and (a) rule-based (Apertium) machine translations.

mas aí testaremos outro até conseguirmos o resultado desejado

(r) si fuera así probaremos con otro hasta lograr el resultado deseado

(s) pero entonces testaremos otro hasta logremos el resultado deseado

(a) pero ahí probaremos otro hasta conseguir el resultado deseado

na região vivem cerca de 40 mil pessoas em nove favelas e três conjuntos habitacionais

(r) en esa región viven cerca de 40 mil personas en nueve favelas y tres conjuntos habitacionales

(s) en región viven alrededor de 40 mil personas en nueve favelas y tres conjuntos habitacionais

(a) en la región viven cerca de 40 mil personas en nueve favelas y tres conjuntos habitacionales

seus textos escritos na linguagem cuneiforme em forma de cunha gravada em tábuas de barro são poesias em homenagem a uma deusa chamada inanna adorada pela sacerdotisa

(r) sus textos escritos en lengua cuneiforme en forma de cuña grabada en tablas de barro son poesías en homenaje a una diosa llamada inanna adorada por la sacerdotisa

(s) sus textos escritos en lenguaje cuneiforme en forma de cunha grabada en tablas de barro son poesías en homenaje la una diosa llamada inanna adorada por sacerdotisa

(a) sus textos escritos en el lenguaje cuneiforme en forma de cunha grabada en tábuas de barro son poesías en homenaje a una diosa llamada inanna adorada por la sacerdotisa
