

STATISTICAL ANALYSIS OF THE FACTORS INFLUENCING THE TRANSLATION QUALITY OF THE DILMANC MT SYSTEM

Rauf Fatullayev¹, Sevinj Mammadova², Abulfat Fatullayev³

¹Institute of Cybernetics of ANAS, Baku, Azerbaijan, *rauf123@mail.ru*

²Institute of Cybernetics of ANAS, Baku, Azerbaijan, *s_mammadova@yahoo.com*

³Institute of Linguistics of ANAS, Baku, Azerbaijan, *fabo@box.az*

One of the most important problems while creation of the machine translation (MT) systems is the increasing of the translation quality and it is possible only with the comprehensive testing process. The formalizing of the traditional grammar rules and using them in the translation process is the small part of the automation of the translation process. The reason of this fact is the impossibility of the development of the grammar which covers all situations of the natural language phenomena. From this point of view, the most of time while creation of the MT system is being spent to the processing of the cases which aren't explained and impossible to explain in the traditional grammar of the natural language. Testing process yields possibility to detect such particular cases and the translation algorithms are being improved in the result of processing of such cases. Except this, the grammatical events that mostly influence the quality of the translation process are detecting and it shows to which area it is necessary to focus the next researches.

Evaluation of the translation quality of an MT system is one of urgent and interesting tasks of the development of the MT system. The evaluation is going on two directions: comparison of the translation quality of two or more MT system for the definition which MT system is better [1], and the definition of the translation quality of any MT system separately [2].

The special methodologies are developed for the evaluation of the translation quality of MT systems, so-called Machine evaluations methods (BLEU Metric, NIST, Meteor, F-Measure, Sentence level evaluation etc.) [3-8]. These methods yields possibility to evaluate the translation quality in general, but it is difficult to define the reason of the erroneous translations on the base of information got after the work of these methods.

During the testing process of the Dilmanc MT system (www.dilmanc.az) for the definition of the factors influencing the translation quality, first the set of test sentences consisting of 1000 sentences is formed. These sentences are selected in a random way from web sites and other sources which are shown in Table 1. During the sentences selection process, the sentences with the different length (number of word-forms in a sentence) are chosen. To this effect, not only English web sites, but the English versions of the Azerbaijani web sites are selected as sources of text (Some of these sources are shown in the Table 1).

Table 1

Sources of selected sentences

Source	Organization	Type of text	Number of sentences
www.bbcnews.com	BBC web site	News	40
www.voanews.com	Voice of America web site	/-/	25
www.anspress.com	ANS corporation web site	/-/	10
www.wikipedia.org	Online encyclopedia web site	/-/	33
www.microsoft.com	Microsoft corporation web site	/-/	5
www.conventions.coe.int	European Union web site	Law	33
www.un.org	United Nations official web site	/-/	84
www.unesco.org	UNESKO web site	/-/	10

www.smallbizarticles.com	Business documents, articles and news	Correspondence	65
www.businessnation	Templates of business documents and forms	/-/	65
www.wikiquote.org	Passages from belles-lettres, citations of famous persons	Belles-lettres	200
Windows Help		Documentation	90
MS Office documentation		/-/	75

Before selection process, we defined the themes on which we want to test Dilmanc MT system. The sentences have been selected on the following themes (Table 2). Number of sentences and their distribution on the themes are shown in the Table 2.

Table 2

Distribution on themes

№№	Themes	Number of sentences
1	News	200
2	Law	200
3	Documentation	200
4	Correspondence	200
5	Belles-lettres	200

Selected sentences have been translated from English into Azerbaijani by using Dilmanc MT system and the results of translation have been investigated. Results of translation have been evaluated on 3 categories: 1) Sentences translated without mistakes; 2) Sentences translated intelligible; 3) Sentences translated wrongly.

1. *Correctly translated sentences.* The meanings of the translations of these sentences correspond to the meanings of original sentences (in English) and they are correctly synthesized on the grammatical rules of Azerbaijani. These sentences compose 23% (228 sentences) of all tested sentences.

2. *Intelligibly translated sentences.* Translations of these sentences on the meanings correspond to the meanings of the original sentences, despite of some mistakes in the synthesis stage of the Azerbaijani sentence. These sentences compose 40% (404 sentences) of all tested sentences.

3. *Incorrectly translated sentences.* In the most of these sentences of this category, there are many serious grammatical errors which are made in the synthesis stage of their translations and as a result of these errors it is difficult to understand the translations of these sentences. But despite of such errors in some cases the meanings of the English sentences in the translations are not distorted. These sentences compose 37% (368 sentences) of all tested sentences.

Distribution on the categories of the classification of the evaluation results are presented in the Fig. 1 as a diagram.

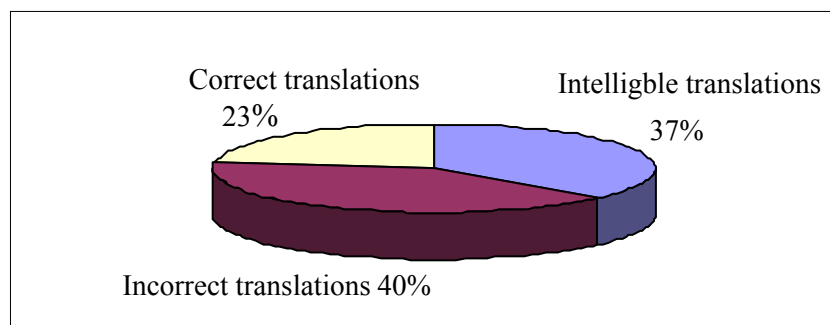


Fig. 1

Analysis of the errors shows the existence of errors both in the analysis module and the synthesis module of the Dilmanc MT system. Result of the classifying of the errors on themes is shown in the following table (Table 3).

Table 3

Distribution of the errors on themes

Cause of error		Code of error	News (200)	Law (200)	Document-tation (200)	Correspon-dence (200)	Bellesl ettres (200)	Total (1000)	
Analysis	1	Tokenization	qrf	2	3	1	0	6	12
	2	Morphological analysis	mrf	12	19	10	9	7	57
	3	Delimitation of set expressions	szb	6	8	3	15	8	40
	4	Homonymy	omo	41	44	38	37	27	187
	5	Delimitation of noun and verb phrases	ifb	12	12	3	10	3	40
	6	Homogeneous parts of the sentence	hmc	16	49	14	10	2	91
	7	Division of complex sentences	mkb	21	20	17	18	15	91
	8	Syntactic analysis	stl	9	12	12	11	8	52
Synthesis	1	Definition of clauses	cmt	17	8	13	13	15	66
	2	Synthesis of imperative sentence	emr	1	0	1	1	1	4
	3	Appositive	app	7	11	3	3	0	24
	4	Definition of the position of adverbs	zrf	6	3	7	6	3	25
	5	Polysemy	pol	9	20	12	12	19	72
	6	Synthesis of suffix chains	shk	12	16	13	15	33	89
Result	<i>Correctly translated sentences</i>			48	26	46	43	65	228
	<i>Intelligibly translated sentences</i>			90	59	78	90	87	404
	<i>Incorrectly translated sentences</i>			62	115	76	67	48	368

Apparently from Table 3, there are errors both in the analysis module of the English sentences and in the synthesis module of the Azerbaijani sentences. These errors are associated with homonymy (in analyses process) and polysemy (in synthesis process). In addition, definition of the homogeneous parts of the sentence and the division of the complex sentences into simple sentences in the analysis process and the generation of the Azerbaijani word-form and definition of the sentence's type cause more errors too (These results are visually represented in the Fig. 2).

Besides definition of the factors influencing the translation quality of the Dilmanc MT system, the results received during the researching process are used for the intensification of the scientific researches on the most of directions which are more “guilty” of the low quality of translation. In addition, researches on the solving of the ambiguity problem (homonymy) are increased and have been already given the start to the works for the solving of polysemy problem. For this purpose, about 3000 more frequently used English verbs which have polysemous meanings are selected and the researches for the development of the formal rules for the choosing their appropriate specific meaning are going on.

Researches on the elimination of the other errors detected in the testing process are continuing too: translation algorithms, lexicons and formal grammatical rules are being improved.

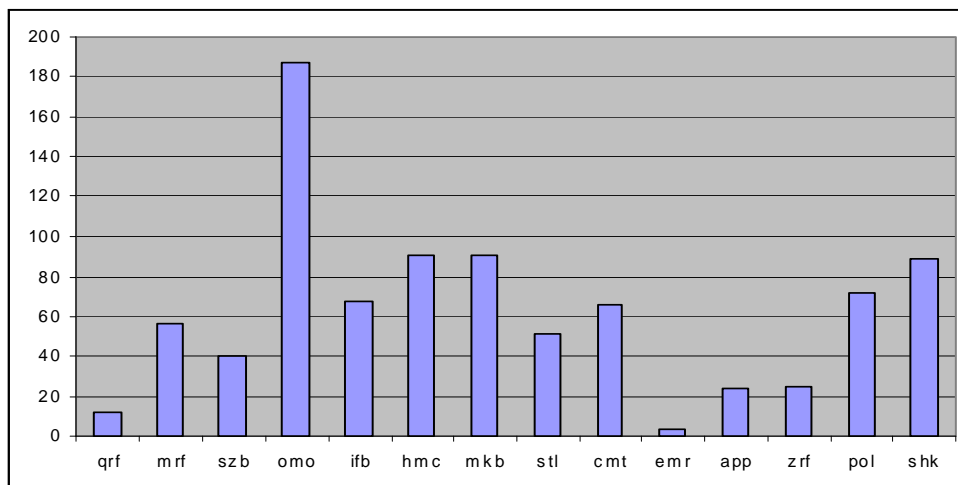


Fig. 2

Such researches for the improving of the translation quality of the Dilmanc MT system are being carried out continuously. Received results are being analyzed and on the results of this work more qualitative translation algorithms are being developed.

References

1. Blekhman M. Comparative Analysis of the Translation Quality Produced by three MT Systems. Part 1 West European Languages to English // *Lingvistica b.v. & Lingvistica 98 Inc.* (www.gutenberg.com).
2. Machine Translation Evaluation (<http://www.ics.mq.edu.au/~szwarths/MT-Evaluation.php>)
3. Papineni, K., Roukos, S., Ward, T., and Zhu, W. J. (2002). "BLEU: a method for automatic evaluation of machine translation" in *ACL-2002: 40th Annual meeting of the Association for Computational Linguistics*. July 6-12, 2002, Philadelphia, PA, USA, pp. 311-318.
4. Doddington, G. (2002) "Automatic evaluation of machine translation quality using n-gram cooccurrence statistics" in *Proceedings of the Human Language Technology Conference (HLT)*, San Diego, CA, pp. 128-132.
5. Denoul, E. and Lepage, Y. (2005) "BLEU in characters: towards automatic MT evaluation in languages without word delimiters" in *Companion Volume to the Proceedings of the Second International Joint Conference on Natural Language Processing* Jeju Island, South Korea, 11-13 October 2005 pp. 81-86.
6. Lee, A. and Przybocki, M. (2005) NIST 2005 machine translation evaluation official results (www.nist.gov/speech/tests/mt/2005/doc/mt05eval_official_results_release_20050801_v3.html)
7. Lin C., Och, F. "Automatic Evaluation of Machine Translation Quality Using Longest Common Subsequence and Skip-Bigram Statistics" in *Proceedings of the 42nd Annual Meeting of the Association of Computational Linguistics*. 21-26 July 2004, Barcelona, Spain, pp. 605-613.
8. Banerjee, S. and Lavie, A. (2005) "METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments" in *Proceedings of Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization at the 43rd Annual Meeting of the Association of Computational Linguistics (ACL-2005)*, Ann Arbor, Michigan, June 2005.