



CiyaTran™

Guidelines to Measure Increase and Decrease in Translation Accuracy

CiyaSoft MT Team
Initial Version 11/2002
Latest update: 2/2017

Major Revisions

A major revision is defined as a substantial improvement to the databases or major changes to the MT engines. A substantial improvement to the databases occurs when 10% or more of database records are improved, or when the size of the databases has increased by 10%, or the combination of improvement/increase in size have affected 10% or more of the records in the databases. Major changes to the MT engines occur when the size of the final executable increases by 5%, due to additions or enhancements to the source code. In making this comparison, modifications to the user interface, search techniques, format conversion, encoding detection and all other non-AI-related code should be excluded.

Sentence Models

Sentence Models (SM) are abstract definitions for sentences, that is the combination of POS's which convey a meaning when present in a given order. The following is an example of an SM:

[Subject] + [Object] + Verb +
[Verb Stem] + [Punctuation]

خوردم. <Verb:Verb Stem>

[I] ate.

A *word* is a combination of one or more alphabetic characters with one or more meanings, and with one or more POS's:

گرانمایه، فردا، از، و

A *sentence* is a group of one or more words, that includes at least one verb, and which satisfies (fits into) one of the SM's.

A *sentence fragment* (SF) is a group of two or more words that partially satisfy one or more SM's.

A *phrase* is a group of two or more words separated by spaces and/or hyphens, which convey a meaning not equivalent to the collection of meanings of each of the component words. Unlike an SF, phrases do not satisfy an SM.

A *Sentence Neighborhood* (SN) is a group of two or more sentences in a document-- not necessarily next to each other--that indirectly carry a concept or an idea from one sentence to the one following it, such that the receiving sentence would be meaningless if it appears alone. The following is an example of an SN:

Ali won the game. This was his third time.

علی بازی را برد. این دفعه سومش بود.

Measurement of Accuracy

Due to fuzzy logic and neural networks, any change to the code or to the databases can drastically affect quality of translation. While a change might fix a bug or enhance a particular area of weakness, it could damage another area. For this reason, Regression Testing (RT) should be run after each new build of databases or compilation of source code to validate and approve the changes.

Running RT makes sure that previously validated SN's have not been adversely affected, and if so, the affected areas are identified. The cycle of RT and corrections to the engines and databases should be repeated until RT passes without error.

In physics, when the speed of an aircraft, or the resistance of a resistor is measured, a +/- percentage of error is often given along with the measurement. In finding CiyaTran's accuracy, finding the error rate is not easy and straightforward. We should try to apply a method which could best resemble the same methods used to determine measurement errors. The following two methods should be used before each major revision:

Absolute Measurement (Manual)

A random document, with random non-domain-specific contents, from a random source (such as Iranian news agencies or the BBC), not exceeding 5KB should be used. Misspelled words, inaccurate punctuations and space misusages should be corrected before the document is translated. In addition, the following options of CiyaTran should be turned off:

User Dictionary Domain Detection

English, Penglish, Classic and Colloquial

The translated document should be manually examined and evaluated sentence by sentence. The translation of each sentence should be graded by a number from 0 to 2:

0 = Missing translation; totally incorrect translation; missing major meanings or concepts
1 = One or more fragments are translated correctly, such that one half or more of the concept can be understood; sentence is translated accurately but with problems in English grammar (tense, person, sign, ...) or punctuation.

2 = Accurate translation. (Errors due to proper nouns should be ignored.)

The number of words in the source sentence should be recorded along with the grade for the given sentence using a table:

Sentence	Size (Number of Words)	Grade (0 to 2)
1	S ₁	G ₁
2	S ₂	G ₂
3	S ₃	G ₃
n	S _n	G _n

The following formula is used by the engine to arrive at the percentage of accuracy:

$$A = (\sum S_n * G_n) * 100 / 2 * \sum S_n$$

Relative Measurement (Automatic)

The collection of over 100,000 sentences with manual translations file should be used. This file was created manually and contains 100 or more examples for each of the supported sentence models. It also contains examples of special cases and of sentences demonstrating morphological or grammar exceptions and problem areas. The percentage of accurately translated sentences in this file, which is expected to increase at each major revision, is a good indicator of how CiyaTran would behave towards random inputs. The percentage calculated by this method is usually within +/-10% of the percentage from Absolute Measurement method described previously.