***IJACSSEJournal**- International Journal of Advanced Computer Systems and Software Engineering*

E-ISSN: 2716-5256 online

A COMPARATIVE ANALYSIS OF TRANSLATION SOFTWARE USING DATA SCIENCE APPROACH TO ARABIC STATEMENTS

Murtaza M. J. Farooque[a], Buthaina S.A. Jadada[a]

[a] *College of Commerce and Business Administration, Dhofar University, Salalah, Sultanate of Oman*

* *Corresponding author: mfarooque@du.edu.om*

ARTICLE HISTORY
Received: 7 June 2021
Accepted: 5 July 2021
Published Online: 31 July 2021

KEYWORDS
Data Mining,
Data Science,
Language Translator,
Decision Tree

ABSTRACT

Readers who find difficulty in Understanding any language take the help of translator. In this era human translators are now replaced by translation software. Many students are using translation software for translating English texts into Arabic. However often these software Applications do not convey the real sense of the original text. It can be misleading several times. In this research an attempt is made to evaluate the reliability of different translation software while translating English into Arabic statements. The Machine translating is Compared with human Translation. Set of statements with three categories, named as simple, moderate and difficult levels. The comparative study has been conducted and the results have been discussed.

1.0 INTRODUCTION

In order to understand any language, a person must know the vocabulary and grammar of that language. Many a time a person, who is not comfortable with a language takes the help of a translator, who translates the spoken or written content into the language which the he understands. The translation gives an equivalent or similar meaning of the original content. With the development of the technology, the human translators are replaced by translation software applications. Translation is a difficult task for human, it remains a challenge for artificial intelligence. However often these software applications do not convey the real sense of the original text. It can be misleading several times.

2.0 REVIEW OF LITERATURE

Data Mining and machine learning has been adopted in various fields by the researchers' like for finding out customer churn in a shopping mall (M A Khan et al., 2016), digital marketing (Alraja & Malkawi, 2015; Jamil & Mohammed, 2015) (Sayyad et al., 2020) study of technology acceptance (ALraja & Aref, 2015; Alraja & Said Kashoob, 2019; Mohammed Aref & Alshahri, 2021), in various fields such as in banking sector, (Malkawi et al., 2010) and in academic for e-examination process (Uddin et al., 2016),(Shanga et al., 2018) in alternative medicine (Junaid Farooque et al., 2016) (Rasheed et al., 2021), and also in detection of faults in information system (M Aref, 2016)(Mohammed Ahmar Khan et al., 2016) (Abdul Rasheed, 2014; Abdul Rasheed & Alraja, 2015; M Aref, 2016; Mohammed et al., 2020), for age prediction (Sable, G., Farooque, M. M. J., & Rajput, 2020) for semantic analysis of opinions in social media (Murtaza M Junaid Farooque & Aref, 2019) (Alraja et al., 2020) (Alkhaldi et al., 2017) (Murtaza Mohiuddin Junaid Farooque et al., 2020) (Alraja et al., 2019) (Junaid Farooque et al., 2016). The interesting studied have been done such as information system audition (Alraja & Alomian, 2013a), empowerment and analysing environment impact (Alraja & Alomian, 2013b; De et al., 2021). The use of technology for sustainable development is utmost important (Alraja, Hussein, et al., 2021; M Aref, 2016; Sayed et al., 2008). The various examples illustrates how techniques (Khan, Sarfaraz; Saayad, Samee; Aref, 2015) supported by technology play an important role in the environment (Mohammed Ahmar Khan et al., 2016)(M A Khan et al., 2016). The study of SMEs via technological support have been studied (Alraja et al., 2020). The Use of IoT and other techniques can be adopted in human lives to make more better and better (Alkhariji et al., 2021; Alraja, Barhamgi, et al., 2021; Bou-ChaayaKaram et al., 2021)(Mohammad Ahmar Khan et al., 2019) (Mahoto et al, 2021), (Hassan et al, 2021)

3.0 METHODOLOGY

The following methodology was adopted to evaluate and compare the different Translation software in comparison with human translators.

3.1 Objective

To find and evaluate the reliability of the Software translation while translating English statements into Maintaining the Integrity of the Specifications.

3.2 Design of the Experiment

A set of 15 different statements in English were prepared out of which 5 were simple, 5 were moderate and 5 were difficult statements. These sentences were given to translation to 15 human translators to translate into Arabic. These statements were also translated using following four different translation software

1. Google Translation
2. Microsoft Translator
3. Free translator.com
4. Babylon Translator

The translated sentences were given to three human experts (E1, E2 and E3) to judge quality of translation. The results have been shown in the following tables.

Table 1A. Simple Statements (Two Factors Analysis without Replications)

Summary	Count	Sum	Average	Variance
5	5	24.73	4.94	0.0053
5	5	24.46	4.89	0.0213
4.41	5	22.80	4.56	0.0630
5	5	24.33	4.86	0.0333
4.75	5	23.43	4.68	0.0507
4.66	5	23.70	4.74	0.0580
4	5	22.80	4.56	0.0317
4.66	5	24.73	4.94	0.0053
E2	8	37	4.62	0.0367
E3	8	37	4.62	0.0367
E1	8	39	4.87	0.0535
E2	8	39	4.87	0.0357
E3	8	39	4.87	0.0357

Table 1B. ANOVA Test Results (Simple Statements)

Z	SS	df	MS	F	P-Value	F Crit
Rows	0.91	7	0.13	7.69	3.40	2.35
Columns	0.60	4	0.15	8.83	9.58	2.71
Errors	0.47	28	0.01			
Total	1.98	39				

Table 2A. Moderate Statements (Two Factors Analysis without Replications)

Summary	Count	Sum	Average	Variance
4.9	5	18.41	3.68	1.1487
4.5	5	20.26	4.05	0.4207
3.16	5	15.30	3.06	2.5402
4.91	5	16.10	3.22	1.8163
4.91	5	16.63	3.32	1.5675
4.5	5	15.91	3.18	1.1125

2.5	5	12.13	2.42	2.5039
4.16	5	19.13	3.82	0.6235
E2	8	36	4.50	0.0825
E3	8	35.4	4.42	0.0380
E1	8	22	2.75	2.3214
E2	8	19.75	2.46	0.4006
E3	8	20.75	2.59	0.3738

Table 2B. ANOVA Test Results (Moderate Statements)

Z	SS	df	MS	F	P-Value	F Crit
Rows	9.07	7	1.29	2.70	0.02	2.35
Columns	33.49	4	8.37	17.442	2.74	2.71
Errors	13.44	28	0.48			
Total	56.00	39				

Table 3A. Difficult Statements (Two Factors Analysis without Replications)

Summary	Count	Sum	Average	Variance
2.58	5	13.51	2.70	1.7040
3.83	5	18.78	3.75	0.2784
2.66	5	14.43	2.88	1.2954
5.00	5	14.61	2.92	1.6960
3.83	5	15.83	3.16	1.6776
2.25	5	12.03	2.40	1.5731
2.08	5	10.56	2.11	1.8492
4.00	5	16.40	3.28	1.3742
E2	8	32.46	4.05	0.0742
E3	8	31.46	3.93	0.0571
E1	8	20.50	2.56	2.1026
E2	8	15.50	1.93	0.4241
E3	8	16.25	2.03	0.4006

Table 3B. ANOVA Test Results (Difficult Statements)

Z	SS	df	MS	F	P-Value	F Crit
Rows	9.25	7	1.32	3.04	0.01	2.35
Columns	33.63	4	8.40	19.36	9.75	2.71
Errors	12.15	28	0.43			

Total	55.04	39

For each statement the expert instructed to give score from 1 to 5 for each of these eight criteria. It was not disclosed to the experts which statement are translated by humans and which are translated by the software. There were in all 19 Translators (15 humans and 4 software) evaluated by 3 experts.

4.0 RESULTS AND DISCUSSION

It was observed that, in case of simple statement there was very minute difference in the scores given by the evaluators to manual translation and translation software. In case of moderate and difficult sentences there are huge difference in the scores given by the evaluators.

To Further reconfirm the analysis two factor Anova without replications was done for all the three classes of the statement (simple, moderate and difficult). The result of analysis is reported in table 1A, 1B, 2A, 2B, 3A and 3B. The results of data mining are shown in table 4 and 5.

Table 4. Evaluation Matrix

Parameters	Value
Correctly Classified Instances	51 (89.47%)
Incorrectly Classified Instances	06 (10.53%)
Total No. of Instances	57 (100%)
Relative absolute error	43.00113 %
Kappa statistic	0.6042
Mean absolute error	0.1183
Root mean squared error	0.3217
Root relative squared error	88.0097

Table 5. Confusion Matrix

	Human	Software
Human	45	3
Software	3	6

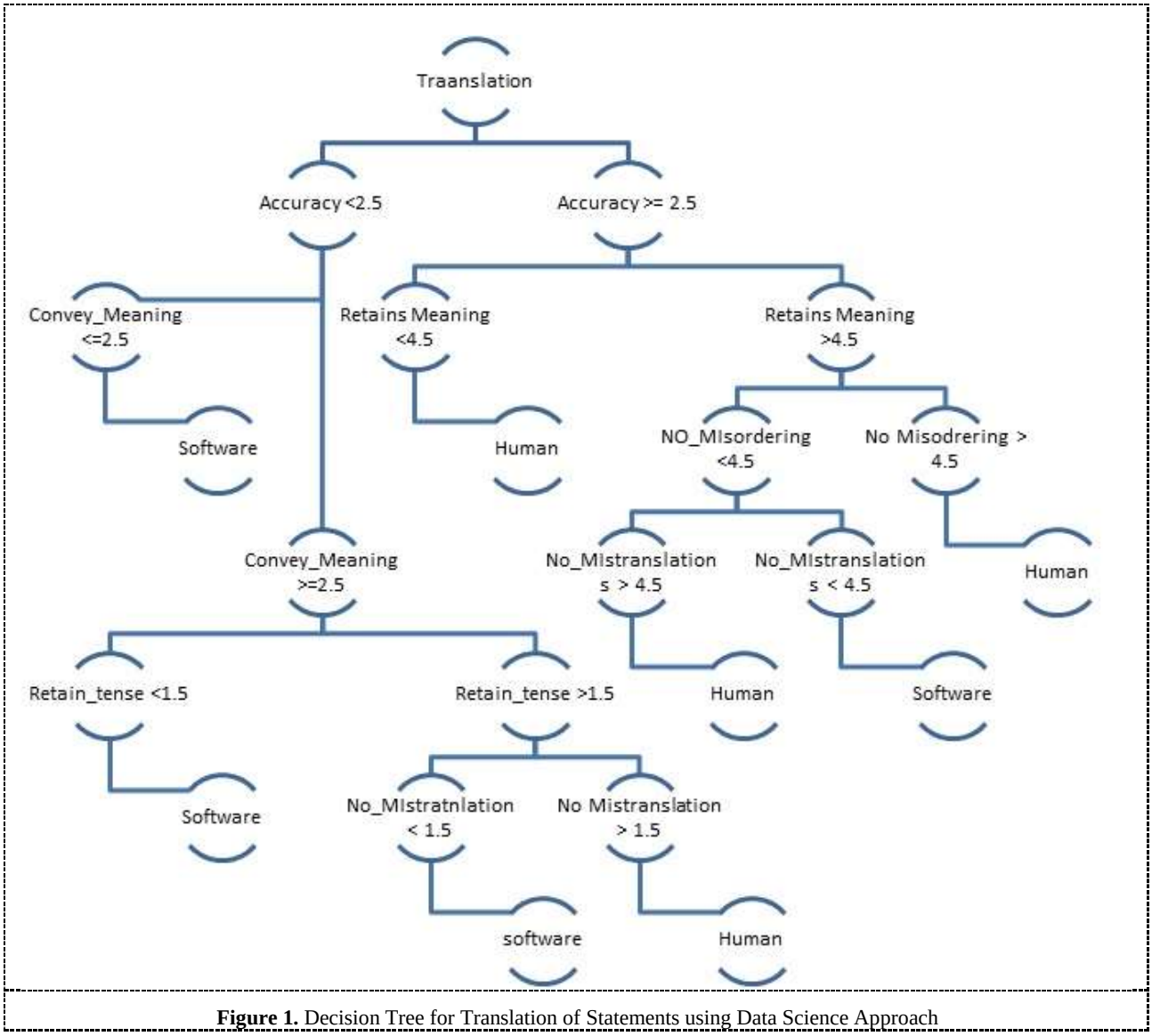

Figure 1. Decision Tree for Translation of Statements using Data Science Approach

4.1 Generation of Decision Tree

The collected data was Merged and converted into CSV format. The Criteria from table were taken as attributes, one more column was added for translator (Human or software). This was taken as class attribute. The decision tree was generated in Weka data mining software using Random-tree algorithm. The generated tree is shown in figure 1. The result of data mining was illustrated in table 4 and 5.

4.2 Recommendation

The correct meaning of sentences, it is recommended to use multiple sources. Translation software applications can be used to translate the words but not documents and sentences.

References

Abdul Rasheed, M. A. (2014). Adoption of Data Mining Technique to find the Condition of an Automobile Machine. *ISSN: 2289-7615Page13International Journal of Information System and Engineering*, *2*(1), 13–19.

Abdul Rasheed, M. A., & Alraja, M. N. (2015). Data Mining Approach To Assess Condition Of Rotating Machine Using Sound Signal. *Journal of Theoretical and Applied Information Technology*, *80*(1), 173–178. www.jatit.org

Alkhaldi, F. M., Hammami, S. M., Kasem, S., Rashed, A., & Alraja, M. N. (2017). Enterprise System as Business Intelligence and Knowledge Capabilities for Enhancing Applications and Practices of IT Governance. *International Journal of Organizational and Collective Intelligence (IJOCI)*, *7*(2), 63–77. https://doi.org/10.4018/IJOCI.2017040105

Alkhariji, L., Alhirabi, N., Alraja, M. N., Barhamgi, M., Rana, O., & Perera, C. (2021). Synthesising Privacy by Design Knowledge Towards Explainable Internet of Things Application Designing in Healthcare. *ACM Transactions on Multimedia Computing, Communications, and*

Applications (TOMM), *17*(2s), 1–29. https://doi.org/10.1145/3434186

Alraja, M. N., & Alomian, N. R. (2013a). The Effect of General Controls of Information System Auditing in the Performance of Information Systems :Field Study. *Interdisciplinary Journal Of Contemporary Research In Business*, *5*(3), 356–370. https://www.researchgate.net/publication/281465359

Alraja, M. N., & Alomian, N. R. (2013b). The Effect Of Information Technology In Empowerment Public Sector Employees: A Field Study. *Interdisciplinary Journal Of Contemporary Research In Business*, *5*(1), 805–815. https://journal-archieves32.webs.com/805-815.pdf

ALraja, M. N., & Aref, M. (2015). Customer acceptance of e-commerce: Integrating Perceived Risk with TAM. *International Journal of Applied Business and Economic Research*, *13*(2), 913–921.

Alraja, M. N., Barhamgi, H., Rattrout, A., & Barhamgi, M. (2021). An integrated framework for privacy protection in IoT — Applied to smart healthcare. *Computers & Electrical Engineering*, *91*, 107060. https://doi.org/10.1016/j.compeleceng.2021.107060

Alraja, M. N., Farooque, M. M. J., & Khashab, B. (2019). The Effect of Security, Privacy, Familiarity, and Trust on Users' Attitudes Toward the Use of the IoT-Based Healthcare: The Mediation Role of Risk Perception. *IEEE Access*, *7*, 111341–111354. https://doi.org/10.1109/ACCESS.2019.2904006

Alraja, M. N., Hussein, M. A., & Ahmed, H. M. S. (2021). What affects digitalization process in developing economies? an evidence from SMEs sector in Oman. *Bulletin of Electrical Engineering and Informatics*, *10*(1). https://doi.org/10.11591/EEI.V10I1.2033

Alraja, M. N., Khan, S. F., Khashab, B., & Aldaas, R. (2020). Does Facebook Commerce Enhance SMEs Performance? A Structural Equation Analysis of Omani SMEs. *SAGE Open*, *10*(1), 215824401990018–215824401990018. https://doi.org/10.1177/2158244019900186

Alraja, M. N., & Malkawi, N. M. M. A. (2015). E-Business adoption in banking sector: Empirical study. *Indian Journal of Science and Technology*, *8*(27), 1–7. https://doi.org/10.17485/ijst/2015/v8i27/70739

Alraja, M. N., & Said Kashoob, M. A. (2019). Transformation to electronic purchasing: an empirical investigation. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, *17*(3), 1209. https://doi.org/10.12928/telkomnika.v17i3.9390

Aref, M. (2016). Fault reporting process of business information systems. *International Journal of Economic Research*, *13*(5), 2277–2283. https://www.researchgate.net/publication/310490290_Fault_reporting_process_of_business_information_systems

Aref, Mohammed, & Alshahri, N. B. (2021). The Effect Of Introjected Perceived Locus Of Control And Trust On Intention To Use Ecommerce Applications. *Access Online Journal IJACSSEJournal-International Journal of Advanced Computer Systems and Software Engineering*, *1*(2), 16–21. https://journal.scientiaca.org/index.php/ijacsse/article/view/202

Bou-ChaayaKaram, ChbeirRichard, Naser, A., ArnouldPhilippe, PereraCharith, BarhamgiMahmoud, & BenslimaneDjamal. (2021). δ-Risk: Toward Context-aware Multi-objective Privacy Management in Connected Environments. *ACM Transactions on Internet Technology (TOIT)*, *21*(2), 1–31. https://doi.org/10.1145/3418499

De, S., Wang, W., Zhou, Y., Perera, C., Moessner, K., & Alraja, M. N. (2021). Analysing environmental impact of large-scale events in public spaces with cross-domain multimodal data fusion. *Computing*, 1–23. https://doi.org/10.1007/s00607-021-00944-8

Farooque, Murtaza M Junaid, & Aref, M. A. R. (2019). Use of Social Networking Sites in Academics: A Review. *Computer Reviews Journal*, *5*, 22–32. http://purkh.com/index.php/tocomp

Farooque, Murtaza Mohiuddin Junaid, Abdul Rashed, M. A., Khan, M. A. I., FarajAllah, T. K. B., Abdul Rashed, M. A., & FarajAllah, T. K. B. (2020). Study Of Web Presence Of Omani Media Houses In Gcc A Data Science Prespective. *Allana Inst of Management Sciences, Pune*, *10*(2), 9–13. http://www.aimsjournal.org/abstract.php?article_id=8946

Jamil, S. A., & Mohammed, A. (2015). Digital Oman-Paradigm Shift for Businesses in Oman. Leadership and Its Role in Preparing the Organization for Unprecedented Change. *2nd International Conference on Leadership and Its Role in Preparing the Organization for Unprecedented Change*.

Junaid Farooque, M. M., Aref, M., Khan, M. I., & Mohammed, S. (2016). Data Mining application in classification scheme of human subjects according to ayurvedic prakruti - temperament. *Indian Journal of Science and Technology*, *9*(13), 1–4. https://doi.org/10.17485/ijst/2016/v9i13/84658

Khan, Sarfaraz; Saayad, Samee; Aref, M. (2015). Performance Based Comparative Study of Sorting Algorithm. *International Journal of Multidisciplinary Research*, *IV*(7 (II)), 49–53.

Khan, M A, Khan, M. I., Aref, M., & Khan, S. (2016). Cluster & Rough Set Theory Based Approach To Find The Reason For Customer Churn. *International Journal of Applied Business and Economic Research*, *14*(1), 439–455.

Khan, Mohammad Ahmar, Khan, S. F., BaOmar, T. A. T. B., & Ba Omar, A. R. M. A. (2019). Development & implementation of smart vehicle over speeding detector using IoT. *Advances in Science, Technology and Engineering Systems*, *4*(2), 170–175. https://doi.org/10.25046/aj040222

Khan, Mohammed Ahmar, Khan, M. A. I., Aref, M., & Farooque, M. (2016). E-marketing a boon for SMEs of Oman. *International Journal of Applied Business and Economic Research*, *14*(1), 233–240.

Malkawi, N. M. M. A., Alraja, M. N., & Alkhayer, T. (2010). Information Systems Auditing Applied Study at Banks Listed in the Damascus Stock Exchange Syria. *European Journal of Economics, Finance & Administrative Sciences*, *21*, 119.

Mohammed, T. S., Rasheed, M. A. A., Al-Ani, M. S., Al-Shayea, Q., & Alnaimi, F. B. I. (2020). Fault Diagnosis of Rotating Machine Based on Audio Signal Recognition System: An Efficient Approach. *International Journal of Simulation -- Systems, Science & Technology*, *21*(1), 1–8.

Rasheed, M. A. A., Junaid Farooque, M. M., Acharya, H. S., Quadri, M. S. A., Abdul Rasheed, M. A., Farooque, M. M. J., Acharya, H. S., & Quadri, M. S. A. (2021). Mathematical Modelling of the Relationship between Two Different Temperament Classifications: During the Covid-19 Pandemic Mohammed. *Emerging Science Journal*, *5*(1), 67–76. https://doi.org/10.28991/esj-2021-01258

Sable, G., Farooque, M. M. J., & Rajput, M. (2020). *Pretrained*

Deep Neural Networks for Age Prediction from Iris Biometrics (First). Taylo & Francis Group, CRC Press. https://www.taylorfrancis.com/chapters/edit/10.1201/9781003079996-13/pretrained-deep-neural-networks-age-prediction-iris-biometrics-ganesh-sable-murtaza-mohiuddin-junaid-farooque-minakshi-rajput

Sayed, B. T., Jabeur, N., & Aref, M. (2008). An Archetype to Sustain Knowledge Management Systems through Intranet. *International Journal of Information and Communication Engineering*, *2*(638), 634–636.

Sayyad, S., Mohammed, A., Shaga, V., Kumar, A., & Vengatesan, K. (2020). Digital Marketing Framework Strategies Through Big Data. In A. P. Pandian, T. Senjyu, S. M. S. Islam, & H. Wang (Eds.), *Proceeding of the International Conference on Computer Networks, Big Data and IoT (ICCBI - 2018)* (pp. 1065–1073). Springer International Publishing.

Shanga, V., Samee, S., Mohammed, A., & Vengatesan, K. (2018). Enhancing Empirical approach in teaching-learning using ICT. *International Journal of Pure and Applied Mathematics*, *118*(20), 2727–2734.

Uddin, M. A., Ahmar, F., & Alraja, M. N. (2016). E-Examinations for Management Students in Oman. *International Journal of Applied Business and Economic Research*, *14*(1), 87–95. https://papers.ssrn.com/abstract=2959732

N. Ahmed Mahoto, R. Iftikhar, A. Shaikh, Y. Asiri, A. Alghamdi, and K. Rajab, "An Intelligent Business Model for Product Price Prediction Using Machine Learning Approach," Intelligent Automation & Soft Computing, vol. 29, no. 3, pp. 147–159, 2021, doi: 10.32604/iasc.2021.018944. [ISI Indexed, 1.647 IF]

S. M. Hassan, A. Alghamdi, A. Hafeez, M. Hamdi, I. Hussain, and M. Alrizq, "An Effective Combination of Textures and Wavelet Features for Facial Expression Recognition," Engineering, Technology & Applied Science Research, vol. 11, no. 3, Art. no. 3, Jun. 2021, doi: 10.48084/etasr.4080. [Scopus, ISI ESCI Indexed]