An-Najah National University
Faculty of Graduate Studies

Language Errors in Machine Translation of Scientific Biological Texts from English to Arabic: The Case of Google Translate

By
Hanan Jamal Alawneh

Supervisor
Dr. Abdel Karim Daragmeh

This Thesis is Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Applied Linguistics and Translation, Faculty of Graduate Studies, An-Najah National University, Nablus, Palestine.
2019

Language Errors in Machine Translation of Scientific Biological Texts from English to Arabic: The Case of Google Translate

By
Hanan Jamal Alawneh

This thesis was defended successfully on 16/5/2019 and approved by:

Defense Committee Members	**Signature**
1- Dr. Abdel Karim Daragmeh/Supervisor	.....
2- Dr. Mahmoud Eshreteh/External Examiner	.....
3- Dr. Hamed Abdelhaq/Internal Examiner	.....

Dedication

To the jewel of Palestine, the eternal capital, Jerusalem

To my first teachers, my dearest mom and dad for their unconditional love and their insistence to make me use the right hand instead of the left one since I was four years, believing that the right one is more blessed. Their vision turned out to be true for I got a blessing from God to finish this thesis.

To my lovely sisters, Rafeef, Raniem, Wa'ed, and Haneen.

To my one and only brother, Ra'if.

To my true friend Zainab who has been by my side since the 1st grade.

To my relatives and friends who ever concern about my thesis with love and passion.

To Google team who have always been a source of inspiration in their good and bad results. This thesis is to assist you in cutting the clouds of low quality results, so Google Translate works better.

Acknowledgement

I would like to express my gratitude to all my teachers who have nourished my intellectual life since kindergarten.

Special thanks to my supervisor, Dr. Abdel Karim Daragmeh for his valuable comments from the first letter of this thesis till the last drop of ink.

Many thanks are due to Dr. Mahmoud Eshreteh for his constructive comments concerning the linguistic part of the thesis.

Very special thanks go to Dr. Hamed Abdelhaq for his feedback regarding the maps and computer science aspects included in this work.

Last but not least, I shall never forget my dearest father and mother who have been a source of continuous support from the very beginning of this journey till the last breath.

الإقرار

أنا الموقعة أدناه، مقدمة الرسالة التي تحمل العنوان:

الأخطاء اللغوية في الترجمة الآلية للنصوص العلمية البيولوجية من الانجليزية الى العربية: ترجمة جوجل

Language Errors in Machine Translation of Scientific Biological Texts from English to Arabic: The Case of Google Translate

أقرّ بأن ما اشتملت عليه هذه الرسالة هي نتاج جهدي الخاص، باستثناء ما تمت الإشارة إليه حيثما ورد، وإن هذه الرسالة ككل، أو أي جزء منها لم يقدم من قبل لنيل أية درجة أو لقب علمي أو بحثي لدى أية مؤسسة تعليمية أو بحثية أخرى.

Declaration

The work provided in this thesis, unless otherwise referenced, is the researcher's own work, and has not been submitted elsewhere for any other degree or qualification.

Student's Name:	**Hanan Jamal Alawneh**	اسم الطالب:
Signature:	.....Hanan..Alawneh...	التوقيع:
Date:	16/5/2019	التاريخ:

Table of Contents

Subject	Page
Defense Committee Members	ii
Dedication	iii
Acknowledgments	iv
Declaration	v
Table of Contents	vi
List of Tables	viii
List of Figures	ix
List of Abbreviations	xi
Abstract	xii
Chapter One: Introduction	
1.1. Introduction	1
1.2. Scientific Translation and Machine Translation	4
1.3.Why the Biology Textbooks?	5
1.4. Problem Statement	7
1.5. Purpose of the Study	10
1.6. Significance of the Study	11
1.7. Limitations of the Study	11
1.8. Research Questions	12
1.9. Thesis Chapters	13
Chapter Two: Literature Review and Methodology	
2.1. Related Literature to Machine Translation	15
2.2. Methodology	24
2.3. Theoretical Framework	27
Chapter Three: Linguistic Discordance in Google Translation	
3.1. Introduction	29
3.2. Errors at Syntactic Level	37
3.2.1. Organization of Constituents at Phrase Level	37
3.2.2. Organization of Constituents at Sentence Level	42
3.2.3. Erroneous Shifts from Verbal to Nominal Sentences in Arabic	49
3.3. Errors at Morphological Level	51
3.3.1 Inappropriate Choice of Suffixes	51
3.3.1.1. Inflections Attached to Sub-headings	51
3.3.1.2. Inflections Attached to the Verb	53
3.3.2. Passive Constructions	53

Subject	Page
3.3.2.1. Failure to Distinguish Between the Simple Past and Passive Inflections	55
3.3.2.2. Passive inflections	58
3.3.3. Unnecessary Derivation for Certain Words	60
3.3.4. Pronouns Translation	62
3.3.4.1. Relative Pronouns Referent/s	62
3.3.4.2. Pronouns Refer to Gender Neutral Nouns	64
3.3.4.3. Phrasal Verbs Meanings along with their Gender Marked Subjects	68
3.4. Conclusion	71
Chapter Four: Cohesive Markers at Paragraph Boundaries in Google Translation	
4.1. Introduction	75
4.2. Areas of Eff./Def. of Lexical Markers in Description Paragraphs	80
4.2.1. Cohesion in Introducing Definitions	80
4.2.2. Cohesion Achieved through Scientific Terms Chains	85
4.3. Areas of Eff./Def. of Lexical Markers in Process Paragraphs	91
4.3.1. Cohesion Achieved through Word-Choice	91
4.3.2. Process-Sequence Verbs	96
4.3.3. Process Signals within Non-past Constructions	99
4.4. Areas of Eff./Def. of Lexical Markers in Causality Paragraphs	103
4.4.1. Word Collocations Deployed in Cause/Result Relations	103
4.4.2. Causal Chains	108
4.5. Conclusion	117
Chapter Five: Conclusions and Recommendations	
5.1. Conclusions	120
5.2. Recommendations	128
References	129
Appendixes	
Appendix (1): The selected texts and their translation as produced by Google Translate	136
الملخص	ب

List of Tables

Table No.	Subject	Page
Table (1)	Errors made at noun phrases level	38
Table (2)	Errors in ordering the noun phrase	40
Table (3)	Errors made at sentence level	43
Table (4)	Errors of turning a sentence to a noun phrase	49
Table (5)	Errors in treating the definite article	52
Table (6)	Errors in the selection of parts of speech	53
Table (7)	Errors in recognizing the passive construction	56
Table (8)	Errors at passive construction arrangement level	57
Table (9)	Errors at passive inflections level	58
Table (10)	Redundancy due to unnecessary repetition	61
Table (11)	Errors in assigning the correct referent	62
Table (12)	Errors in using gender markers	65
Table (13)	Errors made at pronoun level	67
Table (14)	Errors in SV agreement	69
Table (15)	Errors in relative clauses translation	85

List of Figures

Figure No.	Subject	Page
Figure (1)	The process of treating the selected data	27
Figure (2)	Processing of noun phrases	42
Figure (3)	Mapping of nominal sentences to verbal sentences	47
Figure (4)	Mapping of nominal sentences	48
Figure (5)	Mapping of sentences with verb to be	51
Figure (6)	Processing of words with similar forms in plural and present tense	55
Figure (7)	Processing of passive constructions	60
Figure (8)	The process of matching the relative pronoun with its referent	64
Figure (9)	The process of assigning gender to both the sub. and the verb	66
Figure (10)	The process of gender matching between the pronoun and its antecedent	68
Figure (11)	The process of composing the meaning of the phrasal verb "keep"	70
Figure (12)	Parallel structures used in introducing definitions	85
Figure (13)	GT failure in reproducing adjective clauses	85
Figure (14)	The lexical chain used in paragraph 2a	88
Figure (15)	Deficiency at terminological level in paragraph 2a	89
Figure (16)	Suggested translation for paragraph 2a	89
Figure (17)	The process of identifying descriptive paragraphs based on the lexical markers	91
Figure (18)	GT inexact translation for the process in question	95
Figure (19)	GT failure in identifying the verb as a sequence transition	97
Figure (20)	Suggested translation for paragraph 3a	99
Figure (21)	Suggested translation for paragraph 4a	102

Figure No.	Subject	Page
Figure (22)	The process of identifying process paragraphs based on the lexical markers	103
Figure (23)	Errors in employing collocations	105
Figure (24)	'Because' along with its complement in Arabic	106
Figure (25)	Errors in pronoun-referent resolution	107
Figure (26)	GT failure in deploying "wa"	107
Figure (27)	Suggested translation for paragraph 5a	108
Figure (28)	GT failure in reproducing if-structures	111
Figure (29)	Suggested translation for paragraph 6a	111
Figure (30)	The process of identifying causality paragraphs based on the lexical markers	113
Figure (31)	Suggested translation for paragraph 7a	116

List of Abbreviations

AJ : Adjective

AT : Article

Aux. : Auxiliary

AV : Adverb

C : Complement

C/R : Cause/Result

CJ : Conjunction

Def. : Definition

Des. : Description

DT : Determiner

Eff./def. : Efficiency/deficiency

FEMTI : Framework for the Evaluation of Machine Translation in the ISLE

GT : Google Translate

ISLE : International Standards for Language Engineering

L.M. : Lexical Marker

MTE : Machine Translation Evaluation

NLP : Natural Language Processing

NN : Noun

Obj. : Object

PN : Pronoun

Pro. : Process

SL/TT : Source Language/ Target Language

Sub. : Subject

TO : Infinitive Marker to

VB : Verb 'BE'

VM : Modal verb

VV : Verb

Language Errors in Machine Translation of Scientific Biological Texts from English to Arabic: The Case of Google Translate

By
Hanan Jamal Alawneh
Supervisor
Dr. Abdel Karim Daragmeh

Abstract

Machine translation has planted its roots deeply in research domains since it becomes the first aid for survival in this era of "globalization". Thus, the present research explores the areas of efficiency/deficiency in Google Translate performance in scientific biological texts translation from English to Arabic. More specifically, the research aims to test GT performance at two levels: sentence and paragraph levels. Thus, Catford's translation shifts (1965), Halliday and Hassan's model of cohesive devices (1976) and types of paragraphs frequently used in scientific texts are the main tools used to judge GT output. Finally, the researcher attempts to propose solutions for the errors encountered to enhance GT performance in this particular text type to help GT produce translations with high accuracy rates.

Chapter One

Introduction

1.1. Introduction

The 21st century can be best described as a competitive marketplace with two main competing forces. On the one hand, there are companies that work hard in order to put the best products in the hands of their consumers. On the other hand, there are clients who struggle to find an optimal product that both eases their life and saves them time and effort. This leads machines to become like shadows of human beings; if one wants to talk to someone who is far away from him/her, then s/he has to use a machine which is the cell phone in order to communicate with that person; or if one wants to move from one place to another, s/he has to use a car which is also a package of machines. Similarly, if a student, a mother, a father, a tourist, or a beginner translator, wants to learn to read a paragraph, to check the pronunciation, the spelling of certain words, or to translate a short excerpt, a word, a phrase or even a text of whatever kind from one language into another, s/he often uses a machine to perform such tasks. Thus, such trends reflect the fact that machine translation has become a necessity for living in the modern world.

There is a plenty of choices among machine translation software that users often benefit from such as: Bing Translator, which was introduced by Microsoft in 2012 and, provides a multi-lingual translation service as well as Babylon which played an important role in machine translation from and

to Arabic through developing dictionaries that contain acronyms and abbreviations. In addition, there is voice translation software which provides customers with voice to text or text to voice translations by turning a certain message into a unit of translation then producing written or oral translations for it according to the customers' needs. A clear example of voice translation software is Google Translate (GT). This software provides its users with voice translations. Thus, all they need to do is to click on the button "speak" and a written translation of their speech will appear on the screen. Moreover, GT provides translations among 103 languages with over than 200 million users daily (Wikipedia, 2018). Therefore, GT has become the most fashionable, trendy and easily accessible machine nowadays for translation tasks.

However, this software can sometimes be misleading since it is a machine that depends heavily on word-recognition and pattern-matching between the components of the input and the likely equivalence for that input in its translation memory. This framework of translation action was explained by GT team (2012) who stated that:

> When Google Translate generates a translation, it looks for patterns in hundreds of millions of documents to help decide on the best translation. By detecting patterns in documents that have already been translated by human translators, Google Translate can make intelligent guesses as to what an appropriate translation should be. This process of seeking

patterns in large amounts of text is called “statistical machine translation" (Retrieved from: https://urlzs.com/xtgSo).

In other words, the truth that GT is both fast and economical cannot be denied; however, when it comes to its accuracy, the translation product can be inaccurate, incomprehensible and often misleading. It is quite evident that GT comprehension ability is still inferior to human translation. This deficiency is due to the fact that GT has to deal with many languages with different linguistic systems. Thus, Aiken and Balan (2011) stated that: "Although Google Translate provides translations among a large number of languages, the accuracies vary greatly... translations between European languages are usually good, while those involving Asian languages are often relatively poor" (Retrieved from: https://urlzs.com/aPm1L).

This proves that GT is still in its initial stage and thus the door is still open to improve it; the evaluation of its performance is deemed a vital stage in improving the performance of the translation software. Therefore, this thesis aims at evaluating GT performance and pinpointing the problems that it may encounter while translating texts from English into Arabic, particularly scientific biological texts taken from Biology 1 textbook which is taught at the Faculty of Science at An-Najah National University for 1st year students; the machine translation users in this case are 18-19 years old. The texts contain chapters on The Chemical Context of Life, Water and the Fitness of the Environment, The Structure and Function of Large Biological Molecules, An Introduction to Metabolism, Cellular

Respiration, The Cell Cycle, Mendel and the Gene Idea and From Gene to Protein. Finally, the thesis attempts to recommend solutions for the errors encountered in the translation action to enable software developers to enhance their translation program. Such solutions help to reach high accuracy levels when translating such type of texts since they are considered an example of a controlled area covered under the umbrella of scientific genre.

1.2. Scientific Translation and Machine Translation

A scientific text is considered one of the writing modes embedded within the general term called 'scientific genre' for Hatim and Munday define the word 'genre' as "a conventionalized form of speaking or writing which we associate with particular 'communicative events'. Participants in these events tend to have set goals, with strict norms regulating what can or cannot be said within the confines of given genre settings" (2004, p.88). That is to say, the scientific genre is a well-established mode because it employs a set of agreed upon standards and textual norms that regulate the use of both language and message-building within the texts that conform to such genre. In other words, the scientific genre is tied with a language that is characterized by "impersonal style, simpler syntax, use of acronyms, and clarity (Ilyas, 1989, p.109). Accordingly, when it comes to translation, scientific translation is considered to have an informative function. Byrne (2006) stated: "scientific translation primary goal is to deliver scientific information; it aims at presenting well expressed information, that may be

used easily, properly and effectively" (as cited in Soualmia, 2009, p.21). Moreover, Soualmia stated: "Scientific translation is defined as the method employed to help organize thought, procedures and then come into clear, faithful and reliable results, free of subjectivity and personal involvements" (2009, p.19).

However, translators may face a difficulty in translating scientific terms and constructions since Zinaser states that: "Every profession has its growing arsenal of jargon to fire at the lay man and hurls him back from its walls" (1976, p.15). Thus, translators may resort to different procedures while translating texts such as: transliteration, borrowing, or providing footnotes. Thereby, with regard to GT, the situation may be much more challenging for the machine in question may not enjoy enough level of recognition to decide upon which procedure to use. Thus, this may result in low quality translations. This echoes Parikh's words who stated: "machine translation rarely reaches accuracy levels above 70%, while a human translation almost always produces accuracy levels above 95%" (2012. Retrieved from: https://urlzs.com/STgBk). Thus, the present research aims to test the extent GT adheres to the norms associated with the scientific language while translating excerpts taken from scientific biological textbooks.

1.3. Why the Biology Textbooks?

The reasons behind choosing the biology textbooks for students in their 1st year in biology specialization is: First of all, scientific texts are

very challenging; they are "a good example of the most challenging text type … these texts often present information that is conceptually rich but also conceptually dense and abstract. They use terminology that is unfamiliar to many students … using language in ways that students do not encounter in their reading of fictional and narrative texts" (Palincsar, 2013. Retrieved from: https://urlzs.com/ZBv4f).

Secondly, scientific biological texts deal with terminologies and processes related to everyday life activities like sleeping, eating, etc unlike the other branches of science like chemistry and physics which are basically about numbers and statics that are close to the sign language such as the mathematical calculations (+ / - / *); those later texts contain minimum text and therefore they can be easily understood by looking at the symbols. In comparison, the biology text is basically about concepts, descriptions, and terms. Thus, it can be hypothesized that GT can work better with biology.

Finally, students at this stage (1st year in specialization) will take compulsory courses that usually contain introductory and basic concepts about biology in English. These courses will serve as a repertoire for them later on in their specialization. This makes it vital for students to make sure that they understand the ideas and get the accurate equivalent. Therefore, 1st year students who are majoring in biology may choose GT to translate certain texts and terms from English to Arabic to understand the information in their textbooks. In other words, students need to cope with

the language of science which, in turn, uses the English language to express the new experiments/studies in biology; this tendency to use machine translation supports the report issued by GT in 2016 which states that more than 500 million people use GT around the world and that the Arabic language is one of the most widely used languages in the application with more than 100 billion words a day.

1.4. Problem Statement

This research will be concerned with the mistranslations performed by GT while translating scientific biological texts from English to Arabic. When it concerns machine translation, issues related to features and functions may create a challenge for the machine in question, i.e., GT. Thus, this study will explore problems including those named by Wisniewski, Kubler and Yvon (2014) such as: "lexical errors, morphological errors, syntax errors, semantic errors, format errors… " (Retrieved from: https://urlzs.com/CAQbk).

These types of errors may affect the quality of scientific texts since these texts may contain different types of paragraphs including: description paragraphs which aim at describing concepts/objects, process paragraphs that mark the sequence of certain biological processes or causality paragraphs which explain the cause/result of particular phenomena. Thus, in all cases, scientific texts must meet four standards: Syntax, Morphology, Terminology and Cohesion/Naturalness.

First of all, 'Syntax'. In such kind of texts, syntactic structures must guide the machine to one and only one meaning. This means that ambiguity which may threaten the precision of the text is not welcomed in scientific translation so there must be no room for structural ambiguity in the resulting translations made by GT.

The second standard is 'Morphology'. It examines certain morphemes attached to certain words to help the software to get the exact meaning such as connectors, negation, tense and number. However, GT may not benefit from these morphemes to process and understand the stated facts directly since it depends on its intelligent guesses to connect the parts of the text together.

The third standard is 'Terminology' which refers to domain specific terms used heavily in scientific texts. Such terms create a challenge for both GT and students to understand because these technical terms "have one or many meanings in everyday language" while having a different, peculiar and precise meaning in scientific texts (Ali and Ismail, 2006. Retrieved from: https://urlzs.com/FnT8d).

The fourth standard is 'Cohesion/Naturalness' of the resulting translation. In other words, scientific texts vary in the cohesive devices they employ to connect the ideas in a coherent way since they may be descriptive, persuasive, or informative. All these functions aim at putting the information in the hand of the readers without being redundant or consuming much time/effort to get the intended meaning. However, in the

case of GT software, these standards may be demanding. This echoes Al-Asali's words who points out that "the real problem with today's MT systems … is that they do not achieve the appropriate interpretation of certain parts of the source text (ST), which may depend, in one way or another, on the appropriate comprehension of the devices controlling them" (2000:xix).

Therefore, the study will explore problems related to text type; the unique nature of scientific texts leads to machine translation problems when used by students. Thus, the research focus will be on: Firstly, the mistranslations made by GT in areas of describing particular biological processes or biological terms at sentence level including phrasal constituents; e.g. the mistranslation of "inheritance law" in the sentence "Mendel used the scientific approach to identify two laws of inheritance" into "قانون الميراث/الارث" which is a phrase that is used to refer to the process of genes movement from parents to their offspring and it has nothing to do with the concept that refers to the possessions' of a dead person. Secondly, the mistranslations of scientific texts at paragraph level according to the format of such type of texts from English into Arabic such as translating the English text: "water is an excellent solvent for many substances because of its polar nature. Polar substances and ions dissolve in water because opposite charges are attracted to the appropriate ends of water. Strictly hydrophobic molecules, including most lipids, do not mix well with water." into Arabic as:

"الماء هو مذيب ممتاز للكثير من المواد بسبب طبيعته القطبية. تذوب المواد والأيونات القطبية في الماء لأن الرسوم المعاكسة تنجذب الى نهايات المياه المناسبة. لا تختلط الجزيئات الكارهة للماء بشدة ، بما في ذلك معظم الدهون ، بشكل جيد مع الماء."

The Arabic paragraph contains errors such as the underlined short sentence that contains the passive construction "are attracted". The correct translation of the English structure is "الأقطاب المتعاكسة تُجذب الى نهايات المياه المناسبة". In other words, these "opposite charges" do not move by themselves; instead they are moved by an external force. This meaning is not expressed correctly in the Arabic text because the verb "تنجذب" is active not passive. There are also mistranslations of certain terms, despite the context makes their meanings clear such as: "opposite charges" which is translated as "الرسوم المعاكسة" instead of "الأقطاب المتعاكسة (سالبة وموجبة)" and "hydrophobic molecules" which means "نافرة للماء" not "كارهة" since the latter is not a scientific term.

1.5. Purpose of the Study

The present research aims at examining the translation problems GT encounters when translating scientific biological texts found in Biology 1 textbook. It has been observed that GT makes errors in areas such as syntax, morphology, terminology, among others, (Hannouna, 2004, p.450). Thus, the research attempts to detect these errors at sentence and paragraph levels; the researcher will highlight the areas of eff./def. in GT performance to provide useful input about the quality of translation. Then, the researcher will propose solutions for the errors to enhance GT performance. Such

outcomes can be useful for the translation software developers and users alike since Ulitkin (2011) stated that: "despite their efficiency and outlooks, the translation software and electronic means cannot replace the human translator and guarantee high-quality translations". He believes that a good translation is a result of the combination between the translator's talents and experience on the one hand and the electronic technologies on the other hand; therefore, users cannot only depend on the use of machines in translation (Retrieved from: https://urlzs.com/3US1f).

1.6. Significance of the Study

This research is of great importance for it deals with the most widely used machine translation system, GT. Thus, it aims at identifying the challenges GT encounters in scientific biological texts translation; it highlights the areas (syntax, morphology, terminology, cohesion) which are best treated by GT and the ones produced in low quality. Thus, it ends at suggesting recommendations to enhance GT performance concerning the level at which the users/software developers can best use/improve GT in this particular text type translation.

1.7. Limitations of the Study

The present research is limited to GT language errors found in scientific biological texts translated form English to Arabic only. Yet, it does not tackle language errors committed by other machine translation programs. In addition, the present research will be concerned with testing

GT performance at both sentence and paragraph levels since evaluating GT at text level is beyond the scope of the present research. The researcher observed that GT commits errors at both sentence and paragraph levels; thus, it is hypothesized that GT may not perform well at text level since sentence and paragraph levels serve as the basic building blocks of any text. In other words, communicative texts cannot function without strong blocks. Thus, it would be better to evaluate GT performance at smaller levels at first to pave the road for GT evaluation at text level. Moreover, some figures and drawings might be inserted within the text to clarify the information being presented. Thus, users might resort to input GT with only separated/short paragraphs in lieu of longer texts to avoid such visual representations. Finally, the research focus will be on the external characteristics of GT, in particular, eff./def. areas in GT performance regardless of its internal characteristics which include speed, storage, or cost.

1.8. Research Questions

In attempting to evaluate GT performance and investigate the translation problems encountered in scientific biological texts translation, it is important to answer these questions:

1. What are the grammatical errors made by GT when translating biological texts at sentence level?

2. Which cohesive markers are mishandled/mistreated by GT and which of these are reproduced correctly when translating biological texts at paragraph level?

3. What are the possible explanations for making ill-formed translations/inadequate system performance?

4. What are the main recommendations for improving machine translation in this particular text type?

1.9. Thesis Chapters:

The present thesis contains five chapters; the sequence is summarized below.

Chapter One is devoted to introductory information that describes the state of technology in the 21st century in general and machine translation in particular. The chapter also includes: the problem statement, the purpose of the research, the significance of the research, the limitations of the research, the research questions. Finally, the chapters of the thesis.

Chapter Two includes literature review about machine translation and its development. In addition, the methodology, data collection, and the framework in which the data will be treated along with in the present research.

Chapter Three will present the data through analyzing and comparing the source text/input with Google translation/output based on

Catford’s translation shifts (1965). Thus, the researcher will identify the errors made by GT at sentence level then recommend solutions for them.

Chapter Four will discuss the errors made at paragraph level by GT based on Halliday and Hassan's model of cohesive markers (1976) which includes both grammatical and lexical devices, beside the types of paragraphs frequently used in scientific texts. Finally, the chapter will present suggested solutions for the encountered challenges.

Chapter Five presents the conclusions; it is expected that the thesis presents conclusions regarding the quality of GT output, the reasons of its failure, and the effects of mismatches between the source text and GT output. The research also attempts to suggest recommendations for further research that could help in enhancing GT performance in scientific biological texts translation.

Chapter Two

Literature Review and Methodology

2.1. Related Literature to Machine Translation

Machine translation life-cycle is best likened to a baby who starts taking his/her first steps through leaning on one couch and another to follow his/her parents footprints. Yet, once that child balances his/her body and masters the walking skill, his/her parents can hardly catch and control his/her movement. Similarly, machine translation took its early steps after World War **II** drawing on two main factors: First, the invention of the first computer in the 1950s and the desire to benefit from this invention in specific domains. Second, the rising tensions between the two main forces at that time: the United States and the Soviet Union (Russia now) manifested in the Cold War. Accordingly, the American government developed the first version of machine translation to break on the Russian communications and decode their military plans. Thus, machine translation early days were of one function, that is military (Errens, 2019. Retrieved from: https://urlzs.com/95Af3).

However, this mono-function of machine translation started to fade with the need of global communication. In other words, machine translation started to impose itself on civilian domains "because of globalization, the rising of international trade, the expansion of mass media and technology, the increase of migration, and the recognition of linguistic

minorities" (Al-Khawalda, Al-Oliemat, 2014. Retrieved from: https://urlzs.com/XKfcN). In other words, machine translation shifts from being restricted to military interests to serve a number of civilian functions including translation of texts between different languages. Accordingly, the new discipline of "computational linguistics" came into light. This new discipline is defined as:

> a subfield of linguistics and computer science that is concerned with computer processing of human language. It includes automatic machine translation (MT) of one language into another, the analysis of written texts and spoken discourse, the use of language for communication between people and computers, computer modeling of linguistic theories, and the role of human language in artificial intelligence (AI) (Hannouna, 2004, p.53).

Consequently, researchers tend to reflect on this new branch of linguistics through forming linguistic models about machine translation including: the way the machine works, error-tracking or eff./def. identification, accuracy levels of particular text types, etc. Such studies are carried out to see whether the machine could replace human translators, aid them, assist translation theorists who seek to test hypothesis using a particular translation program or software developers who want to promote their translation programs and impress the end users to trust their product (Hatim and Munday, 2004, p.120).

Thus, machine translation life is divided into two main stages: The first generation and the second generation. The former generation, usually referred to as the direct approach, refers to the early days of machine translation where the machine was fed with only a limited number of linguistic rules of each language and a bi-lingual dictionary. Thus, this indicates that machine translation was merely word-for-word replacement at first. In other words, the translation action is done directly between the languages in question provided that the machine has both the necessary rules and vocabulary.

However, this generation received criticism since translation is not just word for word substitution. Yet, it is an art of crafting texts. This echoes Somers and Hutchins's words who stated: "From a linguistic point of view, what is missing is any analysis of the internal structure of the source text, particularly the grammatical relationships between the principal parts of the sentences" (1992. Retrieved from: https://urlzs.com/KpTMD). Moreover, the first generation input was limited to small levels only including: words, phrases and sentences which users cannot edit their output translation. In other words, this direct approach derives its name from the fact that it does not allow the users to interact with the machine for the translation action is done only through literal translation between the source text and the target text/output. This echoes Craciunescu, Gerding-Salas, Stringer-O'Keeffe's words who state: "The first versions of machine translation programs were based on detailed

bilingual dictionaries that offered a number of equivalent words in the target language for each word listed in the source language, as well as a series of rules on words order" (2004. Retrieved from: https://urlzs.com/GDTS5). Moreover, Somers and Hutchins maintain that this approach results in "frequent mistranslations at the lexical level and largely inappropriate syntax structures" (1992. Retrieved from: https://urlzs.com/KpTMD). Such errors may affect the meaning and take the source text away from its intended meaning.

Accordingly, the criticism thrown at the first generation of machine translation led to the evolution of the second generation. In other words, machine translation has developed and it started to view the translation action as a process done along three dimensions: First, the machine decodes the meaning of the ST. Second, it re-encodes this meaning in the target language. In other words, decoding the meaning of the ST in its entirety requires that the machine interprets and analyzes all the elements of the text and transfers them into the target language. Thus, "this process requires in-depth knowledge of the grammar, semantics, syntax etc of the source language and the same in-depth knowledge is required for re-encoding the meaning in the target language" (Dubey, 2013, p.18).

Third, this indirect approach started to allow the end user to interact with the machine. In other words, the direct relationship which holds between the input and the output in the first generation is broken by the interaction of the end user who started to take place in the second

generation since the machine starts to "ask the user to supplement its linguistic information, requesting confirmation of its decisions, or selection from among alternatives" (Somers and Hutchins, 1992. Retrieved from: https://urlzs.com/KpTMD). Errens asserts that "to meet that demand and clean up its data, Google Translate has an improvement function that led users enter suggestions for smoother translations" (2019. Retrieved from: https://urlzs.com/95Af3).

Such mutual procedures between the machine and the end user enhanced the machine performance, in particular, in areas where the tested text type/genre is limited to a set of norms. This echoes Austermuhal's words who states: "the simple but effective system depends on careful pre-editing and the adoption of very controlled lexis and syntactic structures" (2001:163-4 as cited in Hatim and Munday, 2004, p.117). Thus, success stories started to flourish including the well-known story of the Canadian METEO system which was accomplished at the University of Montreal; it translates the weather bulletins automatically from English to French and vice versa for the Metrological Service of Canada. In other words, weather forecasts have specific norms that the machine could easily recognize including: single words and fixed expressions such as: sunny, low 7, wind southwest 10km/h. Moreover, Fromkin and Rodman state: "the greater recognition of the role of syntax and the application of linguistic principles over the past forty years have made it possible to use computers to translate simple texts grammatically and accurately between well-studied languages"

(1995, p.473). This proves that the machine could produce high accuracy rates in cases where the domain is specific enough.

Moreover, this generation yields a number of fruitful concepts including: statistical machine translation and neural machine translation. Pestove states that the former "is based on the idea that if you feed a computer … enough data in the shape of parallel texts in two languages, it will be able to spot and recreate the statistical patterns between them. While the latter means "the source text is the set of specific features. Basically, it means that you encode it, and let the other neural network decode it back to the text, but, in another language". It is a new discipline and it is limited to nine languages only. However, neural machine translations "are helpless when the word is not in their lexicon" (2018. Retrieved from: https://urlzs.com/enQrN).

Consequently, Google team launched their Google Translator Toolkit in 2009. This Toolkit is considered as a platform where translators upload texts and submit them for translation. Thus, Google resorts to use bilingual “parallel corpora”. This corpora consists of a pair of texts, where one text is a translation of the other. This interaction between GT and human translators led Google team to develop their “phrasebook” where users can save their translations. Thus, they started to enjoy the freedom to access their favorite translations of certain phrases and texts. This framework of GT is explained as:

> To translate a text, Google Translate search different documentaries to find the best appropriate translation pattern between translated texts by human. This pattern searching is called SMT. Consequently, the quality of Google Translate depends on the number of human translated texts searched by Google Translate … SMT uses a bilingual text corpora which is a database of the sentences in both source language and target language. A large group of sentences translated from for example English to Persian will be provided for the machine to calculate the probability of the words. If for instance a word like X has probability 75% to be translated into Y, then it will choose Y as the translation of X (Karami, 2014).

However, such new concepts do not indicate that the machine would replace human translators since a lot of research has been done on machine translation including GT. Yet, most of the attempts were sentence level focused and of randomly selected domains. For example, Key mentions types of errors committed by machine translation including: "words with multiple meanings, sentences with multiple grammatical structures, uncertainty about what a pronoun refers to, and other problems of grammar" (1980/2003 as cited in Hatim and Munday, 2004, p.116). In addition, Al-Khawalda and Al-Oliemat (2014) tested GT in translating twelve sentences with different temporal references from English to Arabic. They conclude that GT is confusing for non native English

speakers when it comes to temporal signals (Retrieved form: https://urlzs.com/XKfcN).

Moreover, Al Shehab (2013) tested GT in translating six legal sentences from English to Arabic. Thus, he noted that GT could achieve partial equivalent yet it commits errors in translating the archaic English terms, the passive voice and the modal "shall". Such researches end with no suggested solutions for the errors being identified (Retrieved from: http://www.eajournals.org). In other words, there are errors that are still committed but systematic research on a specific genre may yield fruitful results which may enhance GT performance through developing lexicons containing only technical terms and constructions for the domain in question to reduce problems related to word-choice. In additions, post editing processes could be reduced through minimizing keyboard press rates since Craciunescu et al. (2004) state: "when translation tasks are repeated, … keyboard use can be reduced by as much as 70% with some texts" (Retrieved from: https://urlzs.com/GDTS5).

In connection with machine translation at levels larger than the single sentence, only a handful number of researches attempted to test GT competency in translating long stretches of language between English and Arabic. ElShiekh (2012) examined GT performance at text level. He selected three genres/disciplines which are: advertisement, Koranic and literary texts. Yet, ads contain single words and phrases for they aim to be short, persuasive and eye catching instead of longer paragraphs or texts. In

addition, literary and koranic texts are loaded with emotive words that may pose a difficulty for GT (Retrieved from: http://dx.doi.org/10.5539 /ells.v2n1p56). Abdulhaq believes that "machine translation can handle the parole part of language but it can never master the langue part" (2016, p.8).

ElShiekh classified the errors made by GT at sentence level such as: transliteration and mismatches of polysemous words; however, he neglected paragraph and text levels. In other words, he did not explore the idea of cohesive markers at paragraph and text levels even though his study promised to focus on text-level (2012. Retrieved from: http://dx.doi.org/ 10.5539/ells.v2n1p56). Al-Samawi notes the shortage of studies at paragraph or text levels: "most of the previous studies that tried to use error analysis in machine translation research were at the level of the single word or phrase. Like a rare bird, research on errors of machine translation at the text level may not be easy to find, especially in Arabic English" (2014. Retrieved from: https://urlzs.com/74KMz).

However, Hatim and Munday state: "at present, there is a limited possibility of concordancing the search results or of configuring the search to select the specific text types or genres that are of interests" (2004, p.120). Accordingly, this research will take a step forward in highlighting the norms frequently used in scientific biological texts as a branch of scientific translation to test GT performance in this text type then identify the areas of eff./def. in its output for the features associated with scientific texts may do the mission promising. In other words, Craciunescu et al.

(2004) state: "machine translation is most useful with texts possessing the following characteristics: First, "Terminological homogeneity" which means that the meaning of terms does not vary. Second, "Phraseological homogeneity". It means that the ideas or actions are expressed or described with the same words. Third, short, simple sentences: these increase the probability of repetition and reduce ambiguity" (Retrieved from: https://urlzs.com/GDTS5).

Moreover, Errens states: "It follows that for now, MT delivers best results with scientific and technical writing, anything that adheres more strictly to formulas. Wherever the use of language deviates from standard, where it is more colloquial or artistic, MT falters" (2019. Retrieved from: https://urlzs.com/95Af3). Accordingly, such requirements are available in scientific biological texts. Thus, this will make it easy for the researcher to test the areas of def. in GT performance which prevents it from reaching high accuracy levels then suggest solutions for those defects to enable GT to reach high accuracy rates as much as possible.

2.2. Methodology

This research will adopt the qualitative approach in analyzing the selected data; so to collect relevant data, the researcher uses three successive steps. The first step is the translation of the biological texts using GT. The researcher decides to use ten texts which are taken randomly from the biology textbooks, particularly the Biology 1 textbook that is currently used at the Faculty of Science for 1st year students. The researcher

selects ten texts to reach fair conclusions regarding GT performance and emphasis the fact that the errors committed are not just a coincidence; instead, they serve as indicators that there are serious defects in GT translation program. The texts are technical and they cover topics like the micro/organisms, particularly The Chemical Context of Life (p.1-4), Water and the Fitness of the environment (p.5-8), The Structure and Function of Large Biological Molecules (p.16-31), An Introduction to Metabolism (p.59-67), Cellular Respiration (p.68-78), The Cell Cycle (p.91-99), Mendel and the Gene Idea (p.108-117), From Gene to Protein (p.132-140). These topics contain various biological terms, descriptions and processes expressed in different syntactic structures such as: active and passive constructions, present tense forms, if- structures, etc.

The second step is the examination of the resulting texts; this includes the classification of errors at two levels:

Sentence Level: The research aims to start the evaluation with smaller units such as sentences, then moves gradually to longer stretches of language such as paragraphs. Thus, chapter three traces the recurrent errors made by GT at sentence level in the translation of the selected texts through analyzing them and categorizing the errors in order to pinpoint the semantic shifts that may result and alter the meaning of the scientific text. In other words, the research focuses on both the errors and the extent to which those errors affect or hinder the level of comprehension in each single sentence.

Paragraph Level: The research, in chapter four, aims at testing GT competency in deploying cohesive devices in English to Arabic scientific texts translation at paragraph level. The researcher identifies the paragraph types frequently used in scientific text including: definition paragraphs, process paragraphs and causality paragraphs. Thus, different types of paragraphs taken from the ten texts are inputted into GT to be converted to Arabic to carry out the evaluation. The paragraphs express biological information related to inheritance, water, cell division, enzymes and gene expression.

The researcher examines these levels according to the features of scientific texts as a 'normative genre'. This genre includes universal features of scientific texts such as: technicality which refers to domain specific terms and structural clarity at the sentence/text level which includes issues such as: active and passive constructions and pronoun reference. The researcher also considers the presence of all functional/ morphological items such as connectors that show time, cause, etc in the selected texts.

In the final step, based on GT performance at the above mentioned levels, explanations are given for each type of errors along with suggested solutions to enhance GT performance in this particular text type. Figure (1) gives the areas where GT errors may occur in the translation action.

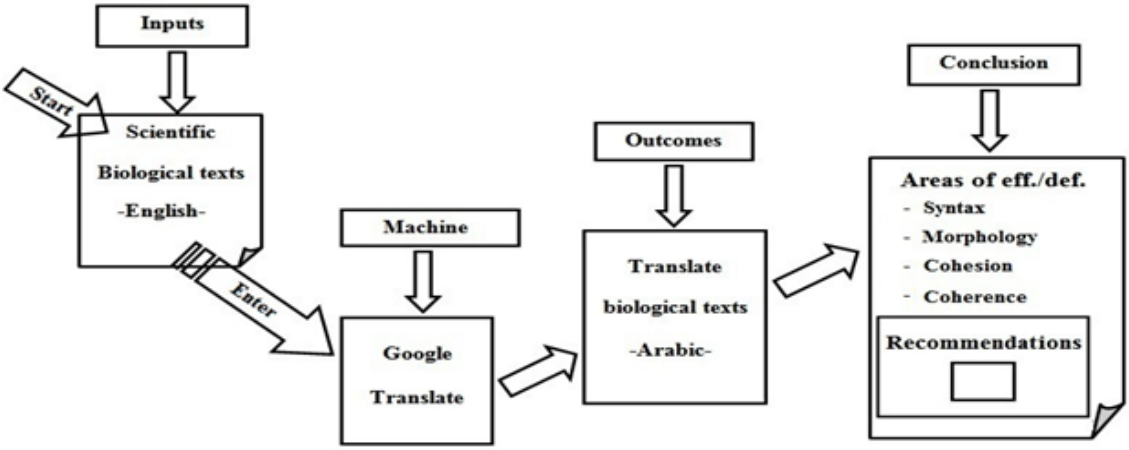


Figure (1): The process of treating the selected data.

2.3. Theoretical Framework

The present research relies mainly on the model of Machine Translation Evaluation (MTE) based on the International Standards for Language Engineering (ISLE's framework of taxonomy 3). The model distinguishes between two types of evaluation; The 'glass box' evaluation which considers GT a ‘glass box’, so the evaluator looks inside the translation engine to see how the translation process is done. While the second type of evaluation is concerned with the relationship between the input and output. In this type of evaluation, GT is treated as a 'black-box', which means that the evaluator has to look at the input and output without taking into account the mechanisms by which the GT engine works (FEMTI, 2003 as cited in Hannouna, 2004, p.115).

Thus, it is the 'black box' evaluation that will be adapted in this research since it helps in identifying areas of errors that may occur in the GT performance and which are deemed to be the main objectives of the present evaluation. In other words, the 'black box' evaluation focuses on the

external quality characteristics of GT. These characteristics of the outcome can be traced by comparing between the input and the output without the need to explore the system functions since the aim of the research is to identify the errors that may result in the translation. However; when the time comes to giving recommendations for the software developers on how to improve it, the research will shed light on the 'glass box' evaluation.

Consequently, the research will draw on Catford's translation shifts (1965) including both 'level shift' and 'category shift' to highlight the grammatical errors that may result at sentence level. Thus, the researcher will compare the source text and GT output to identify the types of shifts that may take place. Yet, when the research shifts to paragraph level, it will examine four types of paragraphs frequently used in scientific texts, descriptive, process, causality and mixed paragraphs, according to Halliday and Hassan's model of cohesive markers (1976) employed to make the text cohesive and coherent. This model includes both grammatical and lexical cohesive devices. Thus, the researcher will trace the cohesive markers deployed in the source text to test whether GT reproduces them correctly in the output based on that model or not.

Chapter Three

Linguistic Discordance in Google Translation

3.1. Introduction

Linguistics can be best described as a musical instrument that is capable of playing different melodies according to a group of tunes set together to form a musical scale that is usually put in front of musicians to follow in big concerts. Similarly, the linguistic system, in any language, does the same function of that musical instrument since linguistics plays a major role in producing stretches of language that sound harmonious and meaningful to both the ears and minds of all language users. In other words, linguistics forms a scientific model that helps language users to form and communicate coherent and innermost thoughts since it controls the way people use their language to express their human experience. However, this experience may be different from one group of language users to another because "language … gives structure to experience, and helps to determine our way of looking at things, so it requires some intellectual effort to see them in any other way than that which our language suggests to us" (Halliday, 1970, p.143).

Therefore, the effort and time spent in understanding the differences between the linguistic systems of all languages worldwide is affected by the components of the linguistic system of the language/s in question. The linguistic system of any language could be seen as an umbrella that covers

different areas, one of which is named 'grammar'. This area consists of two main parts which are: syntax and morphology. The former part deals with the arrangement of words in different language structures while the latter deals with the structure or build up of individual words. Notwithstanding, the division of this umbrella into two parts does not mean that they are different or unrelated to one another; instead, both syntax and morphology are much more interrelated that being contradicted because the term 'grammar' has been used to refer to the two concepts of syntax and morphology by many researchers including Baker who maintains that "grammar is organized along two main dimensions: morphology and syntax. Morphology covers the structure of words", while "Syntax covers the grammatical structure of groups, clauses, and sentences" (1992, p.83).

However, language users do not use the grammatical categories that exist in their languages in the same way since these grammatical elements are not identical to all languages. In other words, each language differs from the other in the way it expresses the same message, for each has its own grammatical patterns that it imposes upon its users. Therefore, in the translation process, the variety among grammatical categories between source and target language poses a difficulty for human translators "because one cannot always match the content of a message in language A by an expression with exactly the same content in language B, because what can be expressed and what must be expressed is a property of a

specific language in much the same way as *how* it can be expressed" (Winter, 1961, p.98).

Thus, the possibility of achieving equivalence at grammatical level has been examined by many researchers who discussed this dilemma in relation to translation. For example, the clear-cut differences between English and Arabic in relation to grammatical categories led Baker to identify five problematic grammatical categories between English and Arabic which are: number, gender, person, tense and aspect, and voice. She maintains that the differences at this level constitute a source of difficulty for human translators because such differences are capable of changing the content of the message in the process of translation. This change may lead to the addition of information which is not found in the original or omitting information from the source text. She concludes that this may happen when "the target language has a grammatical category which the source language lacks", or "if the target language lacks a grammatical category which exists in the source language" (1992, p.83).

Moreover, Catford discussed the process of translation between two languages with different linguistic systems. He maintains that there are some "translation shifts" that may occur in the process of translating a text from a mother language to a target language. He states that such shifts may take place at two main levels: lexical and category. The former shift "occurs when an SL item has a target language equivalence at a different linguistic level from its own (grammatical, lexical, etc.) "while category

shift takes place at four levels which are: class, structure, unit and intra-system shifts. First, "class shift" which involves changing the class of a word, e.g., from an adjective to a noun. Second, "structure shift" which refers to altering the grammatical structure of a sentence, e.g. from active to passive. Third, "unit shift" which refers to switching the rank of: e.g. a clause to a phrase. Finally, "intra-system shift" "which occurs when translation involves selection of a non-corresponding term in the TL system …: e.g. an SL 'singular' becomes a TL 'plural'" (1965 as cited in Hatim, 2001, p.16). Such shifts may take place when the translator cannot adhere to the linguistic forms that exist in the source text.

Therefore, Nord states that "linguistic problems arise from differences of structure in the vocabulary and syntax of second language (SL) and target language TL" (1991, p.88). In the same vein, Abbasi and Karimnia assert that most students commit errors while doing translation tasks at syntactic and morphological levels or what they call "syntactic-morphological errors" such as: errors in the use of the appropriate tense, errors in the use of articles and prepositions, and errors in the use of active and passive voice. They state that students while doing certain translation tasks, they do commit errors at different grammatical levels because they transfer the grammatical rules of their own language into the target language (2011. Retrieved from: https://urlzs.com/wpw5J).

Even so, these problems may be solved since Hannouna believes that human translators can work hard and focus their effort on understanding

and mastering the grammars of the two languages involved in the translation task and they can "draw on general knowledge of the subject matter and the world to arrive at the intended meaning" (2004, p.54). Thus, the research hypothesizes that when it comes to machine translation, deficiencies related to grammatical categories would be more painstaking since machines may not have the ability to analyze all the grammatical categories between the languages involved especially if they are far-distant languages such as English and Arabic; such languages have more differences in their linguistic systems than similarities.

On the grounds of this, linguistic errors may blow in while using machine translation because "MT is often impeded by lexical and syntactic ambiguities, structural disparities between the two languages, morphological complexities and other cross-linguistic differences" (Hannouna, 2004, p.54). In other words, machine translation errors may be attributed to the framework adopted by all machine translation programs that is known as the "Transfer approach". This approach consists of three steps which are: First of all, the scanning and analysis of the ST syntactic structures into their basic building blocks. Secondly, the transfer of those syntactic structures into the TL structure. Finally, the synthesis and restructuring of the output based on that TL structure which may yield one or a number of proposed translations for the same structure. Thus, this approach indicates that the process in machine translation programs is sequential so each step has to pave the road for the next one to take place to

produce optimal output that satisfies the users' desires (Somers, 1998, p.145 as cited in Hatim and Munday, 2004, p.117).

However, in some cases, machine translation programs may commit errors at one or all of the above mentioned levels. This, in turn, leads to many errors in the translations made by the software in question. In other terms, there are shifts that may take place in the translation process which may result in semantic shifts that might change the meaning of the text/s in hand. Such errors may widen the gap between human translators and machine translation as two faces for the same coin since Brown defined errors as "a noticeable deviation from the adult grammar of native speakers, reflecting the inter-language competence of the learner" (2004, p.216). By analogy with humans' competency, errors indicate that the preprocessing mechanisms of the software in question are not doing well, so they need to be enhanced and well-fed. This echoes Al-Samawi's words who states:

> The question whether machine translation would replace human translation was and is still one of the primary concerns of research in machine translation. Researchers, in this regard, are between fear and confidence. Some look at it as a real threat to human translators; others are doubtful and base their doubt on the terrible errors committed by machine translation (2014. Retrieved from: https://urlzs.com/74KMz).

Consequently, many researchers have attempted to identify and classify the errors produced by machines in relation to their linguistic

competency, in particular, grammatical categories at smaller levels such as word and sentence level. For example, Hannouna (2004) states that the machine commits errors in areas such as: category and word class, syntactic arrangement, tense, pronoun translation, suffixes among other areas. She evaluates the quality of three Arabic machine translation systems but her study focuses only on one single level of texts which is the sentence. In addition, Vilar, Xu, D'Haro and Ney (2006) identified five big classes of errors which are: "missing words, "word order", "incorrect words", and "unknown words and punctuation errors". Their study also focuses on smaller units (Retrieved from: https://urlzs.com/XtQ9w).

In addition, Al-Samawi (2014) identified a number of errors made by GT at text level both at syntactic and morphological levels such as: "Violating subject-verb agreement (masculine and feminine; singular, dual, and plural; first, second, and third person)" , "Using a noun in place of a verb", "Using a verb in place of a noun" and "Omitting functional morphemes (i.e. prepositions, articles, conjunctions, pronouns, auxiliary verbs, deixis, etc.)." However, his study focuses on counting the number of errors at the first ten sentences in each text without explaining them or the semantic shift that took place in the texts. Also, he uses 10 texts from 10 different disciplines in his research; this in turn may not be objective or fair enough to make conclusions about GT performance in each field (Retrieved from: https://urlzs.com/74KMz).

Therefore, using GT to do certain translation tasks may yield a number of grammatical errors in different areas because when it comes to a software tool, e.g. GT, and linguistics, the situation may be vague and confusing for:

> Psychologists have told us that individuals acting alone do not normally cause too much trouble; it is only when they form into crowds that they become unmanageable. Similarly, individual lexical items. . . , can only stage sporadic strikes; it is when they group into long syntactic stretches that they begin really to launch all-out assaults on the translator (Wong, 2006, p.130).

If it is so for human translators, then it would be at least the same or even far more challenging for GT since it does not have a sense of judgment or enough intelligence as humans do. Thereby, this chapter aims at detecting the grammatical errors that result in the translations produced by GT then classifying those errors under broad and sub-categories to demonstrate the effect of the grammatical shifts that take place then measure the semantic shifts and their effects at the comprehension level. Finally, this chapter will attempt to draw on the last step in the "Transfer approach" that was further developed by what is called "users' feedback button" nowadays. This button enables the end users to interact with the machine and contribute in enhancing the quality of the output.

In a nutshell, in the last step of the "Transfer approach" that is known as "synthesis", a number of machine translations including GT start to provide the end user with one or a number of suggested translations for the item in question. Accordingly, the end user starts to enjoy the freedom to: **either** accept the proposed translation provided by GT, reject it alright **or** come up with an acceptable translation by his/her own in cases where all the proposed equivalences provided by GT were fuzzy or imperfect.

Thus, this chapter aims at suggesting acceptable translations for the items translated erroneously by GT to be added to the list of options provided by GT in case where those items were re-inputted by a different user. In sum, this chapter seeks to give recommendations to solve each type of errors in an attempt to pursue a good reputation of machine translation and provide the end users with acceptable translations.

3.2. Errors at Syntactic Level

3.2.1. Organization of Constituents at Phrase Level

The Arabic noun phrase is formed when the noun precedes the attributive adjective while in English the attributive adjective precedes the noun according to the naturalness principle that controls the production of well-formed structures in both languages. However, the research observes, in some cases, that GT sticks to the same structure of the source text which results in forms that are irrelevant to the target language since they are incoherent constituents. In other words, it is a well-known fact that when it

comes to phrases, e.g., noun phrases, then the position of the modifier in relation to the modified noun affects the message. For example, Arabic starts with the noun first, then it gives information about it in a direct manner, while English triggers and prepares the readers/listeners for the theme since it describes the object first then reveals and unfolds its identity. These different ways of presenting facts and secrets about the same object in the two languages led GT to commit errors in this area as shown in Table (1) below:

Table (1): Errors made at noun phrases level

Ex.	Source Text	Google Translation
1.	Mendel chose the garden peas for his studies because: garden peas are available in many varieties.	اختار مندل البازلاء الحديقة لدراسته لأن: البازلاء حديقة متوفرة في العديد من الأصناف.
2.	In a chemical reaction, all of the atoms in the reactants must be present in the products. The reactions must be balanced.	في التفاعل الكيميائي ، يجب أن تكون جميع الذرات الموجودة في المواد المتفاعلة موجودة في المنتجات. يجب أن تكون متوازنة ردود الفعل.

The noun phrase in the English text which consists of the noun "garden peas" in **Ex.1** is not translated correctly. In other words, a structure shift takes a place because GT changes the order of the phrase in the translation to appear as a noun phrase that is made up of two nouns which are the "garden" and the "peas", in Arabic "البازيلاء الحديقة". This results because GT resorts to literal translation which, in turn, derives it to treat the word "peas" in its current position in the English sentence as if it were a noun and the garden which is its modifier as an adjective. However, the word "garden peas" in the underlying structure of the noun phrase here serves as a noun which in Arabic means "حبات البازيلاء". The noun here is

used to specify the type of seeds that Mendel selected for his experiments. In other words, the noun phrase is employed to make the idea more specific and precise.

Thus, this shift proves that GT fails to recognize this underlying structure and the way Arabic makes it manifest in its surface structure so this leads to translating the two words as two nouns and this results in a form that is not familiar in the Arabic language which is two consecutive nouns each with the definite article "البازيلاء الحديقة". That is to say, GT fails to analyze the noun phrase "garden peas" as a phrase with one noun, in Arabic "حبات البازيلاء". Thus, the translation produced by GT may lead students to stand for a while to rearrange the sentence and allocate each word in its appropriate position to get the message. This, in turn, may weaken the translation of scientific texts since Ali and Ismail maintain that technical terms create a challenge for students to understand because these technical terms "have one or many meanings in everyday language" but in a scientific text, they have a different, peculiar and precise meaning (2006. Retrieved from: https://urlzs.com/FnT8d).

Moreover, **Ex.2** shows that GT does not stick, in some cases, to the same order that exists in the source text. In other words, GT does its own guesses to translate a certain sentence regardless of how words are combined in the same sentence in the source text. Thus, **Ex.2** shows that GT fails to order the constituents of the sentence in the right way since English starts with the attributive adjective first then comes the noun.

However, this is not the case in Arabic since Arabic starts with the noun which in this case is " ردود الفعل" then comes its adjective which is " متوازنة".

Another dilemma is that GT fails to order the constituents of the noun phrase in the right way as **Ex.3** in Table (2) shows. The noun phrase "electron transport chain" in the second translation provided by GT is not translated correctly since GT fails to recognize its head which is the noun "chain". In addition, providing two different translations of the same noun phrase despite the fact that it is an established scientific term: one that is right while the other is wrong indicates that GT is still unsure about the correct translation.

Table (2): Errors in ordering the noun phrase.

Ex.	Source Text		Google Translation
3.	Electron transport chain accepts electrons from the breakdown products of the first two stages (most of them via NADH) and passes these electrons to an electron transport chain.	1	تقبل سلسلة نقل الإلكترونات الإلكترونات من منتجات التكسير في المرحلتين الأوليين (معظمها عبر NADH) وتمرير هذه الإلكترونات إلى سلسلة نقل الإلكترون.
		2	الإلكترون سلسلة النقل تقبل الإلكترونات من المنتجات انهيار المرحلتين الأوليين (معظمهم عبر NADH) وتمر هذه الإلكترونات إلى سلسلة نقل الإلكترون.

Therefore, producing correct noun phrases requires that GT draw a map for the items in question in order to decide on the function of elements then rearrange them without any loss or distortion that may threaten the quality of the output. For example, in the translation of the above mentioned phrases, GT should have done it without any change in the order

of the constituents since the original is clear and precise, so it would be safer for GT to analyze how phrasal slots are ordered in both English /(Art.)(Adj.)N./ and Arabic /(Art.)N.(Adj.)/ then map unto them to produce correct structures.

Therefore, to handle this phrase-level translation anomaly, the researcher suggests a procedure described in Figure [**2**]. First, the sentence provided by the user is to be split (tokenized) into tokens (words). Then, these tokens are passed on to a Part-of-Speech Tagger (POS) that finds the type of each token, i.e., whether a word is a verb, noun, adverb etc. Using these token types, one can find out whether a sentence complies with the "Art. + Adj. + N." pattern or not. If yes, then the nouns part undergoes the step of bigrams and trigrams extractions, where bigrams and trigrams are phrases consisting of 2 and 3 tokens, respectively. The translation of these noun phrases are looked up from a specialized lexicon. For **Ex.1**, the direct translation of the phrase "garden peas" would be "البازيلاء الحديقة". Now, the user can detect this anomaly and give his/her feedback by suggesting a new translation "حبات البازيلاء", which would then be maintained in the lexicon.

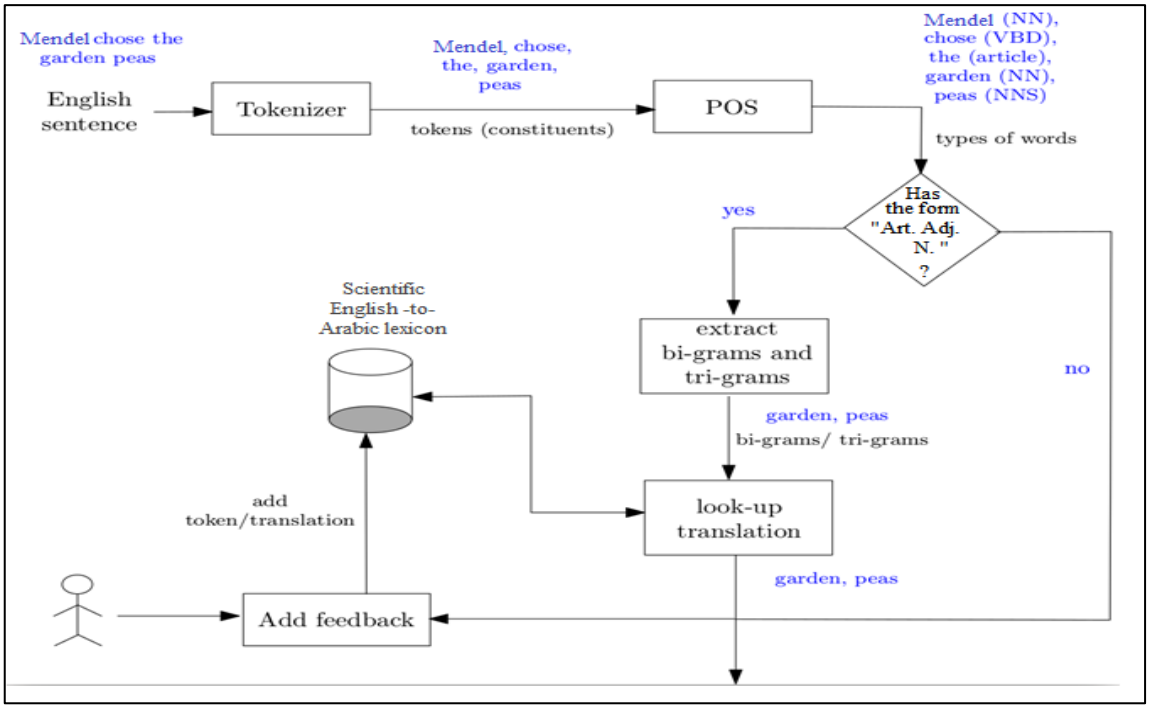


Figure (2): Processing of noun phrases.

3.2.2. Organization of Constituents at Sentence level

The simplest sentence in English consists of SVO/C (subject, verb and object/complement) and conveys a certain message. However, when it comes to GT, it is clear that it commits errors at this level, in particular, with the arrangement of the elements that make up the whole sentence. This is due to the nature of the two languages and the features associated with each of them. That is to say "English is basically an analytic language, i.e., it shows syntactic relationships by word order and function words. Arabic is basically synthetic, i.e., it shows syntactic relationships by its frequent and systematic use of inflected forms" (Hawkins, 1980 as cited in Saraireh 2014). This diversity in ordering the constituents of the sentence may hinder the process of understanding the message since it may drive GT

to commit errors that cause structural ambiguity which in turn yields different interpretations of the same message as shown in Table (3) below:

Table (3): Errors made at sentence level.

Ex.	Source text	Tran.	Google Translation
4.	The number of protons determines the atomic number		يحدد عدد البروتونات العدد الذري.
5.	Electron transport chain accepts electrons from the breakdown products of the first two stages (most of them via NADH) and passes these electrons to an electron transport chain.	1	تقبل سلسلة نقل الإلكترونات الإلكترونات من منتجات التكسير في المرحلتين الأوليين (معظمها عبر NADH) وتمرير هذه الإلكترونات إلى سلسلة نقل الإلكترون.
		2	الإلكترون سلسلة النقل تقبل الإلكترونات من المنتجات انهيار المرحلتين الأوليين (معظمهم عبر NADH) وتمر هذه الإلكترونات إلى سلسلة نقل الإلكترون.

In English, there is only one type of sentences; one that starts with the subject followed by a verb along with its complement and it is called the verbal sentence; while in Arabic, there are two types of sentences: equational and verbal. The former starts with a noun followed by a predicate while the latter begins with a verb followed by a subject and a complement. However, in **Ex.4** in Table (3), the English sentence starts with the subject which is "the number of protons" followed by the verb "determines" and its complement. This sentence follows the unmarked pattern of SVO in English. However, the structure of the English sentence is reversed in the Arabic sentence by GT leading to a semantic shift that results in two readings of the Arabic sentence: it's **either** that the atomic number is the one that decides the number of protons **or** the number of

protons is the one that is responsible for deciding on the atomic number. In the Arabic sentence, both nouns- the number of protons and the atomic number- would stand in both subject and object positions.

Otherwise stated, the problem lies in that readers of this sentence will get confused about the correct meaning of the sentence, especially those readers who are not well-acquainted with the Arabic syntactic rule which states that: if there are two consecutive nouns in a verbal sentence, and the sentence does not use case markers/inflections to distinguish between them, then the subject is the first noun and the object is the second noun. In other words, the process of sorting them is going to be done according to the order in which they appear in a given sentence. To resolve this ambiguity, GT needs to be improved by adding inflections to the Arabic sentence. The inflections (diacritics) to be used in this case are: *damma* to indicate the subject position and *fatha* to indicate the object position; however, as GT does translate the input, the Arabic sentence has two possible readings, the thing that weakens the quality, precision and level of comprehension of the translated text.

Moreover, **Ex.5** in Table (3) shows that reordering the elements in a given sentence may produce redundant stretches of language such as using two similar nouns immediately after one another in the same sentence. Thus, this may give away one meaning from the sentence which, in turn, may change or alter the intended meaning in the source text since repetition may lead to ambiguity. Thus, in **Ex.5**, the English sentence starts with the

subject "electron transport chain" but GT inverted the order in the Arabic sentence leading to two similar forms following one another. Accordingly, when it comes to sorting out these two nouns and assigning them the appropriate inflections (diacritics) in Arabic, the result will be two nouns with the same inflection which is: *Kasrah* "الالكتروناتِ الالكترونات". This redundancy may lead students to realize or perceive it as a typo so they would read the sentence as "تقبل سلسلة نقل الالكترونات". This means that they might omit the second word "الالكترونات" in the Arabic sentence " تقبل سلسلة نقل الإلكتروناتِ الإلكتروناتِ" since they might be deceived or misled by the wrong ordering produced by GT which results in producing redundant words.

Therefore, in the case of translating active sentences where both the subject and the object contain similar nouns and the inflections do not help in clarifying the meaning, it would be better and safer for GT to: **either** maintain the order of the source text to prevent any speculations about how the source text might be like to get the bulk of the message **or** detach the two constituents using a verb. Thus, it would be better for GT to produce nominal sentences that start with the subjects which are: “the number of protons” and "electron transport chain" in the examples then give information about their function.

These examples prove that GT still commits errors in ordering the constituents of both the phrase and sentence due to the differences of the rules that combine these patterns such as NP, SVO, etc. together in the

language/s in hand. These erroneous and random switches between such patterns will lead to errors, such as producing a sentence or a phrase with more than one interpretation; producing forms that do not exist in the target language or wrong ordering of the name of the scientific term. All these errors weaken the level of comprehension and therefore the quality of the translation outcome. In the current situation, the GT users will rely on their intuition to make the sentence sound coherent and cohesive.

Therefore, to solve this problem, GT should adopt a two step procedure described in Figure [**3**]. The first step is a "text preprocessing step" in which GT analyzes the input sentence into tokens, in a "Tokenizer" then those tokens will be marked along with their grammatical categories in a "POS" tagger. This will help in identifying the sentence pattern employed in the input based on the order of the elements in the sentence under study. For example, the sentence in **Ex.4** conforms to the unmarked sentence-pattern in English: /sub.+ v.+ obj./ for it starts with a NP and ends with a NP, "the number of protons" and "the atomic number", respectively.

Secondly, GT needs to decide on whether the two nouns are inflected for case or not. If not, then GT will undergo a second step to preserve the meaning of the ST through allocating both the sub. and the obj. to their correct positions in the present sentence pattern. In other words, the proposed system should split the two nouns "the number of protons" and "the atomic number" in the sub. and obj. slots in Arabic, respectively. Thus, the system should be programmed to map the unmarked pattern of /sub.+ v.

+ obj./ in English to /v.+ sub.+ obj./ pattern in Arabic. Accordingly, users can suggest a precise translation through adding the Arabic diacritics to the sentence such as: *fatha* and *damma* to avoid any ambiguity that may weaken the quality of the output. Thus, such diacritics will be added to a specialized lexicon to be reused again to solve the confusion that may occur in identifying the subject and the object in the GT output as shown in Figure [**3**]:

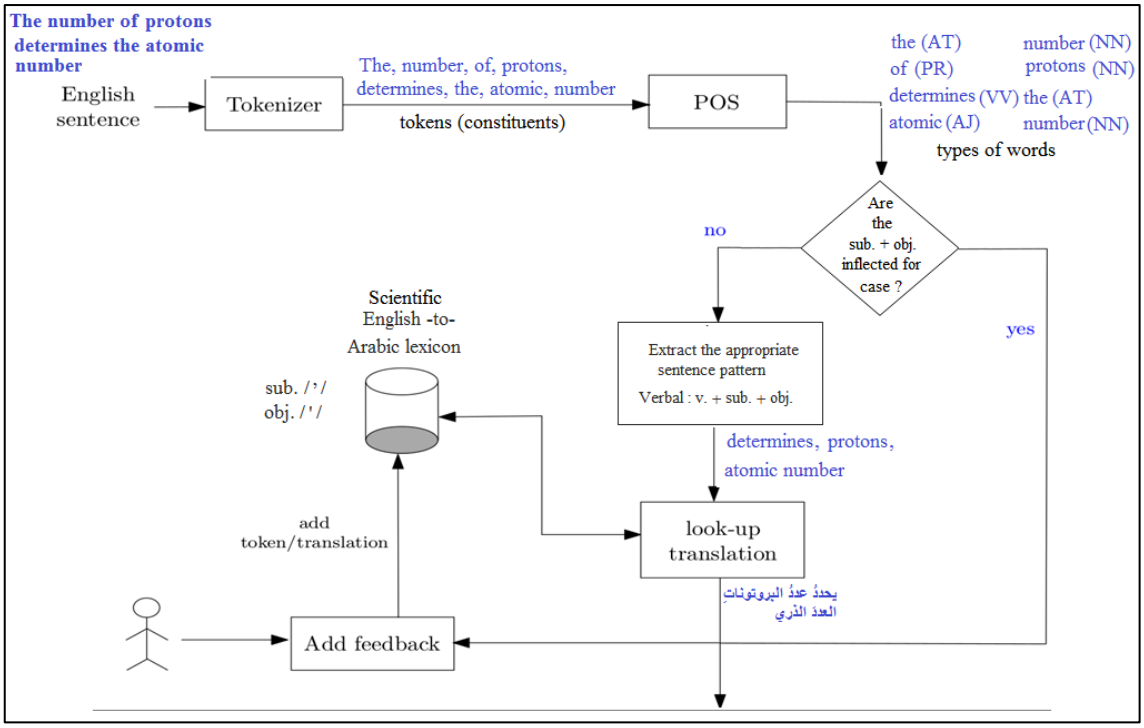


Figure (3): Mapping of nominal sentences to verbal sentences.

However, in cases where the subject and the object contain similar words which in turn may lead to redundancy in the output, GT should be programmed to block those two nouns from following one another in a procedure described in Figure [**4**]. First, GT should undergo the same "text preprocessing step" explained previously to analyze the input sentence.

Second, GT should decide on whether there are similar words in the input between the sub. and the obj. or not. If yes, then GT needs to map the input sentence into a nominal sentence with the pattern /sub.+ v.+ obj./. This sentence pattern will separate the subject from the object by the verb which in turn will reduce the redundancy in the output. At this stage, users can assist GT by providing it with the correct nominal pattern of the sentence under study. Finally, "a diacritics extraction step" will take place to emphasize both the subject and the object. Such procedures will allow the users to enjoy translations of high quality and precision when it comes to translating active sentences.

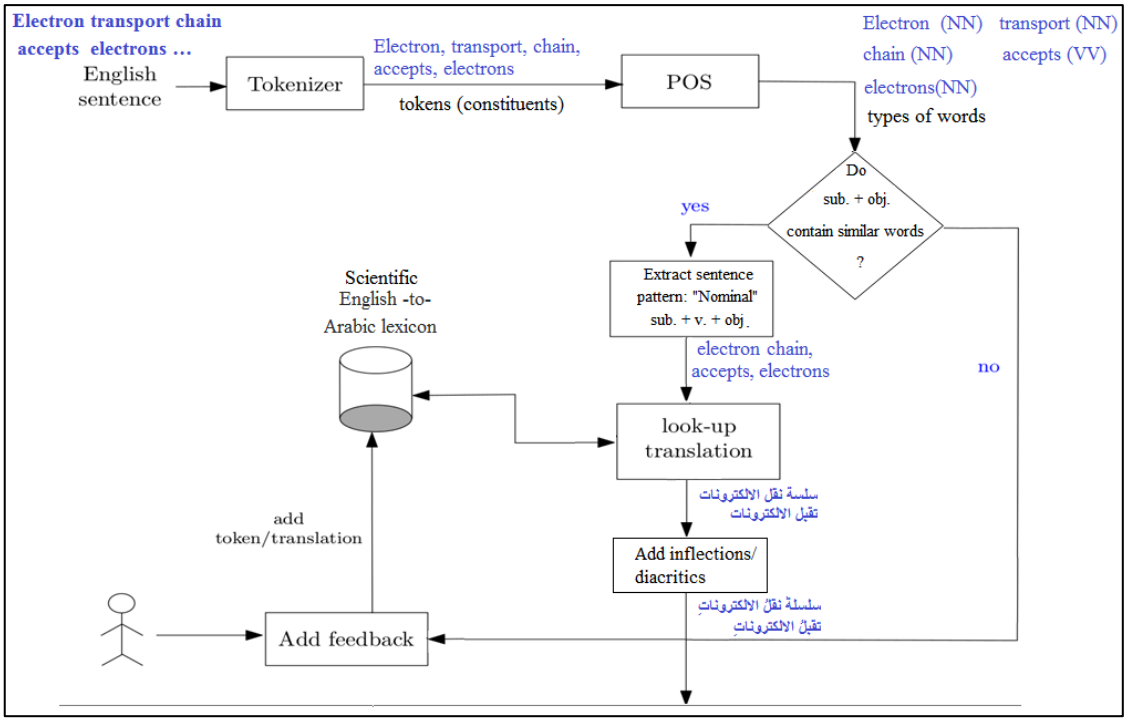


Figure (4): Mapping of nominal sentences.

3.2.3. Erroneous Shifts from Verbal to Nominal Sentences in Arabic

The simplest sentence in any language is made up of different parts of speech such as nouns, verbs, adjectives and adverbs. However, these categories may be problematic to GT since languages differ in the way they derive such parts of speech and the way they combine those elements together to communicate a message as the sentence in Table (4) below shows:

Table (4): Errors of turning a sentence to a noun phrase

Ex.	Source Text	Google Translation
6.	Some isotopes are radioactive	بعض النظائر المشعة

GT fails to order the constituents of the sentence because it does not recognize the auxiliary "are" in **Ex.6**, GT neglects it in the process of translation, this in turn leads to translate the sentence into a noun phrase, "النظائر المشعة"; this leads to a problem in comprehending the sentence because GT replaces the adjective by a noun phrase. In such case, the reader may search for a main verb after the noun but s/he finds nothing since GT drops the auxiliary "are" from the sentence. Thus, this turns the verbal sentence into a noun phrase. These kinds of errors where a verb is not translated directly as a verb; instead, it is turned into a noun are classified as structure shifts.

Therefore, the researcher suggests that GT be programmed to translate the verb to be (Aux.) and the adjective that follows it in a verbal sentence in English into an adjective which makes the sentence equational in Arabic as explained in Figure [**5**]. In other words, Arabic does not use

such type of pseudo-verbs which include: is, am, are, etc. to introduce adjectives. Therefore, GT should undergo the prepossessing step first to identify the aux. and then translate it and the adjective that follows it in English into an equational sentence in Arabic which consists of a subject and a predicate. Second, GT should undergo an extraction step. In other words, the Arabic predicate (adj.) has to be derived from the /aux./ and the /adj./ in the English sentence. However, in **Ex.6**, GT neglects the presence of such pseudo-verbs; this results in "genitive structures" in English, "مضاف ومضاف اليه" in Arabic, such as: "بعض النظائر المشعة". Such form may not help the readers to distinguish or identify the topic of the sentence which is called the theme and the comment that tells the readers more about the theme, that is called the rheme. Therefore, users could add their suggested translation for the sentence which is "نظائر مشعة" to be maintained in the lexicon to be reused again in similar constructions.

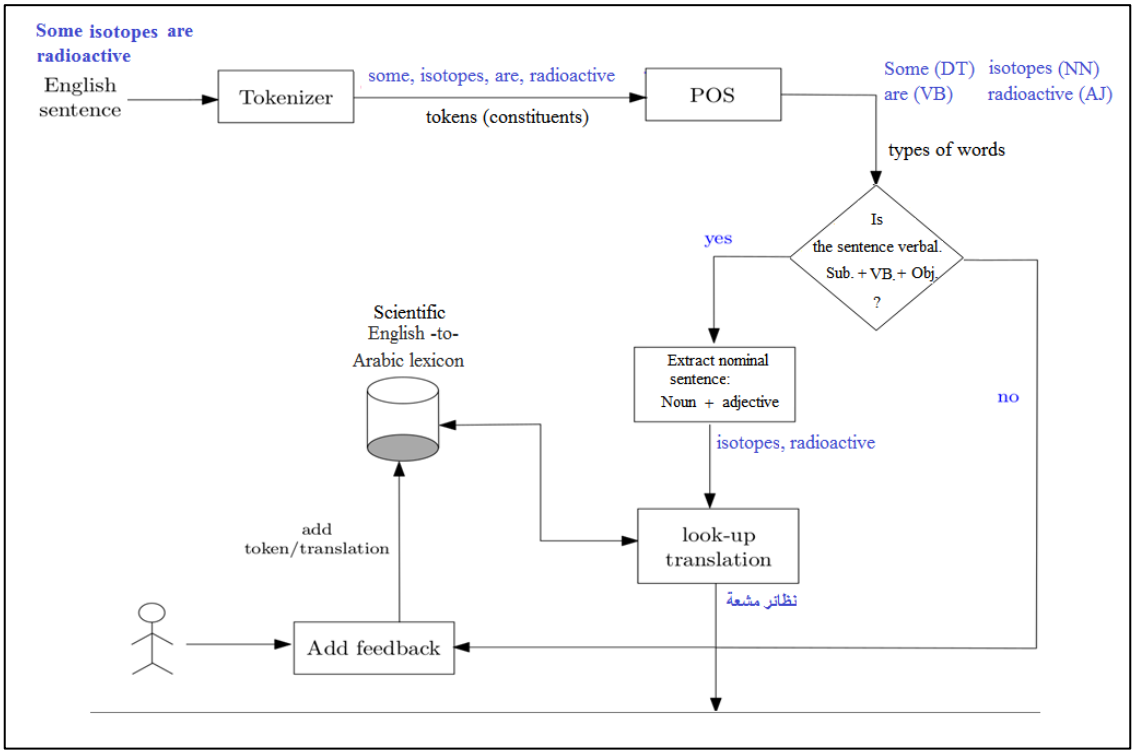


Figure (5): Mapping of sentences with verb to be.

3.3. Errors at Morphological Level

3.3.1. Inappropriate Choice of Suffixes

Affixes in English are of three types: prefixes that are added in front of the word, infixes which are put in the middle of words, and suffixes that come at the end of the word. Each type has a function which helps in constructing a precise meaning.

3.3.1.1. Inflections Attached to Sub-headings

In some cases, GT fails to add the definite article "the","ال", in Arabic. This is due to a well-known fact that in English, when people want to refer to things in general they use the plural form while in Arabic the situation is different. In other words, Arabic employs the "ال" to refer to

things in general while English uses it to refer to or specify the referent/topic.

However, in the example in Table (5), GT fails to add the "ال" to the noun phrase "chemical reactions" in the Arabic text making the noun phrase indefinite since English and Arabic differ in the process of assigning the definite article to nouns according to the function of the sentence. In the English sentence below, the noun phrase which starts with a capital letter /C/ aims to refer to chemical reactions in general for it introduces the topic of the subsequent sentence. In other words, GT should be programmed to attach the "ال" to the noun phrase "Chemical reactions" for it functions as a sub-heading. However, **Ex.7** in Table (5) proves that GT still needs to be enhanced in this area since the Arabic noun phase indicates that the reactions are unknown which in turn makes the topic vague and not specific enough due to the absence of "the" which adds some kind of familiarity and smoothness to the sub-sequent sentence as the example in Table (5) shows:

Table (5): Errors in treating the definite article

Ex.	Source Text	Google Translation
7.	Chemical reactions In chemical reactions, chemical bonds are broken and reformed, leading to new arrangements of atoms.	تفاعلات كيميائية في التفاعلات الكيميائية ، يتم تكسير الروابط الكيميائية وإصلاحها ، مما يؤدي إلى ترتيبات جديدة للذرات

3.3.1.2. Inflections Attached to the Verb

A clear example related to errors in affixation is the main verb in the Arabic sentence below. GT fails to recognize that the /s/ in its current position is used to indicate a verb that is both active and present and has a singular subject. However, GT treats the /s/ as a grammatical category that is used to indicate the plural form of the noun /function/ as shown in Table (6) below:

Table (6): Errors in the selection of parts of speech.

Ex.	Source Text	Google Translation
8.	In a multi-cellular organism, cell division functions to repair and renew cells that die	في الكائنات متعددة الخلايا، وظائف الانقسام الخلوي لإصلاح وتجديد الخلايا التي تموت

This example shows that GT fails to distinguish between words that have similar forms in both plural and simple present cases. In other words, a category shift that changed the category of the word "functions" form a verb in the English sentence to a noun in the Arabic text took place. However, the word *function* is used: **either** as a verb **or** a noun since /function/ could have two forms, this means that *function* could be used both as a verb which means to serve/work or as a noun which means a job/task. Thus, this duality of forms of the same word leads the reader to realize that the output sentence has no verb since it is mistranslated by GT into a noun while in the source text; it is intended to serve as a verb and not a plural noun. This makes the Arabic sentence appear as if it were verb-less which in turn does not help to get the message in the Arabic text since it is

not allowed to identify the verb in a given sentence using one's intuition especially in scientific texts.

Thus, it is important that GT developers feed GT with a procedure described in Figure [**6**] to enable it to handle all the words that have the same form in both plural and present tense, with a 3rd person, singular subject cases. First, GT will undergo the text preprocessing step to decide on the function of the word in question and what it aims to achieve. That is GT needs to process both the position of the word in the sentence and the surrounding elements that shape its identity. For example, the position of the word in question which is "function" in **Ex.8** shows that the word is used as a verb for the subject "cell division". In other words, the verbs "repair" and "renew" could not be the verbs for the subject "cell function" because there is the particle "to" before them. Thus, GT needs to answer this question: "Does the sentence have a verb for the subject “cell division?". If no, then GT will extract a verb that agrees with the present subject in person, number, etc. However, at this stage, GT cannot derive the appropriate form of the verb "function" in Arabic since it translates it as a noun not a verb, in Arabic "وظائف". Thus, at this stage, users can suggest a translation for the word "function" as a verb which is "يعمل على". Accordingly, this translation would be maintained in the lexicon to be reused again in similar circumstances.

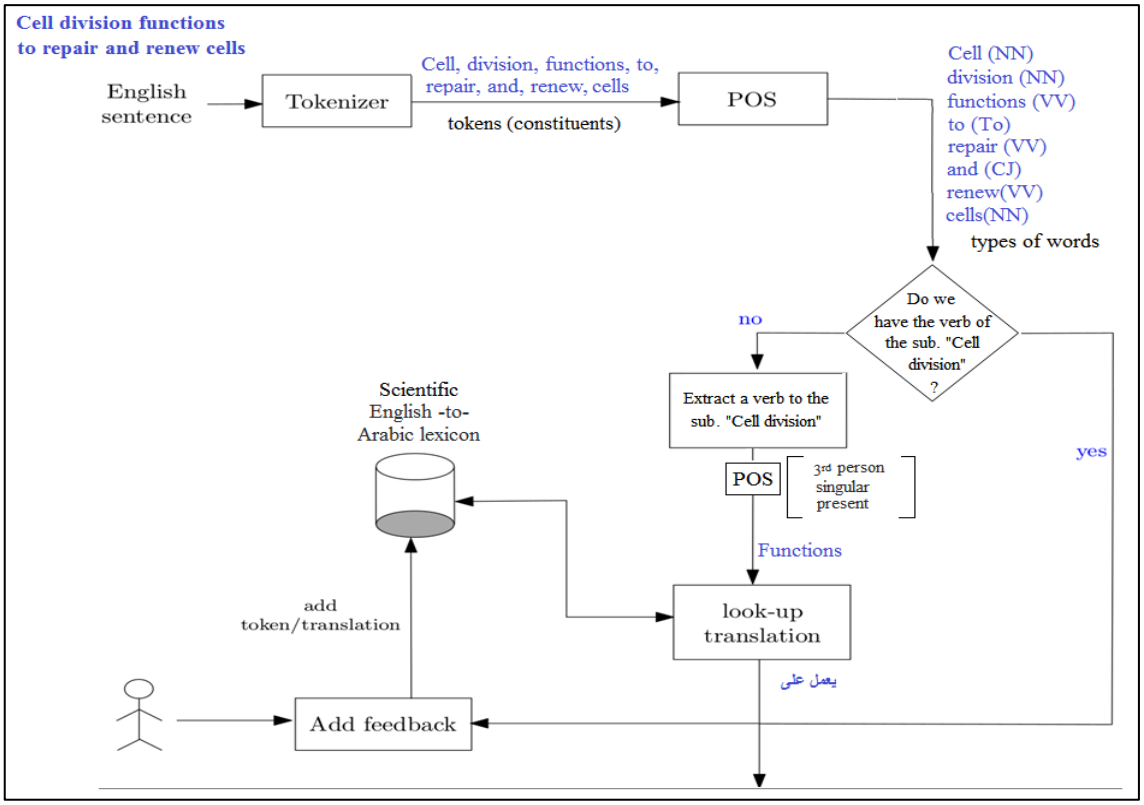


Figure (6): Processing of words with similar forms in plural and present tense.

3.3.2. Passive Constructions

Passive constructions are used heavily in scientific texts to achieve certain purposes. Swales states: "the passive can be used to give the necessary information in the best possible way; impersonally, concisely, objectively, and giving importance to the most important facts" (1971, p.41). However, when considering GT, there is a number of errors that take place in the translation of certain sentences form active to passive. These errors include:

3.3.2.1. Failure to Distinguish between the Simple Past and Passive Inflections

In some cases, GT mistranslates sentences that contain a passive construction by using a simple past form in place of a passive. In other

words, GT fails to distinguish between the simple past form and the participle form that comes after the auxiliary in passive constructions-passive adjectival-. This in turn may affect the truth value of the sentence as Table (7) below shows:

Table (7): Errors in recognizing the passive construction

Ex.	Source Text	Google Translation
9.	a disaccharide consists of two monosaccharides joined by a glycosidic linkage.	يتكون ديساكهاريد من اثنين من السكريات الأحادية التي انضمت إلى ربط glycoside
		يتكون ديساكهارايد اثنين من السكريات الأحادية انضم اليهم الربط غليكوزيدية.

The sentence in **Ex.9** states a fact about the components of "disaccharides", so the verb that is usually used to refer to factual issues in English is the simple present not the simple past since the use of the simple past "joined" may indicate that the components change or that the process of producing disaccharide is done in the past and now it is over. Thus, this is not acceptable in science language since things have to be clear, exact and fixed to establish a kind of mutual trust between the readers and the text/s in hand. In other words, the verb "joined" does not indicate a simple past but it is a passive construction that is erroneously recognized by GT as a simple past. This shows that GT fails to make use of the present key words in the sentence such as the preposition "by" and the verb "consists" to understand that the sentence is talking about actions that happen at a present situation or something that takes place whenever there is a process of disaccharide production.

Basically, GT fails to recover the underlying structure of the sentence to translate it as a passive construction so it goes with the superficial structure which is the simple past. However, this leads to errors in the translation of the passive construction since reading the Arabic sentence may lead to a conclusion about a process that happened in the past due to the use of the simple past form of the verb "انضم الى" and not a process that could be repeated whenever disaccharides are formed since the underlying structure of the sentence is: "are joined", "تنضم الى" not "joined", "انضمت" in Arabic- as a past form. Thus, GT should be programmed to benefit from the words in the textual context in its translation box like: "by" and the verb "consists" in the present case. Such words should help GT recognize the verb in question as a passive form not a past form.

Another issue is that GT neglects the passive construction that is used to describe certain objects in given sentences leading to verb-less sentences that do not have an obvious meaning as the example in Table (8) below shows:

Table (8): Errors at passive construction arrangement level

Ex.	Source Text	Google Translation
10.	Substances dissolved in a solvent are called solutes.	تسمى المواد المذابة في مذيب.

In **Ex.10**, the passive construction is not identified by GT which results in an incomplete sentence since the sentence suggests that there is a name for the materials that dissolve in water but this name is not given in the Arabic sentence for GT fails to put the sentence in the correct order to

come up with a correct passive construction. Thus, readers may expect to find a concept that refers to those substances that dissolve in a solvent but they end up with an incomplete sentence.

This proves that GT fails to parse the relative clause that is used to describe the term "solutes". In other words, GT fails to retrieve the underlying structure of the relative clause which states that substances that are dissolved in a solvent are called solutes. GT fails to come up with a linking word that helps to get a meaningful sentence which in this case could be the linking pronoun "التي". Thus, an acceptable Arabic translation that needs to be inserted among the options list may be: " تسمى المواد التي تذوب في مذيب/محلول مواد مذابة" .

3.3.2.2. Passive Inflections

GT fails to use the appropriate inflections that indicate that the sentence is passive as in the verb "تنجذب" which does not have any inflections to indicate whether it is an active verb "تَنْجَذب" or a passive one "تُجْذب" as the sentence in Table (9) shows:

Table (9): Errors at passive inflections level

Ex.	Source Text	Google Translation
11.	Polar substances and ions dissolve in water because opposite charges are attracted to the appropriate ends of water.	تذوب المواد والأيونات القطبية في الماء لأن الرسوم المعاكسة تنجذب إلى نهايات المياه المناسبة.

The underlying structure of the Arabic sentence is that polar substances go and move out of their will while the English sentence states

that they are moved by an external force. In other words, they do not move out of their will instead they are attracted by a non-mentioned force. This lack of inflections leads to two readings of the constituent "are attracted". However, in light of this structure or other kinds of structures, the word two has to disappear and replaced by oneness particularly in scientific texts as a genre.

Therefore, the researcher suggests that passive constructions should receive double attention from the software developers since such type of texts is loaded with passive structures for the focus in scientific texts is on the scientific facts rather than the ones who came up with those facts. Thus, the researcher suggests a procedure explained in Figure [**7**]. First, the sentence passes through the prepossessing step to analyze all its elements. In other words, GT should parse the sentence correctly through identifying the grammatical subject and object. Second, GT should identify the pattern of the sentence: "whether is it an active /sub. +v. +obj./ or passive /obj.+ v.+ sub./?". Next, if the sentence conforms to the pattern of /obj.+ v. +sub./, an extraction step of the appropriate passive inflections should take place. However, at present, GT cannot insert the appropriate passive inflections.

Thus, users can suggest a translation for the passive sentence below through assigning the appropriate diacritics in Arabic to make the sentence meaningful. The unmarked diacritic used in Arabic to indicate the passive construction is: *damma* which is attached to both the verb and the grammatical subject that follows it. Such suggested translations would be

kept in the lexicon to be reused again by different users having the same input.

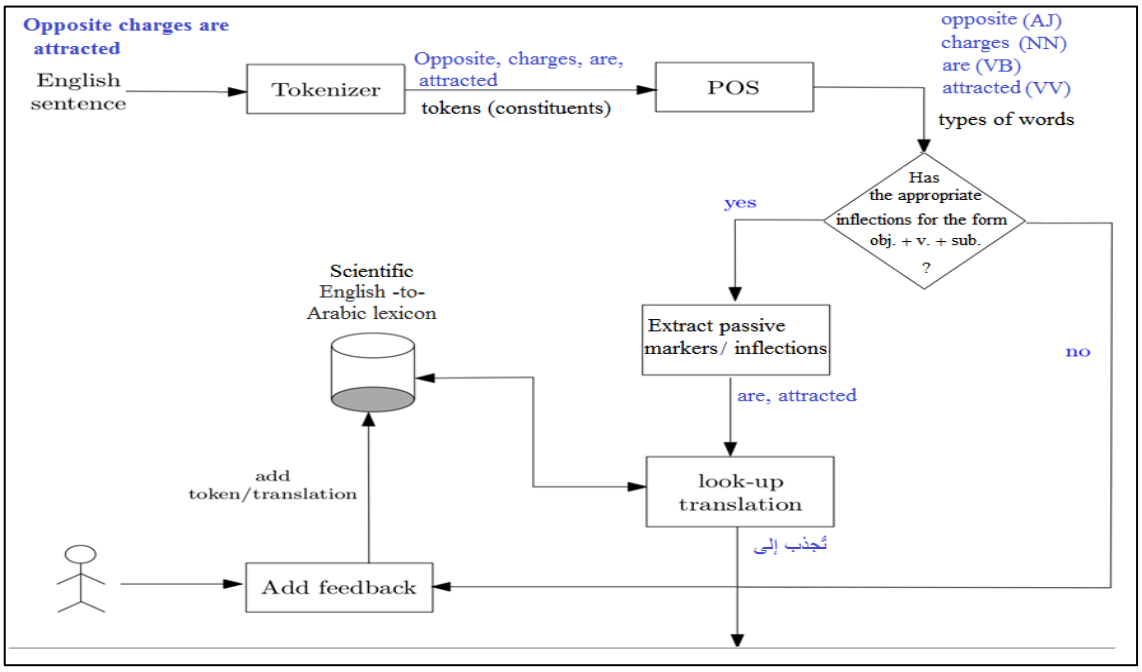


Figure (7): Processing of passive constructions.

3.3.3. Unnecessary Derivation for Certain Words

GT randomly selects a word in the input sentence then derives new forms from that word and inserts those forms in the output. However, this derivation is sometimes done at the expense of other functional/content words in the same sentence. Thus, this may lead to loss in meaning and redundancy in the output such as: "مشحونة الشحنة" and "كمفاعل متفاعل" in the examples in Table (10). Such errors occur when GT fails to identify and choose the correct part of speech to be used and that best completes the sentence.

Table (10): Redundancy due to unnecessary repetition

Ex.	Source Text	Google Translation
12.	Ionic bonds are electrical attractions between oppositely charged ions.	السندات الأيونية هي عوامل جذب كهربائية بين أيونات مشحونة الشحنة
13.	Aerobic respiration consumes oxygen as a reactant to complete the breakdown of a variety of organic molecules (aerobic is from the Greek aer, air, and bios, life).	يستهلك التنفس الهوائي الأوكسجين كمفاعل متفاعل لإكمال تحليل مجموعة متنوعة من الجزيئات العضوية (الهوائية هي من الهواء الجوي والهوائي والسير اليونانية ، الحياة).

In **Ex.12**, the constituent "oppositely charged" is rendered incorrectly as "مشحونة الشحنة" while it means "متعاكسة الشحنة". Thus, these unnecessary derivations from the word "charged" took the place of the word "oppositely". Accordingly, the word "oppositely" was neglected in the translation. This makes the meaning of the sentence not clear since the constituent "oppositely charged" indicates both positive and negative charges. In other words, this idea is lost in the Arabic text with the omission of the word "oppositely".

In addition, the word "reactant" in **Ex.13** is repeated while the sentence indicates that the oxygen is used as an initiator or a reactant in the process of aerobic respiration. These random and unnecessary derivations made by GT and result in omitting content words that contribute to the meaning of the sentence serve as alerts for GT team to reconsider their translation software. Thus, the research suggests that this defect could be solved through developing an automatic checking list that is responsible for making sure that all the elements in the translation box are translated

without any constituent being deleted from the input through scanning and counting both the elements of the input and the output and make sure that all the elements are translated.

3.3.4. Pronoun Translation

Pronouns are used to link and demonstrate the relationships that hold between the constituents of the sentence. In a few words, they are employed in the sentence to indicate to what noun the pronoun refers. The researcher traces a number of errors made by GT at this level. These errors include:

3.3.4.1. Relative Pronouns Referent/s

Some pronouns are used to provide the readers with the exact meaning that help them to understand the relationship that holds between the parts of the sentence. However, in cases where relative pronouns are employed to emphasize the functions associated with their referents and give complete thoughts about them, the researcher observed that GT mistreats such type of pronouns that are used as subordinating tools to the idea/s in question as the sentence in Table (11) shows:

Table (11): Errors in assigning the correct referent

Ex.	Source Text	Google Translation
14.	Electrons are always in motion, found in orbitals located at fixed distances outside of the nucleus called electron shells, which correspond to different energy levels	تكون الإلكترونات دائماً في حالة حركة، وتوجد في المدارات الواقعة على مسافات ثابتة خارج النواة التي تدعى الأصداف الإلكترونية ، والتي تتوافق مع مستويات الطاقة المختلفة

The pronoun "التي" is preceded by three different nouns. This means that the pronoun used in the Arabic sentence may refer to: The last noun as its referent which is "الأصداف الالكترونية", the middle nouns "المدارات" " or "النواة" or the noun at the beginning of the sentence "الالكترونات". Thus, this failure to identify the referent of the pronoun when there is a number of preceding nouns that all could serve as referents for that pronoun reveals that GT is still unable to process the sentence and link both the pronoun and its suitable referent correctly.

Therefore, the research recommends that GT team software their translation program to follow the procedure described in Figure [**8**]. First, GT should undergo the preprocessing step to decide on whether there is a relative pronoun or not. If yes, then GT needs to extract the subject of the pronoun in question through identifying the verb that follows it. Thus, the verb will help GT get the referent of the pronoun in the present case. In essence, the verb "correspond" in sentence **Ex.14** has a plural subject since it does not have the form "corresponds" which requires a singular subject. Thus, "nucleus" will be excluded along with the far-distant nouns from "which": "electrons", "orbitals" and "distances". Third, GT should insert a linking word which emphasizes that the last noun before the pronoun is the intended referent for the pronoun "which". Accordingly, users can suggest translations such as: "والذي بدوره" , "والتي بدورها" to help GT solve this kind of ambiguity. Such constructions indicate that the pronoun refers to the last noun since they add a sequence to the sentence which in turn allows the last

noun to connect with its pronoun without leading to ambiguity. Such suggested translations will be kept in the lexicon to be reused again in similar circumstances.

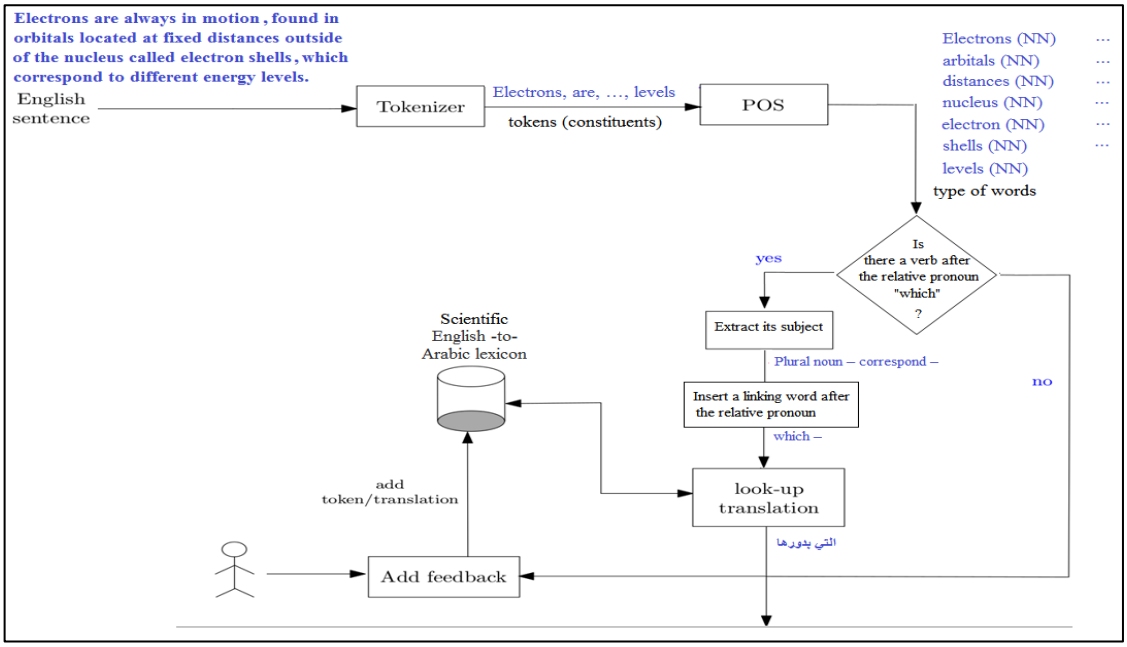


Figure (8): The process of matching the relative pronoun with its referent

3.3.4.2. Pronouns Refer to Gender Neutral Nouns

GT fails to come up with the appropriate gender marker that helps to assign and make clear to which noun the pronoun refers. In some cases, using the appropriate gender marker depends on the preceding noun to decide on the gender marker to be used with the verb. However, GT fails to achieve this connection between the pronoun and the verb since the English word "radiations" is not inflected for gender while the Arabic word "الاشعاع" is classified as a masculine noun. Thus, the existence of gender neutral nouns in English led GT to end up with inappropriate pronouns or gender markers that are attached to the verb. This happens because GT

does not have the ability to process, comprehend and make such connections like human translators as the sentence in Table (12) below shows:

Table (12): Errors in using gender markers

Ex.	Source Text	Google Translation
15.	Some isotopes are radioactive, emitting radiation as they decay.	بعض النظائر المشعة ، التي تنبعث منها الإشعاع لأنها تتحلل.

In **Ex.15**, GT uses "تنبعث" instead of "ينبعث" failing to understand the meaning or retrieve the subject of the sentence since the referent does not come immediately after the verb. In another words, there is a distance between the noun "الإشعاع" and its pronoun "هو". There is a preposition "منها" between the noun "الاشعاع" and the invisible pronoun in the verb "ينبعث" that is "هو" in the Arabic sentence. Thus, GT fails to bring close the verb and the subject in the Arabic sentence and decide on the pronoun to use; it fails to drop the word "منها" and read the sentence as " ينبعث الاشعاع منها"; this may ease the process and focus the attention of GT on the subject and the verb in order to come up with the appropriate invisible pronoun that refers to "الاشعاع". Accordingly, this defect is manifested by using "تنبعث" instead of "ينبعث".

Therefore, to solve this problem, the research suggests a procedure described in Figure [**9**]. First, GT should undergo the prepossessing step to analyze all the tokens that make up the sentence along with their grammatical categories. Second, GT needs to decide on whether the nouns in the sentence are gender neutral in Arabic or not. Thus, if the nouns are

not gender neutral, then GT will undergo a gender extraction step to come up with the appropriate gender markers, feminine/masculine markers, and then map on their grammatical features to come up with the appropriate gender marks to be attached to the verb. Thus, at this stage, users can help GT identify the nouns which require feminine/masculine markers in Arabic through providing GT with the appropriate makers to be attached to the verb by suggesting their own translation which in turn will be added to the lexicon to be reused again by different users.

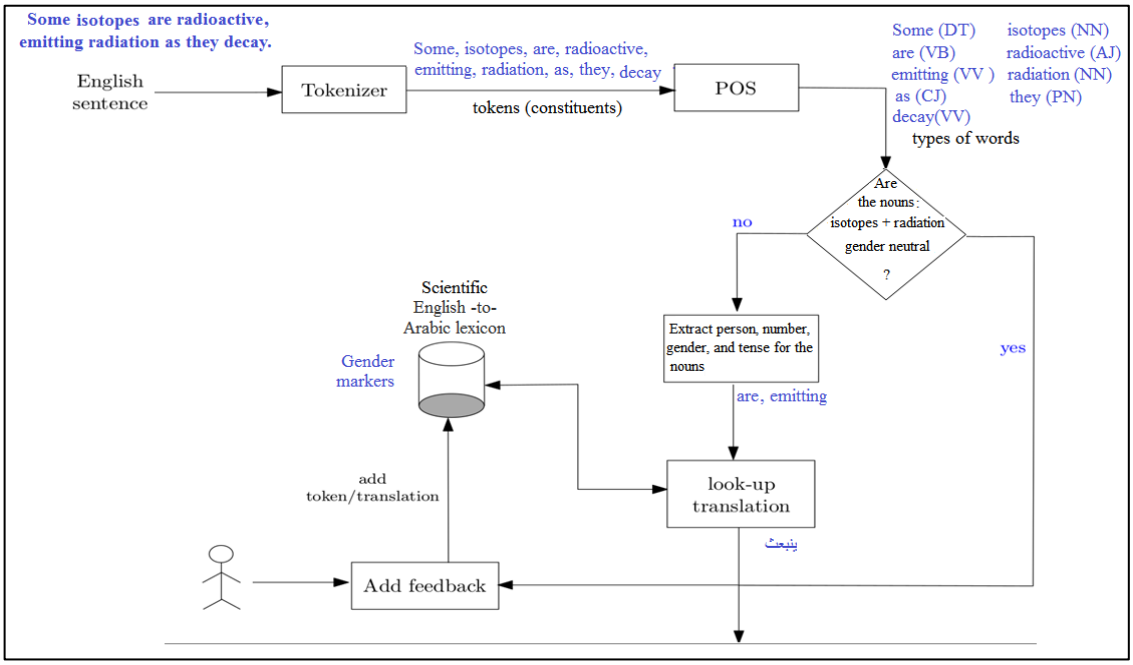


Figure (9): The process of assigning gender to both the sub. and the verb.

Moreover, the translation reveals that the pronoun "them" in the English sentence in **Ex.16** in Table (13) that could be used to refer to both plural masculine and feminine nouns is not rendered correctly in the Arabic sentence. To put it another way, the head noun in the sentence is "a certain type of molecules that attract water and lipids", so the last pronoun in the English sentence "them" refers to this type of molecules. Thus, in the

Arabic sentence, GT shifts this pronoun from a pronoun that is used to refer to a plural noun in English to a plural noun in Arabic. Yet, the problem occurs when GT uses the masculine pronoun "لهم" to refer to those molecules which require a feminine pronoun since they mean "الجزيئات" in Arabic for Arabic is considered a gender-marked language while English is not. In other words, GT fails to come up with the pronoun that marks masculine and feminine aspects with the referent. Thus, the pronoun, "لهم ", is confusing since it takes a masculine noun as a referent while the text contains only feminine referents such as: "المستحلبات ,"الجزيئات" ,"المواد". Such nouns require the pronoun "لها" to be emphasized in a given sentence.

Table (13): Errors made at pronoun level

Ex.	Source Text	Google Translation
16.	Amphipathic molecules make good emulsifiers because they can attract both hydrophobic substances and hydrophilic substances to them.	يجعل amphipathic جزيئات المستحلبات جيدة لأنها يمكن أن تجتذب كل من المواد مسعور والمواد المحبة للماء لهم.

This example shows that GT still errs in making the right connection between the pronoun and its referent. In other words, GT needs to be enhanced in order to be able to mark the feminine and masculine aspects that make the sentence sound coherent and meaningful. Therefore, it would be better for GT team to software their machine translation program in a procedure described in Figure [**10**]. First, GT will undergo the prepossessing step that will enable it to identify both the pronoun and its referent in the first stance; then GT needs to decide on whether the pronoun refers to a gender neutral noun or not. If not, then GT will undergo two

extraction steps: The first one is number extraction where both the pronoun and its referent agree in number while the second is gender extraction step where both the noun and its pronoun agree in gender. Thus, in the present sentence, users could suggest their own translation for the pronoun "them" as "لها" to be saved in the lexicon to reuse it again whenever the pronoun "them" reoccurred with the noun "molecules".

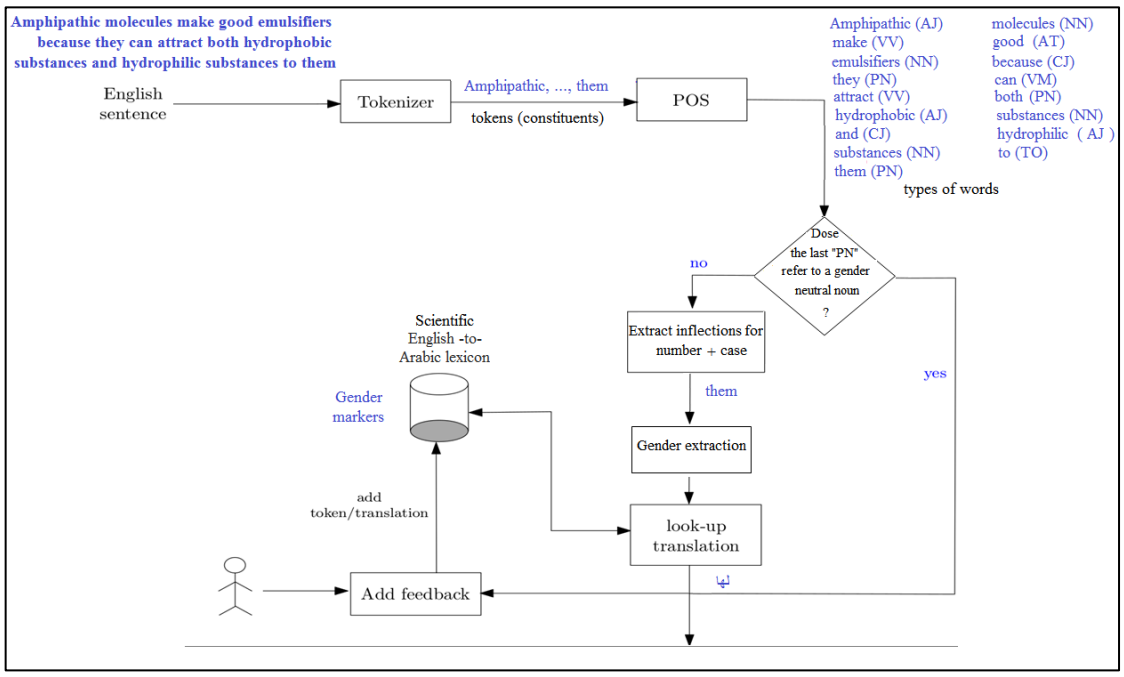


Figure (10): The process of gender matching between the pronoun and its antecedent.

3.3.4.3. Phrasal Verbs Meanings along with their Gender Marked Subjects

In some cases, GT fails to recognize the gender of the subject of the verb. Thus, **Ex.17** in Table (14) shows that the subject of the verb "يحافظ" is the noun "the hydrogen bond" which is a feminine subject not a masculine in Arabic. This means that the verb "يحافظ" requires a feminine marker not a masculine which is in this case "تحافظ". However, GT uses the masculine

marker in the verb which is "يحافظ هو" although there is a feminine subject that is "الرابطة الهيدروجينية" and it requires a feminine marker like "تحافظ هي". This shows that GT fails to make the right connection between the verb and its subject. This might make students confused for they start with the verb "يحافظ" while the key word in the sentence is "The hydrogen bond" which is an established scientific term and it means "الرابطة الهيدروجينية" which is a feminine subject in Arabic as shown in Table (14**)** below:

Table (14): Errors in SV agreement

Ex.	Source Text	Google Translation
17.	The hydrogen bond keeps the molecules far enough apart to make ice about 10% less dense than liquid water at 4 C.	يحافظ الرابط الهيدروجيني الجزيئات على مسافة بعيدة بما يكفي لجعل الثلج أقل كثافة بنسبة 10٪ من الماء السائل عند 4 درجات مئوية.

Moreover, GT fails to come up with the correct meaning of the verb "keeps". It fails to benefit from "word collocations" than may come with the verb such as: the word "مسافة" which denotes the ideas of "a distance from something" in the example above. Thus, in **Ex.17**, GT fails to understand that "keeps" means "يبقي الجزيئات على مسافة بعيدة". In other words, GT goes with the literal meaning of the verb which, in turn, leads to a semantic shift in the output. This happened for GT fails to benefit from the present textual context and draw on words like: "مسافة", "far enough apart" since they indicate that there is an area where the hydrogen bond keeps those molecules away and prevents them from entering it.

Therefore, the researcher suggests that GT be programmed to follow a procedure explained in Figure [**11**]. First, GT should undergo the

preprocessing step to decide on whether the verb is followed by adverbs/prepositions that make its meaning clear or not. If yes, then GT should undergo an extraction step for those adverbs/prepositions. Second, GT needs to compose the meaning of the verb based on those adverbs. Yet, at this stage, GT is unable to extract the meaning of the verb in question, "keep" using the words, "far", "enough" and "apart". Thus, users can assist GT through suggesting a translation for the verb in question such as "يحافظ على مسافة". Such suggested translation will be added to the lexicon to be reused again in similar constructions. Thus, such procedure might help GT come up with the appropriate word that fits in the present context.

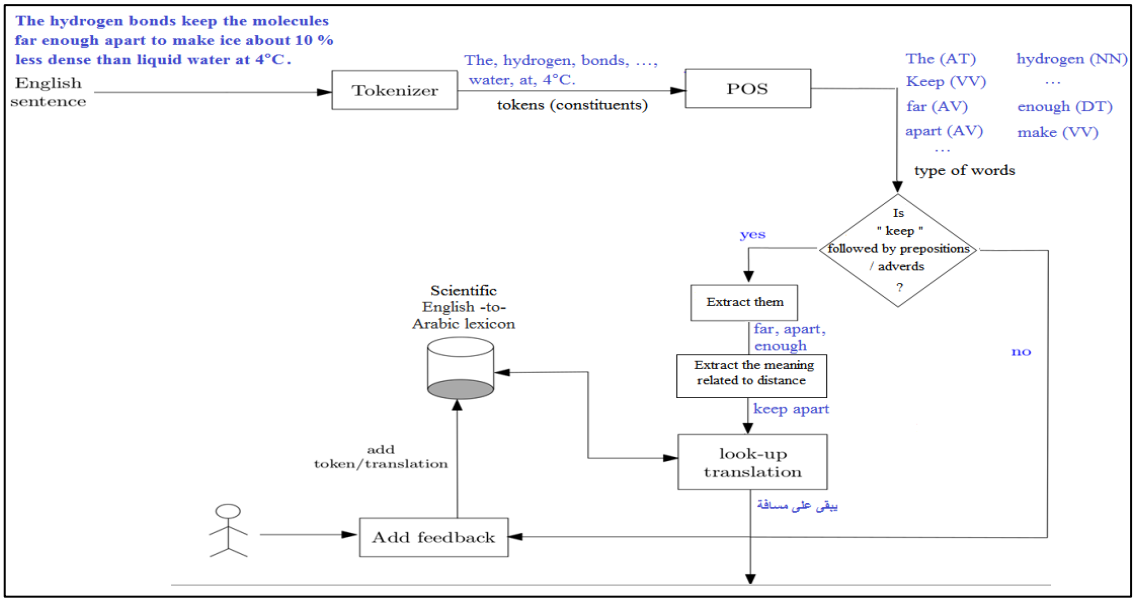


Figure (11): The process of composing the meaning of the phrasal verb "keep"

3.5. Conclusion

The data revealed that using GT to translate single sentences from English to Arabic uncovers a number of efficiencies in areas related to the structure of the sentence and the relationship between its parts. Thus, GT performs well when it comes to pluralization since English makes it clear and easy for GT to recognize it as it forms singular, plural and dual in two ways: morphologically and lexically. The former is done through the use of +/- S while the latter is achieved through lexical words such as two or both. Moreover, GT renders the simple present with main verbs and also modals such as *may* and *can* correctly. No errors were observed in this area except when it comes to the verb *be* that is used to describe the subject of the sentence. In some cases, GT fails to handle it leading to incomplete sentences.

As for machine deficiencies, GT fails to recognize certain structures that help in understanding the intended meaning. Also, inflections that contribute to the meaning of the sentence are mishandled. Such defects may affect the meaning since the meaning of the sentence could be achieved by assigning the appropriate inflections to the word in question for inflections along with their grammatical functions are capable of making the meaning transparent. For example, the Arabic language uses certain inflections or compensational tools such as diacritics (ُ ,ِ) to indicate whether a given word is a subject or object or whether the sentence is active or passive while the English language does not use such kind of

inflections because it depends on word-order to assign the function of the words instead of using inflections which in turn pauses a challenge for GT during the process of translation.

Accordingly, GT still commits errors in cases where there are two successive nouns without inflections since it does not make a distinction between the subject and the object which leads to a sentence with two readings. Another defect is that GT is unable to break the words, phrases, and sentences down into their building blocks correctly then rearrange them since it depends on word-for-word translation which results in wrong pronoun reference. This in turn leads to GT failure in assigning the referent of certain pronouns since there may be many items to which the pronoun refers to beside issues related to the language itself whether it inflects nouns for gender, number, etc. In addition, this chapter shows GT inadequate understanding of the affixes used in a given sentence along with their function since the word may have the same form but its function may differ according to the context in which it is used.

Moreover, errors were observed in passive structures which were translated incorrectly due to GT failure to benefit from the key words in the text that indicate that the sentence is a statement about the past, a relative clause without the relative pronoun and auxiliary, or a passive construction. This is due to the limited competency of GT in relation to understanding which in turn is reflected in its performance or its output.

To conclude, these defects may be contributed to two main reasons which are: Firstly, GT tends to be source text oriented in case where both the SL and TL have the same grammatical category, thus how each language makes that grammatical category manifest confuses it such as: noun phrase constructions in both English and Arabic. Secondly, GT does not have the ability to go forth and back between both the ST and TT to come up with an acceptable translation for the input. In other words, the data revealed that GT fails to shift between SL and TL whenever it encounters a grammatical category that is missing in one of the two languages in question such as: the case of gender neutral nouns in English that are missing in Arabic since Arabic is gender marked. Thus, GT fails to use the appropriate compensational tools in the output which in turn led many errors to float in the translation box.

However, the researcher maintains that these errors may be solved since GT provides its users with "a feedback button" which in turn helps to enhance the quality of the output for the concepts and descriptions present in this chapter that are related to inheritance, nucleus, elements and compounds are all examples of controlled areas where the terms used in expressing ideas related to them are usually the same regardless of place or time. In other words, such branches of science contain very controlled syntactic structures and lexis that are valid for all times and circumstances. Thus, allowing the users to interact with GT and suggesting translations for the existing errors could help in approaching high quality translations

where all the linguistic components of the output are concordant with one another.

Therefore , in an attempt to make it easy for GT, students may decide to apply the subject they are studying- Biology- on GT; this means that students may shift from reading biology to doing biology by feeding GT with paragraphs instead of using singles words, phrases and sentences in an attempt to help GT give them the meaning of those elements according to their function in paragraphs for GT may benefit from the key words that appear in the texts to give correct translations such as time markers like: /s/-present forms, later, last, etc. that may help to decide the tense correctly. This is called "symbiosis" in biology where students and GT try to help and benefit from each other exactly as certain types of organisms do to help and feed each other in order to stay alive but will GT provide students with good/acceptable translations for paragraphs/texts instead of dealing with individual sentences? This question is going to be discussed in the next chapter.

Chapter Four

Cohesive Markers at Paragraph Boundaries in Google Translation

4.1. Introduction

In the third chapter, the researcher examined GT performance at smaller and focused level which is the sentence. This chapter will be devoted to examining GT performance at paragraph level to test the machine efficiency to cope with "Text Linguistics" trends that move the center of attention from single words, phrases and sentences to longer stretches of language such as paragraphs and even whole texts; Beaugrande and Dressler stated that "… there was no established methodology that would apply to texts in any way comparable to the unified approaches for conventional linguistic objects like the sentence." Yet, this discipline takes "…the text as the primary object of inquiry" (1981, p.14).

Since the 1970s, translation scholars have shifted focus to the relationships which bind sentences and paragraphs together. Enkvist states that "a sentence is not autonomous, it does not exist for its own sake but as part of a situation and part of a text" (1978, p.178). Thus, understanding the potential meaning of a sentence depends on the communicative event in which that sentence resides, i.e. the text. Along the same lines, Brown and Yule identified the communicative event as "an instance of language in use rather than language as an abstract system of meanings and relations" (1983, p.6). Moreover, Beaugrande and Dressler identified seven standards

of textuality; they maintained that, for texts to be communicative, these standards must be met: "cohesion, coherence, intentionality, acceptability, informativity, situationality and intertextuality" (1981, p.19).

Subsequently, contrastive translation studies seek to compare how each language deploys such textual standards to communicate information; Hatim maintains that "the word is thus no longer sufficient as a unit of translation" (2001, p.33). Moreover, Baker states that "Every language has its own battery of devices for creating links between textual elements". For example, when conjunctions are used as cohesive devices to express relations, Japanese and Chinese prefer to deploy simpler structures while German tends to use complex structures and subordination (1992, p.188). Moreover, English is a subordinating language that employs connectors to communicate certain types of messages while Arabic is a coordinating language for it uses "and", in Arabic, "و" heavily to compose paragraphs and even whole texts. Yorkey noted that, "Teachers at the American University of Beirut refer to the *wa wa* method of writing because of the Arabic *wa* 'and', which is exceedingly used as a sentence-connector" (1974. Retrieved from: https://urlzs.com/3vqPt). Thus, such comparative studies will be of important value when translating texts between two different languages for each language has its own tools, techniques and rhetoric that it uses to express the same message.

Hence, text cohesion and text coherence are two important dimensions which hold the text together and pinpoint to the relationship

among its elements in English and Arabic alike. The former dimension refers to the grammatical relations which hold between the different sentences or clauses that make up the text. Grabe defined cohesion as "the means available in the surface forms of the text to signal relations that hold between sentences or clausal units in the text" (1985, p.110). Coherence, on the other hand, is concerned with "the conceptual relations that underline the surface text". Accordingly, Halliday and Hasan in *Cohesion in English* relate cohesion to both grammatical and lexical devices. The grammatical devices include: ellipsis, reference, substitution and conjunctions, while the lexical devices include: repetition of lexical items, synonyms, subordinates and collocations (1976).

Thus, the structure of the scientific paragraph may be problematic for GT since scientific paragraphs can function as a descriptive tool, to describe something, or as an explanatory tool, to explain the cause or result of something or they can be used to highlight the sequence of a certain process. However, in all case, scientific texts must show common features like: First of all, technicality which means that the language must be clear, direct and the terms must be related to the field of study. Secondly, texts have to show density and development of information from clause to clause as the text evolves. Finally, there must be connections that present time, causes, conditions, contrast and other linkage (Palincsar, 2013. Retrieved from: https://urlzs.com/ZBv4f).

On that account, this chapter will be devoted to examining areas of eff./def. of cohesive markers in scientific texts translation performed by GT from English into Arabic since each language has its peculiar cohesive patterns that it uses to communicate coherent information in such text type. This creates a challenge for GT in the translation action since GT needs to realize the techniques by which each language tailors both cohesive and coherent texts. In other words, English scientific texts are characterized by: a) rigid word order; b) very few inflections; c) use of abbreviations, formulae, acronyms and registers; d) clear-cut tense aspect distinction; e) adverbs are generally formed by the (ly) affixation to adjectives; and f) technical and scientific terminology covers all relevant domains including: biology, chemistry, medicine, computer science, etc. Arabic, on the other hand, is characterized by: First, flexible word order. Second, highly inflectional. Third, rarely uses abbreviations, formulae and acronyms. Fourth, there is no clear-cut tense aspect distinction. Fifth, adverbs are established by prepositional pre-modification of adjectives and nouns; English prepositions such as *above*, *over*, *before*, *after*, *below*, *under*, *behind*, and *between* are adverbs in Arabic. Finally, lack of scientific and technical terminology which might cover fields of e.g., biology, chemistry, physics, etc (Al-Hassnawi, 2013. Retrieved form: https://fc.lc/ZAS6DlH).

Thus, such differences will likely cause deficiencies in any GT output; the machine will need to recognize the cohesion patterns in each language per se since any lack of awareness of the textual techniques in the

language(s) in question including: lexical terms chains, reference to both time and participants and connectors that show time, cause, result, sequence, etc. may lead to machine errors. This in turn may affect the quality of the translation output. Therefore, this chapter aims to test GT performance primarily at the cohesive and coherent levels then suggest recommendations for GT improvement at these two levels.

Thereupon, the researcher will highlight the cohesive devices employed in the source texts, in particular, reference, conjunctions and lexical cohesion. Substitution and ellipsis involve some kind of deletion so they are not frequent devices in scientific paragraphs; hence, Al-Hassnawi states that "There is no insertion, substitution, or permutation" in scientific texts (2013. Retrieved from: http://www.translationdirectory.com/ article10.htm). Next, the researcher will trace those devices in the translations produced by GT to examine the way GT reconstructs and reproduces such cohesive devices in Arabic. The devices used for reference are highlighted in red, conjunctions in blue and lexical cohesion in green. Finally, the researcher will draw on GT overall performance in the selected paragraphs to pinpoint the cohesive device(s) best treated by GT and the one(s) reproduced in low quality. These findings will help in identifying the areas of def. to suggest solutions for them to enhance GT performance at the cohesive level.

4.2. Areas of Eff./Def. of Lexical Markers in Description Paragraphs

4.2.1. Cohesion in Introducing Definitions

Description paragraphs are used in scientific texts to inform the readers about how things look like or what concepts mean or refer to. Thus, description paragraphs should employ different types of linking words to smooth the travel between the sentences that make up the whole paragraph to help the readers understand the objects/terms being described; these signals are used to introduce definitions and terms within structures like: "is called", "is termed" and "is named". These signals indicate that a definition will follow, so the reader will be prepared and ready to receive the concepts' definitions. Moreover, adjectival clauses are used heavily in description paragraphs after the terms since they aim to give more details about the concepts in question. Such signals play a crucial role in communicating the information in the paragraph in question correctly with optimal consideration of the min-max principle; minimum effort is required to process the information.

However, paragraph **1a** indicates that those signals appear to be problematic for GT; it has been observed that GT fails to recognize the correct structure to introduce definitions in cases where the main concept is fused into the sentence. In other words, when the definition comes first in the paragraph while the concept/term comes at the very end of the sentence, GT mistreats this type of relationship that holds between the definition and the term it defines. This kind of relationship is usually expressed through

the use of relative clauses that include patterns like: "is called" and "is termed" respectively as shown in the English text below:

Paragraph 1a:

Mendel used the scientific approach to identify two laws of inheritance. In the 1860s, George Mendel discovered the basic principles of heredity by breeding garden peas in carefully planned experiments. Mendel chose the garden peas for his studies because: garden peas are available in many varieties. For example, one variety has purple flowers, while another variety has white flowers. A heritable feature that varies among individuals, such as flower color is called a character. Each variant for a character, such as purple or white color for flowers, is termed a trait.

Paragraph 1b:

استخدم مندل المنهج العلمي لتحديد قانونين للإرث. في ستينيات القرن التاسع عشر ، اكتشف جورج مندل المبادئ الأساسية للوراثة عن طريق تربية البازلاء في الحديقة في تجارب مخططة بعناية. اختار مندل البازلاء الحديقة لدراسته لأن: البازلاء حديقة متوفرة في العديد من الأصناف. على سبيل المثال، يحتوي أحد الأنواع على زهور أرجوانية ، بينما يحتوي تنوع آخر على زهور بيضاء. ويطلق على ميزة وراثية تختلف بين الأفراد، مثل لون الزهرة. ويطلق على كل متغير للحرف، مثل اللون الأرجواني أو الأبيض للزهور، سمة.

Accordingly, a comparative analysis of paragraph **1a** and its translation reveals that the ST employs three main cohesive devices to communicate its semantic load correctly. These devices include: First of all, reference that is employed to refer to both participants and time. The reference is manifested in paragraph **1a** through the use of the proper noun "Mendel" in the opening sentences along with its possessive adjective 'his

study' which agrees with its antecedent "Mendel" in both gender and number. The name "Mendel" is repeated three times then it is replaced by the pronoun "his" in the fourth time. In other words, the repetition of the name here is functional since it aims to emphasize the scientist's name and make it salient in the paragraph. Moreover, the ST makes reference to both the past and the present. In other words, when the text talks about an experiment that is done and ended in the past, it employs the past tense which is manifested through the use of the verbs "used", "discovered" and "chose" respectively. Yet, when the ST moves to generalizations about garden peas' location, it shifts to the present tense through using verbs such as: "are" and "has".

Secondly, the ST deploys a number of domain specific terms such as: "inheritance", "trait" and "heritable feature". Such lexical chain keeps the focus of the paragraph and supports the facts being stated in the text; such chain contributes to the lexical cohesion of the paragraph as one unit. Finally, the ST comprises a number of conjunctions used to express reason, provide justifications, mark contrast and introduce examples such as: "for", "because", "while" and "for example". All these lexical markers play a crucial role in making the ST informative, communicative and cohesive.

However, mapping those cohesive devices on GT translation to test its areas of eff./def. in lexical markers reveals that reference as a cohesive device which is used to denote both people and time is reproduced correctly in the translation since GT resorts to literal translation through the

repetition of the word "مندل" and the use of the past tense manifested in the verbs: "استخدم", "اكتشف" and "اختار"; the shift to the present tense in "متوفرة" and "يحتوي" is appropriately conveyed in the Arabic text as well. Moreover, conjunctions and lexical cohesion in Google translation are handled correctly since GT employs literal translation of conjunctions such as: "because" which means "بسبب" and "while" which means "بينما", beside it employs domain specific terms such as:"الأصناف","مبادئ الوراثة" and "ميزة وراثية", except the word "inheritance" that is mistranslated as "الإرث" instead of "الوراثة" in the Arabic text.

However, this relatively good accuracy level in GT treatment of linking words is broken in the last two sentences in paragraph **1a** since the definitions of the concepts of "character" and "trait" are mistreated by GT. In other words, paragraph **1a** contains two main concepts related to two heredity features which are "character" and "trait". Thus, these two terms denote characteristics associated with living organisms according to the inheritance law. However, when comparing the ST with the Google translation, it is observed that GT fails to reproduce the parallel structures used in the English text to introduce those terms/concepts since it uses incomplete or unclear sentences. In other words, GT goes with literal translation of the present parallel structures which in turn leads to errors in the Arabic text as shown in Figure [**12**]:

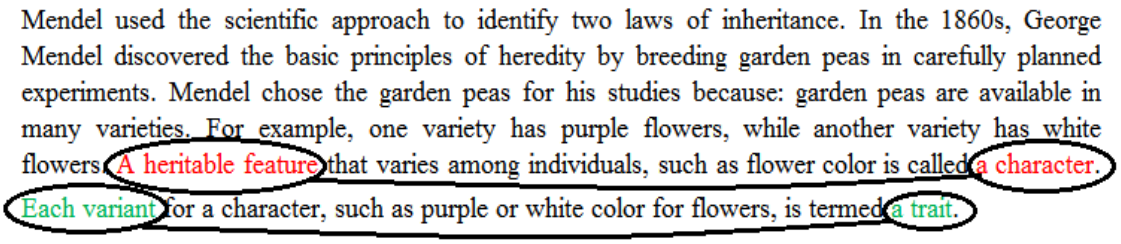


Figure (12): Parallel structures used in introducing definitions

Thus, this confusion happened because the English paragraph deferred the naming of the concepts until the very end in each sentence; GT encountered first the adjectival clause "that varies among individuals", which separates the subject and its predicate, and then the term appears. Accordingly, GT fails to connect them since this gap leads GT to detach the first part of the sentence: "a heritable feature that varies…" from the second part: "is called a character". This results in lack of cohesion in the Arabic text. Moreover, GT fails to provide the appropriate conjunctions such as: "تسمى" and "يطلق عليها" in Arabic. Such words make the message clearer and more cohesive than GT did as shown in Figure [**13**]:

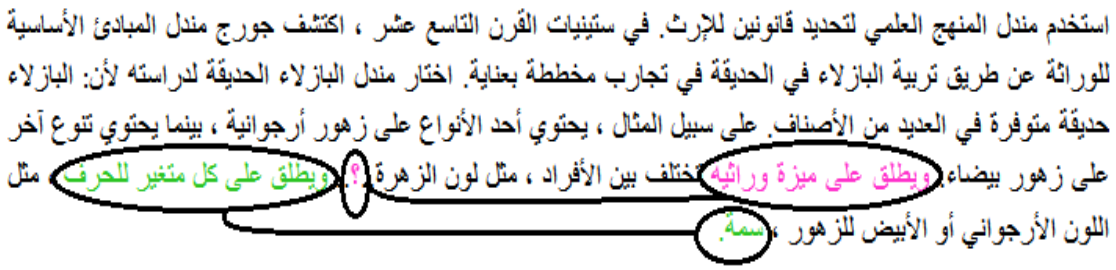


Figure (13): GT failure in reproducing adjective clauses

In such situation, one would assume that feeding the machine with a single sentence might solve the problem since single sentences are shorter than paragraphs so GT would translate them better. However, the problem

persisted and the machine committed the same errors at both sentence and paragraph level as shown in Table **(**15) below:

Table (15): Errors in relative clauses translation

Sentence Num.	Source Text	Google Translation
1.	A heritable feature that varies among individuals, such as flower color, is called a character.	ويطلق على ميزة وراثية تختلف بين الأفراد، مثل لون الزهرة.
2.	Each variant for a character, such as purple or white color for flowers, is termed a trait.	ويطلق على كل متغير للحرف، مثل اللون الأرجواني أو الأبيض للزهور، سمة.

To conclude, GT still does not have the ability to reproduce terms using the appropriate linking words which are frequently used in terminology definition; separating the term from its definition causes the confusion. Accordingly, the research recommends that GT be programmed to rewrite the sentences using the appropriate conjunctions that make the terms transparent. For example, introducing the definitions of both character and trait requires the presence of the appropriate linking words in Arabic which are used for such purposes like: "تسمى" or "يطلق عليها" .

4.2.2. Cohesion Achieved through Scientific Terms Chains

Definition paragraphs tend to employ lexical chains achieved through using synonymies, hyponyms and antonyms. Such chains help the readers trace the concepts being introduced and highlight the technical terms associated with the idea in question. This in turn makes the text transparent and easy for reading since such lexical chains contribute to text

cohesion and unity. For example, paragraph **2a** below introduces terms related to substances and their relationship with water. Thus, the ST establishes cohesion by using three main cohesive devices as highlighted below. Yet, the most salient device in the text is lexical cohesion.

Paragraph 2a:

Water is the solvent of life

Water is an excellent solvent for many substances because of its polar nature. Polar substances and ions dissolve in water because opposite charges are attracted to the appropriate ends of water. Strictly hydrophobic molecules, including most lipids, do not mix well with water. Some molecules have both hydrophobic and hydrophilic ends. Such molecules are said to be amphipathic. Amphipathic molecules make good emulsifiers because they can attract both hydrophobic substances and hydrophilic substances to them. Substances dissolved in a solvent are called solutes.

Paragraph 2b:

الماء هو مذيب الحياة.

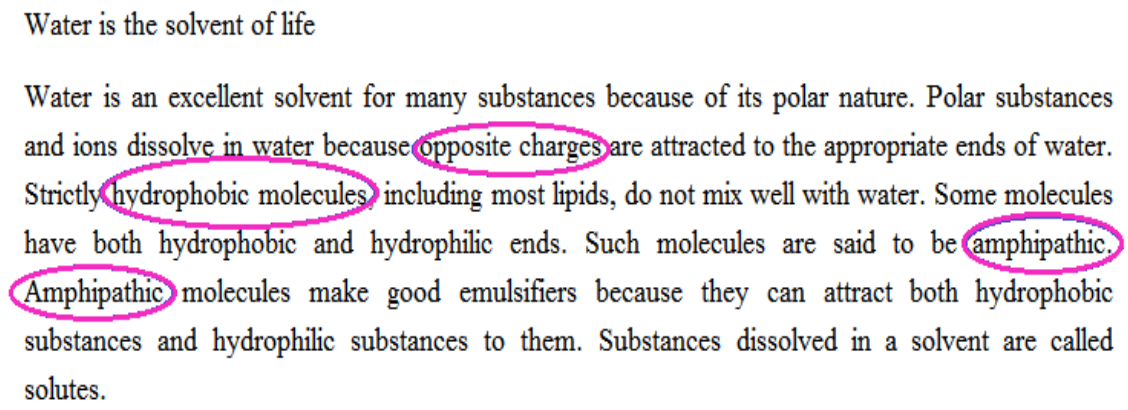
الماء هو مذيب ممتاز لكثير من المواد بسبب طبيعته القطبية. تذوب المواد والأيونات القطبية في الماء لأن الرسوم المعاكسة تنجذب إلى نهايات المياه المناسبة. لا تختلط الجزيئات الكارهة للماء بشدة ، بما في ذلك معظم الدهون ، بشكل جيد مع الماء. بعض الجزيئات لها نهايات مسعرة للماء. ويقال إن هذه الجزيئات تكون amphipathic. جزيئات amphipathic يجعل المستحلبات جيدة لأنها يمكن أن تجتذب كل من المواد مسعور والمواد المحبة للماء لهم. تسمى المواد المذابة في مذيب.

In other words, cohesion in paragraph **2a** is achieved through: First, reference that is manifested through the use of the pronouns "it", "they" and "them" respectively to avoid repeating their antecedents "water" and "amphipathic molecules" which might make the text redundant. Moreover, passive constructions are used as an aid for reference since such constructions minimize the number of participants/referents in any text. In other words, passive constructions hide the participants in cases where they are not necessary to be revealed. Thus, they make it easy for the readers to trace and focus on the terms and ideas being presented. Accordingly, the focus of paragraph **2a** is on "water" which is an inanimate subject so passive constructions are present to refer to characteristics related to water regardless of the scientists' names who come up with those concepts such as: "are said to be" and "are called". Secondly, conjunctions such as: "because" are used to justify why "water" is an excellent solvent and to explain the meaning of the terms presented in the paragraph. Finally, technical terms are used to refer to the features of substances that dissolve in "water" including the scientific terms: "amphipathic" and "opposite charges" and the antonyms: "hydrophobic" and "hydrophilic".

When examining GT performance, it is evident that GT handles the conjunctions correctly except in the last sentence which has the same error committed in introducing definitions explained earlier, "is called". Moreover, pronouns are correctly reproduced except when GT fails to translate the word "amphipathic". Thus, GT fails to assign the pronoun its

correct gender since GT keeps the word in English and does not give its equivalent term. This proves that cohesive markers interact with one another. Thus, any error at one level may lead to errors at other levels.

Furthermore, when it comes to the third cohesive device that is lexical cohesion, GT tends to break such lexical unity that is required to establish strong ties between the key words in the paragraph. In other words, in paragraph **2b**, GT fails to come up with the appropriate words that enable the reader to process the paragraph successfully as shown in Figure [**14**]:

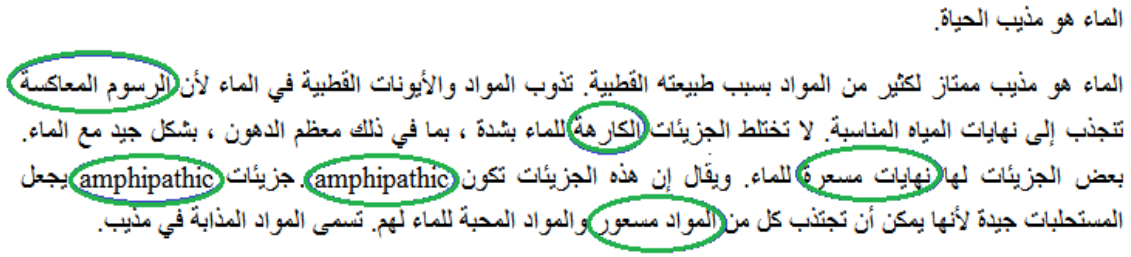

Water is the solvent of life

Water is an excellent solvent for many substances because of its polar nature. Polar substances and ions dissolve in water because opposite charges are attracted to the appropriate ends of water. Strictly hydrophobic molecules, including most lipids, do not mix well with water. Some molecules have both hydrophobic and hydrophilic ends. Such molecules are said to be amphipathic. Amphipathic molecules make good emulsifiers because they can attract both hydrophobic substances and hydrophilic substances to them. Substances dissolved in a solvent are called solutes.

Figure (14): The lexical chain used in paragraph 2a.

Paragraph **2a** states some facts about the properties of "water" as a solvent for many substances. However, GT fails to reproduce the lexical chain that is essential for achieving cohesion across sentences. In other words, GT fails to translate those key words in the English text correctly. Thus, GT is deficient at terminological level and this defect harms the lexical cohesion in scientific texts since GT replaces the scientific terms with words from everyday life in the Arabic text. These words are either

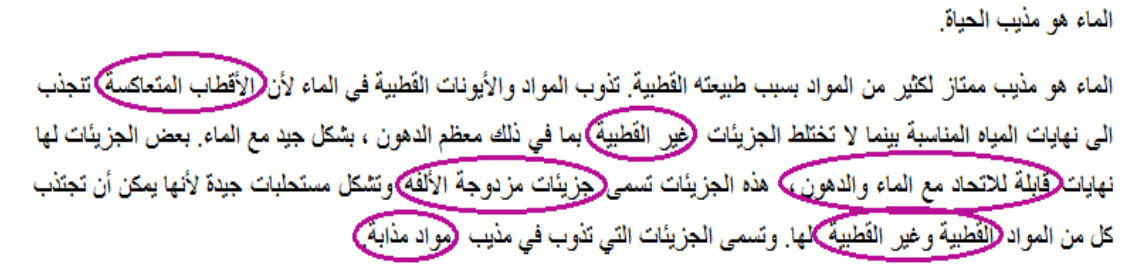
literal or mistranslations like: "الرسوم المعاكسة", "مسعور". Other words are kept in English without being translated in Arabic such as: "amphipathic". Thus, the word-for-word translation seriously harms cohesion as shown in Figure [**15**] below:


الماء هو مذيب الحياة.

الماء هو مذيب ممتاز لكثير من المواد بسبب طبيعته القطبية. تذوب المواد والأيونات القطبية في الماء لأن الرسوم المعاكسة تتجذب إلى نهايات المياه المناسبة. لا تختلط الجزيئات الكارهة للماء بشدة ، بما في ذلك معظم الدهون ، بشكل جيد مع الماء. بعض الجزيئات لها نهايات مسعرة للماء. ويقال إن هذه الجزيئات تكون amphipathic. جزيئات amphipathic يجعل المستحلبات جيدة لأنها يمكن أن تجتذب كل من المواد مسعور والمواد المحبة للماء لهم. تسمى المواد المذابة في مذيب.

Figure (15): Deficiency at terminological level in paragraph 2a.

Thereby, GT should use terms related to the topic in question so it has to be programmed to come up with words taken from dictionaries specialized in science to maintain the lexical cohesion of the descriptive paragraph and its communicative meaning. Therefore, an acceptable translation that may help the readers trace the topic of the paragraph and digest all the supporting sentences that contain key words which in turn refer to the topic in question could be as shown in Figure [**16**]:


الماء هو مذيب الحياة.

الماء هو مذيب ممتاز لكثير من المواد بسبب طبيعته القطبية. تذوب المواد والأيونات القطبية في الماء لأن الأقطاب المتعاكسة تتجذب الى نهايات المياه المناسبة بينما لا تختلط الجزيئات غير القطبية بما في ذلك معظم الدهون ، بشكل جيد مع الماء. بعض الجزيئات لها نهايات قابلة للاتحاد مع الماء والدهون، هذه الجزيئات تسمى جزيئات مزدوجة الألفه وتشكل مستحلبات جيدة لأنها يمكن أن تجتذب كل من المواد القطبية وغير القطبية لها. وتسمى الجزيئات التي تذوب في مذيب مواد مذابة

Figure (16): Suggested translation for paragraph 2a.

Correspondingly, the research recommends that GT be fed with the previously mentioned lexical markers that are frequently used in

description paragraphs in an attempt to help GT come up with the correct scientific terms then use them in the right place whenever it encounters paragraphs of such type. Thus, to solve this type of errors found in description paragraphs translation, the researcher suggests a procedure described in Figure [**17**]. First, the paragraph inputted by the user should undergo a "text preprocessing step" which includes: the process of splitting the paragraph into tokens along with their POS. Next, based on these tokens' categories, the system needs to identify whether there are lexical markers that help identify the type of paragraph in question or not. If yes, then those signals are used to decide on paragraph-type. After that, an extraction step of both the lexical markers and the equivalent terms appropriate to that paragraph type should take place. The translation of those items are looked up from a specialized "scientific lexicon". However, at present, GT direct translation for the term "variant" in paragraph **1a** is: "متغير". Thus, the end user can detect this anomaly by suggesting a new translation that would be maintained in that lexicon to be reused in similar circumstances.

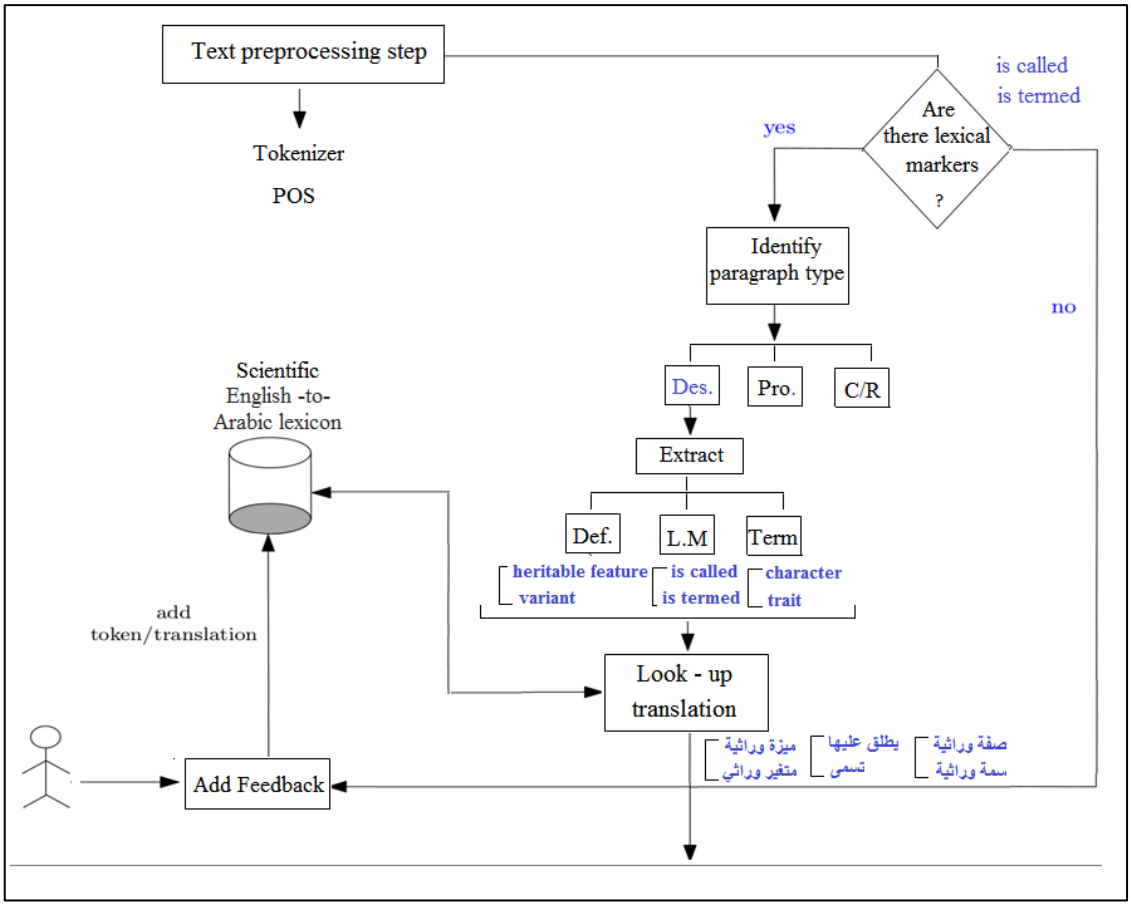


Figure (17): The process of identifying descriptive paragraphs based on the lexical markers.

4.3. Areas of Eff./Def. of Lexical Markers in Process Paragraphs

4.3.1. Cohesion Achieved through Word-Choice

Process paragraphs or "how to" paragraphs are known for their directness and clarity. The relationship between the words and sentences that make up the paragraph one unit are obvious and transparent. This clarity is the essence of scientific writing; Katz maintains that "In science, descriptions must be precise, recipes must be complete, data must be exact, logic must be transparent, and conclusions must be cleanly stated" (2009, p.3). Hence, when it comes to process paragraphs, where the main aim is to

analyze a certain process into a series of stages, getting rid of all slippery words or elements is a must.

Thus, process paragraphs tend to name the process at first, and then they deploy a number of cohesive devices to explain the process in question including: First, transitions that show chronological order such as: first, second, last, etc. Such connectors or sequence transitions help the readers see how things are ordered, connected and related to each other. In other words, the order of events in process paragraphs is important for changing it might change the intended meaning. Second, sentences mostly contain present tense constructions, imperatives or modals to describe each step. Third, content words are sometimes used to mark the process sequence. This could be done through deploying the appropriate word class: noun, verb, acronym, etc to mark the process in question along with its inputs and outputs as shown in paragraph **3a**:

Paragraph 3a:

Translation is the synthesis of a polypeptide, which occurs under the direction of mRNA. In 1956, Francis Crick proposed what he called the central dogma of molecular biology. The central dogma, simply stated, is that DNA codes for the production of RNA,RNA codes for the production of protein, and protein does not code for the production of protein, RNA, or DNA. In Crick's words, "once 'information' has passed into protein it cannot get out again."

In other terms, cohesion in paragraph **3a** is achieved at two main levels: lexical cohesion and conjunctions. The former is done through the use of technical terms and acronyms related to the process of "Translation of mRNA to protein" such as: "central dogma", "protein" and "RNA". Thus, these terms are repeated inside the paragraph for the process being described is done step by step. As a result, paragraph **3a** resorts to structure repetition to avoid any ambiguity which may result in assigning roles to the entities/participants involved in the process being described. In other words, paragraph **3a** contains three main entities that play a major role in the process of translation of mRNA to protein: DNA, RNA and protein.

Thus, they are all used in subject positions in the third line of paragraph **3a** to indicate and emphasize their own roles then complete the roles/outcomes of one another to convey all the steps of the process in question correctly. In addition, those entities/terms are used within the same structure: simple sentence. This structure is repeated three times for it is short and it allows the readers to recover the previous and subsequent steps easily. Conjunctions, on the other hand, are manifested through the use of the verb "codes" that is employed to highlight the subjects, the scientific action and the sequence of the process in question, besides the use of the word "once" in the last sentence to give a summary for the whole paragraph.

However, in regards to GT, it has been observed that GT fails to paraphrase the idea and make it clear since it does not have the ability to

express the meaning using the appropriate words that make the text cohesive. In other words, the researcher observes that GT does not have enough level of recognition in relation to the technical naming of processes and sequence signals when they are expressed through content words such as: verbs, in lieu of temporal signals such as: first, second, next, etc which are employed in the translation of scientific paragraphs to make them cohesive. This means that GT fails to benefit from the parallel structure used to mark the sequence of the given process and it fails to recognize the appropriate word type used in process paragraphs. This affects cohesion since the focus of process paragraphs is on giving instructions or describing steps through using verbs as shown in paragraph **3b**:

Paragraph 3b:

الترجمة هي تركيب بولي ببتيد ، الذي يحدث تحت إشراف mRNA. في عام 1956 ، اقترح فرانسيس كريك ما أسماه العقيدة المركزية للبيولوجيا الجزيئية. العقيدة المركزية ، ببساطة ، هي أن رموز الحمض النووي لإنتاج RNA ، ورموز RNA لإنتاج البروتين ، والبروتين لا يرمز لإنتاج البروتين ، RNA ، أو DNA. في كلمات كريك ، "لمرة واحدة" قد مرت المعلومات إلى بروتين لا يمكن الخروج مرة أخرى ".

Paragraph **3b** shows that GT fails to produce an acceptable paragraph at three main levels: the topic sentence, the process being described and the summary of the process in question. In other words, the first sentence which is the topic sentence in the present paragraph does not refer to a biological process. This proves that GT fails to come up with the right synonym of the word "Translation" since "Translation" is a general

term that could be used to refer to activities other than the biological term or process that denotes "The translation of mRNA to protein".

Thus, this is against the format of scientific texts since the topic sentence has to be precise and narrow enough for it gives the focal point/process that is going to be described and explained in the sub-sequent sentences in that paragraph. This echoes Katz's words who states: "a typical scientific paragraph begins by stating its point, so the lead sentence should tell us the focus of the paragraph." While "the remaining 2-3 sentences … expand on the focal point that was identified in the lead sentence" (2009.Retrieved from: https://books.google.com). Accordingly, this erroneous paraphrasing of the first sentence in the paragraph may weaken the quality of the scientific text since it exists in the topic sentence, an initial position as shown in Figure [**18]**:

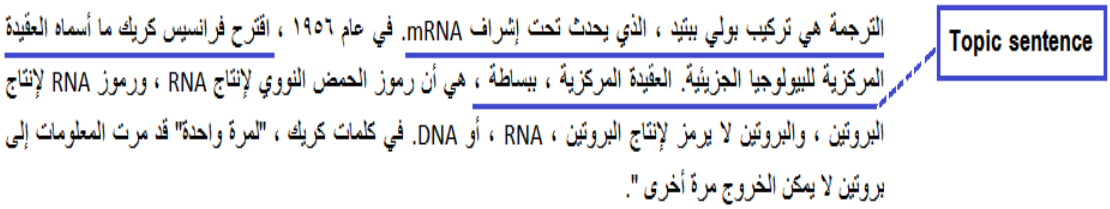


Figure (18): GT inexact translation for the process in question.

In addition, GT fails to use the appropriate words to paraphrase the idea of "central dogma". In a few words, it would be better if GT used the pronoun "التي" to achieve the lexical cohesion and add a smooth touch when moving from the term to its definition rather than repeating the term that makes the sentence redundant as GT did in the present paragraph. However, this results because GT resorts to literal translation as a substitute

for trying to use the appropriate linking words that enable it to paraphrase what the concept means. Thus, the appropriate linking words that could be used in this context are: " العقيدة المركزية تعني , تنص على أن , وباختصار فان العقيدة المركزية تنص على ".

Further, GT fails to echo the given scientist's words because it fails to treat such type of words used for providing a summary for the process in question. In other words, the English text states that: "by the time the information passed to protein which is the last step in the process of translation, it will be blocked inside and cannot get outside again". Thus, the process will stop. Nevertheless, GT fails to paraphrase this type of structures since it goes with the literal meaning of the word "one" which is a number and it neglects the meaning of the word "once" as a linking word for sequence which means "by the time" or "بمجرد" in Arabic. In other words, "one" in such type of paragraphs is different from "one" that indicates numbers. This failure to select the appropriate word category proves that one serious defect is word type recognition which subsequently results in failure to reproduce the process paragraph correctly.

4.3.2. Process-Sequence Verbs

In paragraph **3b**, GT fails to paraphrase the process the paragraph aims to explain since it treats the word "codes" as if it were a noun while it serves as a verb in the present text that means "توجه" or "ترمز". This happened because GT did not read and digest the paragraph from A to Z correctly. In other words, GT did not take into consideration the

complement of the sentence which consists of three steps which are: DNA codes for the production of RNA, RNA codes for the production of protein, and protein does not code for the production of protein, RNA, or DNA as shown in Figure [**19**] below:

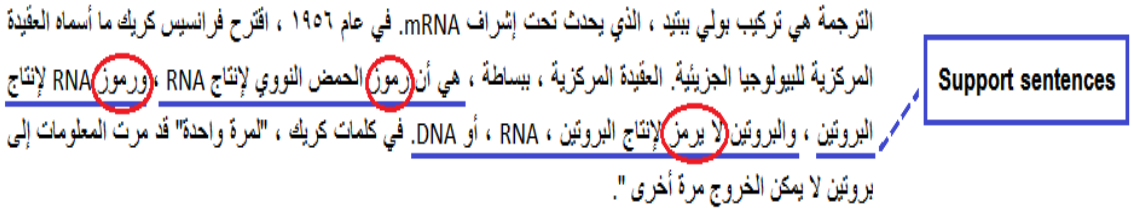


Figure (19): GT failure in identifying the verb as a sequence transition.

Thus, this wrong understanding of the word "code" and the length of the sentences which makes them a one complex support sentence led GT to take the paragraph away from the process that it aims to paraphrase. In other words, GT fails to recognize the fact that "a temporal relation may be expressed by means of a verb such as follow or precede, and a causal relation is inherent in the meaning of verbs such as cause and lead to" (Baker, 1992, p.191). Thus, in paragraph **3b** GT fails to identify such semantic relationship inherent in the verb "code" which shows process sequence even when no explicit signals of such relationship exist in the text. This wrong understanding of the verb "codes" results in two contradicting ideas which are:

In sum, the paragraph describes the sequence of the process of "Translation" so it states that the process of "turning mRNA to protein" goes in one direction and it cannot be reversed. However, this sequential move of the process in a linear way is not clear in GT output due to its failure to recognize the process and also its inability to distinguish verbs from nouns to highlight the process sequence as the arrows show below:

DNA codes for → **RNA** codes for → **protein - Stop- .**

To conclude, word type recognition can play a crucial role in enhancing the quality of GT production since process paragraphs are known for their extensive use of simple sentences in the present tense or imperatives, beside the use of sequence signals in the unmarked case such as: *first*, *second*, *finally*, etc. Moreover, pronouns are rarely used in such type of paragraphs since the focus is on the sequence of events rather than the participants. Thus, these linguistic norms associated with process paragraphs may be used to enhance GT performance through limiting its focus on the verbs that may indicate the sequential process in a given paragraph to maintain the sequence that allows the readers to trace the steps described in a given paragraph through identifying the function of the verb in each sentence/step. Thus, an acceptable translation for paragraph **3a** would be as shown in Figure [**20**]:

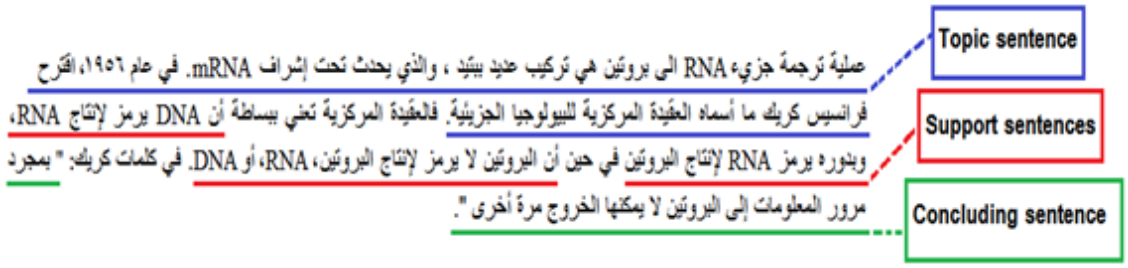


Figure (20): Suggested translation for paragraph 3a.

4.3.3. Process Signals within Non-past Constructions

Process paragraphs tend to employ a number of sequence signals which indicate the actions that take place later on in the process being described such as: *later*, *then*, *next*, etc. Thus, such signals require the presence of non-past constructions to convey the idea that the actions are not done in the past. In other terms, such non-past constructions indicate that the steps are fixed in any process since the non-past tense, in particular, the present tense is used to indicate the idea that the events/steps take place whenever the process in repeated.

Thence, paragraph **4a** describes the process of "cell division". Thus, it deploys three main cohesive devices to make the sequence of the process transparent: First, lexical cohesion which is achieved through naming the process to be described in the topic sentence which is "cell division", then employing a number of scientific terms such as: "chromosomes", "chromatids", "centromere" and "nuclei". Such terms keep the focus on the elements that contribute in the process in question. Secondly, the text employs a number of temporal conjunctions to mark the steps of the process of "cell division" including: "in preparation", "as" and "later" to

help the readers move smoothly from one step to another. Finally, reference to time which is manifested through the use of the present tense including verbs such as: "replicate", "condense", "connect", "shrinks", "separate" and "move" beside passive constructions such as: "joined" and "called" as shown in paragraph **4a**:

Paragraph 4a:

In preparation for cell division, chromosomes replicate, each one then consisting of two identical sister chromatids joined along their lengths by adhesive protein complexes called sister chromatid cohesion. As the chromosomes condense, the region where the chromatids connect shrinks to a narrow area, the centromere. Later in the cell division process, the two sister chromatids of each duplicated chromosomes separate and move into new nuclei.

Paragraph 4b:

في التحضير للانقسام الخلوي ، تتكاثر الكروموسومات ، كل واحد يتألف من اثنين من الكروماتيدات الشقيقة متطابقة انضمت على طول أطوالها عن طريق مجمعات بروتين لاصقة تسمى التماسك الكروماتيني الشقيق. مع تكاثف الكروموسومات، تتقلص المنطقة التي يتواصل فيها الكروماتيدات مع منطقة ضيقة، وهي المركزية. في وقت لاحق في عملية الانقسام الخلوي ، انفصل الشقيقان الشقيقان لكل صبغي من الكروموسومات وينتقلان إلى نويات جديدة.

Accordingly, comparing the ST with Google translation, the researcher observed that GT relatively succeeds in keeping the lexical chain of the paragraph since it uses correct scientific terms for the process in question in Arabic such as: "الكروماتيدات", "الكروموسومات", "الانقسام الخلوي"

"التماسك الكروماتيني" ,"الشقيقة". Moreover, GT resorts to literal translation when it comes to the signals that indicate the sequence such as: "as" and "later". This method works since GT renders them correctly into "ومع" , "في وقت لاحق" in Arabic.

However, with regard to reference manifested in the tense in paragraph **4a**, GT fails to benefit from the sequence signals present in the text to come up with the appropriate tense for the words "joined" and "duplicated". In other words, GT goes with the superficial form of the two words regardless of their underlying structure and the surrounding words that may help GT translate the verb "joined" as a passive construction not a past one and "duplicated" as an adjective not a past tense. In other words, GT fails to benefit from the preposition "by" to retrieve the underlying structure of the verb "joined" which is: "chromatids which are joined by" and "duplicated" as an adjective not a verb since it comes after a quantifier and before a noun: /quantifier+ (adjective)+noun/. In another words, GT is deceived by the "ed" form which in the present paragraph is used to form both passive structures and adjectives since it translates the words as: " انضمت" and "انفصل " in Arabic. However, such errors may be avoided if GT were able to benefits from the sequence word in the text which is "later" and the verbs "separate" and "move". Thus, an acceptable translation of paragraph **4a** is shown in Figure [**21**]:


في التحضير للانقسام الخلوي، تقوم الكروموسومات بنسخ نفسها بحيث تحتوي كل نسخة على اثنين من الكروماتيدات المتطابقة التي ترتبط مع بعضها بواسطة شريط لاصق مكون من مركبات البروتين يسمى الرابط الكروماتيني الشقيق. ومع تكاثف الكروموسومات ، تصغر الجهة التي تتصل فيها الكروماتيدات الى منطقة ضيقة تسمى المنطقة المركزية. وفي وقت لاحق من عملية الانقسام الخلوي، تنفصل كل نسختان شقيقتان من صبغ الكروموسومات وتنتقلان إلى نويات جديدة.

Figure (21): Suggested translation for paragraph 4a.

Therefore, the research recommends that GT be programmed to identify the signals frequently used in process paragraphs including the unmarked temporal conjunctions such as: *first*, *second*, *next*, *then*, etc. However, if there are no apparent sequence signals, then GT needs to identify the verbs in the text since a verb could be bi-functional: a verb, a "content word" that indicates what is being done and a "sequence word" to help the readers understand the process being described. Accordingly, the researcher suggests a procedure explained in Figure [**22**] to translate process paragraphs with high accuracy rates. First, the paragraph should undergo the same "preprocessing step" mentioned earlier: dividing the text into tokens along with their POS. Then, based on the tokens' types (words' types), GT needs to decide on whether there are lexical markers that help recognize the type of paragraph in question or not. If yes, then GT has to use those signals to identify the type of paragraph in question. Accordingly, GT will undergo an "extraction step" to get the unmarked lexical markers used in the present paragraph or extract those markers from the verbs in the paragraph e.g. for paragraph **3a**, the direct translation of the verb "codes" is "رموز". Thus, the end user can detect this error by suggesting a new translation "يرمز الى" , to be kept in the lexicon to be deployed in similar paragraphs.

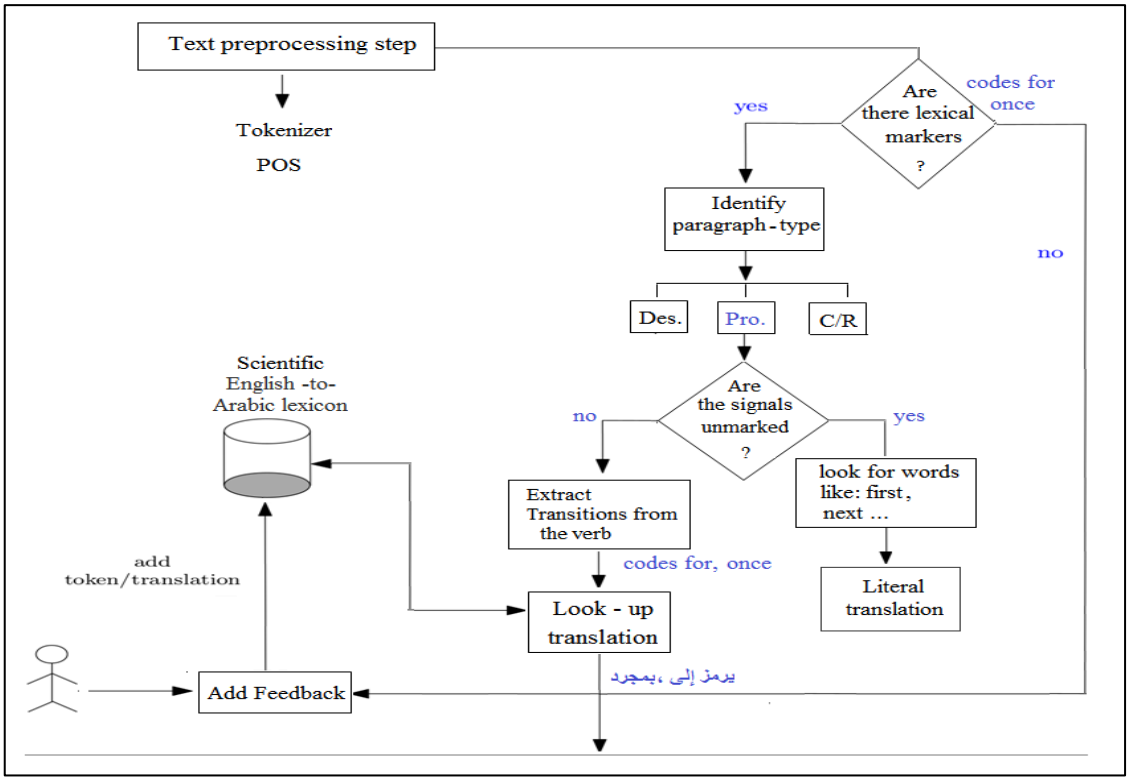


Figure (22): The process of identifying process paragraphs based on the lexical markers.

4.4. Areas of Eff./Def. of Lexical Markers in Causality Paragraphs

4.4.1. Word Collocations Deployed in Cause/Result Relations

Causality paragraphs aim at describing the relationship between two happenings since they aim to answer the question "why". In other words, causality paragraphs study the cause and effect of a particular phenomena. Thus, they make use of synonyms and antonyms since they aim to compare/contrast between the cause and its effect. Moreover, causality paragraphs deploy causal transitions such as: *therefore*, *because*, *hence*, *thus*, *if…then*, etc.

Respectively, paragraph **5a** employs a number of cohesive devices to explain facts along with their reasons about water behavior including: First

of all, reference that is achieved through the repetition of the non-human pronoun "it". Secondly, the text employs a number of collocated words such as: the opposites "warm/cool" and "expand/contradict". Such antonymic relations make it easy for the readers to understand the cause/effect of a particular situation since such lexical devices make the image clear by linking a word with, e.g. its extreme antonym, beside other domain specific terms related to water such as: "ice, freeze and crystalline lattice". Finally, the text employs clausal conjunctions such as: "because" that is used to give reasons about water behavior and "as" to show contrast in water status at different degrees of temperature and to mark the sequence of events that lead to a specific water status. All these contribute to the lexical cohesion of the paragraph as shown below.

Paragraph 5a:

Oceans and lakes don't freeze solid because ice floats.

Ice floats because it is less dense than liquid water. At temperature above 4C water behaves like other liquids, expanding as it warms and contracting as it cools. As the temperature falls to 0C, the water becomes locked into a crystalline lattice, each water molecule hydrogen-bonded to four partners.

Paragraph 5b:

لا تتجمد المحيطات والبحيرات الصلبة بسبب عوامات الجليد.

الثلج يطفو لأنه أقل كثافة من الماء السائل. في درجة حرارة أعلى من 4 Cالمياه تتصرف مثل السوائل الأخرى، وتوسع لأنها ترتفع درجة الحرارة والتقلص لأنه يبرد. عندما تنخفض درجة الحرارة إلى 0 درجة مئوية ، يصبح الماء مقفلًا في شبكة بلورية ، وكل جزيء ماء مرتبط بالهيدروجين لأربعة شركاء .

However, a comparative analysis of the ST with Google translation shows that GT does render most of the technical words correctly such as: "crystalline lattice" and "water molecule" yet when it comes to the collocated words "expand" and "contradict", GT fails to benefit from the causal words that make up the whole text and which help in understanding the cause/effect of a particular phenomena. Thus, GT fails to come up with the appropriate opposite of the word "contracting" in Arabic. In other words, when dealing with water, both "expanding and contracting" have particular equivalents in Arabic which are: "التمدد والتقلص". Yet, GT did not recognize this kind of opposite collocations which results in wrong word collocations in the Arabic text as shown in Figure **[23]**:

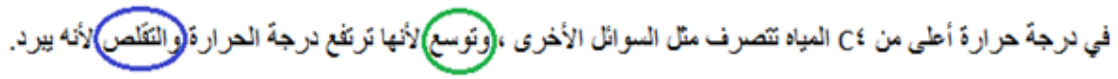


Figure (23): Errors in employing collocations.

Moreover, GT fails to treat the topic sentence and thus translates it erroneously since it fails to recognize the type of complement associated with the causal linking word "because". This in turn leads to GT failure in recognizing the verb in the initial sentence which refers to a very important action in the paragraph. In other words, the scientific action (verb) points out what is going to be done, why and how so it is important to identify it correctly since the readers will search in the topic sentence for the action to be described then look for its causes and effects in the sub-sequent sentences.

In summation, paragraph **5b** shows that GT produces a verb-less topic sentence yet the source text employs the linking word "because" to establish ties between the action and its reason in the topic sentence. In other words, it is well known that the word "because" requires a clause that consists of at least a subject and a verb as its complement. Accordingly, GT fails to understand the function of this conjunction, "because", which is used to provide reasons for the effect/result in question. GT treats its complement as if it were a noun phrase not a clause that is made up of: a subject and a verb. Thus, if GT were able to identify the type of constituents that must follow "because" in the present text, then it would translate the word "floats" as a verb not as a noun phrase since the word "because" must be followed by a clause as shown in Figure [**24**]:

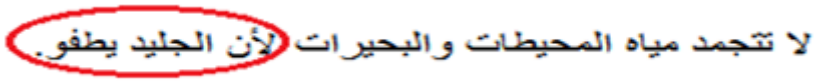


Figure (24): 'Because' along with its complement in Arabic.

In addition, GT fails to keep the chain of reference that exists in paragraph **5a.** In other words, this chain is broken in the Arabic text since GT fails to explain the idea correctly for the pronouns used do not help the readers to identify their referents. In other words, GT uses two types of pronouns which could refer to two types of nouns which are "لأنها" and "لأنه". Such pronouns require feminine and masculine referents, respectively. However, the Arabic text contains "المياه" , "السوائل" which are feminine and "الثلج" which is a masculine noun. Accordingly, this may be confusing to assign the referent/s of "لأنه" and "لأنها" for the source text

refers only to one noun, the same noun, which is "water" while GT uses two pronouns that refer to two nouns. This is against what is mentioned in the source text which only contains one salient noun in the second sentence which is the topic "water" as shown in Figure **[25]**:

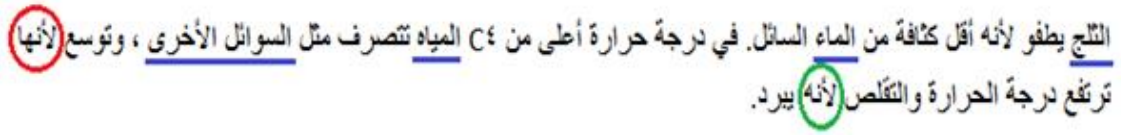


Figure (25): Errors in pronoun-referent resolution.

Furthermore, GT fails to treat constructions such as: "expanding as". This linking word "as" is employed to explain what happens to water at different degrees of temperature; it describes different situations for water at different degrees of temperature. Alternatively stated, there is a case of comparison between what happens to water's shape at high and low degrees of temperature. Thus, reasons are given why water expands/contracts. However, GT fails to connect these ideas in paragraph **5b** together using the appropriate linking word that makes the readers move smoothly between the ideas being described. Thus, it would be better to use "وعندما" to add a sense of continuity to the action so the reader will move step by step to understand and follow the effects/results explained for Arabic resorts to use "wa" heavily to coordinate the ideas in the text together as shown in Figure **[26]**:

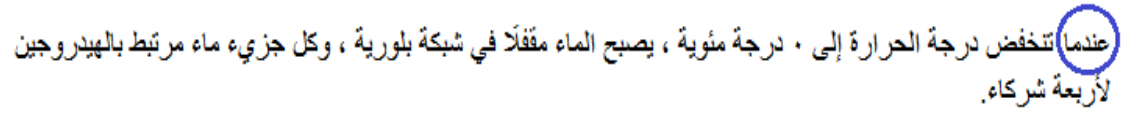


Figure (26): GT failure in deploying "wa".

To conclude, GT should be programmed to keep the chain of reference in causality paragraphs since such chains help the readers trace the events along with their causes and effects in the text in question. Moreover, GT should be programmed to come up with the correct word collocations e.g. opposites to make the text both cohesive and coherent. Thus, a correct translation for paragraph **5a** would be as shown in Figure **[27]**:

الثلج يطفو لأنه أقل كثافة من الماء السائل. في درجة حرارة أعلى من ٤ درجة مئوية المياه تتصرف مثل السوائل الأخرى ، فتتمدد عندما تسخن وتتقلص عندما تبرد ، وعندما تنخفض درجة الحرارة إلى ٠ درجة مئوية ، يصبح الماء مقفلًا في شبكة بلورية ، وكل جزيء ماء مرتبط بالهيدروجين لأربعة شركاء.

Figure (27): Suggested translation for paragraph 5a.

4.4.2. Causal Chains

Most causality paragraphs employ the simple present within if-conditionals to indicate the cause/effect of the action or situation being described. Thus, repeating if- structures within the same paragraph leads to a causal chain. In other words, in such kind of paragraphs, one event causes the next. This in turn creates a causal chain that gives the reader a clear idea about both the causes and their effects in the text in question. Accordingly, paragraph **6a** employs a number of cohesive devices to express information related to the relationship between chemical reactions and the effects of enzymes on them including: lexical cohesion, reference and conjunctions.

In few words, lexical cohesion is achieved through the use of technical words such as: "chemical reactions", "solution", "sterile water" and "hydrolysis". Such terms have peculiar meanings in scientific texts which make them distinct from their everyday use such as: "solution" which means a liquid chemical substance while it means a resolution to a difficult situation in everyday language. Moreover, reference is employed to refer to the time of the causes/effects. Thus, it is expressed through the use of if-structures that contain the non-past tense such as: "will", "may" and "do". Conjunctions, on the other hand, are expressed through the use of transitions that show contrast such as: "however", providing examples such as: "for example" and "if-then" structures that show cause/result relationships as shown below:

Paragraph 6a:

Enzymes:

Spontaneous chemical reactions may occur so slowly. For example, a solution of sucrose dissolved in sterile water will sit for years at room temperature with no appreciable hydrolysis. However, if we add a small amount of the enzyme sucrose to the solution, then all the sucrose may be hydrolyzed within seconds. How does the enzyme do this ?

Paragraph 6b:

الانزيمات:

قد تحدث التفاعلات الكيميائية التلقائية ببطء شديد. على سبيل المثال، سوف يوضع محلول السكروز المذاب في الماء المعقم لسنوات في درجة حرارة الغرفة دون أي تحلل مائي ملموس. ومع ذلك، إذا أضفنا كمية صغيرة من سكروز الإنزيم إلى المحلول، فقد يتم تحلل كل السكروز خلال ثوان. كيف يفعل الإنزيم هذا؟

However, comparing the ST with Google translation, the researcher observed that GT succeeds in translating the technical terms correctly. This in turn helps in maintaining the lexical chain which underlines the paragraph without breaking the lexical cohesion of the text. Thus, the Arabic chain contains terms appropriate to the idea being discussed such as:"تحلل","الماء المعقم", "المحلول", "التفاعلات الكيميائية" . All these words are employed correctly in Google translation.

Notwithstanding, touching on reference to time which is expressed through the use of if- conditionals in paragraph **6b**, GT mistranslates the idea the paragraph aims to express since it fails to allocate the future tense to its appropriate verb in the Arabic text. In other words, the ST causal chain states that if the solution is kept for years at room temperature in sterile water, it will not change. Yet, if an enzyme is added to the solution, then the situation will differ. This means that the future tense "will" needs to be attached to the second clause in the Arabic text for it indicates the result of something not the cause whose meaning is expressed in the present tense manifested in the verb "put" which comes first in the underlying structure of the clausal chain. This contrast between the intended meaning and Google translation is highlighted in Figure **[28**]:

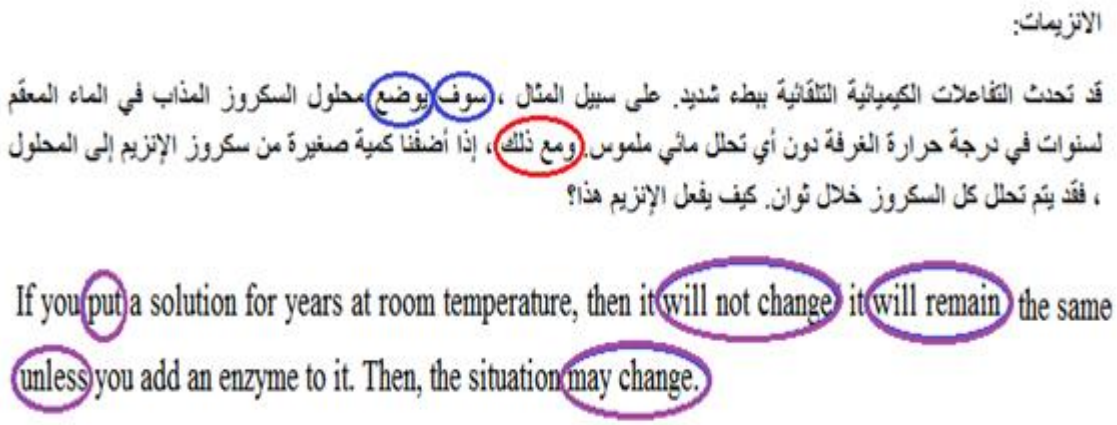


Figure (28): GT failure in reproducing if-structures.

Hence, this failure to assign reference in time to the appropriate verb in the present text leads to GT failure in coming up with the appropriate conjunction that is used to link the present causal chain. In other words, Google translate uses the word "ومع ذلك" while the text talks about two opposite situations, before and after adding the enzyme to the solution. Therefore, the appropriate tense and linking word to be used to express such relationship of leaving something without an enzyme or adding an enzyme to it, is the linking word "لكن" not "مع ذلك" that GT comes up with since there are two opposite situations which require the first linking word that expresses contrast not the second which expresses the situation of being not influenced by something, in English "despite" . Thus, an acceptable translation for the paragraph **6a** could be as shown in Figure **[29]**:

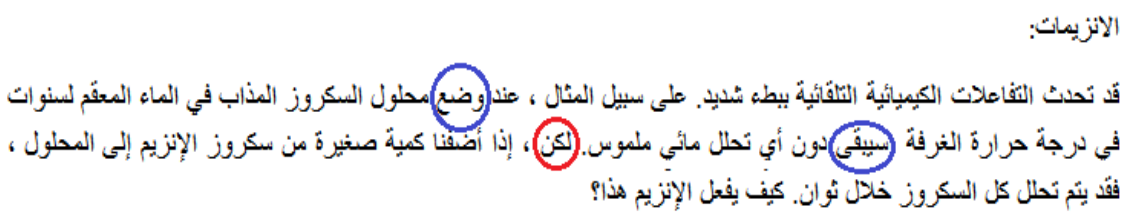


Figure (29): Suggested translation for paragraph 6a.

Therefore, the researcher recommends that GT be programmed to identify the signals frequently used in causality paragraphs including: if-conditionals, collocated words, causal and compare/contrast conjunctions along with their complements since causality paragraphs tend to explain cause/result relationships through using antonyms and collocations. Accordingly, the researcher suggests a procedure explained in Figure **[30]** to translate causality paragraphs with high accuracy levels. First, the paragraph should undergo the same "preprocessing step" mentioned earlier: tokens along with their POS. Then, based on the tokens' types (words' types), GT should identify the lexical markers that help recognize the type of paragraph in question. If GT finds any of the above mentioned signals, then it will undergo an "extraction step" to get the correct structure/form of the lexical markers frequently used in such type of paragraphs e.g. for paragraph **6a** above, the direct translation of the if-structure " if you put a solution in sterile water, then it will remain the same for years at room temperature" is "سوف يوضع محلول السكروز في الماء المعقم بدون أي تحلل مائي". Thus, the end user can detect this error by suggesting a new translation to be saved in the lexicon to be reused again in similar paragraphs such as: "اذا وضعت محلول في ماء معقم , سيبقى بدون تغيير لسنوات" .

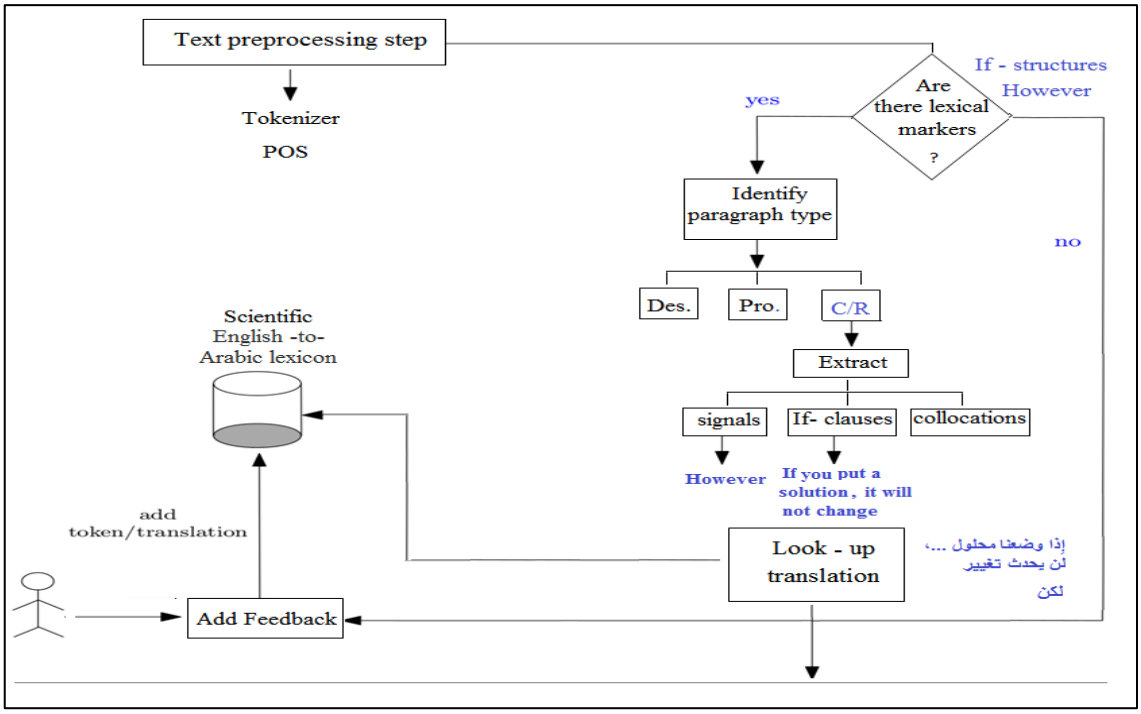


Figure (30): The process of identifying causality paragraphs based on the lexical markers.

Nevertheless, scientific texts may resort to employ a mixture of the three previously mentioned types of paragraphs (definition, process and causality) to make the meaning clear. In other words, scientific texts may shift from using mono-function paragraphs to deploy paragraphs with multiple functions. Such kind of paragraphs is called a mixed or bi-functional paragraph; e.g. a paragraph might start with introducing a definition then moves to describe a certain process or a it might start with describing a process then it shifts to express cause/result relations related to the process in question. This results in mixed paragraphs with a variety of linguistic norms which in turn makes them longer than mono-function paragraphs.

Consequently, when dealing with GT and mixed paragraphs, then GT may encounter a bundle of cohesive norms within a single paragraph

for each cohesive marker has a function to serve to make the paragraph communicative. Thus, GT has to link all the functions in question together in a coherent way as shown in paragraph 7a. The English text represents an example of a mixed paragraph since the paragraph starts with describing the process of respiration, then it shifts to express cause/result relations using if-conditionals as shown below:

Paragraph 7a:

Respiration is a cumulative function of three metabolic stages:

Glycolysis, the initial stage of all glucose metabolism, for aerobic cell respiration or for fermentation, means sugar splitting. Glucose, a six carbon sugar, is split into two pyruvate molecules, a three carbon sugar and produces some ATP via substrate phosphorylation. Glycolysis occurs in the cytosol of the cell. If oxygen is not available, or if the organism lacks enzymes needed for aerobic respiration, the pyruvate molecules will proceed with fermentations, or for some prokaryotes, anaerobic respiration. If oxygen is available and the organism has the enzymes to do aerobic respiration, the pyruvate molecules will be oxidized in the next stages of aerobic respiration.

In other words, the first part of the text is about describing the process of "Glycolysis". The text deploys a number of lexical markers to convey the process in question. First of all, it deploys a number of scientific terms related to the process in question such as "metabolism" and "fermentation". Secondly, the text employs process sequence markers such as: "initial" and "next" beside the verbs: "split", "produce" and "occurs". All these signals indicate that a certain process is being discussed. While

the second part of the text which represents cause/result relations deploys a number of if-conditional parallel structures to express the cause/result relations related to the process of “respiration”.

However, when examining GT performance in the Arabic text, it is observed that GT translates the verbs: "يحدث","ينتج" ,"ينقسم" correctly yet it fails to translate certain terms including: the name of the process “Glycolysis” since it means "تحلل الجلوكوز" and it has nothing to do with "التحلل الجلدي" that GT comes up with. In addition, GT fails to translate the sequence word “the initnal stage” since it translates it as "الأولية" while it means the beginning of the process in the present text, "في المرحلة الأولى". Moreover, GT fails to connect the two functions together (process and causality) since it does not attempt to integrate the two functions to produce a coherent mixed paragraph. This results in a non-communicative text as paragraph **7b** below shows:

Paragraph 7b:

التنفس هو وظيفة تراكمية من ثلاث مراحل الأيض:

التحلل الجلدي، المرحلة الأولية من كل عملية استقلاب الجلوكوز، للتنفس الخلوي الهوائي أو للتخمير، يعني فصل السكر. ينقسم الجلوكوز، وهو عبارة عن ستة سكر من الكربون، إلى جزأين من حمض البيروفيك، وهو عبارة عن ثلاثة سكر من الكربون وينتج بعض الجزيء ATP عن طريق فسفرة الركيزة. يحدث التحلل السكري في العصارة الخلوية للخلية. إذا كان الأكسجين غير متوفر، أو إذا كان الكائن الحي يفتقر إلى الإنزيمات اللازمة للتنفيس الهوائي، فإن جزيئات البيروفات ستستمر مع التخمر، أو بالنسبة لبعض بدائيات النواة، التنفس اللاهوائي. إذا كان الأكسجين متوفراً والكائنات الحية تحتوي على الإنزيمات للقيام بالتنفس الهوائي، فسوف تتأكسد جزيئات البيروفات في المراحل التالية من التنفس الهوائي.

To conclude, GT fails to make the message clear when it involves mixed paragraphs since they tend to be long. In other words, such paragraphs are loaded with a number of linguistic norms. Thus, it is important to use lexical markers appropriate to each part in the paragraph to translate correctly. Accordingly, it might be better to divide the paragraph into two halves to understand the two functions present in the text, then translate them taking into consideration the appropriate lexical markers associated with each half to reach high accuracy rates. Thus, an acceptable translation for paragraph 7a could be as shown in Figure [**31**] below:


التنفس هو وظيفة تراكمية من ثلاث مراحل أيضية:

تحلل الجلوكوز وهي المرحلة الأولى في جميع العمليات الأيضية المتعلقة بتحلل الجلوكوز "السكر". فهي بمثابة فصل السكر في عمليات التنفس الخلوي أو التخمير. حيث ينقسم الجلوكوز والذي يتكون من سكر سداسي الكربون الى جزئين من حمض البيروفيك والذي بدوره يتكون من سكر ثلاثي الكربون وينتج عن هذا التحلل بعض جزيئات الطاقة عن طريق الفسفرة. تحدث هذه العملية في عصارة الخلية. وإذا كان الأكسجين غير متوفر، أو كان الكائن الحي يفتقر الى الانزيمات اللازمة للتنفس الهوائي، فإن جزيئات البيروفات ستستمر بالتخمر، أما بالنسبة لبعض بدائيات النواة، ستلجأ الى التنفس اللاهوائي. واذا كان الأكسجين متوفر والكائنات الحية تمتلك الانزيمات للقيام بعملية التنفس الهوائي، فإن جزيئات البيروفات ستتأكسد في المرحلة التالية من التنفس الهوائي.

Figure (31): Suggested translation for paragraph 7a.

4.5. Conclusion

The data fed into GT and which includes three types of paragraphs, in particular, definition, process and causality paragraphs revealed that GT is still unable to produce optimal cohesive paragraphs where reference, conjunctions and lexical cohesion are all employed in both the right place and time according to the linguistic norms associated with each type of those paragraphs. In other words, the data indicates that GT lacks the ability of understanding, analyzing and reconstructing paragraphs correctly since GT is still unable to break the paragraph into its basic elements then reconstructs it in the target language in a way that makes it meaningful. Thus, a comparative analysis of cohesive devices between the ST and Google translation reveals areas of errors at paragraph level including: non-technical terms usage, misuse of lexical markers in introducing definitions, failure to highlight sequence and failure to express cause/result relationships.

In other words, the data revealed that GT fails to be consistent in its choice of the lexical terms that help the readers remain within the borders of the scientific domain since GT still depends on its own guesses as a means of translating paragraphs in lieu of its focus on scanning the lexical items in relation to the textual context that appears in its translation box. Thus, GT fails to maintain the lexical cohesion deployed mostly in descriptive and process paragraphs since they are basically about concepts and terms. In other words, GT uses words from different disciplines. Such

words break the lexical chain that shapes the domain in question and establishes its identity. This in turn leads to non-technical and low quality translations.

However, such defects may be solved by assisting GT in identifying the paragraph type in question through feeding GT with data about the linguistic norms frequently used in that paragraph type. Moreover, Google team have to feed their translation program with scientific dictionaries that enable the end user to get the correct scientific equivalences for the terms appropriate to the paragraph type in question. Thus, identifying a given paragraph as a scientific one is a step forward in enhancing GT performance since GT will exclude the non-technical meanings of the terms in question and this will help in reaching high accuracy rates when it comes to scientific terms translation.

Moreover, GT fails to connect the ideas between the sentences together in all types of paragraphs for it fails to identify the idea being described whether: **it is** a sequential process, a cause/result situation **or** a description of a particular object. Thus, GT fails to come up with the appropriate conjunctions that allow the readers to trace the information in the paragraph. Such defects may not allow the readers to understand and follow the participants or entities involved in the process in question to get the intended message behind the paragraph correctly. However, achieving cohesion at conjunctions level could be done through limiting the type of conjunctions normally used in each type of the selected paragraphs or what

is called "transitions/signals". Thus, enhancing GT performance in word-type recognition may play a crucial role in producing high quality translations since the same word may be used as a verb, a noun, a temporal signal, etc. Such procedures may pave the road for GT to translate paragraphs with high accuracy rates since paragraphs serve as basic building blocks which make the message coherent and informative.

Chapter Five

Conclusions and Recommendations

5.1. Conclusions

This thesis has mainly relied on descriptive methodologies and contrastive analysis between the ST and its equivalent translation/output as produced by GT at sentence and paragraph levels. Thus, Catford's "translation shifts", Halliday and Hassan's model of cohesive devices and paragraph type frequently used in scientific texts were used to measure the areas of eff./def. in GT performance. GT output was assessed at both sentence and paragraph levels. The thesis arrived at the following conclusions regarding GT performance in scientific biological texts translation from English to Arabic.

At sentence level, the data revealed that GT handles the unmarked sentence pattern in English /sub.+ v. +obj./ correctly in cases where both the sub. and the obj. are semantically salient in the sentence and cannot be used in place of one another. In addition, GT handles the modals, plural and present tense forms correctly. However, in some cases, word-for word translation leads GT to commit errors in grammar, in particular, in areas of: word order, active and passive inflections and affixation. These deficiencies at the grammatical level could be attributed to the diversity of grammatical rules between the two languages and how each language uses its linguistic resources to express relationships. Arabic expresses meaning

through word order and diacritics, while English mostly deploys word order. Thus, the collected data prove that GT is still unable to draw on the contextual information, word senses and real world knowledge about how things are connected or related to each other to compensate for the absence of diacritics in English. Accordingly, errors show up in the output.

Additionally, aspects of gender remain a source of errors in GT performance in case of pronoun translation. In other words, the pronoun should agree with its antecedent in gender, number, etc. However, both English and Arabic differ in their assignment of gender. Arabic, on the one hand, contains two types of nouns based on gender distinction: feminine and masculine. Thus, it expresses gender through inflections like "ة/ات/ون", while English contains three types of nouns: feminine, masculine and gender-neutral. Thus, it uses words like "s/he" and "it" to communicate this aspect. This variance creates a challenge in translating pronouns that refer to gender neutral nouns in cases where the noun is plural. In other words, GT fails to recognize that the same noun could serve as a gender neutral noun in the SL while it should be marked for gender in the TL. Thus, the absence of gender markers from plural nouns in English (English only uses /s/ to indicate the plural form) confuses GT since such nouns should be gender marked in Arabic. Accordingly, errors result in pronoun translation.

At the morphological level, GT fails to decide on the function of the inflections in question. For example, when the inflection /ed/ is used to form e.g.: a past tense, an adjective that ends with /ed/ or a passive

adjectival, GT treats it, in most cases, as an inflection that indicates a past form, neglecting the other two functions. This results in errors in GT output. Such errors could be attributed to the fact that GT does not process all the elements of the sentence correctly which may help GT identify the appropriate function of the /ed/.

At paragraph level, scientific paragraphs manifest universal features related to both technicality and structural clarity since they aim to communicate information that is well-informed, tested and validated. Also, scientific paragraphs are characterized by the use of technical terms and repeated syntactic structures. They display different functions; they could be descriptive, procedural or causal. Each paragraph type deploys a number of cohesive devises to make the meaning transparent and clear. Thus, GT should recognize the cohesive markers peculiar to each paragraph type to produce coherent paragraphs.

Areas of efficiency in relation to cohesive markers at paragraph level are manifest at reference level. In other words, GT performs well when it comes to reference to time and participants in most cases. This could be attributed to the nature of scientific texts which are characterized by repetition, passive constructions and present tense forms used to emphasize the ideas being discussed. Such devices decrease the number of pronouns used in the paragraph in question. This makes the situation easy for GT since it has to deal with only a few numbers of pronouns and tenses in the paragraph in question. Accordingly, at the three types of paragraphs -

description, process and causality- GT shows good performance since only a few errors were observed at reference level and they are attributed to gender aspects discussed earlier.

However, there are a number of deficiencies at lexical markers level. In other words, GT commits errors in lexical chains recognition at paragraph level which prevents it from coming up with translations of high accuracy rates. The data show that GT is still unable to draw on the overall content of the paragraph to come up with correct translations for the scientific terms in question. In other words, GT depends heavily on literal translation which makes it treat every word in the text as if it were a unit on its own. Thus, GT does not relate it to the rest of the words to detect the lexical chain of the text in question. Such literal translations result in non-technical terms which in turn break the lexical chain that shapes the text in question. Thus, such errors in lexical chain recognition are manifested clearly in description and process paragraphs since such types of paragraphs depend heavily on scientific terms to introduce definitions, concepts and processes while causality paragraphs do not depend heavily on lexical chains for the focus is on the scientific action and its causes/results rather than scientific concepts/terms. Thus, no errors in the lexical chain are present in this latter paragraph type except for collocations.

GT fails to appropriately deploy the linking words that connect the ideas in the paragraph correctly. Thus, the examples reveal that deploying

the wrong connectors leads the readers to lose trace of events, descriptions, processes, participants, time, cause, result, etc. However, GT performance in the three types of paragraphs is not identical. More specifically, GT failure in coming up with the appropriate connectors is salient at its most in process paragraphs since lexical markers that help the readers to follow the sequence of the process in question matter a lot in this paragraph type. Yet, having marked process signals manifested in the verb itself in lieu of unmarked process signals such as: first, second, later, etc led to GT errors. In other words, GT fails to deploy the verb as a process signal in cases where there are no explicit/unmarked lexical markers in the text.

Causality paragraphs occupy the second position in misuse of lexical markers since GT fails to make the cause/result relationship clear. It fails to identify and analyze the meaning of the cause and its result at first then select the appropriate lexical markers to connect them together based on the content of the message. In the last place, come descriptive paragraphs where GT fails to introduce the definition correctly. It comes at the end since the main cohesive device employed in such type of paragraphs is the lexical chain compared with conjunctions.

Concerning mixed paragraphs, GT is deficient in analyzing the functions the paragraph aims to communicate since it goes with literal translation which is clear from its output. In other words, mixed paragraphs tend to be longer than mono-function paragraphs since they employ a group of lexical markers that play a crucial rule in moving smoothly form one

purpose to another. However, the output reveals that GT fails to derive the appropriate lexical markers frequently used in mono-function paragraphs then integrate them properly in the translation. This results in a mixed paragraph that appears as a draft containing a group of words put together with no meaningful message.

Having identified these problems in GT performance, the researcher proposed a number of procedures for potential improvements. Such procedures are capable of enhancing GT accuracy rates in scientific biological texts translation. The proposed procedures attempted to draw on the following concepts form natural language processing (NLP):

1. Providing GT with "preprocessing mechanisms" for the input sentence including two steps: First, the analysis of the input to its basic building blocks in a filter called "a tokenizer" then the classification of the elements that make up the input along with their grammatical categories in another filter called a part of speech tagger "POS", is deemed to be necessary. Such mechanisms help GT decide on the appropriate extraction steps to take to get the equivalent translations for the sentence under study.

2. Recognition of the cohesive devices at paragraph level could be improved if GT team enable their translation program to identify the type of the paragraph in question based on the lexical markers frequently employed in each type of the three paragraphs. Moreover, GT should be programmed to encounter a mixed paragraph with

more than one function. In other words, GT has to treat the types of paragraphs as one package since they may be integrated within one mixed paragraph to support the idea and make it salient. Once GT identifies the paragraph type, it will undergo an extraction step to come up with the appropriate lexical markers and use them in a way that suits that paragraph type. Yet, in cases where GT does not recognize the appropriate words/terms to be used in that paragraph type then users can assist GT through suggesting their own translations to correct the errors in the translation. Such procedure may work in enhancing GT performance to reach high accuracy rates at paragraph level.

3. GT needs to be equipped with a dictionary for domain specific terms. Such domain specific dictionary should be linked with the translation action. It may be called "a specialized lexicon" or "a scientific lexicon". Such lexicon could be fed through allowing the end user to suggest his/her translation for the terms in question to be listed in the options list to be reused again with similar concepts/structures. This, in turn, allows GT to draw on frequency rates for terms to decide on the appropriate meaning of the term in question. In other words, terms should be listed in a descending order based on their frequency of occurrence in the field in question. Such word banks may help GT reach close translations of scientific terms.

4. Having an interactive machine by allowing users to interact with GT is a step forward since this will allow GT to build a bridge with the end user. In other words, the end user can interact with the machine in question; make important corrections and changes during the translation action to enhance the quality of the output. Accordingly, saving those changes in a "specialized lexicon" will enable the other users to benefit from such interactive system whenever the system encounters the same terms or constructions for the machine will select the most frequent options for the term in question suggested by the users. This, in turn, will save the users time and effort needed for post editing GT output.

To conclude, there is an extensive need for enduring efforts to improve the quality of GT output to reach the desired goal behind using this software which is close translation. In other words, systems that are limited to special domains, text types or restricted to certain purposes are deemed to be necessary in lieu of general translations provided by GT. Moreover, using GT at levels larger than the single word or sentence is necessary since GT is a machine that is designed to surpass the dictionary which is restricted to translating single words/phrases. However, as it stands at the moment of preparing this research, GT still shows serious deficiency in translating passive constructions, gender neutral nouns, affixes and inability to deploy the cohesive devices peculiar to each type of the three paragraphs (description, process and causality) in the right way.

However, the translation situation can be improved by using the interactive approach between GT and the end user, beside specialized lexicons for scientific translation.

5.2. Recommendations

Focusing on the advantages of analyzing the features of the input, post-editing processes and suggested translations on the quality of GT output is a necessity. Such procedures have become necessary to face machine translation challenges and pursue the idea of "building systems for specific domains". Accordingly, further comparative studies between English and Arabic to get the cohesive norms adopted by each language in the selected kinds of paragraphs is deemed necessary to enable the machine to produce good translations in an attempt to empower it to reflect its name to be fully automatic machine translation and save human translators’ ink.

References

Abbasi, M. & Karimnia, A. (2011). ***An Analysis of Grammatical Errors among Iranian Translation Students: Insights from Interlanguage Theory***. **European Journal of Social Sciences**. 25(4), 525-536. Retrieved March 16th 2019 from <https://urlzs.com/wpw5J>

Abdulhaq, S. Y. I. (2016). **Machine Translation Limits of Accuracy and Fidelity** (Master thesis). Palestine: An-Najah National University. Retrieved 20th March 2019 from <https://scholar.najah.edu>

Aiken, M. & Balan, S. (2011). **An Analysis of Google Translate Accuracy**. Retrieved Oct. 9th 2017 from <https://urlzs.com/aPm1L>

Al Shehab, M. (2013). ***The Translatability of English Legal Sentences into Arabic by Using Google Translation***. **International Journal of English Language and Linguistics Research**. 1(3), 18-31. Retrieved March 20th 2019 from <http://www.eajournals.org>

AL-Asali, S. I. S. (2000). **Computer Assisted Translation of English Scientific Texts into Arabic: Designing a Computer Prototype for Cohesive Ties** (Unpublished PhD Dissertation). Baghdad: Al-Mustansiriyya University.

Al-Hassnawi, A. (2013). **Aspects of Scientific Translation: English into Arabic Translation as a Case Study**. Retrieved March 20th 2019 from <http://www.translationdirectory.com/article10.htm>

Ali, M. & Ismail, Z. (2006**)**. **Comprehension Level Of Non-Technical Terms In Science: Are We Ready For Science In English**. Retrieved November 28th 2017 from <https://urlzs.com/FnT8d>

Al-Khawalda, M. & Al-Oliemat, A. (2014). ***Machine Translation: Deficiency in Translating English Sentences with Different Temporal References into Arabic.*** **Journal of Education and Practice**. *5*(5), 39-46. Retrieved March 20th 2019 from <https://urlzs.com/XKfcN>

Al-Samawi, A. M. (2014). ***Language Errors in Machine Translation of Encyclopedic Texts from English into Arabic: The Case of Google Translate***. **AWEJ. Special Issue on Translation** No.3, 182-211. Retrieved March 16th 2019 from <https://urlzs.com/74KMz>

Baker, M. (1992). **In Other Words: A Coursebook on Translation**. London and New York: Routledge.

Beaugrande, R. de & Dressler, W. (1981). **Introduction to Text Linguistics**. London and New York: Longman.

Brown, G. & Yale, G. (1983). **Discourse Analysis**. Cambridge: Cambridge University Press.

Brown, H. D. (2004). **Principles of Language Learning and Language Teaching**. (4thed). New York: Longman.

Craciunessu, O., Gerding-Salas, C. & Stringer-O'Keeffe, S. (2004). **Machine Translation and Computer-Assisted Translation: A New Way of Translating?** Retrieved March 20th 2019 from <https://urlzs.com/GDTS5>

Dubey, P. (2013). **Study and Development of Machine Translation System from Hindi Language to Dogri Language: An Important Tool to Bridge the Digital Divide** (Doctoral dissertation). India: University of Jammu. Retrieved March 20th 2019 from <https://urlzs.com/xBMU2>

ElShiekh, A. A. (2012). **Google Translate Service: Transfer of Meaning, Distortion or Simply a New Creation? An Investigation into the Translation Process & Problems at Google**. Retrieved March 20th 2019 from <http://dx.doi.org/10.5539/ells.v2n1p56>

Enkvist, N. E. (1978). ***Contrastive Text Linguistics and Translation***. In: Grähs, L., Korlen, G. and Malmberg, B., (Eds.). **Theory and Practice of Translation**, Peter Lang, Berne, 169-188.

Errens, J. (2019). **The Past, Present and Future of Machine Translation**. Retrieved March 20th 2019 from <https://urlzs.com/95Af3>

Fromkin, V. & Rodman, R. (1995). **An Introduction to Language** (6th edition). London: Harcourt Brace Jovanovich.

Grabe, W. (1985). ***Written Discourse Analysis***. In Kaplan R.B. (Ed.). **Annual Review of Applied Linguistics** *5*, 101-123. New York: Cambridge University.

Halliday, M. A. K. & Hasan, R. (1976). **Cohesion in English**. London and New York: Longman.

Halliday, M. A. K. (1970). Language Structure and Language Function, in J. Lyons (Ed.) **New Horizons in Linguistics**. Harmondsworth: Penguin.

Hannouna, Y. H. A. (2004). **Evaluation of Machine Translation Systems: The Translation Quality of Three Arabic Systems (Doctoral dissertation).** Baghdad: Al-Mustansiriyyah University. Retrieved March 20th 2019 from <https://urlzs.com/ZemLJ>

Hatim, B. & Munday, J. (2004). **Translation: An Advanced Resource Book**. London: Routledge.

Hatim, B. (2001). **Teaching and Researching Translation**. England: Pearson Education Limited.

Ilyas, A. (1989). **Theories of Translation: Theoretical Issues and Practical Implications.** Mosul: University of Mosul.

Karami, O. (2014, January). **The brief view on Google Translate machine**. Paper presented at the meeting of the 2014 Seminar in Artificial Intelligence on Natural Language, Germany.

Katz, M. J. (2009). **From Research to Manuscript: A Guide to Scientific Writing.** Retrieved March 19th 2019 from <http://books.google.com/books>

Nord, C. (1991). **Text Analysis in Translation**. Amsterdam: Rodopi.

Palincsar, A. S. (2013). **What Do Teachers Need to Know about the Language of Science Text?** Retrieved March 20th 2019 from <https://urlzs.com/ZBv4f>

Parikh, N. (2012). **Can Machine Translation for Cross- Culture be a Good Choice?** Retrieved March 20th 2019 from <https://urlzs.com/STgBk>

Pestov, I. (2018). **A History of Machine Translation from the Cold War to Deep Learning.** Retrieved March 20th 2019 from <https://urlzs.com/enQrN>

Saraireh, M. A. (2014). ***Common Practice Errors Related to Syntactic Structures in English -into- Arabic Translation***. **AWEJ**, 5(2), 187-205. Retrieved March 20th 2019 from <http://www.awej.org>

Somers, H. L. & Hutchins, W. J. (1992). Basic Strategies. **An Introduction to Machine Translation.** Retrieved March 20th 2019 from <https://urlzs.com/KpTMD>

Soualmia, M. (2009). **Third-year Students' Difficulties in Translating Computing Terms from English into Arabic** (Master thesis). Constantine: Mentouri University. Retrieved March 20th 2019 from <https://urlzs.com/AnArx>

Spethman, M. (2012). **Insight into Google Translate and Machine Translation.** Retrieved June 19th 2018 from <https://urlzs. com/xtgSo>

Swales, J. (1971). **Writing Scientific English: A Textbook of English as a Foreign Language for Students of Physical and Engineering Sciences**. Nelson: London.

Ulitkin, I. (2011). **Computer Assisted Translation Tools: A Brief Review**. Retrieved Oct. 3rd 2017 from <https://urlzs.com/3US1f >

Vilar, D., Xu, J., D'Haro, L. F. & Ney, H. (2006). **Error Analysis of Statistical Machine Translation Output**. Retrieved March 17th 2019 from <https://urlzs.com/XtQ9w>

Wikipedia. (2018, July 21). **Comparison of Machine Translation Applications**. In Wikipedia, The Free Encyclopedia. Retrieved July 30th 2018 from <https://urlzs.com/5utRH>

Winter, W. (1961,1964). Impossibilities of Translation, in W. Arrowsmith and R. Shattuck (Eds.). **The Craft and Context of Translation**. New York: Anchor.

Wisniewski, G., Kubler, N. & Yvon, F. (2014). **A Corpus of Machine Translation Errors Extracted from Translation Students Exercises**. Retrieved March 20th 2019 from <https://urlzs.com/CAQbk>

Wong, L. (2006). ***Syntax and Translatability***. **Babel**. 52(2) , 124-132

Yorkey, R. (1974). **Practical EFL Techniques for Teaching Arabic-Speaking Students**. Retrieved March 30th 2019 from <https://urlzs.com/3vqPt>

Zinaser, W. (1976). **On Writing Well: An Informal Guide To Writing Nonfiction**. New York: Harper & Row.

Appendix (1)

The Selected Texts and Their Translation as Produced by GT

Text 1:

Elements and compounds:

An electrically neutral atom has equal numbers of electrons and protons; the number of protons determines the atomic number. The atomic mass is roughly equal to the sum of protons plus neutrons. The neutron and proton are almost identical in mass. Thus, for atoms and subatomic particles, we use a unit of measurement called the Dalton. Neutrons and protons have masses close to 1 Dalton.

Isotopes: Isotopes of an element differ in their numbers of neutrons. Some isotopes are radioactive, emitting radiation as they decay.

Electron properties:

Electrons are always in motion, found in orbitals located at fixed distances outside of the nucleus called electron shells, which correspond to different energy levels. Each electron orbital holds a maximum of 2 electrons. Each energy level or electron shell has a fixed number of orbitals. For example, the first electron shell, closest to the nucleus, has one orbital. The chemical behavior of an atom depends mostly on the number of electrons in its outermost shell, the valence shell. Electrons in the valence shell are known as valence electrons. Lithium has one valence electron; neon has eight. Elements with a full valence shell are chemically inert.

العناصر والمركبات:

تحتوي الذرة المحايدة كهربائيًا على أعداد متساوية من الإلكترونات والبروتونات. يحدد عدد البروتونات العدد الذري. تكون الكتلة الذرية مساوية تقريبا لمجموع البروتونات زائد النيوترونات. النيوترون والبروتون متطابقان تقريبًا في الكتلة. وهكذا، بالنسبة للذرات والجسيمات دون الذرية، نستخدم وحدة قياس تسمى دالتون. تحتوي النيوترونات والبروتونات على كتل قريبة من 1 دالتون.

النظائر: نظائر العنصر تختلف في عدد النيوترونات. بعض النظائر المشعة، التي تنبعث منها الإشعاع لأنها تتحلل.

خصائص الكترون:

تكون الإلكترونات دائماً في حالة حركة ، وتوجد في المدارات الواقعة على مسافات ثابتة خارج النواة التي تدعى الأصداف الإلكترونية، والتي تتوافق مع مستويات الطاقة المختلفة. ويحمل كل مدار الكتروني كحد أقصى

إلكترونين. يحتوي كل مستوى طاقة أو غلاف إلكترون على عدد ثابت من المدارات. على سبيل المثال ، تحتوي أول قذيفة إلكترونية، الأقرب إلى النواة، على مدار واحد. يعتمد السلوك الكيميائي للذرة في الغالب على عدد الإلكترونات في غلافها الخارجي، قشرة التكافؤ. تعرف الإلكترونات في غلاف التكافؤ بإلكترونات التكافؤ. يحتوي الليثيوم على إلكترون تكافؤ واحد ؛ النيون لديه ثمانية. العناصر ذات قشرة التكافؤ الكاملة خاملة كيميائياً.

Text 2:

Chemical bonds:

Covalent bonds are strong bonds formed when two atoms share one or more pairs of electrons.

Nonpolar covalent bonds are formed when the electronegativities of the two atoms are approximately equal.

When atoms with strong electronegativity (such as oxygen) bond to atoms with weaker electronegativity (such as hydrogen), a polar covalent bond is formed.

Ionic bonds are electrical attractions between oppositely charged ions. Ionic bonds are strong in solids, but weaker when the ions are separated from one another in solution.

Hydrogen bonds are weak electrical attractions. Hydrogen bonds are abundant in water.

Van der waals interactions occur between transiently positive and negative regions of molecules.

Chemical reactions

In chemical reactions, chemical bonds are broken and reformed, leading to new arrangements of atoms. The starting molecules in the process are called reactants, and the final molecules are called products. In a chemical reaction, all of the atoms in the reactants must be present in the products. The reactions must be balanced.

السندات الكيميائية:

الروابط التساهمية هي روابط قوية تتشكل عندما تشترك ذرتان في واحد أو أكثر من أزواج الإلكترونات.

تتشكل الروابط التساهمية غير القطبية عندما تكون الكهربية الكهربية للذيتين متساويتين تقريباً.

عندما تكون الذرات ذات القدرة الكهربية القوية (مثل الأكسجين) مرتبطة بالذرات ذات الكهروسجة الأضعف (مثل الهيدروجين)، يتم تكوين رابطة تساهمية قطبية.

السندات الأيونية هي عوامل جذب كهربائية بين أيونات مشحونة الشحنة. الروابط الأيونية قوية في المواد الصلبة، ولكنها أضعف عندما يتم فصل الأيونات عن بعضها البعض في الحل.

الروابط الهيدروجينية هي نقاط جذب كهربائية ضعيفة. الروابط الهيدروجينية متوفرة في الماء.

تحدث تفاعلات فان دير فالس بين المناطق الإيجابية والسلبية من الجزيئات.

تفاعلات كيميائية

في التفاعلات الكيميائية ، يتم تكسير الروابط الكيميائية وإصلاحها، مما يؤدي إلى ترتيبات جديدة للذرات. تُسمى جزيئات البدء في العملية المتفاعلات، وتسمى الجزيئات النهائية بالمنتجات. في التفاعل الكيميائي، يجب أن تكون جميع الذرات الموجودة في المواد المتفاعلة موجودة في المنتجات. يجب أن تكون متوازنة ردود الفعل.

Text 3:

Oceans and lakes don't freeze solid because ice floats.

Ice floats because it is less dense than liquid water. At temperature above 4C water behaves like other liquids, expanding as it warms and contracting as it cools. As the temperature falls to 0C, the water becomes locked into a crystalline lattice, each water molecule hydrogen-bonded to four partners.

The hydrogen bond keep the molecules far enough apart to make ice about 10% less dense than liquid water at 4 C. Therefore, ice floats on the cool water below and the surface layer of ice insulates liquid water below, preventing it from freezing and allowing life to exist under the frozen surface.

Water is the solvent of life.

Water is an excellent solvent for many substances because of its polar nature. Polar substances and ions dissolve in water because opposite charges are attracted to the appropriate ends of water. Strictly hydrophobic molecules, including most lipids, do not mix well with water. Some molecules have both hydrophobic and hydrophilic ends. Such molecules are said to be amphipathic. Amphipathic molecules make good emulsifiers because they can attract both hydrophobic substances and hydrophilic substances to them. Substances dissolved in a solvent are called solutes.

لا تتجمد المحيطات والبحيرات الصلبة بسبب عوامات الجليد.

الثلج يطفو لأنه أقل كثافة من الماء السائل. في درجة حرارة أعلى من C4 المياه تتصرف مثل السوائل الأخرى، وتوسع لأنها ترتفع درجة الحرارة والتقلص لأنه يبرد. عندما تنخفض درجة الحرارة إلى 0 درجة مئوية ، يصبح الماء مقفلًا في شبكة بلورية، وكل جزيء ماء مرتبط بالهيدروجين لأربعة شركاء.

يحافظ الرابطة الهيدروجينية الجزيئات على مسافة بعيدة بما يكفي لجعل الثلج أقل كثافة بنسبة 10٪ من الماء السائل عند 4 ج. لذلك، يطفو الجليد على الماء البارد إلى الأسفل، وتعزل الطبقة السطحية من الجليد الماء السائل أدناه ، مما يمنعه من التجمد والسماح الحياة في الوجود تحت السطح المتجمد.

الماء هو مذيب الحياة.

الماء هو مذيب ممتاز لكثير من المواد بسبب طبيعته القطبية. تذوب المواد والأيونات القطبية في الماء لأن الرسوم المعاكسة تنجذب إلى نهايات المياه المناسبة. لا تختلط الجزيئات الكارهة للماء بشدة، بما في ذلك معظم الدهون، بشكل جيد مع الماء. بعض الجزيئات لها نهايات مسعرة للماء. ويقال إن هذه الجزيئات تكون amphipathic. جزيئات amphipathic يجعل المستحلبات جيدة لأنها يمكن أن تجتذب كل من المواد مسعور والمواد المحبة للماء لهم. تسمى المواد المذابة في مذيب.

Text 4:

Cells harvest energy via three pathways. Fermentation harvests chemical energy without using either oxygen or any electron transport chain. Aerobic respiration consumes oxygen as a reactant to complete the breakdown of a variety of organic molecules (aerobic is from the Greek aer, air, and bios, life). Anaerobic respiration is used by some prokaryotes whose use substances other than oxygen as terminal electron acceptor in a similar process to that of aerobic respiration without using any oxygen at all; (the prefix an- means without).

Redox Reactions: Oxidation and Reduction.

A reaction in which one substance transfers one or more electrons to another substance is called an oxidation-reduction reaction, or redox reaction. In a redox reaction, the loss of electrons from one substance is called oxidation, and the addition of electrons to another substance is known as reduction.

التنفس الهوائية:

تحصد الخلايا الطاقة عبر ثلاثة مسارات. التخمير يحصد الطاقة الكيميائية دون استخدام الأكسجين أو أي سلسلة نقل إلكترون. يستهلك التنفس الهوائي الأكسجين كمفاعل متفاعل لإكمال تحليل مجموعة متنوعة من الجزيئات العضوية (الهوائية هي من الهواء الجوي والهوائي والسير اليونانية، الحياة). يستخدم التنفس اللاهوائي من قبل بعض بدائيات النواة التي تستخدم موادها غير الأكسجين كمستقبل الكترون طرفي في عملية مشابهة لعملية التنفس الهوائية بدون استخدام أي أكسجين على الإطلاق؛ (البادئة تعني – بدون).

تفاعلات الأكسدة والاختزال: الأكسدة والاختزال.

يسمى التفاعل الذي تنقل فيه مادة واحدة إلكترونًا واحدًا أو أكثر إلى مادة أخرى تفاعل رد فعل الأكسدة أو تفاعل الأكسدة والاختزال. في تفاعل الأكسدة والاختزال، يُطلق على فقدان الإلكترونات من مادة ما أكسدة ، وتُعرف إضافة الإلكترونات إلى مادة أخرى باسم الاختزال.

Text 5:

Cellular respiration: an overview

Respiration is a cumulative function of three metabolic stages:

Glycolysis, the initial stage of all glucose metabolism, for aerobic cell respiration or for fermentation, means sugar splitting. Glucose, a six carbon sugar, is split into two pyruvate molecules, a three carbon sugar and produces some ATP via substrate phosphorylation. Glycolysis occurs in the cytosol of the cell. If oxygen is not available, or if the organism lacks enzymes needed for aerobic respiration, the pyruvate molecules will proceed with fermentations, or for some prokaryotes, anaerobic respiration. If oxygen is available and the organism has the enzymes to do aerobic respiration, the pyruvate molecules will be oxidized in the next stages of aerobic respiration.

Krebs or citric acid cycle or tricarboxylic acid cycle takes place within the mitochondrial matrix of eukaryotic cells or simply in the cytosol of prokaryotes, completes the breakdown of glucose by oxidizing a derivative of pyruvate to carbon dioxide.

Electron transport chain accepts electrons form the breakdown products of the first two stages (most of them via NADH) and passes these electrons to an electron transport chain. At the end of the chain, the electrons are combined with oxygen and hydrogen ions forming water. The energy released in this stage is used to make ATP via oxidative phosphorylation.

التنفس الخلوي: نظرة عامة

التنفس هو وظيفة تراكمية من ثلاث مراحل الأيض:

التحلل الجلدي، المرحلة الأولية من كل عملية استقلاب الجلوكوز، للتنفس الخلوي الهوائي أو للتخمير، يعني فصل السكر. ينقسم الجلوكوز، وهو عبارة عن ستة سكر من الكربون، إلى جزأين من حمض البيروفيك، وهو عبارة عن ثلاثة سكر من الكربون وينتج بعض الجزيء ATP عن طريق فسفرة الركيزة. يحدث التحلل السكري في العصارة الخلوية للخلية. إذا كان الأكسجين غير متوفر، أو إذا كان الكائن الحي يفتقر إلى الإنزيمات اللازمة للتنفيس الهوائي، فإن جزيئات البيروفات ستستمر مع التخمر، أو بالنسبة لبعض بدائيات النواة، التنفس اللاهوائي. إذا كان الأكسجين متوفراً والكائنات الحية تحتوي على الإنزيمات للقيام بالتنفس الهوائي، فسوف تتأكسد جزيئات البيروفات في المراحل التالية من التنفس الهوائي.

إن دورة كريبس أو حامض الستريك أو دورة حامض الكربوكسيليك تحدث داخل مصفوفة الميتوكوندريا للخلايا حقيقية النواة أو ببساطة في العصارة الخلوية ل prokaryotes، تكمل انهيار الجلوكوز عن طريق أكسدة مشتق من البيروفات إلى ثاني أكسيد الكربون.

تقبل سلسلة نقل الإلكترونات تشكيل الإلكترونات لمنتجات انهيار المرحلتين الأوليين (معظمها عبر NADH) وتمرير هذه الإلكترونات إلى سلسلة نقل الإلكترون. في نهاية السلسلة، يتم دمج الإلكترونات مع الأكسجين وأيونات الهيدروجين التي تشكل الماء. يتم استخدام الطاقة الصادرة في هذه المرحلة لصنع ATP عن طريق الفسفرة التأكسدية.

Text 6:

Cell division functions in reproduction, growth, and repair.

The division of a unicellular organism reproduces an entire organism, increasing the population, such as an amoeba. Cell division enables a multicellular organism to develop from a single fertilized egg or zygote egg. In a multi-cellular organism, cell division functions to repair and renew cells that die. For example, dividing cells in your bone marrow continuously make new blood cells.

Distribution of chromosomes during eukaryotic cell division.

In preparation for cell division, chromosomes replicate, each one then consisting of two identical sister chromatids joined along their lengths by adhesive protein complexes called sister chromatid cohesion. As the chromosomes condense, the region where the chromatids connect shrinks to a narrow area, the centromere. Later in the cell division process, the two sister chromatids of each duplicated chromosomes separate and move into new nuclei.

Once the sister chromatids separate, they are considered individual chromosomes. Thus, each new nucleus receives a collection of chromosomes identical to that of the parent cell.

وظائف خلية الانقسام في التكاثر والنمو والإصلاح.

إن تقسيم كائن وحيد الخلية يُعيد إنتاج كائن حي بأكمله، مما يزيد من عدد السكان، مثل الأميبا. يتيح الانقسام الخلوي للكائن الحي متعدد الخلايا أن يتطور من بويضة واحدة مخصبة أو بويضة البيضة الملقحة. في الكائنات متعددة الخلايا، وظائف الانقسام الخلوي لإصلاح وتجديد الخلايا التي تموت. على سبيل المثال، يؤدي تقسيم الخلايا في نخاع العظم إلى تكوين خلايا دم جديدة باستمرار.

توزيع الكروموسومات أثناء انقسام الخلية حقيقية النواة.

في التحضير للانقسام الخلوي ، تتكاثر الكروموسومات، كل واحد يتألف من اثنين من الكروماتين الشقيقة متطابقة انضمت على طول أطوالها من قبل مجمعات بروتين لاصقة تسمى التماسك الكروماتيني الشقيق. ومع تكاثف الكروموسومات، تتقلص المنطقة التي يتواصل فيها الكروماتيدات مع منطقة ضيقة، وهي المنطقة الوسطى. في

وقت لاحق في عملية الانقسام الخلوي، انفصل الشقيقان الشقيقان لكل صبغي من الكروموسومات وينتقلان إلى نويات جديدة.

بمجرد فصل الكروماتين الشقيقة، فإنها تعتبر كروموسومات فردية. وهكذا، تتلقى كل نواة جديدة مجموعة من الكروموسومات مماثلة لتلك الخاصة بالخلية الأم.

Text 7:

Mendel used the scientific approach to identify two laws of inheritance. In the 1860s, George Mendel discovered the basic principles of heredity by breeding gardens peas in carefully planned experiments. Mendel chose the garden peas for his studies because: garden peas are available in many varieties. For example, one variety has purple flowers, while another variety has white flowers. A heritable feature that varies among individuals, such as flower color is called a character. Each variant for a character, such as purple or white color for flowers, is termed a trait.

The feasibility of controlled pollination; the reproductive organs of a pea plant are in its flowers, and each pea flower has both pollen-producing organs (stamens) and an egg-bearing organ (carpel). In nature, pea plants usually self-fertilize: pollen grains from the stamens land on the carpel of the same flower, and sperm released from the pollen grains fertilize eggs present in the carpel.

استخدم مندل المنهج العلمي لتحديد قانونين للإرث. في ستينيات القرن التاسع عشر، اكتشف جورج مندل المبادئ الأساسية للوراثة عن طريق تربية البازلاء في تجارب مخططة بعناية. اختار مندل البازلاء الحديقة لدراسته لأن: البازلاء حديقة متوفرة في العديد من الأصناف. على سبيل المثال، يحتوي أحد الأنواع على زهور أرجوانية ، بينما يحتوي تنوع آخر على زهور بيضاء. ويطلق على ميزة وراثية تختلف بين الأفراد، مثل لون الزهرة. ويطلق على كل متغير للحرف ، مثل اللون الأرجواني أو الأبيض للزهور ، سمة.

جدوى التلقيح المضبوطة؛ إن الأعضاء التناسلية لنبات البازلاء توجد في أزهارها ، ولكل زهرة أزهار تحتوي على كل من أعضاء منتجة لحبوب اللقاح (الأسدية) وعضوية البويضة (الكارب). في الطبيعة، عادة ما تُخصب نباتات البازلاء الذاتية: حبوب حبوب اللقاح من السداة الأرض على كارب الزهرة نفسها ، والحيوانات المنوية المنطلقة من حبوب اللقاح تخصب البيض الموجود في الكارب.

Text 8:

Gene expression

Gene expression is the process by which DNA directs protein synthesis, includes two stages: transcription and translation.

Transcription is the synthesis of message RNA (m RNA) under the direction of DNA.

Translation is the synthesis of a polypeptide, which occurs under the direction of mRNA.In 1956, Francis Crick proposed what he called the central dogma of molecular biology.The central dogma, simply stated, is that DNA codes for the production of RNA,RNA codes for the production of protein, and protein does not code for the production of protein, RNA, or DNA. In Crick's words, "once 'information' has passed into protein it cannot get out again."

The genetic code :

The flow of information from gene to protein is based on a triplet code: a series of non-overlapping, three-nucleotide words. These triplets are the smallest units that can code for all the amino acids. Example: AGT at a particular position on a DNA strand results codes for the amino acid serine at the corresponding position of the polypeptide to be produced.

التعبير الجيني

التعبير الجيني هو العملية التي يوجه بها الدنا تخليق البروتين ، ويتضمن مرحلتين: النسخ والترجمة.

النسخ هو توليف رسالة RNA (م RNA) تحت إشراف الحمض النووي.

الترجمة هي تخليق عديد ببتيد ، والذي يحدث تحت إشراف mRNA. في عام 1956، اقترح فرانسيس كريك ما أسماه العقيدة المركزية للبيولوجيا الجزيئية. العقيدة المركزية ، ببساطة ، هي أن رموز الحمض النووي لإنتاج RNA، RNA رموز لإنتاج البروتين، والبروتين لا رمز لإنتاج البروتين، RNA، أو الحمض النووي. في كلمات كريك، "لمرة واحدة" قد مرت المعلومات إلى بروتين لا يمكن الخروج مرة أخرى ".

الشفرة الوراثية:

ويستند تدفق المعلومات من الجينات إلى البروتين على الشفرة الثلاثية: وهي سلسلة من الكلمات غير المتداخلة والثالثة النوكليوتيدات. هذه الثلاثية هي أصغر الوحدات التي يمكن أن ترمز لجميع الأحماض الأمينية. مثال: AGT في موضع معين على رموز نتائج الحمض النووي للسيرين من الأحماض الأمينية في الموضع المقابل لبولي ببتيد الذي يتم إنتاجه.

Text 9:

Enzymes:

Spontaneous chemical reactions may occur so slowly. For example, a solution of sucrose dissolved in sterile water will sit for years at room temperature with no appreciable hydrolysis. However, if we add a small amount of the enzyme sucrose to the solution, then all the sucrose may be hydrolyzed within seconds. How does the enzyme do this ?

An enzyme is a macromolecule that acts as a catalyst, a chemical agent that speeds up a reaction without being consumed by the reaction.

Every chemical reaction involves both bond breaking and bond forming. To hydrolyze sucrose, the bond between glucose and fructose must be broken and new bonds must form with hydrogen and hydroxyl ions from water. To reach this state, reactant molecules must absorb energy form their surroundings. The energy needed to change reactants into unstable molecular forms(transition –state species) or to push the reactants over an energy barrier so that the reaction can proceed is known as the free energy of activation, or activation energy, abbreviated as EA.

الانزيمات:

قد تحدث التفاعلات الكيميائية التلقائية ببطء شديد. على سبيل المثال، سوف يوضع محلول السكروز المذاب في الماء المعقم لسنوات في درجة حرارة الغرفة دون أي تحلل مائي ملموس. ومع ذلك، إذا أضفنا كمية صغيرة من سكروز الإنزيم إلى المحلول ، فقد يتم تحلل كل السكروز خلال ثوان. كيف يفعل الإنزيم هذا؟

الإنزيم هو جزيء ضخم يعمل كمحفز، وهو عامل كيميائي يعمل على تسريع التفاعل دون أن يستهلكه التفاعل.

كل تفاعل كيميائي يشمل كل من كسر الرابطة وتكوين الرابطة. لتحلل السكروز، يجب كسر الرابطة بين الجلوكوز والفركتوز ويجب أن تتشكل روابط جديدة مع أيونات الهيدروجين والهيدروكسيل من الماء. للوصول إلى هذه الحالة، يجب على الجزيئات المتفاعلة امتصاص الطاقة من محيطها. الطاقة اللازمة لتغيير المواد المتفاعلة في الأشكال الجزيئية غير المستقرة (الأنواع الانتقالية) أو لدفع المواد المتفاعلة فوق حاجز الطاقة بحيث يمكن أن يستمر التفاعل ، تعرف بالطاقة الحرة للتنشيط ، أو طاقة التنشيط ، والمختصرة باسم EA.

Text 10:

Monosaccharides are classified by the number of carbons in the carbon skeleton. Glucose, fructose, and other sugars that have six carbons are called hexoses. Trioses (three- carbon sugars) and pentoses (five- carbon sugars) are also common.

Glucose exists in two forms, the straight chain and the ring. In aqueous solutions, glucose molecules, as well as most other sugars, form rings.

Disaccharides

A disaccharide consists of two monosaccharides joined by a glycosidic linkage.

Maltose is a disaccharide formed by the linking of two molecules of glucose. Also known as malt sugar, maltose is an ingredient used in brewing beer.

Sucrose, table sugar, is formed by joining glucose and fructose. Sucrose is the major transport form of sugars in plants.

Lactose, milk sugar, is formed by joining glucose and galactose.

تصنف السكريات الأحادية حسب عدد الكربون في هيكل الكربون. يطلق على الجلوكوز والفركتوز والسكريات الأخرى التي تحتوي على ستة كربونات اسم hexoses. كما أن هناك ثلاث شحوم (مثل السكريات الثلاث الكربونية) والبنتوسات (خمس سكريات الكربون).

يوجد الجلوكوز في شكلين، السلسلة المستقيمة والحلقة. في المحاليل المائية ، تشكل جزيئات الجلوكوز، وكذلك معظم السكريات الأخرى، حلقات.

المركبات السكرية الثنائية

يتكون ديساكهاريد من اثنين من السكريات الأحادية التي انضمت إلى ربطglycoside.

المالتوز هو ثنائي السكاريد يتكون من ربط جزيئين من الجلوكوز. المعروف أيضا باسم سكر الشعير، المالتوز هو عنصر يستخدم في تخمير البيرة.

السكروز، سكر المائدة ، يتكون من الجلوكوز والفركتوز. السكروز هو الشكل الرئيسي لنقل السكريات في النباتات.

يتشكل اللاكتوز، سكر الحليب، عن طريق الانضمام إلى الجلوكوز والجالاكتوز.

The texts were inputted into GT in September, 2018.

جامعة النجاح الوطنية
كلية الدراسات العليا

الأخطاء اللغوية في الترجمة الآلية للنصوص العلمية البيولوجية من الانجليزية الى العربية: ترجمة جوجل

إعداد
حنان جمال علاونة

إشراف
د. عبد الكريم دراغمة

قدمت هذه الأطروحة استكمالا لمتطلبات الحصول على درجة الماجستير في اللغويات التطبيقية والترجمة، كلية الدراسات العليا، جامعة النجاح الوطنية، نابلس، فلسطين.
2019

الأخطاء اللغوية في الترجمة الآلية للنصوص العلمية البيولوجية من الانجليزية الى العربية: ترجمة جوجل

إعداد

حنان جمال علاونة

إشراف

د. عبد الكريم دراغمة

الملخص

لقد غرست ترجمة الآلة جذورها في مجالات البحث العلمي لتلعب دورا في صراع البقاء على قيد الحياة في عصر العولمة الحالي. لذلك يهدف هذا البحث الى استكشاف مناطق القوة والعجز في ترجمة جوجل للنصوص العلمية البيولوجية من الانجليزية الى العربية وبصورة أدق يسعى البحث لاختبار ترجمة جوجل على مستويين: الجملة والفقرة. لذلك سيكون كلا من نموذج كاتفورد للتغيرات في الترجمة (Catford's translation shifts) ونموذج هالداي وحسن لترابط الأفكار وتماسك أجزاء النص(Halliday and Hassan's model of cohesive devices) وأنواع الفقرات السائدة في النصوص العلمية بمثابة الأدوات التي ستستخدم لتقييم ترجمة جوجل، وفي النهاية سيحاول الباحث اقتراح مجموعة من الحلول للأخطاء الناتجة لتحسين أداء جوجل في ترجمة مثل هذه الأنواع من النصوص وتمكينه من إنتاج ترجمات بدقة عالية.