

# A Pilot Study of Biber's Model for Language Variation Detection: A Language Engineering Approach

Maram Elsaadany

*Pharos University, Alexandria, Egypt*

maram.elsaadany@pua.edu.eg

Sameh Alansary

*Bibliotheca Alexandrina, Alexandria, Egypt*

*Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt*

sameh.alansary@bibalex.org

**Abstract** — *this paper is primarily a translation analysis of Trump's speech and a letter sent to President Trump regarding family separation on "The Leadership Conference on Civil and Human Rights" to locate the differences in translation found between written and speech texts. Google translate is the engine used to translate these two texts into Arabic. The data selected for this analysis is 1000 words in each script. Biber's model (1988) used 67 features to prove that writing is more complicated than speech. The study enforces Biber's claim that writing is more complicated than speech. The findings assure Biber's claim as there were lots of problems in the translation of the speech text into Arabic in comparison to the translated written texts. Our findings strongly support the view that academic writing and conversation have dramatically different linguistic characteristics. The Results clarify the fact that google translate has to be adapted and equipped with a new grammar for speech that is different than the one used for writing to achieve the best outcome for both translations in the same language. This paper is a pioneer in its application as there is no research paper adapted such approach in the field of translation.*

**Key words:** *translation, computational, linguistic variations, speech, writing*

## 1 INTRODUCTION

Computational linguistics (CL) combines resources from linguistics and computer science to discover how human language works. Computational linguistics is a vital field in the information age. According to [1], computational linguists create tools for important practical tasks such as machine translation, speech recognition, speech synthesis, information extraction from text, grammar checking, text mining and more. [2] has stressed the idea that contrastive Analysis (CA) is a method that is connected to Contrastive Linguistics, which is considered a branch of linguistics that focuses on illustrating the differences and similarities among two or more languages at different linguistic levels as semantics, syntax, and phonology.

According to [3] earlier programs have been criticized by the lack of a dictionary; to identify linguistic features, they relied on small lists of words that were built into the program structure itself. These lists included prepositions, conjuncts, pronominal forms, auxiliary forms. Since these word lists were relatively restricted, the grammatical category of many words in texts could not be accurately identified, and therefore these programs could not identify all of the occurrences of some linguistic features. The programs have been designed to avoid skewing the frequency counts of features in one genre or another so that the relative frequencies were accurate. The main disadvantage of this earlier approach was that certain linguistic features could not be counted at all. For example, there was no way to compute a simple frequency count for the total nouns in a text, because nouns could not be identified. For these reasons, the second set of programs has been taking place.

The second stage of program development took place during the years (1985-1986). The approach used in this stage is different from that of the first stage. As a result, a general tagging program to identify the grammatical category of each word in a text was developed. The aim is to develop a program that was general enough to be used for tagging both written and spoken texts. For example, the program could not depend on upper case letters or sentence punctuation. This goal is achieved by using a large-scale dictionary together with a number of context-dependent disambiguating algorithms. The main problem that had to be solved is that many of the common words in English are ambiguous as to their grammatical category. Words like "absent" can be either adjectives or verbs; words like "acid" can be either nouns or adjectives. All past and present participial forms can function as noun (gerund), adjective, or verb. A simple word like that can function as a demonstrative, demonstrative pronoun, relative pronoun, complementizer, or adverbial subordinator.

[4] has developed algorithms to disambiguate occurrences of certain words, depending on their surrounding contexts. For example, a participial form preceded by an article, demonstrative, quantifier, numeral, adjective, or possessive pronoun is functioning as a noun or adjective. That is to say, it is not functioning as a verb in this context; given this preceding context, if the form is followed by a noun or adjective then it will be tagged as an adjective; if it is followed by a verb or preposition, then it will be tagged as a noun. Tagged texts enable automatic identification of a broad range of linguistic features that are major for differentiating between genres in English. The tagged texts are subsequently used as input to other programs that count the frequencies of certain tagged items (e.g. nouns, adjectives, adverbs) and compute the frequencies of particular syntactic constructions (e.g. relativization on subject versus non-subject position). There has also been a debate concerning the need for a linguistic comparison of speech and writing. Historically, academics have regarded writing as the true form of language, while speech has been considered to be unstable and not worthy of study. By the early twentieth century, linguists regarded speech as primary and writing as a secondary form of language derived from speech; thus, only speech was considered worth serious linguistic analysis. In fact, the historical view that written, literary language is true language continues as the dominant perception to the present time. It might well be the case that neither speech nor writing is primary; that they are rather different systems, both deserving careful analysis. This is in fact the view advocated by some researchers studying communicative competence. In this paper, the researcher is comparing the translation of google translate program in two different forms of the same language, speech and writing. In other words, are the two translations accurate? Is one of them better than the other and why?

## 2 RESEARCH QUESTIONS

1. From a language engineering point of view, can a computer program which is capable of translating written texts translate speech texts with the same accuracy?
2. Is the translated text in the written form less erroneous than the text translated in the speech form?

## 3 BIBER'S VARIATIONS

Ref. [4] model is the main model upon which this study is based. The initial step is to collect the English and Arabic texts that are used as the data used in this study. Next, text normalization is crucial for any comparison of frequency counts across texts, because text length can vary widely. Translation studies have only evolved during the last decades [5]. Scientific research in this area is a very recent phenomenon, as stressed by [6]. The call for research in translation is overwhelming as "a whole range of issues seemed to be waiting for examination, and inquiry is overdue". In [7] Grammatical competence is concerned with the linguistic structure of 'grammatical' utterances; communicative competence is concerned with the form and use of all language - both speech and writing. Within this framework, neither speech nor writing needs to be considered primary to the exclusion of the other. Rather, both require analysis, and the linguistic comparison of the two modes becomes an important question.

## 4 THE SAMPLE

The sample used for this pilot study consists of two texts. The first one is the script that represents Donald Trump's speech. It was a telephone conversation with President Zelenskyy of Ukraine. The second text is extracted from a letter written to Donald Tramp regarding the family separation. The machine translation used is Google Translate that translates both texts into Arabic. In order to normalize texts, in this study, the frequency counts of all linguistic features are normalized to a text length of 1000 words so we have to delete some words to normalize the texts to have the same length.

## 5 ANALYSIS

Translation is a complicated task, during which the meaning of the source-language text should be conveyed to the target-language readers. In other words, translation can be defined as encoding the meaning and form in the target language by means of the decoded meaning and form of the source language. [8] listed eight strategies, which have been used by professional translators, to cope with the problematic issues while doing a translation task as translation by a more general word, translation by a more neutral/ less expressive word, translation by cultural substitution, translation using a loan word, translation by paraphrase using a related or unrelated word, translation by omission or by illustration. Cohesion is the network of lexical, grammatical and other relations which provide links to various parts of the text. These relations organize a text and to some extent create it. One example of this is reference to other words and expressions in the surrounding sentences and paragraphs. Google translation tends to misuse references in the speech texts. Most of the time pronouns are not with clear references. This is really clear with the pronoun (It) as there is no clear reference to whom it refers. All the examples given in Table (1) are extracted from the speech extract as the written is translated punctually by google and does not contain any incorrect translation.

In sentence number 1, table (1), google translate misused the pronoun ( يساعدك / تساعدك ). As shown, this is apparent in google translate in the first four sentences. The problem of non-equivalence is also clear in the fact that the source language concept is not lexicalized in the target language. The source language word may express a concept which is known in the target culture but simply not lexicalized. This is evident in sentence number (5) as the word "logically" is translated literally by google translate. In other words [8] stated that the specific term in the target language for this word is misused or mis located in the target language word to express it. Google misuses this by translating it into the literal meaning of the word which is totally incorrect.

In example number (7), presupposed meaning arises from the co-occurrence restrictions on what other words or expressions that we expect to see before or after a particular lexical item. This was clear in google's translate as we spend a lot of effort and a lot of time as collocational restrictions cause many problems and the computer could not identify. These are semantically arbitrary restrictions which don't follow the prepositional meaning of a word. Moreover, Modal verbs are problematic. Their problem lies in the lack of one single, unchangeable meaning for each one of them and in the case of an absence of certain formal markers for the progressive and perfect aspects of the Arabic verb, Arab scholars denote them by using certain auxiliaries before the verb form, such as ( sawfa ), ( sa ) and ( kad ) in terms of tense. In sentence number (8), definitely the implied meaning from this sentence is not "running as in a race" and that is why google misused the word as it should be المحاولة في الاستمرار / [8] referred to problems of non-equivalence above the word level. She suggested lots of strategies to solve such problem as translation by paraphrase. This is also related to google's ignorance of the fixed expressions of the target language. As for sentence number (9), the words connote a different meaning rather than the expressed one. The literal translation of the words together does not deliver the right message as collocations and idiomatic expressions are really problematic in their translations. This according to [8] is referred to as translation beyond the word level. Register is a variety of language that a language user considers suitable to a specific situation. Its variation occurs from variations in the field of discourse as the linguistic choices will vary whether the speaker is taking part in a political speech or something else. Google misinterpreted this formality as in translating main words such as (win big), sentence number (11).

TABLE I: DIFFERENCES BETWEEN GOOGLE TRANSLATE AND CORRECT TRANSLATION

Sentences	Google Translate	Correct Translation
1. Much more than the European countries are doing and they should be helping you more than they are	أكثر بكثير مما تفعله الدول الأوروبية وينبغي أن <u>يساعدك</u> أكثر مما هو عليه	أكثر بكثير مما تفعله الدول الأوروبية وينبغي أن <u>تساعدك</u> أكثر مما هو عليه
2.They are not working as much as they should work for Ukraine	إنهم لا يعملون بقدر ما ينبغي أن <u>تعمل</u> من أجل أوكرانيا	إنهم لا يعملون بقدر ما ينبغي أن <u>يعملوا</u> من أجل أوكرانيا
3. When I was speaking to Angela Merkel she talks Ukraine, but she doesn't do anything.	عندما كنت أتحدث إلى أنجيلا ميركل تتحدث إلى أوكرانيا ، لكنها لا تفعل أي شيء	على سبيل المثال عند التحدث إلى أنجيلا ميركل تتحدث عن أوكرانيا دون فعل شيء <u>من أجلها</u>
4. I told them that they are not doing quite as much as they need to be doing on the issues with the sanctions	إنهم لا يقومون بالقدر الذي يجب عليهم فعله بشأن القضايا المتعلقة بالجزاءات	إنهم لا يقومون بالقدر الذي يجب عليهم فعله بشأن القضايا المتعلقة بالجزاءات
5. It turns out that even though logically, the European Union should be our biggest partner but technically the United States is a much bigger partner than the European Union	أنه على الرغم من المنطوق ، يجب أن يكون الاتحاد الأوروبي أكبر شريك لنا ولكن الولايات المتحدة من الناحية الفنية شريك أكبر بكثير من الاتحاد الأوروبي	على الرغم من ان الاتحاد الأوروبي هو شريكنا الأكبر لكن فعلياً الولايات المتحدة هي أقوى وأهم شريك لنا خاصة عند فرض العقوبات على الاتحاد الأوروبي
6. We are almost ready to buy more Javelins from the United States for defense purposes.	على وجه التحديد نحن على وشك. على استعداد لشراء المزيد من <u>الرمح</u> من الولايات المتحدة لأغراض الدفاع.	ونحن على استعداد لشراء السلاح اللازم من الولايات المتحدة لأغراض دفاعية
7. They are not working as much as they <b>should</b> work for Ukraine.	إنهم لا يعملون بقدر ما ينبغي أن <u>تعمل</u> من أجل أوكرانيا	لقد قمت بالتحدث إلى انجولينا ماركل بشأن الجهود الضئيلة المبذولة في قضايا العقوبات
8. Also, I think I should run more often so you can call me	أعتقد أنني يجب أن <u>أجري</u> أكثر من مرة حتى تتمكن من الاتصال بي	وأعتقد أن على بالاستمرار و مواكبه هذا التقدم لكي يستمر تواصلنا و علاقتنا ببعضنا البعض
9. We all watched from the United States and you did a terrific job. The way you came from behind	شاهدنا جميعاً من الولايات المتحدة وقمت بعمل رائع. الطريقة التي أتيت بها من الخلف	تهانينا على النصر العظيم الذي شهدناه جميعاً من الولايات المتحدة لقد قمت بعمل أكثر من رائع لقد اجتازت الفرصه و انتصرت ببراعه
10.They are not working as much as they should work	إنهم لا يعملون بقدر ما ينبغي أن يعملوا	إنهم لا يقومون بالقدر الذي يجب عليهم فعله
11.We did win big and we worked hard for this	لقد فزنا كبيرة و عملنا بجد من أجل هذا.	لقد كانت فريده من نوعها و لذلك كان انجازنا فريد من نوعه .
12. I think your country is very happy about that	أعتقد أن بلدك ستكون سعيد جداً بهذا.	أعتقد أن بلدتك سعيدة بذلك

In sentence number (12) the problem arises because of verb to be. The existential (be) does not exist in the Arabic language translation as we say / انا بخير / and not \ انا اكون بخير \. Here google translate translated a literal translation in اعتقد أن بلدك ستكون سعيدة . Table (1) addresses all the problems found in google translate. The problem of nonequivalence was apparent in google translate. This level of difficulty can vary tremendously depending on the nature of non-equivalence. Different kinds of non-equivalence require various strategies that the translator should be aware of. However, some strategies are difficult to handle. Grammar is a set of rules which determines the way in which units such as words and phrases can be combined in a language. In translating the speech, google's translate totally neglected the English structure in the way it translated the sentences into Arabic. A language can of course express any kind of information its speakers need to express, but without using the same grammatical structure.

TABLE II: BIBRE'S FACTOR BETWEEN SPEECH AND WRITING

	Google Translate	Correct Translation
Wh question	0.52	0.75
Type token ratio	0.66	0.72
Existential be	0.78	0.78
Amplifiers	0.83	0.47
Private verbs	0.73	0.64
Hedges	0.51	0.39
Contractions	0.62	0.72
Do as a pro verb	0.77	0.36
Word length	0.59	0.33
Past tense verbs	0.77	0.54
2 <sup>nd</sup> persons pronouns	0.89	0.47
Pronoun it	0.42	0.45
Synthetic negation	0.69	0.40
1 <sup>st</sup> person pronoun	0.50	0.78
Wh relative clause subject position	0.64	0.45
Split infinitives	0.58	0.71
Wh relative clause in object position	0.54	0.50
Non phrasal coordination	0.77	0.40
Demonstrative	0.54	0.63
Emphatics	0.62	0.60

Word by word translation is another fatal mistake that google has committed as this way of translating word by word reflects the lack of terms in the target language which makes the final translation very poor and totally incorrect. The tension between accuracy and naturalness is always apparent in google translate as [8] stated that it is not easy for a translator to produce a collocation which is typical to the target language. This ideal cannot be achieved with the source collocation. There are semantically arbitrary restrictions which don't follow the prepositional meaning of a word. This is also related to google misuse and ignorance of the

fixed expressions of the target language. This problem occurs as google is not aware that there are idiomatic expressions that need to be treated as a whole.

As shown in Table 2 each of thirteen features weights more than (0.4). They are wh questions, amplifiers, that deletion, sentence relatives, contractions, 2nd person, private verbs, emphatics, analytic negations, wh relative clause in subject position, 1st person, causative subordination, indefinite pronouns, demonstratives. Private verbs are used for the overt expression of private attitudes, thoughts and emotions. Wh relative clause in subject / object positions are different forms of relative clauses. They are considered as devices for explicit, elaborated identification of referents in a text. First and second person pronoun refer directly to the addresser and the addressee mentioned in a text and thus are used frequently in highly, interactive discourse. As for Wh questions which have weights between 0.5, 0.75, they are used in interactive discourse where there is a certain addressee present to give answers. Finally, emphatic and amplifiers mark extreme feelings.

## 7 FEATURES OF SPEECH

1. Integration: It refers to the way in which a large amount of information is packed into relatively few words in typical writing, because the writer operates under few time constraints and can therefore construct a carefully packaged text. In contrast, typical speech cannot be highly integrated because it is produced and comprehended on-line. Features that are used to integrate information into a text include attributive adjectives, prepositional phrase series, phrasal coordination, and careful word choice.

2. Fragmentation: It refers to the linguistic characteristics of texts produced under severe time constraints, the case for typical speech. Under these conditions, information cannot be carefully incorporated into the text, and the resulting structure is much looser, or fragmented. Linguistic features associated with a fragmented text include clauses strung together with simple conjunctions (e.g., and) or with no connectives at all. For example, I guess you have one of your wealthy people... The server, they say Ukraine has it. Here, the fragmentation is clear and it caused a problem to google as the meaning was not clear to be translated.

3. Involvement: This refers to those linguistic features which reflect the fact that speaker and listener typically interact with one another, while writer and reader typically do not. Due to this interaction, speakers often make direct reference to the listener (by use of second person pronouns, questions, imperatives, etc.), and they are typically concerned with the expression of their own thoughts and feelings (e.g., marked by use of first person pronouns, affective forms such as emphatics and amplifiers, and cognitive verbs such as think and feel). As a result of this concern, speech often has a distinctly non-informational and imprecise character (marked by hedges, pronoun it, and other forms of reduced or generalized content). These features can be considered together as the characteristics of involved text. In contrast, detachment refers to the characteristics of typical writing which result from the fact that writer and reader usually do not interact (e.g., marked by agentless passives and nominalizations). For instance, "I can assure you". Here the usage of the first and second pronouns is obvious which assures the impersonal style and the involvement of the speakers.

## 8 FEATURES OF WRITING

1. Ref. [9] stated that writing is more structurally complex and elaborate than speech, indicated by features such as longer sentences or T- units and a greater use of subordination.

2. As for [10] writing is more explicit than speech, in that it has complete idea units with all assumptions and logical relations encoded in the text.

3. [11] explained that it is more decontextualized, or autonomous, than speech, so that it is less dependent on shared situation or background knowledge.

4. Ref. [9] illustrated that less personally is involved than speech and more detached and abstract than speech.

5. Ref. [12] showed that it is characterized by a higher concentration of new information than speech (Brown and Yule 1983); and more deliberately organized and planned than speech.

6. [12] noted that 'in writing we have time to mold a succession of ideas into a more complex, coherent, integrated whole', whereas speech, because it is produced on-line is more fragmented. [15] has explained that the view of academic writing is that it is grammatically complex, with elaborated

structures, and with meaning relations expressed explicitly. On the other hand, spoken registers, especially conversation have the opposite characteristics.

## 9 CONCLUSION

There is a huge distinction between speaking and writing in their characteristics. There is also a difference in the channel. There may be many sub-channels available in speaking but only the lexical-syntactic sub-channel is available in writing. That is why, the translation provided by google translate to the written text is accurate whereas the translation fulfilled for the speech is not. Also, the opportunity for interaction with the text varies. In one hand, there are no real-time constraints in writing. On the other hand, there are severe real-time constraints that appear in speech. Even these two differences are not absolute. Features such as underlining, bold-face, and certain punctuation marks can be used to represent prosodic or paralinguistic sub-channels in writing. Tape-recorded speech bypasses some of the real-time constraints of speech, more so in comprehension than in production. The recommendation of this study is to build a new grammar to be used in translating speech which is totally different than the grammar used in translating written texts of the same language to avoid the problems that exist due to the differences between the two forms of the same language. It is highly recommended that speech should be differentiated with another system that should put into consideration the features of speech which is different than those of writing. The traditional concern of Arab linguists and translators has been with the behavior of the single word or the constituent part of the sentence rather than with the sentence as a whole. The Arabic sentence in connected discourse has received little attention in linguistic as well as rhetorical thought. In view of the research presented here, variations between speech and writing is still a field of study that requires further advancement based on the use of corpora and the refinement of parameters in register description.

## REFERENCES

- [1] Mitkov, R. (2015). *The Oxford handbook of computational linguistics*. 1st ed. Oxford University Press: Oxford
- [2] Fisiak, J. (1981). Some introductory notes concerning contrastive linguistics. In J. Fisiak (Ed.), *Contrastive linguistics and the Language Teacher* (pp. 1-11). Oxford: Pergamon.
- [3] Towell, R. & Hawkins, R. (1994). *Approaches to second language acquisition*. Clevedon: Multilingual Matters.
- [4] Biber, D. (1988). *Variation across speech and writing*, Cambridge: Cambridge University Press.
- [5] Broeck, R. (1986). Contrastive discourse analysis as a tool for the interpretation of shifts in translated texts. In House, J & Blum Kualka, S (Eds.) *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies*, Tübingen: Gunter Narr P. 37-47.
- [6] Gile, D. (1994). *Beyond testing towards a theory of educational assessment*, London: Falmer Press.
- [7] Simon, S. (1996). *Gender in Translation: Cultural Identity and politics of Translation*. London and New York: Routledge.
- [8] Baker, M. (1992). *In other words: A course book on translation*. London: Routledge.
- [9] CHAFE, Wallace. (1985). Linguistic Differences Produced by Differences Between Speaking and Writing. In OLSON, D., ORRANCE, N. & HILDYARD, A. *Literacy, Language and Learning*. Cambridge: Cambridge University Press, 1985. pp. 105-123.
- [10] Coulmas, Florian. (1996). *The Blackwell encyclopedia of writing systems*. Oxford: Blackwell.
- [11] Dewaele, J.-M. (2000). Gender, social and situational variables in the choice of speech style in native Dutch. Paper presented at the Sociolinguistics Symposium 2000, Bristol
- [12] Coffin, Caroline et al. (2003). *Teaching Academic Writing: A Toolkit for higher education*. Routledge (London).
- [13] Dockrell, J. E., & Connelly, V. (2009). The impact of oral language skills on the production of written text. *British Journal of Educational Psychology* (monograph), 6, 45–62.
- [14] Rosulek, L. (2010). *Language in Society*, 39(5), 710-711. Retrieved September 4, 2020, from <http://www.jstor.org/stable/40925827>

[15} Biber. D, Bethany. G. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness, Journal of English for Academic Purposes, Volume 9, Issue 1.

## BIOGRAPHY



**Dr. Maram Elsaadany** is a lecturer of Translation and Interpretation studies at Pharos University, Alexandria, Egypt. She has attained her PH-D degree in Translation studies from the Institute of Applied Linguistics and Translation (2018), Faculty of Arts, Alexandria University. She has also attained her Master's Degree in Applied Linguistics in (2014). Her main areas of interest are applied linguistics, Computational Linguistics and computational studies in the field of translation.

She is a former English instructor in the Arab Academy for Science and Technology. Furthermore, she is a speaking examiner for IELTS and OET exams.



**Dr. Sameh Alansary:** *Director of Arabic Computational Linguistic Center at Bibliotheca Alexandrina, Alexandria, Egypt.* He is professor of computational linguistics in the Department of Phonetics and Linguistics and the head of Phonetics and Linguistics Department, Faculty of Arts, Alexandria University. He obtained his MA in Building Arabic Lexical Databases in 1996, and his PhD from Nijmegen University, the Netherlands in building a formal grammar for parsing Arabic structures in 2002. His main areas of interest are concerned with corpus work, morphological analysis and generation, and building formal grammars. He is also the head of Arabic Computational Linguistics Center in Bibliotheca Alexandrina. He is supervising and managing the

Universal Networking Language project in Library of Alexandria since 1-6-2005 till now.

Dr. Alansary is the co-founder of the Arabic Language Technology Center (ALTEC), an NGO aims at providing Arabic Language resources and building a road map for Arabic Language Technology in Egypt and in the Middle East. He has many scientific works in Arabic Natural Language Processing published in international conferences and periodicals, and a member in many scientific organizations: (1) Egyptian Society of Language Engineering, Cairo, (2) Arabic Linguistic Society - USA, (3) Association of Computational Linguistics - USA – Europe, (4) Universal Networking Language foundation, United Nations, Geneva, Switzerland.

## دراسة تجريبية لنموذج Biber للكشف عن التباين اللغوي: منهج هندسة اللغة

مرام السعدني

جامعة فاروس- الإسكندرية- مصر

maram.elsaadany@pua.edu.eg

سامح الانصاري

مكتبه الاسكندريه – الاسكندرية - مصر

قسم الصوتيات- كلية الاداب- جامعة الاسكندرية

sameh.alansary@bibalex.org

### ملخص:

هذه الورقة هي في المقام الأول تحليل ترجمة لخطاب ترامب ورسالة مرسلة إلى الرئيس ترامب بشأن انفصال الأسرة عن "مؤتمر القيادة حول الحقوق المدنية والإنسانية". ترجمة جوجل هو المحرك المستخدم لترجمة هذين النصين إلى العربية. البيانات المحددة لهذا التحليل هي 1000 كلمة في كل برنامج نصي. استخدم نموذج ميزة لإثبات أن الكتابة أكثر تعقيداً من الكلام. تفرض الدراسة ادعاء بيبير بأن الكتابة أكثر تعقيداً من الكلام. تؤكد النتائج ادعاءات لأن هناك الكثير من المشاكل في ترجمة نص الكلام إلى العربية مقارنة بالنصوص المكتوبة المترجمة. توضح النتائج حقيقة أن يجب أن يتم تكييفها وتجهيزها بقواعد جديدة للتعبير تختلف عن تلك المستخدمة في الكتابة لتحقيق أفضل نتيجة لكلا الترجمتين في نفس اللغة. تعتبر هذه الورقة رائدة في تطبيقها حيث لا توجد ورقة بحثية تم تكييفها لهذه المطالبة في مجال الترجمة.

### الكلمات المفتاحية:

الترجمة ، الاختلافات الحسابية ، اللغوية ، الكلام ، الكتابة