Oman Royal Speeches Corpus: Compilation and Analysis

Aladdin Al Zahran¹, Rafik Jamoussi, Mahmood Zayid Suwaid Albakri, Anwaar Abduallah Salim Al-Maqbali, Iman Mohammed Ahmed Albuloshi, Ghadeer Rashid Alghefeli, Ebtihal Juma Nasser Albadri, Hiba Khalid Muslem Almandhari, Noof Mohammed Alharrasi

¹Corresponding Author: aalzahran@su.edu.om
Translation Program, Faculty of Language Studies
Sohar University, Oman

Received:08/15/2023 Accepted: 11/03/2023 Published: 12/15/2023

Abstract
For many years, researchers have directed their attention primarily toward developing written corpora, with the consequence that spoken corpora have consistently remained rare compared to written ones. The laborious transcription and annotation tasks make creating and maintaining spoken corpora a challenging endeavor. This project aims to build a transcribed corpus of Oman Royal Speeches and make it available online through a custom-made concordance tool. The study also aims to test the corpus for fundamental corpus-based lexical, stylistic, and discourse-analytical implementations. Compiling the Oman Royal Speeches Corpus is meant to fill a gap by contributing to the development of Arabic spoken language corpora and make available a research tool that can facilitate corpus-based research, uses, and applications in various areas of investigation. The corpus-building process underwent a five-stage process, including data capture, data processing, concordance tool development, testing and evaluation, and online deployment. With 98,511 tokens, the resultant corpus represents a searchable archive of Royal Speeches with a built-in online concordance tool that allows multiple search types and Keyword-in-Context query result display. The corpus has been tested for various corpus-analytic uses and has been found to provide significant findings in these areas. Thus, it has the potential to function as a reliable and authentic record and source of information for researchers and specialists in various fields, as well as a research tool allowing for various applications and analyses in language-related topics.

Keywords: Arabic corpora, Arabic political discourse, corpus analysis, corpus building, corpus linguistics, linguistic and discourse analysis, sentiment analysis, transcribed spoken corpus

Introduction

Corpus linguistics as we know it today has steadily evolved into a major branch of linguistics, and present-day corpora have significantly developed since their inception in the 1980s and 1990s. This development has been aided by rapid and spectacular advances in computer technology (McCarthy & O’Keeffe, 2010). It has equally been informed by research developments in computational linguistics and Natural Language Processing (NLP). Conversely, corpus linguistics and corpora have played a significant role in developing these two fields, illustrating the interrelation among all three disciplines (Dipper, 2008; Xiao, 2010).

The turn of the 21st century witnessed a massive influx of research into corpus linguistics and its applications, especially as a method of linguistic analysis in various fields. This extensive research output motivated some researchers to argue that corpus linguistics would exert a profound impact on the way we study language, causing refinement or redefinition of some language theories or making possible the use of other theories, which were otherwise difficult to demonstrate or apply (McEnery & Hardie, 2012).

Corpora have come to be used as a method of investigation for numerous purposes in various fields, perhaps the most important of which are language research, language teaching, and translation and interpreting. A great deal of research has dealt with the uses and applications of corpora in lexicography, language research, pedagogy and methodology, discourse analysis, semantics, stylistics, pragmatics, sociolinguistics, etc., (e.g., Bonelli, 2010; O’Keeffe & McCarthy, 2022a) and more recently multilingualism and social media (Crosthwaite, Ningrum, & Schweinberger, 2022).

This paper reports on the building of the Oman Royal Speeches Corpus (ORSC) project, which aims to compile a transcribed corpus of Oman Royal speeches and make it available through a custom-made concordancer. The paper also seeks to test this corpus to determine its effectiveness in corpus-based lexical, stylistic, sociolinguistic, and discourse-analytical investigations.

By compiling content that is otherwise scattered and putting it at the disposal of researchers through an easy-to-use online tool, the ORSC constitutes an academic contribution to Arabic Corpus Linguistics in general and Arabic spoken corpora in particular, a field that is still in its infancy, both in the local Omani and broader Arabic contexts.

The research question that this paper attempts to answer is the following:

- How effective can a corpus of oral political speeches be in capturing patterns of linguistic and discourse analytical nature?

Literature Review

A quick survey of some major contributions in the field of corpus linguistics provides an insight into the current trends and directions research into corpus linguistics has taken (see, for instance, Baker, 2012; Bowker & Pearson, 2002; McEnery & Hardie, 2012; McEnery, Hardie, & Younis, 2019; O’Keeffe & McCarthy, 2022a). These can be summarized in several broad categories, including corpus-based and corpus-driven linguistics, “corpus-assisted discourse studies” (Crosthwaite, Ningrum, & Schweinberger, 2022), corpus design, types of corpora, and corpus linguistic applications and uses of corpora.
Corpora

Corpora can be monolingual, parallel, and comparable. They can include spoken, written, or sign language. They can be diachronic or synchronic, annotated or unannotated, and general or specialized (Mc-Enery & Hardie, 2012; McEnery, Hardie, & Younis, 2019; O’Keeffe & McCarthy, 2022b). Obvious examples of corpora in English include the British National Corpus (BNP), with its 100 million words of spoken and written discourse, and the corpus of Contemporary American English (COCA Corpus), which has currently reached one billion words (Davies, 2008. See also Mansour (2013) for a review of the most common corpora in English).

Compared to written corpora, spoken corpora were for many years given lower priority and were, therefore, fewer and smaller in size (Knight & Adolphs, 2022). This outcome is mainly due to the strenuous, “time-consuming, and tedious” transcription and annotation efforts involved in compiling spoken corpora (Reppen, 2022, p. 16). However, recent technological advancements have made spoken data dramatically more accessible, with the consequence that spoken corpora are becoming an essential part of corpus linguistics, providing valuable insights into genuine language use (see Knight & Adolphs, 2022 for examples). Despite these improvements, the category of spoken corpora focusing on political discourse remains relatively rare. Among the outstanding examples in this subgenre, we find the six-million-word HKBU Corpus of Political Speeches (Ahrens, 2015) incorporating political speeches by heads of state, governors, premiers, and other high-ranking officials. Another example is the European Parliament Interpretation Corpus (EPIC), a parallel multilingual corpus of speeches in English, Italian, and Spanish and their corresponding simultaneous interpretations in all possible combinations.

Arabic Corpora

In Arabic, the number of corpora is limited, especially spoken ones. The most extensive written corpus ever reported in the literature is the 1.5 billion words Arabic Corpus (El-Khair, 2016). This corpus was compiled from over five million newspaper articles mined from ten news sources in eight Arab countries and spans over 14 years. El-Khair (2016) also lists other major written corpora with sizes ranging from a few hundred thousand, million, and hundred million to a billion words, which is the size of the 5th edition of the Arabic Gigaword corpus.

The Qur’anic Arabic Corpus (Dukes, 2009-2017) is another prominent example. It is a multimodal corpus that contains three layers of analysis: morphological annotation, a syntactic treebank, and semantic ontology, as well as audio recitations of the verses analyzed. Arifianto (2021) demonstrates how to utilize this corpus in teaching Arabic syntax.

Another example of large Arabic corpora is the International Corpus of Arabic (Bibliotheca Alexandrina, 2013), which is an ongoing project with plans to contain 100 million words of written and spoken material from four sources, including the press, internet articles, books, and academic texts.

Other major examples include the King Saud University Corpus of Classical Arabic (KSUCCA) (Alrabiah, 2013), King Abdulaziz City for Science and Technology (KACST) Arabic Corpus (King Abdulaziz City for Science and Technology, 2020), and finally, KALIMAT Multipurpose Arabic Corpus (El-Haj, 2013), which contains a little over 20,000 articles from Alwatan Omani newspaper.
Linguistic Applications of Corpora

A corpus of political discourse can be investigated from all perspectives implemented in general language corpora. Using classical concordancing tools, different types of information can be generated. These include examining vocabulary and language patterns involving such techniques as frequency, keyword, and collocation analyses. These analyses can inform studies that reveal stylistic, sociolinguistic, and discourse features that may not be current in other language types. The stylistic analysis examines rhetorical devices and figurative language (metaphors, similes, alliteration, repetition), as well as other features that contribute to aesthetic, persuasive, or emotional impact (e.g., Alduhaim, 2019; Al-Sowaidi, Banda, & Mansour, 2017).

Sociolinguistic analysis investigates aspects of language variation, such as gender, social class, etc. It examines the impact of these factors on language use. In Arabic political discourse, some commonly investigated themes or topics include diglossia, gender, identity, and politics (see, for instance, Abdul Latif, 2017; Alduhaim, 2019; Bassiouney, 2020).

With Discourse Analysis (DA), we can investigate the structure, coherence, and argumentation strategies within political speeches. Different aspects, such as discourse markers, speech acts, and argumentation patterns, can be studied to understand how speakers construct arguments to persuade audiences (Mautner, 2022). Political Discourse Analysis (PDA) is an active area of research that applies DA methodologies and tools to the language used in political contexts. With a focus on Arabic, numerous studies have explored the application of discourse analysis to political speeches, covering a wide range of topics and contexts (see, for instance, Abdul Latif, 2017; Al-Sharoufi, 2006; Al-Sowaidi, Banda, & Mansour, 2017; Golfetto, Osti, & Chakrani, 2021; Maalej, 2012).

Sentiment analysis is used to determine the overall sentiment or tone of a text. By classifying statements as positive, negative, or neutral, this type of analysis is commonly used to help identify the emotional appeal or persuasive strategies in news streams, employee surveys, and product reviews (Bing, 2012; Bordolo & Biswas, 2023; Zong, Xia, & Zhang, 2021). Sentiment analysis has also been employed for various manifestations of political discourse. Among the topics addressed here we find the study of polarity or tonality (Haselmayer & Jenny, 2017), prediction of opinion inversion (Matalon et al., 2021), defining political users’ classes (Caetano et al., 2018), opinion in parliamentary debates (Abercrombie & Batista-Navarro, 2020), the analysis of dynamic events, such as elections (Ebrahimi, Yazdavar, & Sheth, 2017), etc. By applying this type of analysis to our corpus, a wealth of information can be identified on how different issues are approached.

Topic models are an automated way to analyze text data, extract the main topics, and infer information without users having to manually read all the data (Boyd-Graber, Hu, & Mimno, 2017; Jagannathan, Roy, & Delhi, 2022; May 2022). Within political discourse, topic modeling has been applied to a variety of genres, such as parliamentary speeches (Curran et al, 2018), the media coverage of a specific topic (Rouhana, 2023), social movement studies (Lindstedt, 2019), etc. Implementation of Arabic text data is gaining momentum and includes Abuzayed & Al-Khalifa (2021) and Rouhana (2023). BERT (Bidirectional Encoder Representations from Transformers) is one of the leading models adopted (see, for instance, Abuzayed & Al-Khalifa, 2021; Alhaj et al., 2022; Rouhana, 2023).

The above literature survey indicates that Arabic language corpora and corpus-based investigations involving Arabic remain rare compared to what is available in other languages, such as English (Mansour, 2013; El-Khair, 2016) and other Western and, more recently, Asian
languages (Crosthwaite, Ningrum, & Schweinberger, 2022). This situation points to the need to develop more Arabic corpora, specifically spoken corpora. The compilation of the ORSC reported in the present article consequently represents a much-needed contribution to the development of Arabic spoken language corpora and fills a gap by making available a research tool that can facilitate corpus-based research, uses, and applications in various areas of investigation.

Methodology

ORSC Compilation

ORSC Design

Building a corpus is subject to specific considerations focusing on sampling, representativeness, and balance, whether this balance is internal, whereby corpus components are of equal size, or external, reflecting the proportions of these components in the actual world (Egbert, Biber, & Gray, 2022; Knight & Adolphs 2022; McEnery & Hardie, 2012; Reppen, 2022). Depending on the purpose of the corpus, the corpus components can be the topics included, the sources, the geographical origins, the chronological periods, etc. These considerations ensure that patterns observed in the corpus are generalizable.

Despite the ever-growing tendency to build larger corpora, there is a noticeable interest in the development of specialized and smaller-size corpora as an effective instrument for the study of specific language uses (Malmkjaer, 1998; Koester, 2022). Specialized corpora involve restricting the texts to be collected to one genre, register, or source, or to a specific time, etc. For the ORSC, the restriction affects the source as it consists exclusively of the Royal Speeches by His Majesty Haitham Bin Tariq and Late His Majesty Sultan Qaboos from 1970 to 2020. With this restriction, a “complete representation of the language” (Reppen, 2022, p. 14) is achieved since all the targeted speeches are included in the ORSC. This restriction also eliminates considerations of balance.

Research Instruments

Corpus investigation and evaluation typically use a concordance tool as a research instrument. Concordancers allow for various types of queries including simple word frequency searches, N-grams, collocations, etc. The keyword-in-context (KWIC) display format used in these tools allows for an easier identification of patterns of different sorts. For the current project, use was made of the free concordance tool AntConc (Anthony, 2023) in addition to the ORSC’s 154 concordance.

Research Procedures

The building of the ORSC project went through five major stages:

Data Capture

This task involved the identification, retrieval, and storing of the Royal Speeches from official sources.

Data Processing

Following collection, the captured data went through the following three sub-stages:
Cleaning up: This stage involved removing any non-text markers and other formatting or coding elements resulting from human error (typos) and capturing from original HTML files for better machine readability, smooth processing, and conversion of the files into XML files.

Validation: This involved several quality assurance checks to ensure that all material is verified and accurate.

Coding: This involved adding standardized speech metadata, including date, time, topic, occasion, speaker, etc., to the files. These categories will represent query filtering criteria available to the concordance tool end users.

Concordance Tool Development
This involved the development of an online concordance tool that allows for several search types using various criteria to exploit the metadata provided. The online concordance tool has been designed to take single or multiple-word queries, allowing for word frequency, collocations, and word usage searches. Results are displayed using the KWIC standard (Bonelli, 2010). In addition to up to 25 words of context on each side of the queried term, the tool allows for extended context by accessing the paragraph that contains the search result with one click and the whole text through a further click on the search result. As indicated under the ORSC design section, the concordance tool allows for different corpus management operations, including incorporating additional future speeches into the database.

Testing and Evaluation
This stage involved testing and evaluating the concordance tool developed and fine-tuning the search options.

Online Deployment
The corpus was made available online for public use at “https://apps.su.edu.om/contool/Default.aspx”.

Results
Following a compiling period of six months, a total of 108 speeches were collected. Since the ORSC is envisaged as a dynamic project, it will incrementally integrate future speeches as they become available. In line with this design principle, the Royal Speeches by His Majesty Sultan Haitham Bin Tariq in 2021 and 2022 have been added to the current analysis. The corpus speeches vary in length from as short as 148 words to as long as 4,537 words, with an average of 915 words per speech. The Annual National Day Speeches were the longest of all speeches, except for His Majesty Sultan Qaboos’s Speech to Sultan Qaboos University Students on May 2, 2000, which was the longest of all speeches in the corpus, amounting to 4,537 words. The corpus material generally falls under the genre of political speeches and covers a wide range of topics (see Figure One for ORSC themes and Figure Two for a visual representation of the most frequent words in ORSC).
Oman Royal Speeches Corpus: Compilation and Analysis

Al Zahran, Jamoussi, Albakri, Al-Maqbali Albuloshi, Alghefelfi, Albadri, Almandhari, & Alharrasi

Figure 1. ORSC Themes

*Figure 2. Visual representation of the most frequent words in ORSC generated by wordclouds.com (Zygomatic, 2023).*
Figure Three provides a view of the user interface, which includes search fields and search criteria, such as speaker or Sultan (that is, His Majesty Haitham Bin Tariq and Late His Majesty Sultan Qaboos) and date. Other search criteria include the nature of the search keyword, that is, whether it is a separate word, a phrase, or part of a group of words and the size of the context before and after the search keyword. Another control determines the number of results that should appear on one page.

![OMAN Royal Speeches Corpus concordance tool](image)

*Figure 3. ORSC concordance tool search interface*

With the ORSC ready and deployed, it went through several tests to assess its value for research. The following lines provide these analyses and applications.

**Lexical Investigations**

One of the key topics that have characterized the national debate in Oman for many years is the diversification of the economy (“تنويع” or sources of revenue”) to reduce reliance on income from oil. If a student of business or political science, a journalist, or any other researcher would like to discuss the topic of diversifying the economy, the ORSC can represent a good starting point where the researcher can easily find the first time this topic was raised in the Royal Speeches, and consequently the circumstances that have led to the discussion of this topic, historical development of the issue, and actions taken by the Government to achieve diversification of the economy since that time until the present, and so on.

A search for the keyword “تنويع” (diversification) returns 46 results related to the topic of diversification in several fields (Figure Four), with the diversification of the economy achieving the highest frequency (40).
Another example of lexical investigations is the preferential use of one item among multiple paradigmatic alternatives, as in the case of synonyms or near synonyms. Such preferential use is often meaningful and symptomatic of regional and local usage. This is the case, for instance, of the two Arabic alternatives for the word ‘year’: “سنة” (15 instances) and “عام” (270 instances). The frequencies of these two alternatives in the corpus clearly show an inclination toward using “عام” (see Table One in Appendix A). Further investigation of the distribution of these alternatives could reveal collocational patterns.

Conversely, when the plurals of these two alternatives are used, the terms “سنوات” and “سنين”, which are the plurals of “سنة” and “عام”, are used more frequently than “أعوام” and “أعوام”，the plural of “عام”.

Another example of the investigation of lexical usage is the ‘royal we’, which is also current in Arabic. As is the case in English, the use of the ‘royal we’ is confined to monarchs and heads of state. Since our corpus is dedicated to royal speeches, the occurrence of this feature is to be expected. The investigation of the corpus confirms using the ‘royal we’, which materializes through two different alternatives; “نحن” (152 instances) and “إنّنا” (489 instances) (see Table Two in Appendix B). However, it is interesting to note that cases of first-person singular are equally available, though less frequently, and usually depict instances that enhance connection with the audience.
Stylistic Analysis

Our corpus analysis has revealed the use of several important stylistic features for various rhetorical purposes. One such feature is the use of the second-person plural pronoun ‘you’ as a direct form of addressing the audience instead of addressing them using the third-person plural to create a sense of engagement, enhance audience interest, and achieve a particular rhetorical purpose. In the first example below, the use of the second-person plural pronoun constitutes an attempt to persuade the audience to view the past achievements as benefits for ‘you’, the people of Oman.

وإذا كانت الأمور تقاس بنتائجها فإنه يمكن القول بأن ما تحصل خلال الحقبة الماضية، بعون منه تعالى، هو إنجاز كبير يشهد به التاريخ لكم أنتم جميعاً يا أبناء عمان.

And if things are measured by their outcomes, then it can be said that what was achieved during the past era, with the help of Allah the Almighty, is an outstanding achievement that history bears witness to for all of you, O sons of Oman.

(1) إن الأمل معقود عليكم أنتم شباب الجامعة [...].

Indeed, hope rests upon you, the university youth [...].

(2) وأنتم... أنتم جميعاً أولئك العاملون [...].

And you, all of you, those who work [...].

A similar persuasive effect is intended in examples two and three, which appear in a speech to the graduates of the fourth cohort of Sultan Qaboos University and the 27th National Day Anniversary Speech, respectively, to persuade the young graduates of the university, and the population of workers to exert more effort in their work because they represent the future of the country.

Example three contains another stylistic device, that is, the repetition of the pronoun “أنتم”, to achieve the desired persuasive effect. Another prominent example of repetition can be seen in the use of the address “أيها المواطنون (الأعزاء)” (Dear citizens, repeated 208 times) and “أيها الإخوة المواطنين (الأعزاء)” (Dear fellow citizens, 19 times) to create a sense of engagement and solidarity with the speaker.

The use of religious references is yet another essential stylistic device that is usually employed in Arabic discourse to appeal to Arab audiences and to achieve unity and solidarity with them. In this sense, Example one in the present section contains an interesting religious reference, relating the achievements to the help of Allah, the Almighty, as a sign of religious belief and submissiveness to the will of Allah to attract the moral approval of the audience.

Sociolinguistic Analysis

In our corpus, sociolinguistic analysis may address the frequent use of national identity markers to appeal to the audience and to create a sense of solidarity, materializing through several variations (see Table Three).

<table>
<thead>
<tr>
<th>S/N</th>
<th>National Identity Variation</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(Children of Oman) أبناء عمان</td>
<td>18</td>
</tr>
<tr>
<td>2</td>
<td>(Omanis / العمانيين (For Omanis))</td>
<td>40</td>
</tr>
</tbody>
</table>
Discourse Analysis

The speeches making up the corpus are mainly ceremonial in nature and are usually addressed to the whole nation. An investigation of the way reference to the country is expressed is indicative of this status. In the corpus, reference to the country is usually made through the name of the country “عمان” (Oman). However, the more frequent use of the alternative and more emotionally loaded “وطن” (motherland) is characteristic of this type of speech (see Table Four).

Table 4. Frequencies of different references to “Oman”

<table>
<thead>
<tr>
<th>Category</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Oman) – عمان</td>
<td>151, including 19 سلطنة عمان (Sultanate of Oman)</td>
</tr>
<tr>
<td>(Motherland) – الوطن</td>
<td>162, including 66 هذا الوطن (this country)</td>
</tr>
<tr>
<td>(Our country Oman) – عماننا</td>
<td>36</td>
</tr>
<tr>
<td>(Our motherland) – وطننا</td>
<td>22</td>
</tr>
</tbody>
</table>

In addition to lexical, stylistic, sociolinguistic, and discourse-analytical investigations, further analyses that take advantage of NLP methodologies can be deployed. These include sentiment analysis.

Sentiment Analysis

In a corpus of political speeches, it is only natural that references to the country’s citizens are made in a positive tone. We have carried out a sentiment analysis on the term “مقيمين” (residents, non-Omanis) using Mazajak: An Online Sentiment Analyser (Abu Farha & Magdy, 2019) to determine whether the references to the term have been positive, neutral or negative. As is evident from Table Five (see Appendix C), the term “مقيمين” has been used ten times throughout the corpus. In most cases, the approach to the term is positive, indicating recognition of the status that non-Omani residents enjoy in Oman and appreciation of the role they play. It is interesting to note that the online sentiment analyzer has achieved 90% accuracy since the context of item number six does not warrant a negative sentiment.

Discussion

The linguistic, stylistic, sociolinguistic, and discourse-analytical investigations and sentiment analysis provide a clear answer to our research question. They adequately demonstrate that a corpus of political speeches is a rich source of authentic data that can provide insights into different areas of investigation (Knight & Adolphs, 2022). Though brief, the investigations undertaken on the ORSC identified interesting patterns that, if investigated further, could evolve into full-fledged projects. These results confirm the worth of the ORSC and make it comparable to corpora used in previous studies, whether in stylistics (Alduhaim, 2019), sociolinguistics (Bassiouney, 2020), discourse analysis (Mautner, 2022), or other.
The ORSC may equally be used in the language learning classroom (Arifianto, 2021) by teachers and researchers in the language, translation, and interpreting industries as a source of authentic examples of grammatical patterns and sentence structures, usage, collocations, terminology, rhetorical devices, etc. The corpus can also help translators and interpreters who are creating terminological databases, journalists, historians, and content writers looking for referenced data, and lexicographers developing dictionaries.

At the current development stage, the project’s outcome is the simplest form the corpus can have. At later stages, other features may be added to enrich the corpus and enhance its productivity for further research opportunities and analyses. For example, part of speech (PoS) tagging and annotation may be added to the corpus to analyze various language, lexical, structural, stylistic, and discourse features, including sentence and structural patterns. Research can focus, for instance, on whether sentences used in the corpus begin with a verb-initial structure or a subject-initial structure and the factors motivating this choice, as using one of these structures may imply the adoption of a specific rhetorical purpose such as emphasis. The results and conclusions could inform comparative research into structural choices and factors affecting such decisions in mediated (simultaneously interpreted speeches and translated texts) and non-mediated (original) discourse. The ORSC can thus be used as a reference corpus by integrating it into further larger corpora of interpreted, translated, and original Arabic speeches to allow for search and research in many different areas, including language, history, media, culture, and so on.

Conclusion

This paper has reported on the process of compiling and testing of the ORSC and its deployment through an online concordancer for carrying out various types of queries on the corpus. The findings have demonstrated that the corpus represents a viable research tool for various lexical, stylistic, sociolinguistic, and discourse-analytical corpus-based investigations, making it comparable to corpora used in other research projects. These findings pave the way for future developments of the corpus, both in terms of size, through the incorporation of new speeches, and productivity, through the implementation of PoS tagging that would further enhance query capabilities within the corpus.

Funding

The research leading to these results has received funding from Ministry of Higher Education, Research and Innovation (MoHERI) of the Sultanate of Oman under the Block Funding Program. MoHERI Block Funding Agreement No MoHERI/BFP/SU/01/2020.

About the Authors

Aladdin Al Zahran serves as an Assistant Professor of Interpreting & Translation Studies at Sohar University in Oman. His primary research focus lies in corpus-based interpreting/translation studies. Currently, Dr. Al Zahran is managing two projects: one on using semantic leads to improve corpus building & terminology extraction, and another centered on creating a multimodal corpus of simultaneously interpreted, translated and original speeches. ORCID: https://orcid.org/0000-0001-5382-6007

Rafik Jamoussi is an Associate Professor of Translation Studies at Sohar University, Oman. He has taught courses addressing terminology management and technology in translation as well as

www.awej.org
ISSN: 2229-9327
courses on literary and legal translation. His research interests include translator training, translation technology, and corpus linguistics. ORCID: https://orcid.org/0000-0003-3036-8968

References


Appendices

Appendix A

Frequencies of Alternatives to “Year(s)”

<table>
<thead>
<tr>
<th>Singular Form of Year</th>
<th>Frequencies</th>
<th>Examples from Corpus</th>
</tr>
</thead>
</table>
| عام | 270 | وسوف يضاعف هذا العدد حسب المخطط الوزاري المعد للعام الدراسي القادم إن شاء الله.  
This number will double according to the ministerial plan prepared for the next academic year; Allah willing.
| | | وفي عام ١٧٩١ صار لدينا (١٥) مستوصفا و(١٥) مركزا صحيا وستة مستشفيات.  
In the year 1971 we had (15) clinics, (15) health centers and six hospitals.
| سنة | 15 | ويسرني أن أقول لكم، إن اقتصادنا الوطني قد حقق تقدما ملموسا في السنة الماضية.  
I am pleased to inform you that our national economy has made tangible progress in the past year. |

<table>
<thead>
<tr>
<th>Plural Form of Year</th>
<th>Frequencies</th>
<th>Examples from Corpus</th>
</tr>
</thead>
</table>
| أعوام | 35 | إن الجهود الجادة والمخلصة التي قام بها المجلس خلال ثمانية أعوام قد أتاحت له أن يمارس مهمة في مختلف المجالات بكل الجدارة والحرص على تقديم المشورة الموضوعية [...]  
The serious and sincere efforts made by the Council over eight years have allowed it to carry out its duties in various fields with full merit and keenness to provide objective advice [...]. |
| سنوات | 81 | [...] ويتكون للعالم أن حياة الشعوب لا تحتسب بالسنين، وإنما تحصى بالإنجازات التي حققتها على طريق التطور الحضاري [...]  
[...] and the world must understand that the life of peoples is not measured by years, but by the achievements it has made through progress. [...]. |
During the past four years, many significant achievements have been made in the transport sector.

Appendix B
Analysis of ‘Royal We’ and Normal ‘I’

Table 2. ‘Royal We’ and normal ‘I’

<table>
<thead>
<tr>
<th>Category</th>
<th>Frequencies</th>
<th>Examples from Corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>إننا</td>
<td>489</td>
<td>واينة أصدرنا الأوامر لوزرائنا بأن يولوا شبابنا عناية خاصة وأن يتيحوا لهم كافة الفرص لكي يؤدوا دورهم على الوجه المرضي في مستقبل بلادهم. Indeed, we have issued orders to our ministers to pay special attention to our youth and to provide them with all opportunities to play their role satisfactorily in the future of their country.</td>
</tr>
<tr>
<td>نحن</td>
<td>152</td>
<td>ونحن على يقين تام، وثقة مطلقة بقدرتِكم على التعامل مع مقتضيات هذه المرحلة والمراحل التي تليها، بما يتطلبه الأمر من بصيرة نافذة وحكمة بالغة وإصرار راـسخ وتضحيات جليلة. We are absolutely certain, and we have full confidence in your ability to deal with the requirements of this stage and the stages that follow, with the required penetrating insight, profound wisdom, firm determination, and great sacrifices.</td>
</tr>
<tr>
<td>إنني</td>
<td>12</td>
<td>إنني سعيد بالحديث إليكم، وتكون السعادة أكمل حينما يكون الحديث عن عمان وما قدمناه لها وفاء وعرفانا بحقها علينا [...]. Indeed, I am happy to talk to you, and the happiness is perfect when the talk is about Oman and what we have offered to it in fulfillment and recognition of its right upon us [...].</td>
</tr>
<tr>
<td>أنا</td>
<td>26</td>
<td>وأننا أقول وأحذر بأن أمن المنطقة كل لا يتجزأ واستقرارها مسؤولية كل حكومات وشعوب المنطقة. I say and warn that the security of the Region is an indivisible whole, and its stability is the responsibility of all the governments and peoples of the region.</td>
</tr>
</tbody>
</table>

Appendix C
Sentiment Analysis of the Term “Residents”

Table 5. Sentiment analysis of “residents”

<table>
<thead>
<tr>
<th>S/N</th>
<th>Term in context</th>
<th>Sentiment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
[1] [...] We invite all citizens and residents of this good land to provide accurate information and answer all their questions [...] Positive

[2] [...] We are developing our brave armed forces and all security agencies and providing them with the means of strength that can be delivered to enable them to perform their role in protecting the state and its territorial integrity and ensure security and tranquility for its citizens and residents. Positive

[3] At the same time, we stress the need for the state's administrative apparatus to review its policies and regulations to ensure [...] the speedy adoption of decisions that guarantee furthering the interests of citizens and other residents who contribute to serving Oman and helping to build it. Positive

[4] What we are proud of is that the citizens and residents of the dear land of Oman live, by the grace of God, under the rule of the state of law and institutions [...]. Positive

[5] We also do not fail to praise the civil sectors, both productive and service, which are keen on sustaining economic activity and providing the necessary services to citizens and residents of this good land [...]. Positive

[6] Our country, especially the governorates of North and South Al Batinah, has been exposed to climatic storms, in which several citizens and residents have lost their lives [...]. Negative

[7] From the first moment, we were keen to follow up on what happened first, and our hearts glowed with supplication to Allah - the Most High, the Almighty - to protect our country, its people, and the residents of its good land from all harm. Positive
<table>
<thead>
<tr>
<th>Page</th>
<th>Arabic Text</th>
<th>English Translation</th>
<th>Sentiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>وإننا [...] نعزي أنفسنا، وذوي من فقدنا من أبناء هذا الوطن العزيز والمقيمين، أثناء الحالة المدارية [...].</td>
<td>We [...] express our condolence to the families of the citizens and residents of this dear country that we lost during the tropical storm [...].</td>
<td>Positive</td>
</tr>
<tr>
<td>9</td>
<td>ونخص بالذكر اللجنة الوطنية لإدارة الحالات الطارئة، وقواتنا المسلحة الباسلة، وشرطة عمان السلطانية، والأجهزة الأمنية، وكافة المواطنين والمقيمين [...].</td>
<td>We especially mention the National Emergency Management Committee, our brave armed forces, Royal Oman Police, the security services, and all citizens and residents [...].</td>
<td>Positive</td>
</tr>
<tr>
<td>10</td>
<td>وحرصا منا على توفير أقصى مستويات الحماية، والرعاية لأبنائنا المواطنين والمقيمين؛ فإننا نوجه الحكومة بالإسراع في دعم وتطوير منظومة الإنذار المبكر [...].</td>
<td>To provide the highest levels of protection and care for our citizens and residents, we direct the government to expedite the support and development of the early warning system [...].</td>
<td>Positive</td>
</tr>
</tbody>
</table>