LCM

LINGUE CULTURE MEDIAZIONI
LANGUAGES CULTURES MEDIATION

8 (2021)
2

Arabic Language and Language Teaching: Policies, Politics, and Ideology

Arabo e didattica dell'arabo: politiche, politica e ideologia

Edited by / A cura di
Marco Aurelio Golfetto, Letizia Osti, Brahim Chakrani

Which Arabic and Why? Policies, Politics, and Teaching 5
Marco Aurelio Golfetto, Letizia Osti, and Brahim Chakrani

Language Tests in the Arabic-speaking World: Between Ideology and Language Policy 13
Cristina Solimando

Arabic as a Language of Politics: A Case Study in Corpus-based Teaching 33
Marco Aurelio Golfetto

Language Ideologies and Policies at Work: Obstructing Arabic Instruction from the French Colonial Era to the Present 65
Brahim Chakrani

Religion, Ideology, and Nation-building in Jordanian Textbooks and Curricula for the Teaching of Arabic Language 81
Nijmi Edres

Decolonising Arabic Language Teaching: A Case Study 101
Alessandro Columbu

News Broadcasts between *fuṣḥā* and Lebanese: Language Choice as an Implicit Comment on National Identity in Lebanon 119
Helge Daniëls

Teaching Arabic as a Foreign Language in Its Diglossic Situation: Is Formal Spoken Arabic an Ideal Solution? 145
Zainab Al-Zaghir and Ghsoon Reda

Literature in Dialect: The Great Absentee 163
Francesco De Angelis

The Status and Future of Arabic Use amid Colonial Languages in the Arab World in Times of Globalization and Advanced Technology: A Political and Sociolinguistic Approach 177
Abdenbi Lachkar

Authors / Autori 211

Arabic as a Language of Politics: A Case Study in Corpus-based Teaching

Marco Aurelio Golfetto

DOI: https://dx.doi.org/10.7358/lcm-2021-002-golf

Abstract

In Arabic studies, corpus linguistics is still a fledgling discipline, but it is bound to become an unavoidable tool for specialists who seek to ground their research in actual language use. This article presents the preliminary results of an on-going research on Arabic applied linguistics. It offers a first approach to the possibility of matching Arabic as a language of politics with corpus-based teaching. The paper briefly introduces the concept of corpus-based teaching, its advantages and potential limits. It then reviews the most important Arabic corpora currently available for the language of politics, by also describing their extension, features, and limitations. Finally, it suggests some classroom activities based on arabiCorpus.

Keywords: Arabic corpora; Arabic language of politics; corpus-based teaching; corpus linguistics; TAFL.

1. Introduction

Over the last few decades, corpus linguistics has widely informed the practice of general linguists, syllabus planners, and teaching material designers, as it allows for the processing of a great amount of data in a short span of time and frees its users from personal bias on the language system, by allowing them to base their work on observation of what language actually *is* rather than on intuition on what language *should be*. Academic research on the use of corpora as a support for teaching activities and teaching tools has developed considerably for English as well as for other European language, especially since the late 1980s (for example Wichmann *et al.* 1997; Knowles *et al.* 1998; Bernardini 2000;

Hunston 2002; Sinclair 2004; Römer 2005; O'Keeffe, McCarthy, and Carter 2007a; Harris and Jaén 2010; Jaén 2010; Reppen 2010; Flowerdew 2012; Friginal 2018; Zaki 2021 [1]). In this sense, specific attention has also subsequently been devoted to ESP (for example Bowker 2002; Gavioli 2005; O'Keeffe, McCarthy, and Carter 2007b; Fuertes-Olivera 2008; Cortes 2012; Hou 2014; Legallois and Prunet 2015; Berber Sardinha 2016).

In the case of Arabic, the study of corpora is still a fledgling but promising discipline, currently "gaining both momentum and prominence within Arabic linguistics more generally" (McEnery, Hardie, and Younis 2019, 9). As such, it is bound to become an unavoidable tool for Arabic linguists and specialists of TAFL, who seek to ground their research in actual language use. Despite the technical challenges that undermined its initial development, it is worth noting that the use of corpora has already started informing the creation of teaching tools for Arabic, although at present this is limited to frequency dictionaries that are based on or cross-checked for frequency on Arabic corpora (see, for example, Buckwalter and Parkinson 2011; Kendall and Mohamed 2020; Abdelsayed 2021; Familiar 2021).

This article presents some results of an on-going research on Arabic applied linguistics. It offers a first approach to the possibility of matching Arabic as a language of politics with corpus-based teaching, by trying to answer the question whether this can be a fruitful match. The first section briefly introduces the concept of corpus-based teaching, its advantages and disadvantages. The second section reviews the most important Arabic corpora available that are relevant for the language field of politics, by also providing information on their extension, features, and limitations. The third section, after postulating and trying to test the existence of an Arabic language of politics, describes some classroom activities based on a specific corpus, i.e. arabiCorpus. The concluding remarks discuss the expected advantages and limits in the use of corpora for teaching this variety of Arabic.

---

[1] For a critical overview of research on corpus linguistics and data-driven learning, see Boulton and Tyne 2013. For a timeline of literature on experimented use of corpora in the classroom, see Boulton 2017. For a metadata analysis of the learning outcomes resulting from language learners' use of the tools and techniques of corpus linguistics, see Boulton and Cobb 2017.

2. CORPUS-BASED TEACHING

Corpus-based teaching exploits different kinds of language corpora in developing study materials for beginner and advanced learners in, potentially, all language areas. Although "corpus-driven" was used as a common label until the end of the 20th century, Tognini-Bonelli (2001) introduced a fundamental methodological distinction from the level of language analysis, drawing a dividing line between the "corpus-driven" approach and the "corpus-based" approach. Corpus-based analyses use corpus data to explore a theory or hypothesis, aiming to validate it, refine it, or refute it. In this perspective, corpus linguistics is seen as a method. On the contrary, corpus-driven linguistics rejects all preconceived theories that are generated from outside the corpora and claims that the corpus itself should be the sole source of hypotheses about language. This perspective rejects the characterization of corpus linguistics as a method, in the sense that "the corpus itself embodies a theory of language" (Tognini-Bonelli 2001, 84-85). The third section of this paper will illustrate activities grounded in the corpus-based method, where hypotheses on the language are authenticated through the corpus rather than being generated from it.

Developing teaching or learning materials from the corpora has undeniable advantages, but potential limits or difficulties should not be underestimated. As Gaskell and Cobb remark, "the various educational uses of concordancing are more talked about than tested with real users" (Gaskell and Cobb 2004, 317). Advantages and disadvantages of this method have been discussed at length over the past decades (see Reference section). The former can be summarized as follows:

1. corpora bypass the teacher's subjectivity in answering questions on the language, thus avoiding their personal bias;
2. tagged corpora are resources for generating frequency lists, concordances, collocations, etc., of language for both general and specific purposes;
3. corpora avoid word isolation, showing vocabulary, lexis, and structures in real language context of use, also allowing to retrieve readily available sentences or whole texts that can be useful for linguistic and content-based teaching;
4. bi- or multilingual parallel corpora may be employed for translation studies and students' training in translation;
5. working on corpora gets students active, by involving them in the practical exploration language;

6. it is possible for teachers, and even for students, to build their own personal, large or small corpus, whether general or focused on specific targets (linguistic areas, subjects, geographic regions, etc.).

At the same time, the limitations in the use of corpora for teaching purposes can be related to the excess of their constraints or, on the contrary, difficulties in managing the amount of data they make available:

1. corpora may have limitations in the use, or require paid subscription that not all educational institutions can afford;
2. the number of available corpora is limited, especially for Arabic on specific subjects and for specific purposes;
3. corpora for general purposes do not always offer the option to select, or even quantify, specific subcorpora and are therefore difficult to use for LSP;
4. using corpora can be time consuming for both teachers and students, as it requires technical knowledge and patience to get acquainted with it;
5. raw corpora (i.e. corpora that consist in simply raw texts, not equipped with concordancers or other tools), that are often specialized ones, are even more challenging;
6. corpora offer a huge amount of data, and dealing with them can be demanding in terms of proficiency level required to retrieve meaningful linguistic information (smaller corpora can be handled more easily, although they might appear less representative);
7. corpora are not always updated regularly, or at all, and especially those dealing with features of contemporary language can quickly become outdated.

3. A survey of existing Arabic corpora relevant for the language of politics

The number of Arabic corpora has grown dramatically over the last two decades. There are three main questions that should be considered when dealing with Arabic corpora. In the first place, one should consider their nature: whether they are general or specialized in a linguistic field, geographical area, or span of time; written or oral; learner corpora (i.e. made up of texts that consist in students' output, and that are aimed at being used for teaching purposes), or parallel corpora (made up of bi- or multilingual materials); raw corpora (i.e. only texts) or annotated corpora (i.e. with grammatical, syntactic, and semantic information), that in turn can be equipped with an interface window and concordance tools. Other important issues to consider are their actual extension and representativeness: whether their size amounts to millions of words (medium or big corpora) or less (small corpora, that can still be useful for research on specific areas), and if they are balanced or not (concerning the types of texts and genres they contain and if these latter are represented in a proportionate way). Finally, one should consider the matters of their availability and usability: whether they are free of charge or require paid subscription; if they incorporate the appropriate kind of tools that the researcher can use; and to what extent they present the user with a level of technical difficulty or ease that he/she can feasibly sustain or learn in an allotted span of time.

Over 115 Arabic corpora of various nature have been surveyed over time, however the general problem is that they are scattered across the web, and thus the task of searching and finding the appropriate resources has long rested on the individual researcher. Recently, some more comprehensive and easily accessible listings of Arabic corpora and more in-dept critical description of the available materials have been made available (Zaghouani 2014; Whitcomb and Alansary 2018; McEnery, Hardie, and Younis 2019; and Eddakouri 2020 [web] [2]). Although they are not completely exhaustive in themselves, these listings integrate each other and enumerate and critically discuss a considerable amount of existing resources.

Zaghouani (2014) reviews 66 freely available Arabic corpora. Of these, 23 are raw text corpora that do not include any kind of annotations and consist of text files; 15 are either error, PoS, and/or syntactically and semantically annotated corpora; 16 are lexical databases or word lists; the remaining 12 are speech, handwriting, or miscellaneous corpora. Unfortunately, some of the links recorded in the article's notes are not updated, and the webpage where the list was originally stored [3] no longer exists.

Whitcomb and Alansary (2018) select and critically discuss a group of seven Arabic corpora that are either free of charge or requires paid subscription. Their paper also provides a basic sketch of teaching activities based on material taken from corpora.

---

[2] https://sites.google.com/a/aucegypt.edu/infoguistics/directory/Corpus-Linguistics/arabic-corpora.

[3] https://www.qatar.cmu.edu/-wajdiz/corpora.html.

The more comprehensive McEnery, Hardie, and Younis (2019) examines different issues related to the building of corpora and their use, also reviewing various noteworthy corpora. Although it provides a shorter list, the book chapters discuss the corpora in great details.

Finally, a helpful and comprehensive resource is the annotated online list of InfoGuistics (Addakouri 2020). It lists 38 corpora of different nature, that the compiler divides into "Web-based (Searchable) corpora" and "Textual corpora (Text Files)". The webpage offers a grid with short descriptions of the corpora, mainly retrieved from their own webpages, and has the considerable advantage of allowing direct (and updated) access to the desired corpus by clicking on its name (hotword). It is worth noting that most of these resources are made available on websites that often offer both text collections and concordancing software combined.

While the majority of the existing corpora do not focus on one specific variety or language area, many are exclusively devoted to the text of the Qur'ān. Thus, experts who want to research either the language of the Qur'ān or general language, or work on the teaching of Arabic for general purposes, have at their disposal a good deal of tools offering a wide perspective on the natural use of language in an inclusive range of settings. However, they have fewer resources for the language of specialized domains, unless these general corpora have marked and searchable subcorpora. Luckily, some of the general corpora make it possible to search for more specific language target areas – typically, modern literature and the language of the Media.

Based both on the aforementioned works and personal research, it was possible to identify five resources which may be fruitfully used to research and devise teaching activities on Arabic as a language of politics.

The **International Corpus of Arabic**[4] is a general corpus created by the Library of Alexandria. It requires registration, after which it is free to use. It contains 100 million words collected between 2006 and 2013, although, so far, it is only 70% completed. The corpus is comprehensive; the creators' purpose is "to build a representative corpus of the Arabic language as it is used all over the Arab world, with the aim of supporting research on such language". Indeed, it aims at reflecting usage of Arabic both inside and outside the Arab world, by selecting materials from every Arabic-speaking country and including texts (especially from newspapers) that are published in other countries but distributed also in the Arab world. In fact, however, Egyptian MSA currently has the lion's share in the composition of the corpus (13 Ml words), followed by Saudi Arabia (8 Ml words), Oman (6 Ml), Palestine and Kuwait (5 Ml each), and UAE and Syria (4,5 Ml each). This offers the opportunity of researching on specific sociolinguistic and regional patterns in the use of regional varieties (Withcomb and Alansary 2018). The corpus includes different types of texts in MSA from newspapers (29%), web articles (20%), books (43%), and academic publications (8%), and numerous genres of literature, strategic and social science, politics, sports and religion, which were collected according to balanced weighting.

---

[4] http://www.bibalex.org/ica/ar/.

The node word query can be carried out also based on lemma, root, and stem. The search options can be set with a very high number of words before and after (it had been tested up to 15 words for the purpose for the present paper) and can be refined by choosing a specific word class (name, verb, adjective, pronoun, particle, and conjunction). Subclasses such as proper names, interjections, adverbs of various types, specific verbal tenses, and stem patterns from a long list can also be selected as further filters, in addition to specifying the number, gender, definiteness, and the noun's collocation as the first word of an *iḍāfa* construct. Ad hoc queries can be submitted by choosing the specific subcorpus of a given country. As for the subcorpora, beside choosing one or more of the aforementioned types of texts, higher levels of specificity can be selected. For instance, the section on web articles offers a politics subcorpus. No information on the size of this specific subcorpus could be retrieved. The search results section displays the number of occurrences, but it does not supply any information on frequency. The search word is highlighted in the center of each result line, preceded or followed by portions of texts of the requested size. The textual context or the sentence that include the search words can be visualized through special commands. Further metadata information is also simultaneously displayed for each text, such as its genre, source, and the country where the text was written or published. Interaction with the users is allowed through the possibility of reporting errors for each single result.

Another relevant corpus is **arabiCorpus**[5]. It is a general corpus that allows free registration (used for statistical purpose). It contains 173.5 million words. Expectations about the didactic use of the corpus are expressed by the developer, Dilworth Parkinson, based at Brigham Young University:

---

[5] https://arabicorpus.byu.edu/.

> This tool can be used to find citations and lexicographical and scholarly purposes, but it was also designed with the advanced students in mind. The hope is that teachers will be able to send students to this site with the instruction to find 5-10 good examples of a particular word, construction or idiom, and it will motivate the students to search about and discover the language in their own. (quoted in Whitcomb and Alansary 2018, 255)

This corpus is mainly based on MSA texts, with some additions of premodern Arabic and Egyptian Arabic. The most substantial part is made up of newswire from about 10 newspapers from Egypt, Jordan, Syria, Morocco, and Kuwait, in addition to a pan-Arab newspaper (*Al-Ḥayāt*) that was published in the UK until 2020. For each newspaper, news is usually collected for a span of one or two years between 1996 and 2012. Additionally, there are four main categories or genres: Modern Literature (1 million words), Nonfiction (27.95 million words), Egyptian Colloquial (160k words), and Premodern texts (9 million words, divided into *Qur'ān*, *1001 Nights*, *adab* literature, grammarians, Medieval Philosophy and Science, and *ḥadīṯ* literature). Beside the newspapers, the most interesting part of the corpus is the Nonfiction category, which includes a large amount of material in the "Islamic Discourse" subcorpus (27.3 million words). It is based on material from the Islamic Discourse website www.Sayd.net (صيد الفوائد). The remaining part of the Nonfiction subcorpus consists of political speeches and some official UN and other diplomatic documents, beside additional heterogeneous material such as literary criticism, that could be better collected separately. Each subcorpus, as well as each single newspaper, can be selected and searched individually or combined with others, by using the Advanced Search mode. This is a very helpful function for navigating texts on politics and Islamic discourse.

While the corpus is not lemmatized or part-of-speech tagged, its results are accurate enough to allow manual wading-through on the part of the researcher (see Parkinson 2019). The query can be easily performed via the interface window, by using alternatively Arabic or Latin scripts (for the latter, the user is given a transliteration chart). A basic part of speech filter, based on morphology, is provided. The query window also allows for search by string, thus making idioms, formulaic expressions, and so forth, readily available. Different types of information can be visualized through the buttons in the top toolbar.

The first basic results displayed are the number of occurrences and frequency (no. of instances per 100k words). Through the citations tool, the user can visualize the search word highlighted in red within its textual context, i.e surrounded by up to 10 words before/after. The results can be sorted according to the word that occurs immediately either before or after the word. A very helpful tool for prospective advanced teaching activities is offered by the possibility of retrieving the whole text in which the search word is contained by clicking on the number on the left side. Additional morphological information is provided by the word-forms button: this collects all the forms in which the word occurs, either definite or indefinite, conjugated, with clitics, with possessives or objective pronouns, and so forth. This is also helpful in spotting false hits (Parkinson 2019). Each instance is interactive and, by clicking on it, the user can view all the occurrences in their context again. The words-before-and-after button lists words occurring at least twice before and after the word, allowing to quickly visualize the most frequent closest collocations. Similarly, but with a wider breadth, the collocates command allows for viewing the most frequent word forms that appear up to four positions on either side of the search term. In this case, only collocates that appear at least four times are listed. In language teaching, collocations are important as they allow to learn and memorize chunks of language, instead of single words that are more difficult to memorize for learners. Finally, the tool also allows for downloading the complete list of found citations with 10 words before/after in a tab-delimited file.

Another relevant resource is **Sketch Engine** (see Alfaifi and Atwell 2016). This software is better known for languages other than Arabic. It is a widely used tool, although it provides a commercial service and requires a paid registration. As for the Arabic language, it includes the tagged Arabic web corpus AraTenTen, which contains 7.4 billion words retrieved from the web in 2012. The size of this corpus is obviously wide and appealing, but it must be noted that, despite its size, it has specific chronological limitations (for example, most of the Arab revolt and post-revolts output is excluded). Sketch Engine also allows users to create and manage their own corpus, from which they can extract concordances, word lists, collocates, and keywords.

Two Arabic corpora with a specifically political subject should be highlighted here. The first one, the **English-Arabic Political Parallel Corpus** (EAPPC) was presented as a case study in translation strategies by Al-Sayed Ahmad, Hammo, and Yagi (2017). The corpus size is 1 million words and consists of 351 Arabic and English original documents, alternatively translated into the opposite language by anonymous translators in the Royal Hashemite Court. The corpus includes 189 speeches, 80 interviews, and 68 letters selected from Jordanian King Abd Allah II's official website. In addition to this, also the Sovereign's book *Our Last Best Chance: The Pursuit of Peace in a Time of Peril* (2011) and its Arabic translation by Shukri Rahim (2011) were included . Provided that the Arabic part amounts to only 488k words, the corpus can be considered one of a small size. However, its interest lies in the fact that it is very specialized: not only does it focus on contemporary political issues and on a specific geographical area, i.e. Jordan, but it is also restricted to one top figure of the Middle Eastern politics, the King of Jordan. It is also a promising corpus as it is equipped with critical elaboration, such as meta-annotation, segmentation, tokenization, English-Arabic alignment, stemming, and PoS-tagging (Al-Sayed Ahmad, Hammo, and Yagi 2017). Furthermore, being a parallel corpus, it allows for comparison of vocabulary that is politically and culturally charged[6].

The second relevant corpus is related to international politics, i.e. diplomacy within the United Nations. The **UN Parallel Corpus** is also a multilingual parallel corpus in the six official languages of UN. A first attempt at processing and aligning documents of this corpus is described in Rafalovitch and Dale (2009). The original corpus described here consists of 2100 UN General Assembly resolutions of about 3 million tokens in average per each of the six official languages extracted from 7 sessions (55 to 62), corresponding to the period 2000-2007[7]. In particular, the Arabic subcorpus consists of 2.7 million tokens, amounting to 17.2 million characters. The texts of this initial subcorpus were preprocessed, formatting-normalized, and aligned across multiple languages at the level of paragraphs. At the time of writing, the United Nations Parallel Corpus v1.0 is stored on a different website[8] and has been considerably expanded. This freely available expanded corpus is now composed of official records and other parliamentary documents of the United Nations that are in the public domain and were produced and manually translated between 1990 and 2014. The texts, most of which are available in all the six languages, are aligned at sentence level and are available in different formats, in plain text bitexts, and as a fully aligned subcorpus (Ziemski, Junczys-Dowmunt, and Pouliquen 2016). The Arabic texts are made up of roughly over 450 million tokens. Their value consists in both the quality of the translations, that serve as official resolutions with legal implications, and the broad range of the political topics covered.

---

[6] The corpus is not available for the public at present.

[7] Among the 20 most frequent tokens reported for Arabic are: *al-umam*, *al-muttaḥida*, *al-duwal*, *al-mu'arraḫ*, *al-qirār*, and *kānūn* (Rafalovitch and Dale 2009, [5]).

[8] https://conferences.unite.un.org/UNCORPUS/en/DownloadOverview#download [01/12/2021].

4. Political keywords in arabiCorpus and corpus-based learning activities

This section presents examples of classroom activities based on one of the corpora described above. For this purpose, the best option is to work on a freely available corpus, already equipped with a query window and tools for extraction. As mentioned above, arabiCorpus is already conceived of as a complete tool for teaching purposes. It is often discussed and positively evaluated in academic literature (for example, Zaki 2017; Whitcomb and Alansary 2018; Parkinson 2019). It also has the advantage of allowing a search for Media language in specific subcorpora, as well as single newspapers from different areas, and it offers the searchable specific subcorpus "Nonfiction", consisting of an "Islamic Discourse" subcorpus and additional miscellaneous political texts. The Islamic Discourse is clearly a sensitive issue when dealing with the politics of most Arabic countries. All these factors make arabiCorpus a politically relevant and appropriate tool for language research and for devising specific teaching activities on the language of politics.

In order to ascertain if arabiCorpus can meet our expectations about the language, we shall now suggest some hypotheses related to the language of politics in terms of vocabulary frequency. To do so, existing tools and materials on the specialized language of politics and diplomacy have been exploited and lexical items have been selected for being tested in the corpus. The lexical items were chosen, in particular, from two specialized vocabularies (Kendall 2012, chapters 1, 2 and 3; Kendall and Mohamed 2020) and two textbooks (Ashtiany 1993, in particular chapter 6; Elgibali and Korica 2007, chapters 1, 2 and 3). The lexical items tested, both common nouns and nouns referring to political key-role figures can be visualized in *Tables 1* and *2*.

Table 1. – Frequency of search words relevant for politics and diplomacy in the subcorpora of Nonfiction, newspapers Al-Maṣrī al-Yawm, Al-Ġād, and Al-Ṯawra[9]*, and Modern Literature. Occurrences are given on the first line, frequencies (i.e. instances of the word/100,000) are given in brackets below (w. = words).*

Modern Literature (Novels; 1 Ml w.)	Newspaper *Al-Ṯawra* [2012?] (Syria; 16 Ml w.)	Newspaper *Al-Ġād* (02) 2010-11 (Jordan; 19 Ml w.)	Newspaper *Al-Maṣrī al-Yawm* 2010 (Egypt; 13 Ml w.)	Nonfiction (mainly Islamic Discourse, and other political texts; 27.9 Ml w.)	Sample search word
21 (2.05)	357 (2.21)	540 (2.75)	269 (1.94)	376 (1.31)	Socialism/اشتراكية (and related adj.)
24 (2.34)	1,958 (12.12)	6,313 (32.16)	4,423 (31.86)	1,064 (3.81)	Opposition/معارضة
59 (5.75)	6,800 (42.1)	12,013 (61.2)	7,765 (55.94)	8,592 (30.75)	Rights/حقوق
139 (13.55)	7,069 (34.76)	18,820 (98.88)	10,804 (77.83)	7,321 (26.2)	Regime/نظام
2 (0.19)	412 (2.55)	1,353 (6.89)	768 (5.53)	200 (0.72)	Reconciliation/مصالحة
3 (0.29)	1,856 (11.49)	5,224 (26.61)	3,834 (27.62)	345 (1.23)	Negotiations/مفاوضات
234 (22.8)	19,399 (120.09)	4,926 (25.1)	2,789 (20.09)	2,959 (10.59)	Revolution/ثورة
157 (15.3)	24,303 (150.45)	27,399 (139.59)	31,876 (229.64)	5,059 (18.1)	Council/مجلس
9 (0.88)	548 (3.39)	819 (4.17)	1,600 (11.53)	648 (2.45)	Propaganda/دعاية
6 (0.85)	989 (6.12)	1,662 (8.47)	149 (1.07)	925 (3.31)	Globalization/عَولمة
0 (0)	209 (1.29)	418 (2.31)	65 (0.47)	174 (0.62)	Sectarianism/مذهبية (and related adj.)
1 (0.1)	538 (3.34)	1,548 (7.89)	834 (6.01)	356 (1.27)	Sectarianism/طائفية (and related adj.)
38 (3.7)	5,208 (32.24)	19,269 (98.17)	21,076 (151.84)	941 (3.37)	Elections/انتخابات
4 (0.39)	802 (4.96)	1,207 (6.15)	808 (5.82)	309 (1.11)	Referemdum/استفتاء
0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	Fake news / أخبار مزيفة
0 (0)	1 (0.01)	0 (0)	0 (0)	1 (0)	Fake news / أخبار زائفة
0 (0)	0 (0)	6 (0.03)	9 (0.06)	1 (0)	Fake news / أخبار كاذبة
0 (0)	5 (0.03)	68 (0.35)	6 (0.04)	71 (0.25)	Populism/شعوبية (and related adj.)

---

[9] ArabiCorpus mostly consists of newspapers from the Eastern Arabic countries. An analysis of newspapers published in the Maghreb will certainly highlight further lexical (as well as grammatical) variation.

Table 1 shows the results for sample search words that were deemed relevant from the point of view of the language of politics and diplomacy in terms of occurrences and frequency (instances of the word/100k words) in five chosen subcorpora. Differences in their frequency have been queried in three newspaper subcorpora and the Nonfiction subcorpus, on one side, and the literary output subcorpus, on the other. Information on the subcorpora, such as their nature, consistency, geographical or intellectual provenience, and time of production, are reported in the table.

Despite their limited number, these preliminary results seem to confirm the core linguistic hypothesis. In particular, the corpora mainly confirm [10] our initial expectations about the language; as predicted, the tested words have a higher frequency in the Nonfiction corpus and the newspaper corpora rather than in the one of modern literature. A noteworthy exception is the word *ṯawra* ثورة, which is more frequent in the Modern Literature subcorpus than in the Nonfiction / Islamic Discourse one.

The Islamic Discourse subcorpus has proved to have its peculiarities, since it is sensitive to some of the tested search words (i.e. *ḥuqūq* حقوق, *niẓām* نظام, *diʿāya* دعاية, *ʿawlama* عَولمة), and by having, at the same time, distinctive frequencies from both the language of the Media and the language of the Modern Literature subcorpus. On the other hand, the language of the Media, i.e. the three corpora consisting in newspapers, contains most of the search words tested. In a way, our idea of language of politics, and especially of diplomacy, seems to be more based on and compliant with that attested in the subcorpora of newswire than any other subcorpus.

Political ideals, trends, or justifications can swiftly vary in modern and contemporary politics and the language is necessarily deeply influenced by this. The low frequency scored by by-now classical concepts such as *ištirākiyya* اشتراكية and *ʿawlama* عولمة and the absence or quasi-absence of recurrences for strictly contemporary (and sometime context-sensitive) expressions such as *maḏhabiyya* مذهبية / *ṭāʾifiyya* طائفية, *šuʿūbiyya* شعوبية and *aḫbār kāḏiba* / *zāʾifa* / *muzayyafa* أخبار كاذبة / زائفة / مزيفة are but initial examples in this sense. Accordingly, the language of politics reflected in the Media seems deeply variable (more than may be expected in a specific ideological setting), and the results can appear accordingly inconsistent, on the basis not only of the specific period of time that is tested, but also, for example, of the form of government of the country where the newspapers are published.

---

[10] But sometimes also surprisingly contradict, for example in the cases of *ištirākiyya* اشتراكية and *maḏhabiyya* مذهبية / *ṭāʾifiyya* طائفية.

Table 2. – Frequency of relevant political key-role figure words in the subcorpora of Nonfiction, newspapers Al-Maṣrī al-Yawm, Al-Ġād, and Al-Ṯawra, and Modern Literature. Occurrences are given on the above line, frequencies (i.e. instances of the word/100,000) are given in brackets below (w. = words).

Modern Literature (Novels; 1 Ml w.)	Newspaper *Al-Ṯawra* [2012?] (Syria; 16 Ml w.)	Newspaper *Al-Ġād* (02) 2010-11 (Jordan; 19 Ml w.)	Newspaper *Al-Maṣrī al-Yawm* 2010 (Egypt; 13 Ml w.)	Nonfiction (mainly Islamic Discourse and other political texts; 27.9 Ml w.)	Sample key-role word
217 (21.15)	1,736 (10.75)	11,892 (60.59)	12,424 (89.5)	12,079 (43.22)	King/ملك
0 (0)	41 (0.25)	245 (1.25)	116 (0.84)	20 (0.07)	Monarch/عاهل
361 (35.18)	40,996 (253.78)	44,621 (227.33)	146,339 (1054.25)	84.77 (30.33)	President/ رئيس[11]
4 (0.39)	2,279 (14.11)	6,169 (31.43)	2,435 (17.54)	307 (1.1)	Prime Minister / رئيس الوزراء[12]
0 (0)	2,423 (15)	82 (0.42)	5.79 (4.17)	52 (0.19)	Prime Minister / رئيس مجلس الوزراء[13]
0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	Prime Minister / وزير الوزراء[14]
0 (0)	72 (0.45)	243 (1.24)	0 (0)	87 (0.31)	Crown Prince / ولي العهد[15]
90 (8.77)	586 (3.63)	906 (4.62)	986 (7.1)	4,036 (14.44)	Sultan/سلطان
62 (6.04)	1,400 (8.67)	4,007 (20.41)	1,484 (10.69)	4.413 (15.79)	Prince/أمير

Table 2 shows the results on sample political key-role figures that are also relevant in the language of politics, given their visibility and decision-making responsibility. They are searched in the same five subcorpora used for *Table 1*.

Also in this case, the results should be treated as provisional. However, at first sight, our initial linguistic hypothesis on lexis provided less predictable results than in the case of *Table 1*. In particular, the specific query of synonyms or quasi-synonyms can confirm hypotheses beyond expectations: the alternative between *malik* ملك and *ʿāhil* عاهل shows how the former prevails on the latter in all subcorpora, thus testifying a greatly more general use. At the same time, there can be a great variability in the use of other synonyms or quasi-synonyms. Their use seemingly follows the preference of use of individual journals or the language in use in the countries where the journals are published. Both the Egyptian *Al-Maṣrī al-Yawm* and the Jordanian *Al-Ġād* prefer *raʾīs al-wuzarāʾ* رئيس الوزراء instead of *raʾīs mağlis al-wuzarāʾ* رئيس مجلس الوزراء, whereas the Syrian *Al-Ṯawra* makes uses of both depending on the country it refers to. The expression *wazīr al-wuzarāʾ* وزير الوزراء appears to be entirely obsolete. The Islamic discourse seems to prefer traditional political figures (*malik* ملك, *sulṭān* سلطان, *amīr* أمير) to other modern ones. Finally, it is worth noting that the raw results concerning lexical items related to classical political key-role figures must be manually filtered, as some nouns describing an office can also serve as proper names (for example, *Sulṭān* سلطان).

---

[11] Without previous further specification.

[12] The query was performed by string as PoS.

[13] The query was performed by string as PoS.

[14] The query was performed by string as PoS.

[15] The query was performed by string as PoS.

4.1. *Corpus-based teaching activities*

The teaching activities described here are aimed at getting students acquainted with both the use of corpora and the language of politics they contain. Corpus literacy is *per se* a learning goal that can be useful for providing basic training for prospective linguists and teachers, who should be introduced to the use of this tool during their linguistic education. From a linguistic point of view, vocabulary building is the first and most natural objective to pursue through corpus-based activities. The activities suggested in what follows require that students have been previously – and at least theoretically – introduced to the scope, structure, and features of Arabic corpora. Therefore, the activities are more likely appropriate for intermediate students of Arabic.

4.2. *Teaching activity one: frequency*

Through this activity, the students will:

- get acquainted with the corpus, explore it, learn about its basic functions, and learn how to use it;
- learn about, and how to search for, occurrences, instances, and frequency;
- use the corpus to extract specific information on frequency of given words;
- expand or consolidate vocabulary.

The task consists in filling in a grid of five selected meaningful words (nouns, at an initial stage, as they are easier to deal with) by retrieving information on their total occurrences and frequency in a specific subcorpus. Each student can be instructed to search for the same word/s in a different subcorpus. Alternatively, students can look for different words in the same subcorpus. At the end of the activity, the retrieved frequencies will be shared in class, inserted in a comparative grid, and collectively discussed to formulate hypothesis on the use of the keywords. For homework, students will devise a frequency wordlist by using the findings from the class activity. The words will be organized according to semantic families in different sheets and in the form of word webs, rather than in simple alphabetical order. *Tables 3* and *4* show the instruction for students, as relates to the task of searching for five different words in a given subcorpus and the related results. See *Figure 1* for the summary of the results for *zaʿīm*.

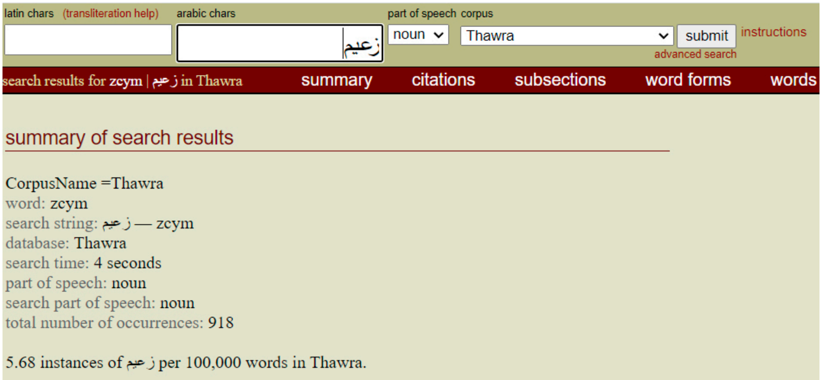
Table 3. – Task based on frequency.

TASK: Find the words listed below in the ***Al-Ṯawra*** *subcorpus. Write down the number of their total occurrences and their frequency.*

Frequency in *Al-Ṯawra*	No. of occurrences in *Al-Ṯawra*	Search word
		Leader/زعيم
		Sovereignty/سيادة
		Committee, Council / لجنة
		Authority/سلطة
		Delegate / (نائب sing.) نواب

Table 4. – Results of the task based on frequency.

Frequency in *Al-Ṯawra*	No. of occurrences in *Al-Ṯawra*	Search word
5.68/100,000	918	Leader/زعيم
18.33/100,000	2,961	Sovereignty/سيادة
76.13/100,000	12,298	Committee, Council / لجنة
26.32/100,000	4,251	Authority/سلطة
8.52/100,000	1,377	Delegate / (نائب sing.) نواب


latin chars
(transliteration help)
arabic chars
زعيم
part of speech
corpus
noun
Thawra
submit
instructions
advanced search
search results for zcym | زعيم in Thawra
summary
citations
subsections
word forms
words

summary of search results

CorpusName =Thawra
word: zcym
search string: زعيم — zcym
database: Thawra
search time: 4 seconds
part of speech: noun
search part of speech: noun
total number of occurrences: 918

5.68 instances of زعيم per 100,000 words in Thawra.

Figure 1. – Summary of results for zaʿīm.

4.3. *Teaching activity two: occurrences and collocations*

Through the second activity, students will:

- become familiar with key terms for offices, forms of government, and present and former political figures of the Arab countries;
- expand their vocabulary in the related domain;
- learn about collocations, both in the form of adjectives and names (apposition);
- explore the use of keywords in a natural and wider language context.

Students are assigned a word/expression for a political or administrative key-role office (keyword) to look up in a specific subcorpus (see task in *Tabs. 5* and *6*). They search for collocations in the form of adjectives of nationality or country names by using the "words before/after" function. Once they have retrieved the desired information, they record it on the grid provided by the teacher in order of frequency, by adding the number of occurrences in brackets. In the plenary session, students share the results of their queries and create a wordlist with all the political offices searched for and compare in which subcorpus they score the highest number of occurrences. For homework, students must complete the grid by adding the official names of the countries that they can search for independently. In addition, they will reuse the new vocabulary learned by utilizing the wordlist created in class and writing five sentences of their own See *Figure 2* for results of the search word *raʾīs*.

Table 5. – Task based on collocations.

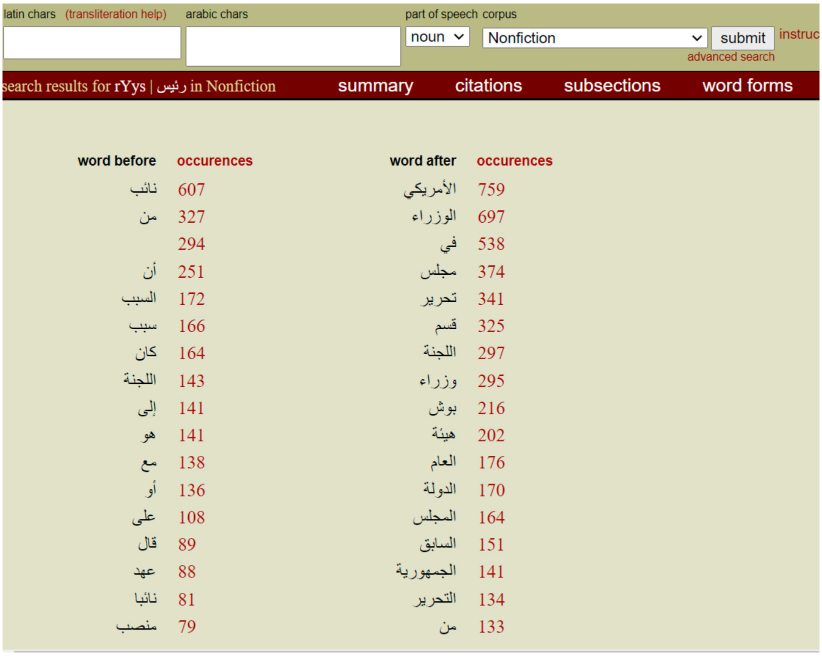
TASK: Find the word listed below in the subcorpus of ***Nonfiction****. Then write down the adjectives of nationality (or the name of the countries) that come immediately after by using the "words before/after" function. Report in brackets the resulting number of occurrences.*

Official name of the country	Adjectives of nationality or country names (add no. of occurrences)	Key-role office and subcorpus for research
		رئيس

For homework, add the official name of the countries that the subcorpus displayed after the word ra'īs. Then write five sentences of your own using different key-role political figures and new and old vocabulary.

Table 6. – Results of the task based on collocations.

Official name of the country	Adjectives of nationality or country names (add no. of occurrences)	Key-role office and subcorpus for research
	الأمريكي (297)	رئيس
	الفرنسي (44)	
	المصري (31)	
	العراقي (28)	
	الإيراني (22)	
	السوري (14)	
	الفلسطيني (10)	
	التونسي (7)	



search results for rYys | رئيس in Nonfiction

word before	occurences	word after	occurences
نائب	607	الأمريكي	759
من	327	الوزراء	697
	294	في	538
أن	251	مجلس	374
السبب	172	تحرير	341
سبب	166	قسم	325
كان	164	اللجنة	297
اللجنة	143	وزراء	295
إلى	141	بوش	216
هو	141	هيئة	202
مع	138	العام	176
أو	136	الدولة	170
على	108	المجلس	164
قال	89	السابق	151
عهد	88	الجمهورية	141
نائبا	81	التحرير	134
منصب	79	من	133

Figure 2. – Results for "words before/after" the search word ra'īs in the Nonfiction subcorpus (detail).

A more complex task, aimed at students already familiar with corpus-based activities, consists in searching for key political figures who hold or have held that specific office (see *Tabs. 7-8* and *Fig. 3* for a sample of the results). This may be carried out by using "citations". With such a task, the students are directed to wading through whole lines of text, thus getting in direct contact with bigger amounts of text. This activity also offers the opportunity of exploring under the teacher's guidance (and even before they are formally presented by him/her) the titles or formulas that are often linked to these offices, such as, for example, *ṣāḥib al-ǧalāla*, *ǧalālatu-hu*, *sumuww*, *ḫādim al-ḥaramayn al-šarīfayn*, *al-maġfūr*, etc. For homework, students are invited to relate the key-role political figures to their own countries. Alternatively, according to their level of proficiency, the students can choose a political figure among those listed and search for his/her short biography. This can be presented in class in the same form (i.e. as a biography, in the third person) or can be reworked in the form of an autobiography (i.e. in the first person) or as an interview, that can be acted out in pairs by students. Additionally, after sharing the results of their queries in class, students are invited to report the complete official names of the Arabic counties also on a blank geographical map and to sketch a map of the forms of government of the Arab countries, by creating a legend and using different colours on the map.

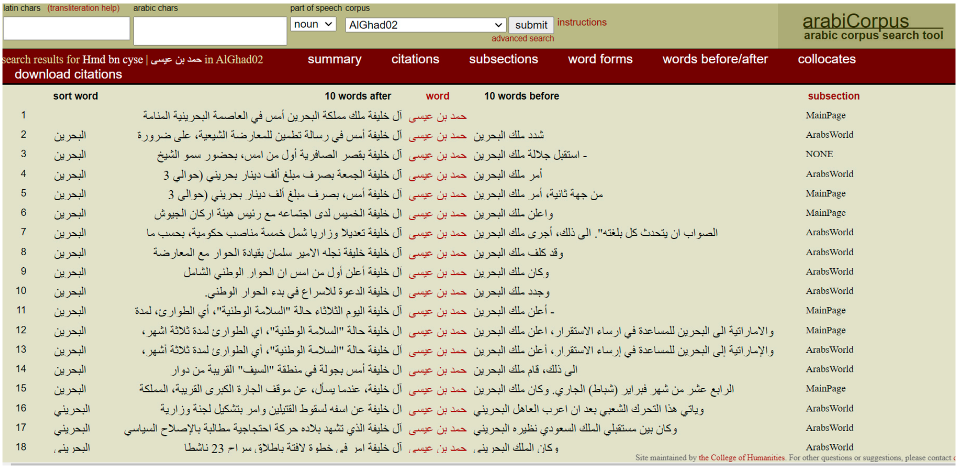
Table 7. – Task based on collocations.

TASK: Now search through the citations of the key-role office "malik" in the ***Al-Ġād*** *subcorpus and find six corresponding names of figures who hold or held that position. Then relate the six key-role figures to their own country by searching for "citations".*

Names of counties ruled by the key-role office	Personal name of the key-role figures who hold or held the office	Key-role office
		ملك

Table 8. – Results for the task based on collocations.

Names of counties ruled by the key-role office	Personal name of the key-role figures who hold or held the office	Key-role office
ملك المملكة الأردنية الهاشمية	عبدالله الثاني	ملك
الملك الراحل للأردن	الحسين بن طلال	
ملك المملكة العربية السعودية	عبدالله بن عبدالعزيز آل سعود	
ملك مملكة البحرين	حمد بن عيسى آل خليفة	
ملك المملكة المغربية	محمد السادس	
ثالث وآخر ملوك العراق (1958-1935)	الملك فيصل الثاني بن غازي	



search results for Hmd bn cyse | حمد بن عيسى in AlGhad02

	sort word	10 words after	word	10 words before	subsection
1		آل خليفة ملك مملكة البحرين أمس في العاصمة البحرينية المنامة	حمد بن عيسى		MainPage
2	البحرين	آل خليفة أمس في رسالة تطمين للمعارضة الشيعية، على ضرورة	حمد بن عيسى	شدد ملك البحرين	ArabsWorld
3	البحرين	آل خليفة بقصر الصافرية أول من امس، بحضور سمو الشيخ	حمد بن عيسى	- استقبل جلالة ملك البحرين	NONE
4	البحرين	آل خليفة الجمعة بصرف مبلغ ألف دينار بحريني (حوالي 3	حمد بن عيسى	أمر ملك البحرين	ArabsWorld
5	البحرين	آل خليفة أمس، بصرف مبلغ ألف دينار بحريني (حوالى 3	حمد بن عيسى	من جهة ثانية، أمر ملك البحرين	MainPage
6	البحرين	ال خليفة الخميس لدى اجتماعه مع رئيس هيئة اركان الجيوش	حمد بن عيسى	واعلن ملك البحرين	MainPage
7	البحرين	آل خليفة تعديلا وزاريا شمل خمسة مناصب حكومية، بحسب ما	حمد بن عيسى	الصواب ان يتحدث كل بلغته". الى ذلك، أجرى ملك البحرين	ArabsWorld
8	البحرين	آل خليفة خليفة نجله الامير سلمان بقيادة الحوار مع المعارضة	حمد بن عيسى	وقد كلف ملك البحرين	ArabsWorld
9	البحرين	آل خليفة أعلن أول من امس ان الحوار الوطني الشامل	حمد بن عيسى	وكان ملك البحرين	ArabsWorld
10	البحرين	ال خليفة الدعوة للاسراع في بدء الحوار الوطني.	حمد بن عيسى	وجدد ملك البحرين	ArabsWorld
11	البحرين	آل خليفة اليوم الثلاثاء حالة "السلامة الوطنية"، أي الطوارئ، لمدة	حمد بن عيسى	- أعلن ملك البحرين	MainPage
12	البحرين	ال خليفة حالة "السلامة الوطنية"، اي الطوارئ لمدة ثلاثة اشهر،	حمد بن عيسى	والاماراتية الى البحرين للمساعدة في ارساء الاستقرار، اعلن ملك البحرين	MainPage
13	البحرين	آل خليفة حالة "السلامة الوطنية"، أي الطوارئ لمدة ثلاثة أشهر،	حمد بن عيسى	والإماراتية إلى البحرين للمساعدة في إرساء الاستقرار، أعلن ملك البحرين	ArabsWorld
14	البحرين	ال خليفة أمس بجولة في منطقة "السيف" القريبة من دوار	حمد بن عيسى	الى ذلك، قام ملك البحرين	ArabsWorld
15	البحرين	آل خليفة، عندما يسأل، عن موقف الجارة الكبرى القريبة، المملكة	حمد بن عيسى	الرابع عشر من شهر فبراير (شباط) الجاري. وكان ملك البحرين	MainPage
16	البحريني	ال خليفة عن اسفه لسقوط القتيلين وامر بتشكيل لجنة وزارية	حمد بن عيسى	وياتي هذا التحرك الشعبي بعد ان اعرب العاهل البحريني	ArabsWorld
17	البحريني	آل خليفة الذي تشهد بلاده حركة احتجاجية مطالبة بالإصلاح السياسي	حمد بن عيسى	وكان بين مستقبلي الملك السعودي نظيره البحريني	ArabsWorld
18	البحريني	آل خليفة امر في خطوة لافتة باطلاق سراح 23 ناشطا	حمد بن عيسى	وكان الملك البحريني	ArabsWorld

Site maintained by the College of Humanities. For other questions or suggestions, please contact

Figure 3. – Results for collocations of al-Malik Ḥamad b. ʿĪsā in the Al-Ġād subcorpus.

4.4. *Teaching activity three: occurrences and collocations*

Through the third activity, the students will:

- revise grammar rules (irregular verb conjugation);
- learn how to search a range of word forms through the dedicated application;
- build their vocabulary (both reinforce old vocabulary and expand it with new one) by wading through lines of collocations;
- explore the use of words in a natural and wider language context.

Students are divided into small groups and assigned an irregular verb in specific tenses, to revise the related rules of conjugation in context through the sub/corpus and reuse its forms. Students will look up the verb in the assigned sub/corpus by selecting "verb" in the PoS menu and then "word forms" in the upper menu. They will find and select the relevant forms of the basic verb (*wazn muǧarrad*) or the augmented ones (*awzān mazīda*), such as, for example, conjugated forms of *tawallā* تولّى/take over, control; *alġā* ألغى/abolish; *nammā* نمّى/develop, promote; and so forth. In the case of verbs, students should be aware that when searching the corpus, defective verbs need two stems, past and present, separated by a comma. The teacher might suggest using the whole corpus, or a specific subcorpus, to make the search meaningful but not dispersive in terms of quantity of results. Subsequently, students report the verb forms in a table handed out by the teacher (see *Tab. 9*). Should the students be in doubt about any of the verb forms, the teacher might suggest looking for confirmation by clicking on the numbers next to the word and wading through the citations. Here, students can tell apart verb tenses using markers such as *mā*, *lā*, *lan*, or *lam*, or other more indirect hints, such as conjunctions *li-*, *in*, *law/la-*, *ʿindamā, lammā*, and the like.

Table 9. – Task based on verbal PoS and word forms.

TASK: Search for the verb تولّى ، يتولّى and select "verb" from the PoS window in the ***Al-Maṣrī Al-Yawn 2010*** *subcorpus. Select the "word forms" facility to search for the manṣūb and maǧzūm forms. Use the forms retrieved to complete the grid. If you need, look for confirmation by exploring the citations of the single word form.*

	المنصوب	المجزوم
أنا		
أنتَ		
أنتِ		
هو		
هي		
أنتما		
هما		
هما		
نحن		
أنتم		
أنتن		
هم		
هن		

Additionally, once they complete the table, students can be invited to wade through the citations and try to work out as many meanings of the verb as possible, by using the old vocabulary they already know. Alternatively, when the meaning/s of the verbs is already known, students are invited to look for the relevant "words before/after" and wade through their citations, so to explore and learn the new vocabulary that surrounds the verbs in the assigned subcorpus. The use of the dictionary is recommended only as a final resource in case students cannot understand the words on their own. For homework, students will be given a gap-fill activity, with sentences extracted from the same subcorpus they have gone thorough in the classroom, to be filled with the appropriate forms of the verbs, or the surrounding high-frequency words that they will select from a list (*Tabs. 10-11* and *Fig. 4*).

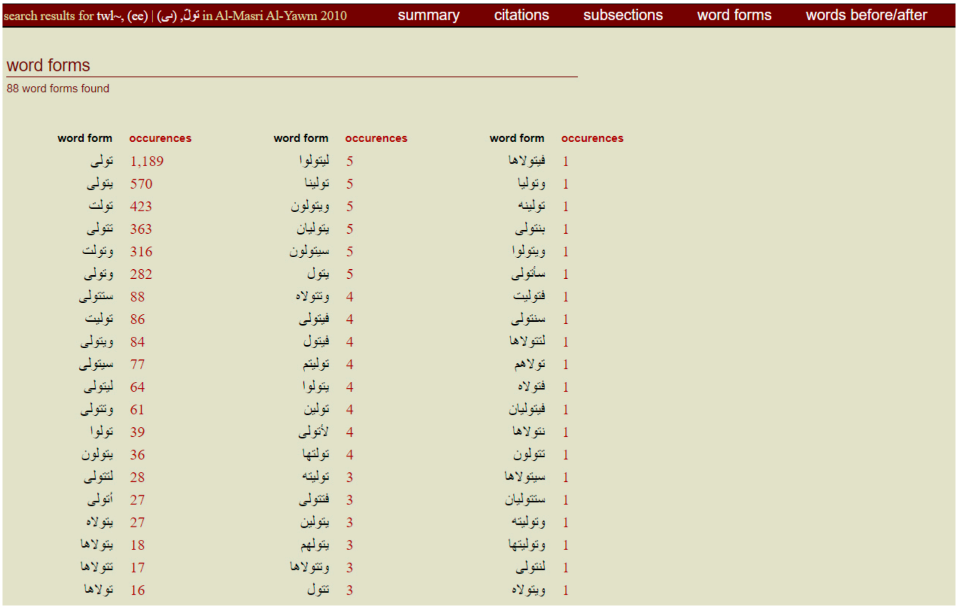
Table 10. – Gap-fill exercise based on collocations.
TASK: Complete the sentences with the words from the list. Egyptian typographic conventions are used here.

إدارة – الرئاسة – والهيئات – قيادة – مسؤوليتها - وزارة

1. تلك التصريحات جاءت بعد مرور عام على تولى ... أوباما.
2. وفى عهد الملك فؤاد تولت ... الأوقاف ... الحكومية إقامة البنايات.
3. وفى ذلك العام تولى ... الجيش العربيي. بقى فى منصب قيادته حتى 1947.
4. رئيس الوزراء هو الذى يحل محل رئيس مجلس الشعب في تولى ... مؤقتا.
5. وجهات النظر والأفكار تحولت إلى تشريعات دافعة للاقتصاد منذ تولى الحكومة الحالية ....

Table 11. – Results of gap-fill exercise based on collocations.

1. تلك التصريحات جاءت بعد مرور عام على تولى إدارة أوباما.
2. وفى عهد الملك فؤاد تولت وزارة الأوقاف والهيئات الحكومية إقامة البنايات.
3. وفى ذلك العام تولى قيادة الجيش العربى. بقى فى منصب قيادته حتى 1947.
4. رئيس الوزراء هو الذى يحل محل رئيس مجلس الشعب في تولى الرئاسة مؤقتا.
5. وجهات النظر والأفكار تحولت إلى تشريعات دافعة للاقتصاد منذ تولى الحكومة الحالية مسؤوليتها.



search results for twl~, (ee) | (ى) تولّ, in Al-Masri Al-Yawm 2010

word forms

88 word forms found

word form	occurences	word form	occurences	word form	occurences
تولى	1,189	ليتولوا	5	فيتولاها	1
يتولى	570	تولينا	5	وتوليا	1
تولت	423	ويتولون	5	تولينه	1
تتولى	363	يتوليان	5	بنتولى	1
وتولت	316	سيتولون	5	ويتولوا	1
وتولى	282	يتول	5	سأتولى	1
ستتولى	88	وتتولاه	4	فتوليت	1
توليت	86	فيتولى	4	سنتولى	1
ويتولى	84	فيتول	4	لتتولاها	1
سيتولى	77	توليتم	4	تولاهم	1
ليتولى	64	يتولوا	4	فتولاه	1
وتتولى	61	تولين	4	فيتوليان	1
تولوا	39	لأتولى	4	نتولاها	1
يتولون	36	تولتها	4	تتولون	1
لتتولى	28	توليته	3	سيتولاها	1
أتولى	27	فتتولى	3	ستتوليان	1
يتولاه	27	يتولين	3	وتوليته	1
يتولاها	18	يتولهم	3	وتوليتها	1
تتولاها	17	وتتولاها	3	لنتولى	1
تولاها	16	تتول	3	ويتولاه	1

Figure 4. – Results for word forms of the verb tawallā in the Al-Maṣrī al-Yawm 2010 subcorpus.

5. CONCLUDING REMARKS

Based on our analysis and proposed activities, the preliminary answer to the question in the Introduction is positive, corpus-based teaching and Arabic as a language of politics may indeed be matched fruitfully. While this paper has focused on one specific corpus, similar positive results are expected from the use of other Arabic political corpora, as they can allow substantial information for linguists, syllabus designers, teachers, and students interested in the field of politics and diplomacy. They are useful for showing students natural usage of language and vocabulary in real context. They can be exploited for collecting information both *on* the language (how it works, its frequency, and other specific variations, thus determining teaching priorities) and *in* the language (selecting content-meaningful texts) of politics.

When used in the appropriate context and with consistent objectives, the deriving teaching activities can be highly motivating for learners. They can be used to encourage students to discover language information before being presented with it, thus activating learning-by-discovery processes, fostering their autonomy, and helping them develop skills of analysis and critical thinking. By offering task-based activities (Ellis *et al.* 2020), corpus-based teaching makes students play an active role in their own process of learning, avoiding automatic repetition and calling learners to search, find, reflect, organize, sequence, and reuse in a more stimulating way than classical teaching materials usually do.

However, at present, even more than in the case of corpora for general purposes (Zaki 2017), one should also admit that the use of Arabic political corpora for fully devised teaching materials is still to come. Some concurrent factors play a role in this: the limited number of Arabic political corpora available; the limitations inherent to some of these (fees required; unexpected unavailability; absence of built-in tools); their limited nature, as, on the one hand, they do not cover all the types of texts and genres in which politics and politicians, diplomacy and diplomats can express themselves and, on the other, they can be representative of only a restricted chronological period, i.e. the span of time in which the texts are collected, especially when we consider how fast the language of political trends and argumentations can change. In this sense, Arabic political corpora are apparently time sensitive and need to be constantly updated. For all these reasons, as a pragmatic choice, corpus-based activities are best thought of and used as complementary to existing textbook-based syllabi, especially those devised for Media Arabic. Additionally, given the general complexity of use, these activities may be more wisely conceived for a specific target of students (i.e. MA students), preferably providing extra-curricular classes to allow a sufficient amount of time for familiarizing with and training on the target corpus and its concordancing tools.

Nevertheless, the teaching activities presented here illustrate the many possible prospective uses of corpora for the purpose of teaching Arabic of politics. The potential for this approach can grow in proportion to both the learners' level of proficiency and autonomy and the availability of appropriate and properly tagged corpora. More complex and refined research can be carried out, which will bring further teaching implications related to the discovery and acquisition of collocations, recognition of word chunks, automatization of frequently recurring structures and phrases, in the lexical, grammatical, and syntactic domains. Corpora consisting of official documents produced by political or administrative institutions can easily highlight standard formulaic expressions, whereas bi- or multilingual aligned official corpora (such as the UN Parallel Corpus) will certainly display recurrent translation solution that can become a standard: all this will be fruitful for teaching students of Arabic and training future translators. When relevant corpora will become available, it will also be possible to finetune research in and teaching of Arabic political discourse, by investigating, for example, leaders' speeches and their ideological constructs and rhetoric. The door is wide open, but the task ahead is huge.

References

Abdelsayed, Ibraam. 2021. *Lessico di frequenza dell'arabo parlato in Egitto (LAPE)*. Pisa: ETS.

Aijmer, Karin. 2009. *Corpora and Language Teaching*. Amsterdam: John Benjamins.

Al-Batal, Mahmoud. 2006. "Playingwith Words: Teaching Vocabulary in the Arabic Curriculum". In *Handbook for Arabic Language Teaching Professionals in the 21st Century*, edited by Kassem Wahba, Zeinab Taha, and Liz England, 331-340. Mahwah (NJ): Lawrence Erlbaum Publishers.

Al-Khalifa, Hend, and Abdulmohsen Al-Thubaity, eds. 2014. *Proceedings of Workshop on Free/Open-source Arabic Corpora and Corpora Processing Tools*. Reykjavik: Springer.

Al-Sayed Ahmad, Alia, Bassam Hammo, and Sane Yagi. 2017. "English-Arabic Political Parallel Corpus: Construction, Analysis, and a Case Study in Translation Strategies". *Jordanian Journal of Computers and Information Technology* 3: 157-171.

Ashtiany, Julia. 1993. *Media Arabic*. Edinburgh: Edinburgh University Press.

Aston, Guy, Silvia Bernardini, and Dominic Stewart. 2004. *Corpora and Language Learners*. Amsterdam: Benjamins.

Atwell, Eric, and Abdullah Hardie, eds. 2013. *Proceedings of WACL'2 Second Workshop on Arabic Corpus Linguistics*. Lancaster: Lancaster University.

Atwell, Eric, Latifa Al-Sulaiti, Saleh Al-Osaimi, and Bayan Abu Shawar. 2004. "A Review of Arabic Corpus Analysis Tools". Dans *Proceedings of TALN04: XI Conference sur le traitement automatique des langues naturelles*, édité par Bernard Bel and Isabelle Marlien, vol. 2, 229-234. Lille (France): ATALA.

Bassiouney, Reem. 2020. *Arabic Sociolinguistics*. Washington (DC): Georgetown University Press.

Beard, Adrian. 2000. *The Language of Politics*. London: Routledge.

Bennett, Gena R. 2010. *Using Corpora in the Language Learning Classroom: Corpus Linguistics for Teachers*. Ann Arbor (MI): University of Michigan Press.

Berber Sardinha, Tony. 2016. "Corpus-based Teaching in LSP". In *Technology-enhanced Language Learning for Specialized Domains*, edited by Elena Martín Monje, Izaskun Elorza, and Blanca García Riaza, 203-215. Abigdon - New York: Routledge.

Bernardini, Silvia. 2000. *Competence, Capacity, Corpora: A Study in Corpus-aided Language Learning*. Bologna: CLUEB.

Boulton, Alex. 2017. "Corpora in Language Teaching and Learning: Research Timeline". *Language Teaching* 50 (4): 483-506.

Boulton, Alex, and Tom Cobb. 2017. "Corpus Use in Language Learning: A Meta-analysis". *Language Learning* 67 (2): 348-393.

Boulton, Alex, and Stephan Wilhelm. 2006. "Habeant Corpus – They *Should* Have the Body: Tools Learners Have the Right to Use". *ASP* 49-50: 155-170.

Bowker, Lynne, and Jennifer Pearson. 2002. *Working with Specialized Language: A Practical Guide to Using Corpora*. London - New York: Routledge.

Braun, Sabine, Kurt Kohn, and Joybrato Mukherjee, eds. 2006. *Corpus Technology and Language Pedagogy*. Frankfurt: Peter Lang.

Buckwalter, Tim, and Dilworth Parkinson. 2011. *A Frequency Dictionary of Arabic*. London: Routledge.

Campoy-Cubillo, Mari Carmen, Begoña Bellés Fortuño, and Maria Lluïsa Gea Valor. 2010. *Corpus-based Approaches to English Language Teaching*. London: Continuum.

Conrad, Susan M. 1999. "The Importance of Corpus Based Research for Language Teachers". *System* 27 (1): 1-18.

Cortes, Viviana. 2012. "Corpora in the Teaching of Language for Specific Purposes". In *The Encyclopedia of Applied Linguistics*, edited by Carol A. Chapelle, [1-7]. Hoboken: Blackwell. [01/12/2021]. https://doi.org/10.1002/9781405198431.wbeal0225.

Ebrahimi, Alice, and Esmail Faghih. 2017. "Integrating Corpus Linguistics into Online Language Teacher Education Programs". *ReCALL* 29 (1): 120-135.

Edo Marzá, Nuria. 2014. "A Practical Corpus-based Approach to Teaching English for Tourism". *International Journal of Applied Linguistics & English Literature* 3 (1): 129-136.

Elgibali, Alaa, and Nevenka Korica. 2007. *Media Arabic: A Coursebook for Reading Arabic News. Luġat wasā'il al-i'lām al-'arabiyya.* Cairo - New York: American University in Cairo Press.

Familiar, Laila. 2021. *A Frequency Dictionary of Contemporary Arabic Fiction: Core Vocabulary for Learners and Material Developers*. Abingdon - New York: Routledge.

Flowerdew, Lynne. 2012. *Corpora and Language Education*. Basingstoke: Palgrave Macmillan.

Frankenberg-Garcia, Ana. 2012. "Raising Teacher's Awareness of Corpora". *Language Teaching* 45 (4): 475-489.

Friginal, Eric. 2018. *Corpus Linguistics for English Teachers: New Tools, Online Resources, and Classroom Activities*. New York: Routledge.

Fuertes-Olivera, Pedro A. 2017. "Pedagogical Application of Specialized Corpora in ESP Teaching: The Case of the UVaSTECorpus". *Scripta Manent* 3 (2): 68-81.

Gajšt, Nataša. 2012. "Technical Terminology in Standard Terms and Conditions of Sale: A Corpus-based Study of High Frequency Nouns and Their Collocations". *Scripta Manent* 7 (2): 33-50.

Gaskell, Delian, and Thomas Cobb. 2004. "Can Learners Use Concordance Feedback for Writing Errors?". *System* 32 (3): 301-319.

Gavioli, Laura. 2005. *Exploring Corpora for ESP Learning*. Amsterdam - Philadelphia: John Benjamins.

Granger, Sylviane, Joseph Hung, and Stephanie Petch-Tyson. 2002. *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins.

Haider, Ahmad S., and Riyad F. Hussein. 2019. "Analysing Headlines as a Way of Downsizing News Corpora: Evidence from an Arabic-English Comparable Corpus of Newspaper Articles". *Digital Scholarship in the Humanities* 35 (4): 826-844.

Harris, Tony, and María Moreno Jaén, eds. 2010. *Corpus Linguistics in Language Teaching*. Bern: Peter Lang.

Heywood, Andrew. 2019. *Politics*. 5th ed. London: Macmillan Education.

Hunston, Susan. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Jaén, María Moreno, Fernando Serrano Valverde, and María Calzada Pérez. 2010. *Exploring New Paths in Language Pedagogy: Lexis and Corpus-based Language Teaching*. London - Oakville: Equinox Publishing Ltd.

Jafarpour, Ali Akbar, Mahmood Hashemian, and Sepideh Alipour. 2013. “A Corpus-based Approach toward Teaching Collocation of Synonyms”. *Theory and Practice in Language Studies* 3 (1): 51-60.

Johns, Tim. 1988. “Whence and Whither Classroom Concordancing?”. In *Computer Applications in Language Learning*, edited by Theo Bongaerts, Pieter de Haan, Sylvia Lobbe, and Herma Wekker, 9-27. Dordrecht: Foris.

Johns, Tim. 1990. “From Printout to Handout: Grammar and Vocabulary Teaching in the Context of Data-driven Learning”. *CALL Austria* 10 (1): 14-34.

Johns, Tim. 1991. “Should You Be Persuaded: Two Examples of Data-driven Learning”. *English Language Research Journal* 4 (1): 1-16.

Keck, Casey. 2020. “Corpus Linguistics in Language Teaching”. In *Encyclopedia of Applied Linguistics*, edited by Carol A. Chapelle. Hoboken: Blackwell.

Kendall, Elisabeth. 2012. *Media Arabic: The Top 1,300 Words for Understanding*. Edinburgh: Edinburgh University Press.

Kendall, Elisabeth, and Yehya A. Mohamed. 2020. *Diplomacy Arabic: An Essential Vocabulary*. Washington (DC): Georgetown University Press.

Kennedy, Graeme. 1998. *An Introduction to Corpus Linguistics*. London: Longman.

King Abd Allah II. 2011. *Our Last Best Chance: The Pursuit of Peace in a Time of Peril*. New York: Viking Press [Arabic transl. by Shukri Rahim, *Furṣatunā l-aḫīra: al-sa‘ī naḥw l-salām fī waqt al-ḫaṭar*. Bayrūt: Dār al-Saqī, 2011].

Larsen-Freeman, Diane. 2012. “On the Roles of Repetition in Language Teaching and Learning”. *Applied Linguistics Review* 3 (2): 195-210.

Leech, Geoffrey. 1997. “Teaching and Language Corpora: A Convergence”. In *Teaching and Language Corpora*, edited by Anne Wichmann, Steven Fligelstone, Tony McEnery, and Gerry Knowles, 1-23. London: Longman.

Leech, Geoffrey. 2001. “The Role of Frequency in ELT: New Corpus Evidence Brings a Re-appraisal”. Papers presented at the *3rd International Symposium on ELT in China 2001*, edited by Wenzhong Hu, 1-23. Beijing: Foreign Language Teaching and Research Press.

Legallois, Dominique, and Anne Prunet. 2015. “Sequential Patterns: A New Corpus-based Method to Inform the Teaching of Language for Specific Purposes”. *Journal of Social Science* 11 (3): 128-138.

Lesniewska, Justyna. 2006. “Collocations and Second Language Use”. *Studia Linguistica* 123 (1): 95-105.

Lewis, Michael. 1993. *The Lexical Approach*. Hove: Language Teaching Publications.

Mansour, Mohamed Abdelmageed. 2013. “The Absence of Arabic Corpus Linguistics: A Call for Creating an Arabic National Corpus”. *International Journal of Humanities and Social Science* 3 (12): 81-90.

Marcos Miguel, Nausica. 2021. “Exploring Tasks-as-process in Spanish L2 Classrooms: Can Corpus-based Tasks Facilitate Language Exploration, Language Use, and Engagement?”. *International Journal of Applied Linguistics* 31 (2, Supplement: *Corpus-based Language Teaching and Learning: Applications and Implications*): 211-228.

Martín-Monje, Elena, Izaskun Elorza, and Blanca García Riaza, eds. 2016. *Technology-enhanced Language Learning for Specialized Domains*. London - New York: Routledge.

McEnery, Tony, Andrew Hardie, and Nagwa Younis, eds. 2019. *Arabic Corpus Linguistics*. Edinburgh: Edinburgh University Press.

McEnery, Tony, and Andrew Wilson. 1997. "Teaching and Language Corpora". *ReCALL* 9 (1): 5-14.

McEnery, Tony, and Richard Xiao. 2010. "What Corpora Can Offer in Language Teaching and Learning". In *Handbook of Research in Second Language Teaching and Learning*, edited by Eli Hinkel, vol. 2, 364-380. London: Routledge.

Meyer, Charles F. 2002. *English Corpus Linguistics: An Introduction*. Cambridge: Cambridge University Press.

Miller, Catherine, François Siino, et Myriam Catusse. 2015. "Les langues du politique. Le regard d'une sociolinguiste". *Revue des mondes musulmans et de la Méditerranée* 138. [05/12/2021]. http://journals.openedition.org/remmm/9263. doi: https://doi.org/10.4000/remmm.9263.

Mukherjee, Joybrato. 2006. "Corpus Linguistics and Language Pedagogy: The State of the Art – and Beyond". In *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*, edited by Sabine Braun, Kurt Kohn, and Joybrato Mukherjee, 5-24. Frankfurt am Main: Peter Lang.

Nesselhauf, Nadja. 2003. "The Use of Collocations by Advanced Learners of English and Some Implications for Teaching". *Applied Linguistics* 24 (2): 223-242.

O'Keeffe, Anne, and Fiona Farr. 2003. "Using Language Corpora in Initial Teacher Education: Pedagogic Issues and Practical Applications". *TESOL Quarterly* 37 (3): 389-418.

O'Keeffe, Anne, Michael McCarthy, and Ronald Carter. 2007a. *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge: Cambridge University Press.

O'Keeffe, Anne, Michael McCarthy, and Ronald Carter. 2007b. "Specialising: Academic and Business Corpora". In *From Corpus to Classroom: Language Use and Language Teaching*, edited by Anne O'Keeffe, Michael McCarthy, and Ronald Carter, 198-219. Cambridge: Cambridge University Press.

Parkinson, Dilworth. 2006. *Using Arabic Synonyms*. Cambridge: Cambridge University Press.

Partington, Alan. 1996. *Using Corpora for English Language Research and Teaching*. Amsterdam: John Benjamins.

Philp, Jenefer, and Susan Duchesne. 2016. Exploring Engagement in Tasks in the Language Classroom. *Annual Review of Applied Linguistics* 36 (1): 50-72.

Pinna, Antonio. 2007. "Exploiting LSP Corpora in the Study of Foreign Languages". In *Languages for Specific Purposes: Searching for Common Solutions*, edited by Dita Gavola, 146-163. Newcastle (England): Cambridge Scholars Press.

Rafalovitch, Alexandre, and Robert Dale. 2009. "United Nations General Assembly Resolutions: A Six-language Parallel Corpus". Paper presented at *MT (International Association of Machine Translation) Summit XII Proceedings*, Ottawa, Canada, August 26, 2009, [1-8].

Rahab, Hichem, Abdelhafid Zitouni, and Mahieddine Djoudi. 2020. "An Enhanced Corpus for Arabic Newspapers Comments". *International Arab Journal of Information Technology* 17 (5): 789-798.

Reppen, Randi. 2010. *Using Corpora in the Language Classroom*. Cambridge: Cambridge University Press.

Roberts, Andrew, Latifa Al-Sulaiti, and Eric Atwell. 2006. "aConCorde: Towards an Open-source, Extendable Concordancer for Arabic". *Corpora* 1 (1): 39-57.

Römer, Ute. 2005. *Progressives, Patterns, Pedagogy: A Corpus-driven Approach to English Progressive Forms, Functions, Contexts and Didactics*. Amsterdam: John Benjamins.

Römer, Ute. 2011. "Corpus Research Applications in Second Language Teaching". *Annual Review of Applied Linguistics* 31: 205-225.

Saad, Motaz, and Wesam M. Ashour. 2010. "OSAC: Open Source Arabic Corpora". Paper presented at *6th International Conference on Electrical and Computer Systems (EECS '10)*, Lefke, North Cyprus, November 25-26, 2010, [1-6]. [01/12/2021]. https://www.researchgate.net/publication/234136250_OSAC_Open_Source_Arabic_Corpora.

Shaw, Peter A. 2009. "The Syllabus Is Dead, Long Live the Syllabus: Thoughts on the State of Language Curriculum, Content, Language, Tasks, Projects, Materials, Wikis, Blogs and the World Wide Web". *Language and Linguistics Compass* 3 (5): 1266-1283.

Sinclair, John. 1991. *Corpus, Concordance, Collocation: Describing English Language*. Oxford: Oxford University Press.

Sinclair, John. 2003. *Reading Concordances*. Harlow: Longman.

Sinclair, John, ed. 2004. *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins.

Thompson, Paul, and Sarah Nasser Alzeer. 2019. "A Survey of Issues, Practices and Views Related to Corpus-based Word Lists for English Language Teaching and Learning". *International Journal of Applied Linguistics & English Literature* 8 (6): 43-53.

Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at Work*. Amsterdam - Philadelphia: John Benjamins.

Vyatkina, Nina. 2013. "Discovery Learning and Teaching with Electronic Corpora in an Advanced German Grammar Course". *Die Unterrichtspraxis* 46 (1): 44-61.

Whitcomb, Alansary. 2018. "Using Linguistic Corpora in Arabic Foreign Language Teaching and Learning". In *Handbook for Arabic Language Teaching Profes-*

sionals in the 21st Century, vol. 2, edited by Kassem M. Wahba, Liz England, and Zeinab A. Taha, 219-231. New York - London: Routledge.

Wichmann, Anne, Steven Fligelstone, Tony McEnery, and Gerry Knowles, eds. 1997. *Teaching and Language Corpora*. London: Longman.

Yuan, Fei, and Shichunzi Zhang. 2021. *The Application of Corpora on the Lexical Approach*. Bristol: IOP Publishing (IOP Conference Series. Earth and Environmental Science, 693).

Zaghouani, Wajdi. 2014. "Critical Survey of the Freely Available Arabic Corpora". In *Proceedings of Workshop on Free/Open-source Arabic Corpora and Corpora Processing Tools*, edited by Hend Al-Khalifa and Abdulmohsen Al-Thubaity, 1-8. Reykjavik: Springer.

Zaki, Mai. 2017. "Corpus-based Teaching in the Arabic Classroom: Theoretical and Practical Perspectives". *International Journal of Applied Linguistics* 27 (2): 514-541.

Zaki, Mai. 2021a. "Corpus-based Language Teaching and Learning: Applications and Implications". *International Journal of Applied Linguistics* 31 (2): 169-172.

Zaki, Mai. 2021b. "Self-correction through Corpus-based Tasks: Improving Writing Skills of Arabic Learners". *International Journal of Applied Linguistics* 31 (2): 193-210.

Zerrouki, Taha, and Amar Balla. 2017. "Tashkeela: Novel Corpus of Arabic Vocalized Texts, Data for Auto-diacritization Systems". *Data in Brief* 11 (C): 147-151.

Zhao, Wen, and Yuanyuan Mu. 2020. "An Empirical Study of Corpus-based Translation Teaching in Higher Vocational Colleges in China". In *Emerging Technologies for Education*, edited by Elvira Popescu, Tianyong Hao, Ting-Chia Hsu, Haoran Xie, Marco Temperini, and Wei Chen, 280-284. Cham (CH): Springer International Publishing.

Ziemski, Michał, Marcin Junczys-Dowmunt, and Bruno Pouliquen. 2016. *The United Nations Parallel Corpus, Language Resources and Evaluation (LREC '16)*, Portorož, Slovenia, May 23-28, 2016.

Copyright (©) 2021 Marco Aurelio Golfetto
Editorial format and graphical layout: copyright (©) LED Edizioni Universitarie

This work is licensed under a Creative Commons
Attribution-NonCommercial-NoDerivatives 4.0 International License.

How to cite this paper:
Golfetto, Marco Aurelio. "Arabic as a Language of Politics: A Case Study in Corpus-based Teaching". *Lingue Culture Mediazioni / Languages Cultures Mediation - LCM* 8, 2 (2021), 33-63. DOI: https://dx.doi.org/10.7358/lcm-2021-002-golf