

Article

Collocational Patterns and Semantic Prosodies of “Saudization” in Modern Standard Arabic Discourse: A Corpus-Based Analysis

Basant S. M. Moustafa 

Department of English Language and Literature, College of Science and Humanities in Al-Kharj, Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia; b.moustafa@psau.edu.sa

Abstract

Since its intensification in the early 2000s, Saudization, perceived as Saudi Arabia’s labor nationalization policy, has profoundly shaped the socio-economic and socio-political discourse in the Arab world. In spite of the extensive policy and economic research on Saudization, it remains linguistically underexplored, particularly through corpus-based analysis of Arabic language discourse. The current study, hence, investigates the collocational patterns of the term “Sa’wadah” (the transliterated Arabic language equivalent of “Saudization”) in a mega-corpus of Modern Standard Arabic (MSA) with the aim of sketching its semantic prosodies and working towards the ideological representations permeating through its collocates. To this end, using the methodological synergy of Critical Discourse Analysis and Corpus Linguistics, the Arabic corpus (arTenTen24) integrated within the online interface of Sketch Engine is scrutinized to extract a statistically significant collocational profile for Saudization and work out its semantic prosodies and their ideological representations. The findings show that the collocational profile of Saudization can be categorized into institutional, economic and evaluative topic domains. Second, the semantic prosodies associated with the term are mainly mixed, with a slight tendency towards negative and critical evaluation. Third, the ideological representations of Saudizations in MSA discourse are marked by ambivalence: the policy is constructed as a national duty whose implementation is a contested practice.

Keywords: Saudization; semantic prosody; modern standard Arabic; labor nationalization; critical discourse analysis; corpus linguistics



Academic Editor: Julien Longhi

Received: 15 July 2026

Revised: 7 September 2026

Accepted: 9 September 2026

Published: 15 September 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Saudization, *سعودة* or “Sa’wadah”, represents Saudi Arabia’s strategic initiative to increase the employment of Saudi nationals in the private sector, reducing dependence on expatriate labor while addressing youth unemployment and economic diversification objectives aligned with Vision 2030 (Peck, 2017). Since the implementation of Nitaqat quota system in 2011, Saudization has evolved from a policy framework into a pervasive discourse shaping economic, social and political narratives across media, government communication and public debate in the Arab World in general and Saudi Arabia in particular. Since language is a powerful and constitutive tool in policy implementation, ideological representations, power relations and social values (Fairclough, 1995; Van Dijk, 2008), the lexical choices, collocational patterns, and semantic prosodies surrounding policy terms

reveal underlying attitudes, evaluative stances and discursive strategies employed by different social actors. In the case of Saudization, understanding how the term and its collocates are used in MSA can highlight evaluative prosodies, whether positive, negative or neutral, as to the nationalization policies. Building on semantic prosodies, the ideological representations of Saudi nationals, expatriate workers and labor market dynamics can also be illuminated.

Utilizing corpus tools, the current study investigates how the term *سعودة* (Saudization), as a linguistically and socially salient term, is collocationally represented in MSA discourse with its concomitant attitudinal and evaluative shades. The particular selection of the current topic as the focus of research is, hence, motivated by its linguistic, social and discursive significance. On the linguistic level, *السعودة* (Saudization) is a neologism derived from the proper noun “as-Sa’ūdiyyah” (Saudi Arabia), which manifests particularly interesting morphological, semantic and pragmatic features in Arabic. The methodological synergy of Critical Discourse Analysis, henceforth CDA, and Corpus Linguistics, henceforward CL, along with the corpus-driven notions of collocation and semantic prosody, have been adopted by various studies, either separately or in combination. However, to date, the use of a corpus-based approach to investigate the notion of Saudization in MSA has not been tackled, to the best of the researcher’s knowledge. Moreover, to date, the use of collocation and semantic prosody as corpus-linguistic analytical approaches to explore the term *سعودة* (Saudization) has received little, if any, scholarly attention, thereby indicating a significant gap in the existing literature.

On the social and discursive levels, with the increasing importance of the Kingdom of Saudi Arabia both locally and internationally and the leading role it plays in the socio-political as well as the socio-economic discourse, crises and tensions in the region, all Saudi-related terms need to be thoroughly studied. Particularly, associated with the national employment strategies and policies of Vision 2030, the concept of “Saudization” has come to acquire predominant importance in media discourse, academic discourse, official statements, labor market discussions, etc. “Saudization” is also an ideologically laden term, frequently co-occurring with evaluative language—either positively (e.g., empowerment, reform, nationalization) or negatively (e.g., inefficiency, unemployment, restriction)—which makes it ideally fitting for examining representational and discursive patterns. Its socio-political and socio-economic weight also allows for the intersection of linguistic analysis with broader questions of identity, political discourse, and social change in the context of the Arab Gulf as well as the whole Arab region. The present study, hence, hypothesizes that the methodological synergy of CDA and CL can provide a robust model for interpreting the ideological representations of the term *السعودة* (Saudization) in an MSA corpus.

The current study aims to answer the following research questions: (1) What collocational profile does the term *سعودة* (Saudization) have in MSA discourse? (2) What are the semantic prosodies associated with *سعودة* (Saudization) based on corpus investigation? and (3) What ideological representations does *سعودة* (Saudization) convey in an MSA mega-corpus? The remainder of this research paper is organized in four sections. The first section is the literature review, where the theoretical underpinnings of analysis and the previous studies are introduced. The second section details the methodology, describing the data set and the analytical procedure employed. The third section presents the analysis, in which the node word is examined in relation to its collocates across concordance lines as a means of identifying semantic prosodies and ideological representations. The discussion section then addresses the findings in relation to the research questions followed by a concluding section that summarizes the study’s key contributions.

2. Literature Review

The aim of this section is to sketch the linguistic field relevant to the present study, encompassing its central topic, Saudization; its methodological approach, namely the synergy of CDA and CL; and its core analytical constructs, collocation and semantic prosody.

2.1. Saudization as a National Policy

Most existing research on Saudization has been conducted from econometric, sociological or policy evaluation perspectives (Randeree, 2012; Hertog, 2014). Cortés et al. (2020) provide rigorous econometric evaluation of Saudi Arabia's Nitaqat quota system, reporting an increase in the Saudi national employment rate, particularly among women. The study also draws attention to the decreasing size of firms, lower productivity and increasing firm exit rates. Rajan and Arokkiaraj (2019) provide a comprehensive overview of Gulf labor nationalization policies, e.g., Saudization, Omanization, Emiratization, Qatarization, etc., across the Gulf Cooperation Council (GCC) countries and conclude that despite different timelines and intensities, the implementation of employment localization in all GCC countries is staggering. They argue that nationalization is linked to post-oil economic planning and economic diversification objectives and that the *kafāla* (labor sponsorship) system shapes migrant–national employment dynamics.

From a linguistics perspective, since MSA is the primary register of government policy documents, official communication and news media, an exploration of a large-scale mega-corpus of MSA discourse contributes essentially towards an understanding of how Saudization is framed, disseminated and represented to the Arabic-speaking population. Although, to date, no study has focused on the neologism “Saudization” in MSA using corpus-based techniques, the term has been examined through a range of theoretical frameworks, models and approaches. For example, Salama (2023) explored the advertising discourse order of “Saudization,” while Salama and Fawzy (2024) examined its construction of the Saudi national identity.

2.2. The Methodological Synergy of CL and CDA

There has been a noticeable increase in the use of methods related to CL in conducting CDA studies over the last two decades. CDA envisages discourse as a social practice subsuming “a dialectical relationship between a particular discursive event and all the diverse elements of the situation(s), institutions(s) and social structure(s) which frame it” (Fairclough et al., 2011, p. 357). The use of corpus tools to analyze large collections of electronically encoded data, on the other hand, provides researchers with analytical potentials, lends greater reliability and generalizations, and practically disproves the “cherry-pick” criticism directed at CDA researchers (Mautner, 2009, p. 34; Widdowson, 2004, p. 102). It also reveals the existence of linguistic patterns and messages which could otherwise go unnoticed (Jaworska & Krishnamurthy, 2012, p. 405).

The combination of CL and CDA has slowly but steadily developed (Mautner, 1995; Stubbs, 1995, 2006; Partington, 2004, 2009; Baker, 2006) into a methodological synergy (Baker et al., 2008), i.e., a synergy that systematically enables the pairing of CL and CDA for the purpose of the objective extraction of collocations depending on collocational statistics. CL has increasingly been used in critical discourse studies to explore the discourse of racism (Krishnamurthy, 1996; Baker et al., 2013; Okeke, 2025), refugees (Baker & McEnery, 2005; Gabrielatos & Baker, 2008; Baker et al., 2008; O'Regan & Riordan, 2018; Griebel & Vollmann, 2019; Cardona, 2022), climate change (Grundmann & Krishnamurthy, 2010), gender issues (Hunt, 2011; Ibrahim, 2022; Moustafa, 2022; Yangxia & Mengdan, 2025), public health (M. Liu et al., 2022; Li et al., 2023), political crises (Xu, 2025) and media discourse (Nguyen, 2023).

2.3. Collocations

The term “collocation” was first introduced by Palmer (1938/1968) in his dictionary *A Grammar of English Words*. The term was then developed and established by Firth (1957, p. 194) when he introduced his “modes of meaning,” particularly meaning by collocation. According to Leech (1974, p. 20), “collocative meaning” comprises “the associations a word acquires on account of the meanings of words which tend to occur in its environment.” Sinclair (1991, p. 170) defines collocation as “the occurrence of two or more words within a short space of each other in a text.” Building on Sinclair’s (1966, 1991) corpus-based collocational framework, Stubbs (2001, p. 29) statistically defines collocation as “frequent co-occurrence.” In this sense, a corpus software package or online interface can verifiably be used to determine “empirically which pairs of words have a statistically significant amount of ‘glue’ between them, and which are hence likely to constitute significant collocations [...]” (McEnery & Wilson, 2001, p. 86). Collocational analysis can be conducted by examining concordance lines, since the mere identification of collocates is far from being sufficient, and hence, the immediate context or co-text of the collocates is to be electronically extracted via a concordancing tool. Collocational analysis has the potential to reveal the hidden messages that implicitly permeate a text (Nattinger & DeCarrico, 1992; Sinclair, 1991; Hunston, 2002). Here, the role of collocation in determining semantic prosody, and ultimately revealing ideological representations, becomes evident.

2.4. Semantic Prosody

Semantic prosody was first introduced and defined by Louw (1993, p. 157) as a “consistent aura of meaning with which a form is imbued by its collocates.” Thus, it primarily serves to denote positive or negative extra shades of meaning which a lexeme acquires through its collocates. Sinclair (1996) discusses semantic prosody within his lexico-syntactic model of extended lexical units of meaning, which includes collocation (lexical choices), colligation (grammatical choices), semantic preference (the relation of patterns to particular semantic fields) and semantic prosody, which functions attitudinally and pragmatically. Its role is vital to the extended unit since the speaker’s rationale for producing the utterance is fundamentally dependent on that pragmatic function. As argued by Bublitz (1996, p. 9), “[w]ords can have a specific halo or profile, which may be positive, pleasant and good, or else negative, unpleasant and bad.” According to Gavioli (2005, p. 46), semantic prosody is “the way in which words and expressions create an aura of meaning capable of affecting words around them,” whereas Berber-Sardinha (2000, p. 93) views it as “the connotation conveyed by the regular co-occurrence of lexical items.” Partington (2004, pp. 131–132) defines semantic prosody as a sort of evaluative meaning that is “spread over a unit of language which potentially goes well beyond the single orthographic word and is much less evident to the naked eye.”

Since semantic prosody pertains to a word’s evaluative or attitudinal meaning (Bednarek, 2008; Begagić, 2013; Mahlberg, 2013; W. Liu, 2020; Ben Ghazlen, 2022), semantic prosody does not depend on a word’s frequency of repetition but on the collocational pattern it has. Hence, the revealing of semantic prosody is heavily dependent on collocation. The collocates of a node word contribute to its meaning since they “convey messages implicitly” (Nattinger & DeCarrico, 1992; Sinclair, 1991; Hunston, 2002). The attitudinal quality of semantic prosody, as well as the idea that words take on meanings, whether positive, negative or neutral, from other words in their vicinity, has been repeatedly emphasized (Sinclair, 1996, p. 87; Stubbs, 2001, p. 65; Hunston & Thompson, 1999, p. 38; Hunston, 2001, p. 142). Malrieu (1999, p. 2) sketches the relation between semantic features and ideological concepts, arguing that semantic effects are fundamentally rooted in ideologies—the most comprehensive context. While ideology is the broadest conceivable

form of evaluative context, it infiltrates all discourse types, albeit in diverse ways; it is most commonly observed within discourses of smaller discursive units, such as sentences, clauses and phrases, where it operates through lexical resonance (the evaluative loading of individual words), syntactic and compositional choices (e.g., nominalization, passivization and modality, which constructs agentivity and responsibility) and surface-level control, whereby connectors and modifiers continually recalibrate local evaluation to sustain overall ideological coherence (p. 4). He further contends that ideologies serve as semantic contexts inasmuch they are mental representations (p. 7).

The relation between semantic prosody and CL is a close one. The majority of research on semantic prosody draws on corpus tools, largely because corpus-based methods, particularly identifying collocations and using concordances, offer a systematic and empirically robust means of detecting it (McEnery et al., 2006, p. 85). This inseparable nature of CL and semantic prosody is reflected even in the way it has been regularly defined. Baker et al. (2006, p. 58) relate semantic prosody to “the way that words in a corpus can collocate with a related set of words or phrases, often revealing (hidden) attitudes.” Similarly, Coffin et al. (2004, p. xxi) define it as “The way in which apparently neutral terms come to carry positive or negative associations through regularly occurring in particular collocations.” In this sense, it has consistently been argued that revealing semantic prosodies can only be enabled computationally through the use of corpus tools and techniques (Bublitz, 1996, p. 9; Louw, 1993, p. 159; Hunston, 2002, p. 142; Tognini-Bonelli, 2004, p. 20; Berber-Sardinha, 2000, p. 93). Previous research on semantic prosody in relation to a specific genre has focused on newspapers (O'Donnell, 2023), online news (Hardiman & Nuraniwati, 2024), novels (Kenny, 2001), business English (Su & Phoocharoensil, 2021), academic research papers (Selmistraitis, 2020), Qur'anic text (Al-Otaibi, 2025), and education (Al-Otaibi, 2022).

2.5. Arabic Language Corpus-Based Studies

In spite of the advent of Arabic language corpus-based studies (Zaghouani, 2014; Alansary & Nagi, 2014; Elghamry, 2025), there is a scarcity of research which uses corpus tools for the investigation of Arabic discourse, particularly concerning Saudi labor market policy. For example, Hussain (2019) used the word sketch tool in the Arabic corpus of Sketch Engine (arTenTen12) to extract the collocational patterns of prepositional phrase (PP) disambiguation and derive collocational rules for distributional patterns. Smadi (2020) provided a quantitative analysis of the abstracts of Arabic journal articles written in the social sciences, exploring their lexical and grammatical features using WordSmith tools on a corpus compiled of 500,000 words and divided according to their respective academic discipline. Alanazi (2020) conducted a corpus-driven critical discourse study on how Saudis are represented on English-language Twitter, specially during the 2015 Mecca crane collapse tragedy. Using a Twitter corpus of 1.9 million tokens, keywords were extracted and concordance lines were examined for the purposes of CDA qualitative analysis; particularly, the Discourse-Historical approach was conducted. Findings indicated that Saudis were consistently negatively represented as associated with war, terrorism, corruption, and incompetence. The recurring negative evaluative patterns pervasive in tweets suggested a negative semantic prosody of Saudis. Moreover, through their lexical choices, the tweets' authors constructed an ideological framing that reinforced Orientalist and Islamophobic tropes.

Qarah (2024) introduced SaudiBERT, an Arabic language model of two Saudi dialect corpora: Saudi Tweets Mega Corpus (social media data) and Saudi Forums Corpus (online discussion forums). The model provides the capability of automated annotation and represents state-of-the-art NLP model development for Saudi contexts. Most pertinently, research has emphasized that Arabic multiword patterns and support-verb constructions

shape meaning in a way that is difficult to capture by simple two-word collocation (Abdulrahim, 2014; Cantos & Almela-Sánchez, 2021). Thus, rich Arabic syntactic and morphological constructions require the extraction of collocates using lemmatization—in order to capture all instances—and the analysis of discontinuous collocations, separated by intervening words (see Section 4.1).

3. Methodology

The present research aims to sketch the collocational profile of the term *سعودة* (Saudization), as a step towards working out its semantic prosodies and the associated ideological representations in an MSA corpus. The study adopts a mixed-methods approach, i.e., the qualitative dimension, enabled through the use of a CDA model, is reflected by the close reading of concordance lines in order to reveal nuanced meanings, while the quantitative dimension, allowed through the use of corpus tools, ensures that the data examined is representative, reduces researcher's bias and guarantees generalizations. In this sense, CDA offers a powerful framework which helps to unfold ideological perspectives in large scale textual data (Baker, 2006; Baker et al., 2008; Partington et al., 2013).

3.1. Data

Investigations of semantic prosody commonly use large general corpora with no focus on particular text types for analysis (Stewart, 2010, p. 17). Since the language variety explored in the present study is MSA, the corpus selected for the examination of *سعودة* (Saudization) is the Arabic corpus integrated within the online interface of SkE. Lemmatized and tagged, the Arabic Web Corpus, or arTenTen24, is collected from varied Arabic web domains in myriads of fields and disciplines and from all Arab countries. The reasons for particularly selecting SkE for the purposes of the current inquiry are as follows: (1) SkE provides access to a huge Arabic corpus in different genres and dialects, comprising more than 5 billion tokens and 4.6 billion words and offering an enormously rich data source. The fact that it is collected from web domains related to all Arab countries and various dialects and genres ensures that the results are representative and generalizable (Mautner, 2009, p. 34; Widdowson, 2004, p. 102). (2) It also offers various tools for advanced linguistic search and analysis, and is, thus, perfectly suitable for retrieving the frequency lists, collocates and concordances needed in the current study. (3) arTenTen24 was last updated in 2024 and, hence, represents the most recent and up-to-date source of data for MSA. (4) It is integrated within the interface of SkE, and thus, saves time and effort.

3.2. Procedure

The procedure adopted in the current investigation follows three steps: collocate extraction, concordance-based analysis, and semantic prosody and ideological representations identification. Since the identification of semantic prosody is based on the lexical co-occurrence within a quite restricted span of 5–6 words to the left and right of the node word (Stewart, 2010, p. 15), the procedure starts with the computational identification of collocates of the node word *سعودة* (Saudization) using SkE corpus methods (Hunston, 2002; McEnery et al., 2006; Baker et al., 2008). The indefinite form *سعودة* (Saudization) is used rather than the definite form *السعودة* (the Saudization) so that the resulting concordances would include all forms of the node word.

Since considering a word's co(n)text is an essential part of understanding and interpreting it (Figureiredo, 2004, p. 218), collocates are examined in their concordance lines to work out the associated semantic prosodies. Whenever more detailed information is needed for better identification of the relevant prosody, reference is made to the broader context of concordance lines, e.g., situational, social, political, economic, etc. The expanded

concordances provide the co-text necessary for “limiting the interpretation” based on the appropriateness and plausibility of contextual elements, i.e., scrutinizing the contextual meaning of a node’s collocates can offer “a semantic analysis of a word” (Sinclair, 1991, pp. 115–116). A semantic prosody can be worked out by considering semantic connections, both paradigmatically, i.e., through various concordance lines, and syntagmatically, i.e., with regard to the relationships between the different elements of individual lines (Stewart, 2010, p. 100). Collocational and concordance investigations offer an attitudinal and evaluative sketch of how سعودة (Saudization) acquires new shades of meaning through its collocates. Based on the semantic prosodies identified, ideological representations of how Saudization is portrayed in MSA discourse are presented (Fischer-Starcke, 2010; Louw, 1993; Baker, 2014).

4. Analysis and Findings

In the following two subsections, the suggested methodological procedure is applied to a corpus-based investigation of سعودة (Saudization) with respect to (1) generating a concordance list of سعودة (Saudization) via SkE tools, (2) calculating its collocates, and (3) identifying semantic prosodies of the extracted collocates and their ideological representations.

4.1. Generating Concordances and Identifying the Collocates of سعودة (Saudization)

The noun سعودة (Saudization) appears in the corpus 17,471 times, including all its grammatical forms, i.e., with and without the definite article (السعودية, السعودية), in singular, dual and plural forms (سعودتين, سعودات), in different cases such as prepositional prefixes (بالسعودية, للسعودية), pronominal suffixes (سعودتها), etc. Using the SkE concordance tool (Kilgariff et al., 2004, 2014), concordances of the lemma سعودة (Saudization) were generated. Then, the SkE collocation tool was used to produce the collocates of سعودة (Saudization), setting the co-text collocational span to $5 \pm$ (Baker, 2006, 2014) and the collocate word type to “lemma” so that all forms of each collocate would be included under the same entry. Based on frequency of occurrence and statistical measures of association, e.g., T-score, MI and LogDice, the collocates are arranged starting with the most salient lexical association and the most stable occurrence patterns to the least. For the sake of analytical depth and interpretability, the top 20 resultant collocates were considered (Table 1). That is, using a fixed number of collocates is methodologically consistent with established corpus-linguistic practice (Baker, 2006; Baker et al., 2013, etc.), and in alignment with the research goal, i.e., identifying the dominant semantic prosodies of the node word سعودة (Saudization).

Table 1 represents the collocates of سعودة (Saudization), as they appear in SkE, ordered from the most to the least statistically significant. The first column indicates the ranked number of each collocate; the second column lists the collocates; the third column contains the English translation of the collocate; the fourth column refers to the frequency of the collocate’s co-occurrence with the node word within the specified co-text span; the fifth column shows the overall number of times the collocate occurred in the corpus; the sixth, seventh and eighth columns represent the values of T-score, MI and LogDice, respectively. Initially, it should be noted that in وتغريمها (And Fining Them), the pronominal suffix -hā (them/it) in Arabic regularly agrees with non-human plural antecedents in the feminine-singular form; in the concordance lines examined, it consistently refers to companies and establishments. As for the collocate استقدام, it literally means “importation”; however, in Saudi administrative and labor-law discourse, it is used specifically to denote the importation of foreign labor; hence, it is glossed accordingly throughout the current study as “importation of foreign labor.” It is also noted that the collocates list reveals a rich network of lexical items relating to labor policies, administrative mechanisms, and labor market regulations and restrictions, and hence, reflects the multifaceted discourse of Saudization in the

investigated corpus. Close sketching of collocates within their linguistic milieu suggests that collocates, such as نطاقات (Nitaqat), نسب (percentages), توطین (localization), إحلال (replacement), قصيبي (Qusaibi) and احتساب (calculation) obviously contribute to the semantic field of institutional governance and administrative regulations, indicating that Saudization is frequently discussed in relation to administrative mechanisms regulating labor market participation. The list also includes a number of lexical items which clearly convey evaluative and critical meanings. Collocates such as وهمي (phantom), تحايل (circumvention), بطالة (unemployment), وتغريمها (and fining them) and تسنر (concealment) reflect consequences related to the policy. Another semantic field preliminarily conveyed through the collocates list is the structure of labor market and employment practices. Collocates such as وظيفة (job), توظيف (employment), عمالة (labor), تأنيث (feminization) and استقدام (importation of foreign labor) bring to the foreground the socio-economic conditions under which the labor policies operate. Reference to certain economic activities, e.g., الليموزين (limousine), خضار (vegetables) and بقالة (grocery shop) suggests that Saudization is frequently discussed in relation to particular labor sectors. It should be noted that the lemma form is cited here per standard corpus convention, though the collocate occurs overwhelmingly, in 38 of 41 concordance lines, in its plural form, البقالات (the grocery shops), consistent with the discourse's treatment of grocery retail as a sector rather than an individual establishment. A preliminary overview of collocates is insufficient to capture the full semantic prosodies of Saudization; therefore, a detailed examination of each collocate in its immediate co-textual environment is required.

Table 1. Top twenty collocates of the node word سعودة (Saudization), their English translations and statistical information.

No.	Collocate	English Translation	Concordances	Candidates	T-Score	MI	LogDice
1	سعودة	Saudization	380	13,005	19.49	13.63	8.67
2	نطاقات	Nitaqat	243	14,547	15.59	12.83	7.96
3	توطین	Localization	848	137,208	29.11	11.39	7.49
4	وهمي	Phantom	898	147,482	29.96	11.37	7.48
5	الليموزين	Limousine	65	3322	8.06	13.05	6.68
6	إحلال	Replacement	166	93,708	12.87	9.59	5.61
7	وظيفة	Job	2005	1,446,010	44.70	9.23	5.49
8	خضار	Vegetables	153	107,305	12.35	9.28	5.33
9	قصيبي	Qusaibi (Proper Noun)	38	14,133	6.16	10.19	5.30
10	تأنيث	Feminization	47	21,759	6.85	9.87	5.29
11	تحايل	Circumvention	66	40,993	8.11	9.45	5.21
12	نسب	Percentages	696	635,050	26.33	8.90	5.13
13	استقدام	Importation of Foreign Labor	108	85,696	10.37	9.10	5.10
14	وتغريمها	And Fining Them (Companies)	19	972	4.36	13.05	5.08
15	توظيف	Employment	477	517,706	21.79	8.65	4.87
16	احتساب	Calculation	127	130,521	11.24	8.72	4.81
17	عمالة	Labor	324	380,237	17.95	8.53	4.74
18	بطالة	Unemployment	256	305,282	15.96	8.51	4.70
19	تسنر	Concealment	75	81,125	8.64	8.65	4.64
20	بقالة	Grocery Shop	41	37,184	6.39	8.90	4.62

4.2. Semantic Prosodies of سعودة (Saudization) and Their Ideological Representations

The concept of representativeness, rather than that of exhaustiveness, has been repeatedly asserted in corpus-assisted discourse research, e.g., Hunston (2002), Sinclair (2003), Louw (1993), Baker (2006). Thus, sampling is a necessary and methodologically justified process, ensuring that the number of concordances to be closely examined is manageable and allowing for a more thorough and detailed analysis. In this sense, a random sample of

50 concordances for each collocate is generated via the SkE sampling tool while all the concordance lines for the collocates with a co-occurrence frequency of ≤ 50 are considered. The particular percentage of each collocation instance contributing towards different semantic prosodies, whether positive, neutral or negative, in relation to all instances of a particular collocate is provided, along with the raw number of occurrences followed by the abbreviation CL, which refers to “concordance line.” For collocates whose total occurrence number is ≤ 50 , the fractions of percentages are approximated.

The first group of collocates includes the top statistically significant collocate of *سعودة* (Saudization), which is the same lexical item since the node word, i.e., *سعودة* (Saudization), collocates with itself, reflecting a metadiscursive pattern where the term is repeatedly used to define, evaluate and problematize the policy of Saudization. Close examination of concordance lines reveals that in 26% of instances (13/50 CL; that is, 13 concordance lines out of 50), a positive semantic prosody emerges, i.e., in contexts emphasizing the success of Saudization and national development, as in example [1], which refers to the high percentage of Saudization and the effectiveness of Saudi youth. Hence, the repetition of *سعودة* (Saudization) constructs a discourse of achievement, competence and progress.

[1] وعن السعودة، قال: نسبة السعودة وصلت إلى 75 بالمائة وأن الشباب السعودي أثبتوا جدارتهم [...] [1]

- (1) Regarding **Saudization**, he stated that the rate of **Saudization** has reached 75 percent, and that Saudi youth have proven their competence.

A neutral prosody is evident in 26% of concordances (13/50 CL), where *سعودة* (Saudization) is used to explain the concept itself. In example [2], the term functions as a technical definition with no explicit evaluation.

[2] السعودة هي حركة تسعى إلى زيادة عدد المواطنين العاملين داخل البلاد، حيث يُطالب أرباب العمل بالوفاء بحصص العمال و السعودة.

- (2) **Saudization** is a movement that seeks to increase the number of citizens working within the country, whereby employers are required to fulfill labor and **Saudization** quotas.

In 48% of cases (24/50 CL), a negative prosody is conveyed by referring to Saudization as inefficient and highlighting cases of fake compliance. Example [3] represents Saudization as an obligatory practice and thus equates it to phantom Saudization, an object of manipulation, creating a discourse of ineffectiveness and delegitimization.

[3] منتقدًا أنظمة وزارة العمل وتخططاتها ووضعها برامج السعودة الإلزامية التي تحولت إلى سعودة وهمية [...] [3]

- (3) Criticizing the Ministry of Labor’s regulations and their inconsistencies, as well as the implementation of mandatory **Saudization** programs that have devolved into phantom **Saudization**.

Taken together, the semantic prosody of *سعودة* (Saudization) is strongly evaluative and polarized, fluctuating between national pride and development on the one hand, and critical judgement on the other, thus reflecting its contested status in public and policy discourse. Within the same semantic domain, the concordance lines of the collocate *نطاقات* correspond to Saudi national labor policy, its regulatory mechanisms, and the consequences of Nitaqat policy. Nitaqat is the Saudi labor nationalization program, administered by the Ministry of Human Resources and Social Development, which classifies private-sector establishment into color-coded compliance bands, ranging from red (denoting the lowest proportion of Saudi employees) to platinum (the highest), according to each firm’s percentage of Saudi workers relative to quotas that vary by company size and industry sector (Peck, 2017). Establishments placed in the lower bands face restrictions on the recruitment of foreign labor, whereas those in the higher bands receive preferential treatment, such as expedited visa processing. A neutral prosody appears in 64% of concordance lines (32/50 CL), where *نطاقات* (Nitaqat) is used descriptively to highlight the mechanisms of Saudization and its regulatory frameworks which relate to quotas, classifications, and compliance.

[4] يكون احتساب نسب توطین الوظائف (السعودة) ضمن برنامج نطاقات للمنشآت [...]

- (4) The calculation of job localization rates (**Saudization**) within the **Nitaqat** program for establishments is determined as follows.

Here, نطاقات is used as a professional technical term, presenting Saudization as a structured and measurable policy tool without evaluation. A total of 34% of concordances (17/50 CL) convey a negative semantic prosody by associating نطاقات with circumvention, exploitation and fake employment, as in examples [5], [6] and [7]. In [5], the very existence of the Nitaqat program is threatened by both the unsustainability of Nitaqat employment and fake employment. In [6], the program is framed as susceptible to manipulation since organizations exploit its regulatory criteria to gain benefits without real compliance, while in [7], broad socio-economic critical impacts of Nitaqat are also implicitly negatively referred to, particularly the rising costs and burdens on both individuals and sectors. Only in 2% of cases (4/50 CL) is the collocate positively represented by stressing efforts of supporting the program and the resultant achievements, as in example [8].

[5] عدم وجود خاصية الاستمرارية في التوظيف الناتج عن نطاقات، وما يُسمى بالسعودة الوهمية، قد يهدد بانهياره.

- (5) The lack of sustainability in employment resulting from **Nitaqat** and what is termed **phantom Saudization** may threaten its collapse.

[6] واستغلال القطاع الخاص لهم في التوظيف الوهمي، كأفراد معوقين يقابل الواحد منهم أربعة مواطنين، ومن أجل الحصول على نطاقات السعودة ورفع فئة تصنيف المنشأة.

- (6) The exploitation by the private sector of individuals in hiring, such as persons with disabilities, where one individual is counted as four citizens, to obtain the **Nitaqat** of **Saudization** status and elevate the establishment's classification category.

[7] يجب استثناء شركات المقاولات من نطاقات السعودة اسعار المقاولات ارتفعت 100% بسبب قرارات وزارة العمل.

- (7) Contracting companies must be exempted from the **Nitaqat** of **Saudization** requirements, [as] construction costs have risen by 100% due to the Ministry of Labor's decisions.

[8] أدخلت وزارة العمل 5 قطاعات مهمة ضمن برنامج "نطاقات" بهدف زيادة نسبة السعودة في هذه القطاعات، خصوصاً أن هناك تخصصات بدأ تزايد السعوديين للعمل فيها مثل التعليم والصيدلة وتقنية المعلومات.

- (8) The Ministry of Labor has integrated five critical sectors into the "**Nitaqat**" program to increase the **Saudization** rate, particularly given the increasing interest of Saudis to work in fields such as education, pharmacy and information technology.

Overall, the semantic prosody of نطاقات (Nitaqat) as a collocate of سعودة (Saudization) is mixed but predominantly negative. While it is neutrally used to reflect the formal regulatory role of Nitaqat, its recurring association with manipulation, economic tensions and fake compliance constructs a discourse of skepticism. Close examination of the collocate توطین (localization) in its concordance lines reveals its recurrent presence in the vicinity of سعودة (Saudization) in coordinated or contrastive constructions. Co-textual evidence proves its relevance to both institutional framework and critical evaluation, i.e., in 50% of concordances (25/50 CL), a positive semantic prosody is evident, as the co-occurrence of توطین (localization) and سعودة (Saudization) is used to emphasize the reinforcement, planning and development of the national workforce, as in example [9]:

[9] لتنفيذ البرنامج بنجاح بما يحقق الأهداف المرجوة منه وتعزيز جهود السعودة وتوطین الوظائف في قطاعات العمل المختلفة.

- (9) To implement the program successfully in a way that fulfills its intended objectives and to bolster efforts toward **Saudization** and the **localization of jobs** across various employment sectors.

Here, *توطين* (localization) is associated with the enhancement of Saudization to achieve success in the targeted goal of nationalization. The collocate, in this sense, highlights national empowerment, indexes the construction of national identity and reflects a broader ideology of state-led labor market reform. A less prevalent neutral prosody, 30% of cases (15/50 CL), emerges in contexts where *توطين* (localization) is used descriptively to refer to regulatory procedures without perspicuous evaluation, as in example [10], which correlates companies' performance with the percentages of localization and Saudization:

[10] في نطاق معرفة نسبة أداء الشركات بما يتعلق **بتوطين سعودة** الوظائف في ظل تفعيل قياس أداء الموارد البشرية ببرنامج (نطاقات).

- (10) Within the scope of determining corporate performance regarding the **localization (Saudization)** of jobs, in light of the activation of human resource performance metrics under the “Nitaqat” program.

A negative semantic prosody is found in 20% of concordances (10/50 CL), where *توطين* (localization) is used to criticize Saudization processes and practices. The collocate in example [11] conveys a discourse of limited effectiveness, suggesting skepticism towards localization outcomes, success and sustainability as to Saudization and even Emiratization. The semantic prosody of *توطين* (localization), in this sense, is a mixture of positive and neutral prosodies, as it combines both institutional legitimacy and labor tensions.

[11] ولا تلقى برامج **توطين** العمالة (السعودية. الأمر، إلخ) إلا نجاحاً محدوداً.

- (11) Workforce **localization** programs (**Saudization**, Emiratization, etc.) have met with only limited success.

The close examination of the concordance lines of *إحلال* (replacement) reveals its central role as a conceptual component of the discourse around *سعودة* (Saudization), frequently co-occurring in definitional, policy-making and evaluative contexts since it directly relates to the essential logic behind Saudization, namely, the replacement of expatriate labor with national workers. A positive semantic prosody is evident in 34% of cases (17/50 CL), where *إحلال* (replacement) is represented as a desirable national goal linked to employment opportunities and economic development. In example [12], *إحلال* (replacement) is embedded in a discourse of achievement and policy success, reinforcing the legitimacy of Saudization as a national project and a strategy for strengthening national workforce and national empowerment.

[12] بهدف رفع تنافسية العمالة الوطنية ودفعاً لسياسات وخطط توظيف أبناء المنطقة وتطبيقاً لبرامج توطين الوظائف وتحقيق **الإحلال والسعودة** بمنشآت القطاع الخاص.

- (12) Aimed at increasing the competitiveness of the national workforce, promoting policies and plans for the employment of local citizens, and applying programs for job nationalization, **replacement** and **Saudization** in private sector establishments.

Here, replacement and Saudization are represented as goals aiming at the national workforce and supporting employment policies and localization programs. A neutral semantic prosody appears in 46% of concordances (23/50 CL), where *إحلال* (replacement) is used to define the concept of Saudization in technical or explanatory terms. In example [13], the collocate functions descriptively, providing a formal definition of the concept without evaluative stance.

[13] **السعودة** مصطلح يطلق على عملية **إحلال** المواطنين السعوديين مكان العمالة الوافدة.

- (13) **Saudization** is a term used for the process of **replacing** foreign workers with Saudi citizens.

A negative semantic prosody emerges in 20% of cases (10/50 CL), where *إحلال* (replacement) is associated with unintended consequences and implementation challenges. Example [14] highlights concerns about inadequate preparation and reduced efficiency as a result of the sudden unplanned replacement of 100% of nurses, representing replacement

as potentially problematic. The semantic prosody of إحلال (replacement), hence, is mostly a mix of positive and neutral, focusing on replacement as a national duty and providing a formal definition of Saudization.

[14] المستشفى الحكومي [...] احتفل بإحلال السعودة مكان الاجانب 100 %، إلا أن هذا الاحلال المفاجئ سبب قصورا في اداء الممرضات والممرضين السعوديين.

- (14) the governmental hospital [...] celebrated the 100% **replacement** of expatriates through **Saudization**, although this abrupt transition resulted in a deficiency in the performance of Saudi nurses.

In the same group of collocates relating to institutional and administrative regulations, the collocate قصيبي (Qusaibi), referring to the former Minister of Labor, Ghazi Al-Qusaibi, functions as a personalized discursive element linking Saudization to political authority and leadership. Unlike all previous technical collocates, قصيبي (Qusaibi) introduces a human agent into discourse, marking a shift towards the evaluation of policy makers and their decisions. A positive semantic prosody can be found in 33% of lines (13/38 CL), where قصيبي (Qusaibi) is associated with leadership and reform initiatives, representing him as an effective figure who raised the visibility of the issues of Saudization and unemployment, as in example [15], where the Minister of Labor is praised for his creative endeavors in popularizing the issue of Saudization, and example [16], where he condemns anti-Saudization voices as unpatriotic. In 37% of cases (14/38 CL), a negative semantic prosody is conveyed as the minister warns Saudi firms and institutions of phantom Saudization, as in example [17]:

[15] أبدع الدكتور غازي القصيبي في جعل مشكلة السعودة والبطالة تتصدر الخارطة الإعلامية السعودية.

- (15) Dr. Ghazi **Al-Qusaibi** excelled in positioning the issues of **Saudization** and unemployment at the forefront of the Saudi media landscape.

[16] وانتقد القصيبي بعض الأصوات المطالبة بإلغاء السعودة [...] وأكد أنها لا تراعي المصلحة الوطنية.

- (16) **Al-Qusaibi** criticized voices calling for the abolition of **Saudization**, asserting that they do not serve the national interest.

[17] حذر وزير العمل الدكتور غازي القصيبي المؤسسات والشركات السعودية من السعودة الوهمية [...]

- (17) Minister of Labor Dr. Ghazi **Al-Qusaibi** warned Saudi institutions and companies against “**phantom Saudization**.”

Concordance lines indicate that the collocate نسب (percentages) is a key element of the regulatory discourse surrounding سعودة (Saudization), i.e., it functions quantitatively to define, monitor and evaluate the Saudization policy. A neutral institutional prosody is highly evident, as it appears in 66% of lines (33/50 CL), where نسب (percentages) is used to describe regulatory requirements and measurement frameworks. In example [18], the collocate functions as part of formal administrative discourse, representing Saudization as a measurable and standardized policy. A positive semantic prosody emerges in 28% of concordances (14/50 CL) which emphasize achievement and progress in increasing national employment. Example [19] highlights the high percentages of Saudization as indicators of policy and development success, and hence represents Saudization as a source of national pride and empowerment.

[18] ويتم احتساب نسب توظيف الوظائف (السعود) ضمن برنامج نطاقات.

- (18) Job nationalization **percentages** (**Saudization**) are calculated within the Nitaqat program.

[19] أشاد وزير العمل [...] بدور الخطوط السعودية في توظيف الوظائف وتحقيق نسب متقدمة في مجال السعودة.

- (19) The Minister of Labor praised the role of Saudi Arabian Airlines in job nationalization and achieving advanced **percentages** in the field of **Saudization**.

A negative prosody appears in 6% of concordances (3/50 CL) which represent cases of manipulative practices or inefficiency in policy implementation, as in example [20], where نسب (percentages) is associated with artificial inflation, suggesting that numerical targets are achieved through malpractices. In sum, نسب (percentages) exhibits a mixed semantic prosody mostly combining neutral administrative measurements with positive evaluations of success and achievement.

[20] شركات كبرى سجلت شبابا عاطلين على أنهم موظفون لديها لرفع نسب السعودة في الشركة.

- (20) Major companies registered unemployed youth as employees to inflate their **Saudization percentages**.

As for the collocate وتغريمها (and fining them), it is recurrently used within regulatory and enforcement discourse in relation to Saudization. Its co-textual environment is highly formulaic, conventionally occurring in reports of violations and sanctions imposed on institutions violating the rules of Saudization. Thus, contributing to a salient evaluative semantic prosody, the collocate represents Saudization within a discourse of monitoring and observation. A neutral routine administrative prosody appears where وتغريمها (and fining them) is used to signal measures taken against institutions committing violations of Saudization requirements and the imposition of financial penalties (example [21]) or as part of a systematic institutional response to malpractice, emphasizing surveillance and accountability (example [22]):

[21] ١١ مخالفة لمؤسسات تأجير لا تلتزم بالسعودة وتغريمها ١١ ألف ريال.

- (21) Eleven violations by rental agencies failing to comply with **Saudization**, and their being **fined** 11,000 riyals.

[22] ملاحقة الشركات التي تلجأ إلى السعودة الوهمية وتغريمها، بل وإيقاف تراخيص عملها.

- (22) Pursuing companies that resort to phantom **Saudization**, **fining them**, and even suspending their work permits.

The collocate احتساب (calculation) is a key procedural term, as it functions as a mechanism for quantifying and regulating compliance with the policy of Saudization. Though its co-textual patterns mainly reflect its institutional nature, they also reveal underlying tensions regarding fairness, representation and manipulation. A neutral prosody is dominant in 76% of instances (38/50 CL), where احتساب (calculation) is used in technical contexts to describe regulatory procedures, as in example [23], which describes the way Saudizing jobs is calculated with the program of Nitaqat. A positive prosody emerges in 16% of concordances (8/50 CL), as the collocate is linked to incentives or supporting mechanisms meant to enhance Saudization policy, as in example [24], which highlights the Ministry of Labor's endeavors in bettering the system to fit the expectations of Saudi citizens. A negative prosody appears in 8% of lines (4/50 CL), where it is associated with policy distortions, as in example [25], which refers to phantom Saudization as the axis of Saudization problems. Broadly, احتساب (calculation) exhibits a mainly neutral semantic prosody, combined with a positive representation of the ministerial role of offering initiatives and opportunities to improve the application of Saudization policy.

[23] يكون احتساب نسب توظيف الوظائف (السعودة) ضمن برنامج نطاقات للمنشآت [...]

- (23) The **calculation** of job nationalization rates (**Saudization**) within the Nitaqat program for establishments is determined as follows.

[24] واستجابت الوزارة مشكورة بأن يوازي توظيف المواطن المعاق بأربعة مواطنين عند احتساب نسبة السعودة.

- (24) The Ministry graciously responded by equating the employment of one disabled citizen to four citizens when calculating the **Saudization** rate.

[25] المشكلة تتمحور في [...] احتساب السعودة الوهمية برفع عدد المشتغلين وبالتالي خفض نسبة البطالة.

- (25) The problem centers on the **calculation** of phantom **Saudization** by artificially increase employment figures, thereby reducing the unemployment rate.

The collocate وهمي (phantom) is the most statistically significant collocate in the second group, in which each collocate suggests evaluative and critical meaning per se. Concordance lines show that وهمي (phantom) forms a highly conventionalized collocation with سعودة (Saudization), most frequently in the expression السعودية الوهمية (phantom Saudization). This collocational pairing saliently pinpoints a widespread criticism of the implementation of Saudization since it is almost exclusively linked to fake compliance with localization quotas that companies superficially meet without providing genuine employment for Saudis, thus undermining the intended goals of labor nationalization policies. In 96% of concordances (48/50 CL) where وهمي (phantom) collocates with سعودة (Saudization), negative semantic prosodies dominate, as exemplified in [26], [27] and [28]. In [26], وهمي (phantom) is used to represent the companies' fraudulent practices of fake compliance with Saudization quotas, whereas in [27], the use of the prepositional phrase على القضاء (eliminating) positions phantom Saudization as a persistent systematic problem requiring intervention. In [28], another negative semantic prosody emerges by linking وهمي (phantom) to broader consequences such as disguised unemployment and economic inefficacy. In 4% (2/50 CL) of concordances, positive prosody emerges; that is, when state-wide official endeavors are exerted to control the phenomenon of phantom Saudization, as exemplified in [29]. On the whole, the collocate وهمي generates a consistently negative semantic prosody where Saudization is constructed as susceptible to manipulation and fake compliance.

[26] ولكن ريثما تحل المشكلة الاولى السعودية الوهمية واستغلال اسمك وراتبك

- (26) however, until the primary problem is resolved, **phantom Saudization** and the exploitation of your name and salary

[27] العمل تصدر قرارا للقضاء على أشكال السعودية الوهمية للمعاقين.

- (27) The Ministry of Labor has issued a decision to eradicate forms of **phantom Saudization** involving persons with disabilities.

[28] لكون التوظيف الوهمي لمجرد تحقيق معدلات السعودية بطلاة مقنعة

- (28) because **phantom** employment, practiced merely to meet **Saudization** quotas, constitutes disguised unemployment

[29] متوقعا أن يسهم برنامج حماية الأجور الذي جرى العمل به منذ عدة أشهر في الحد من السعودية الوهمية

- (29) expecting that the Wage Protection Program, operational for several months, will contribute to curbing **phantom Saudization**

As for the collocate تحايل (circumvention), it strongly and systematically occurs with سعودة in different variations of the construction التحايل على السعودة (circumventing of Saudization), forming a highly evaluative and critical semantic prosody. Concordance lines entirely position Saudization as susceptible to manipulation and avoidance. Thus, تحايل (circumvention) introduces a clear discourse of illegitimacy and systematic abuse, as in the following two examples:

[30] معتبراً ذلك تحايلاً على نظام السعودة [...]، كل ذلك من أجل الحصول على "نطاقات السعودة"

- (30) Describing this as a **circumvention** of the **Saudization** regulations, all to obtain the Nitaqat of Saudization status

[31] لجوء بعض المؤسسات الاقتصادية للتحايل على السعودة إما بالتوظيف الوهمي أو تدوير الوظائف

- (31) the resort of some economic institutions to **circumventing Saudization**, either through phantom employment or job rotation

Example [30] represents Saudization as having fraudulent practices, highlighting intentional manipulation of the Saudization system to gain the benefits of the Nitaqat system. From a socio-economic perspective, in example [31], circumvention is linked to broader labor market malpractices, as manipulations are manifested in the form of fake employment practices and distortions in labor statistics.

The collocates *بطالة* (unemployment) is a central evaluative term in the discourse of Saudization, used in relation to Saudization either as a solution to or a reflection of labor market challenges. A positive prosody appears in 26% of concordances (13/50 CL), where *بطالة* (unemployment) is represented as a socio-economic issue that Saudization seeks to solve. Example [32] discusses the initiatives taken by the Saudi government to support Saudization and reduce unemployment. A negative prosody emerges in 24% of lines (12/50 CL) when the implementation of Saudization as a solution to unemployment is represented as a challenge. Example [33] portrays enhancing Saudization and reducing unemployment as a difficult task:

[32] الخطوة الايجابية التي من شأنها دعم برامج السعودية وتوطين الوظائف والحد من البطالة

- (32) a positive step that would support **Saudization** programs, the nationalization of jobs, and the reduction of **unemployment**

[33] وزارة العمل تواجه تحديا صعبا في مواجهة القطاع الخاص من حيث رفع نسبة السعودية وخفض نسبة البطالة في المجتمع.

- (33) The Ministry of Labor faces a difficult challenge in addressing [the role of] the private sector in raising the **Saudization** rate and lowering the **unemployment** rate in society.

The collocates *تستر* (concealment) is a highly evaluative term in Saudization discourse, i.e., it frequently occurs in contexts where labor market breaches are exposed. The collocates is systematically associated with illegitimate economic practices, thus representing Saudization in the context of regulatory challenges and exploitation. A total of 68% of concordances (34/50 CL) relate to two negative semantic prosodies conveyed through the collocates *تستر* (concealment): (1) a delegitimizing prosody where Saudization is portrayed as contested by the illegal practices of fake compliance, and (2) a causal prosody where concealment is represented as a key factor hindering the success of Saudization, as in [34] and [35], respectively:

[34] رأينا العديد من النقاط التي تم إساءة استغلالها من قبل البعض في سوق العمل مثل التستر التجاري والسعودة الوهمية.

- (34) We have observed several points that have been exploited by some in the labor market, such as commercial **concealment** and phantom **Saudization**.

[35] الصعوبات التي تعيق السعودية في أسواق الخضار وغيرها التستر التجاري، الذي هو أساس كل بلاء، والعقبة الأكبر في فشل التوطين

- (35) The difficulties hindering **Saudization** in vegetable markets and elsewhere are commercial **concealment**, which is the root of all problems and the greatest obstacle in the failure of nationalization.

A positive prosody appears in 32% of concordances (16/50 CL), where the efforts of the Saudi government in regulating the procedures of implementing Saudization and eliminating commercial concealment are highlighted. Thus, in examples [36] and [37], representing concealment as a violation requiring punishment and control, an institutional enforcement prosody which emphasizes regulatory intervention and national endeavors is at work.

[36] تعمل منذ نحو شهر على حملات متتالية لتطبيق نظام السعودية والقضاء على التستر التجاري.

- (36) [The Ministry] has been working for about a month on successive campaigns to enforce the **Saudization** regulations and eliminate commercial **concealment**.

[37] [...] حتى نقضي على التستر ومن يحارب السعودية

- (37) [...] until we eradicate commercial **concealment** and those who fight against **Saudization**

The third group of collocates, relating to the labor market structure, socio-economic conditions and employment practices, starts with وظيفة (job) as the most statistically significant collocate in the category. Concordance lines of وظيفة (job) indicate that the collocate is crucial to the discursive construction of سعودة (Saudization), since it functions as the primary object of Saudization policies. The co-textual patterns relate Saudization to job creation, regulation and labor market restructuring. The semantic prosody is largely shaped by tensions between what Saudization aspires for and its practical outcomes. In 48% of lines (24/50 CL) where وظيفة (job) collocates with سعودة (Saudization), neutral semantic prosody appears, as وظيفة (job) is used with administrative or policy description, functioning as a technical lexeme, as in example [38], where the mechanisms of calculating percentages of women in the workforce are discussed. In 38% of lines (19/50 CL), a positive prosody emerges, as وظيفة (job) is associated with creating jobs and building national capacity, as in example [39], which celebrates the national accomplishment of localizing jobs in teaching. A negative critical prosody is also salient in 14% of concordances (7/50 CL) which highlight the malpractice of phantom Saudization, as in example [40], where وظيفة (job) is linked to fraudulent practices, although national efforts are exerted to correct the illegitimate situation of pseudo compliance. Such context serves to mitigate the negativity inherent in the collocation of وظيفة (job) and سعودة (Saudization).

[38] بشأن آلية احتساب عمل المرأة عن بعد في نسب توظيف الوظائف (السعودة) وضرورة تحديثه

- (38) regarding the mechanism for calculating remote work for women within **job** nationalization rates (**Saudization**) and the necessity of updating it

[39] تمت سعودة ٤٠ ألف وظيفة معلم ومعلمة.

- (39) Forty thousand teaching **jobs** (male and female) have undergone **Saudization**.

[40] غالبية الحالات التي صححت وضعها كانت تعمل في القطاع الخاص بوظائف وهمية (سعودة وهمية).

- (40) The majority of cases that rectified their status were working in the private sector in phantom **jobs** (phantom **Saudization**).

The examination of concordance lines shows that the collocate تأنيث (feminization) is systematically linked to سعودة (Saudization) in coordinated constructions, e.g., والتأنيث السعودة, raising the issue of labor market restructuring by incorporating women. The collocate introduces a gendered dimension to Saudization as it relates to Saudi women's employment and their participation in the nationalization process. Its prosody is shaped by national institutional planning on the one hand, and social and practical implementation on the other. In 67% of concordances (32/50 CL), a neutral prosody appears in institutional and administrative descriptions where تأنيث (feminization) functions as a technical component of labor regulations and policy implementation, as in example [41]. In 31% of lines (15/47 CL) a positive prosody appears, as تأنيث (feminization) is associated with expanding employment opportunities for Saudi women and enhancing their participation in the workforce, as in example [42], where the collocate is used to emphasize the success story of creating jobs for Saudis, especially women, and ensuring their socio-economic empowerment.

[41] ذلك ما أطلقته وزارة العمل متمثلاً في أكثر من 12 قراراً داعماً لتوظيف المرأة وتنظيم عملها في بيع محلات مستلزمات [sic] النسائية وبرنامج سعودة وتأنيث الوظائف الصناعية المناسبة.

- (41) This is what was initiated by the Ministry of Labor through more than 12 decisions supporting the employment of women and regulating their work in shops selling women's accessories, as well as a program for the **Saudization** and **feminization** of appropriate industrial jobs.

[42] عندما تمت السعودة ونفذت خطط التوظيف والتأنيث على المنشآت المتوسطة [...] فذلك إيجابياً قد صنع فرص عمل كثيرة لخريجي [...]

- (42) When **Saudization** was achieved and plans for nationalization and **feminization** were implemented in medium-sized establishments [...], this, being a positive development, created many job opportunities for graduates [...].

The concordance lines of the collocate استقدام (importation of foreign labor) represent a central oppositional situation within the discourse of Saudization. Its contextual patterns almost entirely position it in tension with Saudization both as a competing practice and a regulatory rule, and it is this contrast which shapes the semantic prosody of this collocate. In 68% of cases (34/50 CL), a neutral institutional prosody appears when استقدام (importation of foreign labor) collocates with administrative and regulatory discourse, indexed by lexical signals such as التزام/الملتزمة (commitment/compliant), تسهيل (facilitation), إجراءات (procedures), وزارة العمل (Ministry of Labor) and نسبة (percentage), as illustrated in examples [43] and [44] below. In both examples, the facilitation of foreign labor importation is discussed as a governmental commitment for institutions which correctly adhere to Saudization policy. The collocate, then, is represented as a mechanism within the Saudization policy, suggesting a balanced regulatory relationship between them.

[43] الالتزام بتسهيل إجراءات الاستقدام للمؤسسات الملتزمة بالسعودة

- (43) commitment to facilitating **importation of foreign labor** procedures for institutions that comply with **Saudization**

[44] وزارة العمل تقدم الدعم والمساعدة للمنشآت الملتزمة بتطبيق نسب السعودة من خلال تسهيل إجراءات الاستقدام في حال حاجتها لعمالة وافدة.

- (44) the Ministry of Labor provides support and assistance to establishments that comply with the application of Saudization ratios by facilitating importation procedures, should they require foreign labor.

However, in 28% of lines (14/50 CL), a negative oppositional prosody is salient, where استقدام (importation of foreign labor) is constructed as undermining Saudization objectives, or linked to exploitation and circumvention, or malpractices, as in example [44], which suggests that recruitment practices corrupt the intended goals of Saudization. In this sense, the collocate functions as a discursive counterpoint to Saudization, highlighting tensions between national employment goals and foreign labor importation.

[45] أبدعت الشركات في التغلب على قواعد السعودة بتوظيف سعوديين برواتب رخيصة، أو تسجيل الشركة ضمن نطاق عمل لا يتطلب نسبة كبيرة من السعودة، أو استقدام عمالة تحت مسميات وظيفية خاطئة.

- (45) Companies have excelled in overcoming Saudization regulations by employing Saudis at low salaries, registering the company under an employment scope that does not require a high **Saudization** rate, or **importation of foreign labor** under incorrect job titles.

The concordance lines of the collocate توظيف (employment) are central to the discourse of Saudization, since توظيف (employment) functions as the operational mechanism through which the policy of Saudization is realized. Its co-textual patterns reflect both aspirational and problematic aspects of labor nationalization, hence producing a mixed semantic prosody. In 64% of cases (32/50 CL), a neutral prosody appears in administrative and policy-related contexts, as in example [46], where the collocate is used as a technical bureaucratic term without explicit evaluation. When توظيف (employment) is associated with job creation and national workforce, i.e., in 16% of cases (8/50 CL), a positive prosody is created, as in example [47], which introduces a narrative of success and represents Saudization as effective in increasing national employment opportunities. However, a negative prosody is salient in 20% of lines (10/50 CL) which highlight fake or ineffective employment practices, as in example [48]:

[46] وسيتناول المتحدثون الاستراتيجيات الجديدة لتوطين الوظائف في قطاعاتهم ومستوى التكامل والتعاون في آليات التوظيف والسعودة.

- (46) The speakers will address new strategies for the nationalization of jobs in their sectors and the level of integration and cooperation in **employment** and **Saudization** mechanisms.

[47] هناك نماذج مشرفة حققت أكثر من المطلوب في السعودية وتوظيف المواطنين.

- (47) There are exemplary models that have achieved more than what is required in **Saudization** and the **employment** of citizens.

[48] أهمية تفعيل توظيف الأشخاص ذوي الإعاقة في القطاعات الخاصة التي ترغب [في] توظيفها حتى لا يكون برنامج السعودية وهمي [sic]

- (48) the importance of activating the employment of persons with disabilities in private sectors wishing to **employ** them, so that the **Saudization** program is not a **phantom** one

A key lexical item, the collocate عمالة (labor) functions as the primary object of regulation. Its co-textual environment suggests a strong oppositional representation between national and foreign labor, while its semantic prosodies are shaped by Saudization goals, socio-economic concerns, and critical evaluations of implementation. A positive prosody emerges in 24% of concordances (12/50 CL), as عمالة (labor) is associated with the reduction of foreign labor and the increase in national employment, as in example [49], where the achievement of scoring a high percentage of Saudization is celebrated. While a neutral prosody is evident in 60% of lines (30/50 CL) where the collocate appears in general discussions and descriptions of labor dynamics, a negative prosody emerges in 16% of cases (8/50 CL), where عمالة (labor) is used in relation to policy failure, resistance or unintended consequences, as in examples [50] and [51].

[49] نحن نسعى لتقليل العمالة الوافدة وإحلالها بسعوديين ونسبة السعودية في المشروع تجاوزت 95% والله الحمد.

- (49) We are striving to reduce the amount of foreign **labor** and replace it with Saudis; the **Saudization** rate in the project has exceeded 95%, praise be to God.

[50] وهو يجيب عن أسئلة [...] دارت حولها حول مشكلة البطالة، العمل والتوظيف، التدريب، العمالة الوافدة، العمالة السائبة، السعودية.

- (50) He answers questions [...] that largely revolved around the problem of unemployment, work and employment, training, foreign **labor**, undocumented **labor**, and **Saudization**.

[51] لا نرى أثرا يذكر لعمليات السعودية على أعداد العمالة الوافدة في المملكة والتي تزداد سنويا.

- (51) We do not see any significant impact of **Saudization** operations on the amount of foreign **labor** in the Kingdom, which is increasing annually.

The concordance lines of الليموزين (limousine), خضار (vegetables) and بقالة (grocery shop), being the members of the last group, collectively show their evident relevance to sector-specific discourse, respectively referring to the policy of nationalizing taxi drivers, vegetable markets and the retail sector. Their co-textual patterns reveal their role as salient case studies determining the success, failure and feasibility of Saudization policies. Unlike abstract terms, all three collocates link the discourse to concrete occupational domains. A neutral semantic prosody is observed in 61% of concordances (86/141 CL), as they are used to denote a specific sector targeted by Saudization policies, as in example [52], where the collocate الليموزين (limousine) functions as a technical term referring to a professional domain:

[52] قرار [...] بوجوب سعودة سيارات الأجرة العامة (الليموزين) على شركة مواصلات السعودية

- (52) A decision [...] regarding the mandatory **Saudization** of public taxis (**limousines**) for the Saudi Public Transport Company

Here, the collocate appears within an institutional or enforcement discourse, where Saudization is represented as a regulatory measure targeting the sector of taxi driving. A positive prosody is found in 16% of concordances (23/141 CL), as the implementation of

Saudization in these sectors is represented as successful and beneficial, or as a viable domain for employment opportunities and economic localization, as in examples [53] and [54] respectively, where the collocate contributes towards a nationalistic and optimistic discourse, representing grocery stores as an opportunity for economic benefit and work-force integration.

[53] محلات الخضار والفاكهة التي تم سعودتها وهذا بعد ذاتها كانت نقطة ايجابية [sic] جدا جدا

(53) **Vegetable** and fruit shops that have been subjected to **Saudization**, which in itself was a very, very positive point

[54] فلو تمت سعودة هذه البقالات والسوبر ماركت لكان خيرا عظيما لهم وكذلك خيرا للوطن والمجتمع بأكمله.

(54) If these **grocery shops** and supermarkets had undergone **Saudization**, it would have been a great good for them, as well as for the nation and society as a whole.

A negative semantic prosody is salient in 23% of lines (32/141 CL), as the three collocates are associated with policy inefficiency or impracticality, highlighting financial limitations and portraying the sector as economically unattractive and resistant to successful localization. The concordance lines of بقالة (grocery shop), for example, highlight the debate around its feasibility, economic impact and social acceptability, creating a highly evaluative semantic prosody, as in example [55], which represents the Saudization of low-skilled sectors such as grocery shops as socially and economically misaligned with youth expectations of status and quality jobs.

[55] وانتقد مغردون آخرون القرارات الأخيرة التي تهدف إلى سعودة البقالة، حيث تساءل المغردون هل يعقل أن يقضي شاب سنوات في الدراسة والتحصيل العلمي ليعمل أخيرا في بقالة.

(55) Other tweeters criticized the recent decisions aimed at the **Saudization** of **grocery shops**, asking whether it is reasonable for a young man to spend years in study and education only to end up working in a grocery store.

By relating the discourse of Saudization to concrete, low-skilled economic sectors, the three collocates الليموزين (limousine), خضار (vegetables) and بقالة (grocery shop) generate evaluative interpretations of policy outcomes. The semantic prosody is largely shaped by experiences and tensions of policy implementation. All in all, the semantic prosody of the three collocates in relation to سعودة is predominantly negative, with the sectors frequently used as evidence of policy limitations and implementation challenges.

5. Discussion and Conclusions

The present study investigates the collocational patterns and semantic prosodies of the node word سعودة (Saudization) in MSA discourse, drawing on the mega-corpus arTenTen24 accessed via the SkE online interface. Utilizing the methodological synergy of CL and CDA, the analysis combined the quantitative methods of collocation and concordance generation with the qualitative close exploration of concordance lines. Focusing on the twenty most statistically significant collocates, the study aims to uncover the evaluative meaning and ideological representations related to Saudization in MSA discourse.

The critical research gap identified in the literature review ultimately provided clear methodological foundations across four domains: (1) the synergy of CDA and CL, (2) collocation extraction methodologies and semantic prosody identification, (3) an evaluation of Saudi labor nationalization policies, and (4) discourse analyses of the ideological representations of Saudization in MSA. The findings of the study reveal that the collocational profile of سعودة (Saudization) is both semantically dense and ideologically laden, reflecting its importance within the Saudi national labor market with all its socio-economic and socio-political intricacies. Thus, the linguistic dimension of Saudization is complementary to its social, economic and political dimensions. Responding to the first research question about

the collocational profile of *سعودة* (Saudization), corpus-based investigation revealed that the collocates mainly revolve around four topical domains: (1) an institutional regulatory and administrative domain, for example, *سعودة* (Saudization), *نطاقات* (Nitaqat), *احتساب* (calculation), *نسب* (percentages), *إحلال* (replacement), *توطين* (localization) and *وتغريمها* (and fining them); (2) a critical evaluative domain, for example, *وهمي* (phantom), *تحايل* (circumvention), *بطالة* (unemployment) and *تستر* (concealment); (3) a labor market restructuring domain, for example, *وظيفة* (job), *توظيف* (employment), *عمالة* (labor), and *استقدام* (importation of foreign labor); and (4) a sector-specific domain, for example, *ليموزين* (limousine), *خضار* (vegetables), and *بقالة* (grocery shop). Taken together, these collocational patterns reflect the discursive construction of *سعودة* (Saudization) in relation to national identity, the labor market, and the tensions between institutional regulations and practical realities.

As for the second research question, the semantic prosodies associated with *سعودة* (Saudization) as revealed by corpus examination are predominantly mixed, with a slight skewing towards negative and critical evaluation. The attitudinal meaning is dual, combining positive nationalistic semantic prosodies which advocate national pride and patriotism, and negative prosodies which correspond to the socio-economic complexities of Saudization. While positive and neutral prosodies are present, particularly in contexts relating Saudization to job creation, highlighting national identity, stressing Saudization as a national duty, pride and development, and emphasizing Saudization policy success, they are often overshadowed by critical and delegitimizing patterns. For example, collocates like *نطاقات* (Nitaqat), *استقدام* (importation of foreign labor), and *توظيف* (employment) carry a positive prosody when relating to the increase in employment opportunities and job creation, national empowerment, and state-led reform. However, their co-occurrence with expressions such as *وهمي* (phantom) in the context of nominal employment weighs the scale towards skepticism. Similarly, institutional collocates like *نسب* (percentages) and *احتساب* (calculation) are neutral in lines discussing the process of policy implementation but acquire negative prosodic hues when associated with inflated figures and symbolic compliance. The repeated co-occurrence of *سعودة* (Saudization) with terms such as *تحايل* (circumvention), *تستر* (concealment), and *وهمي* (phantom) creates a consistent semantic prosody of manipulation and illegitimacy as to the procedural implementation of its practices. These patterns demonstrate that semantic prosody is not inherent in individual lexical items but emerges from recurrent contextual associations across the corpus. Close examination of collocates in their concordance lines reveals that the negative semantic prosodies around the concept of Saudization are related to criticizing inefficiency and malpractices of its implementation rather than criticizing the policy per se, i.e., the positive, the neutral and the negative semantic prosodies collectively pour into the Saudi endeavors towards achieving patriotic national duty of localizing the labor market and workforce.

From a CDA perspective, and pertinent to the third research question about the ideological representations that *سعودة* (Saudization) conveys in an MSA mega-corpus, the above findings correspond to a complex ideological representation of Saudization in MSA discourse. The policy is simultaneously constructed as a national imperative and as a problematic practice. On the one hand, the discourse frequently represents Saudization as a necessary strategy for addressing unemployment, improving national participation, and achieving economic stability. This is evident in collocational patterns linking *سعودة* (Saudization) to unemployment, localization and institutional initiatives which collectively contribute towards portraying Saudization as equivalent to reform and development. On the other hand, the appearance of critical collocates such as *وهمي* (phantom), *تستر* (concealment), and *تحايل* (circumvention) reveals a counter-discourse that questions the way Saudization is implemented and highlights discrepancies between policy objectives and actual practices. This duality reflects broader socio-economic tensions such as the restructuring of the labor

market, dependency on foreign labor and the socio-economic challenges of implementation. Moreover, the sector-specific collocates such as بقالة (grocery shop), الليموزين (limousine), and خضار (vegetables) demonstrate how Saudization is often evaluated in particular domains, especially those perceived as low-status. The examination of collocates, then, contributes to an understanding of how Saudization is discursively constructed in MSA, i.e., how Saudization is ideologically represented and conceived as regulation, national duty, nationalization necessity, and patriotic pride but, simultaneously, contested through critical and evaluative commentary as to its implementation and its socio-economic implications.

In conclusion, confirming the study hypothesis, the methodological synergy of CDA and CL provided a capable tool for exploring how السعودية (Saudization) is discursively constructed in MSA; i.e., the close examination of co-occurrence patterns proved that different evaluative meanings are linguistically encoded in collocational patterns and semantic prosodies, and demonstrated how semantic prosody serves as a powerful indicator of latent ideological conceptions. The findings, thus, reveal السعودية (Saudization) as a semantically dense and ideologically layered term in MSA discourse; that is, while consistently legitimized as a national duty and a necessary strategy for economic reform and workforce development, it is simultaneously subject to a counter-discourse that questions its implementation through recurrent associations with circumvention, concealment, and phantom compliance. Beyond their specific contribution to understanding Saudization, these findings underscore the broader value of integrating collocational and semantic prosodic analysis within Arabic corpus linguistics, an area in which such approaches remain comparatively underexplored.

Funding: The author extends her appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2025/02/34090).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The corpus analyzed in this study, arTenTen24, is available through the Sketch Engine online interface at sketchengine.eu, subject to subscription. Extracted data generated for this study are available upon request.

Conflicts of Interest: The author declares no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

- Abdulrahim, D. (2014). Annotating corpus data for a quantitative constructional analysis of motion verbs in Modern Standard Arabic. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. Association for Computational Linguistics. [CrossRef]
- Alanazi, F. (2020). *Saudis in the eyes of the other: A corpus-driven critical discourse study of the representation of saudis on Twitter* [Doctoral dissertation, Lancaster University]. Lancaster University Repository. [CrossRef]
- Alansary, S., & Nagi, M. (2014). The international corpus of arabic: Compilation, analysis and evaluation. In *Proceedings of the EMNLP 2014 workshop on Arabic natural language processing (ANLP)* (pp. 8–17). Association for Computational Linguistics.
- Al-Otaibi, G. M. (2022). Semantic prosody of research verbs: A corpus-informed study. *Journal of Language and Education*, 8(2), 48–65. [CrossRef]
- Al-Otaibi, G. M. (2025). Semantic prosody of deictic verbs in the Holy Qur'ān: A corpus study. *Humanities and Social Sciences Communications*, 12, 88. [CrossRef]
- Baker, P. (2006). *Using corpora in discourse analysis*. Continuum.
- Baker, P. (2014). *Using corpora to analyze gender*. Bloomsbury.
- Baker, P., Gabrielatos, C., Khozravinik, M., Krzyzanowski, M., McEnery, T., & Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society*, 19(3), 273–305. [CrossRef]

- Baker, P., Gabrielatos, C., & McEnery, T. (2013). Sketching muslims: A corpus driven analysis of representations around the word 'Muslim' in the British press 1998–2009. *Applied Linguistics*, 34/3, 255–278. [CrossRef]
- Baker, P., Hardie, A., & McEnery, A. (2006). *A glossary of corpus linguistics*. Edinburgh University Press.
- Baker, P., & McEnery, T. (2005). A corpus-based approach to discourses of refugees and asylum seekers in UN and newspaper texts. *Journal of Language and Politics*, 42, 197–226. [CrossRef]
- Bednarek, M. (2008). Semantic preference and semantic prosody re-examined. *Corpus Linguistics and Linguistic Theory*, 4(2), 119–139. Available online: <https://opus.lib.uts.edu.au/bitstream/10453/10612/1/2008000824.pdf> (accessed on 16 March 2026). [CrossRef]
- Begagić, M. (2013). *Semantic preference and semantic prosody of the collocation make sense*. University of Zenica. Available online: <https://hrcak.srce.hr/file/165524> (accessed on 23 March 2026).
- Ben Ghazlen, B. (2022). A critical overview of semantic prosody. *Legon Journal of the Humanities*, 33(2), 59–79. [CrossRef]
- Berber-Sardinha, T. (2000). Semantic prosodies in English and Portuguese: A contrastive study. *Cuadernos de Filología Inglesa (University of Murcia, Spain)*, 9(1), 93–110.
- Bublitz, W. (1996). Semantic prosody and cohesive company: Somewhat predictable. *Leuvense Bijdragen: Tijdschrift voor Germaanse Filologie*, 85/1–2, 1–32.
- Cantos, P., & Almela-Sánchez, M. (2021). Beyond base and collocate: On the importance of the second-order collocates. *International Journal of Corpus Linguistics*, 26(4), 548–575. [CrossRef]
- Cardona, M. D. (2022). More than just an immigrant: The semantic patterns of (im)migrant/predicate-pairings in news stories about Mexican and Central American (im)migrants to the USA. A corpus-assisted discourse study. *Discourse & Communication*, 16(4), 399–422. [CrossRef]
- Coffin, C., Hewings, A., & O'Halloran, K. (Eds.). (2004). *Applying English grammar. Functional and corpus approaches*. Arnold.
- Cortés, P., Kasoolu, S., & Pan, C. (2020). *Labor market nationalization policies and firm outcomes: Evidence from Saudi Arabia* (NBER Working Paper No. 27737). National Bureau of Economic Research. [CrossRef]
- Elghamry, K. (2025). OPINAR: A multi-source Arabic opinion dataset with keywords, 2004–2025. Mendeley Data, Version 1. Available online: <https://data.mendeley.com/datasets/nbmys7wnm2/1> (accessed on 8 September 2026). [CrossRef]
- Fairclough, N. (1995). *Critical discourse analysis. The critical study of language*. Longman.
- Fairclough, N., Mulderrig, J., & Wodak, R. (2011). Critical discourse analysis. In T. A. Van Dijk (Ed.), *Discourse studies: A multidisciplinary introduction* (2nd ed.). Sage Publications.
- Figureiredo, D. C. (2004). Representations of rape in the discourse of legal decisions. In L. Young, & C. Harrison (Eds.), *Systemic functional linguistics and critical discourse analysis* (pp. 155–169). Continuum.
- Firth, J. R. (1957). *Papers in linguistics 1934–1951*. Oxford University Press.
- Fischer-Starcke, B. (2010). *Corpus linguistics in literary analysis*. Continuum.
- Gabrielatos, C., & Baker, P. (2008). Fleeing, sneaking, flooding: A corpus analysis of discursive constructions of refugees and asylum seekers in the UK Press 1996–2005. *Journal of English Linguistics*, 36/1, 5–38.
- Gavioli, L. (2005). *Exploring corpora for ESP learning*. John Benjamins.
- Griebel, T., & Vollmann, E. (2019). We can('t) do this: A corpus-assisted critical discourse analysis of migration in Germany. *Journal of Language and Politics*, 18(5), 671–697.
- Grundmann, R., & Krishnamurthy, R. (2010). The discourse of climate change: A corpus based approach. *Critical Approaches to Discourse Analysis Across Disciplines*, 4(2), 125–146.
- Hardiman, D. P., & Nuraniwati, T. (2024). Semantic preference and semantic prosody of the collocations of sustainable in the NOW Corpus. 3L: *Language, Linguistics, Literature, The Southeast Asian Journal of English Language Studies*, 30(3), 184–199. [CrossRef]
- Hertog, S. (2014). *Arab Gulf states: An assessment of nationalisation policies* [Gulf Labour Markets and Migration (GLMM) Research Paper 1/2014]. European University Institute and Gulf Research Center. Available online: <https://hdl.handle.net/1814/32156> (accessed on 12 January 2026).
- Hunston, S. (2001). Colligation, lexis, pattern and text. In M. Scott, & G. Thompson (Eds.), *Patterns of text: In honour of michael* (pp. 13–33). John Benjamins.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.
- Hunston, S., & Thompson, G. (Eds.). (1999). *Evaluation in text: Authorial stance and the construction of discourse*. Oxford University Press.
- Hunt, S. (2011). *The discursual construction of female physical identity in selected works in children's literature* [Unpublished doctoral thesis]. Rhodes University.
- Hussain, I. (2019). Using word sketches to resolve prepositional phrase attachment ambiguity in Arabic. *International Journal of Computer Applications*, 177(16), 51–56. [CrossRef]
- Ibrahim, W. M. A. (2022). Breaking the silence: A corpus-assisted analysis of narratives of the victims of an Egyptian sexual predator. *Open Linguistics*, 8(1), 158–188. [CrossRef]
- Jaworska, S., & Krishnamurthy, R. (2012). On the F word: A corpus-based analysis of the media representation of feminism in British and German press discourse, 1990–2009. *Discourse & Society*, 23(4), 401–431. [CrossRef]

- Kenny, D. (2001). *Lexis and creativity. A corpus-based study*. Routledge.
- Kilgarrieff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The sketch engine: Ten years on. *Lexicography*, 1, 7–36.
- Kilgarrieff, A., Rychlý, P., Smrž, P., & Tugwell, D. (2004, November 17). *The sketch engine*. 11th EURALEX International Congress (pp. 105–116), Leiden, The Netherlands.
- Krishnamurthy, R. (1996). Ethnic, racial and tribal: The language of racism? In C. R. Caldas-Coulthard, & M. Coulthard (Eds.), *Texts and practices: Readings in critical discourse analysis* (pp. 129–149). Routledge.
- Leech, G. N. (1974). *Semantics*. Penguin.
- Li, Z., Zhao, R., & Lou, B. (2023). Corpus-based critical discourse analysis of reporting practices in English news reports on public health events in China and the United States. *Frontiers in Psychology*, 14, 1137382. [CrossRef]
- Liu, M., Zhao, R., & Ngai, C. S. B. (2022). Vaccines, media and politics: A corpus-assisted discourse study of press representations of the safety and efficacy of COVID-19 vaccines. *PLoS ONE*, 17(12), e0279500. [CrossRef]
- Liu, W. (2020). A corpus-based study of semantic prosody and semantic preference of “seem”. *Open Journal of Social Sciences*, 8(5), 282–291. [CrossRef]
- Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 157–175). John Benjamins. [CrossRef]
- Mahlberg, M. (2013). Corpus Stylistics. In M. Burke (Ed.), *The routledge handbook of stylistics* (pp. 378–392). Routledge.
- Malrieu, J. P. (1999). *Evaluative semantics: Cognition, language and ideology*. Routledge.
- Mautner, G. (1995). *Only connect: Critical discourse analysis and corpus linguistics* (UCREL technical paper 6). University of Lancaster.
- Mautner, G. (2009). Checks and balances: How corpus linguistics can contribute to CDA. In R. Wodak, & M. Mayer (Eds.), *Methods of critical discourse analysis*. Sage.
- McEnery, T., & Wilson, A. (2001). *Corpus linguistics: An introduction* (2nd ed.). Edinburgh University Press.
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. Routledge.
- Moustafa, B. S. M. (2022). A comparative corpus stylistic analysis of thematization and characterization in Gordimer’s *My Son’s Story* and Coetzee’s *Disgrace*. *Open Linguistics*, 8(1), 46–64. [CrossRef]
- Nattinger, J., & DeCarrico, J. (1992). *Lexical phrases and language teaching*. Oxford University Press.
- Nguyen, M. Q. M. N. (2023). Media presentations of Vietnam’s cybersecurity law: A comparative approach with corpus-based critical discourse analysis. *Computer Law and Security Review*, 50, 105835. [CrossRef]
- O’Donnell, M. (2023). “The pingdemic has been a greater challenge than Covid itself”: Semantic prosodies in UK newspaper articles during the pandemic. *SN Social Sciences*, 3, 146. [CrossRef]
- Okeke, D. E. (2025). Race, violence in the US, and digital (news) discourse: A corpus-based critical discourse analysis of Jordan Neely’s racialization. *Communication, Culture and Critique*, 18(4), 322–337. [CrossRef]
- O’Regan, V., & Riordan, E. (2018). Comparing the representation of refugees, asylum seekers and migrants in the Irish and UK press: A corpus-based critical discourse analysis. *Journal of Language and Politics*, 17(6), 744768. [CrossRef]
- Palmer, H. E. (1938/1968). *A grammar of English words*. Longman. (Original work published 1938).
- Partington, A. (2004). “Utterly content in each other’s company”: Semantic prosody and semantic preference. *International Journal of Corpus Linguistics*, 9/1, 131–156.
- Partington, A. (2009). Evaluating evaluation and some concluding thoughts on CADS. In J. Morley, & P. Bayley (Eds.), *Wordings of war. Corpus assisted discourse studies on the war in Iraq*. Routledge.
- Partington, A., Duguid, A., & Taylor, C. (2013). *Patterns and meanings in discourse: Theory and practice in corpus-assisted discourse studies (CADS)*. John Benjamins.
- Peck, J. R. (2017). Can hiring quotas work? The effect of the Nitaqat program on the Saudi private sector. *American Economic Journal: Economic Policy*, 9(2), 316–347. [CrossRef]
- Qarah, F. (2024). SaudiBERT: A large language model pretrained on Saudi dialect corpora. *PsyArXiv*, arXiv:2405.06239. [CrossRef]
- Rajan, S. I., & Arokkiaraj, H. (2019). Emerging policies in the Gulf and its impact on Asia. In *India’s low-skilled migration to the middle east: Policies, politics and challenges* (pp. 355–370). Routledge. [CrossRef]
- Randeree, K. (2012). Workforce nationalization in the gulf cooperation council states. *Center for International and Regional Studies Occasional Paper*, 9, 1–20.
- Salama, A. H. Y. (2023). Towards a multimodal hermeneutic model: The case of uber-blog-mediated advertising discourse order of ‘Saudization’. *Russian Journal of Linguistics*, 27(4), 886–914. [CrossRef]
- Salama, A. H. Y., & Fawzy, R. (2024). Enchronic cultural discourse analysis: A semio-cultural study of national identity discourse of Saudi Founding Day. *Journal of Multicultural Discourses*, 18(4), 377–400. [CrossRef]
- Selmistraitis, L. (2020). Semantic preference, prosody and distribution of synonymous adjectives in COCA. *GEMA Online, Journal of Language Studies*, 20(3), 1–18. [CrossRef]

- Sinclair, J. (1966). Beginning the study of lexis. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, & R. H. Robins (Eds.), *In memory of J.R. Firth* (pp. 410–430). Longman.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Sinclair, J. (1996). The search for units of meaning. *Textus*, 9, 75–106.
- Sinclair, J. (2003). *Reading concordances: An introduction*. Pearson Longman.
- Smadi, A. M. (2020). A corpus-based analysis of Arabic abstracts in social sciences. *International Journal of Linguistics*, 12(6), 50–80. [[CrossRef](#)]
- Stewart, D. (2010). *Semantic prosody: A critical evaluation*. Routledge. [[CrossRef](#)]
- Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of Language*, 2(1), 23–55.
- Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Blackwell.
- Stubbs, M. (2006). Corpus analysis: The state of the art and three types of unanswered questions. In G. Thompson, & S. Hunston (Eds.), *Style and corpus: Exploring connections* (pp. 15–36). Equinox.
- Su, P., & Phoocharoensil, S. (2021). Semantic prosody and collocation: A corpus study of the near-synonyms persist and persevere. *Eurasian Journal of Applied Linguistics*, 7(1), 240–258. [[CrossRef](#)]
- Tognini-Bonelli, E. (2004). Working with corpora: Issues and insights. In C. Coffin, A. Hewings, & K. O'Halloran (Eds.), *Applying english grammar. Functional and corpus approaches* (pp. 11–24). Arnold.
- Van Dijk, T. (2008). *Society in discourse: How context controls text and talk*. Cambridge University Press.
- Widdowson, H. G. (2004). *Text, context, pretext: Critical issues in discourse analysis*. Blackwell.
- Xu, T. (2025). A corpus-based critical discourse analysis of reuters' news coverage of the 2025 Sino-U.S. Trade war. *International Linguistics Research*, 8(2), 30–39. [[CrossRef](#)]
- Yangxia, R., & Mengdan, Z. (2025). A corpus-driven critical discourse analysis of news reports on disabled women. *Cogent Arts & Humanities*, 12(1), 2465027. [[CrossRef](#)]
- Zaghouni, W. (2014). Critical survey of the freely available Arabic corpora. In *Proceedings of the workshop on free/open-source Arabic corpora and corpora processing tools workshop at LREC 2014* (pp. 1–8). European Language Resources Association.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.