




Saudi Vision 2030 in University English Educational Discourse at Qassim University: A Corpus-Assisted Analysis of Thematic and Publication-Time Maps of Development and Sustainability Lexis in Curriculum and Website Texts

Dr. Bashar Ragheb Hasan Odeh ^{*} 

basher@qu.edu.sa

Abstract:

This article investigates how Saudi Vision 2030 is recontextualized in the official English-language educational discourse of Qassim University. Rather than measuring policy visibility only through direct reference to the Vision, the study traces how development- and sustainability-related priorities are redistributed across institutional lexis, genre, and publication-time patterns. The analysis is based on a specialized institutional micro-corpus of 60 official English-language sources, evenly divided between Web texts and PDF documents. Using a CADS-informed design, the study combines thematic mapping, normalized frequency, document coverage, keyword-informed lexical profiling, exploratory Mann-Whitney U comparison, and qualitative concordance reading. The results show that Education and Skills is the most salient and widely dispersed axis, while the Web and PDF subcorpora perform different institutional functions. Web texts foreground competencies, employability, economic relevance, and innovation as part of an outward-facing academic identity. PDF texts, by contrast, encode alignment more strongly through governance, quality systems, regulations, and procedural documentation. The publication-time profile is interpreted cautiously as evidence of available publication and update cycles, not as a balanced diachronic record. Overall, the study argues that Vision 2030 enters Qassim University's English educational discourse less through repeated explicit citation than through recurrent institutional wording that links quality, competencies, labor-market alignment, sustainability, governance, innovation, and partnership.

Keywords: Institutional discourse, Discourse studies, Corpus-assisted discourse studies, Development and sustainability lexis, Saudi Vision 2030.

* Assistant Professor of Applied Linguistics, Department of English Language and Literature, College of Languages and Humanities, Qassim University, Kingdom of Saudi Arabia.

Cite this article as Odeh, B. R. H. (2026). Saudi Vision 2030 in University English Educational Discourse at Qassim University: A Corpus-Assisted Analysis of Thematic and Publication-Time Maps of Development and Sustainability Lexis in Curriculum and Website Texts, *Arts for Linguistic & Literary Studies*, 8(2): 627 -656. <https://doi.org/10.53286/rpx45z94>

© This material is published under the license of Attribution 4.0 International (CC BY 4.0), which allows the user to copy and redistribute the material in any medium or format. It also allows adapting, transforming or adding to the material for any purpose, even commercially, as long as such modifications are highlighted and the material is credited to its author.



أثر رؤية المملكة 2030 في الخطاب التعليمي الجامعي باللغة الإنجليزية في جامعة القصيم: تحليل مدوّني للخرائط الموضوعية والزمنية لمفردات التنمية والاستدامة في نصوص المناهج والمواقع

د. بشار راغب حسن عودة ^{ID*}

basher@qu.edu.sa

الملخص:

يتناول هذا البحث كيفية إعادة سياق أولويات رؤية المملكة 2030 وتمثيلها لغويًا في الخطاب التعليمي الجامعي باللغة الإنجليزية في جامعة القصيم. ولا يقتصر التحليل على رصد الإحالات المباشرة إلى الرؤية، بل يتتبع كيف تتوزع مفردات التنمية والاستدامة داخل المعجم المؤسسي وأنماط النصوص ودورات النشر والتحديث. ويعتمد البحث على مدونة مؤسسية متخصصة تضم ستين مصدرًا رسميًا باللغة الإنجليزية، موزعة بالتساوي بين نصوص الويب ووثائق PDF. ويجمع التصميم التحليلي بين الخرائط الموضوعية، والتكرارات المعيارية، ومدى الانتشار داخل الوثائق، والتحليل المعجمي المقارن، والمقارنة الاستدلالية الاستكشافية، والقراءة الكيفية للشواهد السياقية. وتظهر النتائج أن محور التعليم والمهارات هو الأكثر بروزًا وانتشارًا، كما تكشف عن تمايز وظيفي بين النصوص؛ إذ تميل نصوص الويب إلى إبراز الكفايات والقابليات للتوظيف والملاءمة الاقتصادية والابتكار ضمن هوية أكاديمية موجّهة للجمهور، بينما ترسخ ووثائق PDF هذا الارتباط من خلال الحوكمة ونظم الجودة واللوائح والإجراءات. ويُفسّر البعد الزمني في هذه الدراسة بوصفه مرتبطًا بتواريخ النشر والتحديث المتاحة للوثائق، لا سجلًا تاريخيًا متوازنًا. ويخلص البحث إلى أن أثر الرؤية يظهر في خطاب الجامعة الإنجليزية عبر مفردات مؤسسية متكررة تربط الجودة والكفايات والمواءمة مع سوق العمل والاستدامة والحوكمة والابتكار والشراكة.

الكلمات المفتاحية: الخطاب المؤسسي، دراسات الخطاب، دراسات الخطاب المدعومة بالمدونات اللغوية، مفردات

التنمية والاستدامة، رؤية المملكة 2030.

* أستاذ اللغويات التطبيقية المساعد، قسم اللغة الإنجليزية وأدائها، كلية اللغات والعلوم الإنسانية، جامعة القصيم، المملكة العربية السعودية.

للاقتباس: عودة، ب. ح. (2026). أثر رؤية المملكة 2030 في الخطاب التعليمي الجامعي باللغة الإنجليزية في جامعة القصيم: تحليل مدوّني للخرائط الموضوعية والزمنية لمفردات التنمية والاستدامة في نصوص المناهج والمواقع، *الآداب للدراسات اللغوية والأدبية*، 8(2): 627-656 <https://doi.org/10.53286/rpx45z94>

© نُشر هذا البحث وفقًا لشروط الرخصة Attribution 4.0 International (CC BY 4.0)، التي تسمح بنسخ البحث وتوزيعه ونقله بأي شكل من الأشكال، كما تسمح بتكييف البحث أو تحويله أو إضافته إليه لأي غرض كان، بما في ذلك الأغراض التجارية، شريطة نسبة العمل إلى صاحبه مع بيان أي تعديلات أُجريت عليه.



1. Introduction

Saudi Vision 2030 has reshaped the policy environment in which Saudi universities define their academic, economic, and social roles. Higher education is now expected to contribute visibly to human-capital development, labor-market readiness, innovation, research productivity, quality assurance, and sustainable development (Almesaar, 2024; Barnawi, 2018; Saudi Vision 2030, 2016). These expectations are not confined to strategic plans or administrative reforms. They also become legible in the language through which universities describe programs, graduate attributes, quality systems, partnerships, and public institutional value.

For this reason, university discourse deserves attention as a site of policy uptake. Program pages, college websites, deanship pages, handbooks, regulations, quality manuals, and strategic documents do more than transmit institutional information. They organize how the university presents legitimacy, audience orientation, and educational purpose (Fairclough, 1993; Gering et al., 2023; Mautner, 2005; Saichaie, 2011). English-language versions of such texts are particularly relevant because they address prospective students, professional communities, accreditation bodies, international partners, and other audiences for whom English functions as an institutional interface (Al-Qahtani, 2021; Alqahtani, 2024).

The present study examines Qassim University as a bounded institutional case. Its concern is not whether Saudi Vision 2030 is repeatedly named in every official text. A more productive question is how Vision-related priorities are reworked into the everyday English lexis of the institution: quality, competencies, learning outcomes, labor-market alignment, sustainability, governance, innovation, partnership, and strategic alignment. These terms are analytically important because they show how national priorities can become ordinary textual routines within university communication.

The central research question is: How are the priorities of Saudi Vision 2030 manifested in university English educational discourse at Qassim University? The question is addressed through thematic mapping and publication-time analysis of lexical fields associated with development, sustainability, education and skills, economy, innovation, governance, society, environment, and partnership. The analysis also compares Web texts and PDF documents, since these two text types are likely to differ in audience, genre, and institutional function.

The study contributes to scholarship on Saudi higher education discourse by shifting the analytical focus from policy intention to institutional representation. Previous research has examined Vision 2030 in relation to educational reform, English language policy, university websites, and modernization (Almesaar, 2024; Al-Qahtani, 2021; Alqahtani, 2024; Barnawi, 2018). Less attention has been given to the concrete



lexical mechanisms through which official university English texts translate national development priorities into program-facing, quality-related, and governance-oriented discourse.

Methodologically, the study is positioned within corpus-assisted discourse studies (CADS), where quantitative patterning and qualitative interpretation are used together rather than sequentially or mechanically (Baker, 2006; Flowerdew, 2012; Gillings et al., 2023; Partington et al., 2013). CADS is appropriate because the study needs both distributional evidence and close reading; frequency can show where a theme is salient, but concordance evidence is needed to explain how a lexical item functions in a specific institutional context.

The corpus consists of 60 official English-language sources from Qassim University, divided equally between Web and PDF texts. This specialized institutional micro-corpus is not designed to generalize statistically to all Saudi universities. Its value lies in its coherence. It permits a close, transparent reading of one policy-relevant institutional environment where national development discourse, English-language representation, and university genre practices intersect.

Overall, the article treats university English educational discourse as an institutional space in which Saudi Vision 2030 is operationalized, adapted, and legitimized through repeated lexical choices. By tracing thematic and publication-time patterns, the study offers a corpus-informed account of how a national transformation agenda enters the textual routines of a contemporary Saudi university.

2. Theoretical Framework

2.1 Saudi Vision 2030 and Higher Education: From Developmental Framework to Discursive Effect

Saudi Vision 2030 provides the national development frame within which Saudi universities increasingly articulate their educational, economic, and social responsibilities (Almesaar, 2024; Barnawi, 2018; Saudi Vision 2030, 2016). For higher education, this frame foregrounds human-capital development, educational quality, research productivity, labor-market alignment, innovation, and sustainable growth. These priorities do not remain at the policy level; they are translated into curriculum descriptions, learning outcomes, quality systems, and public institutional narratives.

The discursive effect of Vision 2030 is therefore approached here as institutional uptake. The study does not assume that national policy language is copied directly into university texts. Instead, it asks how policy priorities are selected, condensed, and reformulated within the communicative genres of a university. This distinction allows the analysis to identify Vision-related meanings even where the Vision itself is not repeatedly named.



2.2 Institutional Discourse in Higher Education as Identity and Legitimacy Work

Institutional discourse in higher education helps universities construct identity, legitimacy, and strategic relevance (Fairclough, 1993; Gering et al., 2023; Mautner, 2005; Saichaie, 2011). In this study, official webpages, program descriptions, quality manuals, and PDF guidelines are treated as institutional texts that position Qassim University as responsive to innovation, sustainability, employability, and state-led modernization.

This framing matters because higher education discourse is not a neutral container for information. Terms such as quality, competencies, employability, partnership, innovation, and sustainability carry rhetorical as well as descriptive force. They help locate the university within a competitive, accountable, and policy-driven higher education environment. The corpus is therefore read as a bounded site where educational work, national priorities, and institutional legitimacy intersect.

2.3 English in Saudi University Discourse: From Instructional Medium to Institutional Interface

In Saudi higher education, English operates not only as a language of instruction but also as an institutional interface through which universities communicate with prospective students, accreditation bodies, collaborators, and professional stakeholders (Almesaar, 2024; Al-Qahtani, 2021; Alqahtani, 2024). English-language pages and documents form part of the university's strategic communication environment.

Accordingly, English is examined here as a semiotic and institutional resource. The study does not focus narrowly on classroom language. It investigates the English-medium representation of curricula, regulations, quality frameworks, and strategic priorities in texts that circulate beyond the immediate classroom.

2.4 Corpus-Assisted Discourse Studies as an Applied Analytical Approach

Corpus-assisted discourse studies provides the methodological foundation for the research because it enables movement between statistical salience and situated textual interpretation (Baker, 2006; Flowerdew, 2012; Gillings et al., 2023; Partington et al., 2013). This movement is crucial in a study of institutional lexis. A frequent word is not meaningful by itself; its analytical value depends on where it occurs, how widely it is dispersed, and what work it performs in context.

In this study, CADS supports a linked sequence of procedures: thematic mapping, normalized frequency, document coverage, keyword-informed profiling, exploratory comparison between Web and PDF texts, and qualitative concordance reading. The design is well suited to the research question because Vision 2030-related meanings may appear through expressions such as learning outcomes, competencies, labor market, quality standards, strategic alignment, partnership, and sustainable development. Such expressions require both corpus-level detection and context-sensitive interpretation.



2.5 Recontextualization as an Interpretive Framework

Recontextualization is the central interpretive lens of the study. Bernstein (1990) explains how pedagogic discourse selects, relocates, and refocuses external forms of knowledge according to institutional rules. Linell (1998) and Wodak (2000) further emphasize that meanings do not travel unchanged across contexts; they are transformed as they move across discourse domains, genres, and institutional purposes.

Applied to the present corpus, recontextualization means that broad national priorities such as development, sustainability, capability formation, competitiveness, and innovation are not expected to appear as policy slogans alone. They may be reworked as program outcomes, graduate competencies, quality standards, labor-market alignment, strategic planning, and governance documentation. Recurrent institutional lexis is therefore treated as evidence of policy recontextualization only when distributional patterns and close reading support that interpretation.

2.6 Publication-Time Distribution and Institutional Change

Publication-time mapping is used as a descriptive layer, not as a claim of balanced diachronic analysis. English-language university texts are shaped by website updates, translation practices, quality-review cycles, strategic planning, and document renewal. These institutional rhythms affect which texts are available in the corpus at a given time.

A translation-studies perspective is relevant because English institutional texts may reflect organizational norms, genre expectations, and audience design rather than literal equivalence alone (Kang, 2008; Koskinen, 2008; Schaffner et al., 2014). Developmental priorities may therefore be condensed, stabilized, or redistributed when represented in English. For this reason, publication-time distribution is interpreted as a map of available publication and update cycles, not as a full historical reconstruction of Qassim University discourse.

2.7 Research Gap and Theoretical Contribution

Although Vision 2030 has generated growing scholarship on educational policy, modernization, and the role of English in Saudi Arabia, fewer studies have examined how its priorities are embedded in university English educational discourse through a CADS lens (Almesaar, 2024; Al-Qahtani, 2021; Alqahtani, 2024). Existing work provides valuable macro-level accounts, but less is known about how policy priorities are lexically reformulated in the routine texts of a single institution.

This study addresses that gap by integrating institutional discourse analysis, recontextualization theory, and corpus-assisted methods. Its theoretical contribution lies in showing how national transformation is translated into university English lexis, while its methodological contribution lies in demonstrating how a carefully bounded institutional micro-corpus can make genre-based patterns visible.

2.8 Conceptual Model of the Study

The conceptual model integrates CADS and recontextualization theory (Bernstein, 1990; Linell, 1998; Wodak, 2000). It treats institutional discourse as an active mediating filter rather than a passive channel. National developmental meanings enter the university's English discourse after being selected, reorganized, and adapted to genre, audience, and institutional purpose.

Within the model, Web and PDF texts represent two communicative pathways in the same institutional environment. Web texts are treated as outward-facing and program-oriented, while PDF texts are treated as more formal, regulatory, and governance-oriented. The model guides the analysis without predetermining the results; the extent and direction of thematic differences are left to the corpus evidence.

Figure 1 illustrates the conceptual model guiding the study by linking Saudi Vision 2030, recontextualization, institutional English discourse, the Web and PDF subcorpora, and the corpus-assisted analytical procedures.

Figure 1. Conceptual model of the study

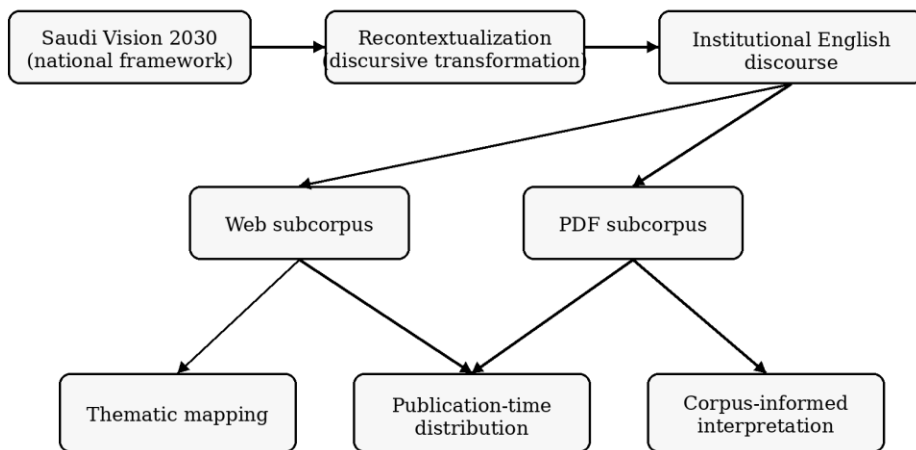


Figure 1. Conceptual Model of the Study

Note. The model links Saudi Vision 2030 as a higher-order national framework to recontextualization, institutional English discourse, the Web and PDF subcorpora, and the study's thematic mapping and publication-time distribution procedures.

2.9 Research Questions

The study is guided by the following main research question:



How are the priorities of Saudi Vision 2030 manifested in university English educational discourse at Qassim University?

This main question is operationalized through four sub-questions:

1. Which thematic axes associated with Saudi Vision 2030 are most salient in the English-language institutional discourse of Qassim University?
2. *How are these thematic axes distributed across the Web subcorpus and the PDF subcorpus?*
3. *Which lexical items and semantic fields distinguish the Web and PDF subcorpora?*
4. How does the available corpus reflect the publication-time concentration of institutional discourse across published update cycles?

2.10 Exploratory Analytical Trends

Rather than relying on rigid predictive hypotheses, this corpus-assisted discourse study is guided by exploratory analytical trends. These trends identify the direction of inquiry without assuming in advance that specific results must appear.

First, the analysis examines whether Vision 2030-related meanings emerge through direct policy citation, through localized academic lexis, or through both. Second, it compares Web and PDF texts to determine whether outward-facing program discourse differs from more internally oriented regulatory and governance discourse. Third, it traces how development- and sustainability-related lexical fields cluster across thematic axes while allowing the corpus evidence to determine which axes become most salient. Fourth, it evaluates subcorpus differences through normalized frequency, document coverage, exploratory inferential comparison, keyword-informed profiling, and qualitative concordance evidence.

3. Methodology

3.1 Research Design

The study adopts a corpus-assisted discourse-analytic design that combines quantitative lexical mapping with qualitative interpretation of institutional discourse. The design is appropriate because the research problem concerns both distribution - how often and where particular lexical fields occur - and discourse function - how those lexical fields construct alignment, legitimacy, and educational value.

Corpus statistics are therefore not treated as self-explanatory. Frequencies, coverage, and inferential comparisons are used as entry points into interpretation, while concordance reading and short excerpts explain how recurrent lexical patterns function in context. This integration follows the logic of CADS, in which quantitative patterning and qualitative interpretation are mutually informing.



3.2 Study Population and Corpus

The corpus consists of 60 official English-language sources produced by Qassim University and collected from its authorized digital environment. The corpus was divided into 30 Web sources and 30 PDF sources in order to compare public-facing, program-oriented discourse with formal documentary, regulatory, and quality-related discourse. The sources include program pages, departmental and college pages, deanship or institutional pages, strategic materials, regulations, handbooks, guidance documents, and quality-related PDF documents.

The dataset is intentionally treated as a specialized institutional micro-corpus. This designation acknowledges its modest size while clarifying its methodological purpose. The study seeks a contextualized and functionally coherent analysis of a policy-relevant institutional environment rather than broad statistical representation (Koester, 2010; Leedham et al., 2021; Vaughan and Clancy, 2013). The bounded corpus allows discursive nuances specific to Qassim University's communicative ecosystem to become visible. Table 1 provides the quantitative profile of the standardized corpus by text type.

Table 1.

Quantitative Profile of the Standardized Corpus by Text Type

Text Type	Number of Documents	of Total Words in Standardized Texts	Extracted	Mean Words	Median Words
PDF	30	3,639		121.30	106.50
Web	30	3,154		105.13	105.00

It is important to note that the word counts presented in Table 1 refer to the standardized, cleaned, and analytically extracted texts used for corpus comparison, not to the full typographical length of the original webpages or PDF files in their raw digital form. This procedure was adopted in order to improve comparability across heterogeneous institutional formats and to focus the analysis on linguistically relevant content.

A source-level genre breakdown was also documented in the corpus log. The Web subset includes program pages, departmental and college pages, and deanship or institutional pages, while the PDF subset includes strategic materials, regulations, handbooks, guidance documents, and quality-related PDF documents. The quantitative comparison retains the Web/PDF distinction because text type is the principal analytical variable used in the corpus-assisted comparison; subtype details are therefore treated as supporting corpus-log metadata rather than a separate statistical grouping.

3.3 Sampling Strategy and Selection Criteria

The study uses purposive criterion-based sampling. Texts were selected because they were official, English-language, institutionally relevant, and analytically connected to curriculum representation, website



discourse, strategic communication, quality, regulation, development, sustainability, innovation, governance, or partnership.

1. The source had to be officially published through Qassim University's authorized digital platforms.
2. The text had to be available in English as a primary or independent version suitable for linguistic analysis.
3. The source had to be directly relevant to one or more of the following domains: programme representation, departmental or institutional identity, quality, strategic planning, regulation, curriculum-related description, development, sustainability, skills, innovation, governance, or partnership.
4. The source had to fall within the broad time frame targeted by the study.
5. The text had to contain sufficient linguistic material to permit thematic and lexical analysis.

The inclusion and exclusion process was recorded in a corpus log. The log documented the title, text type, unit, year or update date where available, URL, inclusion status, and rationale. Its purpose was to make construction of the micro-corpus transparent and inspectable.

3.4 Data Collection and Corpus Preparation

Data collection and corpus preparation followed a staged workflow. Official English-language sources were first identified and classified as Web or PDF texts. Duplicate, non-substantive, or interface-heavy materials were then excluded. The retained texts were cleaned and standardized for comparison while preserving the substantive institutional wording needed for discourse analysis.

Cleaning involved the removal of non-analytical material such as:

- navigation menus,
- footer repetitions,
- technical interface elements,
- duplicated institutional labeling,
- and non-substantive layout content.

The extracted texts were then standardized and saved as independent corpus files using a consistent naming protocol. Each file was accompanied by metadata specifying:

- document ID,
- title,
- text type,
- unit/college,
- year,



- source link,
- inclusion status,
- and inclusion/exclusion rationale.

Figure 2 summarizes the corpus construction and analytical workflow followed in the study.

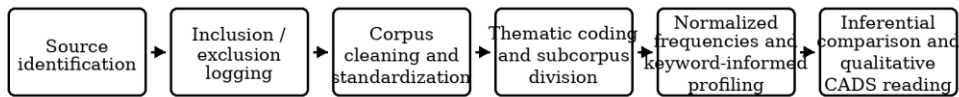


Figure 2. Corpus construction and analytical workflow

Note. The figure summarizes the procedural path from source identification and inclusion logging to corpus cleaning, thematic coding, normalized frequencies, inferential comparison, and CADS-based qualitative interpretation.

3.5 Unit of Analysis

The study operates with more than one analytical unit depending on the level of analysis. At the corpus-comparative level, the unit is the document or page, since each source is treated as one textual instance within either the Web subcorpus or the PDF subcorpus. At the lexical and discourse-analytic level, the units include individual lexical items, multi-word units, recurrent semantic fields, and contextually interpreted concordance evidence.

3.6 Analytical Framework and Thematic Dictionary Construction

For transparency, the study used a bespoke thematically organized dictionary containing lexical indicators, inclusion rules, and overlap-handling notes. The dictionary was not used as an automatic substitute for interpretation. It functioned as a structured guide whose outputs were checked through concordance reading and contextual interpretation, following established qualitative coding principles (Saldana, 2021).

The dictionary was built from three sources: the theoretical literature on institutional discourse and CADS, the developmental vocabulary associated with Saudi Vision 2030, and preliminary inductive reading of the Qassim University texts. This combination allowed the study to retain theoretical relevance while remaining responsive to the actual corpus.

The main thematic axes were Economy, Society, Environment, Governance, Education and Skills, Innovation and Entrepreneurship, and Partnerships. Each axis was linked to lexical indicators and phraseological cues.

- Economy



- Society
- Environment
- Governance
- Education and Skills
- Innovation and Entrepreneurship
- Partnerships

These themes were not selected arbitrarily. They were derived from three converging sources:

1. the theoretical literature on institutional discourse, applied corpus linguistics, and CADS;
2. the developmental vocabulary associated with Saudi Vision 2030;
3. preliminary inductive reading of the university texts included in the corpus.

Each thematic axis was linked to a set of lexical indicators and phraseological cues.

- Education and Skills included items such as education, skills, competencies, curriculum, quality education, and learning outcomes;
- Economy included *labor market, employability, economy, and competitive*;
- Governance included *regulations, policies, standards, and strategic alignment*;
- Sustainability/Environment included *sustainability and sustainable development*;
- Innovation and Entrepreneurship included innovation, technology, and entrepreneurship;
- Partnerships included *partnership, collaboration, and community partnership*.

3.7 Qualitative Discourse-Analytic Procedures

The qualitative component examined how recurrent lexical and thematic patterns worked within their textual surroundings. Concordance lines and short excerpts were used to identify whether a term functioned as program promotion, employability framing, strategic alignment, governance procedure, quality assurance, or developmental positioning. A lexical item was not treated as evidence of recontextualization unless its immediate context connected institutional language to developmental priorities, audience orientation, or genre-specific communicative purpose.

Concordance evidence and short textual excerpts were used to examine how the university represents programmes, skills, quality, innovation, sustainability, governance, and genre-based differences between the Web and PDF subcorpora.

3.8 Corpus-Assisted Quantitative Procedures

The quantitative strand was designed to identify salience, contrast, and thematic distribution within the corpus through a structured set of corpus-assisted procedures.



3.8.1 Data Preparation and Measurement Construction

Because the texts in the corpus vary in length, all lexical and thematic counts were normalized in order to ensure comparability. Normalized frequency was calculated for each relevant item or thematic axis using the following formula:

$$NF = (\text{raw count} / \text{total words}) \times 1,000$$

This procedure made it possible to compare lexical density and thematic prominence across texts of unequal length while minimizing distortion caused by raw frequency alone.

3.8.2 Descriptive Corpus Analysis

The first stage generated descriptive statistics for the corpus and subcorpora.

3.8.3 Thematic Distribution by Text Type

A major analytical goal of the study was to compare the thematic profiles of the Web and PDF subcorpora. To do this, the study calculated normalized thematic density for each document and then compared the resulting distributions across text types.

This made it possible to identify which thematic axes were more strongly associated with the Web subcorpus and which were more closely linked to the PDF subcorpus. These contrasts were treated as analytically meaningful because the two text types serve different communicative and institutional functions.

3.8.4 Publication-Time Distribution

The study also mapped the available distribution of sources across years and examined how thematic concentration relates to the university's published update cycle. The corpus log recorded dated sources from May 2025 to March 2026 where publication or update dates were available; undated strategic sources were retained only when they were active in the same current university digital environment. This component is treated as a map of available institutional publication and update cycles, not as a balanced diachronic corpus. It traces published discursive concentration within the available texts rather than claiming a full historical reconstruction of all university discourse.

3.8.5 Inferential Comparison

To test whether the observed differences between Web and PDF texts were statistically meaningful, the study employed Mann–Whitney U, a non-parametric test appropriate for comparing two independent groups where data distributions may be non-normal and where zero-heavy values are common.

The test was applied to normalized thematic density values at the document level, and results were reported using U values, p-values, and effect size (r).

The use of non-parametric testing is particularly appropriate in institutional corpora of this kind because the documents vary in rhetorical structure, lexical density, and thematic spread. Given the focused



size of the specialized corpus, the inferential step is used cautiously and in an exploratory sense at the document level; accordingly, statistical results are interpreted alongside effect sizes, coverage patterns, and qualitative concordance evidence rather than as stand-alone proof.

3.8.6 Comparative Lexical Profiling (Keyword-Informed)

To identify lexical items that distinguished the two subcorpora, the study carried out a keyword-informed comparative lexical profiling procedure comparing the Web subcorpus and the PDF subcorpus. This step was used to identify words and semantic cues disproportionately associated with one subcorpus rather than the other.

Because the corpus is relatively small and specialized, lexical items were selected for substantive discussion only when they showed a clear directional association with one subcorpus, interpretive relevance to a recurring institutional function, and sufficient statistical salience. Items with log-likelihood values of 5.0 or above were treated as stronger lexical indicators, whereas lower values were retained only as descriptive cues and were interpreted cautiously.

3.8.7 Document Coverage and Dispersion

To avoid overinterpreting items that might be frequent in only one or two documents, the study also considered document coverage, that is, the percentage of documents in which a theme or lexical field appears. This helped distinguish between:

- themes that are highly frequent but narrowly concentrated,
- and themes that are moderately frequent but widely distributed across the corpus.

3.8.8 Secondary Comparison by Academic Unit

In addition to the main comparison between Web and PDF texts, the study included a secondary exploratory comparison across different academic or institutional units where relevant. This step helped determine whether thematic emphasis varied according to program or disciplinary orientation while preserving the primary analytical focus on text type. Table 2 summarizes the statistical procedures and their analytical purposes.

Table 2.

Statistical Procedures and Their Analytical Purposes

Procedure	Analytical Purpose
Raw frequency	To identify the number of occurrences of a lexical item or thematic indicator
Normalized frequency	To compare texts of unequal length
Mann–Whitney U	To test differences between Web and PDF subcorpora
Effect size (r)	To estimate the strength of observed differences
Keyword analysis	To identify lexically distinctive items across subcorpora



Procedure	Analytical Purpose
Document coverage	To assess the distribution of themes across documents
Concordance-based qualitative interpretation	To explain the discourse function of recurrent items

3.9 Tools and Software

The analytical workflow combined close corpus reading with transparent quantitative tabulation. Texts were collected, cleaned, and standardized in document files and spreadsheet logs; concordance retrieval, lexical searching, and keyword-informed inspection were then carried out in AntConc 4.x; normalized frequencies, document coverage, and comparison tables were calculated in Microsoft Excel with supporting manual checks. This combination allowed the study to move from corpus extraction and cleaning to concordance-based checking, frequency normalization, and inferential comparison in a traceable sequence.

3.10 Validity and Reliability

Several measures were adopted to enhance methodological credibility.

First, content validity was strengthened through the careful construction of the thematic dictionary on the basis of theory, Vision 2030-related developmental vocabulary, and preliminary corpus reading.

Second, procedural reliability was supported through standardized extraction, cleaning, naming, and classification procedures. The use of a corpus log and coding log ensured that corpus-building decisions were documented and replicable, while ambiguous cases were rechecked against their immediate co-text before final classification.

Third, interpretive robustness was increased by combining multiple analytical indicators rather than relying on a single measure. The study does not depend solely on frequency counts; instead, it integrates normalized frequency, inferential comparison, coverage, keyword evidence, and qualitative concordance reading.

Where thematic classification involved interpretive judgment, coding decisions were documented systematically in order to reduce arbitrariness and preserve consistency across the analysis. Although formal inter-coder agreement was not calculated, overlap rules were explicitly logged and category decisions were iteratively reviewed to strengthen procedural transparency and coding stability.

3.11 Ethical Considerations

The study is based exclusively on publicly available official university texts. It does not involve human participants, personal data, interviews, or confidential institutional material. Accordingly, the ethical considerations relevant to this research are limited and concern mainly:

- accurate representation of texts,



- faithful citation of institutional materials,
- avoidance of decontextualized quotation,
- and responsible analytical interpretation.

The study does not seek to evaluate the institution. Rather, it examines how developmental priorities are represented within a bounded institutional discourse domain.

4. Results

This section presents the corpus-assisted and discourse-analytic findings. To make the qualitative component transparent, Table 3 presents the initial codes, subthemes, and illustrative corpus evidence that support the thematic interpretation of the corpus. The section then moves to descriptive thematic patterns, publication-time concentration, mean thematic density, document coverage, exploratory inferential comparison, keyword-informed lexical profiling, and discourse excerpts.

Table 3. Initial Codes, Subthemes, and Illustrative Corpus Evidence

Thematic axis	Initial codes	Subthemes	Illustrative corpus evidence and analytical function
Education and Skills	education; skills;	capability formation;	"knowledge, skills, abilities and values"; "in accordance with academic quality standards"
	competencies; curriculum;	academic quality;	- evidence of competence-building and quality-led positioning.
Economy	learning outcomes; quality education	graduate readiness;	
	labor market;	curriculum value	"meet the needs of the labor market";
	employability; competitive;	labor-market alignment;	"requirements of the labor market" - programmes are framed through employability and market responsiveness.
Innovation and Entrepreneurship	knowledge economy;	employability	
	requirements	orientation; economic relevance	Web texts associate innovation and technology with programme goals and future-oriented preparation.
Governance	innovation; technology;	innovation orientation;	
	entrepreneurship;	technology-supported development;	"linked to the program's strategic goals";
Sustainability and Environment	advanced; future-oriented	entrepreneurial readiness	"Quality System of Academic Programs" - PDF texts encode alignment through formal quality systems.
	regulations; policies;	procedural alignment;	
Partnerships and Society	standards; system; strategic goals; quality manual	quality assurance;	
	sustainability; sustainable development;	institutional regulation;	"knowledge economy and sustainable development" - sustainability is linked to institutional planning and programme relevance.
Partnerships and Society	environment; development	strategic governance	
	partnership; collaboration;	sustainability alignment;	
	community; society;	long-term development orientation	Partnership and community-oriented vocabulary frames the university's public responsibility and institutional legitimacy.
	community service	community-facing responsibility; social contribution;	
		institutional partnership	



These codes and subthemes provide the qualitative basis for the descriptive thematic patterns reported in the following section.

4.1 Descriptive Thematic Patterns in the Corpus

Initial corpus-level inspection shows that the thematic profile of the specialized institutional micro-corpus is dominated by a limited set of recurrent developmental and institutional axes. Among these, Education and Skills emerges as the most prominent and widely distributed theme in the corpus as a whole. It is followed, at varying levels of prominence, by Economy, Innovation and Entrepreneurship, Governance, Partnerships, and, to a lesser extent, Environment and Society.

This pattern suggests that the English-language discourse of the university is strongly organized around questions of capability development, programme relevance, quality, and institutional positioning. In other words, the corpus does not simply reproduce general academic language; it reveals a highly patterned institutional discourse in which educational, economic, and strategic priorities are linguistically foregrounded.

From a descriptive perspective, the prominence of these themes supports the assumption that Saudi Vision 2030 enters university discourse less through explicit policy citation than through repeated institutional lexis associated with *competencies, labour market relevance, sustainability, governance, and innovation*.

4.2 Publication-Time Distribution of Sources by Text Type

The publication-time profile shows a strong concentration of available sources in 2025, with fewer texts from earlier years. This distribution reflects the available English-language publication and update cycle in the university's digital environment. It should therefore be interpreted as a map of published discursive concentration, not as a balanced longitudinal archive.

Figure 3 presents the publication-time distribution of the corpus sources by text type.

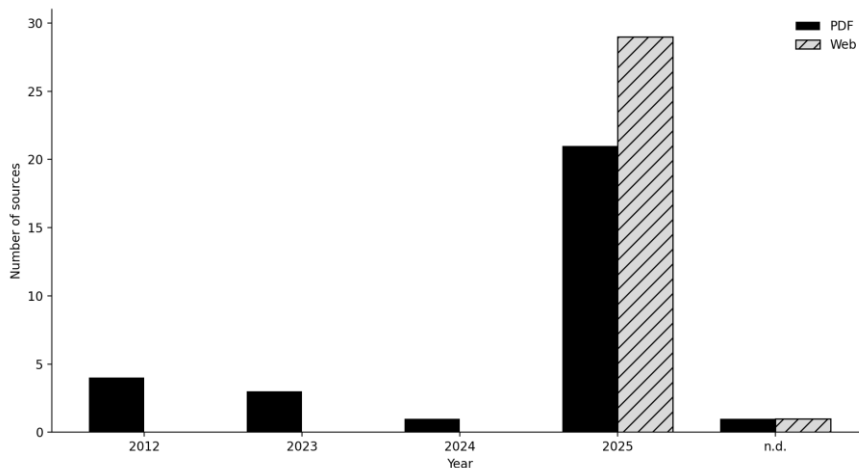


Figure 3. Publication-Time Distribution of Sources by Text Type



Note. Dark bars indicate texts in the PDF subcorpus; patterned light bars indicate texts in the Web subcorpus. The figure captures the published distribution of the standardized corpus rather than a balanced longitudinal archive.

Figure 3 captures the available institutional publication and update cycle, with 2025 representing the most intensive phase of English-language discursive consolidation in the corpus. This pattern is read as a distribution of published discursive concentration rather than as evidence of evenly distributed historical tracking.

This clustering suggests that the lexical prominence of developmental and strategic themes is closely connected to the institution's recent phase of digital self-representation. Attention to publication time helps identify where the current discursive maturity of Vision 2030 alignment becomes most visible.

4.3 Mean Thematic Density by Text Type

Comparison of the two subcorpora reveals substantial differences in mean normalized thematic density across major semantic axes. The overall pattern shows that texts in the Web subcorpus tend to foreground Education and Skills, Economy, and Innovation and Entrepreneurship, whereas texts in the PDF subcorpus show stronger association with Governance and selected strategic-regulatory themes.

Figure 4 shows the mean thematic density of the major axes across the Web and PDF subcorpora.

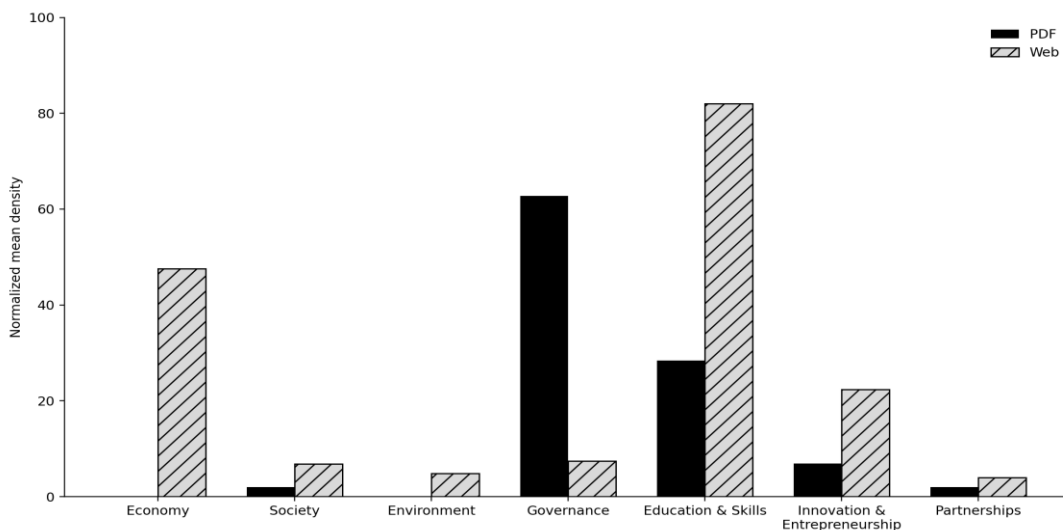


Figure 4. Mean Thematic Density of Major Axes by Text Type

Note. Values represent normalized mean thematic density in the Web and PDF subcorpora. The patterned bars remain distinguishable in grayscale reproduction.



The figure demonstrates that Education and Skills has the highest normalized mean density overall, especially in the Web subcorpus. This suggests that discourse in the Web subcorpus is particularly oriented toward representing programmes in terms of competence-building, curriculum value, and educational relevance. Economy and Innovation and Entrepreneurship also show higher density in the Web subcorpus, which is consistent with the programme-facing and outward-oriented nature of university web communication.

By contrast, Governance is descriptively more prominent in the PDF subcorpus. This reflects the formalized and organizational character of regulations, strategies, guidelines, and institutional procedural texts. In these documents, the language of standards, policy, and institutional order becomes more visible than the more publicly oriented language of programme promotion and employability.

The descriptive contrast across text types supports the broader argument of the study: Saudi Vision 2030 priorities are not represented in a single uniform discourse, but are redistributed according to the communicative function of the institutional genre.

4.4 Document Coverage of Major Axes by Text Type

In order to avoid overinterpreting themes that may be dense in a few texts but absent from many others, the study also examined document coverage, that is, the percentage of documents in which each major thematic axis appears.

Figure 5 reports document coverage by showing the percentage of documents in which each major thematic axis appears at least once.

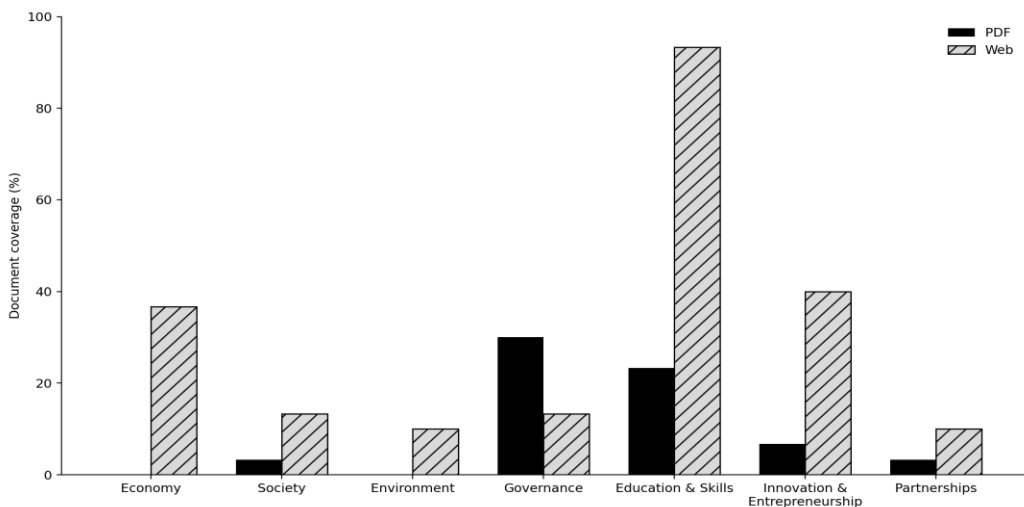


Figure 5. Document Coverage of Major Axes by Text Type



Note. Document coverage indicates the percentage of documents in which each thematic axis appears at least once.

The coverage profile shows that Education and Skills is not only dense but also broadly distributed across the corpus. This indicates that it is a structurally central theme rather than a localized lexical cluster. Similarly, selected developmental themes such as Economy and Innovation and Entrepreneurship appear across a substantial proportion of Web documents, reinforcing the interpretation that these are recurrent institutional priorities rather than isolated textual features.

The distribution of Governance, on the other hand, is more concentrated in the PDF subcorpus. This suggests that governance-related language is tied more closely to formal institutional genres than to programme-facing discourse in the Web subcorpus. Although governance is present in the broader corpus, its distribution is not as widely spread across the Web subcorpus, which helps explain why its institutional profile differs from that of Education and Skills.

Together, density and coverage indicate that the corpus is not merely characterized by thematic salience, but also by patterned thematic dispersion. This distinction is important because it supports the interpretation of some themes as institution-wide discursive tendencies and others as genre-bound lexical concentrations.

4.5 Inferential Comparison Between Web and PDF

To explore whether the observed Web-PDF differences were statistically interpretable at the document level, the study applied Mann-Whitney U to normalized thematic density values. The results are treated as exploratory support and interpreted alongside effect sizes, coverage patterns, and qualitative evidence. Table 4 reports the Mann-Whitney U results for the Web and PDF comparison.

Table 4

Mann-Whitney U Results for Web and PDF Comparison

Theme	Mean W/P	Median W/P	U	p	r	Sig.
Economy	47.53 / 0.00	0.00 / 0.00	615.00	< 0.001	0.315	Sig.
Society	6.80 / 1.96	0.00 / 0.00	494.00	0.180	0.084	Not sig.
Environment	4.81 / 0.00	0.00 / 0.00	495.00	0.082	0.086	Not sig.
Governance	7.46 / 62.78	0.00 / 0.00	358.00	0.060	0.176	Not sig.
Education and Skills	81.83 / 28.32	74.18 / 0.00	723.50	< 0.001	0.522	Sig.
Innovation and Entrepreneurship	22.42 / 6.72	0.00 / 0.00	591.00	0.005	0.269	Sig.
Partnerships	3.76 / 1.96	0.00 / 0.00	478.50	0.338	0.054	Not sig.



The exploratory inferential comparison suggests statistically significant document-level differences in three key thematic axes. First, Education and Skills shows a highly significant difference ($p < 0.001$) in favour of the Web subcorpus, with a relatively strong effect size ($r = 0.522$). This supports the interpretation that programme-oriented discourse in the Web subcorpus foregrounds skills, educational quality, competencies, and related themes more systematically than texts in the PDF subcorpus.

Second, Economy also differs significantly ($p < 0.001$) in favour of the Web subcorpus ($r = 0.315$), indicating that labour-market relevance, competitiveness, and economically oriented educational framing are much more salient in outward-facing university discourse.

Third, Innovation and Entrepreneurship shows a significant difference ($p = 0.005$) again in favour of the Web subcorpus ($r = 0.269$), suggesting that institutional representation in the Web subcorpus is more likely to mobilize innovation language in programme- and public-facing contexts.

By contrast, Governance shows a noticeably higher mean in the PDF subcorpus, but the difference does not reach the conventional threshold for statistical significance ($p = 0.060$). This pattern is therefore interpreted descriptively rather than inferentially. It indicates a tendency for governance language to cluster more strongly in the PDF subcorpus, likely because such language is concentrated in formal regulatory texts; however, the prevalence of zero-heavy thematic values in several PDF variables means that this result should be read as exploratory descriptive support rather than conclusive statistical evidence at $\alpha = 0.05$.

The remaining axes—Society, Environment, and Partnerships—do not show statistically significant differences between Web and PDF texts. Their presence in the corpus is therefore better interpreted as distributed but less sharply differentiated across institutional genres.

4.6 Comparative Lexical Profiling (Keyword-Informed)

Keyword-informed comparative lexical profiling was used to identify lexical items that characterize each subcorpus more strongly. Rather than focusing only on theme-level categories, this procedure highlights the lexical texture through which each institutional genre constructs its priorities. Table 5 presents examples of lexically distinctive items in each subcorpus.

Table 5. Examples of Lexically Distinctive Items in Each Subcorpus

* Indicates exploratory descriptive cues with log-likelihood below 5.0; such items are retained for lexical texture rather than strong statistical distinctiveness.

Lexical item	Log-likelihood	Direction
Economics*	3.95	Web-key
Aims*	3.95	Web-key
National*	3.29	Web-key



Prepare*	3.29	Web-key
Advanced*	3.29	Web-key
Requirements*	2.63	Web-key
Competencies*	2.63	Web-key
Graduates*	2.63	Web-key
Knowledge*	2.63	Web-key
Goals*	2.63	Web-key
Handbook	17.80	PDF-key
Guide	12.71	PDF-key
Student	12.71	PDF-key
System	10.17	PDF-key
Regulations	10.17	PDF-key
Faculty	7.63	PDF-key
Safety	7.63	PDF-key
Descriptions	7.63	PDF-key
Short	7.63	PDF-key
Course	7.63	PDF-key

Note. The table presents lexical contrasts drawn from the standardized corpus. Items with log-likelihood values of 5.0 or above are treated as stronger indicators, whereas lower values are included only as exploratory descriptive cues in a small specialized corpus.

The lexical profile indicates that the Web subcorpus tends to be associated with language emphasizing programmes, competencies, labour-market relevance, educational quality, and employability-related positioning. In contrast, the PDF subcorpus is more likely to foreground terms linked to regulation, policy, standards, and institutional organization. The asymmetry between the two sides is itself analytically meaningful: the PDF profile contains statistically stronger lexical indicators, whereas the Web profile is more diffuse and therefore interpreted more cautiously.

These findings reinforce the interpretation already suggested by the thematic analysis: the two subcorpora participate in the same broad institutional alignment, but they realize that alignment through different lexical and rhetorical pathways.

4.7 Qualitative Evidence of Discursive Recontextualization

A central aim of the study is not only to identify lexical salience but also to show how the priorities of Saudi Vision 2030 are recontextualized in actual institutional discourse. To make the qualitative claims more



transparent, the following short corpus-aligned excerpts illustrate recurrent wording patterns across the two subcorpora.

In programme-oriented Web discourse, developmental priorities are commonly reformulated through employability, quality, and community-facing language. For example, one official programme page states that graduates should “meet the needs of the labor market” (Qassim University, 2026), while the same page also emphasizes education “in accordance with academic quality standards” (Qassim University, 2026). These formulations recontextualize national capability discourse as programme value, graduate readiness, and quality-led academic positioning.

A second Web example from the deanship discourse highlights graduate programmes that “meet the needs of the knowledge economy and sustainable development” and “comply with the requirements of the labor market” (Qassim University, 2025a). Here, development is not framed as abstract policy language, but as a concrete institutional mission linking graduate provision, sustainability, and labour-market alignment.

By contrast, PDF materials tend to recontextualize similar priorities through governance, review cycles, and formal quality mechanisms. In the Quality System of Academic Programs, the discourse refers to a “Periodic survey of the knowledge, skills, abilities and values needed by the labor market” (Qassim University, 2025c). This wording frames labour-market responsiveness as part of an institutionalized quality system rather than as outward-facing promotion.

A parallel governance-oriented pattern appears in the Quality Manual, where the annual operational plan is described as being “linked to the program’s strategic goals” (Qassim University, 2025b). Taken together, these examples show that Web texts tend to translate Vision-related priorities into programme attractiveness and employability, whereas PDF texts encode them through regulation, review, and strategic alignment.

4.8 Reading the Results Through Publication-Time Distribution

The publication-time profile of the corpus does not undermine the analysis; rather, it provides insight into the discursive concentration of the current institutional publication cycle. The prominence of 2025 reflects a period of intensified English-language institutional updating. Accordingly, this pattern should be interpreted as a concentration of the available update cycle rather than as a balanced diachronic archive.

From this perspective, publication-time distribution captures a meaningful institutional moment: the published discursive maturity of alignment with development, skills, innovation, and sustainability within the most active update cycle available in the university's digital environment. This dimension therefore adds interpretive depth by showing where current strategic priorities become most visible.



5. Discussion

5.1 Why Does Education and Skills Emerge So Strongly?

Education and Skills is the most consistent theme in the corpus in both thematic density and document coverage. Its prominence reflects the convergence of the university's core educational function, the developmental logic of Saudi Vision 2030, and the representational demands of institutional English discourse. This finding is consistent with research that links Vision 2030 to human-capital formation, educational reform, and the expanding role of English in Saudi modernization (Almesaar, 2024; Alqahtani, 2024; Barnawi, 2018).

Its significance lies not simply in the recurrence of educational vocabulary, but in the way that vocabulary is framed. Across the corpus, education is repeatedly linked to skills, competencies, quality, labor-market relevance, and program outcomes, suggesting that it is constructed not only as an academic value but also as a developmental and strategic function tied to capability formation and employability.

This pattern extends earlier work on Saudi university representation by showing how policy priorities become visible through ordinary institutional lexis rather than only through explicit Vision 2030 naming (Al-Qahtani, 2021; Alqahtani, 2024). In Qassim University's English discourse, that macro-level orientation is reformulated into recurring language that foregrounds program value, skills development, and educational effectiveness.

Accordingly, the theme functions as a bridge between academic identity and developmental legitimacy: programmes are represented not only in disciplinary terms but also as sites of competence-building, future readiness, and strategic alignment.

5.2 Why Do Web and PDF Texts Differ?

One of the most important findings of the study is the systematic contrast between Web and PDF texts. This difference is not merely lexical; it reflects distinct institutional functions, audience orientations, and genre-specific discourse practices. It also aligns with prior discourse-analytic work showing that university websites and institutional texts construct legitimacy, market responsiveness, and organizational identity (Fairclough, 1993; Mautner, 2005; Saichaie, 2011; Gering et al., 2023).

The Web subcorpus shows stronger representation of Education and Skills, Economy, and Innovation and Entrepreneurship. This pattern is consistent with the outward-facing and programme-oriented role of university websites, which are more likely to foreground employability, competencies, academic attractiveness, and innovation.

The PDF subcorpus, by contrast, shows a descriptively stronger association with Governance and formal institutional organization, although this tendency does not reach conventional statistical significance



in the inferential test. This profile is consistent with the more regulatory and strategic character of PDF texts, which often contain guidelines, policies, regulations, procedural documentation, and planning frameworks.

What emerges is a functional differentiation within the institution's broader alignment with Saudi Vision 2030. Web texts tend to translate developmental priorities into the language of programme value and outward-facing legitimacy, whereas PDF texts more often realize them through regulation, governance, procedural alignment, and institutional structure.

This distinction shows that institutional discourse should not be treated as a single undifferentiated voice. Instead, it is internally organized by genre and communicative purpose, and these internal differences become visible when corpus-assisted comparison is combined with discourse interpretation.

5.3 Recontextualization as the Central Explanatory Mechanism

The results support recontextualization as the study's central interpretive framework. The influence of Saudi Vision 2030 appears less through direct policy quotation than through the institutional reformulation of Vision-related priorities into operational university discourse, a pattern that is consistent with Bernstein's (1990) account of pedagogic discourse and with Linell's (1998) and Wodak's (2000) treatments of meaning transformation across contexts.

This is visible in the movement from macro-level policy values - such as development, sustainability, capability building, and innovation - to localized institutional lexis such as competencies, quality, labor-market relevance, regulations, partnership, and strategic alignment. These lexical and thematic patterns do not merely echo policy language; they re-embed it within genres serving specific institutional purposes.

In this sense, recontextualization is empirically observable rather than merely post hoc interpretation. The same developmental agenda is reformulated differently depending on whether discourse is outward-facing and programme-oriented or formalized and governance-oriented.

From a CADS perspective, this matters because corpus evidence makes recontextualization traceable: normalized frequency, salience, and coverage identify recurring lexical material, while concordance-based reading shows how that material is repurposed in context.

5.4 Institutional Discourse and the Construction of Legitimacy

A broader implication of the findings is that university English educational discourse at Qassim University plays an active role in constructing institutional legitimacy. Across the corpus, the university is presented not only as an academic institution, but as a strategic actor aligned with quality, employability, innovation, governance, and sustainability. This interpretation corresponds with studies of marketization, legitimacy construction, and university self-representation in higher education discourse (Fairclough, 1993; Mautner, 2005; Saichaie, 2011; Gering et al., 2023).



This is evident in both programme pages and formal documents. Web pages do not merely provide curricular information; they position the university as a site of modern relevance, market responsiveness, and developmental contribution. PDF documents, meanwhile, do not merely regulate internal processes; they articulate standards, procedural order, and strategic coherence in ways that reinforce institutional authority.

This dual function is central to contemporary university discourse in contexts of rapid national transformation. Universities are required not only to adapt to policy change, but also to represent themselves as already aligned with it, and English-language discourse becomes one of the means through which that alignment is publicly staged and normalized.

The findings therefore support the broader argument that institutional discourse is not a peripheral communicative layer, but an active component of institutional positioning. Through recurrent lexical choices and genre-specific emphases, the university constructs itself as academically credible, developmentally relevant, and strategically coherent.

5.5 The Significance of the Specialized Institutional Micro-Corpus

The value of the dataset lies in the use of a specialized institutional micro-corpus designed for close examination of a tightly bounded discourse domain.

The corpus focuses on a specific institutional environment, a specific set of text types, and a specific developmental question. This boundedness is methodologically productive because it allows subtle genre-based contrasts and recurrent institutional patterns to become visible within a coherent dataset.

The findings therefore support the methodological value of small specialized corpora when they are constructed transparently, interpreted cautiously, and supplemented with qualitative evidence. In the present case, the corpus is valuable because of its specificity, comparability, and interpretive depth.

5.6 The Publication-Time Dimension as Discursive Concentration Rather Than Historical Balance

The publication-time distribution adds interpretive value by showing where Vision-related institutional discourse becomes most visible within the available update cycle.

The concentration of texts in 2025 suggests a phase of intensified English-language institutional representation rather than a claim of evenly distributed diachronic change.

Accordingly, publication-time mapping helps locate the current representational cycle in which developmental alignment becomes especially visible in university English discourse.



6. Conclusion and Recommendations

This study has shown that Saudi Vision 2030 is not only implemented through policy and planning, but also enacted through university institutional discourse in English. Across the corpus, developmental priorities emerge not primarily through repeated naming of the Vision itself, but through the recurrent recontextualization of education and skills, labour-market relevance, innovation, governance, partnership, and sustainability within genre-specific institutional language.

The findings underscore the centrality of Education and Skills as the dominant thematic axis of the corpus and confirm the analytical importance of distinguishing between the Web subcorpus and the PDF subcorpus. The former tends to foreground programme value, employability, market alignment, and innovation, whereas the latter more strongly encodes governance, standards, strategic organization, and procedural regulation. This contrast demonstrates that the university does not speak with a single undifferentiated institutional voice; rather, its English-language discourse is internally differentiated according to genre, audience, and communicative purpose.

Methodologically, the study demonstrates the usefulness of a specialized institutional micro-corpus for examining a tightly bounded and policy-relevant discourse domain. By combining thematic mapping, exploratory document-level inferential comparison, keyword-informed lexical profiling, document coverage, and qualitative contextual interpretation, the study shows how corpus-assisted discourse studies can generate practically meaningful insights within a specialized corpus design.

The study also contributes to applied corpus linguistics by showing that university English educational discourse can be analysed as a site where educational identity, institutional legitimacy, and developmental alignment are jointly negotiated. In practical terms, the results provide an evidence-based foundation for improving English-language institutional content so that it reflects strategic priorities with greater coherence, clarity, and functional precision.

6.1 Recommendations

Based on the findings, the following recommendations are proposed:

First, universities should review their English-language websites, programme pages, and official documents to ensure closer alignment between linguistic representation and strategic priorities.

Second, institutions should strengthen coherence across genres so that educational quality, employability, sustainability, innovation, and partnership are represented consistently in both public-facing and regulatory texts.

Third, units responsible for translation, quality assurance, web content, and strategic communication should collaborate more closely in shaping precise and developmentally aligned English-language content.



Fourth, English-language institutional texts should reflect strategic priorities through clear, functional, and audience-appropriate wording rather than through declarative statements alone.

Fifth, English-language university educational discourse should be treated as a strategic resource because its quality directly affects the university's external representation. Future research may compare English- and Arabic-language institutional discourse within the same university to examine the degree of cross-language alignment in representing strategic priorities.

6.2 Limitations of the Study

First, the study is based on a specialized institutional micro-corpus of sixty official English-language sources. This corpus was intentionally designed for focused functional analysis rather than broad statistical generalization. Accordingly, the study's added value should be understood primarily in terms of analytical and functional generalization, not wide representational generalization across all higher education discourse.

Second, the corpus is temporally uneven, with a strong concentration of available texts in 2025. This pattern reflects the university's digital publication and update cycle rather than a balanced longitudinal archive. For this reason, the findings linked to publication time should be interpreted as showing the discursive concentration of the available institutional update cycle, not as a fully symmetrical longitudinal history.

Third, the study focuses exclusively on English-language institutional discourse and does not compare it directly with Arabic-language discourse. While this choice is methodologically justified, it means that the analysis does not address possible differences between Arabic and English representations of the same institutional priorities.

Fourth, the study is limited to one university case. Although this allows for focused and coherent analysis, it also means that broader inter-university generalization requires caution. The study does not claim sector-wide generalizability across Saudi higher education; rather, it offers an analytically focused institutional case that can inform future comparative work.

Acknowledgment

The author would like to thank Qassim University, represented by the Deanship of Graduate Studies & Scientific Research, for financial support of this research (QU-ND95-2026).

Appendix A. Corpus Log and Supplementary Verification

The corpus log (Corpus Log, 2025) was prepared as a transparent supplementary record for the 60 official English-language sources analyzed in the study. It documents each source using the following fields: document ID, source title, text type (Web/PDF), academic or administrative unit, year or publication/update date where available, source URL, inclusion status, and inclusion/exclusion rationale.



For submission and technical verification purposes, the full source-level log should accompany the article as supplementary material or be retained by the author for editorial inspection. This appendix records the structure of the log and links it to the methodological procedures reported in Sections 3.2-3.4.

Verification sample of publicly cited sources: Qassim University (2025a) About the Deanship - SRD; Qassim University (2025b) Quality Manual; Qassim University (2025c) Quality System of Academic Programs at Qassim University; Qassim University (2026) English Language and Literature; Qassim University (n.d.) Qassim University Strategic Plan (English version).

The standardized corpus used for comparison contains 60 documents divided equally into 30 Web sources and 30 PDF sources. The word counts reported in Table 1 refer to the cleaned and analytically extracted texts rather than the full typographical length of the original files.

References

- Almesaar, O. A. (2024). English language in Saudi Arabia: Vision 2030 in the historical prism of a clash between cultures. *Frontiers in Communication, 9*, 1205167. <https://doi.org/10.3389/fcomm.2024.1205167>
- Alqahtani, N. (2024). The role of English in implementing the Saudi Vision 2030 programme: A critical discourse analysis [Doctoral dissertation, University of Birmingham]. University of Birmingham eTheses Repository. <https://etheses.bham.ac.uk/id/eprint/14598/>
- Al-Qahtani, A. (2021). Website representations of Saudi universities in Makkah region: A critical discourse analysis approach. *Cogent Arts and Humanities, 8*(1), 1895463. <https://doi.org/10.1080/23311983.2021.1895463>
- Baker, P. (2006). *Using corpora in discourse analysis*. Continuum.
- Barnawi, O. Z. (2018). *Neoliberalism and English language education policies in the Arabian Gulf*. Routledge. <https://doi.org/10.4324/9781315276717>
- Bernstein, B. (1990). *The structuring of pedagogic discourse: Class, codes and control*. Routledge.
- Corpus Log. (2025). *Standardized unified corpus log of Qassim University (60 official sources)* [Unpublished dataset prepared for the present study].
- Fairclough, N. (1993). Critical discourse analysis and the marketization of public discourse: The universities. *Discourse and Society, 4*(2), 133-168. <https://doi.org/10.1177/0957926593004002002>
- Flowerdew, L. (2012). Corpus-based discourse analysis. In J. P. Gee and M. Handford (Eds.), *The Routledge handbook of discourse analysis* (pp. 174-187). Routledge. <https://doi.org/10.4324/9781003035244-11>
- Géring, Z., Tamássy, R., Király, G., and Rakovics, M. (2023). The portrayal of the future as legitimacy construction: Discursive strategies in highly ranked business schools' external communication. *Higher Education, 85*(4), 775-793. <https://doi.org/10.1007/s10734-022-00865-1>
- Gillings, M., Mautner, G., and Baker, P. (2023). *Corpus-assisted discourse studies*. Cambridge University Press. <https://doi.org/10.1017/9781009168144>
- Kang, J.-H. (2008). Institutional translation. In M. Baker and G. Saldanha (Eds.), *Routledge encyclopedia of translation studies* (2nd ed., pp. 141-145). Routledge.



- Koester, A. (2010). Building small specialised corpora. In A. O'Keeffe and M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 66-79). Routledge.
- Koskinen, K. (2008). *Translating institutions: An ethnographic study of EU translation*. St. Jerome.
- Leedham, M., Lillis, T., and Twiner, A. (2021). Creating a corpus of sensitive and hard-to-access texts: Methodological challenges and ethical concerns in the building of the WiSP Corpus. *Applied Corpus Linguistics*, 1(3), 100011. <https://doi.org/10.1016/j.acorp.2021.100011>
- Linell, P. (1998). Discourse across boundaries: On recontextualizations and the blending of voices in professional discourse. *Text*, 18(2), 143-157. <https://doi.org/10.1515/text.1.1998.18.2.143>
- Mautner, G. (2005). The entrepreneurial university: A discursive profile of a higher education buzzword. *Critical Discourse Studies*, 2(2), 95-120. <https://doi.org/10.1080/17405900500283540>
- Partington, A., Duguid, A., and Taylor, C. (2013). *Patterns and meanings in discourse: Theory and practice in corpus-assisted discourse studies (CADS)*. John Benjamins. <https://doi.org/10.1075/scl.55>
- Qassim University. (2025a, May 12). *About the deanship - SRD*. <https://www.qu.edu.sa/en/deanships/srd/about/>
- Qassim University. (2025b). *Quality manual*. <https://www.qu.edu.sa/wp-content/uploads/2025/05/v1/2024-12-12-09-21-58-QUALITY-MANUAL-%281%29.pdf>
- Qassim University. (2025c). *Quality system of academic programs at Qassim University*. <https://www.qu.edu.sa/wp-content/uploads/2025/05/v1/faculty/74/Quality-System-of-Academic-Programs.pdf>
- Qassim University. (2026, March 10). *English language and literature*. <https://www.qu.edu.sa/en/colleges/asc/programs/bachelor-in-english-language-and-literature/>
- Qassim University. (n.d.). *Qassim University strategic plan (English version)*. <https://www.qu.edu.sa/en/colleges/qec/documents/>
- Saichaie, K. (2011). Representation on college and university websites: An approach using critical discourse analysis [Doctoral dissertation, The University of Iowa]. Iowa Research Online. <https://iro.uiowa.edu/esploro/outputs/doctoral/Representation-on-college-and-university-websites/9983776766102771>
- Saldana, J. (2021). *The coding manual for qualitative researchers* (4th ed.). SAGE Publications.
- Saudi Vision 2030. (2016). *Saudi Vision 2030*. <https://www.vision2030.gov.sa/en/overview>
- Schäffner, C., Tcaciuc, L. S., and Tesseur, W. (2014). Translation practices in political institutions: A comparison of national, supranational, and non-governmental organisations. *Perspectives*, 22(4), 493-510.
- Vaughan, E., and Clancy, B. (2013). Small corpora and pragmatics. In J. Romero-Trillo (Ed.), *Yearbook of corpus linguistics and pragmatics 2013: New domains and methodologies* (pp. 53-73). Springer. https://doi.org/10.1007/978-94-007-6250-3_4
- Wodak, R. (2000). Recontextualization and the transformation of meaning: A critical discourse analysis of decision making in EU meetings about employment policies. In S. Sarangi and M. Coulthard (Eds.), *Discourse and social life* (pp. 185-206). Pearson Education.

