INTERNATIONAL JOURNAL OF SOCIETY, CULTURE & LANGUAGE

Discourse Functions and Structural Patterns of the Signalling Noun *Mäsele* in Academic Kazakh

Gulnar A. Sarseke[1*], Samal Ye. Kaliyeva[1], Amangul M. Adilbek[1], Ekaterina V. Iakovleva[2]

[1]L.N. Gumilyov Eurasian National University, Kazakhstan, [2]International Information Technology University, Kazakhstan

Abstract This study investigates the discourse functions and structural patterns of signalling nouns in academic Kazakh. It addresses a gap in research on Kazakh by examining how abstract nouns with context-dependent meanings contribute to the organisation of academic discourse. The analysis draws on a 3.9-million-word corpus of academic Kazakh representing multiple disciplines and genres. Using AntConc, candidate signalling nouns were identified through concordance analysis, and their recurrent structural patterns and discourse functions were examined. Particular attention was given to *mäsele* 'issue', the most frequent signalling noun identified in the corpus. The findings show that *mäsele* functions as a discourse-organising resource, encapsulating established information, anticipating subsequent content, and linking propositions across clauses. Nominal constructions are most frequent, with demonstratives, participles, evaluative modifiers, numerals, postpositions, and modal words contributing to anaphoric and cataphoric relations. The study provides a basis for research on academic Kazakh and has implications for academic literacy and the development of Kazakh for Academic Purposes.

Keywords: *Mäsele, Signalling noun, Academic discourse, Academic Kazakh, Discourse organisation*

[*]**Corresponding Author:**
Gulnar A. Sarseke
sarseke_ga@enu.kz

Received: April 21, 2026
Revised: July 5, 2026
Accepted: August 2, 2026
First Published: August 25, 2026

© 2026 Sarseke, Kaliyeva, Adilbek, and Iakovleva.
This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY).

1. Introduction

Academic discourse relies on linguistic resources that structure information, establish logical relationships, and ensure textual cohesion. Among these, signalling nouns play a particularly important role by connecting ideas within and across clauses and larger discourse units (Flowerdew, 2003). By condensing, summarising, or predicting information, these nouns serve as lexical cues that guide readers through a text's structure. They therefore contribute not only to discourse cohesion but also to the coherence and comprehensibility of academic writing. As explicit indicators of the writer's intended organisation, lexical signals are of considerable importance. They are likely to be among the fundamental means by which readers or listeners correctly decode discourse (Biber, 1985).

https://doi.org/10.22034/ijscl.2026.2089067.4645

Abstract nouns that function to encapsulate, refer to, or organise stretches of discourse have received considerable attention in discourse analysis, corpus linguistics, and academic writing research (Biber et al., 1999; Halliday & Hasan, 1976; Hunston & Francis, 2000; Flowerdew, 1993, 2002, 2003, 2006, 2010; Flowerdew & Forest, 2015; Francis, 1986, 1994; Ivanic, 1991; Schmid, 2000). This phenomenon has been described using a range of terms, including general nouns (Halliday & Hasan, 1976), anaphoric nouns (Francis, 1986), carrier nouns (Ivanic, 1991), labelling nouns (Francis, 1994), shell nouns (Schmid, 2000), and signalling nouns (Flowerdew, 2003; Flowerdew & Forest, 2015). Despite differences in terminology, these approaches share an interest in nouns whose interpretation depends substantially on the surrounding discourse and contribute to the organisation and development of textual information. Recent corpus-based research has further examined these nouns in relation to disciplinary variation, rhetorical organisation and recurrent structural patterns (Casal et al., 2024; Gao et al., 2024; Hyland & Jiang, 2021; Jakub, 2025; Le, 2023; Yang, 2025).

Despite the growing body of research on signalling nouns in various academic contexts, their discourse functions and structural manifestations in Kazakh remain largely unexplored. It should be acknowledged that there has been an increase in interest in academic discourse and corpus-based research within the Kazakh academic community in recent years. Research has focused on the educational content and features of academic texts, as well as the structure of academic writing (Kuzekova, 2025; Ospanov, 2018), and on discourse patterns involving auxiliary verbs and pronouns in Kazakh texts (Sarseke, 2026; Sarseke & Akmagambetova, 2026). However, no studies have yet examined signalling nouns or other linguistic units in Kazakh that perform a signalling function. Given the growing significance of academic Kazakh in education, scholarship, and scientific communication, investigating the discourse features through which academic texts are structured and developed is increasingly important. This study addresses this gap by examining the noun *mäsele* ‘issue’ as an example of one of the most frequently used abstract nouns that performs a signalling function in Kazakh academic discourse. Drawing on a corpus of academic texts from ten disciplines and six genres, the study examines the recurrent structural patterns and discourse functions of *mäsele*, with particular attention to how it establishes links between propositions, encapsulates discourse content, and guides subsequent textual development.

In this study, we adopt the term “signalling noun” following Flowerdew and Forest (2015) to foreground the discourse-organising role of these nouns. We use the term to refer to an abstract noun whose interpretation depends on its relationship to a proposition, event, or stretch of discourse and whose meaning is specified through the surrounding context. Our analysis of *mäsele* therefore focuses on the structural environments in which these discourse functions are realised and on the ways in which the noun contributes to textual cohesion and the development of academic argumentation.

2. Theoretical Framework

Initially, signalling nouns were considered to be closely related to the study of text cohesion and discourse markers. The literature in the area has described them as a cohesive device that shapes the structure of a text. Such nouns have general meanings that are defined by the surrounding context and function mainly as references to previous or subsequent information. In their seminal book, Halliday and Hasan (1976) examined signalling nouns in the context of cohesion. They defined cohesion as the continuity between different parts of a text. They classified cohesion into two categories: grammatical and lexical. The nouns with a signalling function examined in this study were termed “general nouns” and categorised as “lexical cohesion”. Lexical cohesion is created by anaphoric connections between utterances. In this type of cohesion, general nouns refer to a limited number of specific nouns, such as the names of people, places, and facts (Halliday & Hasan, 1976). Examples include *person, man, child, thing, business, move, place, question*, and *idea*, among others. Halliday and Hasan’s study is valuable because it was the first to explore “general nouns” (signalling nouns) in detail, examining how their meaning is determined by context.

Various terms and explanations have been offered in the literature to characterise nouns that we discuss in the study as signalling nouns. Winter (1977) identified some predictive lexical items in written discourse. In particular, Winter examined a class of lexical elements that are typically open-

system terms, such as nouns, verbs, and adjectives, and claimed that they can serve as exponents of a clause relationship, which can have a predictive impact on how written discourse is organised. Winter (1992) uses the term "unspecific nouns". Halliday and Hasan (1976, p. 275) employed the term "general nouns" and described them as "a small set of nouns having generalised reference within the major noun classes". They are primarily interested in the contribution of these nouns to text cohesion. "Anaphoric nouns" (or "A-nouns") is a term coined by Francis (1986). Francis (1986) employs this term to describe nouns that are presented as the given element within a clause containing new information, can be utilised "metadiscursively" within a discourse, and can serve as anaphoric pro-forms. Building upon the research of Winter (1977) and Biber (1985) on lexical signalling, Francis (1986) further reinforces this concept by likening it to signposts. Nouns serve as linguistic signposts, indicating to the reader that specific information is located elsewhere in the text. Schmid (2000) introduces the term "shell noun". This term stems from the realisation that speakers use shell nouns to create conceptual shells for complex information. These shells are conveyed by clauses or larger sections of surrounding discourse (Schmid, 2000, p. 6). Ivanic (1991) prefers the term "carrier nouns". According to Ivanic, the term "carrier" has a dual meaning. The nouns in question "frequently carry a specific meaning within their context in addition to their dictionary meaning" (Ivanic, 1991, p. 95). Conversely, the term "carrier" highlights the affinity with Halliday's Systemic Functional Grammar (Halliday, 1994), in which it refers to subjects in one of two types of clauses that express relational processes.

As we can see, scholars coin different terms and offer various descriptions of their chosen term. While some terms are similar, they are not interchangeable. The reason for these differences is that each scholar has their own theoretical justification. For example, "general nouns" are classified by their lexical meaning. They are nouns with broad meanings, such as *fact, issue, problem, thing,* and *idea*. In contrast, "shell nouns", such as *fact, issue, problem, decision,* and *reason,* are classified by their cognitive and grammatical function. Such nouns encapsulate or package ideas or segments of discourse. "Anaphoric nouns" are classified by their referential function, which refers to a previous context. Like shell nouns, "carrier nouns" package or convey propositional content. "Signalling nouns" perform discourse functions, guiding readers through a text and building its coherence. An abstract noun can be considered a signalling noun if its lexical specification can be found elsewhere in the text. Through its signalling function, an abstract noun organises the text. Signalling nouns are a functional rather than grammatical category that contributes to discourse coherence, occurring in a specific structure (Flowerdew & Forest, 2015).

Signalling nouns are abstract nouns whose meaning is defined by the context. Examples of signalling nouns in English include *problem, reason, method, function, case, question, approach, fact, factor, issue, idea, principle* and *argument,* among others. According to Flowerdew and Forest (2015, p. 21), signalling nouns are "abstract nouns which are non-specific in their meaning when considered in isolation and which are made specific in their meaning by reference to their linguistic context". In other words, the full meaning of a signalling noun is not made clear unless it is used in a specific context. This is why signalling nouns are considered a functional category, as their details can be found elsewhere in the text.

Anaphoric signalling involves nouns that summarise what has been said in the preceding context. In other words, they encapsulate notions or messages that would otherwise be expressed in one or more clauses. In contrast, cataphoric signalling nouns indicate notions or messages that require further explanation in one or more clauses or even paragraphs. They alert the reader that a notion will soon be "unpacked" (Flowerdew & Forest, 2015, p. 1). In anaphoric and cataphoric usages, readers can find the specific meaning of a signalling noun elsewhere in the text.

These examples illustrate that the meanings of signalling nouns are beyond their dictionary semantics; rather, they are determined by the linguistic context in which they appear. In this sense, they can be compared to pronouns, since both perform a referential function, taking their meaning from the surrounding discourse. However, the referential scope of signalling nouns is broad, encompassing clauses or even entire paragraphs. By combining clauses or complex discourse segments into a single word, a signalling noun allows us to condense information that would otherwise be spread over several clauses. They also facilitate the organisation and integration of textual information by establishing

semantic connections between different parts of discourse. Therefore, signalling nouns play an important role in strengthening textual cohesion and organising discourse.

Because the lexical semantics of signalling nouns are inherently context-dependent, they operate on two semantic levels. The first is their dictionary meaning, while the second is the meaning they assume in a particular discourse context. Following Ivanic (1991), I refer to these as *constant meaning* and *variable meaning*, respectively. The constant meaning corresponds to the noun's conventional lexical meaning, whereas the variable meaning is constructed through its interaction with the surrounding discourse. For example, the Kazakh noun *mäsele* 'issue' is defined in the dictionary as "*şeşımın tabudy qajet etetın kelelı ıs-äreket, alda tūrğan mındet*" (a future action that requires a solution, a task ahead) (Sozdikqor, n.d.). In the extract below, *mäsele* retains this conventional lexical meaning while simultaneously acquiring a context-specific interpretation that is derived from the surrounding discourse.

(a)	Endıgı	bır	nazar	audar-yl-atyn	**mäsele**	TŪB-nyñ
	Another	one	attention	pay-PASS-PTCP	issue	TŪB-GEN
	qai	kez-de	öz	qyzmet-ı-n	doğarğandyğ-y	(HIST, diss)
	when	time-LOC	REFL	function-POSS-3-ACC	cease-POSS-3	

'Another **issue** worth paying attention to is when the TNB ceased its activities.'

In this example, the variable meaning of *mäsele* is "when the TNB ceased its activities". Such context-specific meanings cannot be enumerated in a dictionary because they emerge through the nouns' use in a particular discourse context rather than from its conventional lexical semantics. These are therefore transient, discourse-dependent meanings that signalling nouns acquire in use and that vary across contexts.

Researchers consider different features to be fundamental when categorising an abstract noun within the taxonomy of signalling nouns. For instance, Winter (1992) prioritises nouns representing logical or cohesive relations (e.g., *problem* and *fact*). Francis (1986) claims that nouns with thinking or speaking meanings (e.g., *idea* and *argument*) have priority. Schmid (2000) considers abstract nouns whose meanings cannot easily be conveyed by alternative words (e.g., *concept, fact, question, principle, problem/issue*) to be highly important. Flowerdew and Forest (2015) emphasise the importance of the signposting and discourse-organising functions of these nouns. These criteria are fundamental for classifying them as signalling nouns. As this study aims to reveal the discourse functions of signalling nouns in Kazakh academic texts employing a corpus-based approach, we refer more to Flowerdew's theoretical framework that places an emphasis on discourse-organising functions as a core feature of this type of nouns.

3. Methodology

3.1. Research Corpus

The present study employed a corpus-based approach to investigate signalling nouns in academic Kazakh. A corpus-based methodology is particularly well suited to examining the frequency, distribution, and collocational patterns of signalling nouns, while also facilitating qualitative analysis of their contextual use. Corpora provide language data that reflect a speaker's experience of language on a given topic, offering evidence of common patterns in academic texts (Hyland, 2015). Accordingly, the study combines corpus linguistic techniques with manual contextual analysis to identify signalling nouns and examine their semantic and organizational functions in academic discourse in Kazakh.

The analysis is based on a purpose-built corpus of academic Kazakh comprising texts written by native speakers of Kazakh. To ensure broad disciplinary coverage, the corpus includes ten disciplines and six academic genres. Academic genres include journal articles, dissertations, conference papers, textbooks, abstracts, and monographs. When selecting the texts, we considered the representation of genres within each discipline, ensuring that every discipline included at least one example of each genre. This sampling strategy ensured broad genre representation within each discipline, thereby enhancing the

variety of the corpus and the comparability of disciplines within the corpus. Texts were selected from disciplines such as Archaeology and Ethnology (AE), History (HIST), Islamic Studies (ISS), Oriental Studies (ORTS), Philology (PHILG), Philosophy (PHIL), Religious Studies (RELS), Theology (THEO), Turkic Studies (TURKS), and Translation Studies (TS). Texts for the academic corpus were collected according to a set of criteria. First, only publications published within the domestic research avenues were included in the corpus to ensure that the corpus accurately reflects the conventions of academic writing as practised in the Kazakh academic community. Second, all texts should be written by native speakers of Kazakh. This criterion was adopted to minimise the influence of translation and cross-linguistic interference on lexical and discourse patterns. Consequently, translated texts, including Kazakh versions of article abstracts originally written in Russian or other languages, were not selected for the corpus, as such texts may not reflect authentic native academic language use. The next criterion concerned the academic quality and reliability of the data source. To ensure that the corpus consisted of academically validated texts, only peer-reviewed journal articles published in journals included in the List of Journals recommended by the Committee for Quality Assurance of Education and Science of the MES RK were collected. Similarly, dissertations, monographs and textbooks were included only if they had been formally approved for publication by the academic councils of higher education institutions or recognised scientific research organisations. The corpus was also restricted to publications issued between 2000 and 2024. This time span was selected to capture the contemporary stage of academic Kazakh. Restricting the corpus to approximately the last two decades ensures that the linguistic patterns identified in this study are representative of current academic writing practices and enhances the comparability of texts across disciplines. These selection criteria were designed to make the corpus representative, authentic, and methodologically relevant, thereby providing a reliable empirical basis for the analysis of signalling nouns.

The corpus contains 521 texts and a total of 3,905,032 tokens. The size of the texts in each discipline was balanced to facilitate cross-disciplinary comparison. The texts of each discipline consist of approximately 405,000 tokens (ranging from 405,125 to 405,975), while the Theology texts consist of 254,587 tokens (see Table 1). The small size of the Theology texts can be explained by the limited availability of academic texts in this discipline. Theology is a new discipline compared to other disciplines in the corpus. It was recently included in educational programmes in universities in Kazakhstan as a taught discipline. Therefore, available written and published academic texts such as dissertations, monographs, and textbooks are considerably limited. Though articles and abstracts were available to collect (though their numbers are not significant), care was taken to avoid duplicating texts classified under Religious Studies, as this field substantially overlaps with Theology. However, we have not excluded this discipline from the compilation of the corpus with the thought that the analysis of this discipline might reflect some unusual usages of signalling nouns, including *mäsele*. Additionally, since the analyses reported in this study are frequency-normalised rather than dependent on raw frequency counts alone, the small size of Theology texts was considered sufficient for identifying recurrent signalling noun patterns. Nevertheless, this imbalance should be recognised as a limitation when interpreting disciplinary comparisons. The corpus consists of texts published between 2000 and 2024, thereby representing contemporary academic Kazakh.

Corpus analysis was conducted using AntConc (Anthony, n.d.), which was used to retrieve concordance lines, calculate lexical frequencies, and identify collocational patterns of signalling nouns across the dataset. The corpus is morphologically annotated and includes part-of-speech (POS) tagging. Analysed examples in the study are glossed according to the Leipzig Glossing Rules (2015).

Table 1

The Research Corpus

Disciplines	AE	HIST	ISS	ORTS	PHIL	PHILG	RELS	THEO	TURKS	TS	Total
Token count	405657	405593	405625	405504	405975	405567	405522	254587	405125	405877	3905032
Text count	41	33	35	35	48	73	33	127	56	40	521

3.2. Research Design: Identification and Analysis of Signalling Nouns

As Flowerdew and Forest (2015, p. 71) note, the identification of signalling nouns cannot be fully automated "because of the nature of signalling nouns and their specifics". Although the corpus is morphologically annotated and includes POS tagging, it does not provide annotation specifically identifying signalling nouns or their discourse functions. Therefore, the identification and annotation of signalling nouns in the present study were conducted by the two authors as annotators.

A list of candidate abstract nouns that could potentially perform a signalling function in Kazakh academic texts was compiled. The two annotators examined the candidate nouns through the KWIC tool in AntConc and established their raw frequencies in the corpus (Appendix A). The concordances of each candidate were then examined for evidence of a signalling function, with particular attention to pre- and post-modifiers such as *būl* 'this', *sol* 'that', and *mynada* 'in this', as well as numerals. This procedure enabled us to determine whether a noun functioned as a signalling noun and to refine the list of candidates. The frequency data showed that some nouns (e.g., *mänıs, synaq/ğ, perspektiva, argument, boljam/gipoteza, tetık/g*, and *kemşılık/g*) were less frequent, whereas others (e.g., *maqsat, erekşelık/g, qūndylyq/ğ, mäsele, nätije, sebep/b,* and *qyzmet*) occurred more frequently in academic texts. Among these, *mäsele* was selected for detailed analysis because it was the most frequent abstract noun identified as performing a signalling function.

We established the following criterion for identifying a signalling noun: the noun must refer to another proposition, event, or stretch of discourse and therefore have relatively unspecific semantics that require contextual realisation for their interpretation. We employed the following operational definition of a signalling noun: an abstract noun that refers to, labels, or encapsulates information, propositions, arguments, events, or discourse content, thereby helping the writer organise or interpret the surrounding discourse. Guided by the main criterion and operational definition, the two annotators independently categorised occurrences of *mäsele* into four groups: (i) signalling noun; (ii) ordinary lexical noun; (iii) occurrences in a heading or title; and (iv) uncertain or potentially ambiguous case. To assess the reliability of the classification, the two annotators each independently classified a subset of 2,500 occurrences using the four-category coding scheme. The annotators achieved an overall agreement rate of 87%. Disagreements, particularly in potentially ambiguous cases, were subsequently examined in the broader discourse context and resolved through discussion based on the established functional criterion and contextual evidence. Overall, 62% of the occurrences of *mäsele* were classified as signalling uses, while 32% were identified as ordinary lexical uses, 2% occurred in headings or titles, and 4% were classified as uncertain or potentially ambiguous. The occurrences in which *mäsele* functions as a signalling noun were further annotated independently by two annotators and classified into three structural categories: *nominal*, *copular*, and *other*. The inter-annotator agreement was 88,5% for nominal structures, 8.3% for copular structures, and 3.2% for other structure types.

While some signalling uses can be identified relatively readily through structural cues, such as pronominal pre-modifiers, other concordances require examination of the broader discourse context. For potentially borderline cases, the two annotators independently examined the surrounding discourse to determine whether *mäsele* referred to a proposition, event, or stretch of discourse. Where the classification was not immediately clear, the annotators discussed the case and reached a final decision by applying the established functional criterion and considering the available contextual evidence. Cases for which a discourse-organising function could not be established were classified as uncertain and excluded from the final signalling-noun dataset. Cases identified as lexical uses, occurrences in headings or topics, exophoric realisations, or homographs were excluded from the final dataset. Thus, the final set of signalling nouns consisted only of occurrences for which a discourse-organising function could be established through contextual and functional analysis. Following the identification of signalling nouns, the collocation tool in AntConc was used to examine their recurrent pre-modifiers and post-modifiers. This analysis provided evidence for the structural patterns associated with signalling nouns in the corpus.

Due to the agglutinative nature of Kazakh, nouns rarely occur in their basic form and commonly appear with various inflectional affixes. Therefore, the corpus was searched using a wildcard (*) to retrieve morphological variants of the target lexical item. All morphological variants of *mäsele* were

consequently included in the analysis, including *mäseleler* (issue-PL), *mäselesı* (issue-POSS-3SG), *mäselesınıñ* (issue-POSS-3SG-GEN), *mäselege* (issue-DAT), *mäselemen* (issue-INS), *mäselede* (issue-LOC), *mäseleni* (issue-ACC), and *mäselesında* (issue-POSS-3SG-LOC). These morphological variants were annotated according to the same criteria and procedure described above. The concordance and broader discourse context of each form were examined to determine whether it fulfilled a signalling function, and forms classified as ordinary lexical nouns or appearing in headings or titles were excluded from the final signalling-noun dataset.

The specific meaning of a signalling noun, as we noted in the theoretical framework section, emerges from its relationship with the surrounding discourse context. Examples b and c illustrate how the underlined parts of the text specify the meanings of the Kazakh signalling nouns *sebebi* 'reason' and *jait* 'matter/point'. All examples are drawn from the Kazakh academic corpus that has been created for this study. In these and the following examples in the article, a signalling noun is presented in **bold**. The lexical specification of a signalling noun is underlined. The examples are presented with interlinear morpheme-by-morpheme glosses following the Leipzig Glossing Rules (2015). The first three or four lines contain examples in Kazakh, taken from the corpus and transliterated into the Latin script. We used a consistent transliteration scheme for converting Kazakh Cyrillic into Latin script. In particular, Kazakh-specific characters are transliterated as follows: ə - ä, ү - ü, ұ - ū, і - ı, қ - q, ғ – ğ, ө - ö. The same conventions are applied consistently to all linguistic examples discussed in the paper. The last line contains the English translation of the examples:

(b)	O-nyñ	bırınşı	**sebeb-ı**	özge	ūlt	öz	mınez-ı-nde	qai
	3SG-GEN	first	reason-POSS-3SG	other	nation	REFL	character-POSS-3SG-LOC	which
	sapa-ny,	qandai	qasiet-tı	bağala-ityn-y-n	jäne	o-ny	bağalau	täsıl-der-ı-n
	quality-ACC	what	virtue-ACC	value-PTCP-POSS-3-ACC	and	that-ACC	value	way-POSS-3PL-ACC
	qalai	şyğar-atyn-y-n	kör-gı-mız	kel-dı (PHILG, diss)				
	how	get out-PTCP-POSS-3SG-ACC	see-DESID-POSS-1PL	AUX-PAST				

'The first **reason** was that we wanted to see what qualities and virtues other nations value in their own character and how they develop their own ways of evaluating them.'

(c)	Būl	jer-de	öte	mañyzdy	**jait**	būl	bağyt-tar-dy
	This	here-LOC	very	important	point	this	trend-PL-ACC
	batys-ta	dın-dı	ğylymi	tūrğy-da	zertteu-dı	hristiandyq-qa	qyzmet
	West-LOC	religion-ACC	scientific	in terms of-LOC	study-ACC	Christian-DAT	service
	etu-ge	täueldı	qylu-ğa	ūmtyl-ğan	protestanttyq	jäne	katoliktık
	AUX-DAT	subordinate	make-DAT	strive-PTCP	Protestant	and	Catholic
	teolog-tar	qūr-ğan. (RELS, textbook)					

theologist create-
-PL PTCP
'A very important **point** here is that these trends were founded in the West by Protestant and Catholic theologians who sought to subordinate the scientific study of religion to the service of Christianity.'

4. Results

The analysis has shown that signalling nouns tend to appear within specific syntactic structures in academic texts in Kazakh. These structures are common to most of the signalling nouns identified above. By "structure", we mean the combination of a signalling noun with specific pre- and post-modifiers. For example, a signalling noun *mäsele* occurs in the corpus within structures such as *degen + mäsele + boiynşa* 'on the issue of', *osy + mäsele(ler) + jönında* 'about this/these issue/s', *būl mäsele(ler) turaly* 'about this/these issue/s' and *atalğan + mäsele(ler)ge + qatysty* 'regarding the mentioned issue/s' (see more examples in Table 2). In these structures, the noun *mäsele* acts as a signal by combining the prepositional words *degen, osy, būl* and *atalğan* and the postpositional words *boiynşa, jönında, turaly* and *qatysty*. The first four lexemes link the signalling noun to the preceding context and the last four to the subsequent context. These structures can be recognised as patterns of Kazakh academic discourse because, while abstract nouns can fulfil a signalling function on their own, they are much more frequently found within such syntactic structures. The identified structures are given in Tables 2, 3, and 4.

Table 2
*Nominal Constructions of Mäsele**

Construction type	Structural pattern	Frequency
Demonstrative constructions	*būl mäsele (this/these issue/s)*	572
	osy mäsele (this/these issue/s)	296
	sol mäsele (that/those issue/s)	16
	mūndai mäsele (such (an) issue/s)	15
	mynadai mäsele (such (an) issue/s/(an) issue/s like this/these)	14
	ol mäsele (this/these issue/s)	13
	myna mäsele (this issue/s)	12
	osyndai mäsele (such (an) issue/s/(an) issue/s like this/these	11
	sondai mäsele (such (an) issue/s/(an) issue/s like that	1
	ondai mäsele (such (an) issue/s/(an) issue/s like that)	1
Demonstrative/deictic constructions	*mäsele mynada/mynadai/mynalar (issue/s is/are here/such (an) issue/s/ issue/s is/are follows)*	7
Evaluative/adjectival constructions	*özektı mäsele (the most pressing/the most relevant issue/s)*	305
	mañyzdy mäsele (important issue/s)	145
	negızgı mäsele (main issue/s)	121
	(öte, eñ, asa, tym) kürdelı mäsele ((very) complicated issue/s)	104
	basty mäsele (main issue/s)	102
	dauly mäsele (controversial issue/s)	42
	kökeitestı/kökeikestı mäsele (current issue/s)	25
	kelelı mäsele (significant issue/s)	16
	teüitkıldı mäsele (burning issue/s)	14

Construction type	Structural pattern	Frequency
	näzık mäsele (delicate issue/s)	5
	ötkır mäsele (acute issue/s)	4
	tüiındı mäsele (key issue/s)	3
	erekşe mäsele (special issue/s)	1
	polemikalyq (polemic issue/s)	1
Attributive construction	*degen mäsele (the issue/s called/termed .../ the issue/s that is said to be .../ the issue/s referred to as .../*	100
Participial constructions	*atalğan/atalmyş mäsele (mentioned issue/s)*	96
	tuyndağan mäsele (issue/s that arose)	46
	qarastyrylyğan mäsele/ qarastyryp otyrğan mäsele/ qarastyrylyp otyrğan mäsele (considered issue/s)	22
	aitylğan mäsele (raised issue/s)	12
	talqylanğan/talqylanatyn mäsele (discussed issue/s)	11
	kezek küttırmeitın mäsele (urgent issue/s)	9
	qaralğan mäsele (considered issue/s)	4
	şeşılmegen mäsele (unsolved issue/s)	4
Numeral constructions	*bır mäsele/bırınşı mäsele/ekı mäsele/üş mäsele/tört mäsele (first issue/two issues/three issues/four issues)*	122
Postpositional constructions	*turaly mäsele (issue/s about/concerning)*	133
	mäsele qatysty (regarding the issue/s)	103
	mäsele turaly (about the issue/s)	56
	mäsele boiynşa (on the issue/s)	50
	jaily/jaiynda mäsele (issue/s about/concerning)	26
	mäsele jaily/jaiynda (about/regarding the issue/s)	22
	mäsele jönında (about the issue/s)	8
	mäsele bailanysty (related to issue/s)	27
Relational constructions	*mäsele qatarynda (among/in the category/line of issue/s)*	5
	mäsele sanatynda (among/in the category/line of issue/s)	1
	mäsele töñıregında (around/concerning the issue/s)	42

The most frequently occurring pre-modifiers in these structures are the demonstrative pronouns *būl* 'this' and *osy* 'this'. The structure *būl mäsele* 'this issue' is the most common, with 572 occurrences, followed by *osy mäsele* 'this issue', which appears 296 times (see Table 2). Other recurrent patterns include *özektı mäsele* 'current issue' and *turaly/zhaily/jaiynda mäsele* 'issue/s about/concerning' with 305 and 159 occurrences, respectively; *mañyzdy mäsele* 'important issue' and *negızgı mäsele* 'main issue' with 145 and 121 occurrences, respectively; *kürdelı mäsele* 'complex issue' and *basty mäsele* 'main issue' with 104 and 102 occurrences respectively; *degen* mäsele 'the issue/s called/termed…/ the issue/s that is said to be…/ the issue/s referred to as…' and *atalğan/atalmysh mäsele* 'mentioned issue' with 100 and 96 occurrences, respectively and so on. The range of pre-modifiers includes demonstrative pronouns (e.g., *būl, osy, sol, ol*), adjectives (e.g., *özektı, negızgı, kürdelı, basty*), participles (e.g., *degen, atalğan/atalmysh, qarastyrylğan/qarastyrylyp otyrğan*), postpositions (e.g., *turaly/jaily/jaiynda*) as well as numerals and ordinal numerals (e.g., *bır, ekı, bırınşı, ekınşı*).

The distribution of these modifiers also reflects their discourse functions. Pre-modifiers typically establish an anaphoric link by referring back to a previously introduced message, thereby facilitating

thematic continuity across clauses. In contrast, post-modifiers tend to elaborate or specify the signalling noun by introducing new propositional content, contributing to the rhematic development of the discourse while maintaining coherence with the preceding context. From the structural perspective, these constructions can be classified into nominal and copula patterns. The following corpus examples illustrate these structural types:

(1)	Karrami-dıñ Karrami-GEN	negızgı main	pıkır-ı- opinion-POSS-3	Alla Tağala-nyñ God Almighty-GEN	zat-y item-POSS-3
	men	sipat-tar-y-n	jarat-yl-ğan-dar-ğa	ūqsatu-dan (täşbih)	tūr-dy.
	and	characteristics-PL-POSS-3-ACC	create-PASS-PTCP-PL-DAT	imitate-LOC	consist-PST
	Osy	**mäsele**	**boiynşa**	öz	iş-ı-nen
	this	issue	on	REFL	inside-ABL
	Abidi,	Tabni,	Zarini, İsqaqi, Vahidi, Haisani	sekıldı	mäzhab-tar-ğa
	Abidi,	Tabni,	Zarini, İsqaqi, Vahidi, Haisani	like	school-PL-DAT
	böl-ın-dı. divide-PASS-PST	(ISS, monograph)			

"The main opinion of Karrami was to liken the essence and attributes of Allah to His creation (tashbih). **On this issue**, it was divided into schools of thought such as Abidi, Tabni, Zarini, Ishaqi, Wahidi, and Haysani."

(2)	Mūndai Such	özara mutual	kelıspeu-dıñ disagreement-GEN	saldar-y consequence-POSS-3	olar-dy they-ACC	özara mutual
	küpırlık-ke	aiyptasu-ğa	jetele-dı.	**Sol**	**mäsele-ler-dıñ**	iş-ı-nde,
	blasphemy-DAT	accuse-DAT	lead-PST.	That	issue-PL-GEN	among
	«täjsim, incarnation	qidamul qidamul	ğalam» anij	ūğym-dar-y-n concept-PL-POSS-3-ACC	ata-p name-CVB	aitu-ğa mention-DAT
	bol-a-dy. AUX-CVB-3	(RELS, dissertation)				

"The consequences of such disagreements lead them to accuse each other of blasphemy. Among **these issues**, the concepts of "incarnation " and " qidamul anij" can be mentioned."

The variable meaning of the signalling noun *mäsele* is established in the preceding clauses, as illustrated in Examples 1 and 2. The demonstrative pronouns *osy* 'this' and *sol* 'that' perform anaphoric functions by referring back to the preceding discourse. In combination with the signalling noun, these pronouns retrieve and encapsulate the message presented in the preceding clause and carry it forward as a discourse referent into the following clause. In this way, signalling nouns contribute to textual cohesion by establishing an explicit link between previously presented information and subsequent discourse. The postposition *boiynşa* then introduces new information that is related to and elaborates on the discourse content encapsulated by the signalling noun, thereby creating a further link between the established and new messages. For example, in Example 1, the new information *"... öz işınen, Abidi, Tabni, Zarini, İsqaqi, Vahidi, Haisani sekıldı mäzhabtarğa bölındi"* 'it was divided into schools of thought such as Abidi, Tabni, Zarini, Ishaqi, Wahidi, and Haysani' arises from the previously

established *mäsele* and is introduced into the discourse through *boiynşa*. Thus, *boiynşa* not only connects the signalling noun to the following proposition but also facilitates the progression from given information to new information. A similar linking function is performed by postpositions such as *jönınde, turaly, jaily,* and *jaiynda*, which connect signalling nouns to the propositions that follow.

Similar discourse structures are formed by the phrases *atalğan mäsele boiynşa* 'on the mentioned issue', *atalğan mäsele jönınde* 'about the mentioned issue', and *atalğan mäsele töñıregında* 'around the mentioned issue'. In these structures, the participles *atalğan/atalmyş* refer anaphorically to the preceding context and signal that the meaning of *mäsele* is recoverable from the previously established discourse. The postpositional words *boiynşa, jönınde, töñıregında, jaily,* and *jaiynda* then introduce a new re-thematic element and establish a semantic and discourse link between the previously presented message and the information that follows. Thus, the signalling noun *mäsele* functions anaphorically, encapsulating the preceding discourse while simultaneously providing a discourse point of departure for the subsequent proposition. This structure enables the writer to maintain continuity while advancing the discourse by connecting established information with a new message. For example:

(3)	… ış-ı-nde	ejelgı	türkı-ler-de-gı	memlekettılık-tıñ	anyqtama-sy	men
	inside-POSS-3-LOC	ancient	Turk-PL-LOC-DET	statehood-GEN	definition-POSS-3	and
	belgı-ler-ı-n	zerdeleu,	… zertteuşı-ler-dıñ	negızgı	mındet-ter-ı	bolu
	sign-PL-POSS-3-ACC	study	researcher-PL-GEN	main	duty-PL-POSS-3	be
	kerek.	So-nyñ	nätije-sı-nde	baryp	**atalğan**	**mäsele**
	must	that-GEN	result-POSS-3-LOC	go-CVB	mentioned	issue
	jön-ı-nde	ortaq	ūstanym	qalyptas-yp, …	**...**	(HIST, diss)
	about-POSS-3-LOC	common	position	form-CVB	**…**	

"... Therefore, in our opinion, studying the definition and features of statehood among nomadic peoples, including the ancient Turks, should be the main tasks of researchers working on this issue. As a result, a common position **on the mentioned issue** will be formed, and the topic of the formation of a pre-state structure or state by nomadic peoples will find its solution."

In addition to nominal modification and postpositional patterns, *mäsele* occurs in copular and predicative constructions in which its discourse-referential meaning is realised through a predication (see Table 3). These constructions include forms with *bolyp tabylady* 'is/are', *bolyp otyr* 'remain/s', and *bolyp sanalady* 'is/are considered', among others. Compared with nominal constructions, these constructions are considerably less frequent in academic texts.

Table 3
*Copular Constructions of Mäsele**

Construction type	Structural pattern	Frequency
Copular/predicative constructions	*mäsele bolyp tabylady (is/are an issue/s)*	67
	mäsele ainalysady/ainalysty/ainalysuda/ainalysqan/ainalysyp jürgen (deals/dealt/is/was dealing with the issue/s)*	54
	mäsele nazar audaru (pay attention to issue/s)*	31
	mäsele bolyp otyr/mäsele bolyp qala bermek/mäsele bolyp qala berdı/mäsele bolyp qaluda/mäsele bolyp qalup otyr (remain/s an issue/s)	23
	mäsele kelgende (when it comes to the issue/s)*	23

Construction type	Structural pattern	Frequency
	mäsele qarastyrady/qarastyrğan (is/are/was/were considered)	15
	mäsele keler bolsaq/keletın bolsaq (as for issue/s/as for issue/s)*	13
	mäsele köterıledı (issue/s is/are raised)	8
	mäsele bolyp sanalady/sanaldy (is/was considered an issue/problem)	7
	mäsele talqylandy/talqylanady (was/were/is/are discussed)	7
	mäsele sanalady (is/are considered an issue/s)	3
	mäsele köñıl audaru (pay attention to issue/s)*	3
	mäsele köñıl bölu (pay attention to issue/s)*	2
	mäsele + bolyp tabylmaidy (is/are not an issue/s)	1

In copular structures, the signalling function of the noun *mäsele* is typically realised within the same clause in which its semantic content is explicitly specified. A further characteristic of these structures is the use of evaluative adjectives as pre-modifiers of the signalling noun. These pre-modifiers encode the writer's or speaker's evaluation of the issue, foregrounding attributes such as its significance, relevance, complexity, or controversial nature. The most frequently attested evaluative pre-modifiers include *erekşe* 'special' (Example 4), *öte kürdelı* 'very complex', *talasty* 'controversial' (Example 5), and *eñ ozektı* 'the most pressing/the most relevant' (Example 6). By evaluating the issue in this way, the pre-modifiers not only characterise the discourse content encapsulated by *mäsele* but also guide the reader's interpretation of the proposition that follows.

(4)

Sonymen qatar	oqytu	üderıs-ı-ne	senzura	maqūlda-ğan	dını
In addition	teaching	process-POSS-3-DAT	censorship	approve-PTCP	religious
oqulyq-tar-dyñ	engız-ıl-u-ı	nazar	audaru-dy	qajet	et-etın
textbook-PL-GEN	introduce-PASS-INF-POSS-3	attention	pay-ACC	require	AUX-PTCP
erekşe	**mäsele**	bol-yp	tab-yl-a-dy.	(HIST, diss)	
special	issue	be-CVB	find-PASS-CVB-3		

"In addition, the introduction of censorship-approved religious textbooks into the educational process is a special **issue** that requires attention."

(5)

Türkıtanu-da	qoman-qypşaq	taipa-lar-y-nyñ	etnogenez-ı	öte	kürdelı
Turkology-LOC	Koman-Kipchak	tribe-PL-POSS-3-GEN	ethnogenesis-POSS-3	very	complex
jäne	talasty	**mäsele**	bol-yp	tab-yl-a-dy.	(HIST, artic)
and	controversial	issue	be-CVB	find-PASS-CVB-3	

"In Turkology, the ethnogenesis of the Koman-Kipchak tribes is a very complex and controversial **issue**."

(6)

Bıraq	basqa	stil-der-de	būl	mäsele	bırqalypty
But	other	style-PL-LOC	this	issue	smooth
şeşım-ı-n	tau-yp	jat-sa,	BAQ	tıl-ı-nde	«at

solution-POSS-3-ACC	find-CVB	AUX-COND	media	language-POSS-3-LOC	name
qoi-yp,	aidar	tağ-u»	erekşe	eñ	bır
put-CVB	heading	wear-INF	special	very	one
özektı	**mäsele**	bol-yp	tab-yl-a-dy.	(TS, diss)	
pressing	issue	be-CVB	find-PASS-CVB-3		

"But while this issue is resolved fairly easily in other styles, in the language of the media, 'naming and labeling' is a particularly pressing **issue**."

The noun *mäsele* also occurs in several structural patterns in Kazakh academic texts, particularly in combination with the negation word *emes* 'is/are not', the pronouns *ol, sol,* and *bul,* and the modal word *bar* 'is/are' (see Table 4).

Table 4
Structural Patterns of Mäsele – Other Structures*

Structural pattern	Frequency
mäsele (comma or dash) ol/bul (issue/s is/are)	45
mäsele emes (issue/s is/are not)	22
mäsele bar (there is/are issue/s)	28
mäsele joq (there is not/are not issue/s)	3

In these patterns, *mäsele* functions as a discourse signal by referring to previously established information and providing a point of departure for the subsequent message (Examples 7–10). For example, in *Ğalymnyñ tūjyrymy boiynşa ... mañyzdy mäsele – ol ... saiacu şekaralardy joiuy* 'According to the scholar, the most important issue ... is the elimination of political boundaries ...', *mäsele* refers to the preceding context while signalling that its specific content is to be provided in the following message. The subsequent proposition, *saiasi şekaralardy joiuy* 'the elimination of political boundaries', specifies what the *mäsele* refers to in this particular context. Thus, *mäsele* performs a central discourse-organising function by encapsulating previously established information as an issue and creating a cohesive link to the subsequent proposition.

(7)	Ğalym-nyñ	tūjyrym-y	boiynşa	islam	örkeniet-ı-nde-gı	asa
	scholar-GEN	statement-POSS-3	on	Islam	civilization-POSS-3-LOC-DET	very
	mañyzdy	**mäsele** –	**ol**	ortalyqtanğan	islam	memleket-ı-nıñ
	important	issue	it	centralised	Islam	state-POSS-3-GEN
	erkın	intelektualdy	qarym-qatynas-tar-ğa	kedergı	bol-ğan	saiasi
	free	intellectual	communication-PL-DAT	obstacle	be-PTCP	political
	şekara-lar-dy	joiuy.	(ISS, article)			
	boundary-PL-ACC	eliminate				

"According to the scholar, the most important **issue** in Islamic civilisation **is** the elimination of political boundaries that have become obstacles to free intellectual communication within a centralised Islamic state."

In the structure formed with *mäsele ... emes*, the signalling noun *mäsele* links two related messages by signalling a shift from a limited interpretation to a more specific and encompassing one. Rather than simply referring to an issue, *mäsele* establishes the discourse topic, while the following *ol* 'it' connects this topic to the subsequent message and directs the reader to a further specification of its content. For example, in Example 8, *mäsele* is initially associated with *tek terminde emes* 'not only in the term', but the subsequent message specifies that the issue also concerns the meaning of the concept denoted by the term. Thus, *mäsele* functions as a discourse connector, linking the initial proposition with the subsequent elaboration and enabling the writer to extend and refine the interpretation of the issue.

(8)	**Mäsele**	tek	termin-de	**emes**,	**ol**	osy
	Issue	only	term-LOC	COP-NEG	it	this
	termin-men	belgıle-n-etın	ūğym-nyñ	mağyna-sy-nda.	(AE, monograph)	
	term-INS	sign-PASS-PTCP	concept-GEN	meaning-POSS-3-LOC		

"The **issue is not** only **in** the term **but also** in the meaning of the concept denoted by this term."

In structures with the modal word *bar*, *mäsele* functions cataphorically, signalling that its specific content will be elaborated in the following discourse. Pre-modifiers such as *tüitkıldı* 'complex' evaluate the issue and prepare the reader for its subsequent specification (Example 9), while numerals such as *bir* 'one' and *eki* 'two' signal an upcoming enumeration of issues and their explanations (Example 10). Thus, *mäsele* serves a text-organising and metadiscursive function by signalling subsequent developments and guiding the reader through the unfolding argument.

(9)	Mäselen,	bır	ğana	«jäne»	tūlğa-sy-nyñ	öz-ı-ne
	For example	one	only	and	form-POSS-3-GEN	REFL-POSS-3-DAT
	qatysty	qanşama	tüitkıldı	**mäsele**	**bar**.	Bū-ğan,
	regarding	many	complex	issue	is	This-DAT
	bırınşıden,	«jäne»	tūlğa-sy-nyñ	gramatikalyq	funksia-sy-nyñ …	(TS, article)
	first	and	form-POSS-3-GEN	grammatical	function-POSS-3-GEN …	

"For example, **there are** many complex **issues** related to the single form "and". This is due, firstly, to the expansion of the grammatical function of the form "and" over time, and, secondly, to the fact that the linguistic nature of conjunctions has not been fully studied to this day."

(10)	Bıraq	mū-nyñ	arğy	jağ-y-nda	tağy	**ekı**
	But	this-GEN	behind	side-POSS-3-LOC	too	two
	mäsele	**bar**.	Bırınşıden,	bız-de	mūnai	öndır-etın,
	issue	there is	First	We-LOC	oil	produce-PTCP
	aiyr-atyn	baza	bolu-y	tıis.	Ekınşıden,	basqa
	extract-PTCP	base	be-POSS-3	must	second	other
	memleket-ter-men	şart-qa	otyr-ğan-ymyz-da	bız	şikı	mūnai-dy … [8, p. 2].
	state-PL-INS	agreement-DAT	AUX-PTCP-POSS-1-LOC	we	crude	oil-ACC

> "But **there are** two other **issues** behind this. First, we must have a base for producing and extracting oil. Second, when we sign agreements with other countries, we agree to transport crude oil [8, p. 2]."

5. Discussion

Mäsele was the most frequent candidate signalling noun in the corpus, and our analysis confirms that it performs significant signalling functions. The findings demonstrate that *mäsele* functions as more than a general abstract noun in Kazakh academic discourse. Its signalling function emerges from its ability to encapsulate previously established information, anticipate subsequent content, and establish links between propositions. The recurrent structural patterns identified in the corpus therefore provide evidence that *mäsele* contributes directly to the organisation and progression of academic discourse. This supports the view of signalling nouns as discourse-organising resources rather than simply lexical items with abstract meanings.

The analysis identified three broad types of structural patterns involving *mäsele*: nominal, copular, and other configurations. Nominal structures are the most frequent and commonly combine *mäsele* with demonstrative pronouns, participles, evaluative modifiers, numerals, and postpositional expressions. These combinations are important because they help determine how the discourse content associated with *mäsele* is presented and interpreted. In particular, demonstrative elements can establish links with previously mentioned information, while postpositional elements connect *mäsele* with the proposition that follows. The structural patterns therefore support both anaphoric and cataphoric discourse relations.

The findings also show that the discourse function of *mäsele* is closely related to textual progression. In anaphoric patterns, *mäsele* packages information already available in the preceding discourse and presents it as an issue that can be further discussed. In cataphoric patterns, by contrast, it signals that the content of the issue will be specified or elaborated in the subsequent discourse. This distinction is particularly important because it demonstrates that *mäsele* does not merely refer to information; it helps determine how that information is introduced and developed across clauses.

The modifiers accompanying *mäsele* further contribute to this discourse-organising function. Evaluative pre-modifiers such as *kürdelı* 'complex', *mañyzdy* 'important', *tüitkıldı* 'complex', and *basty* 'main' encode the writer's assessment of the issue and consequently influence how the following information is interpreted. Numerals such as *eki* 'two', meanwhile, signal that the discourse content associated with *mäsele* will be divided into identifiable units and subsequently elaborated. These elements therefore perform functions that extend beyond modification: they establish reference, express evaluation, and prepare the reader for the development of the argument.

The findings consequently provide evidence that *mäsele* functions as a textual signpost in academic Kazakh writing. It enables writers to summarise information, maintain continuity, evaluate discourse content, predict subsequent explanations, and direct readers through the development of an argument. These functions demonstrate that *mäsele* acts as a discourse-organising tool, enabling writers to manage the flow of information and direct readers through the progression of an argument.

The findings have practical implications for the development of academic writing instruction in Kazakh universities. The recurrent patterns identified in the corpus can be incorporated into academic Kazakh teaching materials as examples of how signalling nouns contribute to cohesion, information development, and the organisation of extended discourse. For example, teaching materials could include short authentic extracts from disciplinary academic texts in which signalling nouns are highlighted and accompanied by activities that require students to (i) identify the signalling noun, (ii) trace the preceding information to which it refers, (iii) examine how the following proposition specifies, elaborates, or evaluates the information signalled by the noun, and (iv) discuss how replacing or omitting the signalling noun affects the organisation and interpretation of the text. A further activity could ask students to analyse the recurring grammatical and lexical patterns associated with signalling nouns identified in this study, including their use with demonstratives, evaluative modifiers, numerals, and postpositional expressions, and then apply these patterns in their own academic writing. For instance, students could be asked to revise a paragraph by replacing repetitive clauses with appropriate signalling-

noun constructions or to write a short argument in which signalling nouns are used to introduce, summarise, and develop previously presented information. Such activities could be incorporated into modules on academic argumentation, paragraph development, cohesion, and discourse organisation within academic writing curricula. In this way, the findings can provide teachers with corpus-informed examples for developing students' awareness of how lexical and grammatical choices, including signalling nouns, contribute to the construction and development of academic discourse in Kazakh.

Future studies could investigate whether other frequent signalling nouns exhibit similar structural patterns and discourse functions to those identified in this study and how their use varies across different disciplines and genres. Such research would contribute to a more comprehensive understanding of the linguistic resources through which academic Kazakh achieves cohesion, manages information flow, and guides readers through complex arguments.

Disclosure Statement

The authors claim no conflict of interest.

Funding

This work was supported by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan under Grant No. AP23488585 "Digital Humanities: Creating a Corpus of Academic Kazakh".

Data Availability

Due to copyright and licensing restrictions, the compiled dataset cannot be made publicly available. Further information about the data sources and the methodology used for data collection is available from the corresponding author upon reasonable request.

Ethics Declaration

This study used publicly available textual data and did not involve human participants, human-subject research, or the collection of personally identifiable information. Therefore, ethical approval was not required.

References

Anthony, L. (n.d.). *Common statistics used in corpus linguistics*. https://laurenceanthony.net/resources/statistics/common_statistics_used_in_corpus_linguistics.pdf

Biber, D. (1985). On the surface of discourse. *Language*, *61*(3), 734-735. https://doi.org/10.2307/414437

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Longman. https://doi.org/10.1075/z.232

Casal, J. E., Zhang, G., Matouq, G., & Alqabba, H. (2024). These results are inconsistent: 'This/these + shell noun' patterns in engineering theses and research articles. *Journal of Second Language Studies*, *7*(2), 320–346. https://doi.org/10.1075/jsls.00031.cas

Flowerdew, J. (1993). An educational, or process, approach to the teaching of professional genres. *ELT Journal*, *47*(4), 305-316. https://doi.org/10.1093/elt/47.4.305

Flowerdew, J. (2002). *Academic discourse* (1st ed.). Routledge. https://doi.org/10.4324/9781315838069

Flowerdew, J. (2003). Signalling nouns in discourse. *English for Specific Purposes*, *22*(4), 329-346. https://doi.org/10.1016/S0889-4906(02)00017-0

Flowerdew, J. (2006). Use of signalling nouns in a learner corpus. *International Journal of Corpus Linguistics*, *11*(3), 345-362. https://doi.org/10.1075/ijcl.11.3.07flo

Flowerdew, J. (2010). Use of signalling nouns across L1 and L2 writer corpora. *International Journal of Corpus Linguistics*, *15*, 36-55. https://doi.org/10.1075/ijcl.15.1.02flo

Flowerdew, J., & Forest, R. W. (2015). *Signalling nouns in English: A corpus-based discourse approach*. Cambridge University Press. https://doi.org/10.1017/CBO9781139135405

Francis, G. (1986). *Anaphoric nouns (Discourse analysis monograph)*. University of Birmingham.
Francis, G. (1994). Labelling discourse: An aspect of nominal-group lexical cohesion. In M. Coulthard (Ed.), *Advances in written text analysis* (pp. 83-101). Routledge.
Gao, H., Lu, C., & Hu, C. (2024). A corpus-based study of signalling nouns in marketing and economics research articles. *Journal of English for Academic Purposes*, *67*, Article 101318. https://doi.org/10.1016/j.jeap.2023.101318

Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). Edward Arnold.
Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Routledge. https://doi.org/10.4324/9781315836010
Hunston, S., & Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English* (Vol. 4). Benjamins. https://doi.org/10.1075/scl.4
Hyland, K. (2015). Corpora and written academic English. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp 292-308). Cambridge University Press. https://doi.org/10.1017/9781009477888.018
Hyland, K., & Jiang, F. K. (2021). Academic naming: Changing patterns of noun use in research writing. *Journal of English for Academic Purposes*, *49*, Article 100942. https://doi.org/10.1177/00754242211019080
Ivanic, R. (1991). Nouns in search of a context. *International Review of Applied Linguistics*, *29*(2), 93–114.
Jakub, N. (2025). A multi-perspective and functional analysis of English abstract nouns in academic texts. *Journal of Language Science and Practice*, *1*(1), 1–6. https://doi.org/10.65343/jlsp.v1i1.75
Kuzekova, Z. (2025). *Akademialyq jazylym* [Academic writing]. Nazarbayev University.
Le, T. N. P. (2023). Text construction through patterns of signalling nouns: Evidence from the discussion-conclusion section of applied linguistics research articles. *Studies about Languages*, *43*, 46–63. https://doi.org/10.5755/j01.sal.1.43.34785
Ospanov, E. T. (2018). *Akademialyq jazylym: Joğary oqu orynd. student. arn. oqu qūraly* [Academic writing: A textbook for university students]. Ekonomika.
Sarseke, G. A. (2026). Akademialyq diskurstağy dep atalatyn sintaksistık qūrylymdy korpus negızında taldau [Corpus-based analysis of the 'dep atalatyn' syntactic structure in academic discourse]. *Bulletin of the Karaganda University*, *1*(121), 59-68.
Sarseke, G. A. & Akmagambetova, B. E. (2026). Akademialyq diskurs: Esımdıkterdıñ qoldanys erekşelıkterı [Academic discourse: Features of pronoun usage]. *Bulletin of Toraighyrov University*, *1*, 487-504
Schmid, H. J. (2000). *English abstract nouns as conceptual shells: From corpus to cognition*. Mouton de Gruyter. https://doi.org/10.1515/9783110808704
Sozdikqor (n.d.). Mäsele. In *Sozdikqor*. Retrieved February 4, 2026, from https://sozdikqor.kz/
The Leipzig Glossing Rules. (2015, May 31). Conventions for interlinear morpheme-by-morpheme glosses. https://www.eva.mpg.de/lingua/resources/glossing-rules.php
Winter, E. O. (1977). A clause-relational approach to English texts: A study of some predictive lexical items in written discourse. *Instructional Science*, *6*(1), 1-92. https://doi.org/10.1007/BF00125597
Winter, E. O. (1992). The notion of unspecific versus specific as one way of analysing the information of a fundraising letter. In W. C. Mann and S. A. Thompson (Eds.), *Discourse description: Diverse linguistic analyses of a fundraising text* (pp. 131-170). Pragmatics & Beyond New Series 16. https://doi.org/10.1075/pbns.16.07win
Yang, J. (2025). *Exploring cohesive functions of signaling nouns in dissertation writing: A corpus-based analysis of dissertations* [Unpublished doctoral dissertation, Iowa State University].

Appendix A

Candidate Signalling Nouns

In Kazakh	In English	Raw frequency	In Kazakh	In English	Raw frequency	In Kazakh	In English	Raw frequency
mäsele	issue	9972	ideia	idea	2726	model	model	940
qyzmet	function	7308	ülgı	pattern	2458	tüsındırme	explanation	860
jağdai	situation	6733	qūbylys	phenomenon	2417	saldar	condequence	812
jüie	system	6062	täsıl	approach/ way	2326	aspekt	aspect	805

sebep/b	reason	5850	dälel	evidence	2276	qiyndyq/g	difficulty	748
maqsat	purpose/ aim	5506	aiyrma/ aiyrmashylyq/ g	difference	2271	derekköz	source	667
pıkır	opinion	5112	özgerıs	change	1941	problema	problem	653
qatynas	relationship	4938	şart	condition	1891	qadam	step	650
qūndylyq/g	value	4693	fakt	fact	1764	ūsynys	suggestion	555
mysal	example	4575	nysan	object	1696	tüiın	crux	532
erekşelık/g	difference	4502	qağidat/ prinsip	principle	1652	boljam/ gipoteza	hypothesis	510
ūğym	concept	4172	qorytyndy	conclusion	1592	tetık/g	mechanism	340
ädıs	method	4054	närse	thing	1500	jait	point	286
közqaras	attitude	3987	proses	process	1410	ūmtylys	attempt	266
nätije	result	3771	faktor	factor	1360	artyqşylyq/g	advantage	249
deñgei	level	3417	şeşım	decision	1295	synaq/g	challenge	202
tüsınık/g	notion	3285	anyqtama	definition	1172	perspektiva	perspective	88
mümkındık/g	possibility	3265	sūraq	question	1115	mänıs	point	75
qūral	tool	3008	jospar	plan	1081	argument	argument	38
yqpal	influence	2912	element	element	1070			
taqyryp/b	topic/ subject	2856	sipattama	description	1016			
äser	effect/ impact	2808	qabılet	ability	972			