Computational Linguistic Analysis of *al-Mawqiẓah* in Ḥadīth Terminology: Mapping Terminological Load, Commentary Priorities, and Textual Memorability

Hasan M. Al-Rizzo*

*The International Institute for Qurʾānic Computing and Islamic Sciences, USA and Mosul Centre for Strategic and Future Studies, Iraq; Email: halrizzo@gmail.com

Received: 2026-06-06; **Accepted**: 2026-06-16; **Published**: 2026-06-30.

Abstract: Concise manuals of *muṣṭalaḥ al-ḥadīth* often compress complex technical judgments into very few words, making it difficult to identify where commentary should begin. This article examines how computational linguistic analysis can assist the reading of al-Dhahabī's *al-Mawqiẓah fī ʿIlm Muṣṭalaḥ al-Ḥadīth* by mapping terminological density, semantic ambiguity, commentary priorities, and memorization-relevant textual features. Using a qualitative descriptive-computational design, the clean Arabic *matn* is processed through tokenization, morphological and syntactic analysis, semantic-field classification, ambiguity detection, phonological observation, and extraction of compound terminological units. All computational outputs are then validated through the technical framework of *muṣṭalaḥ al-ḥadīth*. The study finds, first, that the terminological load of *al-Mawqiẓah* is not random but organized across eight recurrent ḥadīth fields, especially narration/isnād structure and *jarḥ wa-taʿdīl*. Terms such as *ḥasan*, *ṣaḥīḥ*, *ḍaʿīf*, *mursal*, *shādhdh*, *munkar*, and *ʿillah* cannot be understood lexically, because their meaning depends on their technical position within ḥadīth criticism. Second, the study identifies 180 commentary-priority units, including 60 compound units such as *ḥadīth ḥasan*, *ḥasan ṣaḥīḥ*, and *al-shudhūdh wa-l-ʿillah*, which must be explained as conceptual units rather than isolated words. Third, textual memorability depends not merely on word length or frequency, but on the interaction between terminological density and morphological-semantic stability. The article concludes that Arabic NLP cannot replace traditional commentary, but it can provide a disciplined auxiliary layer for locating where explanation, differentiation, and pedagogical emphasis are most needed.

Keywords: *Arabic NLP; computational linguistic analysis; ḥadīth terminology; al-Mawqiẓah; semantic ambiguity; terminological load; textual memorability.*

Abstrak: Matn ringkas *muṣṭalaḥ al-ḥadīth* sering memadatkan penilaian teknis yang kompleks ke dalam kata-kata yang sangat sedikit, sehingga tidak selalu mudah menentukan bagian mana yang perlu didahulukan dalam syarah. Artikel ini mengkaji bagaimana analisis linguistik komputasional dapat membantu pembacaan *al-Mawqiẓah fī ʿIlm Muṣṭalaḥ al-Ḥadīth* karya al-Dhahabī dengan memetakan kepadatan terminologis, ambiguitas semantik, prioritas syarah, dan fitur tekstual yang

Atsar: Journal of Hadith and Contemporary Studies
E-ISSN: 3164-1830

relevan bagi hafalan. Penelitian ini menggunakan desain kualitatif deskriptif-komputasional. Matn Arab yang telah dibersihkan diproses melalui tokenisasi, analisis morfologis dan sintaktis, klasifikasi medan semantik, deteksi ambiguitas, observasi fonologis, dan ekstraksi unit terminologis majemuk. Seluruh hasil komputasional kemudian divalidasi melalui kerangka teknis *muṣṭalaḥ al-ḥadīth*. Studi ini menemukan, pertama, bahwa beban terminologis *al-Mawqiẓah* tidak tersebar secara acak, tetapi tersusun dalam delapan medan hadis yang berulang, terutama struktur periwayatan/isnād dan *jarḥ wa-taʿdīl*. Istilah seperti *ḥasan*, *ṣaḥīḥ*, *ḍaʿīf*, *mursal*, *shādhdh*, *munkar*, dan *ʿillah* tidak cukup dipahami secara leksikal karena maknanya ditentukan oleh posisi teknisnya dalam kritik hadis. Kedua, penelitian ini mengidentifikasi 180 unit prioritas syarah, termasuk 60 unit majemuk seperti *ḥadīth ḥasan*, *ḥasan ṣaḥīḥ*, dan *al-shudhūdh wa-l-ʿillah*, yang harus dijelaskan sebagai unit konseptual, bukan sebagai kata-kata terpisah. Ketiga, keterhafalan teks tidak hanya bergantung pada panjang kata atau frekuensi kemunculan, tetapi pada interaksi antara kepadatan terminologis dan stabilitas morfologis-semantik. Artikel ini menyimpulkan bahwa Arabic NLP tidak dapat menggantikan syarah tradisional, tetapi dapat menjadi lapisan bantu yang disiplin untuk menunjukkan bagian-bagian yang paling membutuhkan penjelasan, pembedaan, dan penekanan pedagogis.

Kata Kunci: *al-Mawqiẓah; ambiguitas semantik; analisis linguistik komputasional; Arabic NLP; beban terminologis; keterhafalan teks; muṣṭalaḥ al-ḥadīth.*

1. Introduction

The central problem addressed in this article is the absence of a precise analytical tool that can identify, before the process of commentary begins, the locations of terminological density and semantic ambiguity in a concise manual of ḥadīth terminology. Al-Dhahabī's *al-Mawqiẓah fī ʿIlm Muṣṭalaḥ al-Ḥadīth* is written in compact expressions that carry layered technical judgments. These expressions refer to the degrees of acceptance and rejection, the structure of isnād, the states of transmitters, the forms of continuity and discontinuity, and the indicators of hidden defects. Yet the handling of this density has often depended almost entirely on the competence and experience of the commentator. What is still lacking is a computationally assisted linguistic map that can help identify the words, compounds, and recurring expressions that require definition, technical distinction, illustration, or contextual preference.

Al-Mawqiẓah should therefore be approached not merely as a short text suitable for digital processing,[1] but as a concise ḥadīth terminology matn shaped by a pedagogical tradition of memorization, explanation, and repeated study.[2] Such matns combine brevity with technical density: a single word may carry a judgment, a classification, an isnād relation, or a critical rank. Terms such

[1]Aḥmad ibn ʿAlī Ibn Ḥajar al-ʿAsqalānī, *Nuzhat al-Naẓar fī Tawḍīḥ Nukhbat al-Fikar fī Muṣṭalaḥ Ahl al-Athar*, ed. Nūr al-Dīn ʿItr, 3rd ed. (Damascus: Maṭbaʿat al-Ṣabāḥ, 1421/2000), 39–40.

[2]Jalāl al-Dīn ʿAbd al-Raḥmān ibn Abī Bakr al-Suyūṭī, *Tadrīb al-Rāwī fī Sharḥ Taqrīb al-Nawāwī*, ed. Abū Qutaybah Naẓar Muḥammad al-Fāryābī, 2 vols. (Riyadh: Dār Ṭaybah, n.d.), 1:58–59.

Atsar: Journal of Hadith and Contemporary Studies
E-ISSN: 3164-1830

as *ṣaḥīḥ, ḥasan, ḍaʿīf, mursal, shādhdh, munkar,* and *ʿillah* cannot be understood through lexical meaning alone, because they function within the technical system of ḥadīth criticism.[3] This tension between memorability and terminological density makes commentary necessary, and it also makes computational linguistic analysis useful as a preliminary tool for identifying where terminological pressure and commentarial attention are most concentrated.[4]

The article therefore asks a specific question: how can computational linguistic analysis reveal the locations of terminological density and semantic ambiguity in *al-Mawqiẓah* and provide a supporting map for ḥadīth commentary, without replacing the commentator or claiming independent authority in technical questions of ḥadīth terminology? The goal is not to automate commentary or make algorithms the arbiter of technical meaning, but to produce a prior layer of discovery that helps the commentator arrange the order of explanation and locate points requiring scholarly attention.

The article proceeds from the working hypothesis that morphological, syntactic, semantic, and phonological analysis, when connected to ḥadīth knowledge, can distinguish general linguistic ambiguity from ambiguity that directly affects commentary, and can identify terms whose terminological load is especially dense, such as *ṣaḥīḥ, ḥasan, ḍaʿīf, mursal, muʿallal, munkar, shādhdh, mudallis,* and *muḍṭarib*. These terms are not difficult because they are lexically unfamiliar — many are common Arabic words — but because they have been relocated into a technical system in which meaning depends on disciplinary position and relation to neighboring terms, not on everyday usage.

Methodologically, the study is descriptive, analytical, and computational: descriptive in explaining the nature of *al-Mawqiẓah* and its relationship to commentary; analytical in examining the matn's terms and compounds through morphological, syntactic, semantic, and phonological layers; and computational in treating the text as a small corpus, building indicators of terminological density, semantic ambiguity, and commentary relevance.[5]

This study does not claim to measure a student's psychological capacity for memorization. Such a claim would require a separate pedagogical and empirical design. Instead, the study examines textual features that may help explain the matn's receptivity to memorization and instruction: phonological length, terminological density, analytical multiplicity, ease of pronunciation, and structural cohesion. Nor does it claim that Arabic natural language processing can settle technical questions in ḥadīth terminology apart from the science itself. The numerical and computational

[3] Al-Suyūṭī, *Tadrīb al-Rāwī*, 1:59, 1:166, 1:276.

[4] Obeid et al., "CAMeL Tools," 7022–32.

[5] Kareem Darwish et al., "A Panoramic Survey of Natural Language Processing in the Arab World," *Communications of the ACM* 64, no. 4 (2021): 72–81, https://doi.org/10.1145/3447735; Ibrahim Bounhas, "On the Usage of a Classical Arabic Corpus as a Language Resource," *ACM Transactions on Asian and Low-Resource Language Information Processing* 18, no. 3 (2019): 1–45, https://doi.org/10.1145/3277591.

Atsar: Journal of Hadith and Contemporary Studies
E-ISSN: 3164-1830

results are only supporting indicators. They must be interpreted through the established meanings, categories, and commentarial practices of *muṣṭalaḥ al-ḥadīth*.

2. Method

Because the article brings together two disciplines that are often taught separately, it is necessary to make the practical workflow explicit. In operational terms, the study does not require the reader to imagine an abstract technological procedure. It proceeds through a concrete sequence of input, processing, output, and validation.

The input is the Arabic matn of *al-Mawqiẓah*. The first task is text cleaning: separating the matn from marginal notes, editor's symbols, page markers, commentarial additions, and bibliographic material that is not part of al-Dhahabī's concise text. This stage is essential because a computational tool cannot automatically know whether a word belongs to the matn, to a footnote, or to a modern editor's apparatus; without it, term frequency, ambiguity level, and commentary priority would be distorted.

The second stage is tokenization and morphological analysis, dividing the text into analyzable units and examining each for its possible root, lexeme, part of speech, and morphological form. Arabic is morphologically rich, and a single surface form may produce several possible analyses — which is why Arabic NLP tools such as CAMeL Tools, MADAMIRA, and Farasa are useful for preprocessing and segmentation, even though none can by itself determine technical ḥadīth meanings.[6] The third stage is terminological tagging: a preliminary lexicon of ḥadīth terminology is built and used to identify terms belonging to the domains of grading, isnād, matn, hidden defects, narrator criticism, transmission continuity, modes of transmission, and corroborative analysis — a controlled list allowing computational outputs to be reweighted by technical importance rather than mere frequency.

The fourth stage is ambiguity detection, identifying words that shift between general lexical meaning and technical ḥadīth meaning. For example, *ḥasan* can mean good or beautiful in ordinary Arabic but functions as a technical degree between *ṣaḥīḥ* and *ḍa'īf* in the matn; *'illah* may mean cause or illness generally, but in ḥadīth criticism refers to a hidden defect impairing authenticity. The same applies to *shādhdh, munkar, mursal, tadlīs,* and *wahm*. The fifth stage is output generation: tables of central terms, critical terms, semantic fields, compound units, commentary priorities, and memorization-relevant features. The output is not a commentary but a map for commentary, telling

[6]Ossama Obeid et al., "CAMeL Tools: An Open Source Python Toolkit for Arabic Natural Language Processing," in *Proceedings of the Twelfth Language Resources and Evaluation Conference* (Marseille: European Language Resources Association, 2020), 7022–32; Abdelrahim Pasha et al., "MADAMIRA," in *Proceedings of the Ninth International Conference on Language Resources and Evaluation* (Reykjavik: European Language Resources Association, 2014), 1094–1101; Ahmed Abdelali et al., "Farasa," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations* (San Diego: Association for Computational Linguistics, 2016), 11–16, https://doi.org/10.18653/v1/N16-3003.

Atsar: Journal of Hadith and Contemporary Studies
E-ISSN: 3164-1830

the commentator which words and compounds require attention and which should be taught and memorized as conceptual units.

The sixth and decisive stage is ḥadīth-terminological validation: outputs are reviewed through the categories of *muṣṭalaḥ al-ḥadīth* and classical/contemporary commentarial knowledge, since the final technical meaning must be determined by context and disciplinary logic, not by the analyzer alone. The full workflow: clean matn → tokenization → morphological/syntactic analysis → terminological tagging → ambiguity detection → commentary-priority mapping → validation through ḥadīth scholarship. This workflow does not introduce a new ruling in ḥadīth terminology, but offers a digital-humanities model for assisting the teaching, commentary, and sequencing of concise terminology matns.

3. Results and Discussion

3.1. al-Mawqiẓah: Its Position and Terminological Features

Al-Mawqiẓah is important for three related reasons. *First*, it was authored by al-Dhahabī, one of the major critics and historians of transmitters in the eighth/fourteenth century. *Second*, it is a concise matn in the science of ḥadīth terminology. *Third*, its expressions do not simply list technical types; they often reveal a critical sensibility in defining boundaries and clarifying fine distinctions. For these reasons, the text is especially suitable for testing the guiding question of this study: can computational linguistic analysis reveal the terminological load of a concise ḥadīth manual and identify the locations where commentary is especially needed?

Al-Mawqiẓah is not a rigid declarative matn, but a text built around delicate distinctions. It is therefore an appropriate object for studying terminological load, understood not merely as the number of terms in the text, but as the amount of critical, classificatory, and relational meaning that a word or expression carries within the science of ḥadīth. The centrality of the text does not derive from word quantity alone, but from the way its terms are distributed across major domains of ḥadīth studies: narration structure, grading, hidden defects, transmitter criticism, continuity and discontinuity, and corroboration. Mere counting is therefore insufficient. The word ḥadīth may occur frequently, but its importance lies in being the axis around which judgments such as ṣaḥīḥ, ḥasan, and ḍaʿīf turn. Likewise, isnād may occur less frequently, yet it remains one of the key terms by terminological load because it determines how transmission, continuity, and transmitter evaluation are understood.

For this reason, the present study reweights computational results according to ḥadīth field and commentarial function rather than frequency alone. *Al-Mawqiẓah* builds knowledge through binary and relational structures, such as *ṣaḥīḥ/ḍaʿīf, sanad/matn, muttaṣil/munqaṭiʿ*, and *shādhdh/maḥfūẓ*. A term therefore appears not only as an isolated word, but also within a field, an opposite, a qualification, an example, and a judgment.

If matns condense knowledge, commentaries reopen what the matn has condensed. The task of the commentator is not limited to lexical explanation; it includes technical clarification, the listing

Atsar: Journal of Hadith and Contemporary Studies
E-ISSN: 3164-1830

of conditions, the differentiation of closely related types, and the connection of a concise expression to its proper place within works of terminology. Computational linguistic analysis does not compete with the commentator. It offers an exploratory layer: if a word has multiple possible roots, belongs to an ambiguous knowledge field, functions as a critical technical term, or appears inside a compound expression, the result is not a ready-made commentary but an indication that this is a location requiring technical clarification.

3.2. Levels of Linguistic Analysis and the Limits of Computation

The analysis employs two complementary levels. The first consists of outputs associated with Arabic NLP tools, especially preprocessing, morphological analysis, syntactic tagging, and lexical information. The second consists of Python-based procedures designed to add semantic, phonological, and pedagogical layers, such as knowledge fields, initial phonological articulation, number of phonemes, ambiguity-set size, and indicators of commentary and memorization.[7]

This layered procedure is necessary because of the nature of *al-Mawqiẓah*. A ḥadīth term does not appear only in a morphological form or a syntactic function; it emerges through the interaction of word, root, meaning, context, and scientific field. *Ḥasan* may be an adjective in morphological analysis, but it functions as a ḥadīth judgment in terminology. *Mursal* may be a passive participle in form, but in technical usage it refers to a type of discontinuity. Relying on morphology alone would identify roots and forms but would not determine whether a word is technical or general. Relying on semantic classification alone would leave many ḥadīth terms outside their proper disciplinary fields. Relying on phonology alone would describe ease of pronunciation but not conceptual burden. The procedure therefore requires a composite layer: morphological and syntactic analysis, classification through ḥadīth knowledge fields, review through a preliminary ḥadīth terminology lexicon, and evaluation of each term's relevance for commentary or memorization.

The preliminary analysis of parts of speech shows a strong presence of nominal forms, which is consistent with the nature of terminology *matns*. Such texts are built around names of types, judgments, and concepts, such as *al-ṣaḥīḥ*, *al-ḥasan*, *al-ḍaʿīf*, *al-mursal*, *al-munqaṭiʿ*, *al-shādhdh*, *al-munkar*, *al-ʿillah*, *al-isnād*, *al-matn*, *al-rāwī*, *al-thiqah*, *al-ḍabṭ*, and *al-ʿadālah*. The dominance of nouns is therefore terminological as much as linguistic: much of the science is formed through classified names for types of transmission and judgments upon them. Verbs function differently, especially formulas of reception such as *ḥaddatha*, *akhbara*, *samiʿa*, *rawā*, and *qāla*, which may indicate transmission, hearing, and continuity rather than functioning as ordinary verbs. Adjectives also enter the core of judgment: words such as *ṣaḥīḥ*, *ḥasan*, *ḍaʿīf*, *shādhdh*, *munkar*, *maḥfūẓ*, and *maʿlūl* may appear grammatically as adjectives, but terminologically they function as judgments within ḥadīth criticism.

Computational analysis is useful for revealing linguistic possibilities, but it is insufficient by itself for determining technical ḥadīth meanings. An algorithm may supply root, form, part of speech, and semantic field, but it cannot determine the critical differences among ḥadīth types without

[7] Obeid et al., "CAMeL Tools," 7022–32; Pasha et al., "MADAMIRA," 1094–1101; Abdelali et al., "Farasa," 11–16.

Atsar: Journal of Hadith and Contemporary Studies
E-ISSN: 3164-1830

disciplinary validation.[8] For this reason, a layer of ḥadīth-terminological validation is indispensable. Critical terms must be reviewed, general classifications must be corrected, and computational outputs must be reweighted according to the needs of commentary.

This point is especially clear in terms such as *ʿillah*, *ḥasan*, *ṣaḥīḥ*, *munkar*, and *shādhdh*. In ordinary Arabic, *ʿillah* may mean cause or illness, but in ḥadīth criticism it refers to a hidden defect that impairs a report. *Ḥasan* may mean good or beautiful, but in *muṣṭalaḥ al-ḥadīth* it functions as a technical grade between *ṣaḥīḥ* and *ḍaʿīf*. Likewise, *shādhdh* and *munkar* cannot be explained merely through their lexical meanings, because their technical meanings depend on contradiction, transmitter status, and the relation of a report to stronger or weaker transmissions. Computational analysis therefore does not make the final judgment; it only points the commentator to locations where technical judgment is required.

3.3. Terminological Load: Central Terms and Ḥadīth Fields

Terminological load is the core of this study. It refers not merely to the number of terms in the matn, but to the amount of technical meaning carried by a word or phrase, determined by its location in the science and its relation to judgments, defects, transmitters, isnād, matn, continuity, and discontinuity. The word *ṣaḥīḥ*, for instance, is not only an adjective: it opens the door to the conditions of acceptance — continuity of the chain, uprightness of transmitters, precision, and absence of *shudhūdh* and *ʿillah*.[9] For this reason, analysis of terminological load must move from the isolated word to three interconnected levels: central terms, ḥadīth fields, and compound terminological units.

Central terms in *al-Mawqiẓah* do not merely reveal frequent words; they reveal the axes around which the science is organized. Some recur as formulas of performance, others as names of technical doors, and some are central not by frequency but by their place in the construction of ḥadīth meaning. The analysis therefore relies on three combined criteria: occurrence in the clean matn, belonging to a ḥadīth field, and need for commentarial orientation.

Table 1. Central Terms in al-Mawqiẓah

Term	Ḥadīth field	Occ.	Terminological load
al-ḥasan	Grading	25	Grade between ṣaḥīḥ and ḍaʿīf requiring technical clarification
al-ḥadīth	Narration structure	23	Basic material around which acceptance and rejection revolve

[8]Aqil M. Azmi, Abdulaziz O. Al-Qabbany, and Amir Hussain, "Computational and Natural Language Processing Based Studies of Hadith Literature: A Survey," *Artificial Intelligence Review* 52, no. 2 (2019): 1369–1414, https://doi.org/10.1007/s10462-019-09692-w.

[9]Al-Dhahabī, *al-Mawqiẓah*, 24.

Atsar: Journal of Hadith and Contemporary Studies
E-ISSN: 3164-1830

Term	Ḥadīth field	Occ.	Terminological load
ḥaddathanā	Transmission and performance	18	Formula of transmission affecting hearing and narration
al-ṣaḥīḥ	Grading	15	Acceptance conditioned by continuity, ʿadālah, ḍabṭ, and absence of defects
thiqah / al-thiqah	Jarḥ wa-taʿdīl	~16	Critical description relating to uprightness and precision
ḍaʿf / al-ḍaʿīf	Grading	~15	Judgment that may concern the report, transmitter, or route
rawā / riwāyah	Transmission	~13	Act and material of narration connected to modes of performance
ḥasan	Grading	13	Word moving between lexical description and technical rank
samiʿtu / samiʿa	Transmission	~12	Indicator of hearing and continuity
isnād	Narration structure	10	Route of narration and site of continuity and transmitter analysis
ṣaḥīḥ	Grading	10	Technical judgment not reducible to lexical soundness
ijāzah	Transmission	7	Mode of reception and transmission
tadlīs	Continuity and discontinuity	6	Critical term affecting acceptance of ʿanʿanah
infarada	Corroboration	6	Entry point into solitary transmission and comparison
al-matn	Narration structure	6	Content of the report in contrast to the chain
mursal	Continuity and discontinuity	5	Form of discontinuity requiring explanation
al-jarḥ	Narrator criticism	5	Evaluation of transmitters for acceptance or rejection

Atsar: Journal of Hadith and Contemporary Studies
E-ISSN: 3164-1830

The table indicates that the terminological load of *al-Mawqiẓah* is distributed across several domains, not limited to grading alone. Although *ṣaḥīḥ, ḥasan,* and *ḍaʿīf* are highly visible, they are accompanied by terms of isnād, transmission, narrator criticism, continuity, discontinuity, and corroboration — a terminological network linking judgment on the report to the structure of its chain, the condition of its transmitters, and the presence or absence of corroboration.

The term *ḥasan* is especially revealing: linguistically familiar and seemingly simple, but technically one of the terms requiring the most careful clarification, historically a locus of discussion.[10] *Ṣaḥīḥ* likewise carries compound load, understood only through the meeting of uprightness, precision, continuity, and freedom from *shudhūdh* and *ʿillah*.[11] *Isnād, sanad,* and *matn* reveal the structure of narration itself: judgment on a ḥadīth cannot be separated from its route, transmitters, and reported content, which is precisely where the commentator must distinguish "a sound ḥadīth" from "a sound isnād" — soundness of the isnād does not always entail soundness of the matn when hidden defect or contradiction is present.

After central terms, the next step is to examine the fields in which they are organized — the disciplinary domain to which a word belongs, since meaning is not complete outside its field. *Shādhdh* is understood within hidden defects and contradiction; *mursal* within continuity and discontinuity; *thiqah* within *jarḥ wa-taʿdīl*. The preliminary ḥadīth lexicon contained 128 entries distributed across eight domains.

Table 2. Distribution of Preliminary Ḥadīth Lexicon Entries across Fields

Ḥadīth field	**Entries**	**Examples**	**Function in the matn**
Narration and isnād structure	20	ḥadīth, isnād, sanad, matn, riwāyah, ṭarīq, athar, rāwī	Defining the material and structure of transmission
Jarḥ wa-taʿdīl and transmitters	19	thiqah, ʿadl, ḍabṭ, jarḥ, taʿdīl, ḥāfiẓ, shaykh, majrūḥ	Explaining the state of transmitters and its effect on acceptance
Transmission and performance	17	ḥaddathanā, akhbaranā, anbaʿa, samiʿtu, rawā, ijāzah	Identifying formulas and modes of reception
Hidden defects and criticism	17	ʿillah, maʿlūl, shādhdh, munkar, wahm, mukhālafah, maḥfūẓ	Detecting critical impairments and indicators
Continuity and discontinuity	15	muttaṣil, mursal, munqaṭiʿ, muʿḍal, mudallis, ʿanʿanah	Determining whether the route is continuous or interrupted

[10]Al-Dhahabī, *al-Mawqiẓah*, 26.

[11]Al-Dhahabī, *al-Mawqiẓah*, 24.

Atsar: Journal of Hadith and Contemporary Studies
E-ISSN: 3164-1830

Ḥadīth field	Entries	Examples	Function in the matn
Grading	14	ṣaḥīḥ, ḥasan, ḍaʿīf, mawḍūʿ, ṣāliḥ, yaṣiḥḥ	Determining ranks of acceptance and rejection
Transmitter classes and men	13	ḥāfiẓ, tābiʿī, ṣaḥābī, ḥuffāẓ, rijāl	Locating transmitters in scholarly and temporal classes
Corroboration, mutābaʿāt, shawāhid	13	tafarrud, gharīb, ʿazīz, mutābaʿah, shāhid, iʿtibār	Connecting a report to parallels and supports

This distribution shows that *al-Mawqiẓah* does not present ḥadīth terminology from a single angle: it gathers the fields upon which ḥadīth criticism depends — narration, transmitters, formulas, defects, continuity, grading, classes, and corroboration — so apparent categories such as *ṣaḥīḥ, ḥasan,* and *ḍaʿīf* do not operate alone but depend on isnād, transmitters, and corroborative evidence. Narration and isnād structure contains the largest number of entries, as expected, since every judgment must relate to the report's material and route. *Jarḥ wa-taʿdīl* is also prominent, since judgment cannot be separated from the state of transmitters — *thiqah* is not general praise but a critical judgment of uprightness and precision in context. Hidden defects and criticism, while not always numerically largest, is among the most important fields by load: *ʿillah, shādhdh,* and *munkar* are among the hardest to explain because they derive meaning from subtle critical relations, not lexical meaning alone.

3.4. Compound Terminological Units

The analysis of terminological load remains incomplete if it stops at isolated words. Much of the meaning of ḥadīth terminology appears in compounds rather than individual terms — *ḥadīth ḥasan* is not merely *ḥadīth* plus *ḥasan* but a unit of judgment; *al-jarḥ wa-l-taʿdīl* cannot be explained by separating its two words without explaining the system of transmitter criticism as a whole. The preparatory analysis therefore extracted recurring compounds or candidates for compound technical units, not to accept every one automatically, but to show that commentary must often move from word explanation to compound explanation.

Table 3. Compound Terminological Units in al-Mawqiẓah

Compound	Occ.	Ḥadīth field	Commentary importance
ḥadīth ḥasan	4	Grading	Judgment unit not explained by separating its words
ḥasan ṣaḥīḥ	4	Grading	Compound needing special explanation in ḥadīth terminology

Compound	Occ.	Ḥadīth field	Commentary importance
al-shudhūdh wa-l-ʿillah	3	Hidden defects	Combines two impairing conditions affecting judgment
mā infarada	3	Corroboration	Entry into solitary transmission and comparison
infarada bih	3	Corroboration	Indicates a location of critical review
al-jarḥ wa-l-taʿdīl	2	Narrator criticism	Discipline of accepting and rejecting transmitters
marātib al-ṣaḥīḥ	2	Grading	Indicates that ṣaḥīḥ is not one undifferentiated rank
darajat al-ṣiḥḥah	2	Grading	Connects soundness with degree and rank
rutbat al-ṣaḥīḥ	2	Grading	Opens the issue of variation within acceptance
marātib al-ḥasan	2	Grading	Indicates gradation of ḥasan and its relation to ṣaḥīḥ and ḍaʿīf
tābiʿī kabīr	2	Transmitter classes	A class term affecting mursal and discontinuity
lam yaṣiḥḥ	2	Grading	Negative judgment formula not understood from yaṣiḥḥ alone

The table clarifies that many ḥadīth meanings appear in compounds. *Ḥadīth ḥasan* is not explained by saying *ḥadīth* is the noun and *ḥasan* the adjective; it is explained as a rank of judgment — syntactic analysis can describe the adjective-noun relation, but does not by itself show the compound has become a technical unit. *Ḥasan ṣaḥīḥ* is one of the clearest examples: each word has its own meaning, but their combination opens a special issue, particularly in al-Tirmidhī's usage, treated by works of terminology as requiring explanation rather than as a casual combination. *Al-shudhūdh wa-l-ʿillah* is among the deepest critical compounds, since both terms are connected to the conditions of acceptance. Compounds such as *mā infarada* and *infarada bih* belong to comparison and corroboration: solitary transmission is neither automatically rejected nor accepted, its effect depending on the transmitter's state, the location of the report, and whether contradiction or corroboration is present — so the commentator should connect *infarada* to *tafarrud, gharābah,* and *iʿtibār* rather than explaining it only lexically.

From the perspective of memorization, extracting compounds is important because students do not memorize only individual words; they memorize technical units. Memorizing *ḥadīth ḥasan* as a unit is different from memorizing *ḥadīth* and then *ḥasan* separately. The first carries a judgment; the second may remain suspended between lexical and technical meaning.

Atsar: Journal of Hadith and Contemporary Studies
E-ISSN: 3164-1830

3.5. Semantic Ambiguity and Commentary Priorities

After establishing the terminological density of *al-Mawqiẓah*, the study moves to a more precise level: semantic and technical ambiguity. Ambiguity here does not mean that al-Dhahabī's expression is defective; it refers to the plurality of possibilities that arises when an Arabic word moves from general linguistic usage to specialized ḥadīth usage. A word may be clear in ordinary Arabic but require clarification once it becomes a term, or clear inside one field but ambiguous when removed from context.

This is why computational analysis must be joined with ḥadīth validation. Computational analysis reveals multiple roots, morphological forms, fields, and candidate meanings; ḥadīth validation determines whether this plurality is relevant for commentary. A particle such as *fī* may be highly frequent and analytically complex, yet not require independent technical commentary, while a term such as *ḥasan* or *mursal* may occur less often but be highly consequential. Several levels of ambiguity intersect — morphology, lexicon, syntax, semantics, context, and terminology — arising from multiple roots, multiple lexemes, multiple fields, the difference between lexical and technical meaning, dependence on neighboring words, or meaning that only appears inside a fixed compound.

The most important form for this study is technical ambiguity, since it affects understanding of the science itself. *Ḥasan* in the matn is not merely an adjective but a judgment; *shādhdh* is not simply what is unusual but a technical term connected to contradiction and solitary transmission. Al-Dhahabī defines *shādhdh* in relation to contradiction of reliable transmitters, and *munkar* in relation to the solitary transmission of a weak transmitter — a distinction that cannot be determined by lexical meaning alone, but depends on the state of the transmitter and the location of the solitary report.[12] Computational analysis does not produce commentary; it identifies the type of problem, so the explanation can move accordingly — to root and form if morphological, to meaning if lexical, to terminology if technical, to the phrase as a unit if compound.

Critical terms are words that should not be passed over quickly in commentary because they stand at the transition between language and terminology, between more than one ḥadīth field, or between more than one analytical possibility. The preparatory analysis showed that certain terms should receive priority not because they are lexically strange, but because they affect understanding of the matn. Examples include *ḥasan, ṣaḥīḥ, ḍa'īf, mursal, shādhdh, munkar, 'illah, mu'allal, isnād, sanad, tadlīs, wahm,* and *athar*. These terms require attention because they belong to the domains of grading, hidden defects, continuity and discontinuity, narration structure, and narrator criticism.

These examples clarify the practical meaning of orienting commentators: it places before the commentator those locations that should not be left without technical clarification. *Ḥasan*, for example, requires extended explanation because al-Dhahabī himself notes the difficulty of defining it precisely, placing it above *ḍa'īf* and below *ṣaḥīḥ*.[13] *Ṣaḥīḥ* requires explanation of its conditions, not merely translation as soundness, since its technical meaning gathers transmitter reliability,

[12]Al-Dhahabī, *al-Mawqiẓah*, 42.

[13]Al-Dhahabī, *al-Mawqiẓah*, 26.

Atsar: Journal of Hadith and Contemporary Studies
E-ISSN: 3164-1830

precision, continuity, and freedom from contradiction and hidden defect.[14] *Shādhdh* and *munkar* remain the clearest examples of technical ambiguity unresolved by lexical meaning: a reader might assume *shudhūdh* means simple strangeness, but al-Dhahabī connects it to contradiction of reliable transmitters, while *munkar* is tied specifically to the solitary report of a weak transmitter — a distinction the commentator must explain relationally, not in isolation.[15]

The study culminates in a commentary-priority table containing 180 units — individual words, two-word compounds, three-word compounds, and conceptual technical units — linking structure type, ḥadīth field, type of difficulty, and suggested explanatory action (define, differentiate, illustrate, compare, state the condition, or explain as a unit). This step transforms linguistic analysis into ḥadīth work: where processing shows a word is analytically ambiguous or a compound recurrent, the commentarial decision translates that into scholarly action.

Table 4. General Structure of the Commentary-Priority Table

Dimension	Classification	No.	Significance
Structure type	Individual word	120	Most units require individual or technical review
Structure type	Candidate two-word compound	45	Indicates the need to analyze expressions, not words alone
Structure type	Candidate three-word compound	10	Reveals broader conceptual expressions
Structure type	Technical/conceptual compound	5	Units that should not be split in commentary
Commentary priority	Very high	39	Locations requiring extended commentary
Commentary priority	High	81	Locations requiring clear medium or extended explanation
Commentary priority	Explain as conceptual unit	60	Compounds that should be explained as units rather than isolated words
Total	—	180	Guide for the commentator's work

The table shows that most orientation units are individual words, which is expected in a concise *matn*. However, the presence of 45 two-word compounds, 10 three-word compounds, and 5

[14]Al-Dhahabī, *al-Mawqiẓah*, 24.

[15]Al-Dhahabī, *al-Mawqiẓah*, 42.

Atsar: Journal of Hadith and Contemporary Studies
E-ISSN: 3164-1830

technical conceptual units demonstrates that word-by-word analysis is insufficient. *Al-Mawqiẓah* often operates through compound expressions such as *ḥadīth ḥasan*, *ḥasan ṣaḥīḥ*, *al-shudhūdh wa-l-ʿillah*, and *al-jarḥ wa-l-taʿdīl*. The 39 very-high-priority and 81 high-priority units indicate many locations that commentators should not pass over quickly, while the 60 units requiring unit-level explanation confirm that splitting compounds into isolated words may weaken their technical meaning. For example, *ḥasan ṣaḥīḥ* is not simply the combination of two grading terms, but an expression with a specific history of technical usage; likewise, *al-shudhūdh wa-l-ʿillah* is not ordinary coordination, but a compound expression gathering two impairing conditions in the judgment of ḥadīth.

This finding also has implications for textual memorability. In this study, textual memorability does not mean measuring students' psychological capacity for memorization. It refers instead to textual features in *al-Mawqiẓah* that may affect how the *matn* is taught, reviewed, and memorized. These features include phonological length, morphological stability, semantic clarity, compound structure, and terminological density. A word may be short and easy to pronounce, yet conceptually difficult because it carries a dense technical meaning within *muṣṭalaḥ al-ḥadīth*. Terms such as *ḥasan*, *wahm*, *mursal*, and *ʿillah* are therefore not difficult primarily because of pronunciation, but because they must be placed accurately within the fields of grading, hidden defects, continuity and discontinuity, or transmitter criticism.

The pedagogical implication is that memorizing *al-Mawqiẓah* should not be reduced to verbal repetition. Students need to memorize the *matn* together with its technical structure: which words function as judgments, which terms belong to particular ḥadīth fields, and which compounds must be understood as single conceptual units. Computational linguistic analysis can support this process by helping teachers identify which parts of the *matn* require explanation before memorization, which terms need comparison with related concepts, and which expressions should be emphasized as stable technical units.

4. Conclusion

This study concludes that *al-Mawqiẓah* should not be treated merely as a short manual of *muṣṭalaḥ al-ḥadīth*, but as a dense terminological text in which few words carry complex judgments on grading, narration structure, hidden defects, continuity and discontinuity, narrator criticism, and levels of acceptance and rejection. Its brevity does not indicate simplicity. Rather, the concision of the text increases its terminological density and makes certain words and compounds require careful explanation before they can be properly understood, taught, or memorized.

The main findings of this study show that the terminological load of *al-Mawqiẓah* is not randomly distributed, but organized across clear ḥadīth fields. Terms such as *ḥasan*, *ṣaḥīḥ*, *ḍaʿīf*, *mursal*, *shādhdh*, *munkar*, and *ʿillah* cannot be explained through lexical meaning alone, because their meanings depend on their technical position within the science of ḥadīth criticism. The study also shows that many important meanings are carried by compound expressions, such as *ḥadīth ḥasan*,

ḥasan ṣaḥīḥ, *al-shudhūdh wa-l-ʿillah*, and *al-jarḥ wa-l-taʿdīl*. These compounds need to be explained as conceptual units, not as separate adjacent words.

The study further demonstrates that computational linguistic analysis can provide a useful auxiliary layer for reading concise ḥadīth terminology texts. It can help identify where terminological density is concentrated, where semantic ambiguity appears, which compounds require unit-level explanation, and which parts of the text need stronger pedagogical attention. However, Arabic NLP cannot replace traditional commentary or independently determine technical meanings. Its value lies in assisting the commentator, not substituting the authority of *muṣṭalaḥ al-ḥadīth*.

Based on these findings, this article offers a preliminary model for integrating Arabic computational linguistics into the study and teaching of concise ḥadīth terminology *matns*. Future research may develop a specialized digital lexicon of ḥadīth terminology, compare computationally identified priorities with actual commentarial practice, and test the model on other concise *matns* such as *Nukhbat al-Fikar*, *al-Bayqūniyyah*, and *Ikhtiṣār ʿUlūm al-Ḥadīth*. In this way, computational tools can support the discovery, organization, and teaching of technical Islamic texts while preserving the authority of the classical scholarly tradition

Acknowledgements

The author declares no conflict of interest. No external funding was received for this research.

Bibliography

Abdelali, Ahmed, Kareem Darwish, Nadir Durrani, and Hamdy Mubarak. "Farasa: A Fast and Furious Segmenter for Arabic." In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, 11–16. San Diego: Association for Computational Linguistics, 2016. https://doi.org/10.18653/v1/N16-3003.

al-ʿAsqalānī, Aḥmad ibn ʿAlī Ibn Ḥajar. *Nuzhat al-Naẓar fī Tawḍīḥ Nukhbat al-Fikar fī Muṣṭalaḥ Ahl al-Athar*. Edited by Nūr al-Dīn ʿItr. 3rd ed. Damascus: Maṭbaʿat al-Ṣabāḥ, 1421/2000.

Azmi, Aqil M., Abdulaziz O. Al-Qabbany, and Amir Hussain. "Computational and Natural Language Processing Based Studies of Hadith Literature: A Survey." *Artificial Intelligence Review* 52, no. 2 (2019): 1369–1414. https://doi.org/10.1007/s10462-019-09692-w.

al-Baghdādī, Abū Bakr Aḥmad ibn ʿAlī al-Khaṭīb. *al-Jāmiʿ li-Akhlāq al-Rāwī wa-Ādāb al-Sāmiʿ*. Edited by Maḥmūd al-Ṭaḥḥān. Riyadh: Maktabat al-Maʿārif, 1403/1983.

Bashir, Muhammad Huzaifa, Aqil Azmi, Haq Nawaz, Wajdi Zaghouani, Mona Diab, Ala Al-Fuqaha, and Junaid Qadir. "Arabic Natural Language Processing for Qurʿanic Research: A Systematic Review." *Artificial Intelligence Review* 56 (2023): 6801–54. https://doi.org/10.1007/s10462-022-10313-2.

Bounhas, Ibrahim. "On the Usage of a Classical Arabic Corpus as a Language Resource: Related Research and Key Challenges." *ACM Transactions on Asian and Low-Resource Language Information Processing* 18, no. 3 (2019): 1–45. https://doi.org/10.1145/3277591.

Atsar: Journal of Hadith and Contemporary Studies
E-ISSN: 3164-1830

Darwish, Kareem, Nizar Habash, Mourad Abbas, Hend Al-Khalifa, Huseein T. Al-Natsheh, Samhaa R. El-Beltagy, Houda Bouamor, Karim Bouzoubaa, Violetta Cavalli-Sforza, Wassim El-Hajj, Mustafa Jarrar, and Hamdy Mubarak. "A Panoramic Survey of Natural Language Processing in the Arab World." *Communications of the ACM* 64, no. 4 (2021): 72–81. https://doi.org/10.1145/3447735.

al-Dhahabī, Shams al-Dīn Muḥammad ibn Aḥmad. *al-Mawqiẓah fī ʿIlm Muṣṭalaḥ al-Ḥadīth*. Edited by ʿAbd al-Fattāḥ Abū Ghuddah. 2nd ed. Aleppo: Maktab al-Maṭbūʿāt al-Islāmiyyah, 1412.

Habash, Nizar, Owen Rambow, and Ryan Roth. "MADA+TOKAN: A Toolkit for Arabic Tokenization, Diacritization, Morphological Disambiguation, POS Tagging, Stemming and Lemmatization." In *Proceedings of the 2nd International Conference on Arabic Language Resources and Tools*, 102–9. Cairo: The MEDAR Consortium, 2009.

Obeid, Ossama, Nasser Zalmout, Salam Khalifa, Dima Taji, Mai Oudah, Bashar Alhafni, Go Inoue, Fadhl Eryani, Alexander Erdmann, and Nizar Habash. "CAMeL Tools: An Open Source Python Toolkit for Arabic Natural Language Processing." In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 7022–32. Marseille: European Language Resources Association, 2020.

Pasha, Abdelrahim, Mohamed Al-Badrashiny, Mona Diab, Ahmed El Kholy, Ramy Eskander, Nizar Habash, Manoj Pooleery, Owen Rambow, and Ryan Roth. "MADAMIRA: A Fast, Comprehensive Tool for Morphological Analysis and Disambiguation of Arabic." In *Proceedings of the Ninth International Conference on Language Resources and Evaluation*, 1094–1101. Reykjavik: European Language Resources Association, 2014.

Siddiqi, Muhammad Zubayr. *Hadith Literature: Its Origin, Development and Special Features*. Edited by Abdal Hakim Murad. 2nd rev. ed. Cambridge: Islamic Texts Society, 1993.

al-Suyūṭī, Jalāl al-Dīn ʿAbd al-Raḥmān ibn Abī Bakr. *Tadrīb al-Rāwī fī Sharḥ Taqrīb al-Nawāwī*. Edited by Abū Qutaybah Naẓar Muḥammad al-Fāryābī. 2 vols. Riyadh: Dār Ṭaybah, n.d.

Copyright © 2026 by the authors. This publication is subject to the terms and conditions of the Creative Commons Attribution ShareAlike (CC BY SA) license (https://creativecommons.org/licenses/by-sa/3.0/).