

**CONSTRUCTING SCIENTIFIC CLAIMS IN ONLINE HERBAL DISCOURSE:
A COMPUTATIONAL LINGUISTIC ANALYSIS OF *KASKUS* THREADS****Devi Melisa Saragi¹, Prihantoro², Mifthahulkhaerah Anwar³**Universitas Kristen Indonesia¹, Universitas Diponegoro², Universitas Negeri Jakarta³Email: devi.saragi@uki.ac.id¹, prihantoro@lecturer.undip.ac.id²,miftahulkhairah@unj.ac.id³**ABSTRACT**

This study examines how claims are constructed and how scientific tone is linguistically realized in online herbal discourse. Its novelty lies in integrating computational text analysis with discourse analysis to evaluate the scientific quality of online herbal health claims in the Indonesian context. Using a quantitative design, data were collected from fifteen herbal-treatment discussion threads on the Kaskus forum. After Python-based preprocessing, including tokenization, data cleaning, and text normalization, the final corpus comprised 572 files containing 77 claim sentences. Claim construction was analyzed through assertiveness markers and claim types based on Biber's (2006) stance framework. Scientific tone was measured as a continuous numerical score using linguistic features adapted from McNamara et al. (2010). Toulmin's argumentation model, as elaborated by Hample (1977), was applied to interpret the relationship between linguistic form and argumentative justification. The findings show that rhetorical strength does not necessarily reflect scientific tone. Linguistically assertive claims often exhibit low scientific tone because they rely on evaluative and popular language, whereas less assertive claims may achieve a moderate scientific tone through evidential markers and more formal linguistic features. These findings contribute to digital discourse and health communication research by distinguishing claim strength from scientific tone and provide practical insights for improving health literacy and critical evaluation of online health claims.

Keywords: *herbal, kaskus, assertive, scientific tone, claim, computational***INTRODUCTION**

Kaskus has long been recognized as one of Indonesia's most influential online forums, shaping patterns of digital participation and everyday information exchange among its users. Since its emergence in the early 2000s, the platform has functioned as a decentralized public sphere where discussions on diverse topics, ranging from hobbies and technology to health and lifestyle, develop organically through user-generated threads (Lim, 2013; Nugroho, 2020). Among the most active domains within the forum are herbal-related threads, where users commonly exchange experiences, advice, and informal recommendations concerning traditional remedies. These discussions often blend personal anecdotes with fragments of popular health beliefs, forming a dynamic but largely unregulated ecosystem of information circulation. As Indonesia maintains a strong cultural attachment to herbal medicine and traditional healing practices (Sujarwo et al., 2021), the discourses emerging within these online spaces bear significant influence on how ordinary users conceptualize health, risk, and therapeutic efficacy.

Within these herbal-related discussions, health information is rarely presented as systematically verified medical knowledge. Instead, users tend to formulate claims that vary in their degree of assertiveness, evidential grounding, and linguistic alignment with scientific discourse. Such claims frequently combine confident evaluative language with limited empirical support, or alternatively adopt surface features associated with scientific writing, such as references to research, expert authority, or technical terminology, without clearly articulating methodological justification. As a result, the scientific quality of claims becomes ambiguous, not because their factual accuracy can be readily determined, but because the linguistic construction of the claims blurs the boundary between scientific and popular discourse. This ambiguity highlights the need to examine how herbal treatment claims are framed, how strongly they are asserted, and how scientific tone is deployed to position information as credible within online forums such as Kaskus.

Although substantial scholarly interest has been directed toward digital health communication, research specifically addressing user-generated herbal discourse in Indonesian online communities remains limited. Existing studies tend to focus on broader issues such as the spread of health misinformation on major social media platforms, including Facebook, Twitter, and WhatsApp (Maharani and Fuady, 2022; Wirawan and Simarmata, 2023). A smaller number of studies examine online discussions of traditional medicine. However, these analyses are largely qualitative in nature and do not incorporate computational approaches capable of capturing linguistic patterns at scale (Kristiyanti, 2021; Utami, 2022). Moreover, while global scholarship has increasingly applied computational linguistics and natural language processing to assess the credibility, structure, and framing of health-related digital content (Zhang et al., 2020; Miner et al., 2022), such methodological integrations remain rare in the Indonesian context. This gap points to the absence of systematic and data-driven investigations into how herbal knowledge is linguistically constructed and circulated within local digital forums.

The urgency of this research arises from the expanding role of online platforms in shaping public health perceptions. As digital communities increasingly influence individual decision-making related to diet, treatment, and self-care, unverified or weakly supported claims embedded in herbal discussions risk normalizing speculative or inaccurate understandings of health. This concern is particularly salient in Indonesia, where the consumption of herbal products remains widespread and often occurs without rigorous evidence-based evaluation (Windari et al., 2022). Examining the linguistic form of these claims is therefore essential for understanding how epistemic authority and credibility are constructed in participatory digital spaces.

The novelty of this study lies in its use of computational methods to analyze linguistic patterns in herbal discussions on an Indonesian online forum, an approach that remains underutilized in local research. Through Python-based text processing techniques, including tokenization, frequency extraction, and computational scoring of linguistic features, this study provides empirical insights into the discursive structure of herbal threads at a scale unattainable through traditional qualitative analysis. By integrating computational linguistics with discourse-analytic perspectives, the study advances methodological innovation in the analysis of Indonesian digital health communication. The findings not only reveal recurrent themes and lexical tendencies within herbal discussions but also offer a data-informed understanding of how claims,



assumptions, and beliefs are linguistically constructed and positioned as credible knowledge.

The significance of this study extends both academically and socially. Academically, it contributes to research on digital discourse, health communication, and computational humanities by offering a localized case study that enriches global discussions on digital health cultures. Socially, the findings may inform educators, policymakers, and health communicators seeking to design interventions or literacy programs that address public engagement with unverified health claims. By mapping the linguistic landscape of herbal threads on Kaskus, this study provides empirical grounding for understanding the discursive mechanisms through which online claims may influence health perceptions and behaviors. Based on this background, the study is guided by the central research question: *How are claims constructed and evaluated for scientific quality in online herbal treatment discussions found in Kaskus threads?* To address this question, the study examines patterns of linguistic assertiveness, the use of verbal claim markers, and variations in scientific tone as indicators of the scientific quality of claims in user-generated herbal discourse.

This study draws on three complementary theoretical frameworks to analyze the construction and evaluation of claims in discourse: (1) stance as a linguistic resource for expressing epistemic commitment (Biber, 2006), (2) linguistic features of writing quality as continuous numerical measures (McNamara et al., 2010), and (3) argument strength as a function of justification and support in discourse (Toulmin, via Hample, 1977). Together, these frameworks provide a comprehensive basis for examining how claims are linguistically constructed, quantitatively represented, and argumentatively evaluated.

The primary linguistic foundation of this study is Biber's (2006) framework of stance, which conceptualizes stance as the grammatical expression of a speaker's or writer's personal attitudes, evaluations, and assessments toward the information being communicated. In this framework, stance is treated as a core dimension of discourse through which language users position themselves epistemically and evaluatively with respect to propositions, rather than as an abstract or purely pragmatic phenomenon.

A central contribution of Biber's approach lies in its emphasis on explicit, grammatically encoded stance markers. Rather than inferring stance from general lexical choices or contextual interpretation, Biber restricts stance analysis to overt grammatical structures that can be systematically identified and empirically analyzed. This focus allows stance to be operationalized for large-scale corpus-based research while maintaining theoretical clarity and analytical consistency.

Biber (2006) identifies three major grammatical categories through which stance is realized namely *modal verbs*, *stance adverbs*, and *stance complement clauses*. Modal verbs express assessments of *possibility*, *necessity*, *prediction*, or *obligation*, thereby signaling judgments about likelihood or necessity. Stance adverbs provide explicit commentary on propositions and are classified into *epistemic*, *attitude*, and *style adverbs*, with *epistemic adverbs* playing a particularly important role in signaling commitment to *truth*. Complement clause constructions realized through verbs (e.g., *believe that*), adjectives (e.g., *it is likely that*), and nouns (e.g., *the fact that*) allow propositions to be framed as *certain*, *probable*, *assumed*, or *contested*.

Two theoretical principles are central to Biber's framework. First, stance is gradient rather than binary where linguistic expressions encode varying degrees of epistemic commitment, enabling distinctions between strong, moderate, and weak positioning. Second, linguistic stance is analytically separated from epistemic truth. Strong stance markers indicate how confidently a proposition is presented, not whether it is factually correct. This distinction makes the framework particularly suitable for analyzing persuasive, authoritative, or evaluative discourse, including health-related and informational texts.

While Biber (2006) explains how epistemic commitment is expressed linguistically, McNamara et al. (2010) provide a complementary computational framework for representing linguistic features as continuous numerical indices. Using automated discourse analysis tools such as Coh-Metrix, this framework operationalizes writing quality as a gradient construct derived from measurable textual properties, rather than as categorical judgments. In this approach, linguistic features are quantified across multiple dimensions, including cohesion, syntactic complexity, lexical diversity, and word characteristics. Each feature is calculated as a continuous score, allowing subtle variations in textual properties to be statistically modeled and compared across texts. Writing quality is thus conceptualized as an emergent property of interacting linguistic distributions, rather than a holistic or impressionistic assessment.

A key implication of this framework is that linguistic quality is scalar and probabilistic, not absolute. Higher-rated texts tend to exhibit greater syntactic complexity, higher lexical diversity, and lower word frequency, reflecting increased cognitive demand and linguistic sophistication. Importantly, these characteristics are captured through numerical indices rather than subjective interpretation. At the same time, McNamara et al. demonstrate that not all linguistic features contribute equally to quality, highlighting the importance of pattern-based interpretation rather than reliance on single markers. By representing linguistic features as continuous scores, this framework supports flexible analytical practices. Numerical values can be analyzed statistically, correlated with external criteria, and operationalized into interpretive ranges when necessary, without imposing universal or normative thresholds. This makes the framework particularly valuable for studies that seek to integrate quantitative text analysis with discourse-level interpretation.

To connect linguistic expression and numerical representation with epistemic evaluation, this study draws on Toulmin's model of argumentation as explicated by Hample (1977). Toulmin's framework conceptualizes argument strength not in terms of formal logical validity, but as a function of how claims are justified, qualified, and supported in real discourse. At the core of Toulmin's model are six interrelated components namely claim, data, warrant, backing, qualifier, and rebuttal. The claim represents the proposition advanced for acceptance; data provide the evidence supporting the claim; warrants explain the inferential link between data and claim; backing reinforces the warrant; qualifiers indicate degrees of certainty; and rebuttals specify conditions under which the claim may not hold. Together, these components offer a functional structure for analyzing not only what is asserted, but how legitimately it is asserted.

A key principle emphasized by Hample (1977) is that argument strength is relational and proportional. Claims expressed with high certainty require correspondingly strong data and warrants. When strong claims lack evidential support or backing, the

argument becomes epistemically unbalanced. Importantly, Toulmin's model separates the rhetorical form of a claim from its justificatory structure, allowing analysts to identify arguments that appear persuasive but are weak in evidential terms. This makes the framework especially relevant for analyzing persuasive or pseudo-scientific discourse.

Taken together, these three frameworks form a coherent analytical model. Biber (2006) explains how epistemic commitment is linguistically constructed through stance. McNamara et al. (2010) provide tools for representing linguistic features as continuous, quantifiable measures. Toulmin's model supplies the argumentative logic that explains how claims should be supported and qualified to achieve epistemic legitimacy. By integrating these perspectives, the present study situates claim analysis at the intersection of linguistic form, quantitative representation, and argumentative justification.

METHOD

This study employed a quantitative research design supported by a computational text-processing approach to analyze the epistemic reliability of herbal-treatment claims in online discourse. The dataset consisted of discussion threads related to herbal treatments collected from Kaskus, one of Indonesia's most influential online forums. A total of 15 threads published in 2021 were retrieved in HTML format and compiled as the primary corpus.

The initial stage of the study focused on text preprocessing to ensure the consistency and relevance of the linguistic data under analysis. All HTML files were parsed and segmented using the *BeautifulSoup4* library in combination with the Natural Language Toolkit (NLTK) within a Python environment. Tokenization was conducted based on a predefined list of herbal-related lexical items identified in previous studies on traditional medicine discourse. Using the script *batch_tokenizer.py*, the corpus was processed into 572 text files, yielding a total of 15,486 raw tokens. The workflow of this stage is illustrated below this image.

Image 2.a. workflow of script *batch_tokenizer.py*

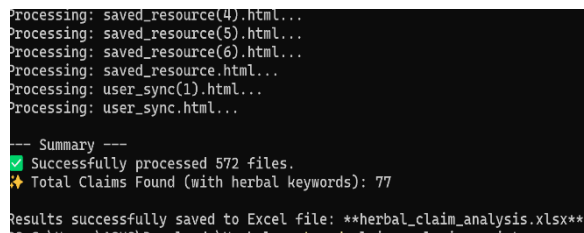
```
Processing: reach_worklet(3).html...
Processing: reach_worklet(4).html...
Processing: reach_worklet.html...
Processing: saved_resource(1).html...
Processing: saved_resource(2).html...
Processing: saved_resource(3).html...
Processing: saved_resource(4).html...
Processing: saved_resource(5).html...
Processing: saved_resource(6).html...
Processing: saved_resource.html...
Processing: user_sync(1).html...
Processing: user_sync.html...

--- Summary ---
✔ Successfully processed 572 files.
🌟 Total combined tokens: 15486
```

Subsequently, a data-cleaning procedure was applied through the script *tokenize_herbal_data.py* to remove noise, eliminate irrelevant strings, and standardize token formats. This process resulted in 572 text files containing 77 validated herbal-

related tokens, which were then organized and further analyzed using the pandas library. A visualization of this process is shown in below.

2.b image of the script *tokenize_herbal_data.py*



```
Processing: saved_resource(4).html...
Processing: saved_resource(5).html...
Processing: saved_resource(6).html...
Processing: saved_resource.html...
Processing: user_sync(1).html...
Processing: user_sync.html...

--- Summary ---
✔ Successfully processed 572 files.
✦ Total Claims Found (with herbal keywords): 77

Results successfully saved to Excel file: **herbal_claim_analysis.xlsx**
```

Following preprocessing, the analysis proceeded to a computational assessment of scientific tone and claim characteristics. In line with established practices in automated discourse analysis (McNamara et al., 2014), linguistic features were represented as continuous numerical scores. These scores reflect the formal and distributional properties of claims, including the presence of scientific markers, assertive claim constructions, and evidential indicators, thereby enabling systematic comparison across texts within the corpus.

For analytical clarity, the continuous numerical scores were operationalized into categorical ranges: scores between 0.0–0.49 were classified as low-quality claims, 0.5–0.79 as medium-quality claims, and 0.8–1.0 as high-quality claims. These ranges are not intended to serve as universal or normative standards of scientific validity. Rather, they function as operational heuristics commonly used in computational text analysis to facilitate the interpretation and comparison of continuous linguistic measures (McNamara et al., 2014). This categorization serves as an initial step in mapping variations in claim quality quantitatively before conducting discourse-level analysis.

Importantly, the numerical classification does not, by itself, determine epistemic reliability. Instead, numerical scores are treated as empirical inputs that require interpretive analysis. Accordingly, the categorized scores were interpreted through Biber's (2006) stance framework, which conceptualizes claim strength in terms of epistemic stance, specifically the degree of certainty, commitment, and authority a writer expresses toward a proposition. Claims associated with higher numerical scores tend to exhibit stronger epistemic commitment, manifested through the use of assertive claim verbs, absolute result predicates, and the absence of hedging devices, and are therefore interpreted as strong claims. Conversely, lower-scoring claims typically display reduced epistemic commitment through hedging, contextual limitation, or the absence of explicit evidential grounding, and are interpreted as weak claims.

By integrating computational scoring with stance-based discourse analysis, this study conceptualizes claim strength as the outcome of an interaction between quantifiable linguistic features and epistemic positioning in discourse. The two approaches serve distinct yet complementary functions: the computational approach systematically captures formal linguistic patterns across the corpus, while stance analysis provides an interpretive account of how these patterns construct epistemic meaning in herbal-treatment discourse. The correlation between numerical claim scores and stance-based interpretations is theoretically grounded in Toulmin's model of argumentation. According to Toulmin (2023), the strength of a claim depends not only on its propositional content but also on how it is warranted, qualified, and supported as credible knowledge. Within

this framework, numerical scores represent the formal realization of claims, whereas stance analysis accounts for their argumentative legitimacy and epistemic force. Consequently, the integration of computational scoring and discourse analysis provides a coherent and theoretically grounded basis for distinguishing between strong and weak claims in online herbal-treatment discourse.

RESULTS AND DISCUSSION

3.a. Table herbal_claim_analysis

Row Labels	Sum of Scientific_Tone_Score
[Diskusi] Ide Produk UMKM_ Jamu Herbal buat Vertigo Ringan, Menarik Juga Nih Gan! _ KASKU	0,35
Kuat (dengan verba klaim)	0,15
Lemah (tanpa verba klaim)	0,2
[HELP] Pengobatan Herbal _ Alternatif tuk penyakit TBC wilayah Jabodetabek _ KASKUS.html	0,7
Kuat (dengan verba klaim)	0,7
6 Obat & Pengobatan Herbal Khas Korea Selatan Yang Ada Sejak Zaman Dahulu _ KASKUS.html	2,4
Kuat (dengan verba klaim)	1,15
Lemah (tanpa verba klaim)	1,25
Honeysuckle_ Bunga Herbal Ajaib untuk Kulit Cerah, Sehat, dan Bebas Masalah _ KASKUS.html	2,6
Kuat (dengan verba klaim)	0,4
Lemah (tanpa verba klaim)	2,2
Ide Produk UMKM_ Jamu Herbal Alami untuk Penyakit Liver _ KASKUS.html	0,8
Kuat (dengan verba klaim)	0,4
Lemah (tanpa verba klaim)	0,4
Ide UMKM 2025_ Jamu Herbal buat Batu Empedu Dini _ KASKUS.html	0,55
Kuat (dengan verba klaim)	0,55
Jamu Herbal Untuk Cadangan_ Tradisi Lama yang Tetap Eksis, Gan _ KASKUS.html	1,15
Kuat (dengan verba klaim)	0,95
Lemah (tanpa verba klaim)	0,2
Pengobatan Herbal Alternatif _ KASKUS.html	1,4
Kuat (dengan verba klaim)	0,7
Lemah (tanpa verba klaim)	0,7
Pengobatan Herbal Efektif Untuk Penyakit Gangguan Pemafasan _ KASKUS.html	1,3
Kuat (dengan verba klaim)	1,3
Pengobatan herbal thyroid (Gondok) _ KASKUS.html	2,6
Kuat (dengan verba klaim)	0,4
Lemah (tanpa verba klaim)	2,2
PENGobatan PENYAKIT TIPES DENGAN HERBAL _ KASKUS.html	2,2
Kuat (dengan verba klaim)	0,6
Lemah (tanpa verba klaim)	1,6
Pengobatan sakit Gigi dengan Herbal _ KASKUS.html	1,1
Kuat (dengan verba klaim)	1,1
Pengobatan Trauma Secara Herbal _ KASKUS.html	0,35
Kuat (dengan verba klaim)	0,35
Pengobatan Tumor Mediastinum Dengan Herbal _ KASKUS.html	0,75
Kuat (dengan verba klaim)	0,75
Resep Jamu Herbal Alami untuk Darah Tinggi, Peluang Emas UMKM 2025 _ KASKUS.html	2,05
Kuat (dengan verba klaim)	0,95
Lemah (tanpa verba klaim)	1,1
Grand Total	20,3

Based on the integrated analytical perspective of epistemic stance (Biber, 2006), continuous numerical representation of linguistic features (McNamara et al., 2010), and argumentative claim structure (Toulmin, via Hample, 1977), the data presented in table 4.a. indicate that claim sentences and scientific tone do not demonstrate a proportional or linear relationship within the analyzed herbal-treatment discourse. In general, the results suggest that the linguistic strength of a claim does not necessarily correspond to a higher level of scientific tone.

The table shows that claim sentences marked as assertive (true) are consistently categorized as strong claims, as they are realized through explicit claim verbs. However, several of these strong claims are associated with relatively low scientific tone scores, indicating that high epistemic commitment does not automatically entail the use of scientific or academically grounded language. These claims tend to be expressed through

persuasive or narrative formulations that lack explicit evidential or empirical markers, resulting in lower scientific tone values. In contrast, the results also reveal that claims categorized as weak (without claim verbs) and marked as non-assertive (false) may exhibit moderate scientific tone scores. This pattern suggests that claims articulated in a descriptive, cautious, and explanatory manner, often resembling academic discourse, can register higher scientific tone values despite the absence of explicit assertive constructions.

Overall, the results demonstrate that scientific tone is more closely related to linguistic presentation than to the assertiveness or categorical strength of claims. This pattern highlights a divergence between epistemic positioning and scientific discourse markers, which provides an empirical basis for the discussion section that follows, particularly with regard to the epistemic reliability of herbal-treatment claims in online forums.

Building on the patterns observed in table 4.a, the discussion now turns to a closer examination of individual claim sentences in order to demonstrate how linguistic assertiveness and scientific tone interact within specific textual realizations.

Thread 3: 6 Obat & Pengobatan Herbal Khas Korea Selatan Yang Ada Sejak Zaman Dahulu _ KASKUS.html

Thread three consists of ten claim sentences, five of which employ explicit assertive constructions and claim verbs, while the remaining five do not exhibit assertive forms or claim verbs. Despite this variation in assertiveness, all ten claim sentences register low scientific tone scores, ranging from 0.20 to 0.45. The explanation is as follow:

The claim sentence (1) “*Gingseng ini dipercaya mampu menghilangkan stres, meningkatkan konsentrasi, menambah stamina, dan mengobati tekanan darah*” is linguistically realized as a strong claim, as it employs explicit assertive constructions and claim verbs that encode a high degree of epistemic commitment. From the perspective of stance in discourse (Biber, 2006), the use of expressions such as *dipercaya mampu* and *mengobati* positions the proposition as credible and authoritative, presenting the therapeutic effects of ginseng as established or generally accepted knowledge rather than as tentative or exploratory information. This linguistic configuration signals strong authorial commitment to the truth of the claim. However, despite its strong assertive stance, the claim registers a low scientific tone score (0.2). In terms of continuous numerical representation of linguistic features (McNamara et al., 2010), this low score indicates a limited presence of linguistic markers typically associated with academic or scientific discourse, such as references to empirical evidence, technical explanation, methodological framing, or cautious qualification. The claim is formulated through a list of broad therapeutic effects without accompanying evidential grounding or explanatory detail, which reduces its overall scientific tone despite its confident presentation.

When examined through the lens of Toulmin’s model of argumentation (via Hamble, 1977), the claim demonstrates an imbalance between epistemic commitment and argumentative support. While the claim exhibits a strong qualifier through assertive language, it lacks explicit data, warrants, and backing that would justify the asserted causal relationship between *ginseng* consumption and multiple health outcomes. As a result, the argument relies primarily on rhetorical assertion rather than on evidential



justification. According to Toulmin's framework, such a configuration weakens the epistemic reliability of the claim, as strong claims require proportional support to be considered argumentatively sound.

Taken together, this analysis illustrates that strong assertiveness does not necessarily correspond to high scientific tone or epistemic reliability. Although the claim presents itself as truthful and authoritative, its low scientific tone score reveals that the claim is constructed through persuasive or narrative language rather than through empirically grounded or academically framed discourse. This finding reinforces the broader pattern observed in the corpus, in which linguistic confidence may function independently of scientific legitimacy in online herbal-treatment claims.

Meanwhile, within the same thread, another claim sentence (3), "*Cara pengobatannya kulit para pasien akan ditusuk dengan jarum di bagian-bagian tertentu,*" is identified for further analysis. This claim only obtains a scientific tone score of 0.2. This scale indicates a low density of linguistic markers associated with scientific discourse. Referring to the numerical measurement approach of linguistic features (McNamara et al., 2010), this low score suggests that the sentence does not employ linguistic elements commonly found in academic or scientific discourse, such as the use of specific technical terminology, explanations of action mechanisms, or evidence-based methodological framing. From a linguistic-structural perspective, this claim is categorized as a weak claim, as it does not employ assertive claim verbs and is characterized by a non-assertive stance. This means that the claim does not directly state the truth, effectiveness, or success of the treatment, since it avoids reinforcing or evaluative verbs such as *treat*, *cure*, *prove*, or *be effective*. As a result, the claim does not demonstrate a strong linguistic commitment to the truth of its propositional content, but instead conveys information in a neutral and descriptive manner.

Within the framework of stance in discourse (Biber, 2006), the sentence functions as a descriptive statement that presents procedural information without explicitly asserting therapeutic effectiveness or epistemic certainty. Consequently, the level of epistemic commitment expressed in the claim remains relatively low. When examined through Toulmin's model of argumentation (via Hample, 1977), the claim exhibits a minimal argumentative structure. The statement does not establish a clear relationship between the proposition and its justificatory basis, as it is not supported by data, warrants, or backing that would explain the relevance of the described procedure to the intended therapeutic outcome. Therefore, the claim functions more as background information than as an argument intended to construct knowledge or persuade the reader. Ultimately, this configuration demonstrates that non-assertive and descriptive claims may still register a low scientific tone when they are not framed using technical, evidence-based language. This finding reinforces the broader pattern observed in the data, indicating that low scientific tone is not solely attributable to cautious claim formulation, but rather to limitations in the linguistic and argumentative strategies used to present information.

Claim Sentence (8), "*Berdasarkan hasil penelitian, diketahui bahwa tanduk rusa mampu meningkatkan sel darah merah dan hemoglobin,*" is still categorized as a weak claim because it does not employ claim verbs that directly assert truth or effectiveness in an absolute manner. Instead, the deer antler claim merely reports findings

without explicitly affirming certainty, and is therefore classified as non-assertive. Nevertheless, this claim obtains a higher scientific tone score (0.45) compared to the two preceding claims (claim sentence 1 and 3), although it remains within the low category. This difference is primarily attributable to the way the claim is linguistically framed rather than to its level of assertiveness.

From the perspective of numerical measurement of linguistic features (McNamara et al., 2010), the increase in scientific tone score is triggered by the presence of scientific discourse markers that were absent from the earlier claims. The phrase “*berdasarkan hasil penelitian*” functions as an evidential signal that implicitly refers to a scientific source of knowledge, while the mention of specific biological variables such as *red blood cells* and *hemoglobin* increases the level of technical terminology in the sentence. These elements raise the density of linguistic features associated with a scientific register, thereby elevating the scientific tone score even though the claim itself is not expressed assertively. From the standpoint of stance in discourse (Biber, 2006), the claim continues to exhibit moderate to low epistemic commitment. Expressions such as “*diketahui bahwa*” and “*mampu meningkatkan*” do not function as claim verbs that directly assert truth; rather, they present information within a reporting framework. As a result, the claim does not position the writer as an explicit guarantor of truth, but instead as a conduit for information derived from particular findings.

When analyzed through Toulmin’s model of argumentation (via Hample, 1977), the claim begins to show an embryonic argumentative structure, particularly in terms of the data component, which is implicitly represented by the reference to research findings. However, the claim still lacks a warrant that explains the causal relationship between deer antler consumption and the increase in biological parameters, as well as backing in the form of methodological references or clearly identified scientific authority. Consequently, although the claim appears more scientific linguistically than the preceding claims, it remains argumentatively weak. In sum, these findings demonstrate that scientific tone may increase without a corresponding increase in assertiveness or claim strength, as long as the claim is framed using language that reflects features of academic discourse, such as references to research and technical terminology. In other words, the variation in scientific tone scores observed in this claim reflects differences in linguistic strategy, rather than differences in the truth value or epistemic reliability of the claim itself.

Thread 4: Honeysuckle_ Bunga Herbal Ajaib untuk Kulit Cerah, Sehat, dan Bebas Masalah _ KASKUS.html

The second thread also contains eleven claim sentences. Of these, nine claims are classified as assertive false, while the remaining claims are assertive true. However, all claim sentences fall within the low scientific tone score range. Claim sentence (2), “*Tanaman ini dikenal dengan bunga kecil berwarna putih kekuningan yang mengeluarkan aroma manis, mirip madu, terutama saat mekar di musim panas,*” functions primarily as a general description of the plant’s morphological characteristics, rather than as a claim that advances specific scientific knowledge or propositions. The descriptive nature of this claim is evident in its focus on observable morphological and phenological attributes, including the small size of the flowers, their yellowish-white



color, a sweet honey-like aroma, and their blooming period during the summer season. These features constitute physical and seasonal characteristics of the plant and are not presented to assert truth, benefit, or any particular implication. Consequently, from a linguistic perspective, the sentence is classified as a weak claim characterized by a non-assertive stance.

The low scientific tone score (0.2) further reflects a mode of expression that more closely resembles everyday observational language than an academic register. Referring to the approach of numerical measurement of linguistic features (McNamara et al., 2010), the description of color, aroma, and flowering season is delivered without the support of specific botanical terminology or an explicit scientific explanatory framework. As a result, the density of linguistic markers associated with scientific discourse in this sentence remains relatively low. From the perspective of stance in discourse (Biber, 2006), the expression “*dikenal dengan*” signals an epistemic distance between the writer and the proposition being conveyed. Rather than positioning the writer as an authority or guarantor of knowledge, the sentence merely reproduces general descriptive knowledge, without explicitly indicating levels of certainty or scientific relevance. This is evident in the absence of epistemic markers such as *consistently*, *based on scientific observation*, *proven*, or *significantly*.

Being examined through Toulmin’s model of argumentation (via Hample, 1977), the statement does not develop an evaluable argumentative structure. The sentence does not articulate a claim that demands justification, nor is it accompanied by data, warrant, or backing that would link the plant description to any biological, medical, or scientific function. Accordingly, the claim serves as contextual background information that enriches the narrative but does not contribute substantially to the construction of scientific knowledge. Thus, this claim illustrates that informative yet non-assertive descriptions tend to yield low scientific tone scores when they are not framed through technical and argumentative linguistic strategies. This variation reinforces the broader pattern observed in the data, namely that the scientific quality of discourse is not determined solely by informational content, but by how that information is positioned and linguistically constructed within the text.

Still within the same discussion, another claim sentence (5) “*Menurut data dari beberapa penelitian herbal modern, honeysuckle mengandung senyawa aktif yang bersifat antibakteri dan antiinflamasi, menjadikannya bahan alami ideal untuk perawatan kulit sensitif atau rentan berjerawat*” . In the herbal claim analysis table, this claim obtains a scientific tone score of 0.45, which remains within the low category, yet is higher than that of the preceding claims within the same thread. This difference indicates an increase in the density of linguistic markers associated with scientific discourse, although the increase is not sufficient to elevate the claim to a medium or high scientific tone category. Referring to the approach of numerical measurement of linguistic features (McNamara et al., 2010), the rise in scientific tone score is primarily influenced by the presence of explicit scientific discourse markers that were absent in earlier claims. The phrase “*menurut data dari beberapa penelitian herbal modern*” functions as an evidential signal, linguistically associating the claim with research practices and scientific sources of knowledge. In addition, the use of technical terms such as “*senyawa aktif*” (active compounds), “*antibakteri*” (antibacterial), and “*antiinflamasi*” (anti-

inflammatory) increase lexical complexity and reinforces the impression of a scientific register within the sentence.

Nevertheless, the scientific tone of this claim remains low because its linguistic framing does not fully conform to the conventions of academic discourse. References to *several studies* are presented in a general manner, without methodological specification, quantitative data, or a clearly defined experimental context. As a result, despite the presence of scientific terminology, the sentence more closely resembles a popular-scientific style than a rigorous scientific presentation. Moreover, the claim is not constructed through analytical or explanatory sentence structures, but rather through an evaluative generalization, as indicated by the phrase “*making it an ideal natural ingredient.*” This evaluation is not accompanied by an explanation of biological mechanisms, such as how the active compounds operate at the cellular level, nor by empirical justification in the form of verifiable quantitative data or experimental context. Consequently, the contribution of this claim to scientific tone remains limited.

These findings indicate that a scientific tone score of 0.45 reflects a position between popular descriptive language and formal scientific language. This claim demonstrates that scientific tone can increase through the inclusion of research references and technical terminology; however, without more rigorous methodological and argumentative framing, such improvement remains partial. These findings reaffirm that scientific tone is determined not merely by the presence of scientific terms, but by how those terms are integrated into the linguistic structure and argumentative construction of the claim.

In contrast, the next statement (10), “*Berdasarkan temuan dari riset bahan alami, honeysuckle memiliki efek anti peradangan kuat yang dapat mengurangi bintik kecil di wajah dan menenangkan kulit yang iritasi,*” is categorized as assertive true and classified as a strong claim with claim verbs. However, this claim obtains a scientific tone score of only 0.2, indicating a low density of linguistic markers associated with scientific discourse. This discrepancy highlights that the level of assertiveness of a claim does not necessarily correspond to its level of scientific tone.

From the perspective of numerical measurement of linguistic features (McNamara et al., 2010), the low scientific tone of this claim is primarily attributable to its evaluative and popular linguistic framing, despite being expressed with a high degree of certainty. Expressions such as “*strong anti-inflammatory effect,*” “*reducing small facial blemishes,*” and “*soothing irritated skin*” represent direct evaluations of outcomes; however, they are not accompanied by technical dermatological terminology, such as references to *pro-inflammatory cytokines*, *facial erythema*, or *epidermal inflammation* (medical dictionary, 2020), nor by quantitative indicators or methodological context that are typical of scientific discourse. As a result, the density of linguistic features associated with an academic register remains low. Within the framework of stance in discourse (Biber, 2006), this claim demonstrates a high level of epistemic commitment, marked by the use of claim verbs that directly assert effectiveness and benefit. The writer positions themselves as a guarantor of the propositional truth by explicitly stating therapeutic effects. However, this strong stance is not supported by adequate scientific framing, causing the high assertiveness to function more as persuasive rhetoric rather than as a scientific presentation.



When analyzed through Toulmin's model of argumentation (via Hamble, 1977), the claim presents a firm assertion but lacks a proportionate justificatory structure. The reference to *"temuan dari riset bahan alami"* (*findings from natural ingredient research*) functions as a general evidential signal, yet it does not develop into explicit data, a warrant explaining the causal relationship between honeysuckle and anti-inflammatory effects, or backing in the form of clear methodological references or scientific authority. Consequently, the claim becomes epistemically unbalanced, as the level of certainty expressed is not aligned with the strength of its argumentative support. Taken together, this configuration demonstrates that claims that are assertive true and linguistically strong may still exhibit a low scientific tone when they are conveyed through evaluative, popular language and lack explicit methodological and argumentative framing. These findings reaffirm that scientific tone is not determined solely by how forcefully a claim is expressed, but by how the claim is linguistically constructed and epistemically supported.

Thread 15: "Resep Jamu Herbal Alami untuk Darah Tinggi, Peluang Emas UMKM 2025 _ KASKUS.html"

This thread contains six claim sentences, of which three are classified as assertive true, while the remaining claims are assertive false. However, five of these claim sentences exhibit low scientific tone scores, ranging from 0.2 to 0.35. Only one claim sentence reaches a moderate scientific tone score (0.75). Based on the herbal claim analysis table, claim sentence (3), *"Pakar herbal juga bilang, bahan alami ini kaya antioksidan yang bisa bantu memperbaiki metabolisme tubuh dan mengatur tekanan darah secara menyeluruh,"* is categorized as assertive true and classified as a strong claim with claim verbs. Nevertheless, this claim obtains a scientific tone score of only 0.2, which indicates a low density of linguistic markers associated with scientific discourse. This finding further confirms that assertiveness and rhetorical strength do not automatically correspond to higher levels of scientific tone.

From the perspective of numerical measurement of linguistic features (McNamara et al., 2010), the low scientific tone of this claim is primarily attributable to its popular and personally authoritative linguistic framing rather than scientific framing. Expressions such as *"pakar herbal juga bilang"* rely on a form of personal authority that is not methodologically defined, while phrases such as *"kaya antioksidan"* (rich in antioxidants), *"bisa bantu memperbaiki metabolisme"* (*help improve metabolism*) and *"mengatur tekanan darah secara menyeluruh"* (regulate blood pressure comprehensively) convey generalized evaluative outcomes without the support of technical terminology, quantitative parameters, or explicit explanations of biological mechanisms. From the standpoint of stance in discourse (Biber, 2006), this claim demonstrates a high level of epistemic commitment. The writer explicitly asserts the benefits and effectiveness of the natural ingredient and positions the source of the claim through reference to the authority of a *herbal expert*. However, this form of assertiveness reflects a narrative authority-based stance rather than an evidence-based stance, rendering the strength of the claim persuasive rather than analytical.

When analyzed through Toulmin's model of argumentation (via Hamble, 1977), the claim presents a strong assertion but lacks an adequate justificatory structure. The

reference to “*herbal experts*” functions as a generalized appeal to authority, but does not develop into verifiable data, a warrant explaining the causal relationship between antioxidants and metabolic improvement or blood pressure regulation, nor backing in the form of clearly identified scientific research or authoritative sources. Consequently, the claim becomes epistemically unbalanced, as the level of certainty expressed is not aligned with the strength of its argumentative support. Thus, this configuration demonstrates that claims that are assertive true and linguistically strong may still exhibit low scientific tone when they are articulated through popular language, rely on personal authority, and lack explicit methodological and argumentative framing. These findings reaffirm that scientific tone is not determined solely by how forcefully a claim is expressed, but by how the claim is linguistically constructed and epistemically supported.

Within the same thread, another claim (7), “*Katanya sih bahan-bahannya alami banget, ada jahe, daun pegagan, sama madu,*” obtains a scientific tone score of 0.2, which is the same as the preceding claim. This similarity in score indicates that, from the perspective of numerical measurement of linguistic features (McNamara et al., 2010), both claims exhibit an equally low density of linguistic markers associated with scientific discourse, resulting in a mode of expression that is closer to a conversational or popular register than to academic discourse. However, despite sharing the same level of scientific tone, this claim differs in important ways with respect to assertiveness and claim strength. The claim is classified as assertive false and constitutes a weak claim without claim verbs, as it is conveyed through informal expressions such as “*katanya sih,*” which signal epistemic distance and uncertainty of source. Within the framework of stance (Biber, 2006), such a formulation reflects a non-assertive stance which means that the writer does not position themselves as a guarantor of the proposition’s truth, but merely reproduces information indirectly in the form of “*hearsay*”.

Moreover, the content of the claim is presented primarily as a list of ingredients (ginger, *Centella asiatica* leaves, and honey) without any scientific framing, such as the classification of active compounds, explanations of mechanisms of action, or the provision of evidential context linking these ingredients to specific therapeutic effects. Such scientific framing is essential for transforming descriptive information into claims with epistemic value. In the absence of this framing, the mention of ingredients functions merely as surface-level information that cannot be evaluated scientifically. In addition, the claim also fails to develop an evaluable argumentative structure (Toulmin via Hample, 1977), as it is not accompanied by data, warrant, or backing that would explain the relevance of these ingredients to therapeutic effectiveness. Accordingly, although the score remains equally low (0.2), this claim exhibits a different type of weakness compared to the preceding claim. It is not only low in scientific tone, but also low in epistemic commitment, due to its non-assertive formulation and reliance on rumor or informal narration. Taken together, these findings confirm that equally low scientific tone scores do not necessarily imply identical claim characteristics. In this case, the low scientific tone is reinforced by an informal and non-assertive mode of expression, causing the claim to function as popular information with minimal epistemic legitimacy.

The last claim discussed is claim (5) “*Rujukan lengkapnya: <https://www.portaljatim24.com/2025/0...nggi-2025.html>*”. This claim obtains a scientific tone score of 0.75 (moderate), despite being classified as assertive false and categorized as a weak claim without claim verbs. This configuration indicates that scientific tone is



not determined solely by the propositional content of a claim, but also by its discursive function and linguistic framing. Within the framework of numerical measurement of linguistic features (McNamara et al., 2010), the relatively high scientific tone score associated with this sentence is not generated by technical terminology or conceptual explanation, but rather by the presence of an explicit evidential marker, namely the provision of a complete reference link. In computational text analysis, the inclusion of references or external sources functions as a strong indicator of informative and quasi-academic discourse practices, as source citation is a conventional feature of scientific and scholarly writing. Consequently, even in the absence of a scientific proposition, the system attributes a higher scientific tone score due to the signal of formal evidentiality. However, from the perspective of stance in discourse (Biber, 2006), this sentence clearly exhibits a non-assertive stance. It does not contain any claim verbs, evaluative language, or statements of truth. Instead, the writer fully defers epistemic responsibility to an external source without articulating a position toward the information referenced. For this reason, the claim is categorized as assertive false and epistemically weak.

Being examined through Toulmin's model of argumentation (via Hamble, 1977), the sentence does not form an evaluable argumentative structure. While the reference link may function as a potential backing, it is not integrated into an argument because the sentence itself lacks an explicit claim, data, or warrant. In other words, the link provides access to information but does not constitute argumentative justification within the text. Overall, this case clearly demonstrates that a moderate scientific tone score can emerge in the absence of an explicit scientific claim when a sentence contains formal markers associated with scholarly discourse, such as references to sources. This finding reinforces the central argument of the study that scientific tone measures linguistic resemblance to scientific discourse, whereas assertiveness and claim type reflect epistemic positioning and propositional strength. Accordingly, this sentence serves as a critical example showing that scientific tone and claim strength operate as distinct and non-equivalent analytical dimensions.

CONCLUSION

Based on the findings and discussion presented above, this study concludes that the analysis of claims in online herbal discourse reveals consistent patterns that are analytically significant for understanding the relationship between linguistic assertiveness and scientific quality. This study examines the strength and scientific quality of claims in online herbal discourse by integrating computational text analysis, stance theory, and argumentation theory. Scientific tone is operationalized as a continuous numerical score and classified into low, medium, and high categories using operationally defined thresholds, following automated discourse analysis practices proposed by McNamara et al. (2014). This approach enables systematic and replicable measurement of claim quality across multiple discussion threads.

The synthesis of nine claim sentences drawn from Threads 3, 4, and 15 demonstrates that variation in assertiveness and the use of verbal claim markers does not correspond to meaningful variation in scientific tone. Across the three threads, both assertive true and assertive false claims were identified, as well as claim sentences constructed with and without explicit verbal claim predicates. However, all of these

claims consistently register low scientific tone scores. This pattern indicates that neither the presence of assertive constructions nor the absence of verbal claim markers guarantees alignment with scientific discourse.

Interpreted through Biber's (2006) stance framework, assertive true claims typically display high epistemic commitment through assertive predicates and minimal mitigation, thereby constructing linguistically strong claims. Nevertheless, within the context of herbal discourse, such stance strategies primarily function to enhance persuasive force within a non scientific register rather than to signal epistemic caution. Conversely, claims categorized as assertive false or lacking verbal claim markers do not exhibit higher scientific tone, suggesting that reduced assertiveness alone does not result in greater scientific legitimacy. From the perspective of Toulmin's argumentation model, the claims analyzed across Threads 3, 4, and 15 exhibit systematic argumentative weakness due to the absence of adequate warrants and backing. As a result, claims may appear rhetorically convincing regardless of their assertive status, yet remain epistemically fragile. The integration of computational scoring and stance analysis provides a coherent explanatory framework for understanding how strong rhetorical positioning can coexist with low scientific quality.

Ultimately, this study demonstrates that the analysis of health related claims in digital discourse cannot rely solely on surface level linguistic assertiveness or the presence of verbal claim markers. A combined approach incorporating computational assessment, stance analysis, and argumentation theory is necessary to distinguish between strong and weak claims based on their scientific and argumentative quality. These findings contribute to applied linguistics research and offer methodological insights for the study of health misinformation in online environments.

BIBLIOGRAPHY

Biber, D. (2006). *Stance in discourse: Subjectivity, evaluation, and interaction*. In S. Hunston & G. Thompson (Eds.), *Evaluation in text: Authorial stance and the construction of discourse* (pp. 97–122). Oxford University Press.

Hample, D. (1977). *The Toulmin model of argumentation*. In J. L. Golden, G. F. Berquist, & W. E. Coleman (Eds.), *The rhetoric of Western thought* (pp. 444–460). Kendall/Hunt.

Kristiyanti, M. (2021). Online discourse of traditional medicine in Indonesian digital communities. *Journal of Indonesian Health Communication*, 5(2), 112–128.

Li, Y., Zhang, J., & Yu, B. (2017, September). An NLP analysis of exaggerated claims in science news. In *Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism* (pp. 106–111).

Lim, M. (2013). Many clicks but little sticks: Social media activism in Indonesia. *Journal of Contemporary Asia*, 43(4), 636–657.
<https://doi.org/10.1080/00472336.2013.769386>



McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, 27(1), 57–86. <https://doi.org/10.1177/0741088309351547>

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.

Miner, A. S., Milstein, A., Schueller, S., Hegde, R., Mangurian, C., & Linos, E. (2022). Smartphone-based conversational agents and responses to mental health crises. *JAMA Internal Medicine*, 182(1), 1–8. <https://doi.org/10.1001/jamainternmed.2021.6454>

Nugroho, Y. (2020). Digital participation and public discourse in Indonesia. *Asian Journal of Communication*, 30(3–4), 219–235. <https://doi.org/10.1080/01292986.2020.1785164>

Sardinha, T. B. (2022). Corpus linguistics and the study of social media: A case study using multi-dimensional analysis. In *The Routledge handbook of corpus linguistics* (pp. 656-674). Routledge.

Sharma, U., Saran, S., & Patil, S. M. (2020). Fake news detection using machine learning algorithms. *International Journal of creative research thoughts (IJCRT)*, 8(6), 509-518.

Sujarwo, W., Arinasa, I. B. K., Salomone, F., Caneva, G., & Fattorini, S. (2021). Cultural erosion of traditional medicinal knowledge in Indonesia. *Journal of Ethnopharmacology*, 274, 114–132. <https://doi.org/10.1016/j.jep.2021.114132>

Utami, N. S. (2022). Narratives of herbal medicine in Indonesian online forums. *Humaniora*, 34(1), 45–58.

Wibowo, A. P., & Hidayat, N. (2023). Eksplorasi Linguistik Komputasional dalam Analisis Bahasa Alami untuk Mengungkap Evolusi Dialek Digital di Era Media Sosial Global. *Journal of New Trends in Sciences*, 1(3), 45-52.

Windari, T., Rahmawati, R., & Prasetyo, A. (2022). Public perception of herbal medicine consumption in Indonesia. *Journal of Public Health Research*, 11(2), 227–235. <https://doi.org/10.4081/jphr.2022.227>

Wirawan, A., & Simarmata, J. (2023). Health misinformation on social media during public health crises in Indonesia. *Journal of Communication Inquiry*, 47(1), 77–95. <https://doi.org/10.1177/01968599221125698>

Zhang, Y., Lyu, H., Liu, Y., Zhang, X., Wang, Y., & Luo, J. (2020). Monitoring misinformation related to COVID-19 in online communities. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1), 1–20. <https://doi.org/10.1145/3392831>