



DOI: <https://doi.org/10.15688/jvolsu2.2025.4.4>

UDC 81'1

LBC 81.002



Submitted: 22.03.2025

Accepted: 05.06.2025

BEYOND CONVENTIONALIZED FORMULAE: A LEXICAL BUNDLE APPROACH TO ANALYZING (IM)POLITENESS

Elena S. Gritsenko

MGIMO University, Moscow, Russia

Olivier Mozard T. Kamou

HSE University, Nizhny Novgorod, Russia

Abstract. The paper introduces a frequency-driven lexical bundle approach to enhance politeness research by addressing the limitations of traditional corpus pragmatics approaches, which assume pre-defined (im)politeness values for linguistic forms. While politeness is widely recognized as a dynamic context-dependent phenomenon, corpus pragmatics research risks overlooking its variability by focusing on pre-selected politeness formulae, such as *thank you, please, sorry*, etc. The lexical bundle approach builds on the concept of conventionalization, defined as the process by which linguistic expressions acquire politeness-related meaning through repeated use in a specific context. The proposed approach focuses on the identification of recurring multi-word sequences (lexical bundles), capturing how they index (im)politeness in discourse. We demonstrate the utility of this approach through a case-study of Wikipedia editors' communication, where politeness is strategically negotiated in collaborative, consensus-driven dialogue. By analyzing discussions from Wikipedia talk pages, we categorize the extracted bundles into two groups: conventionalized politeness formulae, which consistently index collaborative politeness across Wikipedia's editorial discourse (e.g., *what do you think, is it possible to*) and multifunctional expressions, that manifest variable politeness-related meanings (e.g., *why don't you, am I missing something*). This distinction reveals how politeness emerges through both stable norms and context-sensitive strategies. Our findings also highlight the capacity of the new approach to bridge Terkourafi's frame-based theory with empirical analysis, offering a replicable framework for studying (im)politeness as a co-constructed interactional practice.

Key words: lexical bundles, (im)politeness, corpus pragmatics, conventionalization, indexicality.

Citation. Gritsenko E.S., Kamou O.M.T. Beyond Conventionalized Formulae: A Lexical Bundle Approach to Analyzing (Im)Politeness. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 2. Yazykoznanie* [Science Journal of Volgograd State University. Linguistics], 2025, vol. 24, no. 4, pp. 41-52. DOI: <https://doi.org/10.15688/jvolsu2.2025.4.4>

УДК 81'1

ББК 81.002

Дата поступления статьи: 22.03.2025

Дата принятия статьи: 05.06.2025

ПРАГМАТИКА КОНВЕНЦИОНАЛИЗАЦИИ: ЛЕКСИЧЕСКИЕ СВЯЗКИ КАК ИНСТРУМЕНТ АНАЛИЗА (НЕ)ВЕЖЛИВОСТИ

Елена Сергеевна Гриценко

Московский государственный институт международных отношений (университет) МИД России,
г. Москва, Россия

Оливье Мозард Т. Каму

Национальный исследовательский университет «Высшая школа экономики», г. Нижний Новгород, Россия

Аннотация. В статье представлен новый подход к изучению (не)вежливости, направленный на преодоление ограничений традиционно используемых в корпусной прагматике подходов, которые исходят из фиксированных, заранее приписанных языковым формам значений. Несмотря на то что в современной лингви-

стике вежливость признается динамичным явлением, зависящим от контекста, исследования вежливости в корпусной прагматике часто упускают эту особенность, фокусируя внимание на традиционных маркерах вежливости, таких как *please, thank you, sorry* и подобных. Предлагаемый в работе подход опирается на понятие конвенционализации, отражающее процесс, в ходе которого языковые выражения приобретают значения, связанные с вежливостью, благодаря повторяющемуся употреблению в определенных коммуникативных ситуациях. Данный подход основан на выделении из текстов рекуррентных последовательностей слов (лексических связей), способных индексировать (не)вежливость в дискурсивном контексте. Эффективность предлагаемого подхода демонстрируется на примере коммуникации редакторов Википедии, где вежливость стратегически конструируется в рамках дискурса, ориентированного на сотрудничество и консенсус. Выделенные из данного дискурса лексические связи, разделены на две группы: конвенционализированные формулы, стабильно выражающие вежливость (*what do you think, is it possible to*), и многофункциональные связи (*why don't you, am I missing something*), вежливость (или невежливость) которых задается контекстом. Новый подход позволяет связать фреймовую теорию вежливости М. Теркурафи с эмпирическим анализом, расширяя возможности для изучения (не)вежливости как феномена, конструируемого коммуникантами в процессе социального взаимодействия.

Ключевые слова: лексическая связь, (не)вежливость, корпусная прагматика, конвенционализация, индексальность.

Цитирование. Гриценко Е. С., Каму О. М. Т. Прагматика конвенционализации: лексические связи как инструмент анализа (не)вежливости // Вестник Волгоградского государственного университета. Серия 2, Языкознание. – 2025. – Т. 24, № 4. – С. 41–52. – (На англ. яз.). – DOI: <https://doi.org/10.15688/jvolsu2.2025.4.4>

Introduction

Politeness, often perceived as an intrinsic quality of certain linguistic forms, is far from a fixed attribute embedded in language units; rather, it emerges as a dynamic, context-dependent phenomenon co-constructed by interlocutors in the act of communication. Consider, for instance, the phrase *excuse me*, widely regarded as a quintessential marker of politeness in English. In a crowded bus, it might soften a request to pass through, signaling deference and respect; yet, uttered with a sharp tone or paired with an exasperated sigh, it can just as easily convey irritation or annoyance. This duality underscores that politeness is not inherent in the phrase itself but arises from the interplay of situational cues, intonation, and the relational history between speakers. Previous research [Watts, 2003; Terkourafi, 2005] has highlighted this feature, arguing that linguistic expressions gain their polite force through strategic use within specific social contexts. Thus, politeness reveals itself as a collaborative negotiation, shaped continuously by those engaged in the exchange [Haugh, 2007].

However, the long-standing debate over whether politeness is intrinsic in linguistic forms or contextually emergent evolves with new methodologies. While the argument that politeness is not inherently tied to linguistic form, remains valid, it is also undeniable that certain structures routinely carry (im)politeness

meanings; for instance, “You + N” (e.g., *You moron*) is typically offensive and therefore impolite [Culpeper, Gillings, 2018; Van Olmen, Andersson, Culpeper, 2023]. This suggests that linguistic form and context constantly interact in shaping (im)politeness. Corpus-based methods provide a way to explore this relationship quantitatively. Yet, corpus studies often inherit the limitations of the traditional approaches assuming predefined (im)politeness values for linguistic forms. This focus on default interpretations risks overlooking the contextual variability of politeness, making it difficult to identify various politeness-related meanings a form can take in different linguistic and social contexts [Jucker, Staley, 2017].

Our goal in this study is to propose a way to move beyond these limitations by introducing a frequency-driven approach, which uses lexical bundles (recurring multi-word sequences) to identify politeness-related meanings in discourse. Unlike traditional methods, a lexical bundle approach does not presuppose politeness values in linguistic forms; instead, it identifies patterns emerging from the data itself, capturing how specific bundles index (im)politeness across diverse contexts. In what follows, we will argue why the lexical bundle approach offers a promising complementary corpus-based method for studying (im)politeness and demonstrate its utility through a case study of Wikipedia editors’ communication – a context where politeness is strategically

negotiated in collaborative, consensus-seeking discourse.

We will briefly review the history of politeness research – not with the aim of being exhaustive, as this has been thoroughly covered (e.g.: [Culpeper, 2011; Brown, 2017; Locher, Larina, 2019]) – but to situate a lexical bundle approach within the broader field of politeness studies. Next, we outline our methodology and apply it to Wikipedia’s editor discourse. Finally, we discuss the advantages of the proposed approach and its implications for corpus-based politeness studies.

Theoretical background

Waves of politeness research: a brief overview

The study of politeness has evolved through several theoretical shifts, often described in three “waves” [Grainger, 2011; Culpeper, 2011]. The first wave, dominated by the seminal work of Brown and Levinson [1987], framed politeness as a universal strategy rooted in face-saving behaviors, where speakers employ positive or negative politeness tactics to mitigate threats to their interlocutors’ social image. This approach, influenced by Goffman’s [1967] concept of face, prioritized systematic, rule-based models and was later critiqued for ignoring the dynamic, context-dependent nature of politeness and for its assumed universality and Western-centric bias [Eelen, 2001; Mills, 2011].

These criticisms led to the second wave, or post-modern perspective, which shifted focus toward a more discursive approach, with scholars like Watts [2003; 2005] and Locher [2004; 2006] arguing that politeness is a subjective evaluation negotiated in interaction, shaped by power dynamics and cultural expectations. This phase emphasized the fluidity of politeness-related meanings, challenging the universality of earlier frameworks.

The third wave, which has been gaining momentum in recent years, builds on this by viewing politeness as a co-constructed practice embedded in specific communities and contexts [Kádár, Haugh, 2013]. Notable frameworks include Spencer’s [2005] relational approach, Terkourafi’s [2001; 2005] frame-based approach,

Haugh’s [2007] interactional approach, and Kádár’s [2017] rituals, among others. Terkourafi’s approach, which takes a quantitative perspective on politeness by seeking regularities in the use of expressions and their extra-linguistic context, is of special relevance to this study.

Frame-based approach to politeness

Terkourafi critically examined the existing politeness theories and argued that, while differing significantly, they share certain underlying assumptions. Both traditional and post-modern approaches tend to focus on micro-level interactions (specific utterances or exchanges) and often overlook broader social structures or patterns, which limits the scope of politeness research. She proposed to shift attention from isolated utterances to recurring patterns of interaction within specific social contexts or “frames,” arguing that “it is the regular co-occurrence of particular types of context and particular linguistic expressions as the unchallenged realizations of particular acts that create the perception of politeness” [Terkourafi, 2005, p. 248]. Her frame-based view offers a way to study politeness empirically, using observable patterns rather than relying solely on theoretical constructs or subjective interpretations. It incorporates quantitative analysis of linguistic behavior to identify norms and habitual meanings inferred from recurring expressions in context.

In later studies Terkourafi refined this approach further; she argued that when expressions regularly appear in a particular context, they become conventionalized – i.e., accrue socially recognized meanings tied to polite (or impolite) interaction [Terkourafi, 2011; 2015]. These conventionalized expressions, such as ritualized apologies or deference markers, emerge as shared resources within speech communities, combining predictability with adaptability to context. From this perspective, conventionalization involves “already meaningful linguistic expressions acquiring additional meaning that can be put at the service of (im)politeness” [Terkourafi, Kádár, 2017, p. 181].

Though the frame-based approach has been widely recognized, it does not provide a detailed replicable methodology for identifying conventionalized formulae. Previous corpus-based

studies of politeness that have adopted this framework (e.g.: [Culpeper, Gillings, 2018; Van Dorst, Gillings, Culpeper, 2024; Xie, 2024]) focused solely on preselected (conventionalised) politeness formulae, such as *sweetheart*, *good morning*, *thank you*, *please*, and categorized them in accordance with predefined politeness types [Culpeper, Gillings, 2018; Van Dorst, Gillings, Culpeper, 2024]. A lexical bundle approach, which involves identifying and analyzing frequently recurring multi-word sequences in language use, can complement the frame-based approach, providing an objective, data-driven way to study (im)politeness.

***Lexical bundle approach:
methodological implications
for (im)politeness research***

Lexical bundles, as defined by Biber et al. [1999], are recurring multi-word sequences, such as *I don't know if* or *I just wanted to*, identified using corpus analysis software based on frequency and distribution criteria [Biber, 2009, p. 282]. They are typically extracted from spoken or written corpora and reflect recurring patterns tied to specific discourse contexts. Unlike other formulaic expressions, lexical bundles are not inherently complete structural units, yet they exhibit strong grammatical and functional correlates [Conrad, Biber, 2005; Biber et al., 1999]. They function as building blocks of discourse, enhancing its structure and coherence [Biber, 2009; Hyland, 2008a]. Lexical bundles have been extensively studied in academic discourse, where they serve as stance markers, discourse organizers, and referential expressions [Conrad, Biber, 2005; Biber, 2006; Hyland 2008a]. Their potential for (im)politeness research remains underexplored, although some studies suggest that their use in discourse can be shaped by politeness concerns [Conrad, Biber, 2005; Xia, Ai, Pae, 2022; Gritsenko, Kamou, 2024].

Recent developments in corpus pragmatics have facilitated systematic analysis of (im)politeness, particularly the interplay between form and function [Baider, Cislaru, Claudel, 2020; House, Kádár, Leontyev, 2021; Leontovich, Nikitina, 2024]. Terkourafi's approach, which perfectly aligns with corpus analysis, can be complemented by the lexical bundle approach in several ways.

Terkourafi views politeness as emerging from cognitive frames – schemata linking linguistic expressions to minimal contexts (social settings, relationships) based on prior experience. The lexical bundle approach provides a tangible way to identify and study them, grounding politeness in how it actually manifests in everyday interaction. Recurring bundles (e.g., *I'd like to* in suggestions, *would you mind if* in requests) can be seen as linguistic manifestations of abstract frames, reflecting the conventional pairing of form and context.

Terkourafi and Kádár also called for “methodologically sound ways” to identify conventionalized formulae, one of which is “frequency counts” [Terkourafi, Kádár, 2017, p. 190]. Lexical bundles, defined by their frequency and fixedness, directly capture this conventionalization, offering a frequency-driven alternative to predefined (politeness) formulae. Their recurrence in a given context may lead to the “emergence of stable associations” [Zlov, Zlatev, 2024], incorporating not only conventionalized forms but also non-conventionalized expressions that can acquire politeness-related meanings in discourse, as “candidate for politeness become conventionalised to some degree for a particular context of use” [Van Olmen, Andersson, Culpeper, 2023, p. 24].

The lexical bundle approach leverages on their key functions in discourse. Biber stressed that lexical bundles “provide a kind of pragmatic head for larger phrases and clauses, where they function as discourse frames for the expressions of new information” [Biber, 2009, p. 284]. This is directly applicable to (im)politeness, since it emerges not in isolated phrases, but in how the speaker frames them in interaction to achieve an illocutionary goal. Therefore, the proposed approach incorporates not only quantitative but also qualitative analysis, as recontextualizing the identified forms is imperative to determine the pragmatic functions or politeness-related meanings that they take or convey [Jucker, Staley, 2017].

Material and method

Identifying lexical bundles in corpus-based politeness research requires a systematic approach grounded in corpus pragmatics. Comprehensive guidelines for data selection and analysis are detailed by Jucker [2018a; 2018b]. As he notes “The starting point is not a particular

linguistic unit and how speakers use this unit in interaction, but rather particular types of interaction and the effects that such interaction has on the participants. This kind of research looks into the effects of communication and searches for elements that create these effects such as terms of address, specific mitigators or speech acts” [Jucker, 2018a, p. 11].

The next step involves determining the length of the bundles to be investigated and setting the minimum frequency threshold. Four-word bundles are widely adopted as they offer a greater structural and functional variety [Biber, 2009]. The frequency criteria typically range from 5 to 40 occurrences per million, contingent on corpus size and research goals [Biber et al., 1999; Biber, 2006; Cortes, 2004; Hyland, 2008a; 2008b; Flinn, 2023]. To avoid idiosyncratic usage, lexical bundles must recur across multiple texts (e.g., 3–5 texts or 10% of the corpus). Automated tools with N-gram functions, such as AntConc, WordSmith Tool, or Python, facilitate extraction. Post-extraction manual screening is critical to eliminate formulaic but pragmatically irrelevant bundles (e.g., *the rest of the*) and retain those salient to interactional dynamics.

Lexical bundles analyzed in this paper were extracted from the Wikipedia section of the Stanford Politeness Corpus, which consists of 4,353 conversations, totaling 112,277 words [Danescu-Niculescu-Mizil et al., 2013]. To ensure methodological robustness, we identified four-word bundles occurring at least five times across three conversations. Bundles with minimal pragmatic relevance to politeness (e.g., *at the bottom of the*, *the rest of the*) were excluded. Our analysis focused instead on bundles implicated in direct interpersonal negotiation, particularly those functioning within requests, suggestions, and critiques directed at interlocutors. A total of 115 bundles and 1,241 tokens have been extracted from the corpus.

Results and discussion

This section demonstrates how the lexical bundle approach illuminates subtle mechanisms for indexing (im)politeness in collaborative discourse, using conversations among Wikipedia editors as a case study. Wikipedia’s editors constitute a community of practice bound by shared norms, most notably the principle of

“assuming good faith.” This principle, central to Wikipedia’s collaborative culture [Reagle, 2010], obligates editors to interpret other’s actions with good will and engage constructively – a requirement that shapes politeness strategies in their interactions.

The bundles derived from the Stanford Politeness Corpus exhibit varying capacities to convey politeness-related meanings. Accordingly, we categorized them into two primary groups. The first group includes bundles which have been previously identified as conventional politeness formulae (e.g., *thank you for your*, *could you tell me*, *could you please explain*, *might be able to*, *would you like*, *would you be interested*, *would you care to*, etc.). The second group encompasses lexical bundles that have not been previously associated with politeness or impoliteness. In our analysis, however, they can index various politeness-related meanings, depending on the context, tone, and the interlocutors’ intentions – framed as collaborative or confrontational. In this paper, we concentrate on the second group of bundles. This group comprises phrases that consistently index collaborative politeness across Wikipedia’s editorial discourse (e.g., *what do you think*) and bundles that manifest variable politeness-related meanings: in certain interactions, they signify politeness, while in others, they may suggest mild impoliteness or assume a neutral, factual/functional tone (e.g., *why don’t you*). Lexical bundles that are consistently polite can be regarded as discourse-specific conventionalized politeness formulae, while lexical bundles characterized by inconsistent politeness indexing, we designate as multifunctional. This distinction highlights the versatility of the lexical bundle approach in illuminating the nuanced dynamics of politeness in discourse communities.

To demonstrate the applicability of the lexical bundle approach, we will analyze representative examples from both subgroups, highlighting their roles in negotiating politeness or indexing critique and disagreement in Wikipedia’s collaborative discourse.

Conventionalized politeness formulae

A prime example of this subgroup is *what do you think*, the most frequent lexical bundle in

our dataset (94 tokens). In all conversations, it signals respect and cooperative attitude, aligning with Wikipedia consensus-driven ethics. The bundle can accompany suggestions, softening potential imposition (1); request engagement, encouraging feedback rather than demanding commitment (2); invite consensus, fostering group decision-making (3); and solicit agreement, showing respect for collaboration (4).

(1) How about saying that the fastest flow is near the deepest part of the channel, and that in most meanders this is near the outer bank? **What do you think?** (conversation_137);

(2) I see you are quite zealous on Croatian articles, so I am inviting you to support my <url>. **What do you think?** (conversation_1009);

(3) <url> posted a wonderful grammatical analysis of the 'has/had' thing, another editor agreed, so I'm thinking we should make another request for the change, using Bluewave's post as the basis, **what do you think?**" (conversation_1533);

(4) After looking through a bunch of those, I think the lead should be formatted like this: a paragraph explaining who he is, a paragraph summarizing his life and career, and a paragraph summarizing why he is so great. **What do you think?** (conversation_670);

(5) I think the article needs work. For example, I think 'U.S. late 1980s-1990s' should after 'Detroit sound'. 'UK 1990s should after "U.S. late 1980s-1990s". **What do you think?** (conversation_147).

In editors' interactions, this bundle never conveys rudeness. Even in assertive contexts, it mitigates potential conflict (5). It recurs across all conversations as a standard way to close a suggestion, observation, or request, signaling a norm of seeking input in this community. The conventional status of this bundle in Wikipedia talk pages stems from its role in maintaining harmony and progress in a volunteer-driven argumentative space.

The term "conventionalized politeness formulae" refers to a fixed or semi-fixed phrase that has become standardized within a specific community to express politeness reliably and predictably. Such phrases are recognized by speakers due to their frequent use and shared understanding; sometimes they can lose their literal force to serve a pragmatic function. This fully applies to the bundle *is it possible to* (11 tokens), which softens requests, suggestions, or inquiries, indexing politeness through indirectness and deference. While the number of its tokens is

limited, the bundle's uniform politeness suggests that it may be a recurring pattern in Wikipedia talk pages. Its use across varied contexts – technical edits (6), sourcing vandalism protection (7), etc. – supports its potential frequency.

(6) I noticed that you created the Close AFD script. **Is it possible to** edit it so that it works? (conversation_3340);

(7) I have just reported a high level of Vandalism at the page <url>. **Is it possible to** impose a temporary protection to the page (conversation 1105).

In all cases across our dataset, this bundle serves as a polite way to prompt action or seek cooperation. This pragmatic shift – where the literal question becomes a conventional request – mirrors such politeness formulae as *can/could you* or *would it be ok to*. The repeated polite use of this bundle indicates its conventionalization within this specific discourse, where editors implicitly recognize it as a respectful way to engage.

The bundle *do you think it* (22 tokens) normally functions as a polite elicitation of opinion or feedback, indexing respect for the interlocutor's expertise (8). It embeds gratitude and deference, framing the question as a collaborative appeal rather than a demand (9). The bundle is often used to seek validation, reinforcing a cooperative stance (10). This aligns with Brown and Levinson's notion of politeness as a face-saving strategy, where *do you think it* serves as a negative politeness tactic, reducing imposition by appealing to the other person's autonomy.

(8) Good luck with the remaining time for your nomination. BTW: under what section on <url> **do you think it** should appear? (conversation_2491);

(9) "Thanks for the copyedit. Rather than waiting for the PR rigamarole, **do you think it** meets GA criteria and should I submit it?" (conversation_4314);

(10) "I added a clarifying sentence to my user page <url>. **Do you think it** distills the essence? (conversation_2009).

The bundle *I don't see* (27 tokens) indexes a range of politeness meanings, though its effect is context-dependent. The bundle's potential as a discourse-specific conventionalized politeness formula rests on its recurring use as a softened expression of disagreement, uncertainty, or a

request for clarification, mitigating face threats in a collaborative editorial environment. Prefaced by an apology, the bundle softens a disagreement by avoiding direct challenge and inviting further explanation (11). Paired with a courteous request, it signals openness and deference to the other editor's insight, which aligns with a negative politeness strategy of minimizing imposition (12, 13). In these and other cases, *I don't see* softens the implicit challenge to the other person's claim, framing it as the speaker's own limitation (lack of perception) rather than an outright accusation ("It is not there"), which suggests collaborative tone and avoids confrontation. However, the bundle's politeness can wane or shift to neutrality when it expresses skepticism that could be interpreted as a challenge, though its polite force persists if read as an invitation to justify rather than an outright dismissal (14).

(11) I'm sorry, but **I don't see** the value in having different criteria for two versions of the same list, one sorted by name and one sorted by nationality. Every other list and article on space travel uses the altitude definition, why would the sub-list sorted by nationality be different?" (conversation_595);

(12) **I don't see** the vandalism. Could you show me where it is? (conversation_2207);

(13) You mentioned a CU done with a possible result, but **I don't see** it at SPL. Can you kink TC's CU and name the master? (conversation_369);

(14) **I don't see** how you can use this to discount his trustworthiness and oppose him. If there is consensus to promote among several bureaucrats, what's the matter who actually presses the button? (conversation_658).

As a discourse-specific conventionalized politeness formula, this bundle fits partially within Terkourafi's framework of expressions that gain pragmatic weight through repeated use in a specific community. In Wikipedia's editorial context, where precision, consensus, and civility are valued, the phrase recurs as a recognizable way to express non-comprehension or disagreement without escalating tension. Its conventionality is strengthened by frequent pairing with polite markers (e.g., *I'm sorry, could you*) or requests for clarification, embedding it in the community's norm of constructive dialogue.

Similarly, the indexical meaning of *I don't understand* (15 tokens) hinges on its recurring use as a means to navigate uncertainty, seek

clarification, or express disagreement without overt confrontation. The bundle is routinely used to mitigate any potential face threat to the interlocutor, framing the speaker's confusion as a personal limitation and inviting collaboration (15). Yet, it can also lean to neutrality or mild impoliteness when it expresses frustration or disagreement, potentially indexing a challenge to the other's reasoning (16).

(15) I'm afraid **I don't understand** what you are asking. Could you perhaps try to explain it in another way? (conversation_176);

(16) I'm sorry, but there is no discussion on the talk page that indicates that there was a debate about renaming, so the keep vote stands, I'm afraid. There were nine votes to keep, and only one to merge, so **I don't understand** how a single vote trumps nine others? (conversation_278).

In Wikipedia's editorial discourse, this bundle recurs as a recognizable pattern for signaling confusion or requesting assistance, most often with polite intent. Its conventionalization is reinforced by frequent linkage with clarification requests, a characteristic feature of this community's collaborative ethics. However, unlike more definite polite formulae (e.g., *thank you*), the bundle's politeness is not guaranteed; it can also function as a neutral statement of perception or convey a subtle critique. Thus, while it qualifies as a conventionalized expression in this discourse, its status as a politeness formula is context-sensitive.

Multifunctional bundles

While conventionalized politeness formulae reflect stable politeness norms within Wikipedia's collaborative culture, multifunctional bundles demonstrate how politeness is dynamically negotiated through context-dependent linguistic strategies, underscoring the interplay between form and interactional content.

The bundle *do you have any*, the second most frequent one in our dataset (49 tokens), does not function as a politeness marker on its own; its perceived politeness – or lack of it – depends heavily on the surrounding sentence structure and intent of the speaker. Within Wikipedia's editor community, where concise, task-oriented dialogue is common, the phrase seems more functional than

conventional, serving to solicit information, opinions, or resources rather than to signal deference or soften a request. When it functions as a direct query for information or evidence, aligning with Wikipedia's focus on verifiable content, the phrase is neutral – neither markedly polite nor impolite (17). When it follows a statement that implies criticism or frustration, it could be read as a challenge, pressing the recipient to justify themselves (18), and when paired with softer phrasing or an offer of collaboration, the phrase feels more polite (19).

(17) Hi, you recently edited that the Independence Hall was once the tallest building in Philadelphia... **Do you have any** source that supports this?" (conversation_56);

(18) So, you don't like people pointing out the errors you make? Or **do you have any** actual arguments to counter my post? (conversation_609);

(19) Getting this article up to FA, but I'm having trouble finding sources... **Do you have any** advice or would you be interested in working on it together? (conversation_138).

It is noteworthy that the pronoun "any" as part of this bundle introduces indefiniteness, softening the expectation of a specific answer (e.g., *Do you have sources?* vs *Do you have any sources?*). This subtle openness might push the phrase toward politeness in some cases, although it is not a defining feature.

The bundle *why don't you* (19 tokens) presents a compelling case for examining how politeness emerges not as an inherent property of linguistic forms, but as a contextually negotiated outcome shaped by intent, tone, and interactional dynamics. In Wikipedia editor exchanges, the indexical potential of this bundle varies. It spans politeness, neutrality, and impoliteness (mild to strong), with tone and context driving the shift. The bundle can function as a practical suggestion, phrased as advice rather than a demand (20), or a collaborative invitation, softening the directive (21). Yet, in our data the bundle often implies criticism or challenges the interlocutor. This is achieved by framing requests as reprimands (22, 23) or using sarcasm to mock the other person's focus (24).

(20) **Why don't you** mix 2 articles together? (conversation_3708);

(21) Zero has taken a break from Wikipedia, and RK has recused himself from editing controversial articles. and I fix up the <url> article now? **Why don't you** and I fix up the <url> article now? (conversation_4162);

(22) So, I'm a «deletion fanatic»? **Why don't you** refrain from uncivil remarks and focus on adding sources to your unsourced stubs – help the encyclopedia rather than pissing off people?" (conversation_3184);

(23) No, not funny, unless you are a mentally challenged person. **Why don't you** find a better hobby where you don't post lies and offend the people of Tibet? (conversation_2096);

(24) You seem to have plenty of time to discuss, why don't you try to educate <url>? Maybe he needs this more than me?" (conversation_2608).

Although the bundle leans impolite in heated exchanges, and its negative structure (*why not?*) can imply critique, it is not conventionalized as impolite in this setting. Its impoliteness usually emerges if paired with criticism (e.g., *uncivil remarks*, *unsourced stubs*, *pissing off*), not from the bundle alone. In neutral cases, it aligns with Wikipedia's task-oriented collaboration.

The bundle *am I missing something* (6 tokens) is usually seen as a hedge, which serves as a neutral or polite way to seek clarification. In this case, it softens a potential challenge by attributing confusion to the speaker's own perspective, inviting collaboration rather than confrontation (25). However, when paired with sarcasm, frustration, or skepticism – common in contentious editing disputes – the phrase can come across as passive-aggressive, subtly challenging the other party's reasoning or competence. Such cases are prevalent in our data. In contentious conversations, *am I missing something* is read as a rhetorical jab (26) or an accusation (27) rather than a polite inquiry.

(25) You reverted my wording changes here, pointing to the talk page, but the last comment there is my rationale for the rewording from back when I revamped it. **Am I missing something?** (conversation_1125);

(26) I'm no expert on the language, but on the face of it that dab page disambiguates two words which are unambiguous anyway – Zakhmet and Khachpas. **Am I missing something?** (conversation_1998);

(27) Without participating in talk page discussion, you have deleted reliable sources, and rewritten a sentence so that it is supported by no sources at all except your personal opinion. <url> **Am I missing something** here? (conversation_752).

Thus, the indexical meaning of this bundle is flexible, depending on its surroundings. In Wikipedia's editor discourse, it leans toward skeptical or sarcastic, although not outright rude.

The bundle *why did you remove* (14 tokens) in Wikipedia editors' discourse carries a range of tones, from polite curiosity to implied criticism, depending on the context and accompanying language. Unlike *am I missing something*, which probes for clarification about the situation, *why did you remove* directly questions another editor's action, making it inherently more pointed. However, when softened by greetings, gratitude, or collaborative framing and paired with a request for clarification, the bundle reads as a good-faith effort to understand or discuss (28). Impoliteness emerges when the bundle is paired with defensiveness or a tone suggesting that the removal was wrong. For instance, in the example below (29), acknowledging partial agreement *I understand* is conciliatory, but the epithet *excessive* introduces criticism, and emphatic *ALL* adds a dramatic or exasperated flair. Likewise, in the following sample (30), the redundancy of the question *what was the reason?* after *why* adds a touch of frustration, though it is still restrained.

(28) Thanks for your recent edits on <url>, but what does 'inapplicable category' (referring to <url>) in <url> mean? And **why did you remove** that category from <url>? (conversation_383);

(29) I understand the problem with that redundant category – I've hashed that out, I think – but **why did you remove** ALL categories from the Christian shows? That seems... excessive... Or do I just need a nap? (conversation_1781);

(30) **Why did you remove** my comment from your page? What was the reason for that? (conversation_924).

In the collaborative ethos of Wikipedia, which prioritizes constructive dialogue, the last three bundles tend to function as veiled expressions of dissent. This strategic use of language highlights a tension within the editorial community where the need to be polite and maintain a cooperative atmosphere coexists with the inevitable interpersonal friction, revealing how editors navigate disagreement through linguistically nuanced means.

Conclusion

The analysis of lexical bundles in Wikipedia editors' discourse reveals the nuanced ways in which politeness and impoliteness are indexed, extending beyond traditional markers to include multifunctional and context-sensitive phrases. The lexical bundle approach proves effective in uncovering these subtle mechanisms, highlighting how politeness emerges as a dynamic, co-constructed phenomenon rather than a fixed property of language. Discourse-specific conventionalized politeness formulae, such as *what do you think* and *is it possible to*, demonstrate a stable polite force within this community, reflecting Wikipedia's consensus-driven ethos and emphasis on collaboration. These bundles, through frequent and predictable use, have become standardized tools for softening requests, soliciting input, and maintaining harmony in a volunteer-driven argumentative space.

Conversely, multifunctional bundles (e.g., *do you have any*) exhibit a remarkable versatility, indexing politeness, neutrality, or impoliteness depending on tone, context, and intent. This variability underscores the strength of the lexical bundle approach in capturing the fluidity of politeness as a negotiated outcome, aligning with Watts' [2003] discursive perspective. Notably, some bundles in our dataset are commonly used to index disagreement (e.g., *why did you remove*), which can shift from polite inquiry to implied critique based on accompanying language. Similarly, phrases like *am I missing something* and *why don't you* routinely index sarcastic undertones in contentious exchanges, subtly challenging interlocutors' actions or competence while maintaining an appearance of civility. This dual potential – polite or provocative – illustrates how lexical bundles serve as strategic resources for navigating the tension between Wikipedia's collaborative norms and the inevitable conflicts of editorial work.

The concept of conventionalization bridges the lexical bundle approach and politeness research by explaining how certain phrases gain pragmatic weight through repeated use within a specific community. While conventionalized formulae offer reliability, multifunctional bundles reveal the adaptability of language to situational demands, enriching our understanding of politeness as both a norm and a negotiation.

The study showed that the lexical bundle approach enhances politeness research by illuminating unobvious linguistic strategies, particularly in discourse communities where civility and critique coexist. Future research could explore how these patterns evolve across different online platforms or cultural contexts, further testing the interplay between conventionalization and contextual dynamics in expressing politeness.

REFERENCES

- Baidier F.H., Cislaru G., Claudel C., 2020. Researching Politeness: From the 'Classical' Approach to Discourse Analysis... and Back. *Corpus Pragmatics*, vol. 4, pp. 259-272. DOI: <https://doi.org/10.1007/s41701-020-00088-8>
- Biber D., 2006. *University Language: A Corpus-Based Study of Spoken and Written Registers*. Amsterdam, John Benjamins. 261 p.
- Biber D., 2009. A Corpus-Driven Approach to Formulaic Language in English. *International Journal of Corpus Linguistics*, vol. 14, iss. 3, pp. 275-311. DOI: <https://doi.org/10.1075/ijcl.14.3.08bib>
- Biber D., Johansson S., Leech G., Conrad S., Finegan E., 1999. *Longman Grammar of Spoken and Written English*. London, Longman. 1204 p.
- Brown P., 2017. Politeness and Impoliteness. Yan Huang, ed. *Oxford Handbook of Pragmatics*, pp. 383-399. Oxford, Oxford University Press.
- Brown P., Levinson S.C., 1987. *Politeness: Some Universals in Language Usage*. Cambridge, Cambridge University Press. 358 p.
- Conrad S. M., Biber D., 2005. The Frequency and Use of Lexical Bundles in Conversation and Academic Prose. *Lexicographica*, vol. 20, pp. 56-71. DOI: <https://doi.org/10.1515/9783484604674.56>
- Cortes V., 2004. Lexical Bundles in Published and Student Disciplinary Writing: Examples from History and Biology. *English for Specific Purposes*, vol. 23, iss. 4, pp. 397-423. DOI: <https://doi.org/10.1016/j.esp.2003.12.001>
- Culpeper J., 2011. Politeness and Impoliteness. Aijmer K., Andersen G., eds. *Pragmatics of Society*. Berlin, Mouton de Gruyter, pp. 393-438.
- Culpeper J., Gillings M., 2018. Politeness Variation in England: A North-South Divide. Brezina V., Love R., Aijmer K., eds. *Corpus Approaches to Contemporary British Speech: Sociolinguistic Studies of the Spoken BNC2014*. S.l., Routledge, pp. 33-59. DOI: <https://doi.org/10.4324/9781315268323-5>
- Danescu-Niculescu-Mizil C., Sudhof M., Jurafsky D., Leskovec J., Potts C., 2013. A Computational Approach to Politeness with Application to Social Factors. *Proceedings of ACL*. URL: <https://nlp.stanford.edu/pubs/politeness.pdf>
- Eelen G., 2001. *A Critique of Politeness Theories*. Manchester, St Jerome. 280 p.
- Flinn A., 2023. How Often Do Pauses Occur in Lexical Bundles in Spoken Native English Speech? *Corpus Pragmatics*, vol. 7, iss. 4, pp. 303-322. DOI: <https://doi.org/10.1007/s41701-023-00146-x>
- Goffman E., 1967. *Interaction Ritual: Essays on Face-To-Face Behavior*. Garden City, New York, Doubleday. 270 p.
- Grainger K., 2011. "First Order" and "Second Order" Politeness: Institutional and Intercultural Contexts. *Discursive Approaches to Politeness*. Berlin, Mouton de Gruyter, pp. 167-188.
- Gritsenko E.S., Kamou O.M.T., 2024. Academic English Melting Pot: Reconsidering the Use of Lexical Bundles in Academic Writing. *Russian Journal of Linguistics*, vol. 28, no. 3, pp. 615-632. DOI: <https://doi.org/10.22363/2687-0088-39663>
- Haugh M., 2007. The Discursive Challenge to Politeness Theory: An Interactional Alternative. *Journal of Politeness Research*, vol. 3, pp. 295-317. DOI: <https://doi.org/10.1515/PR.2007.013>
- House J., Kádár D.Z., Leontyev V.V., 2021. Ritual Frame Indicating Expressions (An Academic Conversation). *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 2. Yazykoznanie* [Science Journal of Volgograd State University. Linguistics], vol. 20, no. 2, pp. 42-46. DOI: <https://doi.org/10.15688/jvolsu2.2021.2.4>
- Hyland K., 2008a. As Can Be Seen: Lexical Bundles and Disciplinary Variation. *English for Specific Purposes*, vol. 27, iss. 1, pp. 4-21. DOI: <https://doi.org/10.1016/j.esp.2007.06.001>
- Hyland K., 2008b. Academic Clusters: Text Patterning in Published and Postgraduate Writing. *International Journal of Applied Linguistics*, vol. 18, no. 1, pp. 41-62. DOI: <http://dx.doi.org/10.1111/j.1473-4192.2008.00178.x>
- Jucker A.H., Staley L., 2017. (Im)politeness and Developments in Methodology. Culpeper J., Haugh M., Kádár D.Z., eds. *The Palgrave Handbook of Linguistic (Im)politeness*. London, Palgrave Macmillan, pp. 403-429. DOI: https://doi.org/10.1057/978-1-137-37508-7_16
- Jucker A.H., 2018a. Data in Pragmatic Research. Jucker A.H., Schneider K.P., Bublitz W., eds. *Methods in Pragmatics*. Berlin, De Gruyter, pp. 3-36. DOI: <https://doi.org/10.1515/9783110424928-001>
- Jucker A.H., 2018b. Introduction to Part 5: Corpus Pragmatics. Jucker A.H., Schneider K.P., Bublitz W., eds. *Methods in Pragmatics*. Berlin, De Gruyter Mouton, pp. 455-466. DOI: <https://doi.org/10.1515/9783110424928-018>

- Kádár D., 2017. *Politeness, Impoliteness, and Ritual: Maintaining the Moral Order in Interpersonal Interaction*. Cambridge, Cambridge University Press. 427 p.
- Kádár D.Z., Haugh M., 2013. *Understanding Politeness*. Cambridge, Cambridge University Press. 295 p.
- Leontovich O.A., Nikitina A.V., 2024. Communication Monitoring as a Politeness Mechanism. *Training, Language and Culture*, vol. 8, iss. 3, pp. 62-72. DOI: <https://doi.org/10.22363/2521-442X-2024-8-3-62-72>
- Locher M., 2004. *Power and Politeness in Action. Disagreements in Oral Communication*. Berlin, Mouton de Gruyter. xvi, 365 p.
- Locher M., 2006. Polite Behaviour Within Relational Work: The Discursive Approach to Politeness. *Multilingua*, vol. 25, iss. 3, pp. 249-267.
- Locher M.A., Larina T.V., 2019. Introduction to Politeness and Impoliteness Research in Global Contexts. *Russian Journal of Linguistics*, vol. 23, iss. 4, pp. 873-903. DOI: <https://doi.org/10.22363/2687-0088-2019-23-4-873-903>
- Mills S., 2011. Discursive Approaches to Politeness and Impoliteness. *Discursive Approaches to Politeness*. Berlin, De Gruyter Mouton, pp. 19-56. DOI: <https://doi.org/10.1515/9783110238679.19>
- Reagle J.M., 2010. *Good Faith Collaboration: The Culture of Wikipedia*. Cambridge, The MIT Press. 261 p.
- Spencer-Oatey H., 2005. (Im)Politeness, Face and Perceptions of Rapport: Unpackaging Their Bases and Interrelationships. *Journal of Politeness Research*, vol. 1, no. 1, pp. 95-119. DOI: <https://doi.org/10.1515/jplr.2005.1.1.95>
- Terkourafi M., 2001. *Politeness in Cypriot Greek: A Frame-Based Approach*. PhD Thesis. Cambridge, University of Cambridge. 272 p.
- Terkourafi M., 2005. Beyond the Micro-Level in Politeness Research. *Journal of Politeness Research: Language Behaviour Culture*, vol. 1, iss. 2, pp. 237-262. DOI: <https://doi.org/10.1515/jplr.2005.1.2.237>
- Terkourafi M., 2011. From Politeness1 to Politeness2: Tracking Norms of Im/Politeness Across Time and Space. *Journal of Politeness Research*, vol. 7, iss. 2, pp. 159-185. DOI: <https://doi.org/10.1515/JPLR.2011.009>
- Terkourafi M., 2015. Conventionalization: A New Agenda for Im/Politeness Research. *Journal of Pragmatics*, vol. 86, pp. 11-18. DOI: <https://doi.org/10.1016/j.pragma.2015.06.004>
- Terkourafi M., Kádár D.Z., 2017. Convention and Ritual (Im)Politeness. Haugh C.M., Kádár D.Z., eds. *The Palgrave Handbook of Linguistic Politeness*. London, Palgrave Macmillan, pp. 171-195.
- Van Dorst I., Gillings M., Culpeper J., 2024. Sociopragmatic Variation in Britain: A Corpus-Based Study of Politeness. *Journal of Pragmatics*, vol. 227, pp. 37-56. DOI: <https://doi.org/10.1016/j.pragma.2024.04.009>
- Van Olmen D., Andersson M., Culpeper J., 2023. Inherent Linguistic Impoliteness: The Case of Insultive YOU+NP in Dutch, English, and Polish. *Journal of Pragmatics*, vol. 215, pp. 22-40. DOI: <https://doi.org/10.1016/j.pragma.2023.06.013>
- Watts R., 2003. *Politeness*. Cambridge, Cambridge University Press. 328 p.
- Watts R., 2005. Linguistic Politeness Research: Quo vadis? Watts R.J., Ide S., Ehlich K., eds. *Politeness in Language: Studies in Its History, Theory and Practice*. Berlin, Mouton de Gruyter, pp. 11-47. DOI: <https://doi.org/10.1515/9783110199819.XI>
- Xia D., Ai H., Pae H.K., 2022. "Please Let Me Know" Lexical Bundles in Business Emails by Business English Learners and Working Professionals. *International Journal of Learner Corpus Research*, vol. 8, no. 1, pp. 1-30. DOI: <https://doi.org/10.1075/ijlcr.20019>
- Xie F., 2024. Politeness Variation: Politeness in Britain, Australia, and Hong Kong. *Corpus Pragmatics*, vol. 8, pp. 313-333. DOI: <https://doi.org/10.1007/s41701-024-00174-1>
- Zlov V., Zlatev J., 2024. A Cognitive-Semiotic Approach to Impoliteness: Effects of Conventionality and Semiotic System on Judgments of Impoliteness by Russian and Swedish Speakers. *Journal of Politeness Research*, vol. 20, iss. 2, pp. 249-296. DOI: <https://doi.org/10.1515/pr-2022-0017>

Information About the Authors

Elena S. Gritsenko, Doctor of Sciences (Philology), Professor, Department of English no. 3, MGIMO University, Moscow, Russia, elena.s.gritsenko@gmail.com, <https://orcid.org/0000-0001-8669-360X>

Olivier Mozard T. Kamou, Postgraduate Student, Department of Foreign Languages, HSE University, Nizhny Novgorod, Russia, kamouolivier@gmail.com, ototsopkamu@hse.ru, <https://orcid.org/0000-0001-5291-3037>

Информация об авторах

Елена Сергеевна Гриценко, доктор филологических наук, профессор кафедры английского языка № 3, Московский государственный институт международных отношений (университет) МИД России, г. Москва, Россия, elena.s.gritsenko@gmail.com, <https://orcid.org/0000-0001-8669-360X>

Оливье Мозард Т. Каму, аспирант, департамент иностранных языков, Национальный исследовательский университет «Высшая школа экономики», г. Нижний Новгород, Россия, kamouolivier@gmail.com, ototsopkamu@hse.ru, <https://orcid.org/0000-0001-5291-3037>