The Influence of Discourse Topic on the Use of Anglicisms by German University Students

Seoha Lee

***Abstract*—This study investigates how discourse topic influences the frequency and type of Anglicism use in spoken German among university students. Six participants forming three pairs engaged in paired conversation tasks across five discourse topics: Academic, Cultural, Personal, Professional, and Technical. Anglicisms were manually coded into six lexical categories and normalized per 1,000 words per speaker per topic. A one-way ANOVA was conducted to test topic-level differences in code-mixing frequency. Results indicate that Cultural topics elicited the highest rate of Anglicism use while Academic topics elicited the lowest, partially supporting the hypothesis that casual topics trigger more code-mixing. Nouns constituted the most frequent category across all topics, supporting the prediction of lexical-level borrowing dominance. Findings are discussed in relation to indexicality and topic-driven register variation.**

***Index Terms*—Anglicism, code-mixing, discourse topic, EFL, indexicality, spoken German**

I. Introduction

The presence of English-origin words in everyday spoken German has become a pervasive feature of contemporary language use, particularly among young educated speakers. Anglicisms function not only as lexical gap-fillers but as socially meaningful resources that index values such as cosmopolitanism, technical expertise, or in-group belonging [3]. This indexical dimension suggests that Anglicism use is not uniformly distributed across contexts but varies with register, topic, and interactional setting.

Despite this, relatively little empirical work has examined how discourse topic shapes Anglicism frequency in spontaneous spoken German. Most existing studies are corpus-based and do not track topic-by-topic variation in real-time speech. Bhattacharya et al. [1] demonstrated that discourse topic is a significant predictor of code-switching frequency in bilingual speech, but it remains unclear whether a comparable effect holds for monolingual German speakers using English as a foreign language (EFL). The present study addresses this gap.

Two hypotheses are tested. H1 predicts that casual topics (Personal, Cultural) will trigger more code-mixing than formal topics (Academic, Professional, Technical), based on the assumption that informal registers invite greater stylistic freedom. H2 predicts that nouns will constitute the most frequent category of Anglicism use, consistent with the established cross-linguistic tendency for nouns to be the most permeable word class in contact situations [2].

This study makes two novel contributions. First, it collects naturalistic spoken data through structured conversation prompts rather than relying on existing corpora, enabling direct comparison across topics within the same interaction. Second, it applies a fine-grained six-category annotation scheme that captures the full range of Anglicism type, from bare nouns to verb-hybrids and discourse markers, providing greater linguistic granularity than binary classification approaches.

II. Literature Review

A. Anglicisms in German

Anglicisms in German span a continuum from fully integrated loanwords (e.g., “der Computer”) to productive hybrid constructions formed through the attachment of German morphological affixes to English stems (e.g., “gedownloadet,” “streamen”). Onysko [2] distinguishes between established borrowings that have entered the standard lexicon and more recent code-mixing phenomena characteristic of informal spoken registers among speakers with high English proficiency. Nouns are consistently the most frequently borrowed word class, a pattern attributed to the ease of nominal integration and the density of English terminology in technical and cultural domains [2].

The social meaning of Anglicism use has been theorized through the concept of indexicality [3]. Speakers draw on English-origin forms not merely out of lexical necessity but to project particular social identities or align with specific cultural values. This implies that Anglicism frequency is tied to the interactional context, including the topic being discussed.

B. Discourse Topic and Code-Mixing

Topic has long been recognized as a domain-level variable shaping language choice in contact situations. Certain discourse domains, especially technology, social media, and popular culture, are lexically dense with English-origin terminology, creating structural conditions for higher rates of borrowing. Bhattacharya et al. [1] provides direct empirical support for this claim, demonstrating through machine learning classification of the SEAME corpus that topic categories (Academic, Cultural, Personal, Professional, others) are significant predictors of code-switching frequency in bilingual speech.

This study builds on Bhattacharya et al. [1] by applying the same topic taxonomy to a new population, native German speakers with EFL, and by replacing corpus-based classification with structured elicitation prompts. A Technical category is added to capture technology and digital media discourse, a domain expected to yield high Anglicism density given the prevalence of English in the tech industry.

III. Methods

A. Participants

Six university students participated in the study (age range: 18–29). All participants reported German as their first language and were raised in monolingual German-speaking households. English proficiency was self-reported at CEFR B1 or above. Participants represented diverse university majors including business, linguistics, and natural science. They are referred to using anonymized codes SP1–SP6, organized into three pairs: Pair 1 (SP1, SP2), Pair 2 (SP3, SP4), and Pair 3 (SP5, SP6).

B. Data Collection Procedure

Data were collected through a paired conversation task designed to elicit naturalistic spoken interaction. Each pair discussed five structured prompts spanning distinct discourse topics over approximately 35 minutes while being audio-recorded. The five topic categories were adapted from Bhattacharya et al. [1]: Academic, Cultural, Personal, and Professional were retained; a Technical category was added to capture technology and digital media discourse; and the Others category was excluded as it is not operationalizable through targeted prompts.

All prompts were formulated as expository-generation questions using high-token verbs (e.g., "Analysiert," "Erörtert") to elicit extended responses. Prompts were written entirely in German to minimize lexical priming effects and included a mild personal perspective component to encourage naturalistic speech. To control order effects, three prompt orderings were constructed following a Latin Square design and assigned one per pair.

Following the conversation task, participants completed a post-task questionnaire via Google Forms collecting sociodemographic information (age, major, English exposure, social network composition) and self-reported attitudes toward Anglicism use, including perceived frequency, motivations (e.g., clarity, habit, social norms), and identity associations (e.g., cosmopolitanism).

C. Transcription

Audio recordings were transcribed using AssemblyAI's automatic speech recognition service with speaker diarization enabled, using the universal-2 speech model with German as the target language. Each transcript was subsequently reviewed and manually verified by the researcher to correct recognition errors, with particular attention to Anglicisms, hybrid forms, and overlapping speech. Speaker labels were assigned and verified manually.

D. Codebook and Annotation

Anglicisms were identified and annotated according to a six-category codebook developed prior to data collection: (1) Noun [N]: unadapted or pluralized English nouns within a German sentence (e.g., "Deadline," "Major"); (2) Noun-Hybrid [N-HYB]: English nouns with German morphological suffixes (e.g., "Influencern"); (3) Verb-Hybrid [V-HYB]: English verb stems with German inflectional morphology (e.g., "geghostet," "googlen"); (4) Whole Clause [C]: complete English phrases or clauses inserted into German speech (e.g., "I don't know"); (5) Discourse Marker [DM]: English items used as conversational management devices (e.g., "same," "check"); and (6) Adjective/Adverb [ADJ]: English descriptors not covered by the above (e.g., "comfortable," "awkward"). Proper nouns and brand names were excluded.

E. Normalization and Analysis

Raw Anglicism counts were normalized per 1,000 words for each speaker within each discourse topic to account for differences in speech rate and topic duration. Word counts were derived automatically from the verified transcripts. A one-way ANOVA was conducted with discourse topic (5 levels) as the independent variable and normalized Anglicism frequency as the dependent variable. Where a significant topic effect was found, post-hoc pairwise comparisons were performed to identify which topics differed significantly.

IV. Results

A total of 226 Anglicism tokens were identified across 20,302 words of transcribed speech, yielding an overall mean rate of 11.13 tokens per 1,000 words. Token counts varied considerably across speakers (individual overall rates ranged from 5.01 to 19.87 per 1,000 words), reflecting both topic-level and individual-level sources of variation.

A one-way ANOVA revealed a statistically significant effect of discourse topic on normalized Anglicism frequency, $F(4,25) = 5.51$, $p = .003$, $\eta^2 = .47$, indicating that discourse topic accounts for approximately 47% of the variance in Anglicism use.

Table I presents descriptive statistics for each topic. Cultural discourse elicited the highest mean Anglicism rate (mean, M = 22.02; standard deviation, SD = 11.10 per 1,000 words), followed by Professional (M = 12.77, SD = 3.71), Technical (M = 8.81, SD = 6.36), Personal (M =8.59, SD = 8.17), and Academic (M = 3.80, SD = 3.18).

TABLE I: ANGLICISM FREQUENCY BY DISCOURSE TOPIC (PER 1,000 WORDS)

Topic	Mean	Total count	Total words count
Academic	3.8	17	4474
Cultural	21.89	86	3928
Personal	9.21	33	3583
Professional	12.55	50	3985
Technical	9.23	40	4332

Post-hoc Tukey HSD comparisons identified three significant pairwise differences. Cultural differed significantly from Academic (mean diff. = 18.21, p = .001), from Personal (mean diff. =13.43, p = .024), and from Technical (mean diff. = 13.21, p = .027). No other pairwise contrasts reached significance, including Cultural vs. Professional (mean diff. = 9.25, p = .196).

Table III presents individual Anglicism rates per speaker and topic. Marked within-topic variation is evident across all five domains. In Cultural discourse, the highest-frequency topic overall, individual rates ranged from 4.91 (SP4) to 33.78 (SP1) per 1,000 words. SP1 produced the highest overall rate (19.87 per 1,000 words), consistent with her C2 proficiency and extended residence abroad; SP4 produced the lowest overall rate (5.01). This within-sample range suggests that while topic exerts a significant group-level effect, individual speaker factors like proficiency, exposure, and attitudinal orientations substantially modulate output.

Table III: ANGLICISM RATES BY SPEAKER AND TOPIC (PER 1,000 WORDS)

Speaker	Academic	Cultural	Personal
SP1	8.57	33.78	21.47
SP2	2.49	25.42	2.34
SP3	2.87	33.09	3.27
SP4	0.00	4.91	4.13
SP5	6.69	19.58	16.30
SP6	2.21	15.31	4.02
Speaker	**Professional**	**Technical**	**Overall**
SP1	14.03	19.88	19.87
SP2	9.60	7.60	8.67
SP3	14.41	8.58	11.59
SP4	14.34	0.00	5.01
SP5	7.06	8.00	11.63
SP6	17.16	8.79	9.66

Regarding category distribution (H2), Figure 1 presents the overall distribution across all tokens. Nouns [N] constituted the largest share (n = 120, 53.1%), followed by adjectives/adverbs [ADJ] (n = 56, 24.8%), verb-hybrids [V-HYB] (n = 29, 12.8%), discourse markers [DM] (n = 11, 4.9%), whole clauses [C] (n = 5, 2.2%), and noun-hybrids [N-HYB] (n = 3, 1.3%).

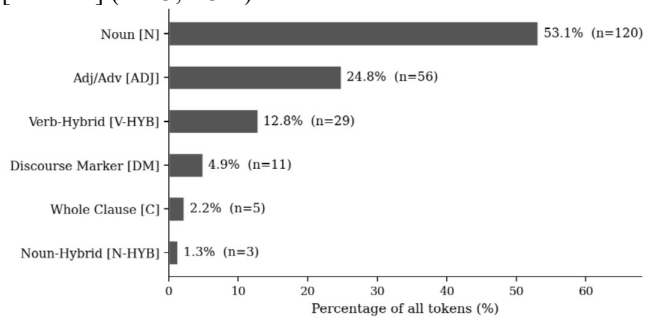

Fig. 1. Overall Anglicism Category Distribution (N = 226).

Figure 2 shows the breakdown by topic. Noun dominance was strongest in Professional (62.0%) and Technical (65.0%) discourse, where English technical terminology lacks established German equivalents. In contrast, ADJ and DM tokens were proportionally most prominent in Cultural and Personal topics (Cultural: ADJ n = 17, DM n = 3; Personal: ADJ n = 16, DM n = 4), consistent with their role as evaluative stance markers in casual interaction. Verb-hybrids were also concentrated in Cultural discourse (n = 14 of 29 total), reflecting the productive morphological integration of social-media-related verbs such as googelst, durchscrollen, and editiert.

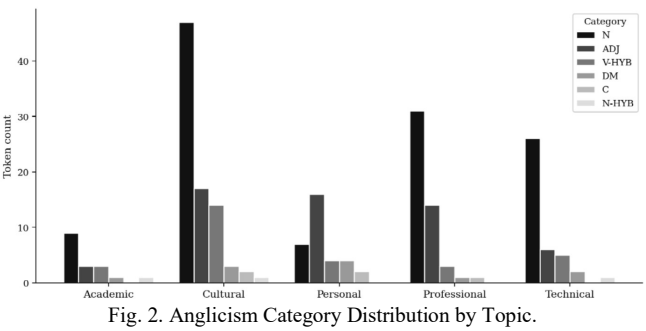

Fig. 2. Anglicism Category Distribution by Topic.

V. Discussion

A. Topic Effects and Discourse Domain

The findings provide partial support for H1. Cultural topics elicited significantly more Anglicism use than Academic, Personal, and Technical topics, consistent with the prediction that casual, culturally oriented discourse invites greater code-mixing. However, the expected casual–formal divide was not fully realized: Professional discourse yielded the second-highest mean rate (12.77 per 1,000 words) and did not differ significantly from Cultural ($p = .196$), while Personal discourse did not differ significantly from any formal topic.

The elevated rate in Professional discourse is better explained by domain-specific lexical density [2] than by register informality. Conversations centered on IT careers, freelancing, remote work, and self-employment are lexically saturated with English terminology (e.g., IT, App, Freelancer, remote, Games Engineering) that lacks widely established German equivalents, motivating nonce borrowing irrespective of formality level. This points to a slightly different explanation: Anglicism frequency seems to depend less on register itself and more on how strongly a topic is structured around English-origin vocabulary.

Cultural topics additionally brought out the widest range of Anglicism types. ADJ tokens (e.g., nice, crazy, jealous) and DM tokens (e.g., safe, true, sorry) were proportionally most prominent in Cultural and Personal contexts, functioning as evaluative stance markers rather than technical borrowings. This aligns with Silverstein's [3] account of indexical order: English-origin expressions in casual discourse do not merely fill lexical gaps but enact social meanings—youthfulness, cosmopolitanism, informal in-group solidarity—thereby reinforcing topic-driven register differentiation.

B. Lexical Category Distribution

H2 is confirmed. Noun dominance (53.1%) is consistent with the established cross-linguistic pattern of greater nominal permeability in contact situations [2]. Nouns were distributed across all topics but were especially dense in Technical and Professional discourse (e.g., Screenshot, RAM, Large Language Model, Code Terminal, IT Support), reflecting the role of English as the global language of technology.

Post-task survey responses provide important qualification to the indexicality account. When asked to rate motivations for mixing (1–5 Likert scale), participants most strongly endorsed naturalness (M = 4.33), technical topic fit (M = 4.17), speed of lexical access (M = 4.00), and general comfort with mixing (M = 4.00). By contrast, identity-oriented motivations scored markedly lower: appearing more international (M = 2.33), appearing cooler or more modern (M = 2.50), and the statement that mixing reflects personal identity (M = 2.83). This pattern suggests that Anglicism use in this sample is primarily driven by habituation and automaticity rather than deliberate identity performance, consistent with the high proficiency levels reported. Lexical gap was rated moderately (M = 3.33), supporting the domain-density explanation for Technical and Professional topics. Notably, SP5 and SP6—whose pair produced the most tokens in Cultural discourse—both attributed their mixing to heavy consumption of

English-language social media, directly corroborating the topic-specific finding. Conversely, SP4, who rated comfort and identity-related motivations uniformly low (all items = 1) while rating speed and technical fit highly (both = 4), produced the fewest tokens among the four female speakers, suggesting that attitudinal disposition toward code-mixing may modulate output independently of proficiency and exposure.

C. Individual Variation and Limitations

Individual variation was notable: speaker-level rates ranged from 0 to 33.78 per 1,000 words. Within each dyad, one speaker consistently produced more Anglicisms than the other, suggesting that interactional dynamics and individual proficiency may modulate topic effects. Future work should examine inter-speaker accommodation and proficiency as covariates.

Several limitations should be acknowledged. The sample is small (n = 6), limiting statistical power and generalizability. All participants were university students with self-reported B1-or-above English proficiency, restricting the range of proficiency and social background represented. Conversation prompts, while naturalistic in delivery, introduce an elicitation artifact not present in fully spontaneous speech. Replication with a larger, more heterogeneous sample is needed to confirm these patterns.

VI. Conclusion

This study investigated the influence of discourse topic on Anglicism use in spoken German among university students. A one-way ANOVA of normalized Anglicism frequencies across five topics revealed a significant topic effect ($F(4, 25) = 5.51$, $p = .003$, $\eta^2 = .47$). Cultural topics elicited significantly more code-mixing than Academic, Personal, and Technical topics. Unexpectedly, Professional discourse also produced high Anglicism rates, attributable to the English-lexical density of technology and career domains rather than to register informality. Noun borrowings dominated across all topics, confirming the cross-linguistic primacy of nominal code-mixing.

These findings extend the topic-based account of code-switching [1] to EFL-context spoken German and demonstrate that domain-specific lexical density (not casual register alone) is a primary determinant of Anglicism frequency. The indexical dimension of Anglicism use, particularly in Cultural discourse, further suggests that speakers deploy English-origin forms as social meaning-making resources [3]. Future research should expand the participant pool, incorporate proficiency measures, and examine accommodation processes within dyadic speech to build a more complete account of topic-driven code-mixing in German.

Appendix

A. Codebook: Anglicism Annotation Categories

All Anglicism tokens were coded according to the following six-category scheme. Proper nouns and brand names were excluded.

[N] Noun	Single English nouns. Examples: Deadline, Level, Gym.
[N-HYB] Noun-Hybrid	English noun roots inflected with German morphological suffixes (-m, -en). Examples: Influencern, Stresslevel.
[V-HYB] Verb-Hybrid	English verb roots combined with German prefixes (ge-) or suffixes (-en, -t). Examples: geghostet, durchscrollen.
[C] Whole Clause	Full English phrases, idioms, or independent clauses. Examples: I don't care.
[DM] Discourse Marker	English items used to structure speech or manage interactional flow, with little referential meaning. Examples: safe, true, sorry, check.
[ADJ] Other/Adjective	English adjectives or adverbs not covered by primary noun or verb categories. Examples: nice, crazy, jealous, random.

B. Elicitation Prompts

The following German prompts were used to elicit conversation across the five discourse topics. Each pair received all five prompts in a counterbalanced order.

- Academic: Erörtert, inwiefern Leistungsdruck an deutschen Universitäten ein Problem darstellt und wie ihr persönlich damit umgeht.
- Cultural: Analysiert den Einfluss von Social Media auf aktuelle gesellschaftliche Trends und euren eigenen Alltag.
- Personal: Beschreibt, welche Werte und Eigenschaften ihr in einer romantischen Beziehung für besonders wichtig haltet.
- Professional: Erörtert eure Vorstellungen und Erwartungen an euren künftigen Berufsweg und die heutige Arbeitswelt.
- Technical: Analysiert, welche Rolle künstliche Intelligenz in eurem Alltag und Studium spielt und wie ihr ihre Zukunft einschätzt.

C. Post-Task Survey: Motivation Items

Participants rated the following statements on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree):

1. Ich verwende englische Wörter, weil sie mir schneller einfallen.
2. Für manche technischen Themen klingen englische Wörter passender.
3. Ich verwende englische Wörter, weil Menschen in meinem Umfeld das auch tun.
4. Ich verwende englische Wörter, um mich präziser auszudrücken.
5. Ich verwende englische Wörter, weil es kein gutes deutsches Äquivalent gibt.
6. Die Verwendung englischer Wörter fühlt sich für mich natürlich an.

References

[1] Debasmita Bhattacharya, Anxin Yi, Siying Ding, and Julia Hirschberg. 2025. Code-switching in Context: Investigating the Role of Discourse Topic in Bilingual Speech Production. In Proceedings of the 6th Workshop on Computational Approaches to Discourse, Context and Document-Level Inferences (CODI 2025), pages 64–80, Suzhou, China. Association for Computational Linguistics.

[2] Onysko, Alexander. Anglicisms in German: Borrowing, lexical productivity, and written codeswitching. Vol. 23. Walter de Gruyter, 2007.

[3] Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. Language & communication, 23(3-4), 193-229.