

Reheat Nachos for Dinner? Evaluating AI Support for Cross-Cultural Communication of Neologisms

Rachel Rudinger Dayeon Ki* Yu Hou*
Hal Daumé III Marine Carpuat Fumeng Yang

University of Maryland
{dayeonki, houyu}@umd.edu

Abstract

Neologisms and emerging slang are central to daily conversation, yet challenging for non-native speakers (NNS) to interpret and use appropriately in cross-cultural communication with native speakers (NS). NNS increasingly make use of Artificial Intelligence (AI) tools to learn these words. We study the utility of such tools in mediating an informal communication scenario through a human-subjects study ($N=234$): NNS participants learn English neologisms with AI support, write messages using the learned word to an NS friend, and judge contextual appropriateness of the neologism in two provided writing samples. Using both NS evaluator-rated communicative competence of NNS-produced writing and NNS' contextual appropriateness judgments, we compare three AI-based support conditions: AI Definition, AI Rewrite into simpler English, AI Explanation of meaning and usage, and Non-AI Dictionary for comparison. We show that AI Explanation yields the largest gains over no support in NS-rated competence, while contextual appropriateness judgments show indifference across support. NNS participants' self-reported perceptions tend to overestimate NS ratings, revealing a mismatch between perceived and actual competence. We further observe a significant gap between NNS- and NS-produced writing, highlighting the limitations of current AI tools and informing design for future tools.¹

1 Introduction

Picture this: You are an international student who learned English as a second language. One day, a native English-speaking friend turns to you: “*When you’re running in the rain and the main character energy starts to hit.*” You know every word—but what does it actually *mean*? How do you respond?

*: Equal contribution; *Reheat nachos* refers to producing a lesser version of an earlier song or album; illustrates one of a failure case of our participant’s incorrect usage.

¹https://github.com/dayeonki/crosscultural_communication

This scenario illustrates a broader challenge in *informal* cross-cultural communication, between non-native English speakers (NNS) 🧑🏫 and native English speakers (NS) 🧑🏻.² Contemporary English is rich with neologisms—newly coined expressions (e.g., main character energy³) or existing forms that have acquired new meanings (e.g., Ohio⁴) (Newmark, 1988), particularly in informal, slang, and internet-mediated discourse (Mattiello et al., 2005). These encode practices and cultural contexts, often to fill lexical gaps within a particular speech community (Santhi, 2010; Khan, 2013). As such, these expressions offer insights into everyday experiences within the communities that use them (McDonald, 2005). This sociocultural richness is precisely what makes these expressions challenging for NNS to understand and use appropriately—yet it is also what makes them appealing to learn.⁵

Traditional learning resources, such as printed dictionaries and textbooks, often fail to capture these rapidly evolving, context-dependent meanings, leading NNS to seek alternative support. They increasingly turn to AI tools for support (Tamkin et al., 2024), which have become go-to resources for language learners (Singh et al., 2024; Saha et al., 2025), who use them to request definitions, rewrite in simpler words, or ask for explanations.

However, existing research has not kept pace with this real-world adoption. Prior studies evaluating AI tools’ capability to understand or process neologisms remain in constrained evaluation formats such as multiple-choice questions (Deng et al., 2024). These settings are far removed from how users actually engage with AI tools when encoun-

²The visual distinction between the two icons is intended solely to aid readability and carries no implication of gender.

³This refers to dramatic self-confidence or self-importance.

⁴This is often used to describe something that is weird, awkward, or cringeworthy.

⁵As an illustrative example, one study found that 98% of NNS surveyed reported positive attitudes toward including English neologisms in language lessons (Rets, 2016).

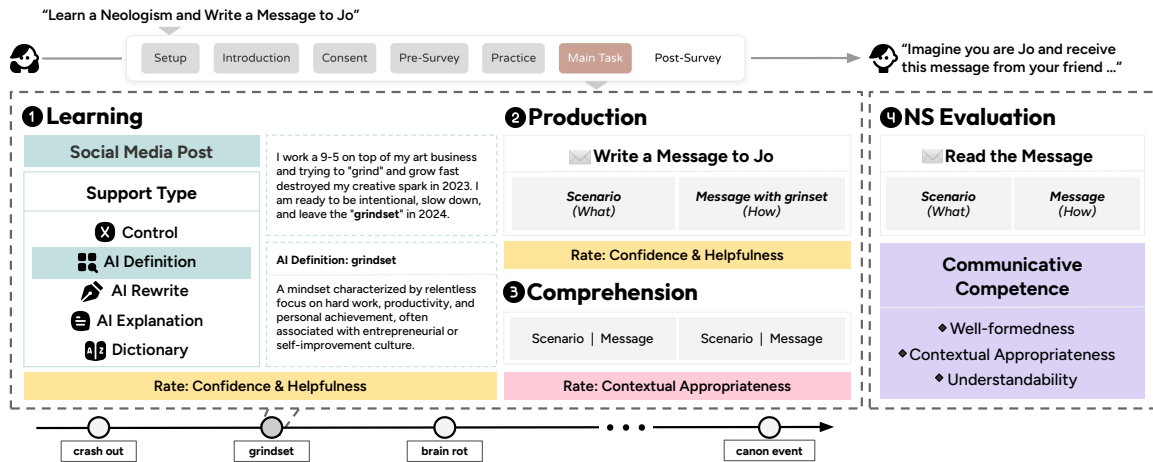




Figure 1: **Overview of our study design.** We set up the communication scenario by having NNS participants learn neologisms and write messages to a hypothetical NS friend, *Jo*. During the main task, for each of the eight neologisms, they complete a three-stage procedure: **1 Learning:** Learn the neologism within a social media post with one of five randomly assigned support types (one control and four treatment); **2 Production:** Write a scenario and a message to *Jo* using the learned neologism; **3 Comprehension:** Rate the contextual appropriateness of the neologism in two provided writing samples. Each NNS-produced writing sample is subsequently rated by two NS evaluators for communicative competence (**4 NS Evaluation**). We color-code the measures throughout the paper using ■ (NS-Rated Competence), ■ (NNS Comprehension Competence), and ■ (NNS Self-Reported Perceptions).

tering neologisms (Liao and Xiao, 2023). Consequently, it remains unclear how useful such AI tools truly are in supporting *real* users navigating the cross-cultural communication of neologisms.

To address this gap, we conduct a between-subjects human study with 234 NNS participants  whose native languages are Spanish, German, or Chinese, along with 144 NS evaluators  residing in the United States. We simulate a real-world scenario in which an NNS encounters a neologism in a social media post, learns the term with different types of support conditions, and uses it to communicate with an NS friend (Figure 1). Motivated by how users typically employ AI tools when encountering new words (Xiao and Zhi, 2023; Klimova et al., 2024), we design four support conditions: (1) AI Definition: providing a dictionary-style definition; (2) AI Rewrite: rewriting the social media post into simpler English; (3) AI Explanation: requesting an explanation of meaning and usage; and (4) Non-AI Dictionary: consulting an official online reference. We select eight neologisms from an online dictionary list. For each neologism, NNS participants first learn it with randomly provided support, then complete tasks assessing their learning goals (Pickering and Garrod, 2013): **(1) Production**, which they describe a brief scenario and write a message to an NS friend using the learned word, and **(2) Comprehension**, which they judge

the contextual appropriateness of the word in two writing samples. Each NNS-produced writing from **Production** is subsequently rated by two NS evaluators along three dimensions of communicative competence: well-formedness, contextual appropriateness, and understandability (Light, 1989).

Our findings show that AI Explanation consistently helps NNS participants achieve higher NS-rated competence across all dimensions compared to no support. Comprehension task performance does not differ significantly across support conditions (§4.1). We also examine the potential of using NNS’ self-reported perceptions as proxies for NS-rated competence, given the practical challenges of directly involving NS in the evaluation. However, what *feels* communicatively effective from the NNS perspective generally overestimates how NS *actually* interprets it (§4.2). We further observe a significant gap between NNS-produced writing and those written by NS evaluators (§4.3). Comments from both parties indicate that NNS’ lack of a mechanism to reliably assess imperfect AI outputs and limited contextualized support are potential causes of this gap (§5.1). Taken together, our findings highlight limitations of current AI tools and inform design for future tools, including designing better support, training on data reflecting real user needs and usage, and communicating uncertainty to help NNS calibrate their reliance on AI (§5.2; §5.3).

2 Background & Research Questions

2.1 Neologisms

Neologisms are newly coined lexical units or existing forms that have acquired new meanings (Newmark, 1988). We focus on internet-derived neologisms (i.e., contemporary slang terms that emerge and evolve in informal discourse) due to both the challenges they pose (Charteris-Black, 1998; Al-Jarf, 2010) and their appeal to NNS (Rets, 2016). Prior work in the context of AI has largely focused on detecting and collecting neologisms (Nguyen et al., 2018; Tomaszewska et al., 2025) or evaluating AI tools' capability to understand and process them (Zheng et al., 2024; Zou and Yu, 2025). However, despite the widespread use of AI by users for informal language learning (Terzimehić et al., 2025), most existing benchmarks operationalize neologism understanding through constrained evaluation formats, such as multiple-choice questions (Deng et al., 2024) or machine translation (Awadh Nasser Munassar and Shafiull, 2020; Zheng et al., 2024). These setups fail to adequately capture how users engage with AI tools *in the wild* (i.e., limited ecological validity (Ethayarajh and Jurafsky, 2020; Liao and Xiao, 2023)). As a result, it remains unclear how effectively AI supports users in learning neologisms for informal cross-cultural communication. Our work aims to address this gap.

2.2 Cross-Cultural Communication

Cross-cultural communication concerns the way people from different cultures interact (Tannen, 1983; Hurn and Tomalin, 2013). We define culture broadly to include *any* communicative context in which interlocutors do **not** share a common **linguistic** or **cultural** background in a given context (e.g., a university lecturer teaching new undergraduate students, or interactions between a native and a non-native speaker) (Thomas, 1983). In such settings, the lack of shared common ground between interlocutors makes communication particularly challenging (Meyer, 2016; Korkut et al., 2018; Carpuat et al., 2025). In response, lay users increasingly turn to AI tools to overcome these communicative gaps (Fu et al., 2024; Sarwari et al., 2024). Prior work has proposed a range of interventions to bridge these gaps (Bourges-Waldegg and Scrivener, 1998; Heimgärtner, 2017), including AI-generated explanations (Ki et al., 2025; Saha et al., 2025; Zhao et al., 2026) or question-answering chatbots (Zhang et al., 2025). How-

ever, human-centered evaluations of such interventions—particularly in *informal* cross-cultural communication settings—remain underexplored. Our work builds on this line of work with a specific focus on neologism communication between non-native and native English speakers. We focus on English, as current AI tools provide stronger support for English than other languages (Bender et al., 2021) and English speakers are more readily accessible for participant recruitment.

2.3 Communicative Competence

As defined by Light (1989), communicative competence is the ability to be *functionally adequate* in daily communication, encompassing sufficient *knowledge, judgment, and skill* to communicate effectively. In this work, we focus on three widely recognized dimensions of communicative competence in the field of language development: (1) grammatical or discourse competence, (2) sociolinguistic competence (Hymes et al., 1972), and (3) strategic competence, referring to compensatory strategies for making the most of one's knowledge (Savignon, 1976; Canale and Swain, 1980). Since communication effectiveness is context-dependent (Hymes et al., 1972), evaluations should be based on outcomes in realistic situations (Canale, 2014). Adopting this perspective, we simulate a cross-cultural communication scenario between NNS and NS and assess NNS' communicative competence through tasks reflecting real-world interactions, with outcomes subsequently rated by NS evaluators. Although directly involving NS is ideal, practical constraints (e.g., limited opportunities) make this challenging (Verhelst et al., 2024). Consequently, NNS often must rely on their own judgment to decide if they are confident in using the neologism in conversations. To approximate this, we also collect NNS self-reported perceptions as proxy measures of competence.

2.4 Research Questions

Informed by the literature, we formalize three research questions (RQs):

RQ1. How effective are different support in helping NNS learn and communicate neologisms?

RQ2. To what extent can NNS' self-reported perceptions serve as reliable proxy of communicative competence?

RQ3. How much can AI support close the gap between NNS and NS communicative competence, and what limitations remain?

Condition	Provided Support	Interaction Patterns	Density	Source	Def.?	Context.?
⊗ Control	No learning support	-	-	-	-	-
🗉 AI Definition	LLM-generated dictionary-style definition of the neologism	AI as traditional dictionaries (Lew et al., 2024)	Low (M=30.5; SD=7.38)	AI (Prompt A.1.2)	✓	✗
🗉 AI Rewrite	LLM-rewritten social media post with the neologism in simpler English	AI-assisted paraphrasing activities (Kim et al., 2025)	Low (M=29.5; SD=10.7)	AI (Prompt A.1.3)	✗	✗
🗉 AI Explanation	LLM-generated explanation of the neologism with contextualized usage examples	AI-assisted explainer (Kohnke et al., 2023)	Mid (M=95.4; SD=5.22)	AI (Prompt A.1.4)	✗	✓
🗉 Non-AI Dictionary	Merriam-Webster page for the neologism	Comprehensive official online reference	High (M=412.3; SD=77.4)	Human experts	✓	✓

Table 1: **Condition details.** **Provided Support** describes the type of learning support given to each condition. **Interaction Patterns** characterizes the key interaction attributes of each support type. **Density** (information density as low/mid/high) is measured by the mean and standard deviation of word count. **Source** indicates where the condition materials are from. **Def.?** denotes whether the condition includes a formal definition of the neologism. **Context.?** denotes whether the condition includes contextualized usage examples.

3 Methods

This section outlines our study design (§3.1), material collection (§3.2), five types of support conditions (§3.3), participant recruitment (§3.4), and the measures and collected data (§3.5).

3.1 Study Design

We examine how different types of AI support influence NNS’ communicative competence with English neologisms using a between-subjects design. NNS participants are situated in a cross-cultural communication scenario where they encounter a neologism in a social media post and wish to communicate with a hypothetical NS friend, *Jo*.⁶ They first complete a practice session and then rate their familiarity with each neologism on a five-point Likert scale (1:Not at all, 5:Very well) (Asahara, 2019; Zheng, 2024). For each neologism, they then complete a three-stage procedure to learn and perform tasks assessing their learning goals (Figure 1):

① **Learning stage.** NNS participants learn a neologism presented in a social media post under a randomly assigned support condition.

② **Production stage.** Participants write (1) a brief scenario and (2) a message using the learned neologism to their NS friend, *Jo*. To support understanding, NS-produced writing samples are shown during the practice session, and four keywords

per neologism, extracted and selected from author-generated writings (Appendix B.1), are provided to aid brainstorming without priming participants. Participants are instructed to make their messages self-contained with at least 10 words, and warnings are displayed for shorter or copy-pasted content.

③ **Comprehension stage.** Participants then rate the contextual appropriateness of the neologism across two provided writing samples by indicating their agreement with “*The use of the word in this message context is appropriate.*” on a ten-point Likert scale (1:Not at all, 10:Completely).⁷ These samples are selected based on 54 NS evaluators’ prior ratings, one considered *poor* (M=5.9/10) and the other considered *good* (M=9.6/10).

④-⑤ process repeats until they finish all eight neologisms. Two attention checks are included throughout the study. Detailed study flow and descriptions are provided in Appendices B.2 and B.3.

Pre-/Post-Task Survey. Before beginning the main tasks, NNS participants complete a brief pre-task questionnaire about their demographics and English social media usage. They later complete a post-task survey asking: (1) perceived confidence, reliance on, and trust in the support for future use (Hoffman et al., 2019); (2) perceived mental burden and task difficulty; and (3) open-ended feedback on the support for the ② **Production** task.

⁶We use a gender-neutral name to avoid implicit assumptions; [https://en.wikipedia.org/wiki/Jo_\(given_name\)](https://en.wikipedia.org/wiki/Jo_(given_name)).

⁷We use a ten-point scale for fine-grained differences, where an one-point improvement is a minimally important difference (Guyatt et al., 1987).

3.2 Materials

Neologisms. We compile a pool of 30 neologisms from the Merriam-Webster slang and trending words list,⁸ and filter them by popularity using Google Trends, retaining those that emerged in 2020 or later. This yields 24 candidate neologisms.

Social Media Posts. For each neologism, we select one representative social media post from Merriam-Webster to convey the scenario. Social media posts provide a realistic, engaging context where neologisms naturally appear and offers an effective, focused learning setting for NNS participants. We present the full list of posts in Appendix Table 10.

3.3 Support Condition Types

We design four support types, in total five conditions as outlined in Table 1. **Control** reflects the standalone usefulness of a social media post without any learning support for neologisms. The three AI-based support conditions differ in the *type* of information provided, corresponding to how people interact with AI in a single turn, and are generated using GPT-4.1.⁹ **AI Definition** illustrates basic cases where NNS use AI as an alternative to dictionaries. **AI Rewrite** covers scenarios where users understand neologisms in the context of a specific example. **AI Explanation** provides explanations in 3–5 sentences of the neologism’s usage, including typical contexts, tone, audience, and connotations. **Non-AI Dictionary** shows the full content from the Merriam-Webster website, which includes the definition, example usages, word origin, and typical contexts.¹⁰ We use the official dictionary condition, which provides the most comprehensive and accurate information, to compare the utility of *imperfect* AI-based support.

From an initial pool of 24 candidate neologisms, we select a final set of eight by ensuring comparable error rates across all three AI-based conditions, measured through both automatic and human annotation against corresponding Merriam-Webster content (Appendix B.4). The final neologisms are: *brain rot*, *canon event*, *cheugy*, *copium*, *crash out*, *delulu*, *grindset*, and *reheat nachos*.¹¹

⁸<https://www.merriam-webster.com/slang>

⁹gpt-4.1-2025-04-14; prompts in Appendix A.

¹⁰For example, <https://www.merriam-webster.com/slang/main-character-energy>; to isolate the text content from distracting layout elements, we extract the text and render it within a controlled, uniform template.

¹¹Our neologisms span diverse linguistic taxonomy (Pinter et al., 2020; Zheng et al., 2024); more details in Appendix B.5.

3.4 Participants

We recruit all participants from Prolific and the sample size is determined using power analysis (expected effect size Cohen’s $f=0.25$; Appendix B.6).

NNS Participants. We recruit 234 participants from around the world who self-identified Spanish, German, or Chinese as their first and primary language and English as fluent language (104, 95, and 35 participants, respectively). We select languages to represent diverse families: Spanish as Indo-European, German as Germanic, and Chinese as Sino-Tibetan. We recruit participants aged 18–44, likely to engage with our communicative contexts (i.e., social media posts), and randomly assign them to five conditions. Each receives USD 6 (equivalent to USD 12/hour); 234 (96.3%) participants who pass both attention checks earn a USD 2 bonus and are included in the analysis.

Of the 234 NNS participants, the mean age was 28.7 years (SD=6.63). Gender was: man (57.3%), woman (38.5%), non-binary (3.0%), and prefer not to say (0.9%). Monthly English social media use was: never (2), rarely (9), sometimes (24, 10%), often (38, 16%), and almost daily (160, 68.7%). Average completion time was 49 minutes (SD=26).

NS Evaluators. We recruit 144 native English speakers to evaluate NNS-produced writing samples during the **Production** stage, simulating the role of the other party in a realistic cross-cultural communication scenario. Evaluators are based in the United States and self-identify English as their first, primary, and fluent language. Each rates 26 NNS-produced writings for one assigned neologism (randomized across evaluators), and is provided the Merriam-Webster dictionary for reference, with additional tools allowed. Evaluators receive USD 6 (equivalent to USD 12/hour), with an average completion time of 32 minutes (SD=9).

Both the NNS and NS tasks were approved by our institution’s IRB, and all participants provided consent prior to the study. Details on the NS evaluation survey and participant demographics are provided in Appendices B.7 and B.8.

3.5 Measures and Collected Data

NS-Rated Competence (RQ1, RQ3). During the **NS Evaluation** stage, NS evaluators rate each NNS-produced writing sample using five questions covering three dimensions of communicative competence (§2.3): **Well-formedness**, assessing

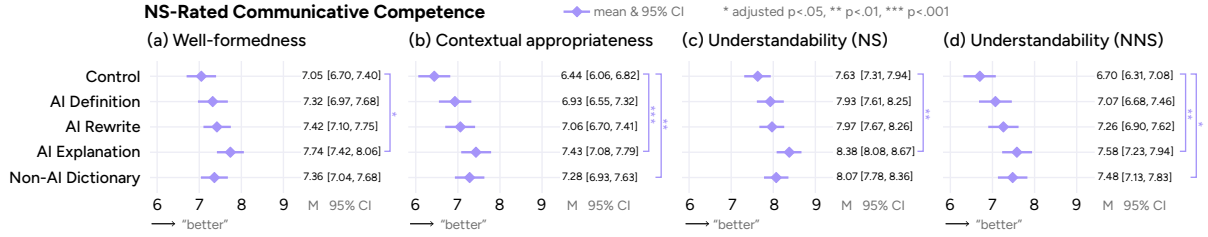


Figure 2: **RQ1+3: NS-rated communicative competence.** All metrics are measured on a ten-point Likert scale (1:Not at all, 10:Completely). x -axis is truncated as no ratings fall below 6 or above 9. (c) Understandability (NS): how much the NS evaluator understands the message; (d) Understandability (NNS): how much the NS evaluator thinks the NNS participant understood the meaning of the neologism through the message. AI Explanation significantly improves communicative competence over Control across all dimensions.

(1) grammatical acceptability and (2) coherence/cohesiveness regardless of meaning, derived from grammatical and discourse competence; (3) **Contextual appropriateness**, evaluating how appropriately the neologism is used in context, capturing sociolinguistic competence; and **Understandability**, measuring (4) ease of understanding the message and (5) whether the NNS correctly understood the neologism’s meaning, reflecting strategic competence. In total, this yields 3,744 observations (1,872 writings from 234 participants \times 8 neologisms, each receiving two ratings per question).

We also collect writing samples from each NS evaluator during the **NS Evaluation** stage, which are subsequently rated by 16 additional NS evaluators on the same dimensions, yielding 288 observations (144 writings \times two ratings each).

▲ NNS Comprehension Competence (RQ1). For each writing sample w shown to NNS participants during the **Comprehension** stage, let \hat{c} be the collected contextual appropriateness rating. We take the difference from the mean of the two NS evaluator ratings, c_1^* and c_2^* , collected prior to the main study:

$$\text{Distance}(w) = \left| \hat{c} - \frac{(c_1^* + c_2^*)}{2} \right|$$

where a lower distance indicates that the NNS participant’s rating is closer to the NS evaluators’. Each NNS participant rates two writing samples per neologism, yielding a total of 3,744 observations.

● NNS Self-Reported Perceptions (RQ2). We collect NNS participants’ self-reported *confidence* in and *helpfulness* of the support during the **Learning** (for correctly understanding the neologism, if any) and **Production** stages (for correctly writing the message, if any), each on a five-point Likert scale (1:Not confident/helpful at all,

5:Very confident/helpful). Helpfulness and confidence ratings are not collected for the Control group, resulting in a total of 1,872 observations.

4 Results

We assess the effectiveness of each support condition (§4.1), measure NNS participants’ self-reported perceptions (§4.2), and compare NS ratings of NNS- to NS-produced writing (§4.3).

4.1 RQ1: How Effective is Each Support?

We employ linear mixed-effects models, where the dependent variables are NS-rated competence and NNS comprehension competence. Fixed effects include support condition, language group, their interaction, English social media usage, and initial familiarity with the neologisms (collected prior to the main task). Random intercepts include NNS participant ID (all models), NS evaluator ID (for NS ratings), and neologism ID (for other outcomes). We report the average marginal effects with Bonferroni-adjusted p -values to address multiple comparison issues.¹² To address the limitations of p -values, we also report 95% confidence intervals (CIs) (Dragicevic, 2016; Cumming and Calin-Jageman, 2024). All means, p -values, and CIs are reported in figures.

◆ NS-Rated Competence. We show results in Figure 2 and summarize key findings below:

(a) Well-formedness.¹³ AI Explanation significantly improves over Control, while no significant differences are observed for other conditions.

(b) Contextual appropriateness. Both AI Explanation and Non-AI Dictionary significantly improves over Control. AI Definition and AI Rewrite

¹²We use R: lme4 (Bates et al., 2015), lmerTest (Kuznetsova et al., 2017), and emmeans (Lenth, 2025).

¹³We average (1) grammatical acceptability and (2) coherence/cohesiveness ratings due to high correlation.

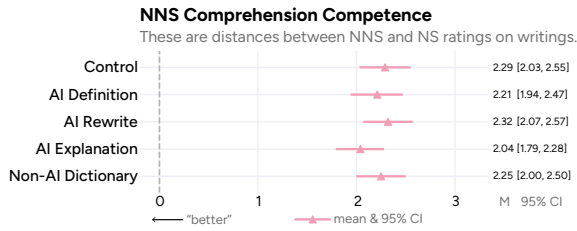


Figure 3: **RQ1: NNS comprehension distance.** Compared to the Control, none of the treatment support conditions significantly reduce the distance.

fall in between these support conditions, with no statistically significant differences.

(c) Understandability (NS). AI Explanation significantly helps NNS participants write messages better understood by NS evaluators compared to Control. No significant differences are observed for other treatment support conditions.

(d) Understandability (NNS). As with Contextual appropriateness, both AI Explanation and Non-AI Dictionary significantly outperform Control, while others show no significant differences. Taken together, compared to no support (Control), AI Explanation consistently helps NNS participants achieve higher communicative competence in all dimensions. In contrast, while Non-AI Dictionary offers the most comprehensive and accurate information, it improves only a subset of dimensions. This is likely due to its high information density (Table 1), which is perceived as higher mental burden and task difficulty (Table 2).

▲ **NNS Comprehension Competence.** As shown in Figure 3, AI Explanation yields the smallest average distance, but the difference is not statistically significant compared to other support conditions. The overall average distance remains relatively high (2.22), suggesting that while some support conditions (AI Explanation and Non-AI Dictionary) yield significantly higher NS-rated competence, they do not necessarily help NNS participants to accurately judge the contextual appropriateness of neologisms in writing samples.

4.2 RQ2: Perceived vs. Actual Competence

We show ● **NNS Self-Reported Perceptions** in Figure 4 and summarize findings below:

(a+b) Confidence (Learning+Production). Non-AI Dictionary, AI Explanation, and AI Rewrite significantly improve confidence over Control, in decreasing order of effect size, while AI Definition shows no significant difference.

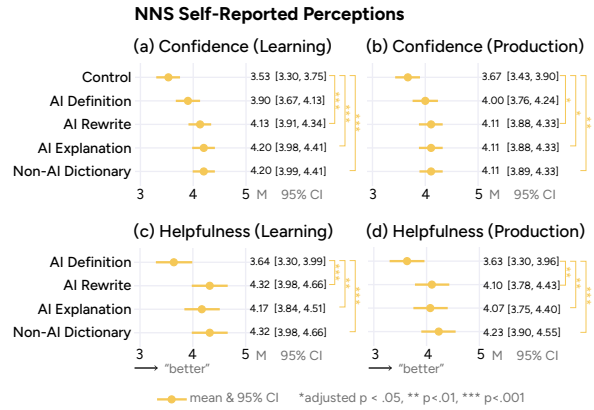


Figure 4: **RQ2: NNS self-reported confidence and helpfulness.** All support conditions except AI Definition significantly improve confidence over Control; all conditions improve helpfulness over AI Definition.

Measures	✕	🗺️	✍️	⚖️	📖
Confidence (↑)	-	3.42	4.02	4.04	4.51
Reliance (↑)	-	2.74	3.60	3.40	4.28
Trust for Future Use (↑)	-	3.56	3.67	3.88	4.44
Mental Burden (↓)	3.67	3.42	3.31	3.44	3.77
Task Difficulty (↓)	3.70	3.66	3.51	3.46	3.65

Table 2: **Post-survey statistics of NNS participants.** **Best** and **second best** scores for each metric are highlighted. All metrics are measured on a 1–5 scale. 📖 Non-AI Dictionary yields the highest confidence, reliance, and trust for future use. Detailed results per language group are provided in Appendix C.3.

(c+d) Helpfulness (Learning+Production). Non-AI Dictionary, AI Rewrite, and AI Explanation all significantly improve over AI Definition, in decreasing order of effect size.

Overall, NNS participants’ self-reports serve as limited proxy for NS-rated communicative competence.¹⁴ While AI Explanation and Non-AI Dictionary align with NS ratings (§4.1), AI Rewrite also boosts confidence and perceived helpfulness without reliably improving actual competence. Combined with open-ended comments (Appendix C.6) and post-task survey results (Table 2), we attribute this over-reliance to its ease of understanding: it presents social media posts in simpler terms, reducing perceived mental burden and task difficulty.¹⁵

¹⁴We further provide AI-rated communicative competence results using LLM-as-judge in Appendix C.1, showing that, in aggregate, they distinguish writings under five conditions while producing higher ratings than NS-rated competence.

¹⁵The effects of the conditions are consistent across all three languages. Results for each measure by language group (Spanish, German, and Chinese) are reported in Appendix C.2.

Scenario	✉ Message to Jo	Support	WF	CA	U
Telling Jo about a botched pizza delivery on a special occasion	The pizza arrived, finally! They took too long so I got it free, which rocks honestly. But I think they pulled a reheated nachos on me, this thing looks like if it has been put through the oven twice! It’s texture is crunchy and dried up. Like it’s still free but... I expected something better, today it’s a big day after all.	❌	1	1.5	1
I’m telling Jo about Tim and his essay	Hey Jo, Tim don’t know how to write an essay so he’s reheating her roommate’s nachos ...	👉	5	6	5
Talking about the entertainment industry	I feel like it’s getting more difficult to create something new - whether that be in music, movies, etc. Everything just sounds and looks like reheated nachos nowadays.	👉	10	10	10
Someone made something that was not as good as the original version.	Taylor Swift was totally reheating Sabrina Carpenter’s nachos on this new album.	NS	10	10	10

Table 3: **NNS- and NS-produced writing samples with NS-rated competence for reheat nachos.** WF: Well-formedness; CA: Contextual appropriateness; U: Understandability (averaged). We color-code the ratings for (low), (mid), and (high). NS denotes NS-produced writing. Examples for other words are in Appendix C.5.

4.3 RQ3: NNS vs. NS Communicative Competence

We compare NS-rated competence between NNS- and NS-produced writing samples. To assess differences across groups, we perform a Kruskal-Wallis test for ordinal ratings and non-parametric data. NS-produced writings receive significantly higher well-formedness ($M=7.94$, 95% CI=[7.70, 8.19]) and contextual appropriateness ($M=7.67$ [7.36, 7.98]) than all NNS support conditions ($p<.001$), except for AI Explanation (n.s.). NS writing also score higher averaged understandability ($M=8.06$ [7.81, 8.31]) than all support conditions: Control and AI Rewrite ($p<.001$), AI Definition ($p=.001$), and Non-AI Dictionary ($p=.011$). Descriptive statistics of NNS- and NS-produced messages are provided in Appendix C.4.

Together, these results show a significant gap between NNS- and NS-produced messages in communicative competence and style, suggesting that no single support alone can fully bridge it.

5 Discussion

5.1 Understanding the NNS-NS Gap

To understand the sources of the gap observed in communicative competence between NNS- and NS-produced writing samples (§4.3), we analyze open-ended comments from NS evaluators on the NNS-produced writings. Many evaluators note that NNS participants appeared to “*have little understanding of what the neologism actually means*” as messages often “*feel awkward*,” contain “*replaced words*” (e.g., *copium* → *cope*, *crash out* → *crashed*), or “*use in secondary meanings*.”

More importantly, evaluators note cases where participants “*used neologisms literally*,” specifically for the word *reheat nachos*. For this word, AI Definition provides only the literal meaning: *reheating prepared nachos to a desirable temperature or texture* (Appendix B.5). As such, AI outputs are often imperfect, yet NNS lack the mechanisms to reliably judge their correctness, leading to inaccurate use of neologisms during communication (first example in Table 3). Feedback from NNS participants confirm this: many noted the support was “*especially helpful when it’s a word that I sort of know, but for an entirely new term, it adds confusion and I cannot judge whether it is correct*.” This further explains the limitations of using NNS self-reports as a reliable proxy for communicative competence (§4.2).¹⁶ Quantitatively, this gap appears in §4.1: NNS participants write effective messages yet struggle to judge appropriate neologism use.

5.2 Implications for Future AI Design

How can we bridge this competence gap? As NS evaluators note, “*unless you find yourself or friends using the term, it can be hard to sometimes know if it’s being used in the proper way*,” reflecting the rich sociocultural references neologisms carry. Yet, direct involvement of NS is often limited by practical constraints: NNS have few opportunities for individualized practice (Verhelst et al., 2024), social discomfort when speaking awkwardly (Nederhof, 1985; Kim, 1994), or prefer AI-based learning tools (Ngo, 2024). Our results show that current AI tools, when used in certain ways, can help improve NNS communicative competence (§4.1).

¹⁶Detailed comments are provided in Appendix C.6.

Richer Contextual Support. Feedback from NNS participants in the AI Explanation group highlights the value of “*various notes on context usage examples, origin, tone, and audience.*” Together with this group’s high NS-rated competence (§4.1), the results suggest that current AI tools can generate informative usage examples that help NNS communicate neologisms effectively.¹⁷ At the same time, requests for “*more examples,*” especially for those “*based on real-world scenarios showing how NS actually uses the word,*” including “*direct quotes,*” “*common scenarios,*” or “*typical mistakes made by NNS*” (i.e., unnatural usage), were common across all conditions. To meet this need, future AI-based support should provide richer, contextualized examples grounded in sociocultural nuance and authentic usage, through: (i) contrastive minimal pairs of natural NS usage and plausible NNS misuses, mined from real-world communities where neologisms appear in situ (e.g., r/Neologisms) or learners seek feedback (e.g., r/English-Learning); (ii) structured pragmatic guidance tied to audience, tone, and connotation (e.g., “Use it when... / Avoid it when...”), making implicit pragmatic cues explicit as AI Explanation already begins to do; and (iii) iterative practice loops in which NNS draft messages and receive targeted feedback, addressing the limitations of single-turn support. The information density of the support and mental load for the user need to be considered together.

Up-to-date Support for Evolving Language.

Beyond support design, the time-sensitive nature of neologisms makes continuously updating AI training data uniquely challenging, often leading to outdated knowledge (§5.1). Future tools should address this on two fronts. First, they should detect emerging neologisms through statistical signals (e.g., atypical likelihood or perplexity patterns) and leverage retrieval-based mechanisms to ground outputs in up-to-date usage examples and definitions. Second, and equally important, they should explicitly communicate uncertainty (Kadavath et al., 2022), both about neologism meanings and about feedback quality, via confidence scores (Lin et al., 2022) or verbal hedging (Prokofieva and Hirschberg, 2014) (e.g., “*I’m not sure, but *reheat nachos* may mean...*”), helping NNS learners calibrate their reliance on AI outputs. This is especially critical given that our preliminary LLM-

as-judge analysis (Appendix C.1) suggests LLM judges tend to overestimate NNS communicative competence relative to NS ratings, highlighting the risk of overly affirmative feedback that may reinforce learner overconfidence.

5.3 Implications for Our Society

While there is clearly room to bridge cross-cultural communication gaps, we caution the community to consider the nature of neologisms, or new slang as studied here, when designing AI tools (Drake, 1980). Informal communication plays an important role in how people build social connections, and the language used in those conversations signifies the social identities of the speakers, further influencing how they identify ingroup or outgroup members to coordinate or build social networks (Yuan et al., 2013; Mendelsohn et al., 2023a,b). Similar to intergenerational gaps in slang (Vizgirdaite, 2009; Sadovskaya, 2023), newly emerging slang brings challenges in communication, but when the technology becomes good enough so a slang term is universally known across cultures and languages, it loses its function of building unique, shared connections between people or within a community. Therefore, carefully evaluating the societal impacts and the degree to which people want AI to help, or not, is necessary when imagining future AI.

6 Conclusion

We explore the utility of AI support for helping NNS use English neologisms in informal interactions with NS. In a between-subjects study, NNS participants complete production and comprehension tasks, with their writings rated for communicative competence by NS evaluators. Across five support conditions, AI Explanation consistently yields the highest NS-rated competence (§4.1). NNS self-reported perceptions are assessed as proxies for competence, but their overestimation limits reliability (§4.2). We further observe competence gaps between NNS- and NS-produced writings (§4.3), largely driven by NNS’ difficulty judging imperfect AI outputs and limited contextualized support (§5.1). Our findings inform design for future AI support and tools (§5.2; §5.3) to help NNS communicate neologisms more naturally with NS, rather than “*like my grandma trying to use slang.*”¹⁸

¹⁷See Appendix C.7 for NS-rated competence ratings for AI-generated, NNS-, and NS-produced writing samples.

¹⁸Quote adapted from one of our NS evaluators.

Limitations

The scope of our human study is limited to the current experimental setup. We specifically focused on *informal* cross-cultural communication of neologisms between non-native and native English speakers, thus our findings may not generalize to other cross-cultural communication contexts. Additionally, our AI-based support types are limited to English-based interventions (e.g., we did not provide AI Explanation in participants' native language), involved only single-turn interactions, and did not encompass the full range of ways people engage with AI tools (e.g., search engine results), which are valuable directions for future research.

Ethics Statement

This study was approved by our Institutional Review Board (IRB). All participants provided informed consent prior to participation and were compensated according to the rates specified in the consent form. Further, NNS participants who passed both attention checks received a bonus.

Acknowledgments

We would first like to thank the members of the CLIP lab at the University of Maryland for their constructive feedback and support. We especially thank Connor Baumler, Nishant Balepur, and Atrey Desai for revising the AI support condition content, and Lorena Calvo-Bartolomé for piloting our annotation as an NNS participant. We also appreciate Yixin Bai, Taehyun Yang, Caleb Holland, Daniel Palamarchuk, Lingjun Zhao, and Navita Goyal for providing early feedback on our project scope and annotation surveys. We would further like to thank Calvin Bao, Kartik Ravisankar, Osvaldo Quinjica, HyoJung Han, and Eric Bennett for their feedback on the draft. Last but not least, Dayeon and Yu would like to extend their gratitude to friends who helped test early stages of the pilot studies: Chenxi Cui, Nuan Wen, Chenghao Yang, Yilin Zhang, Yue Feng, Chen Guo, Yu Miao, Paiheng Xu, Yu Duan, Enwei Wu, and Wei Sun for the Chinese version, and Christopher Sunghun Choi, Nakyung Lee, Daeun Jung, Curie Kim, Joohyung Song, Jeongin Kim, Jisoo Lee, Hano Lee, Hanseul Nam, Dakyoung Heo, Gayoung Lee, Nayeon Kim, Hyojin Lim, and Namhee Kim for the Korean version. We also thank all the Prolific participants who took part in our study. This material is based upon work partially supported by the NSF under

Grant No. 2229885 (NSF Institute for Trustworthy AI in Law and Society, TRAILS), and NSF CAREER Award No. 2339746 (Rudinger). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

- Reima Al-Jarf. 2010. Translation students' difficulties with english neologisms. *Annals of Dunarea de Jos University of Galati*, 24:431–437.
- Masayuki Asahara. 2019. [Word familiarity rate estimation using a Bayesian linear mixed model](#). In *Proceedings of the First Workshop on Aggregating and Analysing Crowdsourced Annotations for NLP*, pages 6–14, Hong Kong. Association for Computational Linguistics.
- Awadh Awadh Nasser Munassar and Khan Ansarul-lah Shafiull. 2020. [Challenges of translating neologisms comparative study: Human and machine translation](#). *Journal of Language and Linguistic Studies*, 16(4):1987–2002.
- Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. [Fitting linear mixed-effects models using lme4](#). *Journal of Statistical Software*, 67(1):1–48.
- Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. [On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?](#) In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, page 610–623, New York, NY, USA. Association for Computing Machinery.
- Paula Bourges-Waldegg and Stephen AR Scrivener. 1998. Meaning, the central issue in cross-cultural hci design. *Interacting with computers*, 9(3):287–309.
- Michael Canale. 2014. From communicative competence to communicative language pedagogy 1. In *Language and communication*, pages 2–27. Routledge.
- Michael Canale and Merrill Swain. 1980. [Theoretical bases of communicative approaches to second language teaching and testing](#). *Applied Linguistics*, 1(1):1–47.
- Marine Carpuat, Omri Asscher, Kalika Bali, Luisa Bentivogli, Frédéric Blain, Lynne Bowker, Monojit Choudhury, Hal Daumé III, Kevin Duh, Ge Gao, Alvin Grissom II, Marzena Karpinska, Elaine C. Khoong, William D. Lewis, André F. T. Martins, Mary Nurminen, Douglas W. Oard, Maja Popovic, Michel Simard, and François Yvon. 2025. [An interdisciplinary approach to human-centered machine translation](#). In *Proceedings of the 2025 Conference*

- on Empirical Methods in Natural Language Processing*, pages 22859–22879, Suzhou, China. Association for Computational Linguistics.
- John Bissell Carroll. 1964. *Language and Thought*. Prentice-Hall.
- Jonathan Charteris-Black. 1998. Compound nouns and the acquisition of english neologisms.
- Geoff Cumming and Robert Calin-Jageman. 2024. *Introduction to the New Statistics: Estimation, Open Science, and Beyond*.
- Hexuan Deng, Wenxiang Jiao, Xuebo Liu, Min Zhang, and Zhaopeng Tu. 2024. **Newterm: Benchmarking real-time new terms for large language models with annual updates**. In *The Thirty-eight Conference on Neural Information Processing Systems Datasets and Benchmarks Track*.
- Pierre Dragicevic. 2016. **Fair statistical communication in hci**. In *Modern Statistical Methods for HCI*, pages 291–330. Springer.
- G.F. Drake. 1980. **The social role of slang**. In HOWARD GILES, W PETER ROBINSON, and PHILIP M SMITH, editors, *Language*, pages 63–70. Pergamon, Amsterdam.
- Kawin Ethayarajh and Dan Jurafsky. 2020. **Utility is in the eye of the user: A critique of NLP leaderboards**. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4846–4853, Online. Association for Computational Linguistics.
- Yue Fu, Sami Foell, Xuhai Xu, and Alexis Hiniker. 2024. **From text to self: Users’ perception of aimc tools on interpersonal communication and self**. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, CHI ’24*, New York, NY, USA. Association for Computing Machinery.
- Gordon H. Guyatt, Marie Townsend, Leslie B. Berman, and Jana L. Keller. 1987. **A comparison of likert and visual analogue scales for measuring change in function**. *Journal of chronic diseases*, 40 12:1129–33.
- Rüdiger Heimgärtner. 2017. Culturally-aware hci systems. In *Advances in culturally-aware intelligent systems and in cross-cultural psychological studies*, pages 11–37. Springer.
- Robert R. Hoffman, Shane T. Mueller, Gary Klein, and Jordan Litman. 2019. **Metrics for explainable AI: Challenges and prospects**. *Preprint*, arXiv:1812.04608.
- Brian J Hurn and Barry Tomalin. 2013. What is cross-cultural communication? In *Cross-Cultural Communication: Theory and Practice*, pages 1–19. Springer.
- Dell Hymes et al. 1972. On communicative competence. *sociolinguistics*, 269293:269–293.
- Saurav Kadavath, Tom Conerly, Amanda Askell, Tom Henighan, Dawn Drain, Ethan Perez, Nicholas Schiefer, Zac Hatfield-Dodds, Nova DasSarma, Eli Tran-Johnson, et al. 2022. Language models (mostly) know what they know. *arXiv preprint arXiv:2207.05221*.
- Mohsin Khan. 2013. Neologisms in urdu a linguistic investigation of urdu media. *Language in India*, 13(6).
- Dayeon Ki, Kevin Duh, and Marine Carpuat. 2025. **Should I share this translation? evaluating quality feedback for user reliance on machine translation**. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 12069–12092, Suzhou, China. Association for Computational Linguistics.
- Min-Sun Kim. 1994. **Cross-cultural comparisons of the perceived importance of conversational constraints**. *Human Communication Research*, 21(1):128–151.
- Yewon Kim, Thanh-Long V. Le, Donghwi Kim, Mina Lee, and Sung-Ju Lee. 2025. **Design opportunities for explainable ai paraphrasing tools: A user study with non-native english speakers**. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference, DIS ’25*, page 1061–1083, New York, NY, USA. Association for Computing Machinery.
- Blanka Klimova, Marek Pikhart, and Layth H. Al-Obaydi. 2024. **Exploring the potential of chatgpt for foreign language education at the university level**. *Frontiers in Psychology*, 15:1269319.
- Lucas Kohnke, Benjamin Luke Moorhouse, and Di Zou. 2023. Chatgpt for language teaching and learning. *Relc Journal*, 54(2):537–550.
- Perihan Korkut, Mustafa Dolmacı, and Burcu Karaca. 2018. **A study on communication breakdowns: Sources of misunderstanding in a cross-cultural setting**. *Eurasian Journal of Educational Research*, 18(78):139–158.
- Alexandra Kuznetsova, Per B. Brockhoff, and Rune H. B. Christensen. 2017. **lmerTest package: Tests in linear mixed effects models**. *Journal of Statistical Software*, 82(13):1–26.
- Russell V. Lenth. 2025. *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.11.2-8.
- Robert Lew, Bartosz Ptasznik, and Sascha Wolfer. 2024. **The effectiveness of ChatGPT as a lexical tool for English, compared with a bilingual dictionary and a monolingual learner’s dictionary**. *Humanities and Social Sciences Communications*, 11:1324.
- Q Vera Liao and Ziang Xiao. 2023. Rethinking model evaluation as narrowing the socio-technical gap. *arXiv preprint arXiv:2306.03100*.

- Janice Light. 1989. [Toward a definition of communicative competence for individuals using augmentative and alternative communication systems](#). *Augmentative and Alternative Communication*, 5(2):137–144.
- Stephanie Lin, Jacob Hilton, and Owain Evans. 2022. [Teaching models to express their uncertainty in words](#). *Transactions on Machine Learning Research*.
- Elisa Mattiello et al. 2005. The pervasiveness of slang in standard and non-standard english. *Mots Palabras Words*, 6(7):41.
- Lucinda McDonald. 2005. The meaning of “e-”: Neologisms as markers of culture and technology. *eTopia*.
- Julia Mendelsohn, Sayan Ghosh, David Jurgens, and Ceren Budak. 2023a. [Bridging nations: Quantifying the role of multilinguals in communication on social media](#). *Proceedings of the International AAAI Conference on Web and Social Media*, 17(1):626–637.
- Julia Mendelsohn, Ronan Le Bras, Yejin Choi, and Maarten Sap. 2023b. [From dogwhistles to bullhorns: Unveiling coded rhetoric with language models](#). In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15162–15180, Toronto, Canada. Association for Computational Linguistics.
- Erin Meyer. 2016. *The culture map: Breaking through the invisible boundaries of global business*. PublicAffairs.
- Anton J. Nederhof. 1985. [Methods of coping with social desirability bias: A review](#). *European Journal of Social Psychology*, 15:263–280.
- Peter Newmark. 1988. *A textbook of translation*, volume 66. Prentice hall New York.
- Kieu Trinh Ngo. 2024. [The use of chatgpt for vocabulary acquisition: A literature review](#). *International Journal of AI in Language Education*, 1(2):1–17.
- Dong Nguyen, Barbara McGillivray, and Taha Yasseri. 2018. [Emo, love and god: making sense of Urban Dictionary, a crowd-sourced online dictionary](#). *Royal Society Open Science*, 5(5):172320.
- Martin J Pickering and Simon Garrod. 2013. An integrated theory of language production and comprehension. *Behavioral and brain sciences*, 36(4):329–347.
- Yuval Pinter, Cassandra L. Jacobs, and Max Bittker. 2020. [NYTWIT: A dataset of novel words in the New York Times](#). In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6509–6515, Barcelona, Spain (Online). International Committee on Computational Linguistics.
- Anna Prokofieva and Julia Hirschberg. 2014. Hedging and speaker commitment. In *5th Intl. Workshop on Emotion, Social Signals, Sentiment & Linked Open Data, Reykjavik, Iceland*.
- Irina Rets. 2016. [Teaching neologisms in english as a foreign language classroom](#). *Procedia - Social and Behavioral Sciences*, 232:813–820.
- Jekaterina Sadovskaya. 2023. [Intergenerational discourse: Its essence and features](#). *European Journal of Language and Literature Studies*, 9(1):25–33. Copyright - © 2023 Jekaterina Sadovskaya, published by Sciendo; Last updated - 2025-01-18.
- Sougata Saha, Saurabh Kumar Pandey, Harshit Gupta, and Monojit Choudhury. 2025. [Reading between the lines: Can LLMs identify cross-cultural communication gaps?](#) In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 8043–8067, Albuquerque, New Mexico. Association for Computational Linguistics.
- D. D. Santhi. 2010. [Translating neologisms](#). <http://dessydys.blogspot.com>.
- Abdul Qadir Sarwari, Muhammad Naeem Javed, Hisham Mohd Adnan, et al. 2024. [Assessment of the impacts of artificial intelligence \(ai\) on intercultural communication among postgraduate students in a multicultural university environment](#). *Scientific Reports*, 14:13849.
- Sandra J Savignon. 1976. Communicative competence: Theory and classroom practice.
- Pushpdeep Singh, Mayur Patidar, and Lovekesh Vig. 2024. [Translating across cultures: LLMs for intralingual cultural adaptation](#). In *Proceedings of the 28th Conference on Computational Natural Language Learning*, pages 400–418, Miami, FL, USA. Association for Computational Linguistics.
- Alex Tamkin, Miles McCain, Kunal Handa, Esin Durmus, Liane Lovitt, Ankur Rathi, Saffron Huang, Alfred Mountfield, Jerry Hong, Stuart Ritchie, Michael Stern, Brian Clarke, Landon Goldberg, Theodore R. Sumers, Jared Mueller, William McEachen, Wes Mitchell, Shan Carter, Jack Clark, Jared Kaplan, and Deep Ganguli. 2024. [Clío: Privacy-preserving insights into real-world ai use](#). *Preprint*, arXiv:2412.13678.
- Deborah Tannen. 1983. Cross-cultural communication.
- Nada Terzimehić, Babette Bühler, and Enkelejda Kasneci. 2025. [Conversational ai as a catalyst for informal learning: An empirical large-scale study on llm use in everyday learning](#). *Preprint*, arXiv:2506.11789.
- Jenny Thomas. 1983. [Cross-cultural pragmatic failure](#). *Applied Linguistics*, 4(2):91–112.
- Aleksandra Tomaszewska, Dariusz Czerski, Bartosz Żuk, and Maciej Ogrodniczuk. 2025. [Neon: A tool for automated detection, linguistic and llm-driven analysis of neologisms in polish](#). In *Computational Science – ICCS 2025*, pages 318–326, Cham. Springer Nature Switzerland.

Eva Verhelst, Ruben Janssens, Thomas Demeester, and Tony Belpaeme. 2024. [Adaptive second language tutoring using generative ai and a social robot](#). In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction, HRI '24*, page 1080–1084, New York, NY, USA. Association for Computing Machinery.

Jurgita Vizgirdaite. 2009. [Filling the child-parent relationship gap via the parent self-education and intergenerational education on internet slang](#). *Social Sciences (1392-0758)*, 64(2):57–66.

Yangyu Xiao and Yuying Zhi. 2023. [An exploratory study of EFL learners' use of ChatGPT for language learning tasks: Experience and perceptions](#). *Languages*, 8(3):212.

Chien Wen Yuan, Leslie D. Setlock, Dan Cosley, and Susan R. Fussell. 2013. [Understanding informal communication in multilingual contexts](#). In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work, CSCW '13*, page 909–922, New York, NY, USA. Association for Computing Machinery.

Yongle Zhang, Phuong-Anh Nguyen-Le, Kriti Singh, and Ge Gao. 2025. [The news says, the bot says: How immigrants and locals differ in chatbot-facilitated news reading](#). In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, CHI '25*, page 1–20. ACM.

Lingjun Zhao, Dayeon Ki, Marine Carpuat, and Hal Daumé III. 2026. [Pragmatics meets culture: Culturally-adapted artwork description generation and evaluation](#). *Preprint*, arXiv:2604.02557.

Jonathan Zheng, Alan Ritter, and Wei Xu. 2024. [NEO-BENCH: Evaluating robustness of large language models with neologisms](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 13885–13906, Bangkok, Thailand. Association for Computational Linguistics.

Wanwan Zheng. 2024. [Estimating word difficulty using stratified word familiarity](#). *Cogent Arts & Humanities*, 11(1).

Yujia Zou and Shan Yu. 2025. [Can large language models understand chinese neologisms?](#) In *2025 International Conference on Asian Language Processing (IALP)*, pages 13–18.

A Prompts

A.1 Support Conditions

We show prompts used for generating AI-based support conditions with GPT-4.1 below. `{word}` entry is populated with each neologism.

Prompt A.1.1: System Prompt

You are a multilingual language expert, who can understand neologisms very well. Neologism is any newly formed word, term, or phrase that has achieved popular or institutional recognition and is becoming accepted into mainstream language. We are particularly interested in internet slangs, which are non-standard or unofficial forms of language used by people on the Internet (such as social media, forums, or messaging apps) to communicate with one another.

Prompt A.1.2: AI Definition

`{system prompt}`

Given the word, provide a formal dictionary definition in English. Return only the definition, no other text.

Word: `{word}`

Definition:

Prompt A.1.3: AI Rewrite

`{system prompt}`

Given the social media post with the word `{word}`, rewrite it into plain English, which is suitable for a general audience. You must rewrite the `{word}` into English. Return only the rewrite, no other text.

Post: `{social media post}`

Rewrite:

Prompt A.1.4: AI Explanation

`{system prompt}`

Given the word `{word}`, explain in English how it is used, including typical situations, tone, intended audience, and connotations. Return only the explanation in 3–5 sentences, no other text.

Word: `{word}`

Explanation:

A.2 AI-Rated Communicative Competence

Prompt A.2: AI-Rated Communicative Competence

Task Instruction: Imagine your friend is telling you about something that happened in their day, and they use a specific neologism to describe it.

Neologisms are newly created terms or phrases that have gained popularity and are starting to be accepted into everyday language.

As a native English speaker familiar with neologisms, we'd love your help in evaluating these scenarios.

You will be evaluating scenarios for one neologism word, so please make sure to understand the neologism using our provided reference dictionary before proceeding.

For each scenario, your friend will share:

1. What happened (a short description)
2. A Message they sent to you using the word.

If you're unsure about the meaning of the neologism, please feel free to check the reference dictionary page provided in the beginning or use any resources you trust!

Neologism: {word}

Reference dictionary page: {dictionary URL}

What happened: {scenario}

Message from your friend: {message}

Please indicate how much you agree with each statement from 1:Not at all to 10: Completely.

Well-formedness: Grammatical correctness, coherence, cohesiveness

Q1. The grammar of the message is acceptable.

Q2. The message is coherent and cohesive in English, regardless of its meaning.

Contextual appropriateness: Appropriateness of the word's usage in context

Q3. The use of the word "{word}" in this message context is appropriate.

Understandability: Ease of understanding

Q4. I understand what my friend is trying to say in this message context.

Q5. I think my friend understood the meaning of the word "{word}" from this message.

Also provide your rating for confidence from 1-5:

Q6. How confident do you feel in the judgments you provided?

5: Very confident

4: Confident

3: Somewhat Confident

2: Not confident

1: Not confident at all

Return the response in JSON format with keys "Q1" through "Q6" and values for your judgment only. Do not add any extra text.

A.3 AI-Generated Writing Sample

We use the same system prompt (A.1.1).

Prompt A.3: AI-Generated Writing Sample

```
{system prompt}
```

Imagine something just happened, and you want to tell your native English-speaking friend Jo about it. Please (1) briefly describe what happened in the What section, and (2) write a message with the word {word} you would send to Jo.

To get started with the scenario, you can think about:

- something that happened to you (or a common acquaintance), when and where?
- something you just saw in the news
- ... anything you'd love to share with Jo

Try to make your message self-contained so that Jo can understand it easily, and write at least 10 words.

Please generate 5 scenarios and return the response in JSON format with keys "what" and "message" for each scenario.

Output Example:

```
[
  {
    {
      "idx": "1",
      "what": "...",
      "message": "..."
    },
    ...
  ]
```

B Study Design Details

We present details on our human study design.

B.1 Keyword Extraction Process

During the **Production** stage, participants are instructed to write a brief scenario and a message to their NS friend. Across three pilot studies, NNS participants consistently reported that the writing was the primary bottleneck, leading to completion times that were longer than initially expected.

To support the writing process, we generate five example writing samples per neologism and extract four keywords to present to participants. Keywords are selected according to the following criteria to avoid priming participants:

- Avoid revealing or at least implicitly conveying the meaning of the neologism (e.g., the keyword "phone" for "brain rot").
- Span diverse everyday contexts, including social, work, and leisure domains.
- Support multiple plausible scenarios to avoid constraining to a single predictable narrative.
- Exclude sensitive or controversial topics (e.g., politics, gender).

Neologism	Keywords
brain rot	video, music, song, game
canon event	meme, small talk, gym, boss
cheugy	wedding, cafe, high school reunion, sport
copium	job interview, concert, airport, stock
crash out	during lunch, tournament, moving, amusement park
delulu	office, singer, class, museum
grindset	gym, marathon, classroom, home
reheat nachos	during a trip, roommate, metro, after dinner

Table 4: **Extracted keywords for each neologism.** The same set of four keywords per neologism is shown identically to all participants.

NNS participants are not required to use the keywords but are encouraged to draw on them if they run out of ideas (Figure 11). The complete set of keywords for each neologism is shown in Table 4.

B.2 Study Interface

We built a custom study interface, with screenshots shown in Figure 11 following the task flow: (1) Study setup, (2) Introduction with task instructions, (3) Consent to participate, (4) Pre-task survey, (5) Practice session, (6) Familiarity check, (7) Main task (1 Learning to 3 Comprehension), and (8) Post-task survey. NNS participants were required to complete all pre- and post-task survey questions and a practice session before the main study to ensure they understood the associated tasks.

Pre-Task Survey. The pre-task survey includes the following nine questions:

- **Native language:** What are your native language(s)?
- **Age:** What is your age?
- **Gender:** What is your gender?
 - Man
 - Woman
 - Non-binary
 - Prefer not to say
- **Nationality:** What is your nationality?
- **English proficiency:** What is your level of proficiency in English? Please select the option that best describes how you use English in your daily life.¹⁹
 - 0: No Proficiency
 - 1: Elementary Proficiency
 - 2: Limited Working Proficiency
 - 3: Professional Working Proficiency
 - 4: Full Professional Proficiency
 - 5: Native or Bilingual Fluency

¹⁹https://en.wikipedia.org/wiki/ILR_scale

- **Years spent in English-speaking country:** How many years have you lived in an English-speaking country?
- **Information sources:** When you encounter an unfamiliar English neologism or slang term, which sources do you typically use to learn its meaning? Select all that apply.
 - AI tools (e.g., AI Overview produced by Google, ChatGPT, Claude)
 - Translation tools (e.g., Google Translate)
 - Online dictionaries (e.g., Merriam-Webster, Urban Dictionary)
 - Reference sites or news outlets (e.g., Wikipedia, BBC, New York Times)
 - Social media or online communities (e.g., TikTok, X/Twitter, Reddit)
 - Other (please specify)
- **English social media usage:** In the past month, how often have you read English posts on social media (e.g., Twitter/X, Reddit, Facebook) in your daily work and life?
 - Never: Never in the past month
 - Rarely: Fewer than once a week
 - Sometimes: Two or three times a week
 - Often: More than three times a week, but not every day
 - Habitually: Almost everyday
- **English writing frequency:** In the past month, how often did you write in English for the following activities? (Never/Rarely/Sometimes/Often/Habitually)²⁰
 - Texting or messaging friends
 - Academic assignments
 - Work-related communication
 - Social media posts
 - Personal writing (e.g., journaling, creative writing)

Post-Task Survey. The post-task survey includes three sections. We dynamically replace CONDITION with the participant’s assigned support condition. The Control group is not asked for the first section and have different set of questions for the third section.

- **[1] What do you think about the CONDITION assistance?** (1:Strongly disagree, 5:Strongly agree)
 - I am confident in the CONDITION. I feel that it works well.
 - The CONDITION is very reliable. I can count on it to be correct all the time.
 - I will use the CONDITION again in the future.

²⁰<https://bilingualism.northwestern.edu/leapq/>

- [2] **What do you think about the task?** (1:Very low, 5:Very high)
 - How mentally demanding was the task?
 - How hard did you have to work to accomplish your level of performance?
- [3] **Is there anything else you want to share with us?** (Control)
 - If you could have had any type of assistance during the writing task, what would have been most useful to you?
- [3] **Is there anything else you want to share with us?** (Treatment)
 - If the CONDITION assistance was helpful in doing the writing task, what aspects were most helpful to you?
 - If the CONDITION assistance wasn't quite what you needed for the writing task, what would have made it better?

B.3 Comprehension Writing Collection

In this section, we describe the collection process for the writing samples used in the **Comprehension** stage. We use writing samples produced by NNS participants during three pilot studies, which were subsequently assessed by two NS evaluators per sample along the same dimensions of NS-Rated competence: well-formedness, contextual appropriateness, and understandability. Based on these ratings, we select two writing samples per neologism that exhibit comparable levels of well-formedness but markedly different levels of contextual appropriateness—yielding one *good* sample with higher contextual appropriateness rating and one *poor* sample with lower rating. On average, the two samples received well-formedness ratings of 8.3/10 and 9.4/10, respectively, and contextual appropriateness ratings of 5.9/10 and 9.6/10, respectively.

B.4 Error Rate Annotation

From an initial pool of 24 candidate neologisms, we select eight by controlling error rates across the three AI-based support conditions (AI Definition, AI Rewrite, AI Explanation). Specifically, we ensure that each condition includes a relatively balanced number of examples with both correct and incorrect AI outputs. Since none of the conditions directly predict the task outcomes—both NS-rated competence and NNS comprehension competence—we employ both automatic and human annotation to obtain proxy measures of support quality, as described below.

Neologism	AI Definition		AI Rewrite		AI Explanation	
	Auto	Human	Auto	Human	Auto	Human
brain rot	0.345	4.67	0.943	4.33	0.889	5.00
canon event	0.410	5.00	0.801	4.33	0.863	4.67
cheugy	0.130	5.00	0.867	5.00	0.893	5.00
copium	0.282	4.67	0.883	4.67	0.914	4.33
crash out	0.512	1.00	0.903	4.67	0.826	1.00
delulu	0.370	5.00	0.991	5.00	0.912	4.00
grindset	0.281	5.00	0.924	5.00	0.918	5.00
reheat nachos	0.295	1.33	0.701	5.00	0.847	4.00

Table 5: **Automatic and human annotation results of error rates.** **Auto:** Automatic annotation (0–1 scale); **Human:** Human annotation (1–5 scale). We color-code the ratings for each quality group: (low), (mid), and (high).

Automatic annotation. We first assess the *relevance* of each AI-based support condition using cosine similarity between its output and a corresponding reference from the Merriam-Webster dictionary.²¹ The reference varies by condition:

- **AI Definition:** similarity to the dictionary’s short definition entry
- **AI Rewrite:** similarity to the original social media post
- **AI Explanation:** similarity to a GPT-4.1-generated summary of the full dictionary content

Human annotation. We then evaluate the *accuracy* of each AI-based support condition through human annotation. Recruitment proceeded in two stages to ensure annotator familiarity with the neologisms: (1) a pre-screening meaning-mapping quiz covering ten neologisms, and (2) a main annotation task restricted to participants achieving at least 80% accuracy in pre-screening. For each condition, we recruit three NS annotators (nine total) to rate how accurately the AI output matches the Merriam-Webster dictionary content on a five-point Likert scale (1:Bad, 5:Perfect). **Figure 7** shows the pre-screening quiz and the main annotation task, which are both implemented in Qualtrics.²²

Finally, we combine relevance and accuracy scores from automatic and human annotations to cluster AI support into three quality groups (high, mid, low) per condition. We apply proportional sampling across clusters to select a balanced set of eight neologisms. Results are shown in **Table 5**.

²¹For example, <https://www.merriam-webster.com/slang/main-character-energy>.

²²<https://www.qualtrics.com/>

B.5 Study Material Details

We illustrate how each type of support is presented to NNS participants in Figure 8. We also show the AI-based support conditions (AI Definition, AI Rewrite, AI Explanation) and original social media posts used for each neologism in Table 10. Non-AI Dictionary group participants are provided with the corresponding Merriam-Webster dictionary page.

B.6 Power Analysis

We determine the sample size for the main human study with NNS participants using a power analysis. Given five conditions (one control and four treatments), we use $F_{TestAnovaPower}$ ²³ with $\alpha=0.05$ and $power=0.8$. Assuming a medium effect size (Cohen’s $f=0.25$), the required total sample size is 196, corresponding to approximately 40 participants per condition.

B.7 NS Evaluator Survey Details

We detail the survey used in the **4 NS Evaluation** stage, which was implemented in Qualtrics. Each NS evaluator rates 26 NNS-produced writing samples associated with one neologism; the neologism assignment is randomized across evaluators. Figure 9 illustrates the task instructions and main workflow. Evaluators first familiarize themselves with the target neologism using a provided Merriam-Webster dictionary entry, along with any additional tools as needed. They then write their own scenario and message using the neologism (later used to compare NNS- and NS-produced writing samples; §4.3). Finally, evaluators rate each NNS-produced writing sample on communicative competence across five questions and report their confidence. After the main evaluation task, we also collect open-ended comments on their general thoughts and reactions while reading the messages.

Pre-Task Survey. The pre-task survey includes the following five questions:

- **Age:** What is your age?
- **Gender:** What is your gender?
 - Man
 - Woman
 - Non-binary
 - Prefer not to say
- **Nationality:** What is your nationality?

²³<https://www.statsmodels.org/dev/generated/statsmodels.stats.power.FTestAnovaPower.html>

- **Native language:** What are your native language(s)?
- **English social media usage:** In the past month, how often have you read English posts on social media (e.g., Twitter/X, Reddit, Facebook) in your daily work and life?
 - Never: Never in the past month
 - Rarely: Fewer than once a week
 - Sometimes: Two or three times a week
 - Often: More than three times a week, but not every day
 - Habitually: Almost everyday

B.8 Participants Details

All human studies, including the main study with NNS participants (§B.2), human annotation for error rate control (§B.4), and the NS evaluation (§B.7) are conducted on the Prolific platform.²⁴

NNS Participants. To ensure high-quality responses, we limit participation to Prolific users with an approval rate above 95% and at least 10 prior submissions. Participants receive a base payment of USD 6 for 30 minutes of participation, with an additional USD 2 performance-based bonus for those who passed both attention checks. Including Prolific platform fees, the total cost of the main task was USD 2,400. Detailed pre-survey statistics by language group are reported in Table 11.

NS Evaluators. We restrict participation to Prolific users with an approval rate above 98% and at least 100 prior submissions, and limit each evaluator to a single evaluation task for one neologism. Evaluators receive a base payment of USD 6 for 30-minute task. Including Prolific platform fees, the total cost was USD 1,411.

Among the 160 NS evaluators (144 for evaluating NNS-produced writing and 16 for NS-produced writing), the mean age was 32.8 years ($SD=6.46$). Gender identities are reported as man (81, 50.6%), woman (75, 46.9%), non-binary (3, 1.9%), and prefer not to say (1, 0.6%). Reported nationalities included American (148, 92.5%), Caucasian (3, 1.9%), African American (3, 1.9%), Mexican (2, 1.3%), Fijian (1, 0.6%), Nigerian (1, 0.6%), Puerto Rican (1, 0.6%), and Asian (1, 0.6%). Native language were reported as English (156, 97.5%), English and Spanish (3, 1.9%), and English and Telugu (1, 0.6%). Monthly use of English social media content was heavily skewed toward higher frequencies: no evaluators reported never or rarely engag-

²⁴<https://www.prolific.com/>

Condition	NS-rated				AI-rated			
	WF	CA	U	C	WF	CA	U	C
Control	7.05	6.44	7.17	4.17	8.93	7.52	8.18	4.10
AI Definition	7.32	6.93	7.50	4.23	9.03	8.16	8.56	4.20
AI Rewrite	7.42	7.06	7.62	4.24	9.04	7.77	8.31	4.12
AI Explanation	7.74	7.43	7.98	4.23	9.18	8.24	8.59	4.19
Non-AI Dictionary	7.36	7.28	7.78	4.23	9.02	7.72	8.32	4.12
Avg.	7.19	7.04	7.34	4.17	9.04	7.88	8.39	4.15

Table 6: **NS-rated vs. AI-rated competence.** **WF:** Well-formedness; **CA:** Contextual appropriateness; **U:** Understandability (averaged); **C:** Confidence. **Best** and **worst** scores for each metric are highlighted.

ing, 10 (6.3%) reported sometimes, 17 (10.6%) often, and 133 (83.1%) reported using English social media almost daily.

C Detailed Results

C.1 AI-Rated Communicative Competence

Having NS evaluators to rate the NNS-produced writing samples is ideal, but involving them is also practically challenging. We investigate using NNS self-reported perceptions as an alternative measure of communicative competence (§4.2), but find they are not reliable proxies. As a middle ground, we test AI tools (i.e., LLM-as-judge) for rating NNS writing. We prompt GPT-5²⁵ to rate each of the three communicative competence dimensions by answering the same five questions used in the **4 NS Evaluation** stage (§3.5). We collect two ratings per question, resulting in 3,744 observations. The prompt is in Appendix A.2. We use the instructions and questions given to NS evaluators (Figure 9).

In Table 6, we compare the mean NS- and AI-rated competence scores for NNS-produced writing samples across support conditions. AI ratings are higher on average for well-formedness, contextual appropriateness, and understandability, whereas NS evaluators report higher confidence. Although the exact rankings differ, AI ratings correctly identify the best (AI Explanation) and worst (Control) conditions, matching NS-rated competence.

C.2 Language Group-specific Results

We perform a Type III Analysis of Variance (ANOVA) using Satterthwaite’s method. The interaction between condition and language is not significant for all the mixed-effects models, indicating that the effect of the condition is consistent across three languages. We present results specific to each language group on NS-rated communica-

²⁵`gpt-5-2025-08-27`

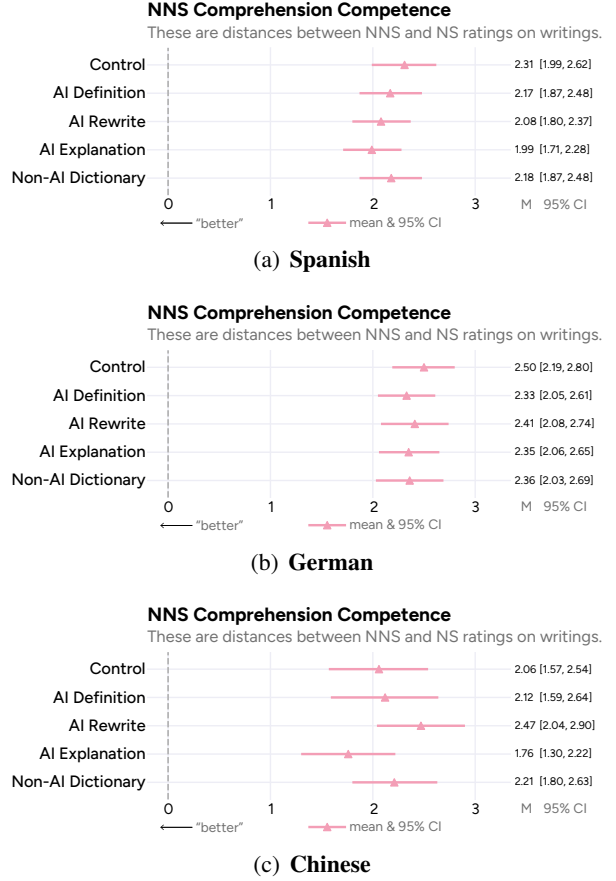


Figure 5: **NNS comprehension distance for each language group.**

tive competence (Figure 10), NNS comprehension competence (Figure 5), and NNS self-reported perceptions (Figure 6).

C.3 Post-task Survey

We present detailed post-task survey statistics for each language group in Table 7.

C.4 Descriptive Statistics of Messages

We report descriptive statistics for NNS- and NS-produced messages in terms of length and lexical diversity, measured via Corrected Type Token Ratio (CTTR) (Carroll, 1964). NNS-produced messages are generally longer than those produced by NS: the average word count is 27.0 for Control, 34.2 for AI Definition, 29.5 for AI Rewrite, 32.7 for AI Explanation, and 27.4 for Non-AI Dictionary, compared to 22.0 for NS messages. NS-produced messages are also lexically less diverse overall (CTTR=2.93) than NNS messages across all conditions (3.19 for Control, 3.50 for AI Definition, 3.34 for AI Rewrite, 3.44 for AI Explanation, and 3.25 for Non-AI Dictionary).

Measures	Language	Control	AI Definition	AI Rewrite	AI Explanation	Non-AI Dictionary
Confidence (↑)	Spanish	-	3.50	4.27	4.29	4.60
	German	-	3.48	3.53	3.76	4.57
	Chinese	-	2.80	4.25	4.00	4.22
Reliance (↑)	Spanish	-	3.00	3.90	3.75	4.40
	German	-	2.60	3.13	2.90	4.14
	Chinese	-	2.40	3.63	3.71	4.22
Trust for Future Use (↑)	Spanish	-	3.75	4.05	3.92	4.65
	German	-	3.32	3.00	3.71	4.29
	Chinese	-	4.00	3.88	4.29	4.22
Mental Burden (↓)	Spanish	3.94	3.30	3.23	3.38	3.75
	German	3.30	3.44	3.33	3.43	3.79
	Chinese	4.17	3.80	3.50	3.71	3.78
Task Difficulty (↓)	Spanish	4.00	3.80	3.59	3.46	3.70
	German	3.20	3.48	3.47	3.33	3.79
	Chinese	4.50	4.00	3.38	3.86	3.33

Table 7: **Post-survey statistics for each language group.** All metrics are measured on a 1–5 scale.

Measures	Control	AI Definition	AI Rewrite	AI Explanation	Non-AI Dictionary
Keyword Inclusion Rate	0.262	0.125	0.239	0.144	0.173
Similarity to Social Media Post	0.313	0.297	0.311	0.299	0.304

Table 8: **Descriptive statistics for NNS-produced messages.** Highest scores for each metric are highlighted.

We also compute two metrics for NNS-produced messages: (1) Keyword inclusion rate, which quantifies the proportion of messages containing any of the keywords we provide to support writing during the **2 Production** stage (Table 4), using exact string matching; and (2) Similarity to the social media post (Table 10), which we compute via embeddings²⁶ and measure with cosine similarity.

As shown in Table 8, the keyword inclusion rate is highest for the Control group (0.262), likely because participants, having no additional support, rely more on the keywords to initiate their writing. Similarly, message similarity to the social media post is highest for the Control group (0.313), followed by AI Rewrite (0.311) and Non-AI Dictionary (0.304).

C.5 Qualitative Examples

In Table 12, we present examples for each neologism, showing NNS- and NS-produced writing that received differing NS-rated competence.

C.6 Open-ended Feedback

We present detailed comments from NNS participants on (1) which aspects of each condition were helpful and (2) which aspects fell short and how

²⁶<https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2>

Condition	N	Mean	95% CI	p-value
NNS (Control)	656	6.44	[6.06, 6.82]	
NNS (AI Definition)	800	6.93	[6.55, 7.32]	n.s.
NNS (AI Rewrite)	736	7.06	[6.70, 7.41]	n.s.
NNS (AI Explanation)	848	7.43	[7.08, 7.79]	$p < .001$
NNS (Non-AI Dictionary)	704	7.28	[6.93, 7.63]	$p = .028$
NNS (Avg.)	3,744	7.04	[6.99, 7.29]	n.s.
AI	80	7.12	[6.37, 7.87]	n.s.
NS	288	7.67	[7.36, 7.98]	$p < .001$

Table 9: **Contextual appropriateness dimension of NS-rated competence for AI-generated vs. NNS- vs. NS-produced writing samples.** N: Number of observations; **Mean:** Mean values; **95% CI:** 95% Confidence Intervals; **p-value:** p -values in comparison to the NNS (Control) condition.

they could be improved. Representative feedback for both parts per condition is shown in Table 13.

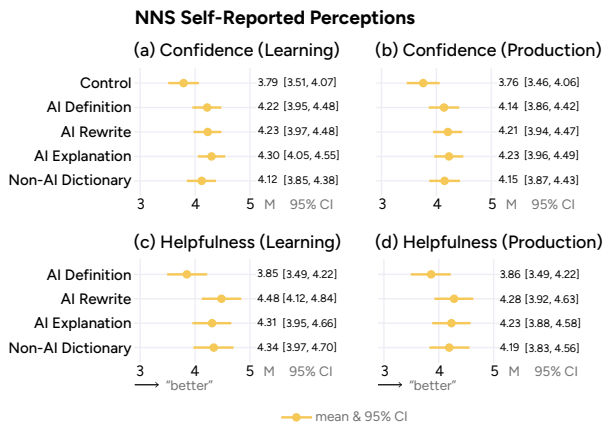
C.7 AI vs. NNS vs. NS Writing Samples

To evaluate the effectiveness of current AI tools in generating contextual usage examples, we extend our comparison of NS-rated competence (§4.3) to include AI-generated writing samples. Using GPT-5, we generate five samples per neologism, each rated twice, yielding 80 observations. The prompt is shown in Appendix A.3. We use the same instructions given to NNS participants during the **2 Production** stage (Figure 11, (g)).

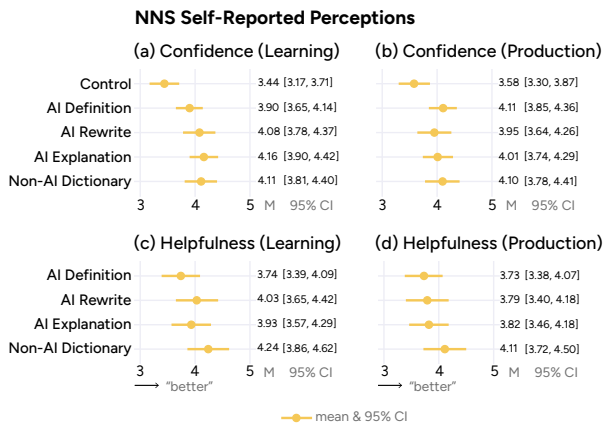
As shown in Table 9, we compare NS-rated contextual appropriateness across all NNS condition groups, AI-generated, and NS-produced writing samples. Differences are assessed using a Kruskal-Wallis test for ordinal, non-parametric data. AI-generated samples have a higher average than Control, AI Definition, and AI Rewrite, but the difference is not statistically significant, likely due to their high variance (i.e., wider 95% CI). This suggests that, despite current AI tools generating plausible examples, these are in fact perceived as similarly or even less communicatively competent than the top-performing NNS support condition (AI Explanation).

D Usage of Large Language Models

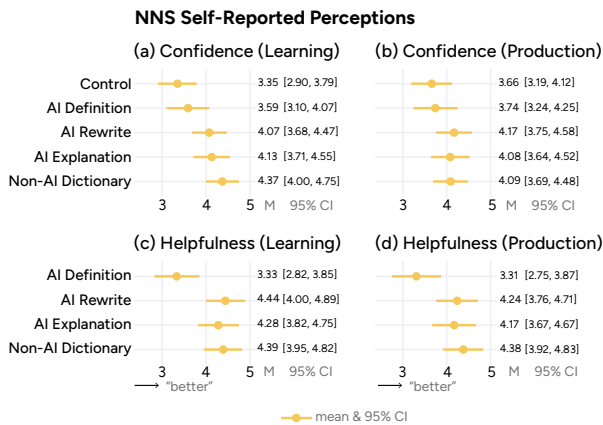
We used LLMs to support and refine the writing of our work. Importantly, we did not rely on them to generate content or sentences from scratch. Instead, we employed them primarily to polish the clarity and expression of how we presented our results. In addition, we used them for stylistic adjustments, such as improving readability and removing layout issues (e.g., widows and orphans).



(a) Spanish



(b) German



(c) Chinese

Figure 6: NNS self-reported confidence and helpfulness perceptions for each language group.

In this short test, you will be given **10 neologisms** and 10 meanings. Your task is to match each neologism with its correct meaning.

Neologism is any newly formed word, term, or phrase that has achieved popular or institutional recognition and is becoming accepted into mainstream language. We are particularly interested in internet slangs, which are non-standard or unofficial forms of language used by people on the Internet (such as social media, forums, or messaging apps) to communicate to one another. These words often reflect the culture, humor, or communication shortcuts.

This qualification task helps us ensure that participants are familiar with evaluating unfamiliar word explanations and can accurately judge language appropriateness in the main study.

If multiple submissions are made, only your first submission will be considered. The estimated time for completion is 5 minutes.

If you qualify, you will be invited to participate in the next phase of the study, where you will evaluate AI-generated meanings of neologisms.

Next

(a) Introduction (Pre-screening)

Please map each meaning to its correct neologism.

	Neologism									
	brain rot	cheugy	copium	delulu	grindset	manosphere	no cap	preppy	rizz	tradwife
For real, seriously	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Unrelenting self-dedication	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Makes fun of being uncool	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Misogynistic male Internet groups; online media seen as promoting toxic masculinity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mindless digital content; the fixation on it and harmful mental effects of it	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
romantic appeal or charm	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
An Internet taunt for a delusional loser	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
A dutiful housewife active on social media	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Delusional; audaciously self-confident	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
A cheery, sporty style popular among girls involving lots of pink; a girl who follows this style	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

(b) Task (Pre-screening)

Figure 7: Screenshots of our NS annotator survey used to compute error rates. We show examples from the AI Definition condition, including task instructions and main study content for the pre-screening quiz (a–b) and the main annotation task (c–d).

In this survey, you will be shown **28 definitions** of words.
That word might be a neologism, a newly created term or phrase that has gained popularity and is starting to be accepted into everyday language.

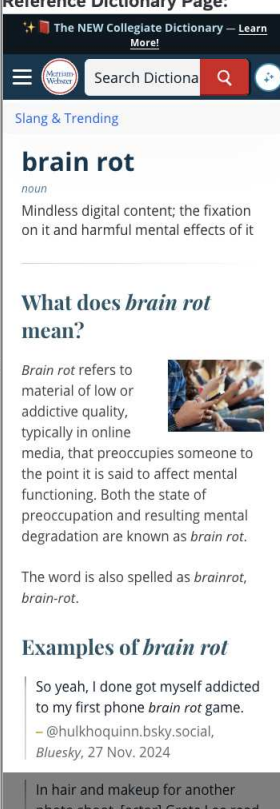
For each definition, you will be asked to make two judgments:
(1) **Accuracy**: Does the definition accurately reflect the meaning of the word?
(2) **Adequacy**: Does the definition include key points in the reference dictionary definition?

To assist you, we provide a **reference dictionary page** that includes a concise definition, part of speech, and example usages of the target word. You may consult this page at any time during the survey to help inform your judgment.

Estimated completion time: ~30 minutes


Next >

(c) Introduction (NS Annotation)

<p>Word: brain rot</p> <p>Definition: A state of mental decline or reduced cognitive function, often attributed to excessive consumption of low-quality or mindless digital content, such as social media, memes, or online videos.</p>	<p>Reference Dictionary Page:</p>  <p>The screenshot shows a mobile app interface for 'The NEW Collegiate Dictionary'. The word 'brain rot' is highlighted under the 'Slang & Trending' section. The definition is: 'Mindless digital content; the fixation on it and harmful mental effects of it'. Below the definition is a section titled 'What does brain rot mean?' which explains that it refers to material of low or addictive quality, typically in online media, that preoccupies someone to the point it is said to affect mental functioning. It also mentions that the word is also spelled as 'brainrot' or 'brain-rot'. There is an example of a tweet: 'So yeah, I done got myself addicted to my first phone brain rot game. -@hulkhoquinn.bsky.social, Bluesky, 27 Nov. 2024'. At the bottom, there is a partial sentence: 'In hair and makeup for another photo shoot, actor Greta Lee read'.</p>	<p>*Q1. Does the definition accurately reflect the meaning of the word?</p> <p><input type="radio"/> 5: Excellent</p> <p><input type="radio"/> 4: Very good</p> <p><input type="radio"/> 3: Good</p> <p><input type="radio"/> 2: Ok</p> <p><input type="radio"/> 1: Bad</p> <p>Please leave any additional comments for Q1, if applicable. (e.g., errors in the definition, newly introduced meanings, or altered meanings)</p> <div style="border: 1px solid gray; height: 20px; width: 100%;"></div> <p>*Q2. Does the definition include key points in the reference dictionary definition? (i.e., "What does __word__ mean?" section or any other definition entry)</p> <p><input type="radio"/> 5: All</p> <p><input type="radio"/> 4: Most</p> <p><input type="radio"/> 3: Much</p> <p><input type="radio"/> 2: Little</p> <p><input type="radio"/> 1: None</p> <p>Please leave any additional comments for Q2, if applicable. (e.g., missing details from the reference dictionary page, newly added details in the definition)</p> <div style="border: 1px solid gray; height: 20px; width: 100%;"></div>
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Next >

(d) **Main Annotation** (left: annotation material, right: questions)

 **AI Definition**


main character energy
 A confident and self-assured attitude characterized by behaving as if one is the central or most important figure in a situation, often with a focus on personal growth, self-expression, and the belief that one's actions and experiences are significant, similar to the protagonist in a story.

(a)  **AI Definition**

 **AI Rewrite**


Feeling entitled to be the most important person in other people's lives while you're barely playing an important role in your own is wild.

(b)  **AI Rewrite**

 **AI Explanation**

main character energy
 Main character energy refers to someone behaving as if they are the protagonist of a story, often exuding confidence, self-assurance, or a sense of uniqueness. It is commonly used on social media to describe people who unapologetically prioritize themselves or stand out in a crowd. The tone is usually playful, admiring, or aspirational, and the phrase is popular among younger internet users. It can be both complimentary and slightly tongue-in-cheek, suggesting someone is living life boldly or dramatically.

(c)  **AI Explanation**

 **Dictionary**

main character energy
noun
 Dramatic self-confidence; obtrusive self-importance

What does *main character energy* mean?
Main character energy is an informal Internet expression for self-assured bearing or behavior. The phrase is used both to compliment self-confidence as well as criticize its excesses.

Examples of *main character energy*
 ... here are 31 birthday nail art ideas, from understated "quiet luxury" vibes to bold glitter Barbie French tips, that will give *main character energy*

(d)  **Non-AI Dictionary**

Figure 8: **Support conditions provided to NNS participants.** We illustrate each condition using the practice-session neologism “main character energy.”

Neologism	Social Media Post	AI Definition	AI Rewrite	AI Explanation
brain rot	So yeah, I done got myself addicted to my first phone brain rot game.	A state of mental decline or reduced cognitive function, often attributed to excessive consumption of low-quality or mindless digital content, such as social media, memes, or online videos.	So yeah, I ended up getting myself addicted to my first mindless phone game.	Brain rot is internet slang used to describe the feeling of mental numbness or decline after consuming excessive, often low-quality or repetitive online content, such as endless social media scrolling, memes, or mindless videos. It is typically used in a humorous or self-deprecating tone among younger audiences, especially on platforms like TikTok, Twitter, or Discord. The term carries a negative connotation, implying that the content is "rotting" one's brain or making them less productive or intelligent. It is often used to jokingly criticize one's own habits or the addictive nature of certain online trends.
canon event	i think my october canon event was someone spreading a rumor that i died and then having someone i knew for 3 days confess their love to me	A significant or pivotal occurrence, often unavoidable, that fundamentally shapes a person's character or life trajectory, typically referenced in online discussions to describe moments perceived as essential to someone's personal development or storyline.	I think the important turning point in my life in October was when someone spread a rumor that I died, and then someone I had only known for three days confessed their love to me.	Canon event refers to a significant or pivotal moment in someone's life that is seen as inevitable or necessary for their personal development, often drawing from storytelling where "canon" means official or essential to the plot. It is typically used humorously or ironically online, especially among younger audiences, to describe awkward, embarrassing, or formative experiences (like a bad breakup or a fashion disaster) that everyone seems destined to go through. The tone is usually lighthearted or self-aware, and the phrase often carries a sense of resigned acceptance, as if these moments are unavoidable parts of growing up or being human.
cheugy	I don't care how cheugy it is, I'm sharing my favorite quotes!	A term used to describe someone or something that is out of date, trying too hard to be trendy, or no longer considered cool, especially in reference to millennial trends or aesthetics that have fallen out of favor.	I don't care how outdated or uncool it is, I'm sharing my favorite quotes!	Cheugy is a slang term used to describe someone or something that is outdated, trying too hard, or no longer considered cool—often in reference to trends, fashion, or behaviors that were once popular but now seem passé. It is typically used by younger people, especially Gen Z, to gently poke fun at millennials or anyone perceived as out of touch with current trends. The tone is usually playful or teasing rather than mean-spirited, and it often carries a sense of lighthearted judgment. Common situations include commenting on decor, clothing, or social media habits that feel behind the times.
copium	Victors writing history is the biggest lie ever. Frequently the losers are the ones who feel the need to write the history because they are huffing extreme amounts of copium.	A fictional substance or concept humorously referred to as something people metaphorically "inhale" or rely on to cope with disappointment, failure, or unfavorable outcomes, especially by rationalizing or denying reality; often used online to mock or satirize someone perceived as making excuses.	Victors writing history is the biggest lie ever. Frequently the losers are the ones who feel the need to write the history because they are desperately trying to comfort themselves and cope with their loss.	Copium is a blend of "cope" and "opium," used online to mockingly describe someone who is using denial or rationalizations to deal with disappointment or failure. It's often employed in a sarcastic or teasing tone, especially in gaming, sports, or political discussions when someone refuses to accept an unfavorable outcome. The term implies that the person is metaphorically "inhaling copium" to numb the pain of reality. Its intended audience is typically internet-savvy users familiar with meme culture. The connotation is generally dismissive or mocking, suggesting that the person is deluding themselves rather than facing facts.
crash out	there's a dude crashing out on twitter because he can't understand why so many colleges found him stuck up and rejected him, it's so funny	To fall asleep suddenly and deeply, often due to exhaustion or fatigue, typically in an unplanned or informal setting.	There's a guy having a meltdown on Twitter because he can't understand why so many colleges thought he was arrogant and rejected him, it's so funny.	Crash out is an informal slang phrase used to describe falling asleep suddenly or passing out from exhaustion, often after a long day or intense activity. It is commonly used among friends or peers, especially in casual conversations, texts, or social media posts. The tone is relaxed and colloquial, with no negative connotations—it's typically used to convey relief or humor about being extremely tired. Typical situations include someone coming home late and immediately falling asleep, or dozing off unexpectedly at a party or while studying.
delulu	They believe in and live by a sci fi book, I'm not surprised that they're this delulu and make up weird scenarios.	A slang term derived from "delusional," used to describe someone who holds unrealistic or overly optimistic beliefs, often about themselves, situations, or relationships, especially in a humorous or self-aware manner.	They believe in and live by a sci fi book, I'm not surprised that they're this delusional and make up weird scenarios.	Delulu is a playful slang term derived from "delusional," often used online to describe someone who has unrealistic or overly optimistic beliefs, especially about relationships, celebrities, or fandoms. It is typically used in a lighthearted, self-aware, or teasing tone, rather than as a harsh insult. The term is popular among younger internet users, particularly in K-pop and stan communities, to poke fun at themselves or others for entertaining far-fetched fantasies. While it can be affectionate or humorous, it sometimes carries a mildly mocking connotation.
grindset	I work a 9-5 on top of my art business and trying to "grind" and grow fast destroyed my creative spark in 2023. I am ready to be intentional, slow down, and leave the "grindset" in 2024.	A mindset characterized by relentless focus on hard work, productivity, and personal achievement, often associated with entrepreneurial or self-improvement culture.	I work a 9-5 on top of my art business and trying to work relentlessly and focus only on hustling and productivity destroyed my creative spark in 2023. I am ready to be intentional, slow down, and leave the mindset of constant hard work and hustle in 2024.	Grindset is a blend of "grind" and "mindset," used to describe a mentality focused on relentless hard work, hustle, and self-improvement, often in the context of entrepreneurship or personal success. It is commonly used on social media, especially among young adults and those interested in business or self-development. The tone can be both aspirational and ironic—sometimes genuinely praising dedication, other times mocking the excessive glorification of nonstop work. The intended audience is typically ambitious individuals or those involved in "hustle culture." Connotations range from admiration for determination to criticism of unhealthy work-life balance.
reheat nachos	"Reheating nachos" is so funny because some nachos are somebody else's reheated nachos. Pharrell reheated Off The Wall & prince's nachos for this sound. Lmao.	To attempt to restore previously prepared nachos to a desirable temperature and texture, typically using a microwave or oven, often resulting in diminished quality compared to their original state.	"Reusing old ideas" is so funny because some ideas are just somebody else's reused ideas. Pharrell reused Michael Jackson's "Off The Wall" and Prince's ideas for this sound. That's hilarious.	Reheat nachos is used online to describe a situation or content that is being recycled, reused, or brought up again after losing its original appeal, much like reheating leftover nachos that are no longer fresh. It often carries a mildly humorous or dismissive tone, suggesting that the repeated topic or meme is stale or unoriginal. The phrase is typically aimed at internet-savvy audiences familiar with meme culture and online trends. It can be used to gently mock someone for bringing up old news or to comment on the lack of novelty in a discussion.

Table 10: **AI-based support conditions and social media posts for each neologism.** **AI Definition:** Formal dictionary definition of the neologism; **AI Rewrite:** Rewrites of the original social media posts into simpler English; **AI Explanation:** Natural language explanations of the neologism's meaning and usage.

Imagine your friend is telling you about something that happened in their day, and they use a specific *neologism* to describe it. Neologisms are newly created terms or phrases that have gained popularity and are starting to be accepted into everyday language.

As a native English speaker familiar with neologisms, we'd love your help in evaluating these scenarios.

You will be evaluating 26 scenarios for **one** neologism "**brain rot**", so please make sure to understand the neologism using our provided reference dictionary before proceeding.

For each of the 26 scenario, your friend will share:

1. **What** happened (a short description)
2. A **Message** they sent to you using the word.

If you're unsure about the meaning of the neologism, please feel free to check the reference dictionary page provided in the beginning or use any resources you trust!

Task Instructions

Please follow these steps when completing the task:

1. Read and understand the meaning of the neologism.
2. Proceed to the task once you feel you're comfortable giving evaluation.
3. Read each scenario and provide judgments on its well-formedness, contextual appropriateness, understandability, and your confidence. If applicable, feel free to leave us any comments or notes!

Total estimated time for the study is 30 minutes. You will receive a compensation of \$6.

Good luck! 🍀

Next

(a) Introduction

🍀 **Your friend is writing a scenario using the neologism: brain rot**

Please make sure you understand the neologism using the reference dictionary page below or other necessary tools (e.g., Google search, AI chatbots).

[Reference Dictionary Page](#)

* You only need to read and learn the page for "brain rot"!

The screenshot shows the Merriam-Webster website interface. At the top, it says "The NEW Collegiate Dictionary, 12th Edition Over 5,000 words added — Learn More!". Below that is a navigation bar with "Dictionary", "Thesaurus", and a search bar. The main content area is titled "Slang & Trending" and features the word "brain rot" as a "noun". The definition is: "Mindless digital content; the fixation on it and harmful mental effects of it". There is an image of people looking at a screen. Below the definition, it asks "What does *brain rot* mean?" and provides a detailed explanation: "Brain rot refers to material of low or addictive quality, typically in online media, that preoccupies someone to the point it is said to affect mental functioning. Both the state of preoccupation and resulting mental degradation are known as *brain rot*." At the bottom of the page, it says "The word is also spelled as *brainrot*, *brain-rot*." On the right side, there is a "WORD OF THE DAY" section for "senescence". At the bottom of the screenshot, there is a question: "Do you feel confident enough to read the messages containing the word?" with a "Yes" button.

Next

(b) Familiarize with Neologism

Figure 9: **Screenshots of our NS evaluation survey, organized according to the task flow.** NS evaluators first consult the dictionary page to familiarize themselves with the neologism (b), then provide their own writing sample (c) and evaluate NNS-produced writing samples (d).

Neologism: brain rot

Reference dictionary page: <https://www.merriam-webster.com/slang/brain-rot>

If you were the one writing this scenario, how would you phrase it? *

* Please use the neologism "brain rot" in the "Message to your friend" part.

🗨️ What happened

✉️ Message to your friend

Next

(c) Collect NS-produced Writing Sample

Example 1 / 26

Neologism: brain rot

Reference dictionary page: <https://www.merriam-webster.com/slang/brain-rot>

🗨️ **What happened:** I saw a tik tok video.

✉️ **Message from your friend:** Hey Jo, I just saw this brain rot tik tok video with the cat chasing its own tail. I can't stop watching it.

Please indicate how much you agree with each statement.

	1: Not at all	2	3	4	5	6	7	8	9	10: Completely
Well-formedness: Grammatical correctness, coherence, cohesiveness										
Q1. The grammar of the message is acceptable. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Q2. The message is coherent and cohesive in English, regardless of its meaning. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Contextual Appropriateness: Appropriateness of the word's usage in context										
Q3. The use of the word "brain rot" in this message context is appropriate. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Understandability: Ease of understanding										
Q4. I understand what my friend is trying to say in this message context. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Q5. I think my friend understood the meaning of the word "brain rot" from this message. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q6. How confident do you feel in the judgments you provided? *

5: Very confident

4: Confident

3: Somewhat Confident

2: Not confident

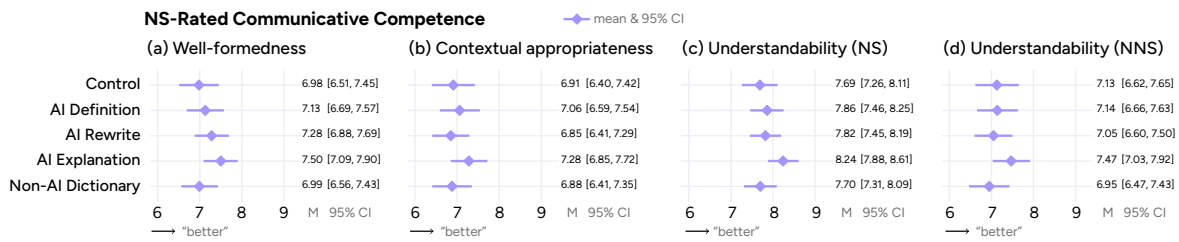
1: Not confident at all

Next

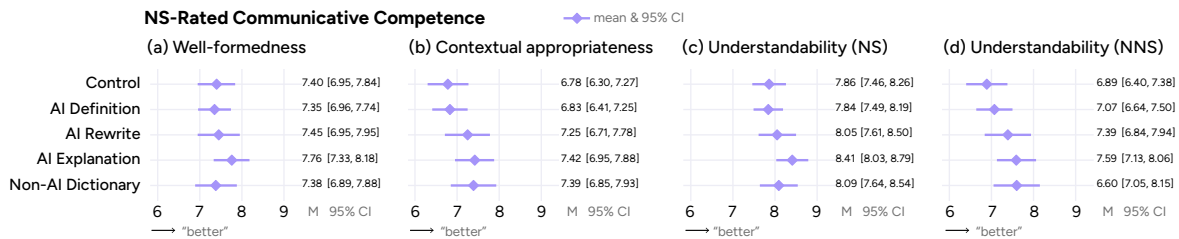
(d) Main Evaluation

Measures	Spanish	German	Chinese
N	104	95	35
Condition Group	Control (16.3%), AI Definition (19.2%), AI Rewrite (22.1%), AI Explanation (23.1%), Non-AI Dictionary (19.2%)	Control (17.1%), AI Definition (14.3%), AI Rewrite (22.9%), AI Explanation (20.0%), Non-AI Dictionary (25.7%)	Control (18.9%), AI Definition (26.3%), AI Rewrite (15.8%), AI Explanation (23.2%), Non-AI Dictionary (15.8%)
Native Language(s)	Spanish (89.4%), Spanish & English (7.69%), Spanish & Catalan (2.88%)	German (79.0%), German & English (4.21%), German & Russian (3.16%), German & Polish (2.11%), German & Bosnian (2.11%), German & Romanian (1.05%), German & Spanish (1.05%), German & Swiss (1.05%), German & Farsi (1.05%), German & Turkish (1.05%), German & Arabic (1.05%), German & Slovak (1.05%), German & Albanian (1.05%), German & Greek (1.05%)	Chinese (77.1%), Chinese & Cantonese (8.57%), Cantonese (5.71%), Chinese & Taiwanese (2.86%), Chinese & English (2.86%), Chinese & English & French (2.86%)
Age	28.9 (S.D.=5.91)	27.8 (S.D.=7.05)	30.5 (S.D.=7.21)
Gender	Man (48.5%), Woman (43.7%), Non-binary (6.8%), Prefer not to say (1.0%)	Man (80.0%), Woman (20.0%)	Man (22.9%), Woman (74.3%), Prefer not to say (2.9%)
Nationality	Mexican (56.3%), Spanish (21.4%), Chilean (15.5%), Mexican-American (1.9%), American (1.9%), Cuban-American (1.0%), Honduran (1.0%), Ecuadorian (1.0%)	German (87.4%), Austrian (6.3%), Swiss (2.1%), German-Albanian (1.1%), German-Greek (1.1%), German-Spanish (1.1%), Romanian (1.1%)	Chinese (54.3%), Canadian (11.4%), Taiwanese (8.6%), Australian (8.6%), British (5.7%), Malaysian (5.7%), American (1.0%), Kongese (1.0%)
English Proficiency	3.56 (S.D.=0.946)	3.35 (S.D.=0.920)	3.49 (S.D.=1.07)
Years	2.80 (S.D.=6.77)	1.08 (S.D.=2.93)	10.7 (S.D.=9.10)
Info. Source	AI tools (18.0%), Translation tools (22.3%), Dictionary (28.2%), Reference/News (6.3%), Social media (25.2%)	AI tools (24.4%), Translation tools (30.5%), Dictionary (25.0%), Reference/News (3.7%), Social media (14.0%), Ask around (2.4%)	AI tools (37.5%), Translation tools (25.0%), Dictionary (18.8%), Reference/News (7.8%), Social media (9.4%), Ask around (1.6%)
Social Media	Habitually (81.6%), Often (11.7%), Sometimes (5.8%), Rarely (1.0%), Never (0.0%)	Habitually (65.3%), Often (18.9%), Sometimes (9.5%), Rarely (4.2%), Never (2.1%)	Habitually (40.0%), Often (22.9%), Sometimes (25.7%), Rarely (11.4%), Never (0.0%)
Writing (Texting)	Habitually (25.2%), Often (19.4%), Sometimes (26.2%), Rarely (19.4%), Never (9.7%)	Habitually (24.2%), Often (21.1%), Sometimes (23.2%), Rarely (16.8%), Never (14.7%)	Habitually (31.4%), Often (28.6%), Sometimes (25.7%), Rarely (14.3%), Never (0.0%)
Writing (Academic)	Habitually (15.5%), Often (17.5%), Sometimes (22.3%), Rarely (20.4%), Never (24.3%)	Habitually (26.3%), Often (14.1%), Sometimes (23.2%), Rarely (22.1%), Never (13.7%)	Habitually (22.9%), Often (37.1%), Sometimes (2.9%), Rarely (14.3%), Never (22.9%)
Writing (Work)	Habitually (33.0%), Often (17.5%), Sometimes (22.3%), Rarely (15.5%), Never (11.7%)	Habitually (20.0%), Often (24.2%), Sometimes (25.3%), Rarely (12.6%), Never (17.9%)	Habitually (51.4%), Often (28.6%), Sometimes (11.4%), Rarely (2.9%), Never (5.7%)
Writing (Social Media)	Habitually (39.8%), Often (20.4%), Sometimes (18.4%), Rarely (15.4%), Never (5.8%)	Habitually (26.3%), Often (28.4%), Sometimes (15.8%), Rarely (11.6%), Never (17.9%)	Habitually (8.6%), Often (34.3%), Sometimes (28.6%), Rarely (17.1%), Never (11.4%)
Writing (Personal)	Habitually (20.4%), Often (24.3%), Sometimes (24.3%), Rarely (15.5%), Never (15.5%)	Habitually (10.5%), Often (8.4%), Sometimes (22.1%), Rarely (28.4%), Never (30.5%)	Habitually (14.3%), Often (11.4%), Sometimes (31.4%), Rarely (25.7%), Never (17.1%)

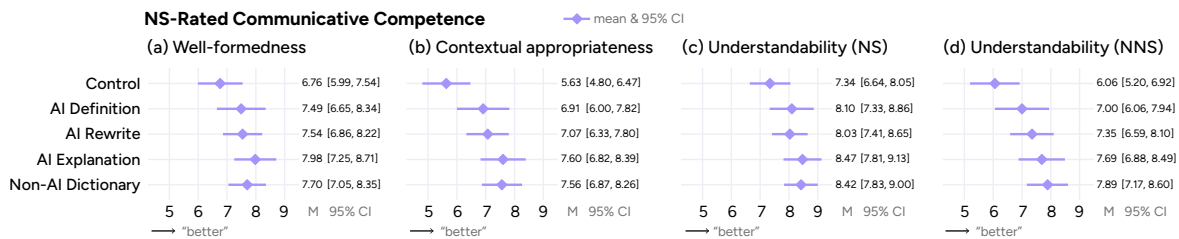
Table 11: **Pre-survey statistics of NNS participants for each language group. Info. Source:** Information sources when participants encounter an unfamiliar English neologism or slang term (includes free-form “Others” responses); **Social Media:** Monthly English social media usage; **Writing:** Monthly English writing frequency for each activity.



(a) Spanish



(b) German



(c) Chinese

Figure 10: NS-rated competence for each language group.

Neologism	Scenario	✉ Message to Jo	Support	WF	CA	U
brain rot	I use tiktok a lot of times in a day	The videos I watched on tiktok are so brainrot that make me feel tired	✘	1.5	1	3
	Recently I noticed my attention span has decreased and I'm blaming social media.	So I finally set a timer on all my socials, to make sure I spend less time engaging with brain rot content and more time in something productive. I promise you I will be more attentive, Jo	☒	4.5	5	5
	I carry my phone with me all the time and kept checking on message updates for nothing.	Hi Jo, I am gonna uninstall Tiktok as it's such a brain rot APP and disrupting my daily work.	☒	9	10	10
	I spent my entire Sunday afternoon scrolling through short, meaningless videos on social media instead of doing the yard work I had planned.	I really need to get off my phone. I spent the entire day consuming so much brain rot there there wasn't enough room in my mind for me to remember to take care of the yard work that I had been planning to get done all week.	NS	10	10	10
canon event	doing wrong exercises in the gym is a canon event as a beginner.	Hey Jo, i looked at your gym plan and there are some exercises i would change. But dont worry its a canon event for beginners.	☒	2	1	1.5
	Describing my health concerns	I'll be getting my fourth and fifth eye surgeries next month. I guess I can look at it as a canon event	☒	5	6	5
	At the gym today, I tripped over a yoga mat right in front of everyone, and instead of being embarrassed, I just laughed it off because it felt like one of those life-defining a "funny momment"	I totally wiped out at the gym today—guess that's my canon event for December!	☒	8.5	9	9
	this one time I was so late for a class, that I ran into my professor while trying to get to the class, my books fell, and that drew a lot of attention to me, I ended up getting scolded by him which was embarrassing as I had managed to get the attention of a lot of my peers.	getting scolded by my professor in front of my peers was my canon event , since then I make sure I am always 10 minutes early to every class.	NS	10	10	10
cheugy	i was so cheugy at the wedding of our friend	hey jo i was so cheugy at the weeding of brian, did you noticed that?	☒	1	3	1
	Talking about a gym member	Hey Jo, i noticed a guy in my gym wearingcrazy gym outfits every day, freaking out about his training plan and protein intake. Thats just way too cheugy imo. Just train and stop overthinking dude!	✘	5.5	4	4
	I went to a high school reunion and noticed some people still dressing and acting exactly like they did back then. . .	Hey Jo, the reunion was sooo wild! Some outfits and hairstyles were so cheugy , it felt like time froze back in high school.	✘	10	10	10
	We were at a cafe and someone played an old motivational playlist while posting a quote graphic with cursive fonts and hashtags from 2016.	I didn't want to say anything in the moment, but the playlist and quote posts felt a little cheugy , like trying too hard to be inspirational instead of just being natural.	NS	10	10	10
copium	i am going home after a concert	hey wanna hang out? i need a way to copium with the post concert depression	☒	3.5	1	2
	i need to tell someone about my job interview	Hello Jo, so my job interview went a little strange i was really unprepared for the amount of copium in the room.	✘	5.5	5	4
	Jo and I are talking about our favorite singer's last concert	Your are on a dose of copium my friend, his last concert was so ugly I cannot accept we wasted our savings on someone like him	☒	10	10	10
	I was walking outside when some children started throwing eggs at me! Everyone around me was laughing as I shouted them down and showed how unruly and poorly behaved they were!	That sounds like pure copium on your part. They were laughing because you lost your cool dealing with admittedly unruly children. You definitely don't come out of this looking like the winner dude.	NS	9.5	10	10
crash out	Was partying with friends and one went totally crash out when his favorite artist entered the stage	Was partyin with Tom and he went crash out as Jason Derulo entered the stage	☒	1	1	1
	Some guy lost the TCG Pokemon tournament and started yelling and screaming to everyone that it was rigged	hey Jo, I just came back from the TCG Pokemon tournament! I had a great time but there was a dude that started legit crashing out when he lost	☒	5	6	7
	A freelancer that works with us delivered poor work (again)	Jo you wont believe how bad the footage was this guy send us today. I nearly crashed out in front of my boss.	☒	10	10	10
	My professor announced at the very end of class that the midterm was moved up by a full week, even though we already have two other exams.	I was fine all day, but when my professor moved the midterm up a week at the last minute, I actually crashed out	NS	10	10	10
delulu	Fritzen got a new job and hopes to drive his new bosses Ferrari	Hi there Jo! My old friend Fritz (the redhead) got a new job and is completely delulu about being allowed to drive his new bosses Ferrari to get an oil change.	☒	1	1	1
	Rabid fans crafting and basing their lifestyles on a fictional concept that's not even real.	Hey Jo, I understand you are a Star Wars fan, as I am one too. But there is no need to go all delulu about it and live your life like a hermit.	☒	5.5	5	5.5
	I discuss the latest news on the erratic behavior of Katy Perry.	Jo, she's just delulu : no one in their right mind would feel heroic for being sent to the moon to record tiktok videos.	☒	10	10	10
	A guy made eye contact with me and I thought it meant something	Call me delulu but I made eye contact with a guy and I think hes my soulmate	NS	10	10	10
grindset	A friend of mine had a long sickness and missed some weeks in school. No I tell Jo how he managed to learn all the stuff he had missed	Hey Jo, you remember Carl? He had a real grindset after his infection, no he gets good grades!	☒	1	1	1
	I am launching a coaching course this fall. I have been working on the final details.	Hey Jo, next week I am finaly launching my course, remember I you said I had that grindset look. Tanks for the encouraging.	☒	5	4	3
	I stayed up all night working on a uni project, skipped breakfast, and still went to the gym in the morning	Bro I pulled an all nighter and still hit the gym pure grindset vibes, but I might pass out soon lol	☒	10	10	10
	I got up super early, went to the gym, and then finished a big chunk of my project before most people were even awake.	I woke up at 5am, hit the gym, then finished my project, grindset is kicking in	NS	10	10	10

Table 12: NNS- and NS-produced writing samples with NS-rated competence. WF: Well-formedness; CA: Contextual appropriateness; U: Understandability. We color-code the ratings for (low), (mid), and (high). NS denotes NS-produced writing.

Condition	Helpful	Needed
Control	N/A	<ul style="list-style-type: none"> • Video explaining the context and examples about the words. • More examples, especially ones showing how the word is used in casual conversations. • AI tools (e.g., Google search or ChatGPT), especially in interactive settings. • Explanation of the context.
AI Definition	<ul style="list-style-type: none"> • Concise explanation, not overwhelming or overly lengthy. • Words are easily understandable, simple, short, and use basic concepts. • Very much hit/miss, but when hit, nailed the concept according to own knowledge. • Definitions got me something in the right direction, but not really about the usage. • Helpful to grasp general meaning quickly and gave good starting point for using new words in context. 	<ul style="list-style-type: none"> • Cross-reference is needed to confirm. • Include examples of usage and context (e.g., “canon event” is mostly used for a computer game/fantasy story plot). • AI should also tell where the origin of the word, or the metaphoric use came from (e.g., “copium”). • It could have been more concise since it sometimes over-explained. • 30% of the times it mistook the concept or explained it wrongly made me trust it little, after all I can’t tell if it’s lying or not, especially when it’s a word I don’t know anything about.
AI Rewrite	<ul style="list-style-type: none"> • Changed the neologisms to words that I know. • Helped in understanding the slang tone and adapting the writing to sound natural for a native English speaker. • I think it did a good job of only changing the word in question, while keeping the rest of the statement. 	<ul style="list-style-type: none"> • Perhaps present 2 alternative versions of rewrites. • AI was too formal, I would prefer to have some real people explanations.
AI Explanation	<ul style="list-style-type: none"> • Helped a lot when mentioned the audiences of the words. • Examples, colloquial language in explaining the meaning. • Context and tone was helpful—it was helpful in explaining not only the meaning but also who uses the term, and the context in which it is used. • Good explanations, starting with the origin or etymology of the word, explaining in a couple of sentences and then talking about the usual use nowadays. 	<ul style="list-style-type: none"> • Visual examples. • Perhaps also give dictionary definition. • Want to search for real world examples because some didn’t sound real. Ideally some direct quotes of more real life cases (like talking to a real person). • Sometimes it gave too many meanings and tried to be safe, like saying that words gets mostly used in a joking manner when in reality they are mostly used to be mean. • It could be even shorter to reduce the information to a minimum and make the explanation even more effective. • Could have included at least a couple of examples very different to each other to have a good spectrum of the meaning. • I didn’t like that in almost every explanation it mentioned that the words were used by “younger audiences.”
Non-AI Dictionary	<ul style="list-style-type: none"> • Examples were really helpful. • Context, evolution of the term (origin) was really helpful. 	<ul style="list-style-type: none"> • Make it more concise. • Actual personal explanation from someone who uses those slang terms. • Maybe specifically mention if the word is towards negative or positive tone? Some words will be confusing if they are used in sarcasm. • More cultural notes, such as examples from social media, explanations of the playful/sarcastic tone slangs often carry (some section like “Common scenarios” or “How people use this online” would make it really easier to apply the word naturally in writing context).

Table 13: **Feedback from NNS participants for each support condition.** **Helpful:** Aspects of the support participants found helpful; **Needed:** Aspects they felt were missing or would have improved the support. We do not ask the **Helpful** question for the Control group.

Learn a Neologism and Write a Message to Jo

You find a neologism on social media and want to learn how to use it when chatting with your native English friend, Jo!

Setup Introduction Consent Pre-Survey Practice Main Task Post-Survey

Customize Your Study

In this study, you'll be learning some neologisms and sending messages to your native English friend, Jo.

Which picture of Jo would you like to use throughout the study? *



Continue


(a) **Study Setup** (All subsequent pages display the character selected by the NNS participant.)

Learn a Neologism and Write a Message to Jo

You find a neologism on social media and want to learn how to use it when chatting with your native English friend, Jo!

Setup Introduction Consent Pre-Survey Practice Main Task Post-Survey

Welcome!


In this study, you will see short social media posts with English neologisms (e.g., main character energy), which are newly created words that are becoming popular. You will learn the words and write messages to your native English friend Jo , with additional questions to check your understanding of the words.

How the Study Works

1. **Intro & Consent:** Read a brief overview and provide consent to participate.
2. **Pre-Survey:** Answer a few questions about your background and language proficiency.
3. **Practice:** Go through one example to get familiar with the interface and task.
4. **Main Task:** Learn 8 neologisms from social media posts. There will be a familiarity check before you start and an attention check halfway through.
5. **Post-Survey:** Share your thoughts and feedback. Note that there is an attention check question included. Your completion code will appear after you finish.

Important Notes

- There are **two attention checks**, one during the main task and one in the post-survey. **Passing both confirms your payment.**
- Please complete the tasks on your own and do not use outside help or tools (e.g., ChatGPT or Google). **Your payment might be affected if you violate this rule.**
- Please try to complete the study in one sitting and avoid reloading the website.
- Once you move to the next page, you won't be able to go back, so take your time and choose your answers carefully.

Total estimated time for the study is 35-45 minutes. The base compensation is \$6, with a \$2 bonus for passing the attention checks, plus an extra bonus for original and good quality responses (no need to be perfect, as long as you try your best!). Have fun learning the neologisms and having a fun conversation with Jo .

Good luck! 🍀

Continue

(b) **Introduction**

Figure 11: Screenshots of our annotation interface, organized according to the task flow. We exclude Consent to Participate page for anonymization.

Learn a Neologism and Write a Message to Jo

You find a neologism on social media and want to learn how to use it when chatting with your native English friend, Jo!

- Setup
- Introduction
- Consent
- Pre-Survey
- Practice
- Main Task
- Post-Survey

Pre-Study Survey

Please answer a few questions about your background and language proficiency.

1. What is your Prolific ID? *

Enter your answer

2. What are your native language(s)? *

Enter your answer

3. What is your age? *

Enter your answer

4. What is your gender? *

Select an option

5. What is your nationality? *

Enter your answer

6. What is your level of proficiency in English? Please select the option that best describes how you use English in your daily life. *

Select an option

7. How many years have you lived in an English-speaking country? *

Enter your answer

8. When you encounter an unfamiliar English neologism or slang term, which sources do you typically use to learn its meaning? Select all that apply. *

Add a source...

9. In the past month, how often have you read English posts on social media (e.g., Twitter/X, Reddit, Facebook) in your daily work and life? *

Select an option

10. In the past month, how often did you write in English for the following activities? *

	Never	Rarely	Sometimes	Often	Habitually
Texting or messaging friends	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Academic assignments	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Work-related communication	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Social media posts	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Personal writing (e.g., journaling, creative writing)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Continue

(c) Pre-task Survey

Learn a Neologism and Write a Message to Jo

You find a neologism on social media and want to learn how to use it when chatting with your native English friend, Jo!

Setup Introduction Consent Pre-Survey **Practice** Main Task Post-Survey

Practice Session

Before the main session, let's go through an example to get used to the format. This won't count toward your final results.

Practice Example 1 of 1

- ★ Word you are learning is: **main character energy**
- ★ Assistance you are getting is: **AI Definition**, where a dictionary definition of the word is provided.



AnonyUser
@user1234567890 · Aug 8

Feeling entitled to main character energy in other people's lives while you're barely showing up as a supporting cast member in your own is wild.

💬 5

↻ 10

❤️ 100

👁️ 1.1K



AI Definition

main character energy

A confident and self-assured attitude characterized by behaving as if one is the central or most important figure in a situation, often with a focus on personal growth, self-expression, and the belief that one's actions and experiences are significant, similar to the protagonist in a story.

Task 1/3: How confident are you in understanding the meaning of main character energy correctly?

Not confident at all

1

2

3

4

5

Very confident

How helpful was the AI definition in understanding the meaning of main character energy correctly?

Not helpful at all

1

2

3

4

5

Very helpful

Next Task

(d) Practice Session

Learn a Neologism and Write a Message to Jo

You find a neologism on social media and want to learn how to use it when chatting with your native English friend, Jo!

Setup

Introduction

Consent

Pre-Survey

Practice

Main Task

Post-Survey

Rate Your Familiarity with Each Word

	Not at all	1	2	3	4	5	Very well
1. Do you know what crash out means?				Not much	Not sure	Kind of	
2. Do you know what grindset means?							
3. Do you know what brain rot means?							
4. Do you know what copium means?							
5. Do you know what reheat nachos means?							
6. Do you know what delulu means?							
7. Do you know what cheugy means?							
8. Do you know what canon event means?							

Start Main Session

(e) Familiarity Check

Learn a Neologism and Write a Message to Jo

You find a neologism on social media and want to learn how to use it when chatting with your native English friend, Jo!

- Setup
- Introduction
- Consent
- Pre-Survey
- Practice
- Main Task**
- Post-Survey

Main Session

Welcome to the main task! Please don't use any outside help or tools. Answer the questions for each post as best you can!

Example 1 of 8

- ★ Word you are learning is: **crash out**
- ★ Assistance you are getting is: **AI Definition**, where a dictionary definition of the word is provided.



AnonyUser
@user1234567890 · Aug 8

there's a dude crashing out on twitter because he can't understand why so many colleges found him stuck up and rejected him, it's so funny

🗨 5

↻ 10

❤ 100

📊 1.1K



AI Definition

crash out

To fall asleep suddenly and deeply, often due to exhaustion or fatigue, typically in an unplanned or informal setting.

Task 1/3: How confident are you in understanding the meaning of **crash out correctly?**

Not confident at all

1

2

3

4

5

Very confident

How helpful was the AI definition in understanding the meaning of **crash out correctly?**

Not helpful at all

1

2

3

4

5

Very helpful

Next Task

(f) Main Task (📌 Learning)

Learn a Neologism and Write a Message to Jo

You find a neologism on social media and want to learn how to use it when chatting with your native English friend, Jo!

Setup Introduction Consent Pre-Survey Practice **Main Task** Post-Survey

Main Session

Welcome to the main task! Please don't use any outside help or tools. Answer the questions for each post as best you can!

Example 1 of 8



- ★ Word you are learning is: **crash out**
- ★ Assistance you are getting is: **AI Definition**, where a dictionary definition of the word is provided.

Task 2/3: Imagine something just happened, and you want to tell your native English-speaking friend Jo  about it. Please (1) briefly describe the scenario in the  **What section**, and (2) write a message to Jo  with the word **crash out**.

To get started with the scenario, you can think about or imagine:

- something that happened to you (or a common acquaintance), when and where?
- something you just saw in the news, or anything!

If you are running out of ideas, maybe try a scenario around: during lunch tournament moving amusement park


- 💡 Tip 1: The  **What section** explains why you're sending the message or adds a little background context.
- 💡 Tip 2: Write a message  you'd actually send to Jo (e.g., start with "Hey Jo" or mention Jo but not required!). Make it self-contained and at least 10 words.
- 💡 Tip 3: Please write your own words. Jo loves to see your unique writing style!

 **What happened**

Briefly describe what happened

 **Message to Jo**

Write a message to Jo with the word you are learning

How confident are you in using the word **crash out** correctly in writing the message to Jo  ?

Not confident at all Very confident

How helpful was the AI definition in using the word **crash out** correctly in writing the message to Jo  ?

Not helpful at all Very helpful

Next Task

(g) Main Task (🎯 Production)

Learn a Neologism and Write a Message to Jo

You find a neologism on social media and want to learn how to use it when chatting with your native English friend, Jo!

Setup Introduction Consent Pre-Survey Practice **Main Task** Post-Survey

Main Session


Welcome to the main task! Please don't use any outside help or tools. Answer the questions for each post as best you can!

Example 1 of 8

- ★ Word you are learning is: **crash out**
- ★ Assistance you are getting is: **AI Definition**, where a dictionary definition of the word is provided.

Task 3/3: Please rate your agreement with the contextual appropriateness (i.e., the word's usage in context) for each scenario.

What happened: A friend collapsed when his cell phone stopped working on his first trip to Japan.

Message to Jo  : Hi Jo, I wanted to tell you that a friend of mine crash out when his cell phone stopped working on his first trip to Japan. He didn't know what to do at that moment.

The use of the word **crash out** in this message context is appropriate.

Not at all Completely

What happened: I am telling Jo about how no matter how much I study, I cannot understand this course.

Message to Jo  : Jo, I am crashing out over this course I chose to take. I can't understand anything.

The use of the word **crash out** in this message context is appropriate.

Not at all Completely

Next Example

(h) Main Task (📌 Comprehension)

Learn a Neologism and Write a Message to Jo

You find a neologism on social media and want to learn how to use it when chatting with your native English friend, Jo!

Setup Introduction Consent Pre-Survey Practice **Main Task** Post-Survey

Attention Check

This is the first attention check. Please select the word you just learned from the list below.

Which neologism did you learn in the writing task you just completed?

- brain rot
- copium
- grindset
- crash out

Continue

(i) Main Task (Attention Check)

Learn a Neologism and Write a Message to Jo

You find a neologism on social media and want to learn how to use it when chatting with your native English friend, Jo!

Setup

Introduction

Consent

Pre-Survey

Practice

Main Task

Post-Survey

Post-Study Survey

Before you finish the study, please share with us your experience and feedback!

What do you think about the AI definition assistance?

Please answer the following questions based on your experience of using the assistance in this study.

I am confident in the AI definition. I feel that it works well.*

Disagree strongly Agree strongly

The AI definition is very reliable. I can count on it to be correct all the time.*

Disagree strongly Agree strongly

I will use the AI definition again in the future.*

Disagree strongly Agree strongly

What do you think about the task?

Please answer the following questions based on your experience of doing the task.

How mentally demanding was the task?*

Very low Very high

How hard did you have to work to accomplish your level of performance?*

Very low Very high

Please leave this question blank.

Very low Very high

Is there anything else you want to share with us?

We greatly value your thoughts on creating better language learning tools and experiences for the future!

If the AI definition assistance was helpful in doing the writing task, what aspects were most helpful to you?*

Enter your response

If the AI definition assistance wasn't quite what you needed for the writing task, what would have made it better?*

Enter your response

Any additional comments or feedback about the study?

Enter your response

Complete Study

(j) Post-task Survey