# Multilingual Datasets for Augmentative and Alternative Communication

**Keith Vertanen**
Department of Computer Science
Michigan Technological University, Houghton, MI, USA
vertanen@mtu.edu

## Abstract

There is a lack of text resources for building predictive Augmentative and Alternative Communication (AAC) interfaces in many languages. We share a compilation of four multilingual datasets in hopes of improving the situation. We started with four English datasets designed to match the conversational-style text AAC users may want to produce in person-to-person conversations. Next, we used a large language model (LLM) to correct errors and augment the original English data. Finally, we used Google's translation LLM to convert the datasets' 9,154 individual communications and 37,486 dialogue turns into 33 other languages.

## 1 Introduction

People with little or no functional speech due to a disability may rely on Augmentative and Alternative Communication (AAC) to participate in everyday conversations. For literate AAC users, an interface that uses a language model to make character or word predictions is often used. Some example predictive AAC systems include:

- **Direct selection keyboards** — Displays characters and word predictions that are directly selected in some way (e.g. by using a finger or by using an eye-tracker to detect gaze fixations).
- **Scanning keyboards** — Highlights characters and word predictions in some order. Users take some action when their desired target is highlighted (e.g. pushing a button or twitching a muscle). A common form is row-column scanning where targets are in a grid. The first action selects a row and the second selects a target.
- **Dasher** — In Dasher (Ward et al., 2000), users write letter-by-letter by navigating a world of nested letter boxes. Letters appear in alphabetical order from top to bottom on the right of the interface. Users navigate by pointing in some way (e.g. a mouse or eye-tracker). Pointing up and down controls what letter is written next. Pointing left and right controls navigation speed. The size of a letter box is proportional to its probability under a language model.

But even with a predictive AAC system, entry rates are often below 10 words-per-minute (Polacek et al., 2017; Higginbotham et al., 2007). This is far below a speaking communication partner who might speak at 150 words-per-minute. Making good predictions is crucial for closing this gap. But this requires language models optimized for this task. For privacy reasons, there are no corpora containing large amounts of text written by AAC users engaged in conversations. Past research has found training (Vertanen and Kristensson, 2011) or fine-tuning (Gaines and Vertanen, 2025) language models using informal corpora such as Twitter, written dialogues (Li et al., 2017), or short messages (Song et al., 2014) improves the prediction of AAC-like text. Crowdsourcing has also been used to create datasets of communications (Vertanen, 2013; Vertanen and Kristensson, 2011) and dialogues (Vertanen, 2017). But conversational text resources tend to be in English, hindering progress for interfaces such as Dasher which might otherwise easily support communicating in many different languages.

Further, AAC system developers may want to evaluate different pretrained language models on their base ability to predict conversational text. But even if test text can be found in each target language, differing levels of conversational style or semantic content can make precise comparison tricky.

Our contribution is to organize, improve, and provide translated versions of four existing conversational datasets. Our datasets[1] have Creative Commons licenses allowing for commercial (three datasets) or non-commercial use (one dataset).

[1] https://zenodo.org/records/23565608

| | | |
|---|---|---|
| English | Dutch | Polish |
| Afrikaans | Finnish | Portuguese |
| Albanian | French | Russian |
| Arabic | German | Slovak |
| Basque | Greek | Slovenian |
| Bengali | Hebrew | Spanish |
| Catalan | Italian | Swahili |
| Chinese (Simplified) | Japanese | Swedish |
| Chinese (Traditional) | Mongolian | Turkish |
| Croatian | Norwegian | Ukrainian |
| Czech | Persian | Welsh |
| Danish | | |

Table 1: The 34 languages included in our datasets.

## 2 Dataset Details

### 2.1 Process Overview

We selected datasets based on: 1) being publicly available, 2) having a permissive license, 3) demonstrated utility for AAC text entry, and 4) created in English by humans. Two datasets had isolated single communications and two had dialogues between two people. All datasets were created by writing on a computer (as opposed to, for example, being transcribed from speech recordings).

For some datasets, we performed correction and augmentation tasks using Llama 4 Maverick (400 B total parameters).[2] We hosted the model via Ollama[3] and ran it on eight 40 GB A100 GPUs. See Appendix A or our repository for the exact prompts we used for the correction and augmentation tasks.

We selected our 34 supported languages (Table 1) based on the languages in the open-source AAC interface Dasher (Ward et al., 2000) and in the products of the assistive technology company Smartbox. We used Google's translation LLM to convert the English communications and dialogues to each language. When making our API calls, we considered a single communication and all turns of a dialogue to be a "document". All translations were performed in July of 2025. We now describe the details of each of the four datasets we curated.

### 2.2 Multilingual COMM2

For evaluating an AAC system, Venkatagiri (1999) had people write statements and questions in response to ten everyday communication situations (e.g. talking to your mother). We copied this approach and had Amazon Mechanical Turk workers create a 1,506 communications (Vertanen, 2013).

First, we had the Llama 4 LLM "correct any obvious mistakes in spelling, punctuation, or capitalization" in the original dataset's communications. See Appendix A for our exact prompt. 15% of the communications were corrected. Of the 232 corrections, 86 corrections were exclusively attributable to punctuation changes with the most common being adding commas (56) or converting them to semicolons (19). Other corrections were exclusively attributable to changing words to fix grammar or spelling issues (43), changes to casing (36), or changes to whitespace (12). The remaining 55 corrections involved multiple types of changes.

To verify the corrections, we marked whether each of the original 1,506 communications needed correction. We did this without knowledge of whether each communication had been corrected by the LLM. We marked 6% as needing correction. Of these 90 communications, the LLM had corrected 81. Of the nine we marked that the LLM did not correct, seven involved subjective changes involving the use of hyphens, commas, and end of sentence punctuation. The remaining two involved the grammatically awkward sentence "May you please give me directions to the museum?" and a sentence we flagged due to the word "double-double". The latter is the name of a type of hamburger at a popular fast-food chain.

We reviewed the 151 corrections flagged only by the LLM. These were classified as punctuation only (62), word use only (34), whitespace only (11), casing only (8), and multiple types of changes (36). Common punctuation corrections including adding commas (39) and replacing commas with semicolons (18). The word-use corrections included clear errors we had missed, such as changing "withdrawal" to "withdraw" in "I would like to withdrawal all my money." Other word-use corrections were more subjective, such as adding words such as "that" or "of". The LLM also caught spelling errors we had missed, such as changing "we're" to "were". Overall, we judged the correction process to be accurate. Our repository contains the human scoring and the automatic corrections.

We also used the LLM to create a more concise version of each communication that did not change its content or impair the conversation partner's ability to understand it. See Appendix A for our exact prompt. The original and corrected communications averaged 7.8 words while the concise version was 3.7 words. Table 2 gives examples including the corrected and concise versions.

[2] https://ollama.com/library/llama4:maverick
[3] https://github.com/ollama/ollama

| Type | Version | Text |
|---|---|---|
| statement | original | I would like some fried with that. |
| | corrected | I would like some fries with that. |
| | concise | Fries with that. |
| question | original | When date is this assignment due? |
| | corrected | When is this assignment due? |
| | concise | When is the due date? |

Table 2: Examples from the original COMM2 dataset as well as our corrected and concise versions.

### 2.3 Multilingual Turk

We previously created dialogues by first asking Amazon Mechanical Turk workers to invent a question they might ask someone (Vertanen, 2017). These questions were shown to another worker who first decided if the question was plausible, and if so, to invent a response. Two turns were then shown to another worker who added a third turn, and so on until the dialogue had six turns.

The original dataset contains 1,419 dialogues. The original dataset was split into a training set (60%), a development set (20%), and a test set (20%). We retained the same split.

Similar to COMM2, we had the Llama 4 LLM first correct the text in the original dialogue turns. 28% of dialogue turns were corrected with 90% of dialogues needing at least one turn corrected.

We also asked the LLM to invent a new conversation between two people about a similar topic as each of the corrected dialogues. Table 3 shows an example of a corrected dialogue and the augmented similar version. We asked for about six dialogue turns but did not strictly enforce this. The invented dialogues were occasionally longer with 5.6% having 7 turns, 15.2% having 8 turns, and 0.8% having 9 or 10 turns. The invented dialogue turns were also more verbose, averaging 12.2 words versus 7.7 words in the corrected dialogue turns.

### 2.4 Multilingual Imagine

We previously used Amazon Mechanical Turk to create a large set of fictional AAC-like communications (Vertanen and Kristensson, 2011). Workers were asked to invent communications as if they were using a row-column scanning AAC interface for communication. The original data contained 6,142 communications and was split into a training set (80%), development set (10%), and test set (10%). We retained the same split.

As with previous datasets, we corrected the communications using Llama 4. We found this dataset

| Person | Dialogue turn |
|---|---|
| A | Did you call the theater? |
| B | So sorry, I forgot to call the theater. |
| A | You can just go online. |
| B | That's true, I'll do that now. |
| A | What movie is it that you want to see? |
| B | The Lord of the Rings. |
| A | Do you want to check the showtimes for tonight? |
| B | I was going to do that, but I got distracted. |
| A | You can just look it up on your phone, it's easy. |
| B | Yeah, that's a good idea, I'll do it now. |
| A | Are you thinking of watching any Marvel movies? |
| B | Actually, I was thinking of watching Spider-Man. |

Table 3: Examples of a corrected Turk dialogue (top) and the similar dialogue (bottom) created by Llama 4.

| Error type | Original | Corrected |
|---|---|---|
| spelling | Please be quit. | Please be quiet. |
| grammar | Let's go for a dinner. | Let's go for dinner. |
| casing | please let this work! | Please let this work! |
| whitespace | Don't let the bugs bite ! | Don't let the bugs bite! |
| punctuation | Could I have some milk please? | Could I have some milk, please? |

Table 4: Examples of the original and corrected communications from the Imagine dataset.

was particularly dirty with 46% of communications needing correction. The corrected communications averaged 5.1 words. Table 4 shows some example communications with different types of errors.

### 2.5 Multilingual DailyDialog

Li et al. (2017) crawled websites designed to teach English. They harvested examples of dialogues occurring between two people about various topics such as school or going out to eat. The original data contains 13,118 dialogues with 1,000 reserved as a development set and 1,000 as a test set. We retained the same split, but due to translation costs, we used only the first 20% of each set (2,224 training, 200 development, 200 test). Following the original license of DailyDialog, our version also does not allow commercial use (CC BY-NC-SA).

We found that the turns in DailyDialog were mostly free of spelling, grammar, and capitalization errors; we did not use Llama 4 to correct the original dialogues. In our 20% subset of the original data, the average turn contained 11.3 words. Our subset contained dialogues with 1–31 turns, with 7.6 turns on average. Table 5 shows an example dialogue from the original dataset.

| Person | Dialogue turn |
|---|---|
| A | What a great weekend that was! |
| B | My feelings exactly. |
| A | I really enjoy the beach in the summer. |
| B | I couldn't agree with you more. |
| A | We should come here more often. |
| B | You are absolutely right. |
| A | I'd like to come back next weekend. |
| B | That's exactly the way I feel. |

Table 5: Example from the original DailyDialog dataset.

### 2.6 Dataset format

Each dataset contains a top-level folder with the original communications or dialogues. All text data is stored in JSON format. Below the top-level folder are subfolders for each of the 34 languages. The en subfolder contains the corrected or truncated original data, and any augmented variants.

## 3 Dataset Analysis

### 3.1 Translation expansion

The length of each *utterance* (either a communication or a dialogue turn) varied widely between the four datasets. Imagine had the shortest utterances at 19.9 non-whitespace Unicode characters in the correct English utterances. COMM2 was the next longest at 24.6 characters. Turk and DailyDialog were similar at 42.2 and 47.9 characters respectively. We saw similar length trends between the datasets within each language's translated data. See Table 7 in Appendix B for a complete list of lengths in each dataset for each language.

We computed the ratio between the non-whitespace characters in the target language versus English for each utterance in every target language and dataset. We did this including any dataset splits or augmented versions. We found that the single independent communications tended to expand more when translated. The median ratio was 1.085 for COMM2 and 1.091 for Imagine versus 1.030 for Turk and 1.034 for DailyDialog. Figure 1 shows this trend was true for most languages.

We conjecture the more verbose translations for COMM2 and Imagine were due to the lack of surrounding dialogue context as compared to Turk and DailyDialog. The informal nature of the dialogues may have encouraged translations that were less formal and expressible in fewer characters in many languages. We analyzed this in detail for Spanish by looking for identical utterances that occurred in both Imagine and Turk. There were 471 distinct

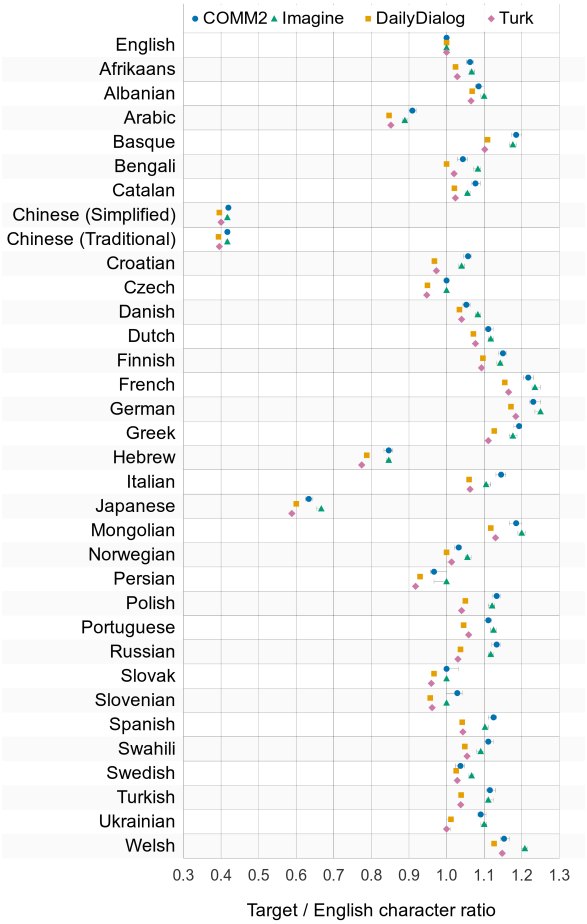


Figure 1: Median target-to-English non-whitespace character ratios by language and dataset. Error bars show 95% bootstrap confidence intervals.

communications of which 260 had different translations. Among the 260, the translation was typically longer in Imagine for 153, shorter for 45, and equal for 62. For distinct utterances occurring more than once, we compared the median translation length across occurrences. Table 6 shows some examples.

### 3.2 Preserved conciseness

Recall we used the Llama 4 LLM to create a concise version of each English communication in COMM2. This version was on average 53% shorter. To investigate whether the translated versions preserved the same level of conciseness, we calculated a ratio to measure the preserved conciseness:

$$\frac{T_{\text{concise}}/T_{\text{full}}}{E_{\text{concise}}/E_{\text{full}}} \quad (1)$$

where $T_{concise}$ and $T_{full}$ are the non-whitespace characters in the translated concise and original communications respectively. $E_{concise}$ and $E_{full}$ are the characters in the English concise and orig-

| Imagine Spanish | Turk Spanish |
|---|---|
| ***Let me check.*** | |
| Permítame verificar. | Déjame ver. |
| ***Can I get a ride?*** | |
| ¿Puedo tomar un viaje? | ¿Me llevas? |
| ***Can you lend me a hand?*** | |
| ¿Puedes echarme una mano? | ¿Me ayudas? |
| ***Can I have some wine?*** | |
| ¿Puedo tomar un poco de vino? | ¿Me traes vino? |
| ***I love you.*** | |
| Te amo. | Te quiero. |

Table 6: Different Spanish translations of identical English utterances in Imagine and Turk. All but "I love you." were shorter when embedded in a Turk dialogue during translation.

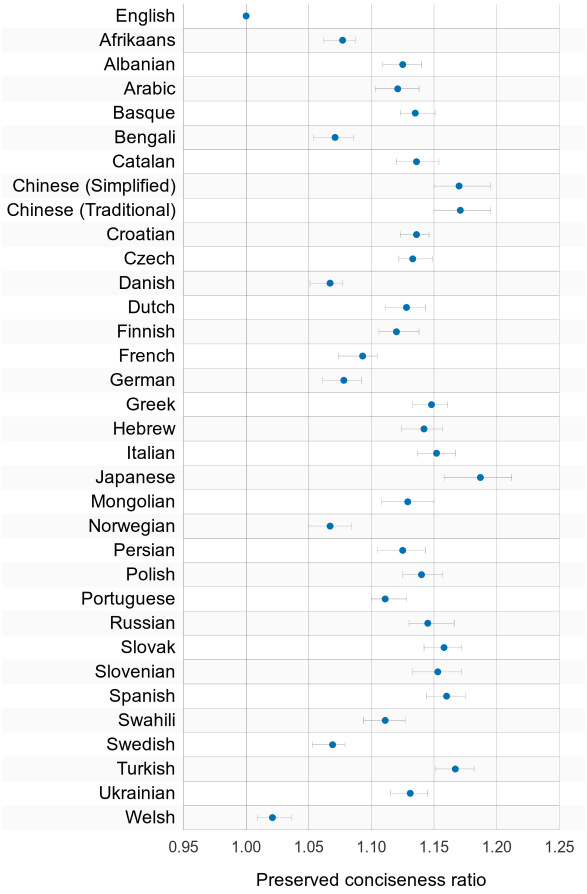


Figure 2: Preservation of COMM2 conciseness after translation. Points are the median preserved-conciseness ratio. Error bars show 95% bootstrap confidence intervals. A ratio of 1 means the proportional reduction from full to concise was preserved exactly; values above 1 indicate some conciseness was lost.

inal communications respectively. A ratio of 1.0 means the translation maintained a similar level of shortening as in English. A ratio above 1.0 means the translation had less shortening.

We found all 33 target languages had a ratio above 1.0, with a median ratio of 1.13. As shown in Figure 2, ratios varied significantly between languages with a minimum of 1.02 for Welsh and 1.19 for Japanese. This translation expansion of shorter English source texts is similar to our previous finding on COMM2 and Imagine communications. In this case, it would be interesting to compare the approach of translating the concise version generated by the LLM in English versus using a LLM to make a concise version of the non-concise translation.

## 4 Conclusions

We have presented four multilingual parallel datasets for improving predictive text entry interfaces. We started with four existing English datasets, improving and expanding the datasets via correction and augmentation using a large language model. We then translated each dataset into 33 languages commonly used in commercial and open-source AAC interfaces. Our datasets all focus on text that might be used in everyday conversations. We hope our datasets can be used to improve the ability of non-speaking individuals to converse in a variety of languages.

## Limitations

The original conversational utterances were written in English and automatically translated into 33 other languages. These translations may contain errors or exhibit unnatural style. While it would be ideal to collect conversational data from native speakers of each language, this can be expensive. It also does not control for the semantic content of the dataset in each language. Human expert translation from a single source language dataset into each target language might address this concern, but would also be expensive.

The COMM2 and Imagine datasets consisted of single communications invented by Amazon Mechanical Turk workers. With no surrounding context, this may have made accurate translation more difficult than when communications are translated when embedded in a larger conversation. It also seems to have resulted in less concise translations compared to translations of multi-turn dialogues.

Due to translation costs, we only translated the datasets into 33 languages. While the chosen languages represent perhaps 45% of the mother tongues of the world's population, expansion to more languages might allow improved predictive

AAC to more of the world in their native language.

Our datasets are all based on datasets that are available online. In some cases, the data has been available for a substantial period of time (e.g. since 2011 for Imagine). Thus, we caution that pretrained language models may have been contaminated by the text in our datasets. If these datasets are used for evaluating different pretrained models, it could lead to overly optimistic performance estimates for contaminated models.

All our data is based on humans producing individual communications or turns in a written conversation. They were not based on real-world, real-time conversations where some participants may speak instead of write. Further, they were not sourced from actual AAC users.

It would be interesting to study the simulated text entry performance of different LLMs over the set of languages in our datasets. For an interface like Dasher (Ward et al., 2000), computing the perplexity or bits per character would be a good proxy for potential user performance. But computing an accurate distribution over the next character requires marginalizing over different possible tokenizations (Gaines and Vertanen, 2025). Doing this robustly over many LLM tokenizers, model architectures, and target languages could be challenging.

Our datasets only have conversations in a single language. In some communities, code switching between several languages is common (e.g. Spanglish). Building AAC interfaces to support such users is an interesting challenge that may require curating new datasets.

## Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. IIS-1750193 and IIS-2402876.

## References

Dylan Gaines and Keith Vertanen. 2025. Adapting large language models for character-based augmentative and alternative communication. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 15273–15291, Suzhou, China. Association for Computational Linguistics.

D. Jeffery Higginbotham, Howard Shane, Susanne Russell, and Kevin Caves. 2007. Access to AAC: Present, past, and future. *Augmentative and Alternative Communication*, 23(3):243–257.

Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. 2017. DailyDialog: A manually labelled multi-turn dialogue dataset. In *Proceedings of The 8th International Joint Conference on Natural Language Processing (IJCNLP 2017)*.

Ondrej Polacek, Adam J. Sporka, and Pavel Slavik. 2017. Text input for motor-impaired people. *Universal Access in the Information Society*, 16(1):51–72.

Zhiyi Song, Stephanie Strassel, Haejoong Lee, Kevin Walker, Jonathan Wright, Jennifer Garland, Dana Fore, Brian Gainor, Preston Cabe, Thomas Thomas, Brendan Callahan, and Ann Sawyer. 2014. Collecting natural SMS and chat conversations in multiple languages: The BOLT phase 2 corpus. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 1699–1704, Reykjavik, Iceland. European Language Resources Association (ELRA).

Horabail Venkatagiri. 1999. Efficient keyboard layouts for sequential access in augmentative and alternative communication. *Augmentative and Alternative Communication*, 15(2):126–134.

Keith Vertanen. 2013. A collection of conversational AAC-like communications. In *ASSETS '13: Proceedings of the ACM SIGACCESS Conference on Computers and Accessibility*.

Keith Vertanen. 2017. Towards improving predictive AAC using crowdsourced dialogues and partner context. In *ASSETS '17: Proceedings of the ACM SIGACCESS Conference on Computers and Accessibility (poster)*, pages 347–348.

Keith Vertanen and Per Ola Kristensson. 2011. The imagination of crowds: Conversational AAC language modeling using crowdsourcing and large data sources. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP'11, pages 700–711. ACL.

David J Ward, Alan F Blackwell, and David JC MacKay. 2000. Dasher—a data entry interface using continuous gestures and language models. In *Proceedings of the 13th annual ACM symposium on User interface software and technology*, pages 129–137. ACM Press.

## A LLM Prompts

We used the Llama 4 Maverick LLM to correct obvious problems in the original text of the communications in COMM2 and Imagine datasets. Figure 3 gives the exact prompt we used for doing the correction. We also corrected turns in the Turk dialogues. Figure 4 shows the prompt for correcting these dialogues.

We prototyped a variety of prompts for making a communication in COMM2 more concise. This included asking for a certain percentage reduction in the total text. We got the best results by explaining the motivation about why an AAC user might

```
Here is the first turn in a conversation:

__TURN__

Correct any obvious mistakes in spelling,
punctuation, or capitalization. Do not otherwise
change the turn in the conversation.

Only output the corrected turn in the
conversation, do not describe what you did. Your
result should consist of a single line of text.
```

Figure 3: Prompt used for correcting errors in the original communications in the COMM2 and Imagine datasets.

```
Here is a conversation between two speakers
labeled A and B:

__TURNS__

Correct only obvious mistakes in spelling,
punctuation, or capitalization.
Do not describe your thinking or what you
corrected.
Do not change the conversation.
Do not add information.
Do not add additional turns to the conversation.
Only output the six turns of the corrected
conversation.
```

Figure 4: Prompt used for correcting errors in the original Turk dialogue data.

want to be concise. We found it also important to be clear that the communication needed to remain understandable in order to avoid too aggressive compression. Figure 5 shows our final prompt.

We created a similar dialogue for each original dialogue in the Turk dataset. We initially tried asking Llama 4 to generate new dialogues without basing it on an existing dialogue. However, we found this led to many dialogues being about the same topic such as visiting the local coffee shop.

```
Here is something a person said to someone else
in a conversation:

__TURN__

Imagine the person has ALS and uses a computer
to communicate. It takes the person several
seconds to enter each letter.

Without changing the content of what the person
said, make the conversation turn more concise
such that the other person would still
understand exactly what the person means.

Do not discuss your thought process or reasoning.
Output only 1 line containing the final text of
the new concise turn.
```

Figure 5: Prompt used for making a concise version of a COMM2 communication.

```
Here is a conversation between two speakers
labeled A and B:

__TURNS__

Invent a new conversation between two speakers
labeled A and B.
The new conversation should be about a similar
topic.

Each turn should appear on a separate line.
Each line should start with the label of the
speaker.
Speaker A should start the conversation.
Turns should alternate between speaker A and B.
The conversation should have about 6 turns.
The conversation should be in a conversational
style similar to the previous conversation.
The conversation should use good punctuation,
good spelling, and be in __LANG__.
```

Figure 6: Prompt used for making a similar dialogue based on an existing Turk dialogue.

Anchoring on an existing dialogue topic allowed augmentation over the diverse topics in the original human generated dialogues, but with the LLM introducing substantial variability in the exact turns and the words used.

We experimented with having Llama 4 do both the dialogue augmentation and translation to the target language at the same time. However, we found Llama 4 struggled to both invent the new dialogue and translate at the same time. Often it generated a mixture of English and the target language. This was especially true for low-resource languages and may reflect the lack of sufficient text in those languages in Llama 4's training data. Figure 6 shows our final prompt.

## B Corpus statistics

Table 7 gives details of the average length of communications and dialogue turns for each language and dataset. In rare cases, Google Translate failed to return any result for a given utterance. This occurred most often in Arabic (41 out of 46,640).

For each language, we counted the distinct characters that appeared anywhere in the four datasets. Different types of whitespace were treated as a regular space. See our repository for the character inventories. We caution these inventories are based solely on our four relatively small datasets. For languages with large character sets such as Chinese, these inventories are likely missing many characters. The inventories may also be idiosyncratic due to how the datasets were originally created or by the characters produced by Google Translate.

Table 7: Mean non-whitespace Unicode characters per utterance in each multilingual collection. A COMM2 or Imagine message counts as one utterance; in DailyDialog and Turk, each individual dialogue turn counts as one utterance. Unique characters is based on the union of the characters observed in the four datasets. Empty utterances is the number of utterances containing no text.

| Language | ISO | Mean characters per utterance | | | | | Unique | Empty |
|---|---|---|---|---|---|---|---|---|
| | code | COMM2 | DailyDialog | Imagine | Turk | Overall | characters | utterances |
| English | en | 24.6 | 47.9 | 19.9 | 42.2 | 40.6 | 98 | 0 |
| Afrikaans | af | 26.2 | 49.5 | 21.4 | 43.6 | 42.1 | 107 | 0 |
| Albanian | sq | 27.0 | 51.5 | 21.9 | 45.2 | 43.6 | 97 | 1 |
| Arabic | ar | 22.8 | 40.5 | 18.1 | 36.5 | 34.9 | 150 | 41 |
| Basque | eu | 29.0 | 53.3 | 23.5 | 46.6 | 45.3 | 99 | 12 |
| Bengali | bn | 25.9 | 48.4 | 21.6 | 43.3 | 41.5 | 159 | 1 |
| Catalan | ca | 26.8 | 49.5 | 21.3 | 43.5 | 42.0 | 114 | 2 |
| Chinese (Simplified) | zh-CN | 10.1 | 18.5 | 8.2 | 16.7 | 15.9 | 3279 | 21 |
| Chinese (Traditional) | zh-TW | 10.0 | 18.4 | 8.2 | 16.6 | 15.8 | 3322 | 21 |
| Croatian | hr | 25.9 | 46.7 | 20.7 | 41.3 | 39.9 | 104 | 1 |
| Czech | cs | 24.9 | 45.7 | 19.7 | 40.1 | 38.8 | 123 | 1 |
| Danish | da | 26.2 | 49.8 | 21.6 | 44.1 | 42.4 | 104 | 0 |
| Dutch | nl | 27.5 | 51.8 | 22.3 | 45.8 | 44.1 | 106 | 0 |
| Finnish | fi | 28.1 | 52.8 | 22.6 | 46.2 | 44.8 | 100 | 6 |
| French | fr | 30.4 | 55.3 | 24.8 | 49.3 | 47.4 | 120 | 1 |
| German | de | 30.5 | 56.6 | 24.9 | 50.4 | 48.4 | 106 | 0 |
| Greek | el | 29.2 | 54.6 | 23.4 | 47.2 | 46.1 | 162 | 1 |
| Hebrew | he | 20.6 | 37.7 | 16.8 | 32.7 | 32.0 | 124 | 8 |
| Italian | it | 28.3 | 51.5 | 22.3 | 45.3 | 43.8 | 102 | 0 |
| Japanese | ja | 15.4 | 28.1 | 13.1 | 24.9 | 24.1 | 2313 | 9 |
| Mongolian | mn | 28.9 | 53.0 | 23.7 | 47.4 | 45.5 | 159 | 1 |
| Norwegian | no | 25.6 | 48.5 | 21.1 | 42.9 | 41.3 | 105 | 1 |
| Persian | fa | 23.8 | 45.0 | 19.5 | 39.1 | 38.1 | 157 | 18 |
| Polish | pl | 28.0 | 50.7 | 22.5 | 44.2 | 43.1 | 113 | 5 |
| Portuguese | pt | 27.4 | 50.2 | 22.3 | 44.9 | 43.1 | 119 | 2 |
| Russian | ru | 27.8 | 49.9 | 22.3 | 44.0 | 42.6 | 159 | 0 |
| Slovak | sk | 25.3 | 46.7 | 20.2 | 40.7 | 39.6 | 127 | 1 |
| Slovenian | sl | 25.2 | 46.2 | 20.1 | 40.9 | 39.4 | 102 | 1 |
| Spanish | es | 27.7 | 49.9 | 22.0 | 44.3 | 42.7 | 113 | 5 |
| Swahili | sw | 27.2 | 50.5 | 21.7 | 44.5 | 43.0 | 94 | 0 |
| Swedish | sv | 25.7 | 49.5 | 21.2 | 43.6 | 42.0 | 102 | 0 |
| Turkish | tr | 27.7 | 49.9 | 22.4 | 44.1 | 42.6 | 108 | 22 |
| Ukrainian | uk | 27.2 | 49.3 | 22.2 | 43.0 | 41.9 | 162 | 1 |
| Welsh | cy | 28.5 | 54.2 | 24.1 | 48.5 | 46.4 | 112 | 0 |