

A Collection of Conversational AAC-like Communications

Keith Vertanen
Montana Tech
Butte, Montana, USA
kvertanen@mtech.edu

ABSTRACT

We contribute a public test set of everyday conversational communications. The communications were written in response to ten hypothetical situations given to workers on the crowdsourcing site Amazon Mechanical Turk. After quality control, our public dataset consists of 1,506 unique communications. These communications can be used to help design and evaluate text-based predictive communication aids. The collection also provides a common public test set for research into predictive conversational text entry.

Categories and Subject Descriptors

K.4.2 [Computers and Society]: Social Issues - assistive technologies for persons with disabilities.

Keywords

Augmentative and Alternative Communication, Evaluation

1. INTRODUCTION

For non-speaking individuals with good literacy skills, text-based augmentative and alternative communication (AAC) devices can aid in participation in everyday conversations. To increase communication rates, AAC devices often predict letters and words based on the user's previous text.

Researchers often evaluate predictive text entry interfaces by testing how well the system performs on a set of test sentences either in offline experiments or in a user evaluation where users transcribe the test sentences. While ideally a test set of actual AAC communications would be available, to our knowledge no such public test set exists. In this paper we contribute a large, public test set of conversational-style communications for use by AAC researchers.

2. DATA COLLECTION

Our collection methodology parallels the one described by Venkatagiri [2]. Users were presented with 10 different hypothetical communication situations and asked to enter two responses, one in the form of a statement (giving information

or opinion) and the other in the form of a question. Each of the 10 situations asked the user to imagine they were communicating with a different type of person: friend, father, mother, sister/cousin, brother/cousin, teacher, counterperson (people who attend to counters such as bank tellers and fastfood servers), stranger, salesperson, or acquaintance.

We collected a large set of test sentences from a diverse group of participants via a web-based version of Venkatagiri's [2] collection methodology. Our approach was inspired from previous work in which we successfully leveraged workers' creativity on the crowdsourcing site Amazon Mechanical Turk to create plausible AAC-like communications [3].

The web data collection proceeded as follows:

1. *Basic Information Sheet*: Workers were asked to provide basic information about themselves, such as age, sex, country and English proficiency.

2. *Sentence Collection*: Workers were presented with the 10 situations in random order (Figure 1). Workers were asked to reproduce utterances they had recently used, or to make up a sentence they might use in the given situation. After entering each sentence, workers indicated whether they had actually used the utterance in real life. Workers were instructed not to include private information. Workers were prevented from pasting text into the text result field.

3. *Basic Information Verification*: Workers were asked to provide basic information about themselves again, but this time in permuted order. We used the two sets of information to filter out workers who did not provide consistent information. Providing inconsistent information suggests the worker may have been "gaming" the task to obtain a quick monetary reward rather than providing well thought out responses.

3. RESULTS

We conducted the data collection task via a human intelligence task (HIT) on Amazon Mechanical Turk. We restricted the HIT to workers in the United States who had a 95% accepted HIT rate. The HIT paid \$0.45 and took workers on average 10 minutes to complete. In total, 80 unique workers completed the task, 66% self-reported as being female, and 96% self-reported being native English speakers.

We used a semi-automated process to flag workers who submitted garbage, repeated communications, or had discrepancies in their two sets of basic information. We removed three workers due to such violations. Overall we found the work-

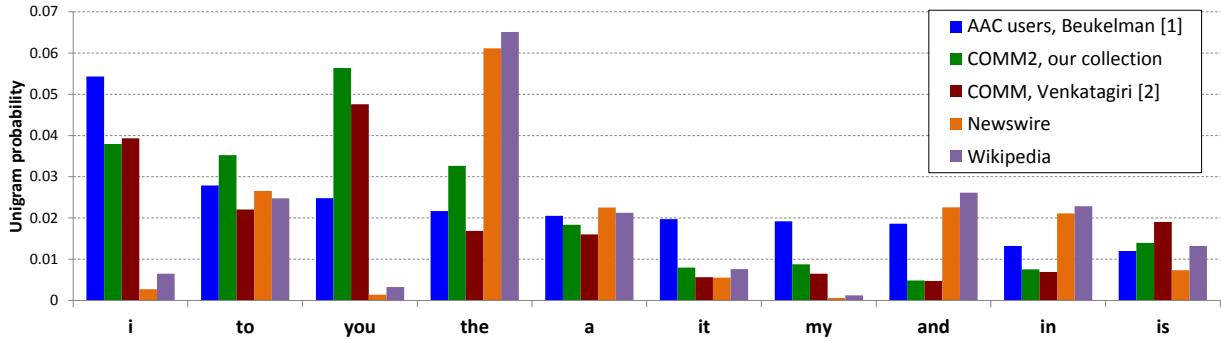


Figure 2: The unigram probabilities of the top-10 words in five different datasets.

Situation 1/10
 For the given communication situation, write two sentences. One sentence will be in the form of a statement (giving information or opinion) and the other will be in the form of a question.
 As best you can, **reproduce utterances you have recently used in a similar communication situation**. If this is not possible, make up a sentence you might likely use in the given situation.

Do NOT include any private information (such as real email addresses, phone numbers, or names).
 Do NOT cut and paste (we will reject work done in this way).
 Do NOT use texting abbreviations or slang.
 Please write **complete sentences with good grammar and spelling**.

Communication situation: Talking to your mother.

With respect to the above communication situation, write a sentence in the form of a **statement** (giving information or opinion):

I have actually used this statement I might use this statement in the future

With respect to the communication situation, write a sentence in the form of a **question**:

I have actually used this question I might use this question in the future

[Continue >>](#)

Figure 1: Screenshot of the data collection task.

ers’ communications to be well-written and plausible (Table 1). We collected 773 pairs of statements and questions. We manually reviewed the communications in Microsoft Word, correcting obvious mistakes in spelling and grammar. We also eliminated 40 communications that were not sensible. The dataset is available online¹.

Our final set consists of 1,506 communications written by 77 crowdsourced workers. The workers indicated they had actually used 76% of the communications. Our corpus has 12,015 words (1,768 unique), with each communication having between 1 and 25 words (mean 7.8 sd 3.3), and 2 to 130 characters including spaces (mean 38.2 sd 16.7). For comparison, Venkatagiri [2] collected 260 communications from 13 students. The students indicated they had actually used 58% of the communications. Venkatagiri’s corpus had 1,865 words (644 unique) with a mean length of 7.2 words, with 12 to 91 characters per sentence (mean 36.4).

To compare our test against genuine AAC user data, we used information provided in a study by Beukelman et al. [1]. This study analyzed the communications made by five non-speaking adults over a 14-day period using a tape-typewriter AAC device. We compared the unigram probabilities of the top-10 words reported in [1] with our collection and with Venkatagiri’s original dataset [2]. As shown in Figure 2, in both our collection and Venkatagiri’s, the word “you” ap-

¹<http://aactext.org/comm2>

Table 1: Example communications.

- Good to see you!
- How are you doing today?
- I want to go on the rollercoaster.
- Is there a sale?
- Looks to be boring.
- Are you new to Chicago?
- I’m sorry, I don’t have any extra money to give you.
- That is a nice car you drive.
- Could you make me some breakfast?
- I heard it was going to rain today.
- What time is your appointment?

pears much more frequently than in the genuine AAC user data. We believe this resulted from users often writing questions directed at the other individual in the stated situation, e.g. “When are *you* leaving?” and “Do *you* have the time?”. Both our collection and Venkatagiri’s [2] show better similarity to the genuine AAC-data than non-conversational text sources such as newswire text or Wikipedia articles.

4. CONCLUSIONS

The development of improved predictive AAC interfaces can help accelerate the communication of literate AAC users. To aid this effort, we have contributed a public collection of conversational-style communications. By having anonymous workers invent communications, we were able to create a public resource allowing AAC researchers to compare results using a common test set. We believe further crowdsourcing efforts such as ours could yield even richer and more diverse sets of data useful for improving AAC.

5. REFERENCES

- [1] D. R. Beukelman, K. M. Yorkston, M. Poblete, and C. Naranjo. Frequency of word occurrence in communication samples produced by adult communication aid users. *Journal of Speech and Hearing Disorders*, 49:360–367, November 1984.
- [2] H. Venkatagiri. Efficient keyboard layouts for sequential access in augmentative and alternative communication. *Augmentative and Alternative Communication*, 15(2):126–134, 1999.
- [3] K. Vertanen and P. O. Kristensson. The imagination of crowds: conversational AAC language modeling using crowdsourcing and large data sources. In *Proceedings of the ACL Conference on Empirical Methods in Natural Language Processing*, pages 700–711. ACL, 2011.