Low resource scalar implicature in te reo Māori

Caleb Moses
caleb.moses@mail.mcgill.ca
dept. of Computer Science, McGill University

Abstract

In this paper, I designed an implicature game based on a teaching methodology designed to teach low resource languages in a multi-modal way.

1 Introduction

The Silent Way (Gattegno, 2010) is a language teaching methodology which emphasises the subordination of teaching to learning. Lessons are conducted entirely in the target language, with a minimal vocabulary established in the early stages which incorporates simple utterances, instructions as well as hand-gestures. Cuisenaire rods are used to establish a world-state, which becomes a point for discussion where utterances can be hypothesised by the student and tested. When done in groups, lessons are overseen by a teacher but students take turns demonstrating their understanding to each other, and are able to correct and learn together.

In New Zealand, as well as in other countries, the Silent Way has been adopted by the Māori people as a dominant teaching methodology, especially for second-language learners of the Māori language (te reo Māori). The methodology was specifically adapted to te reo Māori (Mataira, 1980; Ka'Ai, 2008) by Dame Kāterina Te Heikōkō Mataira and Ngoingoi Pewhairangi where it is referred to as "Te Ataarangi". In the classroom, students in Te Ataarangi are encouraged to master grammar by experiencing the language primarily through speaking and listening, and doing practical exercises with other students. The classroom operates according to the following rules:

- Speaking in English is not allowed
- Belittling other students is not allowed
- Students are not allowed to prompt other students
- Students can only reply when a question is posed directly to them
- Students must practice empathy with their peers

Because the classroom proceeds to the next topic only once every student has demonstrated their understanding, students cannot stall for time, expecting the teacher to lose patience and move the class on. Because of this, Te Ataarangi has been effective for many students who do not function well in a traditional classroom.

The experiments in this paper have been designed with the aim of replicating the Te Ataarangi method of teaching using computational techniques. One long-term goal of doing this could be to design computational language agents that are capable of teaching the language to new students using the method. In this paper, we aimed to investigate the feasibility of training an agent to deliver a Te Ataarangi lesson by carefully formulating the problem, defining the requirements of the agent and then testing out the components one by one. At the end, we would like to demonstrate that our computational model works *because* it follows the principles of the original method.

Following this line of thinking, we observe that a typical Te Ataarangi lesson generally takes the form of a scalar-implicature game, wherein the teacher uses cuisenaire rods and hand gestures to specify a world state, or a *context* (c) for discussion. They then give an *utterance* (u) corresponding to that context, and over time the job of the student is to identify the relevant grammatical rules and vocabulary so that they can generalise the given utterance to other contexts. The student can then demonstrate their knowledge by carefully constructing their own contexts, and proving they have the correct utterance to the teacher.

The lesson bears a resemblance to a Socratic dialogue and incorporates aspects of pragmatics.

Since the teacher must think about what the student knows or does not know about the language, and construct examples in order to falsify incorrect assumptions, or to add missing context that may also be leading the student astray. We therefore propose to model the lesson with a pair of Rational Speech Actors according to the Rational Speech Act (Frank and Goodman, 2012; Goodman and Frank, 2016) framework. We define two Rational Speech Actors, one each for the teacher and the student. Both agents implement a literal listener, a pragmatic speaker and a pragmatic listener. The literal listener is a model that defines a mapping from contexts (c) to utterances (u). The teacher agent is pre-trained, such that it has prior knowledge of the rules of the language. The student agent then learns through a dialogue with the teacher agent. The dialogue is set up in the following way:

- The teacher selects an (context, utterance) pair and provides it to the student by way of demonstration.
- The student then selects an (context, utterance) pair and conjectures an utterance for that state for the teacher to evaluate.
- The student receives feedback from the teacher, either by way of correction, or the conversation may move on indicating that the selection was correct. The feedback is provided as an additional (context, utterance) pair.
- This loop continues until the teacher is satisfied that the student has acquired the required knowledge.

The contexts under consideration are defined as an arrangement of cuisenaire rods. Each rod has a colour and a length, and they are sampled at random and arranged into a row. From the sample, a selection is made and the utterance is constructed to accurately identify the selected rods among their neighbours. In English, utterances therefore look may look like “The red rod”, “The blue rod on the left”, or “All of the rods” if all of the rods in the context are selected.

Throughout this work, we operate under the assumption that we are working with an under-resourced language. So we place an emphasis on methods that could reasonably be applied by a language community in the absence of large training datasets containing millions of tokens. Te reo Māori is used as the language for all experiments, although it would be easy to run these experiments in any language, and would require the hand labelling of no more than 1,000 examples. As a speaker of the Māori language, I was able to prepare the dataset for these experiments alone, in only a few hours.

We begin the project by training a recurrent neural network to solve this scalar implicature problem, using a small training dataset. Initially, it was proposed to use a small dataset for parsimony reasons, as the task as designed resembles a very early beginner Māori language lesson, requiring minimal prior knowledge of the language. This problem proved to be more difficult than initially anticipated, and so various strategies, such as data augmentation were attempted in order to pre-train a competent teacher agent for this task.

2 Related Work

2.1 Scalar implicature

Scalar implicature (Grice, 1975) is a fundamental problem class for any language learning system. As such, learning a new language involves solving implicature tasks of a range of types, in increasing complexity. While not sufficient on its own, mastering scalar implicature will teach a student about grammar, vocab and syntax that are common in everyday life. It is clear that in order for a conversational agent to be an effective teacher, it first must have a mastery of the target task.

The task of learning natural language quantifiers from abstracted world states was investigated in (Sorodoc et al., 2016), where the authors trained and tested whether neural networks were able to learn natural language quantifiers such as *no*, *some* and *all*. In their task, the model was trained on images containing 1-16 circles of up to 15 different colours. They generated all possible colourings in their dataset. They compared an RNN with a quantised memory network (qMN) and found that the latter outperformed the RNN with a model accuracy of 88.8% on familiar examples vs 65.7% for the RNN.

In (Zheng et al., 2021) the authors prepare a dataset of sample dialogues, where implicature is a present, and models are then trained in order to resolve the implicature to an explicit form by answering targeted questions constructed based on the sample dialogue. This bears some resemblance to the language instruction problem, in the sense that the question-answer pairs are given after a dialogue in order to check understanding. The formulation of the n-round dialogue as a sequence of QA-pairs $\{(Q_1, A_1), (Q_2, A_2), \cdots, (Q_n, A_n)\}$, where Q_i is the question raised by the first agent, and A_i is the response provided by the second agent, which may contain an implicature. The dataset contains 15,000 dialogues, where 6,000 were used for training, 4,000 for development and 5,000 for testing. This is considerably more than the dataset we are currently using, and also the dialogues are quite rich and varied, as compared to ours which focus on a very specific task.

2.2 Dialogue agents

Dialogue agents are often envisioned as a method for teaching new information. However, to convey information well it is necessary for a model to know when the information has been acquired by the target or not. This is a necessary part of teaching to any kind of curriculum. So the purpose of our use of dialogue agents in this paper is to track the progress of the student and make recommendations about what they need to learn next.

Previous work has trained rational pragmatic models for correctly generating and following natural language instructions for complex tasks (Fried et al., 2017). Indeed, sequential language games (Khani et al., 2018) have been explored using training data acquired via Amazon Mechanical-Turk. Personality modelling (Yang et al., 2020) is related to our goal task because it is necessary for the teacher to have an implicit model of the student's progress in order to make sensible world state recommendations. Other work has focused on making agents that communicate with humans in natural language (Lazaridou et al., 2020).

2.3 Under-resourced language modelling

Low resource language work in pragmatics often uses the term "low-resource" to refer to "not having a lot of data" as opposed to data from or relating to under-resourced language communities. In this way, there are a lot of low-resource works that have little resemblance to work on language revitalisation. Some work has focused on implementing pragmatic reasoning under computational constraints (Van Arkel et al., 2020), which the authors referred to "frugal pragmatism". Possibly this gap in the literature could be reflective of the fact that many under-resourced languages have yet to acquire models which are powerful enough to demonstrate strong pragmatic capabilities.

In some ways, many of the well known pragmatics problems are readily expressed in English as well as in other languages, and so to keep the work accessible to the largely English speaking research community, it is to be expected that pragmatics work in other languages is still quite rare. Perhaps relatedly, a large proportion of the under-resourced language language modelling work focuses on translation compared to solving task specific goals like question-answering. Generally speaking, the goal appears to be to translate out of the under-resourced language into a high resource language like English, and then to do any question-answering or other complex tasks in the high resource language, before translating it back.

3 Method

3.1 World state generation

The world states under consideration in this task are referred to as "contexts". Each context comprises of a sequence of randomly chosen cuisenaire rods, each of which have a given colour (chosen from a set of 8) and length (ranging from 1-10). A context can have up to 13 rods, and from the context a random subset is selected.

The goal of the task is to write an utterance which correctly identifies the selected rods from among the rest in the context. This problem was constructed to be similar to a typical first language lesson according to the Te Ataarangi method.

Sampling uniformly from a large world state space would result in a large proportion of states with high entropy. This would be undesirable, since the goal is to teach the language so we chose to start from simpler world states and gradually increase the complexity. For that reason, we used the following procedure to generate arrangements of cuisenaire rods:

- An entropy budget is decided in advance. This entropy is computed as the sum of the entropy of the colour and height.
- A rod is then drawn, with random height and colour.
- The entropy of the current configuration including the new rod.
- If it is above the threshold, then end otherwise add another rod.

3.2 Labelling world states

During labelling, the entropy budget begins set to 0.5, and then every 66 examples the budget was increased by 0.5 to a cap of 8.0 resulting in a sample size of 990. For certain easy cases, such as when all rods are selected, a pre-computed utterance such as "ngā rākau katoa" ("all of the rods") or "te rākau" ("the rod") would automatically be entered on behalf of the user. This happened in approximately 30% of the examples in the dataset.

3.3 Conversational Agents

In this work, we aim to implement a pair of conversational agents. A teacher $A_{teacher}$, and a student $A_{student}$. The agents comprise of two models, the first maps contexts to utterances. In the case of the teacher, this model is pre-trained to be correct, while the student begins with a randomly initialised model.

The second model is designed to suggest world states for demonstration. For the teacher, this functions as a curriculum that takes in the history of (context, utterance) pairs from the student, and suggests world states that will help the student arrive at correct conclusions. In the case of the student, the world states are chosen to demonstrate that it has acquired the ability to produce correct utterances for the kinds of world states that the teacher has demonstrated so far, as well as to make conjectures that will demonstrate that the patterns it has learned generalise in the right way.

3.4 Models

3.5 The dataset

3.6 The experiments

4 Results

Figures 1, 2 and 3 show the politeness, planning discourse, and sentiment levels respectively in emails of the POIs, executives, and normal employees.

The politeness of the execs and regular employees appeared to remain more or less constant throughout. However, in the POI emails, there was a distinct increase in politeness prior to the collapse, followed by a sharp drop thereafter, confirming the results in (Niculae et al., 2015).

The planning discourse frequencies in Figure 2 similarly showed a concordance with the features in (Niculae et al., 2015). There is a clear spike in planning discourse which happens in March 2001 – a time that coincides with the first public questioning of Enron's solvency in a *Fortune* article (McLean, 2001). While it is impossible to assert a causal relationship, it should be noted that all data anomalies observed do line up with major events in the Enron collapse, and are isolated to only POI emails and not the general executive emails.

Figure 3 did show a departure from the findings of (Niculae et al., 2015) where no clear increase in positive sentiment among the POI emails was seen.

Table 1 gives the model parameters and evaluation metrics for the best performing classifier trained. As seen, the model was competent at identifying POIs. This is promising given the presence

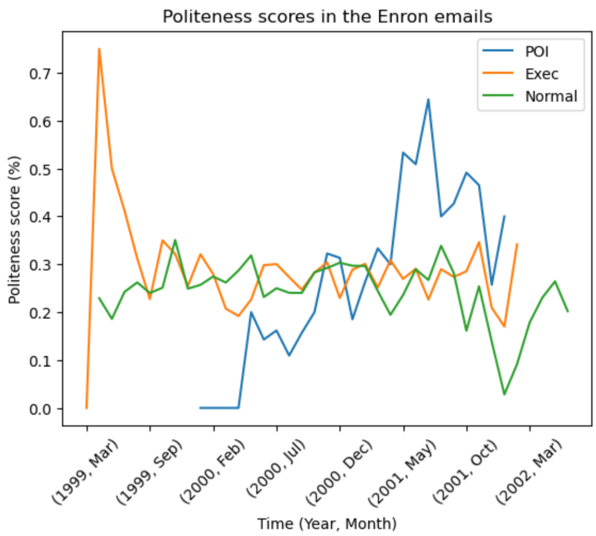

Figure 1: Politeness scores produced by the Stanford politeness classifier

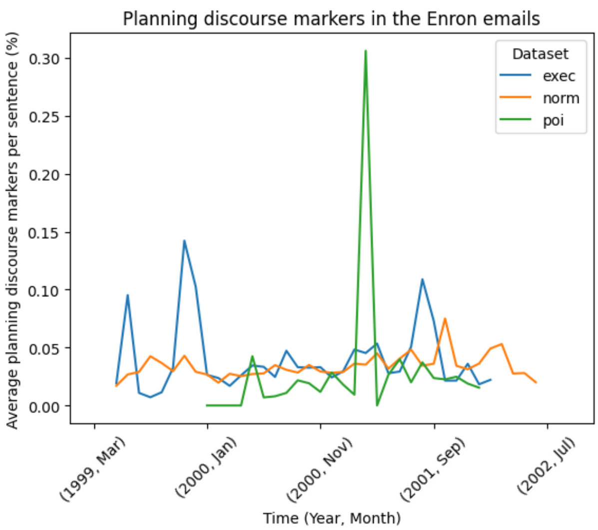

Figure 2: Percentage of planning discourse markers per sentence in the Enron emails

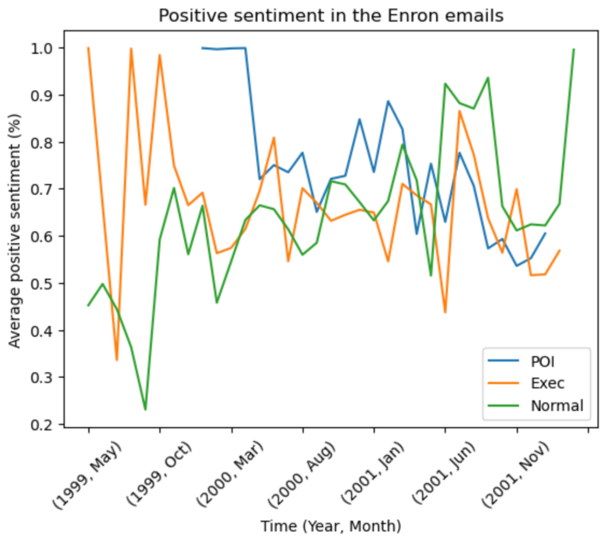

Figure 3: Sentiment fluctuation in Enron emails

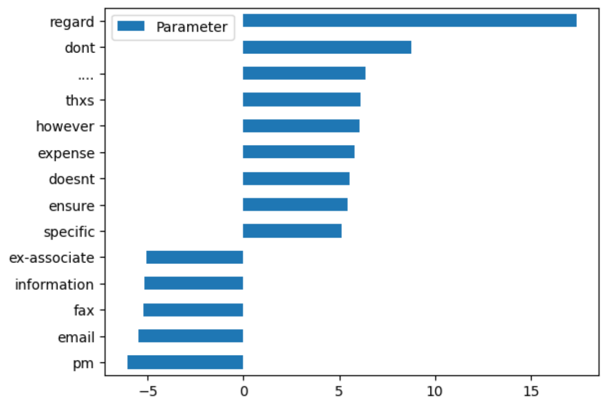

Figure 4: Parameter weights for the top 14 features deemed most useful for detecting emails written by fraudulent/deceptive actors, as suggested by our best classifier

of other executives in the data, and thus shows the models are able to identify features pertinent to POIs but not executives more broadly.

Figure 4 shows those features that were most useful for detecting fraudulent activity. Several features ('dont', 'however', 'doesnt') seemed to suggest a pattern of negative sentiment may be related to fraud, however it is interesting to note that this was not reflected in Figure 3

5 Discussion and Conclusion

Results in Figures 1 and 2 align well with the hypotheses in (Niculae et al., 2015), namely that increased politeness and planning discourse markers are cues related to deception. This gives credence to the conjecture that these linguistic cues are indeed useful in more complex environments beyond games, and are general traits related to deceptive behaviour.

Most notably, it is interesting that planning discourse markers increase dramatically at the very beginning of the collapse and then subside, even as further events are unfolding, whereas politeness increases gradually up to the very collapse. This is an interesting addition to the results in (Niculae et al., 2015). Diplomacy covers a short time span, whether measured by physical time or by number of events leading up to the deception. It is therefore not surprising that all deception markers overlap. In the case of Enron, the collapse was unfolding over a year, leaving more space between events, and therefore more time for deception markers to develop.

Metric	Value
Preprocessing	Lemmatization + Unigram
Model Class	Logistic Regression (C=10)
Precision	0.8204
Recall	0.8335
F1 Score	0.8170
Accuracy	0.8335

Table 1: Best Model Performance Metrics

The fact that our results align with the ones in (Niculae et al., 2015), though the alignment is not perfect, gives us some food for thought. In particular, since the emails analyzed are not directed at the targets of the deception (i.e., the shareholders and authorities), our findings suggest that an individual's general level of politeness increases when that individual engages in deception more generally.

Furthermore, the results from our classifier training gives some interesting insights into the word usage of deceptive and non-deceptive actors. Most notable are the words with negative weights (therefore indicative of a non-POI email). There we observe words like "fax" and "email" which may be more associated with non-executive individuals. Additionally words like "pm" suggest potential meetings which may be more commonplace for non-executive individuals.

These words may not be applicable to other settings, however additional work may be needed to confirm this.

5.1 Limitations

One major limitation of this work is that our classification of POIs only occurs based a single email, and does not account for prior context or email threads within its prediction. Doing so in such a situation could potentially give even more insightful cues in deception and potentially manipulation.

Another limitation is that a large part of the deception at Enron was targeted at the general public. As such, interviews and public communications by POIs would have added valuable linguistic cues that are not included in our analysis.

5.2 Future work

Incorporating models that can handle email context and email threads within their predictions would be an important step towards better understanding deception.

Another interesting direction would be to test whether toxicity relates to fraud. This would be consistent with the results of our classifier and to some extent with (Eckhaus and Sheaffer, 2018) seeing as toxicity and hubris can be linked.

5.3 Conclusion

We explored the possibility of identifying emails written by a fraudster based solely on their language using the Enron email dataset. We found that politness and planning discourse markers increased in frequency around the collapse, but sentiment did not appear related to deception. Our linear classifer acheieved good F1 score indicating that it is possible to identify POIs even among other executives. More research is required to validate the features identified by our models.

References

Eyal Eckhaus and Zachary Sheaffer. 2018. Managerial hubris detection: the case of enron. *Risk Management*, 20(4):304–325.

Michael C Frank and Noah D Goodman. 2012. Predicting pragmatic reasoning in language games. *Science*, 336(6084):998–998.

Daniel Fried, Jacob Andreas, and Dan Klein. 2017. Unified pragmatic models for generating and following instructions. *arXiv preprint arXiv:1711.04987*.

Caleb Gattegno. 2010. *Teaching foreign languages in schools: The silent way*. Educational Solutions World.

Noah D Goodman and Michael C Frank. 2016. Pragmatic language interpretation as probabilistic inference. *Trends in cognitive sciences*, 20(11):818–829.

Herbert P Grice. 1975. Logic and conversation. In *Speech acts*, pages 41–58. Brill.

Tania Ka'Ai. 2008. *Ngoingoi Pēwhairangi: An Extraordinary Life*. Huia Publishers.

Fereshte Khani, Noah D Goodman, and Percy Liang. 2018. Planning, inference and pragmatics in sequential language games. *Transactions of the Association for Computational Linguistics*, 6:543–555.

Angeliki Lazaridou, Anna Potapenko, and Olivier Tieleman. 2020. Multi-agent communication meets natural language: Synergies between functional and structural language learning. *arXiv preprint arXiv:2005.07064*.

Katerina Mataira. 1980. *The effectiveness of the silent way method in the teaching of Maori as a second language*. Ph.D. thesis, The University of Waikato.

Bethany McLean. 2001. Is enron overpriced?

Vlad Niculae, Srijan Kumar, Jordan Boyd-Graber, and Cristian Danescu-Niculescu-Mizil. 2015. Linguistic harbingers of betrayal: A case study on an online strategy game. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1650–1659, Beijing, China. Association for Computational Linguistics.

Ionut-teodor Sorodoc, Angeliki Lazaridou, Gemma Boleda Torrent, Aurelie Georgette Geraldine Herbelot, Sandro Pezzelle, Raffaella Bernardi, et al. 2016. "look, some green circles!": Learning to quantify from images. In *ACL 2016 5th Workshop on Vision and Language (VL'16): Proceedings of the Workshop*, pages 75–79. Association for Computational Linguistics.

Jacqueline Van Arkel, Marieke Woensdregt, Mark Dingemanse, and Mark Blokpoel. 2020. A simple repair mechanism can alleviate computational demands of pragmatic reasoning: Simulations and complexity analysis. In *the 24th (Virtual) Conference on Computational Natural Language Learning (CoNLL 2020)*, pages 177–194. The Association for Computational Linguistics.

Runzhe Yang, Jingxiao Chen, and Karthik Narasimhan. 2020. Improving dialog systems for negotiation with personality modeling. *arXiv preprint arXiv:2010.09954*.

Zilong Zheng, Shuwen Qiu, Lifeng Fan, Yixin Zhu, and Song-Chun Zhu. 2021. Grice: A grammar-based dataset for recovering implicature and conversational reasoning. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 2074–2085.