Institute of Informatics
Logics and Security Studies

Translating a Cue-Based Dialogue Act Classifier

Nick Webb

October 1, 2009

ILS Tech Report: 008

Available for download from:
`http://www.ils.albany.edu`

1 Task

We have a cue-based Dialogue Act classifier [WHW05a, WHW05b, WL08]. This classifier works by applying a set of automatically extracted cue phrases (single words or short phrases) directly to unseen utterances of dialogue. For detail on our classifier, see Section 2. We have experimented using several hand-labelled dialogue corpora, including the SWITCHBOARD corpus [JBC+97] and the ICSI-MRDA corpus [SDB+04], and have achieved a good level of classification accuracy using a set of cue phrases that overlap between these two corpora.

We find that the total number of overlapping cues phrases between these two corpora is around 33k. Of those, most phrases predict a <STATEMENT> category. The remaining 719 shared cues are spread across the remaining categories. This is a very small number of cue phrases. Using just these cue phrases, we achieve a reasonable level of classification accuracy. Therefore we have made a very strong statement about the power of these 719 phrases.

What we want to know is if the same is true in different languages. The task is to take the 719 English cue phrases, and translate them into a target language (such as Chinese). Using Google Translate [1], create software that can automatically translate English cues into some target language. Then apply these cues as a classifier to chat data in that target language. This may need the student to acquire such a corpus, tag it, and assess tagging accuracy manually.

2 Background

Dialogue Acts are an extension of work in Speech Acts, notably the work of [Aus62] and built upon by [Sea69]. The work of [Aus62], defined Speech Acts to encompass the idea that meaning can be explained by action, rather than in concepts like reference and truth conditions. Often, these action utterances could contain a performative verb, such as 'apologise', or else the performance of an action may remain implicit. A request such as "Can you close the window?", under these conditions, is an utterance through which the speaker is attempting to enact some change in the world, through the actions of the hearer. In many of the examples [Sea69] addressed, the utterances contained cue words or phrases, which could be used to indicate the presence

[1] http://translate.google.com/

of particular Speech Acts. For example, a number of directives were realised by utterances beginning with *wh*-words, or containing subject-aux inversion (*"Can you..."*). The Literal Meaning Hypothesis [Gaz81] is a strong version of this observation, suggesting that every utterance has an illocutionary force built into its surface form, therefore the selection of surface realisation of our utterances is so intentional as to communicate the illocutionary force we wish to display. It indicates that an analysis of utterances that relies on surface lexicalisations should perform well in identifying the underlying Speech Act. Within the field of computational linguistics, recent work, closely linked to the development and deployment of spoken language dialogue systems, has focused on the some of the more conversational roles such acts can perform. For example, if an automatic system or agent asks a question, and a user replies with another question, it is possible for this to be seen, in context, as a clarification of the original question. Less important then are the notions of pragmatic force, and more important is their role in a progression of dialogue and how the recognition of some sort of act can aid in the interpretation of a user's utterance. These acts are called Dialogue Acts [Bun94] or in some literature Conversational or Dialogue Moves [Pow79]. This level of analysis concentrates on how what you say commits you to some action, such as accepting or rejecting some offer. [Lew00] characterises this as a stress on "what you say committing you, rather than your committing to what you say", as in the focus of the early work by [Aus62] and [Sea69]. There is a possible interpretation of the work of [Lew00], in common with the literal meaning hypothesis, that there is much literal meaning in the words chosen to express an utterance that can also be used in interpretation. This stands in contrast to the need to construct a complex mental model of your conversational partner in order to accurately process utterances. Our work explores to what extent you can perform an interpretation of utterances based solely on the words in the utterance (and a few other simple lexical features). This is called the cue-based model.

In this model, the *hearer* uses cues (or simple indicators, either alone or in combination) in the utterances (both individual and in context of the wider dialogue) to decide on an interpretation. This model captures the idea that the surface form of an utterance provides all manner of cues as to what this interpretation could be. Cues can be lexical, collocational, syntactic, prosodic or based on a deeper conversational structure. The cue model relies on ideas similar to the literal meaning hypothesis [Gaz81] we discussed earlier, that is the literal meaning hypothesis is strong prior claim for cue-based modelling to be successful. The key to the cue model from a computational perspective is that these cues can be probabilistically associated with specific Dialogue Acts, and we shall exploit this later.

Cue phrases (also called 'discourse cues', 'discourse markers' or 'clue words') are words and phrases that explicitly signal the structure of a discourse or dialogue, and in turn can be used to determine the intention of the speaker. Phrases include single word cues, such as *"well"* or multiword cues, for example *"in any case"* (or for that matter *"for example"*). Importantly, in most prior work concerning cue phrases once phrases have been identified, either manually or automatically, they are passed to some later process as one of many possible features, that are then used to identify Dialogue Acts (as in the work of [SCVS98]), or many other uses.

2.1 Lexical and Syntactic Cues

Lexical cues (specific words or phrases) are the most widely understood set of cues and have appeared in literature across languages and conversational styles [Coh84, War85, GS86, HL93, MH95]. Syntactic cues (such as specific grammatical constructions) have the benefit that they can be cross-linguistic, like the strong correlation between sentence-initial or sentence-final particles, special verb order, and general word order that can often be seen in <Yes-No Questions> [JM08]. Given a large corpus, [JM00] were able to find many other such syntactic correlations.

When [SCVS99] performed a broader analysis of discourse processing literature, they found a total of 687 different cue phrases listed or mentioned as useful for either intention or structural analysis of discourse and dialogue. Most cue phrases from the literature were recorded as being generally useful for spotting discourse structure or function. [SCVS99] decided to attempt to learn automatically cue phrases that were good indicators of Dialogue Acts, using the annotated verbmobil corpus as source material. [SCVS99] took their resulting ranked list of cue phrases (that is, the cue phrases that remain once the measures have been applied, leaving an ordered list of cue phrases, those which convey the most information first), and subjected the phrases to lexical filtering, removing duplicate or overlapping phrases. The cue phrases were evaluated for effectiveness, by passing them as a feature to a subsequent machine learning method.

2.2 Cue Phrase Selection

We were interested by the work of [SCVS99] automatically identifying potential cue phrases from a corpus. In their experiments, [SCVS99] constructed all n-grams of lengths 1 through 3 from the corpus, and then applied a range of measures which effectively pruned the n-gram list, until only candidate cue phrases remained. In order to test the effectiveness of these automatically acquired cue phrases, [SCVS99] passed them as features to a machine learning method, in their case Transformation Based Learning (TBL).

We began our experiments with an intuition as to what constituted a good cue phrase. We hypothesised that a cue phrase would be a word or phrase in the corpus that would serve as a reliable indicator of a particular Dialogue Act. If we look at cue phrases that regularly co-occur with individual Dialogue Acts, we could see if the presence of that particular phrase was a reliable predictor of that Dialogue Act. Each cue phrase could predict many different Dialogue Acts, but we are interested only in that DA that is maximally predicted by the cue phrase in question.

For example, if the n-gram *"hello there"* occurs in the corpus a total of 100 times, and of those 100, 95 instances occur in utterances that are annotated as <CONVENTIONAL-OPENING>, then we could say that *"hello there"* has a maximal *predictivity* of 95% with <CONVENTIONAL-OPENING>, and we discard any other relationships this particular n-gram has with any other categories (as it is certain to be the highest association).

More formally, we can describe our criteria, predictivity, for selecting cue phrases from the set of all possible cue phrases in the following way. The predictivity of phrase c for DA d is the conditional probability $P(d|c)$, where:

$$P(d|c) = \frac{\#(c\&d)}{\#(c)}$$

We represent the set of all *possible* cue phrases (all n-grams length 1–4 from the corpus) as C, so given $c \in C$: c represents some possible cue phrase. Similarly, D is the set of all dialogue act labels, and $d \in D$: d represents some dialogue act label. Therefore $\#(c)$ is the count of (possible) cue phrase c in corpus, and $\#(c\&d)$ is the count of occurrences of phrase c in utterances with dialogue act d in the training data. The *maximal predictivity* of a cue phrase c, written as $mp(c)$, is defined as:

$$mp(c) = \max_{d \in D} P(d|c)$$

IFor our experiments, the word n-grams used as potential cue phrases during are automatically extracted from training data. All word n-grams of length 1–4 within the data are considered as candidates. The maximal predictivity of each cue phrase can be computed directly from the corpus. We can use this value as one threshold for pruning potential cue phrases from our model. Removing n-grams below some predictivity threshold will improve the compactness of the model produced. For example, taking our earlier example of *"hello there"*, 95 instance of this cue phrase occurred in utterances annotated as <CONVENTIONAL-OPENING>. For the other 5 occurrences, we may as well discard this information, because the resulting predictivity (even if all 5 instances occur in utterances marked with the same DA) can only be 5%. Thus the maximal predictivity for the n-gram *"hello there"* is 95% in this instance.

Another reasonable threshold would appear to be the frequency count of each potential cue phrase. Phrases which have a low frequency score are likely to have very high predictivity scores, possibly skewing the model as a whole. For example, any potential cue phrase which occurs only once will de-facto have a 100% predictivity score. We can use a minimal count value ($t_{\#}$) and minimal predictivity thresholds (t_{mp}) to prune the set C^* of 'useful' cue phrases derived from the training data, as defined by:

$$C^* = \{c \in C \mid mp(c) \geq t_{mp} \wedge \#(c) \geq t_{\#}\}$$

The n-grams that remain after this thresholding process are those we identify as cue phrases. For our initial experiments, we used a predictivity of 30% and a frequency of 2 as our thresholds for cue extraction, and applied the mechanism to the SWITCHBOARD corpus.

3 Cue-Based DA Classification

Having defined our mechanism to extract cue phrases from a corpus, we need some way to evaluate their effectiveness. [SCVS99] passed their cue phrases as a feature to a machine learning method. We chose instead a method where the cue phrases extracted from a corpus could be used *directly* as a method of classification. If our extracted cues are indeed reliable predictors of Dialogue Acts, then a classifier that uses these cues directly should perform reasonably well. If, on the other hand, this mechanism did not work, it would not necessarily mean that our cue phrases are not effective, only that

`Speaker A: DA="yes-no-question":` $\boxed{\textbf{would you ever}}$ **have thought of that**

Example n-gram	*Total Count*	*Predicts* DA *(count)*	*Predictivity*
would	4563	statement-non-opinion (2368)	51.9%
you	38905	statement-non-opinion (16342)	42.0%
would you	157	wh-question (53)	33.8%
would you	157	yes-no-question (40)	25.5%
would you ever	3	yes-no-question (3)	100%
would you ever have	1	yes-no-question (1)	100%

Figure 1: SWITCHBOARD: Example cue-based classifier

we need to pass them to a subsequent machine learning process as others had done. The benefit of our direct classification approach is that it is very fast to apply, and gives us immediate feedback as to the possible effectiveness of our automatically extracted cue phrases.

The predictivity of a cue phrase can be exploited directly in a simple model of Dialogue Act classification. Intuitively, when we see the phrase *"hello there"*, we want to always assign the category <CONVENTIONAL-OPENING> and all other things being equal we'll be correct 95% of the time. To apply this method, we can train a classifier on a part of a corpus, extracted potential cue phrases as described in the previous section. The cue phrases selected using our measure of predictivity are used directly to classify unseen utterances in the following manner. We identify all the potential cue phrases a target utterance contains, and determine which has the highest predictivity of some dialogue act category, then assign that category. Given the notation we define above, we can obtain the DA predicted by a particular cue ($dp(c)$) by:

$$dp(c) = \underset{d \in D}{\mathsf{argmax}}\, P(d|c)$$

In the example shown in Figure 1, the utterance *"would you ever have thought of that"* has been annotated by a human as a <YES-NO-QUESTION>.

In Figure 1, we show the predicitivity scores for a selection of the n-grams in the utterance. We can see the 1-grams, such as *"would"* and *"you"*, have a high *total count*, but equally these phrases are found in utterances marked with a range of different DAs, as indicated by the low predicitivity scores (for all the following examples, only the maximum predicitivity score is shown, unless otherwise indicated). Neither of the 1-grams shown have a maximum predictivity score that would indicate the correct DA for this utterance, and if we used only 1-grams, we would incorrectly assign the category <STATEMENT-NON-OPINION> to this utterance. If we look at a selection of 2-grams, such as *"would you"*, we can see that this n-gram is maximally predictive of the DA <WH-QUESTION>, but at a lower predictivity than the 1-grams we examined earlier, so even using this mix of 1- and 2-grams, we would not assign the correct category. We can see that *"would you"* has a low rate of predictivity for <YES-NO-QUESTION> at a level below our predictivity threshold, and so this n-gram phrase would not be retained by our classifier. It is only when we look at 3-grams and above that we see really useful results. The 3-gram *"would you ever"* is 100% predicitive of the correct DA category. That is, the three times that this 3-gram occurs in this corpus, they are all in utterances annotated by humans as a <YES-NO-QUESTION>. We can also see that the 4-gram *"would you ever have"* has a 100% predictivity of the correct category, but only occurs once in the entire corpus, so would be discarded for falling below our frequency threshold.

If multiple cue phrases share the same maximal predictivity, but predict different categories, we select the DA category for the phrase which has the higher number of occurrences (that is, the n-gram with the highest frequency). If the combination of predicitivity and occurrence count is insufficient to determine a single DA, then a random choice is made amongst the remaining candidate DAs. If $ng(u)$ defines the set of ngrams of length 1..4 in utterance u, and C^*_u is the set of n-grams in the utterance u that are also in the threshold model C^* then C^*_u is defined as:

$$C^*_u = ng(u) \cap C^*$$

Given our thresholds, the $mpu(u)$ (the utterance maximal prediction, or mp value for the highest scoring cue in utterance u) is defined as:

$$mpu(u) = \max_{c \in C^*_u} mp(c)$$

The maximally predictive cues of an utterance ($mpcu(u)$) are:

$$mpcu(u) = \{c \in C_u^* \mid mp(c) = mpu(u)\}$$

Then the maximal cue of utterance ($mcu(u)$), i.e. one of its maximally predictive cues that has a maximal count (from within that set), is:

$$mcu(u) = \underset{c \in mpcu(u)}{\mathsf{argmax}} \; \#(c)$$

Finally, for our classification model, $dpu(u)$ utterance DA prediction — the DA predicted by model for utterance u, is defined as:

$$dpu(u) = dp(mcu(u))$$

If no cue phrases are present in the utterance under consideration, then a default tag is assigned, corresponding to the most frequent tag within the training corpus, which for the SWITCHBOARD corpus is the tag <STATEMENT-NON-OPINION>.

References

[Aus62] J. L. Austin. *How to Do Things with Words*. Oxford University Press, Oxford, 1962.

[Bun94] Harry Bunt. Context and dialogue control. *THINK*, 3:19–31, 1994.

[Coh84] Robin Cohen. A computational theory of the function of clue words in argument understanding. In *Proceedings of the 1984 International Computational Linguistics Conference*, pages 251–255, California, 1984.

[Gaz81] G. Gazdar. Speech Act Assignment. In A. Joshi, B. Webber, and I. Sag, editors, *Elements of Discourse Understanding*, pages 64–83. Cambridge University Press, Cambridge, England, 1981.

[GS86] Barbara Grosz and Candace Sidner. Attention, Intentions, and the Structure of Discourse. *Computational Linguistics*, 19(3), 1986.

[HL93] Julia Hirschberg and Diane Litman. Empirical Studies on the Disambiguation of Cue Phrases. *Computational Linguistics*, 19(3):501–530, 1993.

[JBC+97] D. Jurafsky, R. Bates, N. Coccaro, R. Martin, M. Meteer, K. Ries, E. Shriberg, A. Stolcke, P. Taylor, and C. Van Ess-Dykema. Automatic detection of discourse structure for speech recognition and understanding. In *Proceedings of the 1997 IEEE Workshop on Speech Recognition and Understanding, Santa Barbara.*, 1997.

[JM00] D Jurafsky and J Martin. Speech and language processing: an introduction to natural language processing, computational .... *MIT Press*, Jan 2000.

[JM08] Daniel Jurafsky and James H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics*. Prentice-Hall, 2nd edition, 2008.

[Lew00] I. Lewin. A formal model of conversational game theory. In *Fourth Workshop on the Semantics and Pragmatics of Dialogue: Gotalog 2000*, pages 115–122, 2000.

[MH95] Daniel Marcu and Graeme Hirst. A uniform treatment of pragmatic inferences in simple and complex utterances and sequences of utterances. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, pages 144–150, Boston, MA, 1995.

[Pow79] Richard J. D. Power. The organisation of purposeful dialogues. *Linguistics*, 17:107–152, 1979.

[SCVS98] Ken Samuel, Sandra Carberry, and K. Vijay-Shanker. Dialogue act tagging with transformation-based learning. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, Montreal, 1998.

[SCVS99] Ken Samuel, Sandra Carberry, and K. Vijay-Shanker. Automatically selecting useful phrases for dialogue act tagging. In *Proceedings of the Fourth Conference of the Pacific Association for Computational Linguistics, Waterloo, Ontario, Canada*, 1999.

[SDB+04] E. Shriberg, R. Dhillon, S. Bhagat, J. Ang, and H. Carvey. The ICSI meeting recorder dialog act (MRDA) corpus. In *Special Interest Group on Discourse and Dialogue (SIGdial)*, Boston, USA, 2004.

[Sea69] J. R. Searle. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, Cambridge, 1969.

[War85] Richard G. Warner. *Discourse Connectives in English*. Garland Publishing, Inc. , 1985.

[WHW05a] Nick Webb, Mark Hepple, and Yorick Wilks. Dialogue Act Classification Based on Intra-Utterance Features. In *Proceedings of the AAAI Workshop on Spoken Language Understanding*, 2005.

[WHW05b] Nick Webb, Mark Hepple, and Yorick Wilks. Empirical determination of thresholds for optimal dialogue act classification. In *Proceedings of the Ninth Workshop on the Semantics and Pragmatics of Dialogue*, 2005.

[WL08] N. Webb and T. Liu. Investigating the Portability of Corpus-Derived Cue Phrases for Dialogue Act Classification. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008)*, Manchester, United Kingdom, 2008.

A Cue phrases shared between the SWITCHBOARD and ICSI-MRDA corpora, listed by DA label

AGREE-ACCEPT			
Cue Phrase	*Utterance Length*	*Frequency*	*Predictivity*
think so too	Long	4	100%
so too <finish>	Long	4	100%
think so too <finish>	Long	3	100%
<start> yes	Long	3	58.9%
agree with that <finish>	Long	4	57.1%
i agree with that	Long	3	50%
that's true <finish>	Long	6	46.2%
agree with that	Long	4	40%
that is true <finish>	Medium	4	100%
that is true	Medium	4	100%
<start> i i agree	Medium	3	100%
<start> that is true	Medium	3	100%
absolutely right	Medium	3	100%
absolutely right <finish>	Medium	3	100%
is true <finish>	Medium	7	87.5%
is true	Medium	7	87.5%
absolutely <finish>	Medium	2	84.6%
<start> that's true	Medium	5	81.7%
<start> that's true <finish>	Medium	5	81.5%
<start> absolutely	Medium	4	80%
<start> well that's true	Medium	4	80%
exactly <finish>	Medium	5	77.1%
that's true	Medium	1	76.9%
absolutely	Medium	3	76.9%
that's true <finish>	Medium	9	76.4%
exactly	Medium	6	72.3%
true <finish>	Medium	1	71%
true	Medium	1	70.3%
<start> that's right <finish>	Medium	1	67.1%
well that's true	Medium	4	66.7%
well that's true <finish>	Medium	4	66.7%
<start> that's right	Medium	1	66.5%
i agree <finish>	Medium	4	64.5%
<start> exactly	Medium	9	64.3%
that's right <finish>	Medium	1	63.8%
that's right	Medium	1	62.9%
<start> you're right	Medium	5	62.5%
<start> you're right <finish>	Medium	5	62.5%
<start> sure	Medium	1	61.5%
<start> i agree <finish>	Medium	2	61%
agree <finish>	Medium	4	60%
i i agree	Medium	3	60%
i i agree <finish>	Medium	3	60%
i agree	Medium	4	58.6%
yep	Medium	4	55.6%
agree	Medium	4	55%

Cue Phrase	*Utterance Length*	*Frequency*	*Predictivity*
<start> i agree	Medium	2	53.1%
you're right <finish>	Medium	1	52.2%
you're right	Medium	1	52.2%
definitely <finish>	Medium	7	50%
oh yes <finish>	Medium	9	47.4%
<start> oh yes <finish>	Medium	8	44.4%
me too	Medium	4	44.4%
it's true	Medium	4	44.4%
<start> yeah you	Medium	4	44.4%
it's true <finish>	Medium	4	44.4%
me too <finish>	Medium	4	44.4%
<start> it is <finish>	Medium	3	42.9%
it does <finish>	Medium	5	41.7%
oh yes	Medium	9	40.9%
definitely	Medium	1	40%
yeah you	Medium	4	40%
of course <finish>	Medium	9	39.1%
course <finish>	Medium	9	39.1%
<start> oh yes	Medium	8	38.1%
course	Medium	9	36%
of course	Medium	9	36%
yes <finish>	Medium	2	34.4%
right	Medium	3	33.6%
it does	Medium	6	33.3%
<start> of course	Medium	5	33.3%
<start> of course <finish>	Medium	5	33.3%
right <finish>	Medium	2	33.2%
<start> exactly <finish>	Short	9	84.1%
exactly <finish>	Short	9	84.1%
exactly	Short	9	84.1%
<start> exactly	Short	9	84.1%
<start> true <finish>	Short	9	81.8%
<start> true	Short	9	81.8%
true	Short	9	81.8%
true <finish>	Short	9	81.8%
<start> definitely	Short	1	76.9%
definitely	Short	1	76.9%
definitely <finish>	Short	1	76.9%
<start> definitely <finish>	Short	1	76.9%
absolutely	Short	2	68.3%
<start> absolutely	Short	2	68.3%
absolutely <finish>	Short	2	68.3%
<start> absolutely <finish>	Short	2	68.3%
<start> yes <finish>	Short	1	60.7%
yes <finish>	Short	1	60.7%
yes	Short	1	60.7%
<start> yes	Short	1	60.7%

MAYBE

Cue Phrase	*Utterance Length*	*Frequency*	*Predictivity*
could be <finish>	Medium	2	64.5%
could be	Medium	2	52.5%
maybe that's	Medium	3	42.9%
maybe	Short	4	90%
<start> maybe <finish>	Short	4	90%
<start> maybe	Short	4	90%
maybe <finish>	Short	4	90%

REJECT

Cue Phrase	*Utterance Length*	*Frequency*	*Predictivity*
<start > no <finish>	Medium	5	100%
no <finish>	Medium	5	100%
actually no <finish>	Medium	3	100%
no	Medium	8	88.9%
<start > no	Medium	8	88.9%
well no <finish>	Medium	7	87.5%
<start >well no <finish>	Medium	6	85.7%
well no	Medium	7	77.8%
<start >well no	Medium	6	75%
actually no	Medium	3	75%
no <finish>	Medium	1	55.5%
uh no	Medium	1	37.1%
<start >uh no	Medium	1	35.3%
uh no <finish>	Medium	1	34.5%
<start > nope	Short	1	85.7%
nope	Short	1	85.7%
nope <finish>	Short	1	85.7%
<start > nope <finish>	Short	1	85.7%
no	Short	3	68.2%
<start > no	Short	3	68.2%
no <finish>	Short	3	68.2%
<start > no <finish>	Short	3	68.2%

BACK-CHANNEL			
Cue Phrase	*Utterance Length*	*Frequency*	*Predictivity*
<start> uhhuh uhhuh <finish>	Medium	2	83.3%
uhhuh uhhuh <finish>	Medium	2	81.8%
<start> uhhuh uhhuh	Medium	2	81.2%
uhhuh uhhuh	Medium	2	77.8%
<start> uhhuh	Medium	6	56.8%
uhhuh <finish>	Medium	4	56.5%
uhhuh	Medium	1	55.1%
<start> uhhuh yeah <finish>	Medium	1	52%
uhhuh yeah <finish>	Medium	1	51.9%
huh	Medium	1	50.6%
yeah	Medium	3	50%
<start> uhhuh yeah	Medium	1	48.3%
uhhuh yeah	Medium	1	46.9%
<start> right yeah <finish>	Medium	5	38.5%
<start> yeah	Medium	7	36.8%
<start> right yeah	Medium	5	33.3%
right yeah <finish>	Medium	5	33.3%
uhhuh	Short	4	90.8%
<start> uhhuh <finish>	Short	4	90.8%
<start> uhhuh	Short	4	90.8%
uhhuh <finish>	Short	4	90.8%
huh	Short	1	86%
huh <finish>	Short	1	86%
<start> huh	Short	1	86%
<start> huh <finish>	Short	1	86%
yeah <finish>	Short	5	59%
<start> yeah	Short	5	59%
yeah	Short	5	59%
<start> yeah <finish>	Short	5	59%
<start> yep <finish>	Short	2	47.7%
yep	Short	2	47.7%
yep <finish>	Short	2	47.7%
<start> yep	Short	2	47.7%
<start> right <finish>	Short	1	37.2%
right	Short	1	37.2%
right <finish>	Short	1	37.2%
<start> right	Short	1	37.2%

APPRECIATION			
Cue Phrase	*Utterance Length*	*Frequency*	*Predictivity*
well that's a good	Long	4	100%
<start> wow that's	Long	3	100%
wow that's	Long	3	100%
well that's a	Long	4	80%
<start> well that's a	Long	3	75%
that's a good idea	Long	1	66.7%
<start> wow	Long	4	57.1%
that's that's a good	Long	5	55.6%
good point <finish>	Long	8	53.3%
sounds good	Long	3	50%
good to know	Long	4	40%
like a good	Long	4	40%
that's great	Long	5	33.3%
<start> good point <finish>	Medium	1	100%
a good point	Medium	1	100%
that's a good point	Medium	1	100%
a good point <finish>	Medium	1	100%
<start> that's cool <finish>	Medium	8	100%
<start> that's cool	Medium	8	100%
that's a great	Medium	7	100%
<start> that's a great	Medium	7	100%
that's amazing <finish>	Medium	6	100%
<start> good idea	Medium	6	100%
that sounds good	Medium	6	100%
that sounds good <finish>	Medium	6	100%
<start> good idea <finish>	Medium	6	100%
how interesting <finish>	Medium	6	100%
that's amazing	Medium	6	100%
how interesting	Medium	6	100%
<start> that's amazing	Medium	5	100%
<start> that sounds good	Medium	5	100%
<start> that's amazing <finish>	Medium	5	100%
that's wonderful	Medium	4	100%
fascinating	Medium	4	100%
fascinating <finish>	Medium	4	100%
<start> how interesting	Medium	4	100%
<start> how interesting <finish>	Medium	4	100%
that's wonderful <finish>	Medium	4	100%
that's nice <finish>	Medium	4	100%
<start> wow that's	Medium	3	100%
<start> that's wonderful	Medium	3	100%
<start> what a	Medium	3	100%
what a	Medium	3	100%
<start> that's nice	Medium	3	100%
<start> well that's interesting	Medium	3	100%
well that's interesting <finish>	Medium	3	100%
well that's good <finish>	Medium	3	100%

Cue Phrase	Utterance Length	Frequency	Predictivity
<start> that's wonderful <finish>	Medium	3	100%
well that's interesting	Medium	3	100%
wow that's	Medium	3	100%
that's not bad	Medium	3	100%
that's not bad <finish>	Medium	3	100%
<start> well that's good	Medium	3	100%
well that's good	Medium	3	100%
<start> that's nice <finish>	Medium	3	100%
<start> that's a good	Medium	3	97.5%
that's a good	Medium	3	97.5%
a good idea <finish>	Medium	2	95.8%
a good idea	Medium	2	95.8%
that's a good idea	Medium	1	94.7%
good point <finish>	Medium	3	94.6%
<start> that's great	Medium	3	94.4%
<start> good point	Medium	1	94.4%
<start> that's great <finish>	Medium	3	94.3%
<start> that's good <finish>	Medium	3	94.1%
good idea	Medium	3	93.8%
good idea <finish>	Medium	3	93.8%
<start> that's interesting	Medium	2	93.1%
<start> that's interesting <finish>	Medium	2	92.9%
good point	Medium	3	92.3%
that's interesting <finish>	Medium	4	92.2%
that's great	Medium	4	91.8%
that's great <finish>	Medium	4	91.7%
that's interesting	Medium	4	90.6%
<start> that's good	Medium	3	90%
that would be great	Medium	8	88.9%
that's cool	Medium	8	88.9%
that's cool <finish>	Medium	8	88.9%
a good	Medium	4	88.2%
very interesting <finish>	Medium	7	87.5%
<start> wow	Medium	7	87.5%
very interesting	Medium	7	87.5%
very good	Medium	7	87.5%
<start> that sounds	Medium	7	87.5%
cool <finish>	Medium	2	83.9%
very good <finish>	Medium	5	83.3%
interesting <finish>	Medium	1	82.1%
amazing <finish>	Medium	9	81.8%
amazing	Medium	9	81.8%
be great	Medium	2	81.5%
be great <finish>	Medium	2	81.5%
cool	Medium	3	81.1%
interesting	Medium	1	80.9%
sounds good <finish>	Medium	1	80%
sounds good	Medium	1	80%

Cue Phrase	*Utterance Length*	*Frequency*	*Predictivity*
neat <finish>	Medium	8	80%
would be great <finish>	Medium	8	80%
would be great	Medium	8	80%
<start> that's funny <finish>	Medium	4	80%
<start> that's funny	Medium	4	80%
that's nice	Medium	4	80%
that's good <finish>	Medium	5	79.7%
that's good	Medium	5	78.6%
good question	Medium	7	77.8%
good question <finish>	Medium	7	77.8%
<start> that would be	Medium	2	76.9%
that sounds	Medium	1	76.9%
that's funny	Medium	6	75%
that's funny <finish>	Medium	6	75%
<start> that's very	Medium	6	75%
that's so	Medium	3	75%
exciting	Medium	3	75%
really neat <finish>	Medium	3	75%
<start> that's kind	Medium	3	75%
exciting <finish>	Medium	3	75%
really neat	Medium	3	75%
<start> that's kind of	Medium	3	75%
great	Medium	1	74.3%
nice <finish>	Medium	1	73.1%
<start> good	Medium	4	72.9%
great <finish>	Medium	1	72.8%
neat	Medium	8	72.7%
wow	Medium	2	71.8%
wow <finish>	Medium	2	70%
wonderful <finish>	Medium	9	69.2%
<start> that's really	Medium	9	69.2%
that would be	Medium	2	69%
<start> that's a	Medium	5	68%
be good	Medium	1	66.7%
be good <finish>	Medium	1	66.7%
good	Medium	2	65.7%
<start> that'd	Medium	2	65.6%
<start> that'd be	Medium	2	65.6%
that's really	Medium	1	65.2%
nice	Medium	2	64.3%
<start> sounds	Medium	7	63.6%
idea <finish>	Medium	5	62.9%
that'd be	Medium	2	62.5%
weird <finish>	Medium	5	62.5%
that's a	Medium	5	61.9%
that'd	Medium	2	61%
funny <finish>	Medium	1	60.9%
good <finish>	Medium	1	60.8%

Cue Phrase	*Utterance Length*	*Frequency*	*Predictivity*
wonderful	Medium	9	60%
be fun <finish>	Medium	3	60%
be fun	Medium	3	60%
idea	Medium	5	59.6%
funny	Medium	1	56%
sounds	Medium	2	55.3%
point <finish>	Medium	3	52.1%
not bad	Medium	5	50%
not bad <finish>	Medium	5	50%
god <finish>	Medium	5	50%
<start> that's pretty	Medium	3	50%
deal <finish>	Medium	3	50%
god	Medium	6	42.9%
deal	Medium	3	42.9%
helpful	Medium	3	42.9%
helpful <finish>	Medium	3	42.9%
pretty good	Medium	3	37.5%
pretty good <finish>	Medium	3	37.5%
oh that's	Medium	1	35.8%
<start> oh that's	Medium	1	35.3%
<start> oh my	Medium	5	33.3%
oh my	Medium	5	33.3%
that's pretty	Medium	4	33.3%
is good <finish>	Medium	3	33.3%
hline *<start> nice*	Short	7	100%
nice	Short	7	100%
<start> nice <finish>	Short	7	100%
nice <finish>	Short	7	100%
interesting	Short	3	84.4%
<start> interesting <finish>	Short	3	84.4%
interesting <finish>	Short	3	84.4%
<start> interesting	Short	3	84.4%
man	Short	3	75%
<start> man <finish>	Short	3	75%
man <finish>	Short	3	75%
<start> man	Short	3	75%
great <finish>	Short	6	62.3%
<start> great	Short	6	62.3%
<start> great <finish>	Short	6	62.3%
great	Short	6	62.3%
<start> cool	Short	3	60.4%
<start> cool <finish>	Short	3	60.4%
cool <finish>	Short	3	60.4%

Cue Phrase	*Utterance Length*	*Frequency*	*Predictivity*
cool	Short	3	60.4%
wonderful <finish>	Short	6	60%
wonderful	Short	6	60%
<start> wonderful	Short	6	60%
<start> wonderful <finish>	Short	6	60%
<start> wow <finish>	Short	5	53.2%
wow	Short	5	53.2%
wow <finish>	Short	5	53.2%
<start> wow	Short	5	53.2%
good <finish>	Short	7	51.1%
<start> good	Short	7	51.1%
<start> good <finish>	Short	7	51.1%
good	Short	7	51.1%

BACKCHANNEL-IN-QUESTION-FORM

Cue Phrase	*Utterance Length*	*Frequency*	*Predictivity*
<start> oh really	Medium	2	78.1%
<start> oh really <finish>	Medium	2	78.1%
oh really	Medium	2	78.1%
oh really <finish>	Medium	2	78.1%
<start> isn't that	Medium	3	42.9%
isn't that	Medium	3	37.5%
really	Short	4	80%
<start> really <finish>	Short	4	80%
<start> really	Short	4	80%
really <finish>	Short	4	80%

RESPONSE ACKNOWLEDGEMENT			
Cue Phrase	*Utterance Length*	*Frequency*	*Predictivity*
<start> oh okay	Medium	2	96.6%
oh okay	Medium	2	96.5%
<start> oh okay <finish>	Medium	2	96.5%
oh okay <finish>	Medium	2	96.3%
i see <finish>	Medium	3	94.8%
i see	Medium	3	94.2%
<start> i see <finish>	Medium	2	93.5%
<start> i see	Medium	2	92.6%

SIGNAL NON UNDERSTANDING			
Cue Phrase	*Utterance Length*	*Frequency*	*Predictivity*
the what	Medium	3	100%
pardon	Medium	3	100%
<start> what's that <finish>	Medium	4	50%
<start> excuse me <finish>	Medium	9	47.4%
what's that <finish>	Medium	4	44.4%
<start> what's that	Medium	4	44.4%
excuse me <finish>	Medium	1	43.5%
<start> excuse	Medium	9	42.9%
<start> excuse me	Medium	9	42.9%
excuse me	Medium	1	40%
what's that	Medium	4	33.3%
<start> pardon <finish>	Short	3	100%
pardon	Short	3	100%
pardon <finish>	Short	3	100%
<start> pardon	Short	3	100%
<start> what <finish>	Short	3	40.7%
what <finish>	Short	3	40.7%
what	Short	3	40.7%
<start> what	Short	3	40.7%

OFFER			
Cue Phrase	*Utterance Length*	*Frequency*	*Predictivity*
<start> why don't	Long	1	67.9%
<start> put	Long	6	54.5%
<start> why don't you	Long	6	50%
why don't	Long	4	47.4%
<start> you should	Long	3	37.5%
why don't you	Long	1	35.9%

APOLOGY			
Cue Phrase	*Utterance Length*	*Frequency*	*Predictivity*
sorry i	Long	1	66.7%
<start> i'm sorry	Long	6	60%
i'm sorry i	Long	5	55.6%
sorry about	Medium	3	100%
<start> sorry	Medium	2	95.5%
<start> i'm sorry	Medium	4	80.7%
<start> i'm sorry <finish>	Medium	3	77.6%
sorry	Medium	1	72.2%
i'm sorry	Medium	6	71.6%
i'm sorry <finish>	Medium	5	68.8%
sorry <finish>	Medium	8	67.2%
excuse me <finish>	Medium	9	39.1%
<start> excuse me <finish>	Medium	7	36.8%
sorry	Short	8	86.6%
sorry <finish>	Short	8	86.6%
<start> sorry <finish>	Short	8	86.6%
<start> sorry	Short	8	86.6%

THANKING

Cue Phrase	*Utterance Length*	*Frequency*	*Predictivity*
thanks for	Long	3	75%
<start> thank you <finish>	Medium	3	100%
<start> thank	Medium	4	97.6%
<start> thank you	Medium	4	97.6%
thank	Medium	4	83.3%
thank you	Medium	4	83%
thank you <finish>	Medium	3	81.4%
<start> thanks	Short	2	100%
thanks <finish>	Short	2	100%
thanks	Short	2	100%
<start> thanks <finish>	Short	2	100%

AFFIRMATIVE NON YES ANSWERS

Cue Phrase	*Utterance Length*	*Frequency*	*Predictivity*
<start> i did <finish>	Medium	4	80%

OTHER ANSWERS

Cue Phrase	*Utterance Length*	*Frequency*	*Predictivity*
i have no idea	Medium	8	100%
have no idea	Medium	8	100%
have no idea <finish>	Medium	8	100%
i have no	Medium	1	90.9%
<start> i have no	Medium	1	90.9%

RHETORICAL QUESTIONS

Cue Phrase	*Utterance Length*	*Frequency*	*Predictivity*
how can	Long	6	40%
who knows	Medium	3	50%
who knows <finish>	Medium	3	50%
why not <finish>	Medium	3	33.3%

OPEN QUESTIONS			
Cue Phrase	*Utterance Length*	*Frequency*	*Predictivity*
what about the	Long	3	33.3%

OR CLAUSES			
Cue Phrase	*Utterance Length*	*Frequency*	*Predictivity*
or is it just	Long	4	100%
<start> or are you	Long	3	100%
<start> or are	Long	1	90.9%
<start> or do you	Long	1	83.3%
<start> or should	Long	6	75%
or are you	Long	3	75%
<start> or do	Long	1	73.7%
<start> or is that	Long	7	70%
<start> or is it	Long	9	64.3%
<start> or is	Long	2	63.6%
or is that	Long	8	53.3%
or do you	Long	1	52.6%
or are	Long	1	45.8%
or should	Long	7	38.9%
or is	Long	2	38.6%
or are they	Long	3	37.5%
or is it	Long	1	37.1%
or do	Long	1	36.8%
<start> or was	Medium	5	100%
or was	Medium	5	100%
<start> or are	Medium	3	100%
or are	Medium	3	100%
or is it <finish>	Medium	4	80%
or is it	Medium	6	75%
<start> or is it	Medium	6	75%
or is	Medium	1	65%
<start> or is	Medium	1	65%

WH-QUESTIONS			
Cue Phrase	*Utterance Length*	*Frequency*	*Predictivity*
long does	Long	1	100%
how long does	Long	1	100%
does it take	Long	9	100%
long does it	Long	9	100%
long does it take	Long	9	100%
how long does it	Long	9	100%
<start> how long does	Long	5	100%
<start> when is	Long	5	100%
why do you	Long	5	100%
<start> and what are	Long	4	100%
<start> how does	Long	3	100%
how is the	Long	3	100%
how big is	Long	3	100%
so what was	Long	3	100%
<start> how is	Long	3	100%
big is	Long	3	100%
<start> how often	Long	3	100%
<start> so what was	Long	3	100%
<start> what what are	Long	3	100%
<start> what does	Long	8	88.9%
what are they	Long	7	87.5%
when is	Long	6	85.7%
<start> how many	Long	1	84.6%
<start> how much	Long	1	82.4%
<start> what kind	Long	9	81.8%
<start> what kind of	Long	8	80%
<start> what did	Long	4	80%
<start> and how do	Long	4	80%
it take	Long	1	78.6%
and how do you	Long	3	75%
they doing	Long	3	75%
how much do	Long	3	75%
<start> where where	Long	3	75%
what's your	Long	3	75%
<start> how long	Long	8	72.7%
how is	Long	1	71.4%
<start> how do you	Long	1	71.4%
and what are	Long	5	71.4%
<start> what was the	Long	5	71.4%
<start> so how do	Long	5	71.4%
<start> what was	Long	7	70%
<start> what do you	Long	1	68.8%
<start> so what's	Long	1	68.8%
<start> what are	Long	1	66.7%
<start> what do	Long	1	66.7%
<start> what's the	Long	1	66.7%
<start> but how	Long	8	66.7%

Cue Phrase	*Utterance Length*	*Frequency*	*Predictivity*
<start> how how	Long	6	66.7%
what do you mean	Long	1	65%
how did	Long	7	63.6%
<start> and what's	Long	5	62.5%
where does	Long	5	62.5%
so what's the	Long	5	62.5%
what's what	Long	5	62.5%
<start> so what's the	Long	5	62.5%
and how do	Long	5	62.5%
so how do you	Long	5	62.5%
so what's	Long	1	60%
would you do	Long	3	60%
what's it	Long	3	60%
<start> so how	Long	1	59.1%
<start> so what are	Long	4	57.1%
<start> how	Long	9	56.5%
why would	Long	6	54.5%
how does	Long	1	52.6%
what does	Long	2	52.5%
what did	Long	1	52.4%
<start> how do	Long	1	51.7%
<start> what what	Long	2	51.3%
what are you	Long	9	50%
how often	Long	6	50%
what what are	Long	5	50%
what did you	Long	5	50%
so how do	Long	5	50%
what would you	Long	4	50%
how much is	Long	4	50%
what sort	Long	3	50%
what what do you	Long	3	50%
what does that mean	Long	3	50%
what sort of	Long	3	50%
<start> why do	Long	3	50%
<start> how did	Long	3	50%
when do	Long	3	50%
what do you	Long	2	49.2%
what was the	Long	1	48.1%
<start> what is	Long	9	47.4%
<start> what's	Long	1	45.2%
<start> who	Long	9	45%
you doing	Long	4	44.4%
what does that	Long	4	44.4%
what are	Long	5	43.7%
do you mean	Long	1	42.9%
<start> well how	Long	3	42.9%
how would	Long	1	41.7%
so what are	Long	5	41.7%

Cue Phrase	*Utterance Length*	*Frequency*	*Predictivity*
what's the	Long	3	39.8%
what what is	Long	5	38.5%
what do	Long	3	38.4%
so how	Long	1	37.8%
what's a	Long	3	37.5%
<start> wha	Long	3	37.5%
<start> what	Long	1	36.6%
<start> what if	Long	8	36.4%
what's what's	Long	7	35%
what is the	Long	2	33.9%
how do we	Long	1	33.3%
would it be	Long	4	33.3%
<start> so where	Long	4	33.3%
where did	Long	3	33.3%
what's your	Medium	4	100%
how does	Medium	3	100%
what are	Medium	1	80%
where's	Medium	4	80%
<start> what are	Medium	1	78.6%
<start> what do you	Medium	6	75%
what do you	Medium	6	75%
so what's	Medium	3	75%
<start> where's	Medium	3	75%
what's the	Medium	6	66.7%
how many	Medium	6	66.7%
<start> how many	Medium	4	66.7%
<start> what's the	Medium	4	66.7%
what about <finish>	Medium	4	66.7%
<start> what do	Medium	7	63.6%
what do	Medium	8	61.5%
<start> what's	Medium	2	60%
what's	Medium	3	57.1%
<start> what what	Medium	4	57.1%
what's that	Medium	6	50%
where is	Medium	4	50%
<start> where is	Medium	3	50%
what is	Medium	1	48.1%
<start> what	Medium	7	47.7%
<start> what was	Medium	8	44.4%
what is it <finish>	Medium	3	42.9%
<start> who's	Medium	3	42.9%
<start> what is	Medium	7	41.2%
what was	Medium	9	37.5%
what is it	Medium	3	37.5%
who's	Medium	3	37.5%
what's that <finish>	Medium	3	33.3%
<start> what's that	Medium	3	33.3%
<start> why <finish>	Short	1	68.4%

YES-NO QUESTIONS			
Cue Phrase	*Utterance Length*	*Frequency*	*Predictivity*
<start> are there	Long	4	100%
and do you have	Long	3	100%
<start> have you	Long	6	85.7%
i mean do	Long	4	80%
i mean do you	Long	4	80%
<start> is it the	Long	4	80%
<start> but are	Long	4	80%
there any	Long	9	75%
is there any	Long	6	75%
<start> do you think	Long	6	75%
<start> do you know	Long	6	75%
<start> and do you	Long	3	75%
<start> was it	Long	3	75%
<start> i mean do	Long	3	75%
you think that would	Long	3	75%
<start> is there	Long	1	70.6%
do you know what	Long	7	70%
<start> do they	Long	4	66.7%
mean do you	Long	4	66.7%
and do you	Long	4	66.7%
did you have	Long	4	66.7%
mean do	Long	4	66.7%
<start> is there a	Long	4	66.7%
<start> do we	Long	8	61.5%
<start> do you have	Long	1	61.1%
<start> but do you	Long	3	60%
but do you	Long	3	60%
do you have a	Long	3	60%
<start> is that the	Long	3	60%
<start> does that	Long	3	60%
do you see	Long	3	60%
<start> does	Long	2	57.1%
is it like	Long	4	57.1%
do they have	Long	4	57.1%
<start> is it	Long	2	56.8%
<start> should we	Long	5	50%
do you think that	Long	4	50%
<start> so do you	Long	3	50%
<start> would you	Long	3	50%
do you know	Long	1	48%
<start> do you	Long	3	47.1%
<start> should	Long	1	45.5%
<start> do	Long	5	45.2%
<start> isn't	Long	4	44.4%
but are	Long	4	44.4%
<start> did	Long	1	42.9%
<start> but do	Long	3	42.9%

Cue Phrase	*Utterance Length*	*Frequency*	*Predictivity*
do you do you	Long	3	42.9%
do you have	Long	1	42.5%
<start> so are	Long	5	41.7%
so are	Long	6	40%
<start> is this	Long	6	37.5%
is it the	Long	6	37.5%
so do you	Long	3	37.5%
<start> has	Long	3	37.5%
<start> did you	Long	4	36.4%
<start> is the	Long	4	36.4%
<start> so do	Long	4	36.4%
<start> are	Long	2	36.1%
<start> are we	Long	5	35.7%
<start> would	Long	8	33.3%
<start> does it	Long	3	33.3%
but do	Long	3	33.3%
is that what	Medium	3	100%
ever <finish>	Medium	3	100%
<start> is that what	Medium	3	100%
<start> is that <finish>	Medium	1	94.1%
<start> do you	Medium	2	85.2%
<start> do you <finish>	Medium	5	83.3%
<start> did you	Medium	5	83.3%
that what	Medium	4	80%
<start> is there	Medium	4	80%
<start> do you know	Medium	4	80%
<start> are you	Medium	7	77.8%
<start> do they	Medium	3	75%
<start> is is	Medium	5	71.4%
have you	Medium	5	71.4%
do you know	Medium	4	66.7%
<start> was it	Medium	4	66.7%
<start> did	Medium	1	61.1%
<start> wasn't	Medium	3	60%
do you know <finish>	Medium	3	60%
<start> have you	Medium	3	60%
have you <finish>	Medium	3	60%
<start> can you	Medium	3	60%
<start> is that	Medium	3	57.7%
are you	Medium	1	55.6%

Cue Phrase	*Utterance Length*	*Frequency*	*Predictivity*
did you	Medium	6	54.5%
<start> do	Medium	3	53.4%
<start> can	Medium	1	52.6%
<start> is it	Medium	2	51.3%
<start> are	Medium	1	50%
<start> was	Medium	7	50%
is is	Medium	5	50%
is there	Medium	5	50%
can you	Medium	4	50%
do you	Medium	2	49%
do you <finish>	Medium	7	46.7%
<start> are they	Medium	4	44.4%
<start> is	Medium	6	43.3%
do they	Medium	3	42.9%