

TOPIC CHANGE AND LOCAL PERPLEXITY IN SPOKEN LEGAL DIALOGUE

P.E.Kenne and Mary O’Kane

The University of Adelaide
South Australia 5005
Australia
{pek,mok}@dvcr.adelaide.edu.au

ABSTRACT

The notion of local perplexity of a corpus (over a range of language models) is examined, and the local behaviour of perplexity (over several court transcripts) is examined to determine if there is any correspondence between local variations in perplexity and some event in the dialogue, such as topic change.

1. INTRODUCTION

For a given language model and corpus

$$w_1 w_2 \dots w_N$$

the perplexity PP of this corpus with respect to the language model is defined as

$$PP = \exp\left(-\frac{1}{N} \Pr(w_1 w_2 \dots w_N)\right).$$

The perplexity may be thought of as approximately the average branching factor in the language model. Most of the literature reports a single value of PP for a language model/corpus combination. However, there is no reason why a number of values of perplexity (corresponding to different parts of the corpus) may not be calculated.

2. TRAINING AND TEST DATA and LANGUAGE MODELS

The transcripts from two Australian court cases were used for training language models. These transcripts do not include non-speech events such as "um" etc, but they do include repetitions, false starts etc. Table 1 gives details of these data sets.

	c1	c2
Training size (words)	180K	250K
Test size (words)	20K	38K
% Coverage	85	95

Table 1: Training/test set details

The training sets c1 and c2 are the first fourteen days and first twenty days of two different trials (from different jurisdictions). The test sets are the next two and three days of transcript.

Several language models were tested: word bigram, class bigram and word-phrase bigram [4], all using a linear back-off strategy [2]. For the work described here, we combined both training sets in building the different language models.

3. THE EXPERIMENTS

A number of approaches are available for calculating local perplexity. The first approach involves using a moving, fixed-size window, where the window size and overlap (both in words) could be varied to observe the effect of different size windows. The perplexity is calculated for each window. Figure 1 shows the local perplexity for the initial part of the test set of c1 using a bigram model using window sizes of 100, 400 and 600 words with no overlap. (The corpus perplexity with respect to the bigram language model is 19.)

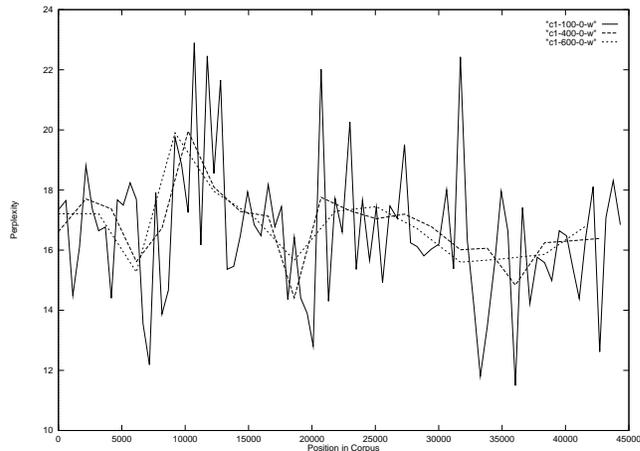


Figure 1: The effect of window size on local perplexity.

The effect of window overlap is illustrated in figure 2. This figure shows the local perplexity of a segment of the same corpus as figure 1. The selected region (windows 600 to 800)

is of interest as this represents the part of the trial in which witnesses are first called. (The earlier parts of the case only have interactions between the judge and several lawyers, and the lawyers' opening statements.) We have fixed a window size of 600 words and the figure shows the local perplexity obtained with no overlap and that obtained by using 90% overlap of the windows.

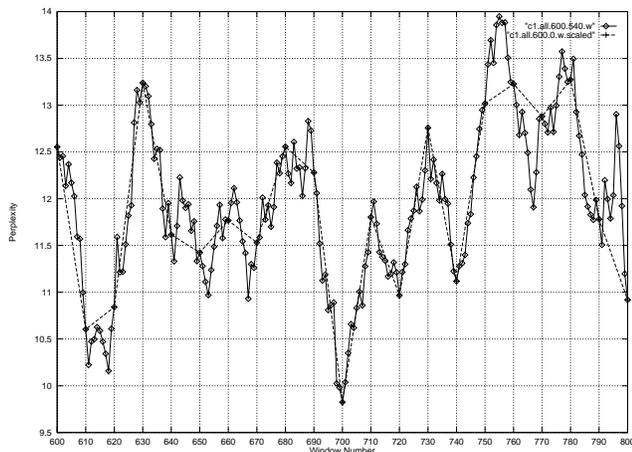


Figure 2: 0% and 90% Window Overlap.

Another approach to calculating local perplexity is to use window size of a number of speaker turns, either a fixed number of turns, or the number of turns determined by a requirement on minimum number of words. The major difference between this approach and that described above is that a window is now a variable size, but a window boundary is on a speaker change. It is possible to obtain windows of very different sizes. A further restriction on the window size (by imposing a maximum size as well) is possible. This approach was not investigated, and all results below used a fixed window size of 600 words, with 90% overlap.

Figure 3 shows part of the transcript near the beginning of window 700 (the location of a local minimum is local perplexity in figure 2). In this transcript the responses of the witness are preceded by <. For a further characterization of Australian court transcripts see Kenne *et al* [3].

Does the type of language model affect the local perplexity? We trained three class bigram models and compared the local perplexities obtained from each. Figure 4 shows the results for these models. (The window size is 600 and there is a 90% overlap.)

The solid line in figure 4 is for a language model where the classes were obtained by part-of-speech tagging using a tagger developed by Brill [1]. The language models for the two dashed lines were obtained from the part-of-speech tagger and arbitrarily combining a number of classes.

Is there any correspondence between utterance length, local perplexity and topic change? Figure 5 shows the utterance

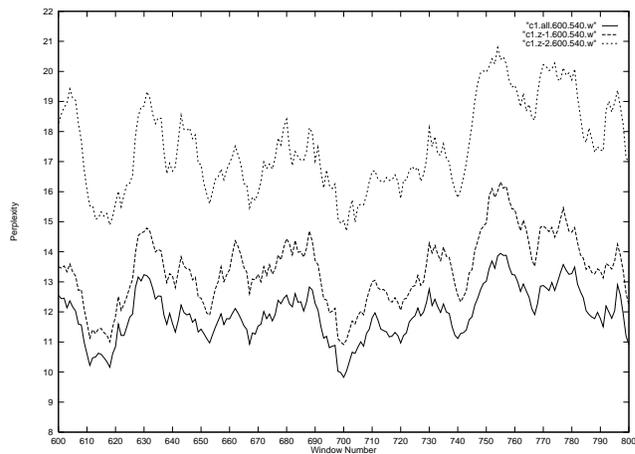


Figure 4: Local Perplexities for Different Class Bigram Models

length (in words) for each utterance. (This figure illustrates the phenomenon of long opening remarks by lawyers.)

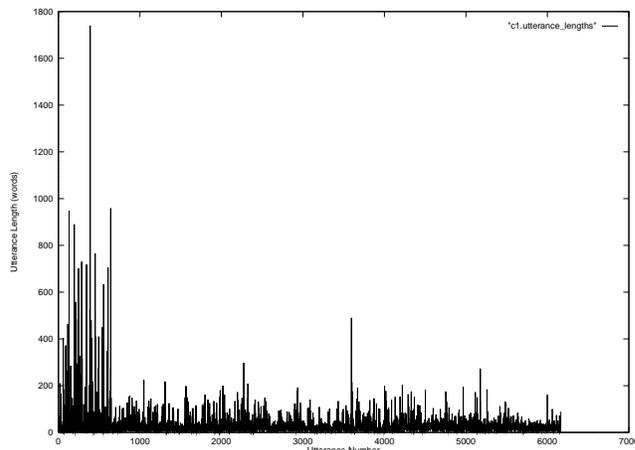


Figure 5: Utterance Lengths in Case c1

We computed the correlation coefficient of local perplexity and utterance length. Figure 6 shows the correlation coefficient over a small range of windows. Figure 7 shows local perplexity and the absolute value of correlation coefficient (with a threshold of 0.5), showing that there is little significant correlation between local perplexity and utterance length. This is repeated over the whole corpus.

Neither utterance length nor local perplexity correlate well with topic change over the entire corpus. Topic change was obtained by hand annotation of parts of the corpus, together with automatic location by means of keyword spotting. Such annotation is somewhat arbitrary (for example, it is possible to assign a number of topics to the transcript around window 700), and it is not clear that a more uniform, and less coarse labelling than that obtained by keywords, would yield better results.

And do you know - if you don't say so - do you know if there are employers that pay in in relation to off-site area , into that superannuation scheme ?

< Well , I know that there is .

Now , Mr Trohear , you are also a member of the divisional conference and divisional executive at a national level of the BWIU division of the CFMEU .

Is that right ?

----- Beginning of Window number 700 -----

< I am , yes .

And prior to its amalgamation , you were in fact a member of the national conference and national executive of the BWIU .

Is that

right ?

< I have been .

And you are also a member of the national conference and national executive of the CFMEU ?

< I am .

Yes .

And you are that by virtue of your - you hold those positions by virtue of your position in Queensland , that is the Queensland divisional branch secretary .

Is that right ?

< Yes .

You were in the courtroom this morning when I opened in relation to the formation of the CFMEU .

Do you recall that ?

< Yes .

Now , can you just explain to the commission , briefly , the purpose of the formation of the CFMEU , in terms of industry unionism ?

< The purpose is to bring - is to build the union , based on industries , and at this point in history those industries are the construction industry , the forestry and forest related products , the energy industry and the mining industry , and our purpose is to allow for each of those sections within the union , the maximum amount of democracy in determining their direction . And in fact the rules of the CFMEU are quite unique in that the national body of the union , who say they are over-riding the conference , can't make decisions which purely would restrict a single division .

So , for example , the conference would have difficulty making decisions which would effect the quality , singularly of , say , the BWIU division or later to be the construction division . So , it's quite a unique union in that concept .

In relation to the current situation , there is a division for the BWIU ?

< Yes .

Figure 3: Segment of transcript near window 700

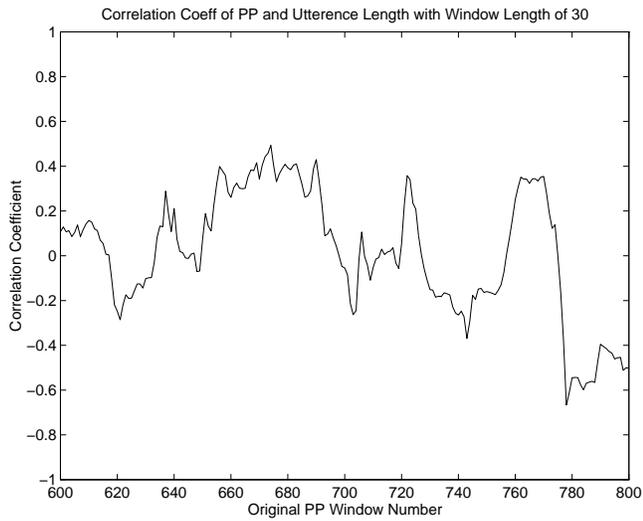


Figure 6: Correlation Coefficient Over Whole Corpus

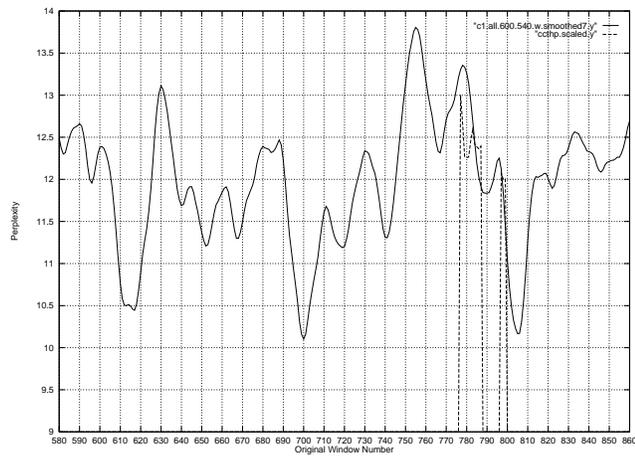


Figure 7: Section of the Smoothed Local Perplexity Graph Superimposed with the Thresholded Correlation Coefficient

4. ACKNOWLEDGEMENTS

We would like to thank Jonathan Cross for performing a number of the computations.

5. REFERENCES

1. Brill E. Some advances in rule-based part of speech tagging. In *Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI-94)*, 1994.
2. Ney H., Essen U., and R. Kneser. On structuring probabilistic dependences in stochastic language modelling. *Computer Speech and Language*, 8:1–38, 1994.
3. Kenne P.E., O’Kane M., and Pearcy H.G. An australian speech database derived from court recordings. In *Proceedings of the Fifth Australian International Conference*

on Speech Science and Technology, pages 808–813, December 1994.

4. Kenne P.E., O’Kane M., and Pearcy H.G. Language modeling of spontaneous speech in a court context. In *Proceedings Eurospeech 95*, pages 1801–1804, September 1995.