

1 The data

Results when the limit is 200

	correct	incorrect
Europarl-10	19	11
Europarl-30	27	3
Europarl-90	29	1

-Bigrams-

	correct	incorrect
Europarl-10	22	8
Europarl-30	29	1
Europarl-90	30	0

-Trigrams-

2 Trend

In the table above, we see that when the sentences are longer, the system is better at guessing the language. The trigrams are performing slightly better in all the categories. This is congruent with what we have learned about ngrams. When a incorrect guess is made, the incorrect language is often very similar to the actual language.

3 Exceptions

Interestingly when both bigrams and trigrams are wrong, they choose their incorrect guesses are often different languages. Moreover there are a few instances where the bigrams guess correct and the trigrams do not.

4 Manipulating the limit

With a limit of 10, the trigrams will be outperformed by the bigrams. Either they perform the same or worse than bigrams. When we set the limit to 20, their performance is more or less the same.

5 Results

5.1 Limit of 200

2-gram models for 10-word sentences: 19 correct, 11 incorrect

2-gram models for 30-word sentences: 27 correct, 3 incorrect

2-gram models for 90-word sentences: 29 correct, 1 incorrect
3-gram models for 10-word sentences: 22 correct, 8 incorrect
3-gram models for 30-word sentences: 29 correct, 1 incorrect
3-gram models for 90-word sentences: 30 correct, 0 incorrect

5.2 Limit of 20

2-gram models for 10-word sentences: 10 correct, 20 incorrect
2-gram models for 30-word sentences: 22 correct, 8 incorrect
2-gram models for 90-word sentences: 24 correct, 6 incorrect
3-gram models for 10-word sentences: 12 correct, 18 incorrect
3-gram models for 30-word sentences: 19 correct, 11 incorrect
3-gram models for 90-word sentences: 24 correct, 6 incorrect

5.3 Limit of 10

2-gram models for 10-word sentences: 10 correct, 20 incorrect
2-gram models for 30-word sentences: 14 correct, 16 incorrect
2-gram models for 90-word sentences: 21 correct, 9 incorrect
3-gram models for 10-word sentences: 5 correct, 25 incorrect
3-gram models for 30-word sentences: 14 correct, 16 incorrect
3-gram models for 90-word sentences: 20 correct, 10 incorrect