

# Measuring Old English syntax

Alisabet Ruffle

1 March, 2006

# Old English: the Questions

- ▶ The Question: *Can syntactic property X be applied to Old English?*
- ▶ The (often heard) Answer: *Not necessarily.*
- ▶ The Second Question: *Why not?*
- ▶ The (A) Second Answer: *Because of translation issues.*

# Translation issues?

- ▶ The only surviving examples of Old English we have are written pieces.
- ▶ A significant proportion of these pieces are translations from originals in other languages.
- ▶ There may be an influence of the original language's syntax upon the sentence structure observed in those pieces.
- ▶ If so, we cannot draw meaningful conclusions about the syntax of Old English.

# Old English sentences

Let's have some examples:

- ▶ þā dyde se cyning swā swā hē cwæþ, and se bisceop þæt geþafade
- ▶ then do-pst the king just-as he say-pst, and the bishop that consent-pst
- ▶ adv V det N adv proN V, conj det N proN V (Old English order)
- ▶ Then the king did just as he said, and the bishop consented to it.
- ▶ adv det N V adv proN V, conj det N V P proN (Modern English order)

# Old English sentences

- ▶ Hēr Sigerīc wæs gehālgod tō arcebisceope, and Ēadwine abbod forþfērde, and Wulfgār abbod fēng tō þām rīce.
- ▶ Here Sigeric was consecrate-pst to archbishop, and Eadwine abbot die-pst, and Wulfgar abbot succeed-pst to that office.
- ▶ adv Sigeric V-V P N (?), conj Eadwine N V, conj Wulfgar N V P proN N (Old English order)
- ▶ In this year Sigeric was consecrated to the archbishopric, and Eadwine the abbot died, and Wulfgar succeeded to that office.
- ▶ adv Sigeric V-V P det N, conj Eadwine det N V, conj Wulfgar V P proN N (Modern English order)

## More Questions

The scenery has now changed!

- ▶ Is the syntax of our present corpora of Old English significantly influenced by these translation issues?
- ▶ How can we determine this? Can we measure this influence?

# The Corpora

I will use the *Nijmegen Corpus*: a subset of the Helsinki Corpus of English Texts.

- ▶ 148 files; about 50 “documents”
- ▶ Already tagged
- ▶ Historically spread
- ▶ Contains both translated and non-translated works!

# The Analysis

- ▶ Using the Wiersma-Nerbonne tools
- ▶ Are there any other methods?

## An idea from coding theory

The *Hamming distance* of two vectors is a way of measuring how “different” they are:

$$v_1 = \{2, 4, 5, 7, 1\}, v_2 = \{2, 4, 7, 1, 5\}$$

Then the Hamming distance is

$$d(v_1, v_2) = 3$$

Can we adapt the Hamming distance to this project?

# Hamming distances: speculation

What would we need to take into account?

- ▶ Differences in sentence lengths
- ▶ Differences in sentence boundaries
- ▶ Assigning a value to the elements of each sentence
- ▶ Deriving the Hamming distance of a document from the Hamming distances of each sentence

# The End which is A Beginning

**Questions?**