Formulaic Sequences (FSs) in Early Modern English (EMoDE) 
A Corpus-assisted Historical Pragmatic Study

Ding Huang (M.A.)
PhD candidate at Anglistisches Seminar, Ruprecht-Karls-Universität Heidelberg | ding.huang@stud.uni-heidelberg.de

WHY IT IS INTERESTING
- Creativity vs. fixedness: Even though language is creative, people also use many fixed, clichéd, situation-based multi-word units to say things.
- A prevalent phenomenon: FSs in Present-Day English (PDE), being studied exhaustively, are found to make up a large proportion of both spoken (58.6%) and written (52.3%) discourse in PDE.
- Status of English in Early Modern England (1500-1700):
  - o Introduction of print press to England (1476) ➔ Standardisation
  - o The rise of English ➔ Multi-functional, ranging from private to official.
  - o Synthetic to a analytic ➔ flexible to fixed word order, e.g. Shakespeare liked to move words around in order to rhyme.
  - o Grammaticalization, e.g. “be going to” in EMoDE: literal, moving from place A to B
- PDE: grammatical, future tense

RESEARCH QUESTIONS
- What are the forms and functions of in EMoDE communicational texts?
- How do they characterise different types of communications and texts?

DEFINITION

FSs: relatively fixed multi-word units which frequently occur in a certain type of texts and serve as conventional pair of form, meaning, and function.

Figure 1: relationship between FSs and Construction Grammar

Table 1: 500 lexical bundles that have the highest expression rate were analyzed in depth.

<table>
<thead>
<tr>
<th>Type</th>
<th>Examples</th>
<th>Type/occurrence</th>
<th>Occurrence</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>verb+aux</td>
<td>I am (124)</td>
<td>12/4,036</td>
<td>9/212</td>
<td></td>
</tr>
<tr>
<td>noun+aux</td>
<td>my poor lord (14)</td>
<td>124/5,381</td>
<td>148/5,381</td>
<td></td>
</tr>
</tbody>
</table>

*They are found as part of longer FSs.

**Empirically, FSs are found as nothing outstanding but certain types of constructions, hence fall in the framework of Construction Grammar.

PILOT STUDY

Purpose: to test the methodology

Corpus: 37 plays by Shakespeare - 0.82 million tokens; modern spelling; not tagged.

Table 2: Categorization on the basis of the degree of schematicity from high to low.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Grammatical constructions (111)</th>
<th>Semantic constructions (156)</th>
<th>Situational formulae (51)</th>
<th>“Big word” (25)</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>FS (24)</td>
<td>Uncertain (90)</td>
<td>FS (24)</td>
<td>FS (10)</td>
<td>Examples</td>
</tr>
<tr>
<td>Examples</td>
<td>I have been X</td>
<td>I dare not X</td>
<td>I will not X</td>
<td>I know not X</td>
<td>Must needs be</td>
</tr>
</tbody>
</table>

***To which extend that sequences require “the addition of a single, semantically or formally restricted semantic element” (i.e. X) in order to be syntactically and semantically complete, and “if more Xs were allowed in a sequence, the degree of measured schematicity would be higher”.

This study restricts the number of compulsory X to one, but there is no limit for the number of optional X.

****“Uncertain” bundles are included in this analysis in order to determine if they should be seen as FSs.

INTERPRETATION: FSs are found in all 4 categories, but most of them are situational formulae and semantic constructions. Bundles in the “uncertain” table (see Table 3) tend to be syntactic and semantic incomplete, i.e. grammatical constructions (“G+L” in Figure 1) and semantic constructions (“L” in Figure 1).

Corpora

Corpus of spoken communication/speech-related texts:
- A Corpus of English Dialogues 1560–1760 (CED) - 1.2 million words; 177 test files;
- ‘authentic dialogue’ (Trial Proceedings and Witness Depositions), ‘constructed dialogue’,(Drama Comedy, Didactic Works, and Prose Fiction), Miscellaneous

Corpus of written communication/letters:
- Parsed Corpus of Early English Correspondence (PCEC) — c. 1410-1681
- 2.2 million words, 84 collections, 666 informants
- E.g. private, official, business letters, etc.

Figure 2: New texts are needed in both spoken and written corpora because CED and PCEC are not even in size and do not entirely cover the Early Modern period.

Procedures

Step 1: Computer assisted retrieval of LBs ➔ corpus-driven
- Frequency; length

Step 2: Manually identify FSs from the list of LBs ➔ corpus-based
- Completeness in syntax and semantics; fixedness; idiomatization

Reference:
10. Unicode .txt version by Mike Scott [http://www.lexically.net]