

# The Dickens Lexicon Digital and the Study of Late Modern English

Tomoji Tabata<sup>\*</sup>, Masahiro Hori<sup>\*\*</sup>, Osamu Imahayashi<sup>\*\*\*</sup>,

Miyuki Nishio<sup>\*\*\*\*</sup>, Kiyonori Nagasaki<sup>\*\*\*\*\*</sup>

## Abstract

Our project to create The Dickens Lexicon Digital is based on the compiled index cards of the late Dr. Tadao Yamamoto (1904-91). Dr. Yamamoto first conceived a plan for the compilation of the Dickens Lexicon, and published *Growth and System of the Language of Dickens: An Introduction to A Dickens Lexicon* in 1950, as an introduction to the Dickens Lexicon, which he was awarded the Japan Academy Prize for this work in 1953. In order to compile the Dickens Lexicon he collected the materials not only from all of Dickens's works but also from his letters and speeches. Unfortunately, in 1991 he died without seeing his vision accomplished.

The project for the Dickens Lexicon was organized in 1998 by a research group of 20 (currently 22) scholars. Our ultimate aim has been to compile the Dickens Lexicon from the approximately 60,000 cards which Dr. Yamamoto elaborately drew up and left to us, and he believed that even if the work was not issued in his lifetime, his pupils and successors would be able to publish it at some future date.

Our Dickens Lexicon is neither book nor document-based, but is rather designed as a web-based reference resource. Users will be able to search and retrieve lexical data (idioms, definitions, sources, quotations, and notes), that were recorded in the original card-database of approximately 60,000 indexed entries, without the need to install extra software (apart from a web browser) on their computers. Some of the types of information on Dickensian idioms should prove quite valuable for non-native researchers of English in particular, as certain idiomatic expressions in English which are common to native-speakers of English may be not noticed as idioms, or not understood as ironical or collapsed idioms. The lexicon will also be implemented with a multifunctional information retrieval system. In addition to the indexed entries, the lexicon will make it possible to retrieve frequency information on lexical items (from single words to phrases, including multi-word units) drawing upon the full corpus of Dickens' texts, and an additional set of comprehensive 18th and 19th century fictional

---

\* Professor, University of Osaka. Email: tabata@lang.osaka-u.ac.jp.

\*\* Kumamoto Gakuen University. Email: hori@kumagaku.ac.jp.

\*\*\* Hiroshima University. Email: imahaya@hiroshima-u.ac.jp.

\*\*\*\* Kinki University. Email: nishio@hiro.kindai.ac.jp.

\*\*\*\*\* International Institute for Digital Humanities. Email: nagasaki@dhii.jp.

texts. A range of functions such as concordance display, sort capability, distribution charts, and statistical data with t-scores and MI-scores and log-likelihood ratios will be available in a user-friendly interface. Therefore, a close scrutiny of idioms appearing in the Dickens Lexicon, incorporating this multifunctional information retrieval system, will not only make us more aware of the ways idioms provided an important characteristic in Dickens' usage of English (compared with those presented in almost comprehensive 18th and 19th century fictional texts), but will also provide greater insight into the characteristic structure of idiomaticity in the English language.

The Dickens Lexicon Digital is planned to be available on the web by March 2017, although it will be partially functional and available on the web from 2014.

This session provides an introduction to the Dickens Lexicon project, including its practical use for research.

**Keywords: Lexicon, Dickens, Digital, Idioms, E-texts, Collocation, Phraseology, Language Study**

2 0 1 4 D A D H

## **I. Osamu Imahayashi: “Dr Tadao Yamamoto and the Dickens Lexicon Project”**

This paper is an interim report on the Dickens Lexicon Project, discussed in the introduction. The team that has brought it to fruition consists of approximately twenty scholars who graduated from either Hiroshima University or Kumamoto University

It was not until the beginning of World War II that Dr Tadao Yamamoto first developed a version for the compilation of the Dickens Lexicon. As wartime conditions turned progressively worse, the completion of the Lexicon was left for the future. However, he decided to write an “Introduction” to the work in the early spring of 1944, and in the same year presented it to the University of Tokyo as his doctoral thesis under the title of Growth and System of the Language of Dickens: An Introduction to A Dickens Lexicon. Several joint research projects for the compilation of the Lexicon were organised, but unfortunately on the 28th of July in 1991, Yamamoto died without seeing it accomplished.

Our ultimate aim is to complete the compilation of the Dickens Lexicon from the cards drawn up by Tadao Yamamoto, who believed “even if the work was not issued in my life time, my pupils and successors would be sure to publish it sometime.” Our Lexicon is expected to be released as “Dickens Lexicon Online” on an internet website with the multifunctional search engine by 2014 in honour of the bicentennial of the birth of Charles Dickens if everything is “in a concatenation accordingly.”

## **II. Tomoji Tabata and Kiyonori Nagasaki: “Digital Enhancements to the Dickens Lexicon”**

The Dickens Lexicon is designed to tap into the digitized resources so as to facilitate related language research in historical stylistics and linguistic variation in late modern English: specifically, diachronic studies of phraseology in eighteenth- and nineteenth-century British fiction. The database of idiomatic expressions in Dickens’s writings and a corpus of fiction texts will help locate a particular idiom in Dickens’s texts together with its occurrence data based on the reference corpus of late modern British fiction. Users will be able to search and retrieve lexical data stored in the original card database of c. 60,000 indexed entries without installing extra software (apart from a web browser) on their computers. The lexicon will also be implemented with a multifunctional information retrieval system. In addition to the indexed entries, the lexicon will make it possible to retrieve frequency information on lexical items (from single words to phrases, including multi-word units) drawing upon a full-text corpus of Dickens texts and a reference corpus of major eighteenth- and nineteenth-century fiction texts. At the time of writing, the Dickens sub-corpus encompasses 24 sets of texts, amounting to approx. 4.8 million word tokens, while the eighteenth- and the nineteenth-century sub-corpora incorporate 33 works (approx. 5 million ) and 193 works (approx. 29.8 million tokens), respectively. A range of functions such as

concordance display, sort, occurrence distribution chart will be made available in a user-friendly interface.

### **III. Miyuki Nishio: “The Dickens Lexicon Digital: Definition of Idioms”**

Dickens adopts an abundance of colloquial phrases and expressions in his works. Studying the idioms that Dickens uses in his novels and letters allows us to deeply understand Dickens’ language as well as his works. Dr. Yamamoto began to compile the Dickens Lexicon, studying Dickens’s language, in particular, idioms.

Idioms have been studied from the viewpoint of grammaticalization over the last couple of decades and a precise definition of what idioms are remains elusive. Sixty years ago, Dr. Yamamoto, clearly ahead of his time, defined them as “expressions which are delimitable units of language” and explained them using Dickens’s language.

Bybee and Hopper (2001) note that “the more often two elements occur in sequence, the tighter will be their constituent structure.” The same applies to the use of idiomatic phrases in Dickens’s works. Dickens exploits various kinds of idiomatic expressions in his novels and sometimes develops them into transferred or metaphorical expressions. Yamamoto (1951) proposed a three stages cline of idiomatic development (change → development → stability) in Dickens’s works, which was later expanded by Hori (2011) adding the fourth stage of “deviation.” This presentation deals with the idiomatic expression “many’s the NP” construction; a phrase that is still in the stage of development in Dickens’s works.

### **IV. Masahiro Hori: “A Practical Use of The Dickens Lexicon Digital”**

As a practical tool The Dickens Lexicon Digital will be able to contribute toward the following research topics:

- (1) A study of the language and style of Charles Dickens.
- (2) A comparative study of idiomatic expressions between Dickens and other writers in the 18th and 19th centuries.
- (3) A chronological research of idioms.
- (4) A research into the language and style of 18th and 19th century fictional texts.
- (5) A research into the history of the English language.
- (6) An interdisciplinary research between information science and language study.

In this presentation several examples from these research topics will be discussed.

For example, according to Yamamoto's *Growth and System of the Language of Dickens: An Introduction to A Dickens Lexicon* (1950), idioms of "line," widely distributed in Dickens's writings, go through several grades of idiomatic development. He argues that the phrase "line of business" represents the first grade of development, from which "line," as the second grade, is detached and established idiomatically in the phrase "in the (potato) line." For the third grade of the idiom, he goes on to argue that the form of idiom has developed a transferred and figurative meaning.

These three grades of idiomatic development described by Yamamoto, are closely reviewed and examined through use of a machine readable version of the Dickens Lexicon, based on the approximately 60,000 cards that Yamamoto compiled, with a multifunctional search engine. The ability to investigate this database with new technologies provides us not only with the validation of Yamamoto's thesis but also new discoveries of Dickens's chronological development and an understanding of his consistency in using idiomatic expressions. Such a study on the grades of development of idioms in Dickens presents new and promising insights into a history of the English language as a practical use of the machine readable Dickens Lexicon and suggests a model for the lexicons of other authors.

## References

- Imahayashi, Osamu, Masahiro Hori, Akiyuki Jimura, and Tomoji Tabata (2008) "The Dickens Lexicon Project." ERA, Vol.25, The English Research Association of Hiroshima, 43–53.
- Yamamoto, Tadao (1950 [2003]). *Growth and System of the Language of Dickens: An Introduction to A Dickens Lexicon*. Hiroshina, Japan: Keisuisha.
- Yamamoto, Tadao (1954) "A Memoir of the Joint Research for the Compilation of the Dickens Lexicon," *Anglica*, Vol. 1, No. 5, 438–51.