




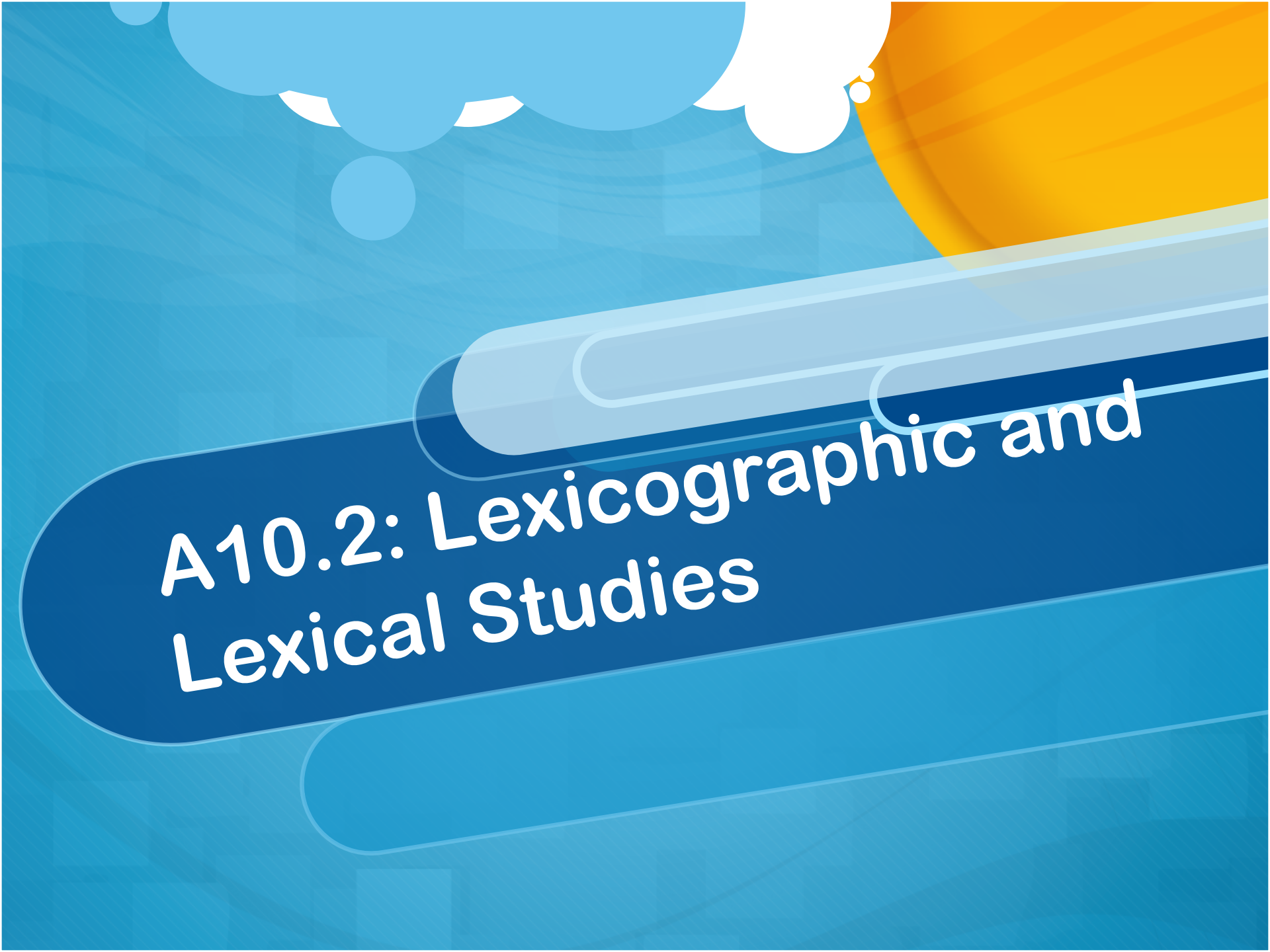
Unit A10: Corpora and applied linguistics

Misaki Shinkawa
I.D. #6113913
Period 3
June 20th, 2013

The background features a vibrant blue sky with stylized white and light blue clouds. A large, bright yellow sun is partially visible in the upper right corner. A blue pen with a silver clip is positioned diagonally across the middle of the frame. Below the pen, a dark blue banner with rounded ends contains the text 'A10.1: Introduction' in white.

A10.1: Introduction

- 
- ◊ Use of corpus data in language studies.
 - ◊ The following sections address:
 - ◊ A10.2-A10.8: Areas of linguistics where corpora have been typically used.
 - ◊ A10.9-A10.14: Other areas of studies that have started using corpus data.
 - ◊ A10.15: Limitations of using corpora in linguistic analysis.



A10.2: Lexicographic and Lexical Studies

A10.2: Lexicographic and Lexical Studies

- For lexicographic(辞書研究) and lexical(語彙研究) studies, corpus data is invaluable in a sense that it has contributed to the development of dictionaries (1990s~).
- Corpora are useful for the following five reasons:
 1. Machine-readable (Refer to Unit A1)
 - Extract authentic examples of the usage of a lexical item from a large body of text.
 - Speed of processing.
 - Accurate and consistent; avoid human bias in analysis.
 2. Frequency info. and quantification of collocations
 - Useful for materials writers and language learners.
 3. Corpus mark-up and annotation (Refer to Unit A3 and A4)
 - Textual (domain, genre, register, etc.) and sociolinguistic (age, sex, race, etc.) metadata ⇒ allow lexicographers to give accurate descriptions of usage.
 4. Monitor corpus (Refer to Unit A7)
 - Track minor changes in the meaning and usage of lexical item ⇒ dictionaries in up-to-date.
 5. Complement☺ or refute☹ intuitions of lexicographers ⇒ more accurate.

○ **S. Hunston** (2002: 96) summarized changes brought about by corpora in terms of five 'emphases':

1. An emphasis on frequency
2. An emphasis on collocation and phraseology
3. An emphasis on variation
4. An emphasis on lexis in grammar
5. An emphasis on authenticity



○ An important area of lexicographic study is **loanwords** (借用語, 借入語, 外来語).

○ Past: Intuitions > Corpus data.

○ Present: Research on 'Russianism' in English (Podhasek and Puitrowski 2003).

○ Comparison between BNC, *Oxford English Dictionary* (OED), & *Longman Dictionary of Contemporary English*.

○ 2 findings:

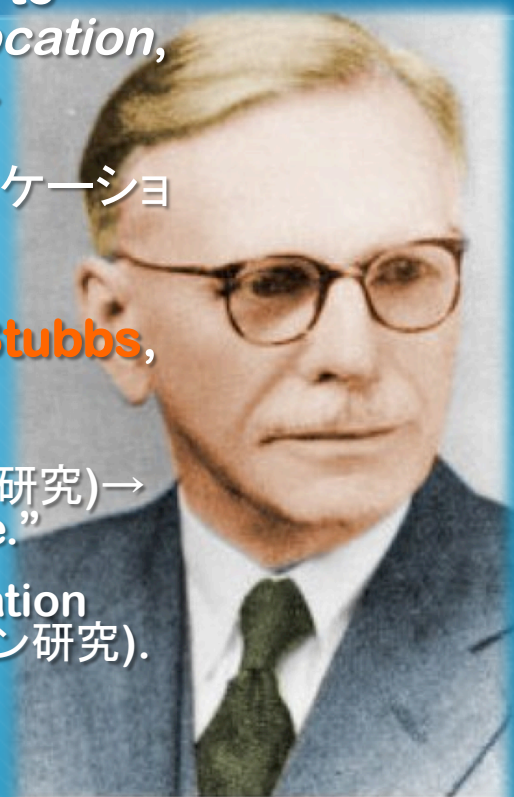
○ 1) Half of 360 Russian loanwords occurred only once in the BNC and only a few were very frequent.

○ 2) Items selected by OED on the basis of etymology(語源学) exhibit the same distribution pattern as items selected on the basis of frequency in the BNC → Collocational patterns growing around loanwords are significant.

⇒ Intuition and corpus data **do not** always lead to different conclusions.

⇒ "[a]ssimilation criteria based on frequency counts have proved to be less reliable than previously thought, and alternative criteria such as metaphor should also be taken into account" (Kurtboke and Potter 2000: 99).

- Corpora have been used extensively in lexical studies, with an emphasis on **collocation** (コロケーション, 共起表現) and **collocational meaning** (意味領域との共起関係).
- 50 years of research.
- Primarily introduced by **J.R. Firth** (1957): 'I propose to bring forward as a technical term, meaning by *collocation*, and apply the test of *collocability*' (Firth 1957: 194).
- Quantitative, statistical approach (統計に基づいたコロケーション研究).
- Accepted by Halliday, **Greenbaum**, Sinclair, **Hoey**, **Stubbs**, Partington, McEnery, Wilson, and **Hunston**.
 - Sinclair: lexical collocation studies (語彙的コロケーション研究) → Ideas of "semantic prosody" and "semantic preference."
 - ⇔ Greenbaum: lexical, grammatical, and semantic collocation studies (語彙の面だけでなく、文法的、また意味的なコロケーション研究).



J.R. Firth

Different approach on lexical studies

a. Statistical:

- **Greenbaum** (1974: 82) defined *collocation* as ‘a frequent co-occurrence of two lexical items(語彙項目) in the language.’ → But no specification on how frequently the co-occurrence of two lexical items should be to be considered a collocation.
- **Hoey** (1991: 6-7) uses the term *collocation* only if a lexical item appears with other items ‘with greater than random probabilities(ランダムな確率) in its (textual) context.’



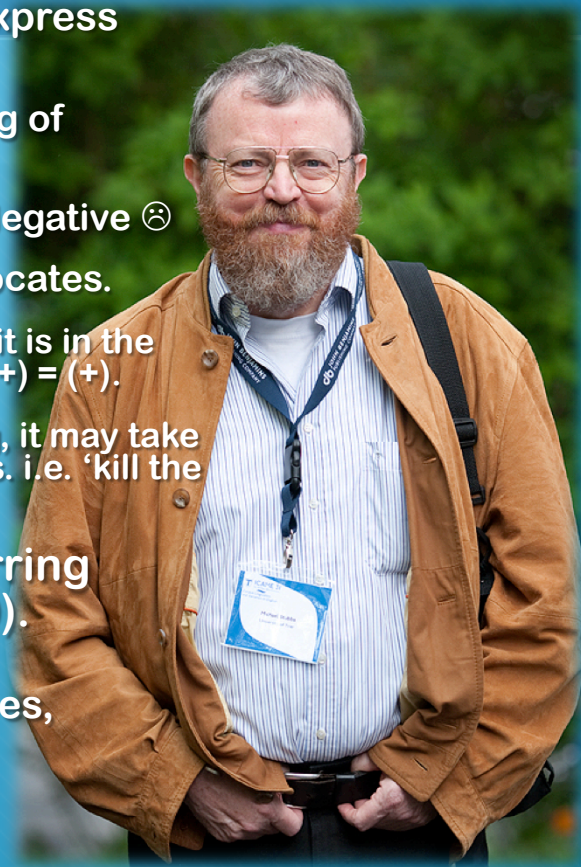
B. Grammatical:

- **Herbst** (1996: 382) defines *collocation* as a frequent co-occurrence with particular grammatical class of words.
- The collocation of a node word(中心語) with a particular grammatical class of words is normally referred to as *colligation* (1960s).

- Without corpus linguistics, no collocation research would have been possible!
- Intuition as a poor guide to collocation (Refer to A1.5).
 - Two noticeable examples of collocations:
 - A. 'He has been doing research in corpus linguistics for 25 years.'
 - B. 'He has been making research in corpus linguistics for 25 years.'
 - A. 'We visited schools here and there.'
 - B. 'We visited schools there and here.'
 - Why?
 - 'Each of us has poor knowledge of the language, we have prejudices and preferences, our memory is weak, our imagination is powerful (so we can conceive of possible contexts for the most implausible utterances), and we tend to notice unusual words or structures but often overlook ordinary ones' (Krishnamurthy 2000: 32-33).
 - Therefore, statistical evidence is a key in language studies😊

Michael Stubbs on Semantic meaning

- 'There are always semantic relations between node and collocates, and among the collocates themselves' (2002: 225).
- Negative, rarely positive.
 - i.e.: 'Commit'... suicide, a crime, a sin, etc. 'End up'... hating, crying, sleeping, etc.
- *Semantic Prosody* (\rightarrow an idea introduced by Sinclair): to express speaker/ writer attitude or evaluation (Louw 2000: 58).
 - Strongly collocational in that it operates beyond the meaning of individual word.
 - i.e.: 'Personal' and 'Price' \Rightarrow Neutral \ominus 'Personal price' \Rightarrow Negative \ominus
 - Result of the interplay between the item and its typical collocates.
 - The item does not appear to have an affective meaning until it is in the context of its typical collocates. i.e. 'awfully clever.' $\rightarrow (-) + (+) = (+)$.
 - \Leftrightarrow If a word has typical collocates with an affective meaning, it may take on that affective meaning when used with atypical collocates. i.e. 'kill the time.' $\rightarrow (-) + (+) = (+)$.
- *Semantic Preference*: 'a lexical set of frequently occurring collocates [sharing] some semantic feature' (ibid.: 449).
 - i.e.: 'large' \rightarrow 'quantities and sizes.' i.e. number(s), quantities, amount(s) etc.




Semantic Preference

- Relates the node item to another item from a particular semantic set.
- Feature of collocates.
- ‘Contributes powerfully’ to building semantic prosody (Partington 2004: 151).

Semantic Prosody

- Interaction between a given node word and its collocates.
- Affect wider stretches of text.
- Feature of the node word.
- ‘Dictates the general environment which contains preferential choices of the node item’ (Partington 2004: 151).

- Partington (1998: 68), Stubbs (2001a: 449) and Hunston (2002: 142): Semantic prosody is connotational (言外の意味).
- ⇔ Louw (2000: 49-50): 'the force behind semantic prosodies is more strongly collocational than the schematic aspects of connotation.'
- Authors' view: Connotation can be collocational or non-collocation whereas semantic prosody can only be collocational.
- Note that lexical studies is important for morphological analysis(形態分析) at the sub-lexical level not only for morphologists, syntacticians, and lexicographers, but also for all language learners.
- i.e. Distinction between '-ic' and '-ical.'
 - 'Classical music' or 'Classic music'?
 - 'Classical literature' or 'Classic literature'?



A10.3 Grammatical Studies (文法研究)

- Just like lexical studies, grammatical studies has also exploited corpus data.
- Balanced representative corpus provides a reliable basis for qualifying syntactic(統語論的) features of a language or a language variety; useful in testing hypothesis derived from grammar theory.
- *A Comprehensive Grammar of the English Language* as a milestone → *Longman Grammar of Spoken and Written English* as a new milestone.
- Gives ‘a thorough description of English grammar, which is illustrated throughout with real corpus examples, and which gives equal attention to the ways speakers and writers actually use these linguistic resources’ (Biber et al. 1994: 45).

Characteristics of new corpus-based grammar

- The availability of spoken English corpora has provided new insights into spoken grammar.
- CANCODE corpus has ‘allowed authors to discover many features of spoken grammar that are absent or marginalized in written grammar.’
- Grammar of spoken language is radically different from that of written language and thus need to be modelled on the basis of spoken data with no prior assumption that spoken and written grammars share the same framework.

Conclusion

- All the linguists specified in this section has given prominence to the close association between pattern and meaning → ‘a resource for vocabulary building in which the word is treated as part of a phrase rather than in isolation’ (Hunston 2002: 106).

neocogitus a number year old neophyte in the hallowed benedictine monastery in wittenberg was expelled from the order yesterday for disobeying the abbot dishonoring the time honored traditions of the ancient order and making pacts with the devil the young neophyte was charged by his abbot father orthodoxius paleologus with shirking his sacred duties in the scriptorium malingering and spreading heretical ideas click read more for full post young neocogitus is not suited to holy work said paleologus from the moment he entered the brotherhood he was a force of discord and disobedience not an hour passed without neocogitus doing something to disrupt the solemn traditions of the order after much soul searching fasting and praying over the problems caused by this wayward youth the holy spirit finally spoke to me and said for the good of the order neocogitus must go a promising beginning the trouble began on the first day the young neophyte was brought for training in the scriptorium the vast chamber in which monks of the order have been hand copying the holy writ for the last number years according to friar melonymous tuck the monk charged with training young neocogitus in the science of holy transcription the neophyte was arrogant sarcastic and insubordinate within a minute of entering the hall neocogitus beheld the glorious sight of these dozens of god servants backs bowed and heads down in pious meditative labor dutifully performing god work said tuck and he had the audacity to snicker i smelt a whiff of sulfur and knew we had a heretic in our ranks but neocogitus mastered his spleen said tuck and went on to prove himself an unusually able apprentice there no denying the young man was uncommonly quick to learn tuck continued all of the finer points of the book copying arts copying in neat script in straight lines spelling correctly and above all staying awake and alert neocogitus mastered within a day indeed by the end of his first eight hour duty the young man had produced a flawless reproduction of the book of genesis a task that took three times as long for far more experienced scribes more remarkable still by the end of his first month as an apprentice scribe neocogitus achieved what had never been done before he had produced an entire copy of the bible all number books plus the apocrypha it was a miracle said tuck it had never been done before and it was perfect flawless i personally checked each and every line for errors and there were none abbe paleologus heard the news with joy it seemed a sign from heaven paleologus said there were so many heathens living in the darkness helpless to see the light without god word yet because there were so many more heathens than there were monks to copy the holy writ and because it normally took six months for a scribe to produce one correct copy it seemed we would never be able to rescue all the heathens from their ignorance but this young monk neocogitus the abbot continued seemed sent to improve our chances i had heard that his conduct was often troublesome irreverent and lacking in humility but at the time i thought this was one more instance of god mysterious ways this whelp would help us spread god word with godspeed little did i know he concluded that this was not god work at all but the devil the devil work neocogitus first bible aroused curiosity throughout the monastery the order was abuzz with talk about its inerrancy legibility elegant script above all however the talk focused on this question how had the young man produced it so fast was it really possible to produce an accurate copy of the entirety of the scripture in one short month to get to the bottom of this mystery father paleologus summoned neocogitus to his chambers for a private interview what is the secret to neocogitus miraculous powers learn this and more in the next episode on sale at blogstands soon

References

- 堀正広(2009)『英語コロケーション研究入門』 研究社。

