Making useful wordlists for ELT
Simon Smith, Ming Chuan University
ssmith@mcu.edu.tw

We use corpora from the WWW to make vocabulary lists.

Outline
• Importance of learning natural English
• Wordlists in English learning
• Making relevant wordlists
• Using two corpus analysis tools
  – WebBootCat
  – Sketch Engine
• Conclusions and future plans

The problem
• Learning non-authentic English
  – It’s raining cats and dogs!
  – Long time no see!
• In Taiwan, all students learn these
• They may believe they are authentic
• But English speakers hardly use them!

Word and phrase lists
• Students must learn vocabulary
• It is best to learn vocabulary through practice:
  – Reading
  – Speaking to American people
  – Interacting in the language
• That is difficult for Asian students
• In Taiwan, students must learn vocabulary from lists

From the Taiwan Education Ministry
• 6000 word high school list
  – Probably useful for policy makers
  – May be useful for teachers
  – Not useful for learners
• Better to organize wordlists by topic?

Vocab lists are only useful if organized in some way, probably by topic. We still need to take care over what words are included. It is difficult to think of vocab items, related to a topic, by brainstorming or introspection.

We generated a good vocab list using WebBootCat, contrasted with a vocab page from Simon’s university textbook.

From my university’s textbook
• It is not easy to make up a good vocabulary list for an abstract topic
• Try these topics:
  – Unit 1: Getting started at University
  – Unit 2: Family and Hometown
• Please
  – Choose a topic
  – Write down some good keywords
• Better use computer to help us!

So, we should teach vocabulary by topic?

Students are sometimes asked to learn words and phrases which are not authentic. Vocab lists are not the best way to learn language, but we should still try to make the vocab meaningful and useful.
The steps involved in generating a corpus from the web are shown.

We can analyse the corpus to get concordances and other output, as well as wordlists.

We can use the output wordlist as a set of seed words to bootstrap an even better wordlist. Also, we can include multi-word terms in the wordlist.

We sum up the advantages of using corpus-generated wordlists.

Advantages of automatic wordlist creation

• contain relevant, topical vocabulary
• created easily and conveniently
• of course, we can select the words manually, from the automatic list!

Disadvantages of manual wordlist creation

• It is difficult to get inspiration to make good wordlists manually.
• Manual wordlists may include rare or unnecessary vocabulary.

Right now, we are working on ways to generate cloze exercises from corpora. Thank you for coming! Please go to www.sketchengine.co.uk and make your own corpora and wordlists!