How can language learners benefit from corpora (or not)

Ana Frankenberg-Garcia

Three gateways to corpora

First

Second

Third

Published corpus-based tools and resources
Spell checkers and autocorrect options

<table>
<thead>
<tr>
<th>absence</th>
<th>absence</th>
</tr>
</thead>
<tbody>
<tr>
<td>accessories</td>
<td>accessories</td>
</tr>
<tr>
<td>accident</td>
<td>accident</td>
</tr>
<tr>
<td>accommodate</td>
<td>accommodate</td>
</tr>
<tr>
<td>according to</td>
<td>according to</td>
</tr>
<tr>
<td>across</td>
<td>across</td>
</tr>
</tbody>
</table>

Automatic redirection of web searches

![Automatic redirection of web searches](image)

Machine Translation

![Machine Translation](image)

Corpus-based ELT materials

![Corpus-based ELT materials](image)

New generation dictionaries

- Word frequency
  - Which headwords to include
  - Which senses to present first
  - Which words to use in definitions
- Grammatical properties of words
- Corpus examples
- Collocations
- Learner corpus data

![New generation dictionaries](image)

person - definition

1. an individual human, usually an adult. The plural is people, but in formal or official language the form persons is used.

- Rooms cost £20 per person.
- She's the type of person who always offers to wash up after a party.
- Every single person in the room stopped talking.
- Some people hate camping.
- elderly/eligible/unemployed persons

http://www.macmillandictionary.com/dictionary/british/person
Grammars

Longman Grammar of Spoken and Written English (Biber et al. 1999:214)

Textbooks

Touchstone Student Book 1 (McCarthy et al. 2005:25)

Textbooks

Touchstone Student Book 1 (McCarthy et al. 2005:46)

Published corpus-based tools and resources

- Language selected and edited by experts
- Information from raw corpus data processed by experienced professionals
- Learners benefit from final product
- No hassle
Published corpus-based tools and resources

- Amount of language in polished format is limited
- Language is infinitely bigger and more complex
- Learners (and teachers) have questions about language not covered or not covered in sufficient detail in dictionaries, grammar books, etc.

Direct uses of corpora

- One-to-one student teacher corpus consultations
- Data-driven learning (Johns 1991)
- Use of computers not widespread
- Corpora not accessible to general public

Direct uses of corpora

- Twenty years later ...
- Anyone with a computer and a connection to the internet can access corpora

And...

- There are books and websites about the use of corpora in language learning

If teachers start using corpora

Learners can come into direct contact with corpora too

1. Without computers
   - The soft approach (Gabrielatos 2005)
   - Hands-off (Boulton 2010)

2. With computers
   - The hard approach (Gabrielatos 2005)
   - Hand-on (Boulton 2010)

Using corpora hands-off

- Learners do not need to learn how to use concordancing software
- Teachers can prepare countless different types of hands-off corpus-based exercises for them
- But the mere fact that a classroom handout or activity is based on corpora is no guarantee of success ...
1. A smooth cheek, nearly hidden beneath grimple and sunglasses.
2. Protected by a grimple of bone, the brain was an organ of mind-boggling complexity.
3. Shoshana noticed her hair was so perfectly sleek it looked like a blond grimple.
4. Megan flopped down next to me and yanked off her grimple.
5. I heard my own laughter, echoing inside the grimple.
6. A woman with a ponytail sneaking out from under her grimple.
7. The balaclava can also help counteract chills from your vented grimple.
8. You can be pretty anonymous with a grimple on, I keep it that way.
9. She said he wore a grimple with the visor down.
10. During rehab, she will probably wear a grimple.

Grimple exercise
• Interesting from a linguistic perspective
• Raises language awareness
• Shows meaning can be derived from context
• But why bother?
• Isn’t it easier and faster to look up grimple in a dictionary?
• Grimple = concordances for helmet in COCA*


“grimple” in MacMillan Dictionary online

helmet - definition

NOUN [COUNTABLE] Pronunciation /ˈhelmɪt/

a hard hat that you wear to protect your head

a motorcycle helmet

Thesaurus entry for this meaning of helmet

“grimple” in Oxford Portuguese Dictionary Online

Another example

Many of them were ______ workers, you know support staff, people who cleaned, cooked.

Employment? – All ______ jobs, answered Claire. ‘dishwasher, building janitor…’

…he does ______ tasks vaguely tied to insurance.

If they do find a job, it’s ______ labour.


Missing word exercise
• Interesting from a linguistic perspective
• Raises language awareness
• Shows meaning can be derived from context
• But do learners have time for this kind of analysis when they only have a few hours a week to learn a foreign language?
• Won’t they feel frustrated if they can’t guess that the missing word is menial?
Another example

Exercises

1. Study the following concordance lines with can and classify them according to whether they relate to:
   1. Ability
   2. Possibility
   3. Permission
      (There are three of each.)
   a. I can cook anything because I’m a Gordon Ramsay chef.
   b. You can also reach the museum by train from any station on the suburban line.
   c. How often you, Pierre, and your can join me in a bit of shopping?
   d. Chimps can create artworks, use tools in a quite sophisticated fashion, and understand the concept of zero.
   e. Anyone who turns up on the day can join in the fun.
   f. Many fishing tasks can be seen as stochastic.
   g. Tropical fish can live without food for six weeks.
   h. Many credit cards can now be used to draw cash from ATMs around the world.
   i. You can keep it. I’ve got others.

From: Natural Grammar (Thornbury 2004:31)

Can exercise

- Interesting from a linguistic perspective
- Raises language awareness
- No evidence that classifying different meanings of *can* has a positive effect on learning
- Conscious analysis could even be detrimental

Aisle exercise

(Frankenberg-Garcia 2012:40)
1. The air hostesses inquired what I was making and a man passing in the *aisle* quite genuinely complimented me on my work.
2. I arrived at Salisbury Cathedral, just as the bride was about to go up the *aisle*.
3. As she looked around she felt a twinge of sadness that in a carriage where 70 per cent of the commuters were men there were five women forced to stand in the *aisle*.
4. He hurried up the *aisle* of the church.
5. She picked up her suitcase and made her way along the *aisle*.
6. The layout of the store, with wide *aisles*, gives customers room to move around.
7. I spend much of my time at the shops; wandering through the *aisles*, faltering, never knowing what to buy.

Grimple, menial, can and aisle

- On the surface, concordances for *grimple, menial, can* and *aisle* are very similar
- All focus on a given word
- All draw attention to context
- But *aisle* concordances were different

Aisle

- Not a word at random
  - learners were curious about *aisles* because of authentic classroom situation
- Concordances used to expand previous one-off contact with the word
- New input exponentially enhanced with concentrated doses of word in context
- Learners figured out there are *aisles* not just on planes, but also on buses, trains and in shops, churches, cinemas...
- Learners noticed difference between *aisle* and *corridor*
Hands-off activities are not just about language comprehension

Write a composition about something that happened on a bus (Frankenberg-Garcia 2012:43)

1. Back at the bus stop, the other people were furious.
2. You missed the bus, “ she said.
3. When the bus pulled up to our school on Lincoln Street, I stood up.
4. I was sitting on the bus to school, I was looking out the window.
5. I saw a man trying to catch the bus as it pulled out of the station.
6. We got on and off the bus together.
7. She reached the gate just as the bus driver was collecting the last of the tickets and closing the door.
8. The people in the bus queue are going to work.
9. He walked to the bus shelter at the roadside, and waited.
10. Benny knew they would come to meet her off the bus.
11. I’m usually late for the bus.
12. I used to ride the bus by myself when I was your age.

Hands-off activities need not be just on paper

- Learners working on a textbook unit about payment methods
- Many preposition errors noted
- Slides with concordances for lemma=pay for whole class participation and discussion

First, concordances with lemma=pay leading learners to figure out

- You pay people (no preposition)
- You pay for things in general
- But you pay the bill/tax/rent (no preposition)
- You pay (in) cash
- You pay in dollars/euros etc.
- You pay by cheque/credit card

Next, gapped concordances for practise Source: COCA

- Americans living in Europe but being paid …… dollars feel as if they’ve taken a big pay cut.
- Brenna’s grandmother paid …… the trip with what would have been Brenna’s college money.
- He foolishly agreed to pay …… Daniel the outrageous sum of fifteen thousand dollars to represent him in court.

- I never paid …… the book. I stole it.
- She paid …… the bill, and they left the restaurant, full but dissatisfied.
- I paid …… her $5 to cut the grass.
- Etc.
Lemma=pay slides

- Learners had problems using prepositions with pay
- Massive doses of pay in context
- Because concordances were not bland, exercise-book sentences
- Stimulating conversation topics to discuss as a group
- Not just another boring, solitary grammar practice exercise

Hands-off activities need not be just on paper

- With user-friendly freeware like HotPotatoes (http://hotpot.uvic.ca/)
- Teachers can create customized interactive CALL exercises for learners to work at their own pace

Hands-off activities need not be just on paper

- Learners of English for Tourism having trouble understanding Americans using inches, gallons, ounces, etc.
  – setting is directly related to what is going on in the classroom
- CALL exercise with concordances from COCA to help learners become familiar with non-metric units used in the USA

Hands-off need not be just concordances

- Collocation data can be used to boost vocabulary related to any topic (Frankenberg-Garcia 2012)
- Word Sketches (Kilgarriff et al 2004)

Using corpora hands-on

- More problematic than hands-off (Gabrielatos 2005, Boulton 2010)
- Learners need to grapple with
  – New technology
  – Raw corpus data
  – New approach to learning

Word Sketch for "train" - enTenTen corpus (Jakušček et al 2013)
Using corpora hands-on

From the view point of technology
- Language classrooms don’t usually take place in a computer lab
- Language learners may not like technology
- All sorts of things can go wrong with computers
- Internet connection can be slow or unavailable
- Concordancers are not particularly intuitive or user-friendly

Using corpora hands-on

From the perspective of the corpus, learners need to
- Decide which corpus to use
- Understand the strengths and limitations of the chosen corpus
- Interpret their results in the light of the above

Using corpora hands-on

From the perspective of learning, learners may feel overwhelmed and overburdened with
- The amount language at their fingertips
- Raw corpus output, i.e., unedited language
- Having to figure out for themselves what to look out for in a sea of data

Using corpora hands-on

And if this is not enough to discourage language learners...
- Many hands-on classrooms activities described in the literature are for linguists, not for learners
- Interesting from the viewpoint of language research, but dissociated from the reality of language learning

Using corpora hands-on

Reppen (2010:44)
- Which texts have the highest normed counts for well in MICASE?
- Is well being used as a filler, a discourse marker or an adverb?
Reppen (2010:62)
- Create frequency lists for a few newspaper texts and scan resulting word lists to determine level of difficulty of each text

Using corpora hands-on

Hands-on corpus consultation needn’t be discarded
- Particularly suitable for learner’s individual questions
  - What preposition/verb/adj/adv should I use here?
  - How do you say X in English?
  - Why is X wrong?
  - Which word sounds better, X or Y?
  - Can I say X here?
Using corpora hands-on

• If all learners are asking the same questions, worth preparing a hands-off activity for the whole class
• If only individual learners ask, show them how to look up the answers autonomously
  – But dictionaries are quicker (easier?)
  – If not in dictionaries, use corpora

What preposition can I use here?

I received a message in/on/at my cell phone?

Monolingual learners’ dictionary

http://www.oxfordadvancedlearnersdictionary.com/

Try this: http://corpus.byu.edu/coca/

Another hands-on example

• Business English class
• Looking at closing lines in business letters
  – Should you need any further information, please do not hesitate to contact me
  – Once again, I apologize for any inconvenience
  – I look forward to your reply
  – I look forward to hearing from you
  – I look forward to seeing you next week
  – Etc.

Last year’s teacher said you can’t say “I look forward”, you have to say “I am looking forward”
Well, let’s check

- This site (http://www.someya-net.com/concordancer/) contains Business Letters in British and American English

![Business Letter Corpus](image)

Looking forward to hearing

When using corpora hands-on

- Teachers can demonstrate to learners how they can use corpora
- To find out answers to questions that are in the forefront of their minds
- For which there are no ready-made answers in dictionaries and other conventional references
- Relevance and usefulness of corpora is immediately obvious

When using corpora hands-on

- One computer only in class can be used like a shared dictionary
- It’s unlikely that language learners will become power users of corpora — But is that what they want?
- After a few demonstrations learners can begin to use corpora on their own to answer simple questions that are relevant to them

Recent developments

- SkELL (Baisa & Suchomel 2014)
- Sketch Engine for English Language Learning
  - > 1 billion words
  - texts from news, academic papers, Wikipedia articles, open-source (non-)fiction books, webpages, discussion forums, blogs etc.

http://skell.sketchengine.co.uk/run.cgi/skell
Some recommendations

- Corpus-based exercises and activities should be relevant to what is going on in the classroom
  - Language learners don’t have time to and may not be interested in analysing language for it’s own sake
- There is no need to use the terminology of corpus linguistics
  - Don’t say corpus, say website or tool
  - Don’t say concordance, say sentence or text extract
  - Don’t say collocation, say words that go well together

Some recommendations

Learners need specific instructions on what to do with corpora

- When using corpora hands-off
  - Printing out concordances and asking learners to observe and discuss them is too vague
  - Tell them exactly what to look out for
  - If they notice other things as well, great.
- When using corpora hands-on
  - Show that a corpus can help
  - Show how to carry out relevant query
  - Make sure learners can understand the results

Some recommendations

When dealing with corpus data

- If hands-off
  - Select which data to use and edit it if necessary
  - Leave out language that is too difficult or inappropriate for the classroom
- If hands-on
  - Encourage learners not to be intimidated by unedited language and to focus instead on the answers to the question that prompted their query
  - But be there to help learners with any clarifications that they feel are important
  - Sometimes one query leads to another one...

References


And to conclude

- When discussing whether or not data-driven learning actually works
- Remember there are different ways of introducing learners to data-driven learning
- What is fascinating to a linguist may not be relevant or useful to a learner
- For data-driven learning to work, it must focus on what is interesting to the learner rather than on what is interesting about the data
References