

The aim of the Trinity Lancaster Corpus project is to create a large corpus of learner speech which will become a leading research tool in a wide range of contexts including Second Language Acquisition, language testing and L2 pedagogy and materials development

What is a language corpus?

A language corpus is a collection of texts which is compiled for the purpose of language analysis.

In a spoken corpus, recorded speech is coded with tags so that users can examine the texts to determine how language is used in particular contexts (eg in formal or informal situations), by specific groups (eg different ages, different mother tongues), for specific purposes (eg for academic purposes or for social purposes).

Who are the speakers in the corpus?

The data comes from speakers taking Trinity's GESE test, which focuses on communicative skills and allows test takers some choice in their contributions. This means that the Trinity Lancaster Corpus can offer unique insights into how learners manage interaction and build meaning based on their own identity.

The corpus samples the language of learners from a variety of L1 backgrounds, representing English speakers from Italy, Spain, Mexico, Argentina, Brazil, China, India, Sri Lanka and Russia.

How can the corpus findings be used?

The findings of such analyses can be used for many practical and pedagogic purposes:

- ▶ devising teaching materials
- ▶ constructing tests and other assessment procedures
- ▶ compiling accurate dictionaries
- ▶ improving communication among different social or cultural groups.

It will also allow the development of locally focused teaching materials and test support activities.

Corpus milestones so far

- ▶ 300 hours of transcribed speech
- ▶ 3.5 million words
- ▶ 1,500 L2 English speakers
- ▶ 3 proficiency levels
- ▶ 9 linguistic & cultural backgrounds
- ▶ large age range (9-72)

I think...
I think...
I think...
I think...

What does a line of text from a corpus (a concordance) look like?

Look at this example where **I think** is the key phrase:

ut dreams 0:01:32.9 E: ah **S: erm I think** dream **is** what you like E: ah S: an
I want to be a translator because **I think** foreigner **is** very friendly E: yeah r
to? S: erm E: to be a translator **S: I think** I should be er study hard E-English
you? S: yes I need I need **and** but **I think** health **is** the impo= the most impos
s the best 0:07:01.7 E: mm **S: but I think** they are silly E: yeah S: because t
buy a good doctor S: but er **erm I think it's** more important than <unclear=
:s the rules when he's driving **S: I think it's** no because he's drive us <.> g

Corpus findings

Example

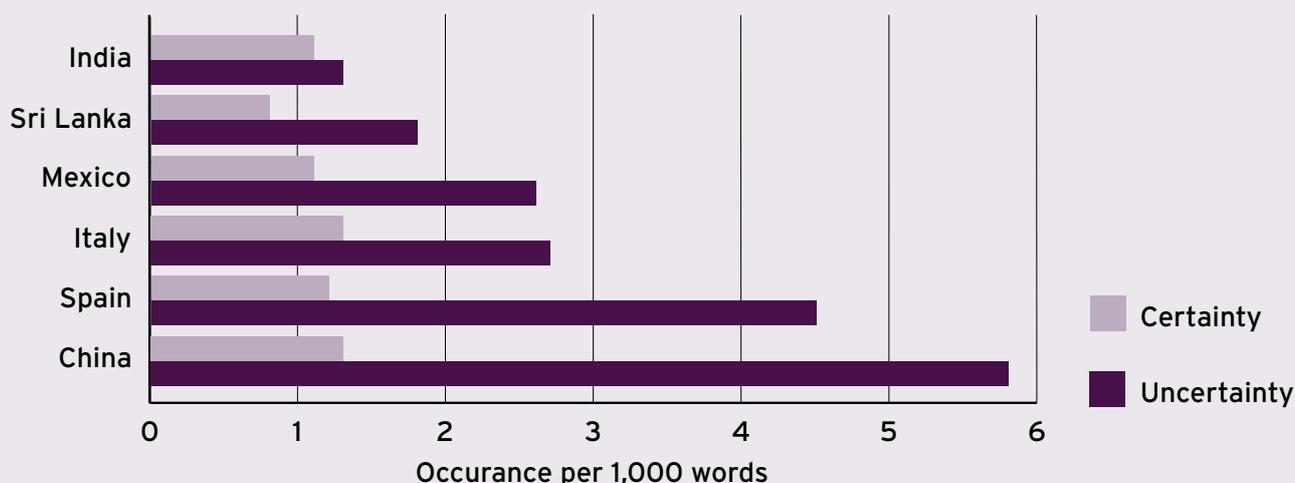
Did you know: active listenership is a feature of better performance in our tests?

Research finding: More successful communicators are more engaged as listeners.

Communicative implications: Successful communication does not rest solely on speaking skills – appropriate listening skills are of equal importance in helping the conversation to flow naturally. Good listeners help to co-manage the conversation by sending signals to the speaker. These signals are, for example: **mm, okay, yes, yeah, uhu, oh, no.**

<E> *have you heard of t'ai chi?*
<S> **er**
<E> *it's a very old Chinese exercise*
<S> **uhu** *no I don't know it*
<E> *and it looks very similar but it works on an inside energy*
<S> **ah**
<E> *in the body so you are you are bending*
<S> **mm**
<E> *in a very good position your pelvis*
<S> **mm**

Different first languages express uncertainty differently



Corpus findings: Look at how much certainty and uncertainty the speakers with different cultural and linguistic backgrounds express. Expressing certainty seems to be fairly universal in terms of its quantity. Uncertainty – which is often connected with politeness – varies considerably depending on the speaker's background.

Read more at http://cass.lancs.ac.uk/?page_id=1327

Find out more about the Trinity Lancaster Corpus by visiting trinitycollege.com/corpus