Description of the Trinity Lancaster Corpus (version for Early Data Grants)

The corpus made available for the Early Data Grants will contain approximately two million words from the Trinity Lancaster Corpus. Below are some of the key characteristics of the corpus.

1. Speaker characteristics: Linguistic background and proficiency level of speakers

The corpus contains data from L2 speakers of English from different linguistic and cultural backgrounds at three broad levels of language proficiency (B1, B2 and C1/C2 level of the Common European Framework of Reference). At each proficiency level (e.g. C1) speakers can be further distinguished according to the overall grade they received for their performance into more successful (i.e. those with ‑Distinctionø and ‑Meritø) and less successful speakers (those with ‑Passø grade). The table below offers an overview of the number of speakers from each background at each proficiency level.

L1 & cultural background	English proficiency		
	B1	B2	C1/C2
China	80	70	30
India	100	100	30
Italy	100	100	60
Mexico	100	70	60
Russia	40	28	-
Spain	100	100	60

The data from India contains speakers of different first languages (and their combinations) such as Hindi, Gujarati, Marathi, Tamil and Bengali. The sample released for the Early Data Grants will contain at least the following number of speakers of Hindi, Gujarati and Marathi.

Linguistic background	English proficiency		
	B1	B2	C1/C2
Hindi	50	50	6
Marathi	15	15	7
Gujarati	30	26	3

2. Speaker characteristics: gender and age

The corpus is also searchable according to the age of speakers and their gender. In terms of age, the corpus contains speakers ranging from 9 to 72 years of age. The speakers can be either categorised into larger groups (e.g. young speakers vs. older speakers) or speakers of specific age and age range can be searched for (e.g. 9-10 year olds, 30-40 year olds).

3. Speaking tasks

The corpus contains spoken data from different speaking tasks in which one L2 speaker interacts with one L1 speaker of English. Each L2 speakers participates at least in two different tasks. The number of tasks is related to the proficiency level of the speakers and the

examination grade they opted for. The table below provides an overview of the speaking tasks for each proficiency level.

Speaking tasks		Proficiency band		
		B1: Threshold	B2: Intermediate	C1 & C2 Advanced
Monologic	Presentation			✓
Dialogic	Interactive task		✓	✓
	Discussion	✓	✓	✓
	Conversation	✓	✓	✓

In the first task, the *presentation,* the L2 speaker talks about a topic of his/her own choice. The *presentation* is semi-formal or formal in nature. The three tasks following the presentation are dialogic and represent different interactive situations: *Discussion* involves the examiner and the candidate discussing the topic of the *presentation*. In the *interactive task* the candidate is required to take a proactive role in the conversation with a series of questions or comments following a prompt from the examiner. The final task, *conversation,* involves talking about two topics of general interest. For C1 candidates these topics are selected from a list previously known to the candidate; for C2 candidates any topic can be chosen by the examiner. The four tasks thus represent a variety of settings in which spoken language is used with different degrees of formality, interactiveness, topic familiarity and different interlocutor roles.

Speaking tasks: Individual speaker contribution in each task

	Average no. of words per L2 speaker		Duration
	Mean	SD	
PRES	655.67	122.59	up to 5 min.
DISC	469.04	125.34	up to 5 min.
INT	379.49	97.06	up to 5 min.
CONV	684.71	162.32	up to 6 min.

Speaking tasks: Discourse characteristics

	Topic familiarity	Interlocutor roles	Interactiveness	Formality
PRES	pre-selected topic	candidate-led	monologic	semi-formal/formal
DISC	pre-selected topic	jointly-led	dialogic	semi-formal
INT	general topic	candidate-led	dialogic	semi-formal
CONV	general topic	jointly-led	dialogic	semi-formal

For a more detailed description of the tasks and examples of the tasks, see the following:

- Trinity College London (2016). Exam Information: Graded Examinations in Spoken English (GESE). Available at http://www.trinitycollege.com/site/?id=368 along with videos of the speaking tasks

- Gablasova, D., Brezina, V., McEnery, T. & Boyd, E. (2015). Epistemic stance in spoken L2 English: The effect of task type and speaker style. *Applied Linguistics* (Advance Access).
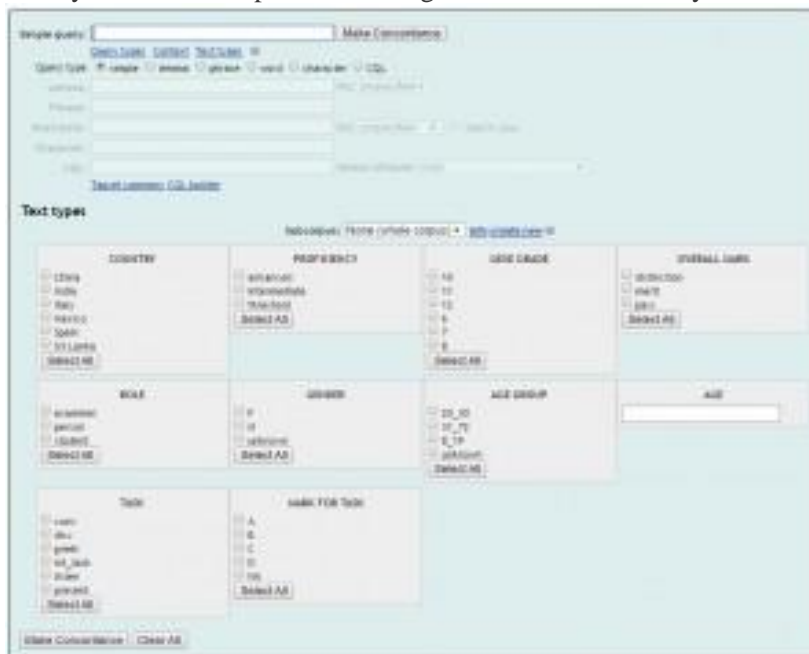
- Gablasova, D. & Brezina, V. (2015). Does speaker role affect the choice of epistemic adverbials in L2 speech? Evidence from the Trinity Lancaster Corpus. In Romero-Trillo, J. (Ed.) Yearbook of Corpus Linguistics and Pragmatics. Springer.

3. Search environment and analyses

The SketchEngine (SE) interface allows the corpus to be searched according to different variables (e.g. age of speakers, proficiency) and their combinations (see the Figure below for an example of SE interface). It is possible to search for single words and specific expressions as well as for collocations and n-grams. Researchers can also search for various grammatical and lexico-grammatical constructions (e.g. passives). Using SketchEngine, it is possible to obtain data for

- groups of speakers which can be used in statistical analyses using statistical tests such as chi-squared.
- individual speakers which can be used in statistical tests taking individual differences between speakers into consideration (e.g. Mann-Whitney U, ANOVA, t-test).

Trinity Lancaster Corpus: SketchEngine search functionality

4. Examples of transcripts: Excerpts from the Trinity Lancaster Corpus

These excerpts illustrate the nature of language produced in the individual tasks (Note: only parts of each task are included).

PRESENTATION
[L2 Speaker (S): Age: 49; Gender: female; Exam location: India; Examiner (E)]

S: erm my topic for presentation today is the social conditioning I will first of all speak a bit about definition of what social conditioning means and secondly I'm going to speak about the pros and the cons of social conditioning and what I believe erm are the risks and benefits
E: mm
S: and finally I will conclude with my opinion and observation of how we can tackle social conditioning and how <1.26> relate erm social conditioning is natural behaviour or phenomenon which has been handed down through centuries of tradition and erm certain progressive changes through generations er now social conditioning infor= er reflects in human behaviour it's said that a child up to the age of three er i-in that child right brain is dominant so the child has natural adjustment to the world around him or her but after the age of three the left brain er takes dominance and then the child needs to be socially conditioned and er has developed socially conditioned self er socially conditioned ego and then when the ego er comes into play then the child er begins to have conflict between his er natural character
E: mm
S: or natural behavioural tendencies and those expected of him or her socially so erm this might reflect in the child's judgement or adult's judgement behaviour and way of thinking and now erm it creates conflict in the individual to a certain extent but erm er the <unclear=individual's> er not really his or her true self but it's er is actually role playing you know it's erm er creates a a self that's in tune with the conditions erm to take <unclear=a> society now I believe that some rules are necessary like for example paying taxes to the government or observing road rules but some rules are really ridiculous for example in our society in India we say er a ch= a girl must get married by the time she hits her twenties I think that's ridiculous or if for example you don't have to serve before dinner I think those rules are not necessary and er this kind of thing which creates conflict you know leads to a bit of emotional imbalance so much so that a fragi= a fragile man can even collapse because not really sure how to behave […]

DISCUSSION
[L2 Speaker (S): Age: 14; Gender: male; Exam location: China; Proficiency: C1.1 (Advanced); Examiner (E)]

E: okay well thank you very much that was interesting yes interesting I'm a bit surprised that you chose Roosevelt as an example of a planned economy <.> erm because I associate planned economies with er Russia and China in the period
S: yeah
E: nineteen forty nine to nineteen seventy five you know
S: yeah I I also think so however I think that what China's done with planned economy is not really good it is not good enough to be a positive example
E: mm
S: and Roosevelt he used planned economy he made a great success with planned economy
E: mm
S: and China is not as great but it's er unique and special as well
E: sure but er Roosevelt never nationalised any industries did he he didn't take over Ford and General Electric and er

S: he allowed a certain level of in-inde= er a certain level of freedom <unclear=6.14>

E: mm

S: and that enable his success

E: mm

S: and I suppose that pla= a total planned economy is not very good and what he did with planned economy is er different from the economy that we are using right now it's

E: mm

S: er a bit more controlled yet I suppose er different economies

E: mm

S: suit different situations

E: mm

S: so it's hard to say that which kind of economy is better because they suits different er times and circumstances so what do you think?

E: well er <LAUGH> er I'm not sure what to think that's the problem but erm some commentators argue that the present economic crisis is just as bad as erm the crisis in the nineteen thirties I mean erm I'm very angry that er the tax payer has had to rescue the banks from bankruptcy

S: ah well that is a common problem

E: mm

S: with all governments but I suppose that sacrifice is very necessary if you don't rescue the banks how who will rescue you when you are in urgent need of money banks make money er er can can give money <.> make it more <.> current you can use it when you are in urgent need for it so with banks a country is alive and without banks if you see if just left banks away you cannot survive

E: mm well I think they should have allowed the big banks to go bankrupt and should not have rescued them what might have followed if er if the big banks had disappeared?

S: well I think that if the big banks have disappeared people will the econom= the economy of a country

E: mm

S: will go down

E: mm but the but that has happened in any case

S: yeah and but if er the tax payers

E: mm

S: sacrifice themselves for the banks

E: mm

S: so the sacrifice is great but it's not as great as it would <unclear=have done> when the banks fell apart because <.> it's so hard to explain believe me

E: sure erm in a sense China is erm succeeding during the current difficulties they are not affecting China would you agree?

S: er I suppose every country is affected yet I suppose to degree that China has been affected there's not so great

E: uhu

S: but I suppose there should be some impact as well

E: mm do you think the RMB should er the exchange rate should er should should change?

S: er well this is a very controversial point

E: that's why I ask <LAUGH>

S: er according to President Obama's idea he think that this should definitely change but according to me according from the from the degree of our country

E: mm

S: from the welfare of our country I I do not think that we should change it […]

INTERACTIVE TASK
[L2 Speaker (S): Age: 28; Gender: male; Exam location: Mexico; Proficiency: C1.1. (Advanced); Examiner (E)]

S: but do you think that technology is good for the society for for us?
E: erm well I think it's a tool I'm not sure that it's good for us but it's like any tool it's the people isn't it?
S: right I I think so er I find the technology is really useful for me er I think that that's my thinking because of my work
E: mm
S: er I teach and I like to teach with technology
E: mm mm
S: er do you use your t= the technology in your work?
E: in my teaching yes absolutely all the time
S: what kind of er techniques do you apply with technology?
E: erm well I I I use I use it in in teaching all aspects of English
S: right
E: spoken written er listening and yes so and pronunciation erm but I don't think it's always necessary to have the latest
S: right erm for er some people er it's necessary yeah but for me it's not necessary er because er my students are not er able to buy the latest technology
E: mm
S: in the world and but I think that they like to work with that do you think so?
E: erm I I think yes erm erm it's it's great it's a motivating tool erm but I suppose what I'm really talking about is you know this sort of <.> desire very strong desire people fee-feel to rush out and buy whatever is new
S: alright < LAUGH >
E: < LAUGH >
S: er well er do you think that technology is er good for students in school at home or er in their work if their work?
E: I think if they are showed how to use it yes
S: yeah <unclear=16.28>
E: but I think they're er especially young people they they don't really think about how to use it they just like
S: I think they do
E: a nice shiny new thing
S: < LAUGH > yeah young people always just thinking technology I think erm er they find it interest interesting er they can enjoy it a lot < LAUGH >
E: < LAUGH > mm
S: and for older people it sometimes is boring do you agree?
E: wh= er what technology is boring
S: do do you think technology is boring?
E: well erm let's think of an example
S: er for example to be online with the tablet and maybe two three hours like adolescents or young people
E: oh I wouldn't spend that long would you?
S: er not only twenty thirty minutes per day
E: right I'm oh really as little oh I spend more time than that
S: really < LAUGH >
E: but I'm that's because I'm searching for information or I have work to do
S: oh right […]

CONVERSATION
[L2 Speaker (S): Age: 39; Gender: male; Exam location: Spain; Proficiency: C1.2 (Advanced); Examiner (E)]

S: although the consumers are <.> upset sometimes with so many adverts on tv
E: uhu okay
S: er what do you think about it?
E: erm I find advertising annoying <.> I find it annoying
S: d= I not d= I'm not for for the same opinion sometimes the advertising are t= good are funny or entertaining
E: mm
S: they they have a a history behind the so <unclear=20.27>
E: yeah but don't you think it's it's overkill the the amount of advertising that they have
S: yeah I
E: mm for example on er if if you're trying to watch a tel= if you're trying to watch a film on tv <.> if you're trying to watch a film on tv
S: you waste a lot of time because of the adverts
E: mm
S: but what is part of our lives tv is part of our lives as well
E: mm okay but er mm your mm do you think it's absolutely necessary to have six minutes of advertisements on <unclear=21.07>
S: no of course not <.> but companies trying to take the the most then you have you are relaxed in your so <unclear=coach>
E: yeah
S: so is the the the the time of the day you are more open to
E: mm okay you were saying that some of the <.> some of the adverts are are v= interesting give me an example of
S: one er well I can talk about <unclear=21:35>
E: that's that's not normal < LAUGH >
S: well for example the there are adverts that <unclear=21.45> big producer a bit simple a-a-and people remember after years
E: mm
S: for example I I remember an advert from er phone company when the <unclear=boy> phones started <.> and it was a kid with the mobile er the idea and the idea again to many people and only saying hello how are you and Happy Christmas
E: mm
S: but all the people remember that advert
E: okay erm did that convince you to go out and buy a mobile phone?
S: sorry
E: did that convince you to go out and buy a mobile phone?
S: maybe not to buy a mobile phone but […]