CHECKLIST

How to evaluate a corpus website
"Er, is this corpus website any good? Should I use it?"

The safe and general answer will be: depends what you want from it.
But to help you think through what that means for you, ask yourself (or your group) the following questions from the checklist:

QUESTION 01

Is there information about the corpus and its content?
Can you easily find information on the website about what's in it?

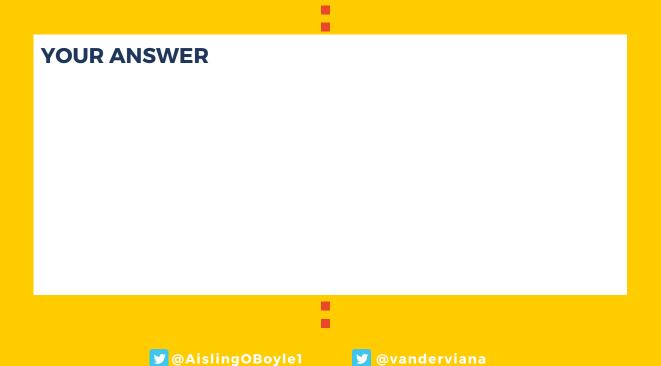| YES	NO
Good to go! You should proceed to consider the following questions in this checklist. Many corpus websites already have a guide or a link to an academic publication. Irrespective of the kind of corpus, it is crucial to have some sort of information about the corpus. Otherwise, how do you know what you are looking at!?	Don't rely on it! There is probably little mileage in considering this corpus website any further. It would be difficult to make sense of any finding if you do not know little about the corpus.

YOUR ANSWER

QUESTION 02

What kind of English language varieties are included?
E.g. American English, Scottish English?

Think about which language varieties are represented in the corpus. Most corpora will aim at representing a specific language variety rather than all varieties of English. This variety is generally (but not necessarily always) identified at national level such as American, British, Canadian and Scottish English. If you work in a context where you are requested to teach one particular variety, you might want to choose your corpus accordingly. Remember that you can use corpora to sensitize your students to a range of varieties of English.

YOUR ANSWER

@AislingOBoyle1   @vanderviana

CHECKLIST

"Er, is this corpus any good? Should I use it?"

QUESTION
03

What does the corpus represent? What's in it?
Café conversations? Physics textbooks? Student essays?

Does it include a range of genres/registers or just one? For example, is it a corpus of newspapers, TV programmes, everyday conversations, speeches, novels, lectures, manuals and so on, or is it a collection of all of these? Think about what will be most useful for your English language learning and teaching needs. For example, a corpus of newspaper texts might be helpful if you are teaching a course to journalists or if you are teaching a lesson on news stories to learners. However, this corpus would be of very little use to students of Chemistry.
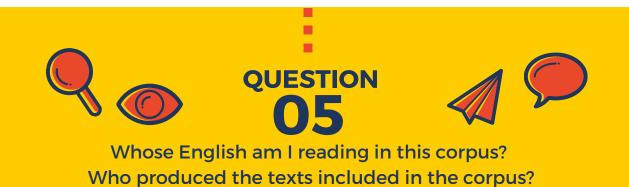
YOUR ANSWER

QUESTION
04

When were the texts included in the corpus produced?
Do you want **Good morrow** or **What's up**?

How relevant are the texts in the corpus to your needs from a time perspective? For most EFL/ESL contexts, historical corpora – i.e. corpora containing texts which do not represent present-time English language use – may be of little relevance. Most EFL/ESL students and teachers are studying the language not to be linguists but to learn how to communicate in English in our current times.

YOUR ANSWER

QUESTION
05

Whose English am I reading in this corpus?
Who produced the texts included in the corpus?

There are a number of ways to think about this question. Consider, for example, the person's educational level or socio-economic status. Is this important? Why does it matter? Perhaps the most important aspect for English language learners and teachers to consider is the language background of text producers. Most English corpora contain texts produced by speakers of English as a first language. This has been a point of contention in the pedagogical application of corpora and has been criticized for reinforcing linguistic imperialism in the classroom. As is the case with any material used for learning and teaching, you need to be aware of the affordances and drawbacks of these materials and use them sensitively in the language classroom. It is important to highlight that there are corpora which represent the use of English as an international language. These are regarded as more appropriate resources in some English language learning and teaching contexts.

YOUR ANSWER

CHECKLIST

"Er, is this corpus any good? Should I use it?"

QUESTION 06

What kind of linguistic information (if any) is available with the corpus?
Will it tell me if I'm looking at a noun or a verb?

Some corpora will contain linguistic information other than the words that were spoken/written. This linguistic information may relate to different language systems, but grammatical information is by far the most commonly available information. When a corpus has been POS (part-of-speech) tagged, it means that the words have had their classes (e.g. adjective, noun, verb, preposition) identified.

Having this type of information available may be useful in certain language teaching contexts because users will be able to search for combinations of words + word classes. Some examples of these combinations are the preposition(s) that follow the noun **raise (e.g. money, awareness)**, the adjectives which are used with the adverb **utterly (e.g different, new)**, and the adverbs that are used to qualify the verb **to yell (e.g. loudly)**.

YOUR ANSWER

QUESTION 07

When was the corpus and the website last updated?

Most corpora will be static ones – i.e. once the texts have been included, no data are added or removed. This means that it might be a while since the corpus was last updated. But you need to consider the relevance of the texts to your students, as discussed in #2. The website update date is an important indicator of how much time and effort the person in charge of it invests in ensuring that everything is running smoothly. This is especially important in this case because technology evolves quite fast.

YOUR ANSWER

QUESTION 08

Can you freely search the corpus? Really, for free?

YES

Okay, then think about how much access you are given to the original texts. Is this enough for your teaching purposes? Most corpus websites will allow users to have access to concordance lines – short snippets of text around a specific search word, and these lines should be enough for teaching purposes as the focus is on identifying language patterns. In some cases, if the focus is on meaning, perhaps an expanded co-text will be needed.

NO

Are you, the students and/or your institution able to afford the cost? Several corpora are available online for free. However, there are some where users have to pay in order to access them. Before suggesting a corpus to students and/or adopting one for teaching purposes, it is vital to consider whether either the end-users or their institutions can cover these costs.

REMEMBER: Check how many searches you and/or your students can perform in a single day. Some corpora will be available for users, but the access may be limited to a number of queries per day from any one IP. This needs to be taken into account because this limitation might make it difficult for class work to be undertaken by all the students at the same time from any single institution.

YOUR ANSWER

CHECKLIST

"Er, is this corpus any good? Should I use it?"
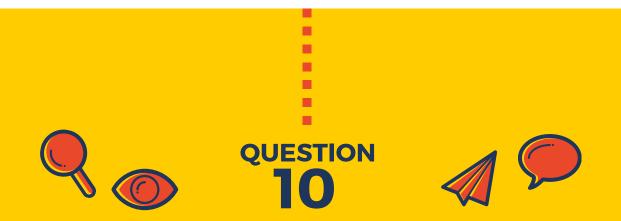
QUESTION
09

How user-friendly is the corpus website?

How comfortable are you using the website? It is important that you become familiar with how the website works. This may take a few goes. This is important so that you can guide your classmate/students/teachers through it if needed. Remember, sometimes the ugliest looking corpus, is actually really useful once you get the hang of it!

Think about others. Consider your students, will they find it easy or difficult to use the website? Newer generations (is that X? Y? Z?) of students seem to have much less trouble navigating corpus websites than their teachers. But remember- do not make too many assumptions. Learning and teaching English can take place at any age, in all sorts of contexts, with varying access and digital skills. Some may benefit from simpler and more straightforward corpus websites. These websites are likely to have limited functionalities, but this doesn't mean they aren't useful!

YOUR ANSWER

QUESTION
10

How much guidance is there for the corpus website users?

This is an important aspect in learning about the corpus and the website. The guidance available may vary in form – from explanatory texts to video demonstrations. Some corpus websites, especially those aimed for language teaching and learning, may also indicate what kind of queries may be relevant and or provide examples of potential learning and teaching activities.

YOUR ANSWER

NOW, WHAT DO YOU THINK?

IS IT ANY GOOD?

SHOULD YOU USE IT?