

Lecture 8

Corpus Linguistics

مجمع لغوي

- A corpus is a collection of language material, made in some principled way (not haphazardly), either on tape or written in hard copy (e.g. books, student essays) or in electronic form. We are concerned only with the last type.

المجمع عبارة عن مجموعة من المواد اللغوية، ركبت في بعض الطرق ذو مبادئ (ليس جزافاً)، أما في الشريط أو مكتوبة في نسخ مطبوعة (مثل الكتب والمقالات الطلابية) أو في شكل إلكتروني. ونحن نهتم فقط مع آخر نوع.

- Such collections are used in many different ways by different people. We are concerned mainly with use

ومثل هذه المجموعات تستخدم بطرق عديدة ومختلفة من قبل مختلف الناس. وأساساً نهتم باستخدام:

1) by linguists to help describe language, and test theories

(1) بواسطة اللغويين للمساعدة في وصف اللغة، واختبار النظريات

2) by teachers and learners to aid language learning (i.e. a form of CALL).

(2) بواسطة المعلمين والمتعلمين للمساعدة تعلم اللغة (أي شكل من أشكال CAAL).

- To perform any electronic corpus-based task directly you need two things - a **corpus** and a **search engine**.

لتنفيذ أي مهمة إلكترونية تستند إلى المجمع مباشرة تحتاج أمرين: - أ المجمع ومحرك البحث.

- A **corpus** itself is just text (a form of data), which may have been originally written, or be transcribed speech.

المجمع نفسه هو النص فقط (شكل من أشكال البيانات)، والتي قد يكون كتب في الأصل، أو يكون خطاب مدون.

- Corpora are not all stored in the same format (though often they are in the plainest of DOS or ASCII text), and they may have coded information (**tags**) added in and out of the text, to show e.g. who

was speaking, the register of the text, or the part of speech of each word.

المجاميع ليست كلها مخزنة في نفس الشكل (على الرغم من أنها غالباً تكون في أبسط نص من الدوس أو ASCII) وأنها قد تكون مشفرة المعلومات (العلامات) وأضيفت داخل وخارج النص، إظهار مثلاً الذي كان يتحدث، سجل النص، أو جزء من الكلام لكل كلمة.

- To use a corpus for any task you have to access it by using a **search engine** - a program which generally runs through the text (or a precompiled index to the text) and broadly does one of two things:

لاستخدام المجمع لأي مهمة لديك للوصول إليه باستخدام محرك البحث - البرنامج الذي يمتد عموماً من خلال النص (أو مؤشر المترجمة مسبقاً للنص) ويفعل على نطاق واسع واحداً من أمرين :

✓ **USERS OF CORPORA**

مستخدمي المجاميع

- **Dictionary makers** - e.g. to find out how words are actually used, and how often, and improve dictionary entries

صانعي القاموس - على سبيل المثال لمعرفة كيفية استخدام الكلمات في الواقع، وكيف في كثير من الأحيان، وتحسين مداخل القاموس

- **Descriptive grammarians** - e.g. to improve their descriptions to fit the facts of actual use of constructions

وصف النحويين - على سبيل المثال لتحسين أوصافها لتناسب وقائع الاستخدام الفعلي للتركيب

- **Stylisticians** - e.g. to see what differences there are in how frequently different authors use certain words

Stylisticians - على سبيل المثال لنرى ما هناك من الاختلافات في مدى تكرار استخدام كلمات معينة لمبدعين مختلفين.

- **Sociolinguists** - e.g. to see how frequent certain constructions are in conversation

علماء اللغويات الاجتماعيين - على سبيل المثال لنرى كيف تكرر بعض التركيب في المحادثة

- **Computational linguists** - e.g. to see if their grammatical parsing programs will work on naturally occurring language
اللغويات الحاسوبية-مثلاً لترى إذا كانت برامج التحليل النحوية ستعمل على طبيعية اللغة.
- **Language learning researchers** - e.g. to see how often learners with a particular L1 get something wrong
باحثوا تعلم اللغة - على سبيل المثال لمعرفة عدد المرات المتعلمين مع L1 خاصة الحصول على شيء خاطئ
- **Writers of teaching syllabuses** - e.g. to see how often the passive really occurs in academic English
كتاب في مناهج التدريس - على سبيل المثال لترى كيف في كثير من الاحيان ماتكون سلبية الحدوث حقا في الأكاديمية الانجليزية
- **Writers of teaching course materials** - e.g. to incorporate authentic examples into their material
كتاب في تدريس المواد الدراسية - على سبيل المثال لدمج أمثلة أصيلة إلى وضعها المادي
- **Teachers making class tasks, or even learners directly themselves** - e.g.
جعل المعلمين مهام الفئة ، أو حتى المتعلمين بشكل مباشر بأنفسهم - على سبيل المثال
- ✓ **to supply additional clues for context guessing word meaning**
لتوفير أدلة إضافية لسياق التخمين معنى الكلمة
- ✓ **for guidance on how to use word when writing**
للحصول على إرشادات حول كيفية استخدام الكلمة عند الكتابة
- ✓ **to help prompt self-correction**
للمساعدة في موجه التصحيح الذاتي
- ✓ **for word study**
لدراسة الكلمة
- ✓ **for 'language awareness' work on grammar**
"توعية اللغة للعمل على قواعد اللغة"

- **History of Corpora**

تاريخ المجاميع

<http://privatewww.essex.ac.uk/~scholp/corpintro.htm#hist>