

Tutorial

Computer-Assisted Corpus Analysis: An Introduction to Concepts, Processes, and Decisions

—SUSAN LANG (0000-0002-0858-9214), DUNCAN A. BUELL (0000 0002 4668 5848), AND NORBERT ELLIOT (0000-0002-3933-2008)

Abstract—Problem: *This tutorial aims to guide readers through key concepts, basic processes, and common decision points that inform computer-assisted corpus-based research in technical, professional, and scientific communication (TPSC). **Key concepts:** Based on our collaborative experiences and an example developed for this tutorial, key concepts of corpus analysis useful to TPSC researchers and practitioners include the following: corpus location, text preparation, and programming language and software selection. **Key lessons:** These key concepts can be used to establish basic process and decision points that, in turn, yield lessons related to the usefulness of lexicogrammatical language models and the significance of multidisciplinary. **Implications:** Although corpus research is a growing and important part of the field of TPSC, challenges remain in terms of language model variety and ethical considerations. At least in part, these challenges can be met, respectively, by alignment between corpus and analytic tools and reference to the Common Rule and related international standards.*

Index Terms—*Computer methods, corpus analysis, ethics, language models, lexicogrammatical analysis, multidisciplinary.*

教程

计算机辅助语料库分析：概念、过程以及决策的介绍

—SUSAN LANG (0000-0002-0858-9214), DUNCAN A. BUELL (0000 0002 4668 5848), AND NORBERT ELLIOT (0000-0002-3933-2008)

提要

问题：本教程旨在指导读者通过主要概念、基本流程及常见决策点来学习技术、职业及科学传播（TPSC）领域中的计算机辅助语料库研究。**主要概念：**基于我们的合作经验及一个为本教程开发的案例，对于 TPSC 领域的学者及从业者而言，有用的语料库分析中的主要概念包括语料库位置、文本准备及编程语言和软件的选择。**主要经验：**这些概念可以被用来制定基本流程和决策点，继而提供关于词汇语法的语言模型的有用性及跨学科的重要性的相关经验。**启示：**尽管语料库研究正在 TPSC 领域中蓬勃发展并日益重要，但是在语言模型的多样性及伦理考虑方面仍然存在挑战。至少在一定程度上，这些挑战可以通过将不同语料库和分析工具协调一致及参考《通用规则》和相关国际标准来应对。

索引词：计算机方法；语料库分析；伦理；语言模型；词汇语言分析；跨学科

—Translated by Chenxing Xie

Tutorial

Análisis de corpus asistido por computadora: introducción a los conceptos, los procesos y las decisiones

—SUSAN LANG (0000-0002-0858-9214), DUNCAN A. BUELL (0000 0002 4668 5848) Y NORBERT ELLIOT (0000-0002-3933-2008)

Resumen—Problema: *Este tutorial tiene por objeto guiar a los lectores a través de los conceptos clave, los procesos básicos y los puntos de decisión comunes que rigen la investigación basada en corpus asistida por computadora en la comunicación técnica, profesional y científica. Conceptos claves:* Sobre la base de nuestras experiencias colaborativas y de un ejemplo desarrollado para el presente tutorial, los conceptos claves del análisis de corpus útiles para los investigadores y profesionales del ámbito de la comunicación técnica, profesional y científica incluyen los siguientes: *ubicación del corpus, preparación del texto y selección del lenguaje de programación y el software. Lecciones claves:* Estos conceptos claves se pueden utilizar para establecer procesos básicos y puntos de decisión que, a su vez, proporcionan lecciones con respecto a la utilidad de los modelos de lenguaje léxico-gramaticales y la importancia de la multidisciplinariedad. **Implicaciones:** Aunque la investigación de corpus representa una parte creciente e importante del campo de la comunicación técnica, profesional y científica, todavía existen determinados desafíos en términos de variedad de los modelos lingüísticos y consideraciones éticas. Estos desafíos pueden, al menos en parte, resolverse, respectivamente, mediante la alineación entre el corpus y las herramientas analíticas y la referencia a la Common Rule y a las normas internacionales conexas.

Índice de términos: *métodos informáticos, análisis de corpus, ética, modelos de lenguaje, análisis léxico-gramatical, multidisciplinariedad*

—Translated by Bruno Salgado