



Language as a Detective Story

The elements of written language – words, punctuation marks, and sentences –¹ cannot not only be used to write thrillers, but they can also become the source of detective stories, as the well-known British author JK Rowling and two linguists at EURAC can testify.

by Egon Stemle und Alexander Onysko

In early summer this year, the novel “The Cuckoo’s Calling” of a certain Robert Galbraith was well received by critics. It was not a bestseller. Still, some readers were puzzled: the book was too good to be written by an author of whom nobody had ever heard before. Especially², since there were no previous texts, no photos, and no evidence, that³ the author had a background as a writer. One surprising detail in the book, for example, was the accurate description of women’s clothing, which didn’t⁴ fit well with the author’s supposed former occupation as a military policeman.

People became suspicious, and a reader wrote in an online review, “This book is so well written that I suspect that some years down the road we will hear the author’s name is a pseudonym of some famous writer.”

One thing led to another, and the Sunday Times editor Richard Brooks eventually confronted the publisher, but not before he investigated the similarities between Galbraith and Rowling. Patrick Juola, a professor of computer science at Duquesne University in Pennsylvania, was called in to analyze the text. “The idea of looking at people’s language to know who they are goes back to the Book of Judges,” Juola says, where the pronunciation of the word *shibboleth* was used to identify an enemy tribe in the biblical story of the Ephraimites. Today, the identity of the author is unmasked with the help of a computer recognizing the similarities in the writings of Galbraith and Rowling.

Not quite as thrilling but methodologically similar and inspired by a related question, linguists from the EURAC Institute for Specialised Communication and Multilingualism try to automatically determine the first language background of learners of English. Using an electronic collection of English test essays written by learners of eleven different language backgrounds (Arabic, Chinese, French, German, Hindi, Italian, Japanese, Korean, Spanish, Telugu, or⁵ Turkish), the researchers, Egon Stemle and Alexander Onysko, make predictions on the language background of the learners. In automatic classification, each of the learner groups uses specific features, which are brought to the surface by computational calculations. Some of these identifying features, which can be thought of as a linguistic fingerprint, are transferred from the learners’ first languages. The results of this research have implications for language teaching as they can provide targeted feedback to language learners. Infact⁶, some of the peculiar characteristics of English texts by German-speaking and Italian-speaking learners are hidden in this text –¹ would you’ve⁴ guessed that?

You can find more information in the comments next to the text.

By the way, Galbraith also had the same agent and the same publisher as Rowling so that the true identity of the author would probably⁷ have been revealed in any case. However, Juola collected a lot of evidence pointing to Rowling as the author and nothing specifically to the opposite. Finally, the Sunday Times came out asking, “Did J.K. Rowling write *The Cuckoo’s Calling*?” Less than a day later, Rowling confirmed through a spokesman that she had indeed written the novel –¹ and the rest is history. ☁

1 Im Deutschen kann ein Binde- oder Gedankenstrich gesetzt werden um Phrasen und Satzteile einzufügen oder um Gedanken zu verbinden. Obwohl diese Funktion auch im Englischen gegeben ist, wird von ihr dort üblicherweise weniger Gebrauch gemacht, sodass die Verwendung in Lernertexten ein Zeichen für den Erstsprachhintergrund Deutsch sein kann.

2 Das Wort „especially“ kommt in englischen Texten, die von Lernenden mit Deutsch als Erstsprache geschrieben werden, signifikant öfter vor als bei Personen mit anderen Sprachhintergründen. Sehr wahrscheinlich ist dies deshalb der Fall, da das deutsche Wort „speziell“, das sich aus demselben Ursprungswort gebildet hat wie „especially“ (also als kognat oder urverwandt bezeichnet werden kann), im Deutschen sehr häufig ist und als Diskurspartikel verwendet wird. (Diskurspartikel sind Wörter, die das Gespräch steuern und keine eigentliche semantische Bedeutung haben, wie „also“ oder „ähm“).

3 Die Regeln der deutschen Grammatik verlangen vor der Konjunktion, die einen Finalsatz einleitet („dass“), ein Komma. Dasselbe gilt aber nicht für die Konjunktion „that“ im Englischen. Deswegen erhöht sich die Wahrscheinlichkeit, dass Englischlernende mit Erstsprache Deutsch dieses Muster ins Englische übernehmen.

4 Gli italiani tendono ad utilizzare regolarmente i clittici - cioè le forme abbreviate - nell'ortografia standard anche se il loro uso in inglese dovrebbe essere limitato a un contesto più informale. È il caso di “it’s” al posto di “it is”.

5 Während “or” im Englischen nur im Wortsinn der Exklusion verwendet wird, kann die entsprechende deutsche Konjunktion „oder“ freier verwendet werden, um etwa bei Aufzählungen Wörter zu verbinden, ähnlich dem „und“.

6 “Infact” può segnalare che chi scrive parla italiano perché questa espressione è una traduzione del termine “infatti”, simile per significato all’inglese “in fact”.

7 Rispetto agli studenti con altre madrelingua, gli italiani che imparano l’inglese tendono ad eccedere nell’uso di “probably” perché il corrispondente “probabilmente” è molto frequente nella lingua italiana. “Probably” e “probabilmente” sono tecnicamente parole imparentate, cioè parole che hanno la stessa origine.



<http://languagelog.ldc.upenn.edu/nll/?p=5315>