

Plagiarism

RUI SOUSA-SILVA

Defining Plagiarism

General language dictionaries define plagiarism as “the process or practice of using another person’s ideas or work and pretending that it is your own” (Cambridge University Press & Assessment, 2025) or as “the action or practice of taking someone else’s work, idea, etc., and passing it off as one’s own; literary theft” (Oxford University Press, 2025). These definitions have offered the public a fundamental understanding of plagiarism, shaping their general perceptions of it as an illegitimate use of another person’s intellectual property. At the institutional level, definitions of plagiarism may vary slightly, depending on the nature of the material lifted (most definitions primarily focus on text, while others also consider other forms of work, ideas, designs, methods, etc.), or on the degree of intentionality of the suspect. Conversely, they unanimously agree that the submitted work must be substantially or partially produced by someone else, while failing to acknowledge the original sources or acknowledging them deficiently to be plagiarism. Given the serious implications of plagiarizing, a precise definition is required. Plagiarism can be systematically defined as intentionally or knowingly reusing words, works or ideas from another source without a proper, clear, and unambiguous acknowledgment (Sousa-Silva, 2021, p. 577). Consequently, a suspect work must meet a minimum of three criteria to be considered plagiarism.

The primary and necessary criterion for plagiarism is the reuse of an original work from another source. If the suspect work does not originate from an external source, plagiarism cannot be asserted. Self-plagiarism, which involves the recycling of one’s own previous work without acknowledgment, falls under this category: since the suspect author draws upon their own work rather than an external source, the reuse remains illegitimate and can be interpreted as a lack of integrity, but it cannot be considered plagiarism. The originality of a source depends on various factors, including its purpose and genre. For instance, terms and conditions are meticulously crafted to adhere to established legal criteria binding two parties. They are highly formulaic and adhere to a structured pattern, allowing for the expected lack of singularity. Conversely, scientific, academic, and literary works, to name just a few, are expected to exhibit originality.

The second criterion for plagiarism, in addition to the first criterion, is the absence of acknowledgment of the sources in one or more of the following forms: (a) the original sources are not cited (i.e., references are omitted entirely); (b) citations are included but fail to clearly identify the (parts of) the work that belong to the plagiarist and those that belong to the sources; and/or (c) sources are cited ambiguously, leading the reader to believe that the work belongs to both the plagiarist and the author cited simultaneously. This suggests a collaborative effort or a shared idea between the plagiarist and the cited author.

The third criterion for determining plagiarism is the author’s intention to deceive. An author acts intentionally when they either make a deliberate attempt to deceive readers or act carelessly, by not taking action to prevent the illegitimate reuse of their sources. Intention is a key factor, particularly in cases involving students who are still learning academic writing skills and may not yet

The Encyclopedia of Applied Linguistics. Edited by Carol A. Chapelle
(Section Editor: Nicci MacLeod).

© 2026 John Wiley & Sons, Ltd. All rights reserved, including rights for text and data mining and training of artificial intelligence technologies or similar technologies. Published 2026 by John Wiley & Sons, Ltd.
DOI: 10.1002/9781405198431.wbeal0914.pub2

fully comprehend the intricacies of academic writing (Howard, 1995). Therefore, they should not be punished on the same grounds as intentional plagiarists.

Although plagiarism can exist in literary and nonliterary works, recipes, films, music, or works of art, among others, language plays a pivotal role in both producing and detecting plagiarism, making it a subject of applied linguistics.

Plagiarism, Law, and Language

Plagiarism and the law are inextricably linked, despite the prevalent yet erroneous notion that plagiarism is immoral but not illegal (Garner, 2009, p. 1267). The Bern Convention for the Protection of Literary and Artistic Works (1971), initially adopted in 1886, has facilitated the adoption of provisions for the protection of works and the rights of their authors in the signatory countries, in the form of copyright laws. While copyright infringement and plagiarism are not synonymous, there is a degree of overlap between the two concepts (Pecorari, 2012).

Copyright laws typically encompass the protection of both moral rights and financial rights. Moral rights confer upon the author the right to the integrity of their work, while financial rights grant the exclusive right to financially explore it (Pereira, 2003, p. 19). A notable distinction between the two is that financial rights are transferrable, whereas moral rights are not. Furthermore, financial rights are limited in time, so that publications are not financially protected when a specific period elapses. Conversely, moral rights are inalienable, nontransferable, and perpetual, ensuring that authors permanently retain the rights to the integrity of their work. Journal publications are illustrative of this distinction: when publishing an article, authors typically cede their rights to the publisher, granting permission for actions such as copying, storage, and publication, as well as rental, sale, transmission, and distribution of the work, but are not authorized to alter the work as they deem fit. The distinction between financial and moral rights underscores the fact that plagiarism can persist even when financial rights are not infringed.

Cases of nonacademic plagiarism are more frequently brought to court due to the potential for financial remedies (Ascensão, 1992), but instances of academic plagiarism tried by courts of law abound. For instance, in 2019, the mayor of a Portuguese municipality was charged with plagiarism in his PhD thesis (Sousa-Silva, 2022). The prosecution asserted that the suspect had copied from 16 sources, but the forensic linguist working on the case determined that he had plagiarized from over 50 different sources. The mayor's conviction by the lower court demonstrates that plagiarism is not merely an academic or ethical issue that can be resolved solely by academic institutions, but also a legal matter.

A crucial exemption from copyright infringement when reusing another source's work is fair use, which consists of a "reasonable and limited use of a copyrighted work without the author's permission" (Garner, 2009, p. 676). Use of quotations from a book in a scientific publication or book review counts as fair use. Another exemption is parody, which can serve not only as a form of mockery, ridicule, or irony but also as a form of creation, recreation, and self-reflexivity, by giving new life to an older (creative) work (Hutcheon, 2000). Determining whether a work constitutes plagiarism or fair use hinges on four major criteria (Garner, 2009, p. 676): the purpose of the use; the nature of the work subject to copyright; the extent of the work reused; and the financial impact of the reuse. An example of fair use involved TS Eliot (Coulthard et al., 2017, p. 174). When accused of plagiarizing Shakespeare in one of his poems, Eliot argued that any educated reader would readily recognize the extract and, therefore, acknowledgment was unnecessary given the nature and purpose of his own poem.

A limitation imposed on both plagiarism and copyright infringement is that the law protects the expression of an idea, but not the idea itself (Pereira, 2003). To be granted legal protection and subsequently prevent a derivative work from accusations of counterfeiting, it is not sufficient to have an idea, as ideas are intangible and therefore nearly impossible to prove. Instead, an author

must demonstrate that they originated the idea, and therefore the idea must be fixed in a tangible form, such as in writing, or in visual, audio, or oral support.

Plagiarism Detection and Analysis

The prevailing public perception that any work can be susceptible to plagiarism has been largely shaped by media coverage of high-profile cases involving prominent figures such as politicians, journalists, chefs, musicians, and artists, who were perceived as gaining an unfair advantage over others. The media attention drawn to these cases has contributed to raising public awareness about the issue but has also often led to the erroneous assumption that any text that bears a superficial resemblance to a pre-existing work constitutes plagiarism. To avoid making false allegations against innocent people, precise criteria are required when establishing plagiarism.

Establishing plagiarism typically involves two approaches: detection and analysis. Although the two terms are often used interchangeably, they differ significantly in scope and objectives. In a globalized world, controlling the exploitation of a published work is challenging because of its ubiquity. Any published work can fall victim to plagiarism at any time and in any location, regardless of the author's awareness. Consequently, instances of plagiarism are frequently identified through intuition. Experienced educators possess the ability to intuitively detect plagiarism. For instance, Johnson's (1997) suspicion that the work submitted by three of her students was the result of collusion was triggered by her initial intuition. Naturally, educators possess professional reading skills, are familiar with their students' abilities, and are knowledgeable about their writing proficiency. But while ordinary readers may lack the expertise of educators, they often detect plagiarism through chance encounters. In 2006, a newspaper reader experienced a sense of *déjà vu* after reading a particular article. After careful consideration, he realized that he had encountered the same content on Wikipedia (Sousa-Silva, 2022). To enhance the likelihood of detecting plagiarism with a probability greater than chance (or intuition), specialized systems commonly referred to as "plagiarism detection software" have been developed. These systems systematically screen written content for potential instances of plagiarism, irrespective of any prior suspicions raised (Maurer et al., 2006).

Detecting plagiarism is a relevant initial step, but not the most important. Suspect texts often yield false positives, making it risky and unethical to publicly accuse a possibly innocent person of plagiarism without conducting a thorough examination of the suspect texts (Reich, 2009, p. 188). This is the primary objective of plagiarism analysis. Establishing plagiarism is thus more intricate than simply identifying matching text and establishing a direct comparison. To address this complexity, forensic linguists have deployed various criteria to assist, not only academic institutions but also courts of law in determining cases of plagiarism.

One criterion is directionality. In plagiarism disputes, establishing the directionality of the lifting is key to determine who committed the offense. This process is straightforward when the two texts are published on two different occasions, as the original text is typically the one published first. Conversely, establishing the directionality can be difficult in cases of contemporary production (Turell, 2004). To ascertain this direction, forensic linguists rely on linguistics principles, such as the concept of idiolect and linguistic uniqueness. The solution proposed by Turell, based on Johnson (1997), to establish the directionality of the lifting is to first compare the plagiarized and plagiarizing texts with a third, non-suspect text to determine whether the plagiarized and non-suspect texts share more similarities with the plagiarized text than with each other (Turell, 2004).

The methods employed by forensic linguists to detect and analyze plagiarism vary widely, depending on the plagiarism strategies used (Sousa-Silva, 2021). These methods primarily follow two main approaches adopted by computational linguists: external and intrinsic plagiarism detection (Oberreuter et al., 2011).

External plagiarism detection consists of enabling a direct comparison between the suspect text and the original source, so they are employed when both texts are known. This approach is

prevalent in cases of literal, word-for-word plagiarism, where the original is copied with minimal alterations, so the source can easily be found. Intrinsic plagiarism methods, on the contrary, are employed when source texts are unavailable for comparison. In such cases, an internal analysis of the text is conducted to identify shifts and inconsistencies in style indicative of multiple authorship. This method proves useful, for instance, in detecting AI-generated plagiarism, where a comparative analysis is not feasible.

Plagiarism strategies have undergone significant transformations over the years. As text-matching software and online search engines became more sophisticated, detecting word-for-word plagiarism became easier. Consequently, plagiarists have resorted to alternative approaches. One such alternative is translingual plagiarism, where the plagiarist copies text in one language, translates it into another, and subsequently passes it as their own. Detecting translingual plagiarism presents a distinct challenge because the plagiarized and plagiarizing texts are in two different languages, rendering text-matching techniques ineffective. The method proposed to address this strategy involves identifying uncommon linguistic features in the suspect text that may be indicative of originating from another language or language family. The suspect text can then be machine translated into the anticipated language and subsequently submitted to a text-matching software or online search engine (Sousa-Silva, 2014).

Another challenging task involves analyzing plagiarism in two translations of the same original. Translated texts are constrained by the original, limiting the translators' individual choices to those of the source text. Therefore, the similarity between target language texts tends to increase. For the analyst, determining whether similarity is attributable to translational options or plagiarism becomes more challenging, but not impossible. In her analysis of plagiarism in a translation of Shakespeare's *Julius Caesar* into Spanish, Turell (2004) discovered that, in the original text, the sentence "Good Morrow, Brutus" was translated—whether inadvertently or by conscious choice—as "Buenas noches, Bruto" ("Good evening, Brutus"). The highly individualizing sentence was reused by the suspect translation, despite all previous translators adhering to a literal interpretation of the original text.

A major challenge to plagiarism detection is generative artificial intelligence (AI). The launch of ChatGPT in 2022 enabled ordinary users to generate text on virtually any topic instantly and at no cost. Since AI-generated texts are often random and closely imitate human writing (Bender et al., 2021), they are likely to be unique and therefore overlooked by text-matching software. This is especially the case when generative AI is combined with paraphrasing software (Weber-Wulff et al., 2023). An intrinsic forensic linguistic analysis of this form of plagiarism is able to identify those instances by focusing on the systematic utilization of highly patterned language (Sousa-Silva, 2024), even when language models attempt at mimicking individual writing styles.

Plagiarism and Technology

Technology and plagiarism are closely interconnected. As argued by Coulthard et al. (2017), the technologies that facilitate plagiarism can also be used for plagiarism detection. Most educators have utilized rudimentary tools, such as online search engines, to identify plagiarized texts or, occasionally, the text comparison feature of word processors to ascertain textual similarity. However, the most widely employed tools are text-matching software, commonly referred to as "plagiarism detection software." These software applications possess the advantage of systematically screening written assignments for plagiarism, thereby eliminating the reliance solely on intuition. They operate by comparing texts submitted by the user against works in proprietary digital libraries and databases, as well as open-access and academic repositories, to identify identical or similar sequences of text. Subsequently, the software calculates the percentage of overlap of the suspect text with other text(s).

One notable limitation of this software is that it is only capable of detecting texts that are accessible within the databases used. Thus, if text is extracted from a source that is exclusively available in hard copy or in other databases, the software is likely to overlook it. Moreover, when substantial alterations are made, such as in word order, chains of words are broken and so the software is likely to misidentify the lifting. In both instances, false negatives are generated. Additionally, unless explicitly specified, quoted material will be included in the text-matching analysis, which leads to false positives.

Publicly accessible large language models (LLMs) and generative AI systems have raised concerns about new forms of plagiarism that are missed by traditional similarity detection systems, prompting the development of AI-generated plagiarism detection tools (Odri & Ji Yun Yoon, 2023). However, these tools have generally yielded disappointing results, as exemplified by GPTZero, which erroneously classified the American Constitution as AI-generated (Edwards, 2023). While other generative AI tools have demonstrated moderate success in detecting generative AI plagiarism, such systems are susceptible to failure, particularly when employed in conjunction with other tools such as paraphrasing software (Weber-Wulff et al., 2023). Furthermore, as writing assistants powered by AI become increasingly prevalent, including in educational institutions, generative AI detection systems are prone to flagging texts that utilize writing assistants as AI-generated. Only redesigned plagiarism detection tools that are grounded on linguistic analyses can provide a fair and just plagiarism detection service.

SEE ALSO: AI in Academic Writing; Authorial Voice in L2 Writing; Ethics in Research; Forensic Linguistics: Overview; Technology and Language: Overview; Authorship and Stylistics

References

- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). *On the dangers of stochastic parrots: Can language models be too big?* Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610–623). <https://doi.org/10.1145/3442188.3445922>
- Cambridge University Press & Assessment. (2025). Plagiarism. In *Cambridge dictionary (online)*. Cambridge University Press & Assessment. <https://dictionary.cambridge.org/dictionary/english/plagiarism>
- Coulthard, M., Johnson, A., & Wright, D. (2017). *An introduction to forensic linguistics: Language in evidence*. Routledge.
- Ascensão, J. O. (1992). *Direitos de autor e direitos conexos*. Coimbra Editora.
- Edwards, B. (2023). Why AI writing detectors don't work. *Ars Technica*. <https://arstechnica.com/information-technology/2023/07/why-ai-detectors-think-the-us-constitution-was-written-by-ai/>
- Garner, B. A. (2009). *Black's law dictionary* (9th ed.). West.
- Howard, R. M. (1995). Plagiarisms, authorships, and the academic death penalty. *College English*, 57(7), 788–806. <https://doi.org/10.2307/378403>
- Hutcheon, L. (2000). *A theory of parody: The teachings of twentieth-century art forms*. University of Illinois Press.
- Johnson, A. (1997). Textual kidnapping – A case of plagiarism among three student texts? *The International Journal of Speech, Language and the Law*, 4(2), 210–225.
- Maurer, H., Kappe, F., & Zaka, B. (2006). Plagiarism—A survey. *Journal of Universal Computer Science*, 12(8), 1050–1084.
- Oberreuter, G., L'Huillier, G., Ríos, S. A., & Velásquez, J. D. (2011). Approaches for intrinsic and external plagiarism detection: Notebook for PAN at CLEF 2011. *Notebook Papers of CLEF 2011 LABs and Workshops*.
- Odri, G.-A., & Ji Yun Yoon, D. (2023). Detecting generative artificial intelligence in scientific articles: Evasion techniques and implications for scientific integrity. *Orthopaedics and Traumatology: Surgery and Research*, 109(8), 103706. <https://doi.org/10.1016/j.otsr.2023.103706>

- Oxford University Press. (2025). Plagiarism. In *Oxford English dictionary*. Oxford University Press. <https://www.oed.com/search/advanced/Meanings?textTermText0=plagiarism&textTermOpt0=WordPhrase&tl=true>
- Pecorari, D. (2012). Plagiarism. In C. A. Chapelle (Ed.), *The Encyclopedia of applied linguistics* (1st ed.). Wiley. <https://doi.org/10.1002/9781405198431.wbeal0914>
- Pereira, A. L. D. (2003). Problemas actuais da gestão do direito de autor: Gestão individual e gestão colectiva do direito de autor e dos direitos conexos na sociedade da informação. In *Estudos em Homenagem ao Professor Doutor Jorge Ribeiro de Faria – Faculdade de Direito da Universidade do Porto* (pp. 17–37). Coimbra Editora.
- Reich, E. S. (2009). *Plastic fantastic. How the biggest fraud in physics shook the scientific world*. Palgrave MacMillan.
- Sousa-Silva, R. (2014). Detecting translingual plagiarism and the backlash against translation plagiarists. *Language and Law/Linguagem e Direito*, 1(1), 70–94.
- Sousa-Silva, R. (2021). Plagiarism: Evidence-based plagiarism detection in forensic contexts. In M. Coulthard, A. May, & R. Sousa-Silva (Eds.), *The Routledge handbook of forensic linguistics* (2nd ed., pp. 364–381). Routledge.
- Sousa-Silva, R. (2022). Forensic plagiarism detection and analysis. In I. Picornell, R. Perkins, & M. Coulthard (Eds.), *Methodologies and challenges in forensic linguistic casework* (pp. 77–92). Wiley-Blackwell.
- Sousa-Silva, R. (2024). *Detecting plagiarism in the era of generative AI: a forensic linguistics approach* [Conference presentation]. 5th European Conference of the IAFL: Applying Linguistics, Improving Justice, Birmingham, UK.
- Turell, M. T. (2004). Textual kidnapping revisited: The case of plagiarism in literary translation. *The International Journal of Speech, Language and the Law*, 11(1), 1–26.
- Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity*, 19(1), 26. <https://doi.org/10.1007/s40979-023-00146-z>
- World Intellectual Property Organization. (1971). *Berne convention for the protection of literary and artistic works*.

Suggested Readings

- Angèlil-Carter, S. (2000). *Stolen language?: Plagiarism in writing*. Longman.
- Pecorari, D. (2008). *Academic writing and plagiarism: A linguistic analysis*. Continuum.
- Sousa-Silva, R. (2019). Plagiarism across languages and cultures: A (forensic) linguistic analysis. In S. Brunn & R. Kehrein (Eds.), *Handbook of the changing world language map* (pp. 1–21). Springer International Publishing. https://doi.org/10.1007/978-3-319-73400-2_191-1
- Turell, M. T. (2008). Plagiarism. In J. Gibbons & M. T. Turell (Eds.), *Dimensions of forensic linguistics* (pp. 265–299). John Benjamins.
- Woolls, D. (2012). Detecting plagiarism. In L. Solan & P. Tiersma (Eds.), *The Oxford handbook of language and law*. Oxford University Press.