ON TRANSLATED PLAGIARISM IN ACADEMIC DISCOURSE

Diana Yankova
New Bulgarian University, Sofia, Bulgaria

Abstract

Cross-language plagiarism is increasingly being accorded the interest of academics, but it is still an underresearched area. Rather than displaying linguistic similarity or identity of lexemes, phrases or grammatical structures within one language, translated plagiarism is viewed as the theft of ideas involving two languages. Two instances of translated plagiarism will be discussed - lifting a text from language A, translating it in language B to reuse it as one’s own text, and back-translation: lifting a text verbatim from language A, translating into language B and then re-translating back into language A. The emphasis will be on non-standard structures and inappropriate linguistic choices violating source language norms which could go some way towards assisting in the detection of translated plagiarism, a task heretofore not resolved either by linguists or by computer specialists. The topic is of seminal importance to non-English speaking academic contexts.

Keywords: academic plagiarism, back-translation, translated plagiarism, illegal text lifting detection

Article history:
Received: 24 June 2020;
Reviewed: 29 July 2020;
Revised: 8 December 2020;
Accepted: 9 December 2020;
Published: 20 December 2020

Copyright © 2020 Diana Yankova

This open access article is published and distributed under a CC BY-NC 4.0 International License which permits non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited. Permissions beyond the scope of this license may be available at dyankova@nbu.bg. If you want to use the work commercially, you must first get the authors’ permission.

Citation: Yankova, D. (2020). On Translated Plagiarism in Academic Discourse. English Studies at NBU, 6(2), 189-200. https://doi.org/10.33919/esnbu.20.2.1

Funding
This study was financed by the Alexander von Humboldt Foundation, Germany, as part of a larger project entitled “Text Plagiarism in the Social Sciences vis-à-vis Ethical Aspects and Common Practices” and realized within the framework of the Research Group Linkage Programme of the foundation in the period 01.01.2017 – 30.06.2018. Ref. 3.4 – 1062413 – BGR – IP.

Diana Yankova, PhD is Professor of Linguistics and current head of the Languages and Cultures Department, New Bulgarian University, Sofia. She teaches graduate and undergraduate courses in text linguistics, translation of legislative texts, American and Canadian culture studies. She is the author of several monographs and numerous articles on legal language with special emphasis on culture and genre-specific characteristics of Common law and Continental legislation, points of convergence between legal studies and linguistics, terminological and structural considerations in translating supranational law, approximation of legislation, teaching EALP.

E-mail: dyankova@nbu.bg https://orcid.org/0000-0003-4524-882X
Aims of the study

A topic that has fairly recently began to be accorded academic interest is that of translated plagiarism, also known as plagiarism in translation (Turell, 2008), translingual plagiarism (Sousa-Silva, 2014), cross-language plagiarism, among others. Most research so far has focused on textual plagiarism and based on Coulthard’s (2004) conception that the idiolect - the lexical, syntactic, and pragmatic use of a language, is idiosyncratic to an individual.

This would preclude the likelihood of two speakers producing two identical stretches of discourse and can therefore be detected by linguistic analysis which ascertains whether two texts are textually identical or similar. Translated plagiarism, on the other hand, cannot be detected by standard text comparison. Rather than displaying linguistic similarity or identity of lexemes, phrases, or grammatical structures within one language, translated plagiarism is considered plagiarism of ideas involving two languages.

It is proposed that translated plagiarism is investigated using Selinker’s interlanguage theory expounding that the same meaning is not expressed identically by native and non-native speakers of a language. This results in language transfer and influences the translation process. It is presumed, therefore, that such a text will display non-standard structures and inappropriate linguistic choices violating source language norms, lending a feeling of foreignness to a text. Norm deviations found in a case study of 20 student papers will be discussed according to the level of language in which they occur.

Plagiarism: definition, reasons, types

The phenomenon of plagiarism has for the past few decades become increasingly prominent in public life, and although intellectual property rights were considered initially a Western or an Anglo-Saxon concept (c.f. Flowerdew & Li, 2007, p. 162), the unlawful and uncredited borrowing of someone else’s work pretending it is your own is becoming more and more widespread across countries, disciplines, and types of users. Although not restricted only to tertiary institutions, it is of special concern for higher education establishments because of the very nature of the learning process: in many of disciplines, mainly the humanities - students are required to produce original pieces of argumentative writing based on an effective and analytical scrutiny of several academic sources incorporating and interweaving the ideas and concepts in those sources within their own writing. For novice writers this proves to be a hard task indeed since they
have not yet fully mastered the discourse features and strategies of the respective genre (e.g., an argumentative essay) and at times a very thin line can be found between own research and what would be considered borrowing without giving credit (cf. Angélil-Carter, 2000; Devlin & Gray, 2007; Elander et al., 2010).

Therefore, plagiarism is both a complex and a fuzzy concept, spawning debate, contention, and dispute in academe. A recent study (Vassileva & Chankova, 2019) of attitudes to plagiarism among university professors in Germany and Bulgaria showed differing opinions on such aspects of the phenomenon as what constitutes plagiarism, which types of plagiarism are acknowledged, how should academe react to text appropriation, etc. Numerous definitions have been put forth, but the fact is that this form of academic misconduct has proved to be an abstruse concept. It has been described as ‘the theft, or unacknowledged use, of text created by another’ (Coulthard, 2004, p. 433), a phenomenon that encompasses ‘From simple errors in citation to patchwriting and to downloading or purchasing whole essays’ (Shi, 2006, p. 264), as ‘language re-use’ (Flowerdew and Li, 2007), among others. Pecorari (2008, p.6) offers three criteria, which if present, point to textual plagiarism - linguistic relationship: a text contains words and/or ideas of another text; non-coincidental similarity between texts: words and ideas are repeated from another text; and failure to attribute relationship to another text. In addition, the third criterion is determined by the expectations of the discourse community and the understanding of the reader.

Another issue that has been widely discussed is the intercultural context of plagiarism or the cross-cultural differences in discourse culture and conventions. Arguments have been put forth that compared to Western, individualistic cultures, in more collectivist societies (e.g., in Asia) individual accomplishment is not highly esteemed and encouraged. Within such discourse communities, ideas are considered common heritage and knowledge, hence the absence of a need to credit sources (Deckert, 1993; Sowden, 2005). However, there has been a call to ‘guard against essentializing culturally conditioned views of plagiarism’ and move away from cultural stereotyping (Flowerdew & Li, 2008, p. 166) (cf. also Pecorari, 2008; Pennycook, 1996).

In discussing plagiarism, a question that always arises is whether text appropriation has been effectuated inadvertently or intentionally and a distinction has been attempted at delineating intentional and unintentional plagiarism (cf. Casanve,
Based on whether intentional deception is present or not, Pecorari (2015) distinguishes between what she terms prototypical plagiarism and patchwriting (a notion proposed by Howard, 1993), the former demonstrating intention to deceive (however difficult it is to determine intention), while with the latter there is absence of intention and is engendered by inexperience. Patchwriting is sometimes considered to be an important stage of the student learning process with positive pedagogical value and it has been recommended that novice writers and students make maximum intellectual use of it and then move beyond it (Howard, 1995, p. 796).

Academic dishonesty has been analysed from different perspectives. Ethical, legal, behavioural, technological angles have been used by researchers to explain why students cheat. A pivotal role is played by the exigencies of the current situation in learning environments: students are expected to produce more work than previously under higher pressure, with stricter deadlines, while leading more dynamic lifestyles in the present busy, high-powered, fast world. Due to greater performance expectations, less time for studying, high competitiveness, initial lack of writing skills, scantier contact classes, more group work and easy access to electronic resources, students sometimes opt for an easier solution to their academic assignments – appropriating someone else's work and submitting it as their own. In addition, a facilitating factor is the distance learning environment that is being more and more adopted by higher education institutions across Europe and North America. Mental fatigue and pure laziness on the part of students can also contribute to the practice of plagiarism (cf. Bennett, 2005; Marsden et al., 2005). Powell (2012, p. 9) has developed a detailed theoretical model to explain why plagiarism occurs, in which she lists contributing personal traits and situational factors. Personal traits include goal orientation or the desire to succeed and the fear of failure, academic integration, and degree of ethical reasoning, while situational variables can be student specific – external family, financial or time pressure, academic performance and prior learning experience, and institution specific: defiance or objection to the task and levels of satisfaction with course and teacher.

Sousa-Silva et al. (2010, pp. 6-11) analysed paraphrasing and referencing verbs as textual strategies that students apply to appropriate text. They singled out the most common textual devices for paraphrasing: replacing lexical items with resort to e.g., synonyms, antonyms, superordinates, i.e., paradigmatic semantic variation; deleting or
adding specification; adding words to reused verbatim strings; omitting and replacing elements at the lexical, syntactic or stylistic level to make the text coherent; changing the spelling and the morphological categories and characteristics of an element; changing the word order. They also noted the verbatim use or making some changes to the referencing verbs.

Some of the strategies used by students to use sources inappropriately can be detected by antiplagiarism detection software, such as Turnitin™, Grammarly, CopyCatch, CatchItFirst, SafeAssign, PlagTracker, PlagScan, Unicheck (the last three supporting Cyrillic). These software programmes, however, cannot detect all forms of plagiarism, but can only complement the work done by the linguist. More often than not an experienced lecturer can detect traditional types of plagiarism with the naked eye: some signs are easy to recognize, because they usually consist of thematic, linguistic or pragmatic discrepancies, such as orthographic deviation (font, spelling styles, capitalization), stylistic variation (paragraphs with different degree of proficiency in writing skills), variation in content (arguments not germane to the thesis), syntactic variation (sentence length, syntagmatic structures), etc.

Cross-language plagiarism

The strategy of plagiarising through translation proper is a phenomenon that is gaining momentum and is facilitated by the wealth of academic research in English easily available online. This is a practice that affects mostly non-English speaking countries and thus provides a new context and entails different variables. So far, research has mainly focussed on plagiarism within the English-speaking world and not much academic endeavour has been directed to plagiarism through translation. Several aspects of this new non-Anglophone environment need scrutiny, such as the moral and ethical concept across different cultures, the attitude of students and faculty, and the university policies, among others. Also, detection in such cases is much more difficult, since lexical, grammatical, syntactic, etc. overlap cannot be established easily or unequivocally when two language systems are involved.

The strategies to conceal text plagiarism by means of translation can be divided into two groups: translation proper and back translation. Since the predominant language of texts on the Internet and generally in academic journals is English, in most
cases the translation is from English into another language and thus translation proper
is usually resorted to in non-English speaking contexts (but not necessarily).

Back-translation involves translating a text in one language to another language
and then translating it back again into the original language. Here is an example of a
short text from a random academic article originally in Spanish:

El papel que juega la seguridad es central en las dinámicas de integración que se llevan a
cabo, sobre todo en el plano económico, y que llevan a una sensación de confianza, a un
entrelazamiento que imposibilitaría una acción armada entre los distintos Estados
inmersos en el proceso de integración. (Orozco Restrepo, (2016).

Google-translated into English returns the following:

The security role is central to the dynamics of integration They carried out, especially in
the economic sphere, and They lead to a sense of confidence, to preclude an interlace
armed action between the various immersed States the integration process.

When the English text is translated back into Spanish, it yields a different version
of the source text:

El rol de seguridad es fundamental para la dinámica de la integración. Se llevaron a cabo,
sobre todo en el ámbito económico, y Conducen a una sensación de confianza, para
evitar un entrelazado acción armada entre los distintos Estados sumergidos el proceso
de integración.

With minor editing the resultant text can be considered standard Spanish. It is
highly unlikely to be detected by antiplagiarism software and will therefore be
considered original. Since this process is automated, it is not even necessary to know a
foreign language, therefore this strategy can be employed by both native and non-native
students but is especially common with English native speakers (cf. Jones, 2009).

Detecting plagiarism through translation is a hard task indeed. Several
possibilities have been offered in aid of identifying the infraction. Jones & Sheridan
(2015, pp. 717-720) have proposed several strategies, starting with analysis of the
writing style of a student by means of collecting writing samples as proof of student
writing skills and as a point of reference for establishing the presence or absence of
plagiarism in future, as well as monitoring of the development of their writing skills.

Sousa-Silva proposes a model for detecting translingual plagiarism based on
Coulthard’s idiolect, or that speakers “make typical and individuating co-selections of
preferred words” leading to what he terms linguistic fingerprinting (Coulthard, 2004, p.
and hence able to be used in identifying the author of a given text, and Selinker’s (1972) concept of a speaker’s interlanguage whose linguistic characteristics differ with native and non-native text producers and get transferred from the source to the target text (Sousa-Silva, 2014, p. 91). His model includes five categories: borrowing transfer, convergence, shift, restructuring transfer and attrition (Sousa-Silva, 2014, p. 84). Borrowing transfer implies the addition of a new element from the source language to the target language text, and convergence is observed when the final text consists of constituents that do not belong to either the source or the target language. A shift encompasses cases when elements and values diverge from those of the target language approximating source language aspects, restructuring transfer consists of introducing source language elements into the target language text. The fifth category in the framework, attrition, is found in instances when due to influence of the source language some target language elements are lost. The detection steps he proposes are a reversal of the plagiarism process: translating the target text back into the source text, checking the translated text for nonstandard linguistic forms, and comparing the translated text with other texts using Google.

Another possible model has been developed by Şahin et al. (2014) based on quantitative data collected by means of plagiarism detection software and qualitative analysis on the micro level (lexical/morphological, phrase level, sentence/clause level, tense/mood/aspect, textual level) and features on the macro level (discoursal level, translation strategies, page layout, etc.). The authors highlighted the fact that the qualitative analysis they performed detected the more subtle instances of plagiarism which would be overlooked by detection software, underscoring the need for developing such computer programmes and the indisputable role of the human factor in detecting plagiarism.

**Case study**

My interest in the topic of academic plagiarism focuses on the form of academic dishonesty through translation of texts because of the context in which I teach, namely BA and MA students majoring in English, Business Communication and Economics at New Bulgarian University. As a rule, tuition in the former programme is in English, barring translation and interpretation classes, where for obvious reasons both source and target languages are used. With Business Communication and Economics, the assignments are in Bulgarian. Most of the continuous assessment assignments that
students are required to complete are submitted through the electronic platform Moodle where each lecturer can turn on or off plagiarism detection software for each task. When students submit assignments in English the system takes seconds to establish if they have made inappropriate use of texts without citing the relevant sources but PlagScan cannot detect dishonest practices of textual borrowing from another language.

In attempting to tackle this issue and to test some of the Sousa-Silva's model a case study was carried out at New Bulgarian University in which 20 essays suspected of plagiarism written in Bulgarian and submitted by BA and MA students were analysed. The first stage was to look for inconsistencies, which were detected on different linguistic and conceptual levels. A contrastive linguistic analysis and error analysis was performed, and the most common techniques of interlingual transfer were highlighted applying Sousa-Silva's (2014) framework.

**Orthography and punctuation:**
- using capital letters for nationalities and days of the week, which are not capitalized in Bulgarian, clearly an interference from English as in Понеделник, Май, Български when not the first word in a sentence;
- missing commas, in places where they are essential in Bulgarian, and generally used much more often than in English;
- wrong format of dates: year/month/date as typically American instead of date/month/year as the format is in Bulgarian;
- wrong rendition of place names, e.g., Kremlin as Кремлин, instead of Кремъл.

**Syntax:**
- resorting to passive voice in Bulgarian much more than appropriate;
- wrong use of possessive pronoun: неговото дело, instead of делото му;
- leaving a pronoun as a subject in the sentence - Bulgarian is a pro-drop language;
- wrong use of articles;
- hardly any instances of paraphrasing;
- no reordering of sentence structure;
- overuse of the -ing form: дали не нарушава закона, използвайки мобилния си телефон по време на шофиране (to check he was not breaking the law by using his mobile while driving). In Bulgarian it is
more common to say дали не нарушава закона, като използва мобилния си телефон;
- sentence length variation.

**Lexical devices:**
- unnaturally sounding collocations;
- not so much lexical substitution (hypernyms, hyponyms, synonyms and antonyms);
- lack of understanding of context and hence wrong choice when translating a word;
- style shift.

Another aspect that merits attention is sentence length in student essays which usually divides experienced and novice writers. It has been widely acknowledged that at the start of their tertiary level studies, in their academic essays, students employ shorter sentences (e.g., 15-17 words according to Harris (2001)) and with the progression of their studies as they master more and more the conventions of the academic genre, they resort to longer sentences and more complex ones both in terms of structure and content. Thus, at the initial stages of university study sentence length can also be considered another factor in weighing in whether it is students’ own work or language re-use.

Having detected the above language inconsistencies in some of the essays the second stage of the procedure was to translate the Bulgarian texts into English using Google translate so that we could have the text submitted by the student and the supposed text they copied from in one and the same language – in this case English. Then a search was carried out for English-language texts approximating the translations. The third stage consisted of running the ‘original’ texts and the texts with the highest incidence of overlap through PlagScan to establish the exact amount of copied text in each essay.

**Conclusions**

The small-scale case study conducted corroborates findings of previous research and demonstrates that technology alone cannot be used for the detection of deceitful text appropriation between two or more languages. Due to the lack of reliable plagiarism detection software which works across languages, all the methods for detection of translated plagiarism rely heavily on the astute eye of the university professor.
The models proposed so far all involve a combination of applying software for establishing similarities in texts, software for machine translation and the active participation of the practitioner whose skills can complement technology and compensate its fallibility. Our approach to discerning cases of plagiarism in translation is to analyse the textual strategies resorted to by students to appropriate text.

The detection process can be made simpler in future, if, and when machine translation is improved to a degree that can convert felicitously texts across languages. Pataki (2012) presented a method whereby information retrieval was used, and an algorithm was elaborated which can detect a stretch of text of 10 sentences across German-English and Hungarian-English. However, until such methods can be reliably applied to other pairs of languages we have to rely on our intuition and skills in applying contrastive linguistic analysis.

Another important aspect of plagiarism detection be it within one language or across two or more languages is a concerted effort on the part of university governing bodies and university professors for its prevention. How to curtail plagiarism is a pressing issue in current academic discourse. The practice of appropriating someone else’s work is becoming more common and more widespread geographically not only among students but also among academics. Any discussion of the phenomenon must undoubtedly start with plagiarism awareness and then prevention. This applies for both domestic and international students.

Some measures that can be appropriate in this context can be delineated in three groups. First, administrative measures, or for Universities to adopt, and more importantly enforce a policy of academic integrity with an explicit definition of plagiarism and implementing and enforcing penalties in cases of ascertained text appropriation. The ethical and legal side of the act of plagiarism should be clarified and disseminated with a clear message that it is immoral and illegal (copyright law). Second, practical measures, or developing software tools to assist in the detection of the currently time-consuming process of ascertaining translated plagiarism ‘beyond reasonable doubt’. And last but by no means least - academic measures on the part of lecturers and professors with a conscious effort geared towards developing students’ academic writing skills focussed on strategies and techniques of developing own ‘voice’, using, and integrating sources, synthesizing ideas, as well as how to reference correctly.
References


Sousa-Silva, R., Grant, T., & Maia, B. (2010, June 21-23). “I didn’t mean to steal someone else’s words!”: a forensic linguistic approach to detecting intentional plagiarism. [Conference presentation]. Fourth International Plagiarism Conference, Newcastle upon Tyne, UK.


