

# How much can you Copy?

Mary Davis and John Morley

## Published in:

*Student plagiarism in higher education: Reflections on teaching practice*

[ISBN: 9781138055155] / edited by Diane Pecorari and Philip Shaw (Routledge, 2019).

## Introduction

Copying is often associated directly with plagiarism; in a survey of 54 universities, Pecorari (2001) found the copying of words or 'verbatim repetition' to be one of the widely listed practices that constituted plagiarism according to institutional definitions. It is therefore unsurprising that a great deal of attention is paid to copying in academic writing, and the associated problems are of considerable concern to staff and students. Students frequently ask tutors 'How many words are OK?' and their tutors are likely to find this a difficult question to answer. Indeed, as Shi (2008, p. 6) concluded, 'There is no consensus on how much language must be copied to be deemed plagiarism'. Consequently, explicit guidelines about levels of acceptable copying are rare, and the tendency to equate copying directly with plagiarism can result in anxiety and confusion (Davis, 2013). Furthermore, text-matching software has encouraged educators to focus increasing scrutiny of possible instances of copying in student work (Angélil-Carter, 2000; Pecorari, 2013). Thus, it is important to examine the practice of copying from a number of perspectives, beyond just the copying of words: in this essay, we will discuss why we copy and different practices of copying, before examining possible distinctions between acceptable and unacceptable practices. In this way, we aim to raise the reader's awareness about copying and to offer some guidelines to draw lines between acceptable and unacceptable practices.

## Why do we copy?

Imitating what others do and how others behave seems to be a fundamental characteristic of human nature. In psychological sciences and in biology, unconscious behavioural imitation is known as mimicry. The tendency for individuals to imitate or mimic the behaviour of fellow humans undoubtedly had survival benefits in our evolutionary past. It is not difficult to imagine a scene in which a group of early human-like creatures are busy

collecting fruits and berries. Suddenly, one or two startled members of the group stop collecting and run for cover, emitting alarm calls. Any individual who does not run, mimicking the flight response of the predator detector or detectors, would stand a high chance of being eaten by that predator. If such situations occurred frequently in our evolutionary past, then it is not surprising that the instinct to imitate others would have been selected as a survival trait. Indeed, such collective flight behaviour is commonly observed in many mammal and bird species. Similarly, among our primate relatives today, we can observe that young apes imitate older and more experienced apes using tools to crack open nuts or to extract termites. In this very basic sense then, individuals who were effective copiers had a higher chance of survival and of passing on their genes to the next generation. On another level, it is thought that imitating others, whether verbally, facially, emotionally or behaviourally, increases rapport, linking, and empathy between interaction partners. It thus strengthens group bonds (Paukner, Suomi, Visalberghi, & Ferrrai, 2009; Stel, van Baaren, & Vonk, 2008; Stel & Vonk, 2010). This is also likely to be an evolutionary trait which enhances survival prospects, since groups that bond closely are more likely to be successful by caring for others, sharing and making a collective defence.

Imitation of verbal behaviour has been shown to occur in the very first few days of life. In one well-known study, for example, Simner (1971) showed that infants as young as two to four days old will cry in response to another infant's crying. Imitation is also considered to play a significant role in human language development. Whilst few today would agree with the simplistic behaviourist view that all language constructions must be imitated before they can be incorporated into an individual's repertoire of linguistic behaviours (for example, Mowrer, 1960), it is accepted that exposure to language forms, whether we are thinking of phonological, syntactic or phraseological patterns, is the principal driver of language development. These forms are imitated, sometimes with novel variations, as subsequent language productions in the form of vocalisations. In fact, studies show that by one year of age, infants' spontaneous utterances are imitations of the language patterns that they perceived in their environment (Kuhl, 2000). Furthermore, it would seem that verbal imitation is a continuing aspect of language development which does not end with the onset of adulthood. Adults tend to adopt each other's accents, latency to speak, rate of speech, utterance durations (Cappella & Planalp, 1981) and sentence syntax (Bock, 1986). Moreover, it is easy to observe that in adulthood, unconscious verbal imitation and conscious copying of syntactic and phraseological patterns occur quite frequently. One only has to consider how,

when new phrases are coined, they may quickly become widely used and understood within the adult population: 'cyber attacks', 'fake news', 'hard Brexit'.

Imitation as a form of learning holds an important place in developmental psychology. In fact, imitation is thought to be fundamental to how a developing human learns to live within their own culture. For the individual, it provides the resources necessary for acting and participating within, initially, their family, and later, society. One influential theory is that learning is a cognitive process that takes place in a social context and can occur purely through observation. Social learning theory (Bandura, 1977) draws heavily on the concept of modeling, or learning by observing the behaviour of others. Children pay attention to the way models behave and encode their behaviour. Later, they may imitate the behaviour they have observed, and if it is appropriate, it is likely to be reinforced by the approval of others. Other theories of human development also draw heavily on imitation as a learning process. Piaget (1962), for example, observed that during the first two years of a child's life, their learning seems to prioritise the imitation of observed actions (what he termed 'sensorimotor imitation'). Then, around the end of the second year of life, and continuing into the early school years (to about age seven), imitation enters a cognitive stage. This is where a child has a mental image of the behaviour to be imitated (what Piaget called 'representative imitation'), even if the child is not conscious of this. More conscious forms of imitative learning are well recognised: learning to play music, sing, dance, pronounce a new word or phrase, or manipulate a new grammatical structure in a second language. We might term this imitation and memorisation of 'how to do something'. Another form of conscious imitative learning is memorisation of textual material with a view to later reproduction. We will discuss memorisation in relation to one educational culture later in this essay.

As imitation is such a natural part of human learning behaviour, it is unsurprising to find students imitating others by reusing their written words, especially in the early stages of their study. Howard (1993, p. 233) defines the practice of patchwriting as 'copying from a source text and then deleting some words, altering grammatical structures, or plugging in one-for-one synonym-substitutes'. Howard (1993, 1995, 1999) has long argued that patchwriting should be considered a natural stage of development, when students do not have enough language or knowledge about a subject to write about it in their own way. Many scholars agree with her (Bloch, 2012; Pecorari, 2003; Sutherland-Smith, 2008). For example, Pecorari (2003, p. 338) asserts: 'Today's patchwriter is tomorrow's competent academic writer'.

Students' intention in using the words of others may often come from a legitimate desire to engage more deeply in the discourse of the community but there could be a mismatch between what students attempt to do, and how their actions are perceived by the community around them. A tutor-reader would only see the result of using other authors' words, but not the attempt or the process behind it, which can be a normal stage towards competence as an academic writer. Where the attempt to paraphrase is at a patchwriting level, Introna and Hayes (2004) call this 'grey plagiarism', because of reader uncertainty about whether it can be seen as acceptable or unacceptable.

Having introduced the practice of copying and imitation, discussed why it is an important evolutionary trait and begun to consider how students may use it, we will next briefly review historical and cultural practices, before examining acceptable and unacceptable copying practices of academic writing in more detail.

### **Historical and Cultural Practices**

Copying has been recognised as part of authorial practice throughout history. Many scholars of plagiarism argue that the historical context of plagiarism needs to be brought into perspective (Angélil-Carter, 2000; Bloch, 2012; Carter Simmons, 1999; Howard, 1999; Pecorari, 2008; Stearns, 1999; Sutherland-Smith, 2008). For example, Sutherland-Smith (2008) discusses the 16<sup>th</sup> century Romantic notion of authorship in which 'each author was an individual genius who created an original work' (p. 40). However, as Blum (2009) points out, Shakespeare's plays often retell other tales, and Homer produced collaborative works that were passed down orally to some extent, meaning that even historically, the absolute originality of a unique author comes into question. It is also important to consider the history of the legal perspective on plagiarism. Sutherland-Smith (2008) cites the *Statute of Anne* of 1710 which decreed that the thoughts and ideas of authors were 'property' and misappropriating them was an offence. This developed into the copyright laws used since this time. Howard (1999) points out that our insistence on absolute authorial originality and autonomy is a fairly recent phenomenon, and what is understood by authorship is contingent.

In many parts of the world, accurate copying of text through memorisation has been highly valued both as a way of transmitting knowledge and as an important part of the learning process. In this volume, Li and Flowerdew discuss Chinese cultural practices in detail, and how socio-historical factors influence educational approaches to the use of source text. We draw here briefly on the example of the Middle East, which has been less widely

studied in this respect. Nasr (1992) explains how memorisation as a way of transmitting knowledge has played and still plays a significant role in the educational philosophy and intellectual tradition of the Middle East. This can be traced back to the origins of Islam when, it is believed, the Quranic revelation was received by Prophet Mohammad orally and subsequently memorised before being transmitted to his followers, who in turn memorised the verses. The memorisation of the *Quran* (which actually means recitation) and *Hadith* (traditions) are to this day school subjects that students are tested on at the end of each academic year, and this applies to all countries in the Arab region. In fact, there are many specialist schools for the memorisation and recitation of the *Quran* in all Arab countries. These specialist schools are called *tahfiz* (literally memorisation), and they train students in memorising the *Quran* part by part until it is all memorised by heart. Memorising it all is considered a great accomplishment, which also brings the memoriser or *hafiz* (keeper) a high status in society that is celebrated by the individual's community.

The extent to which this tradition might have influenced judgements about the acceptability of copying other texts is difficult to determine with any degree of certainty. We do know that copying texts was viewed fairly positively during the Islamic Golden Era; it was common practice for libraries to allow individual scholars to hand-copy their treasured manuscripts with a view to permitting the knowledge to be shared (Majali, 1976). However, this was before the advent of printing, and even in the West, up to this time, accurate textual copying by hand was also highly valued. We also know that from the earliest times, there was large pool of oral poetry which singers were able to draw on and reproduce; at the same time, there is evidence that, even within the oral poetic tradition, many poets were proud of the fact that they were not using the material of others (Schoeler, 2006). Even if memorisation is encouraged in the educational systems of the Middle East, there is recognition that, by itself, it is neither sufficient, nor is it the goal of education. Nakosteen (1964), for example, suggests that, in addition to the principle of memorisation and repetition, what has been valued in Islamic educational institutions is reflection and application of the content. Furthermore, it is not difficult to find online discussions about the unacceptability of plagiarism in universities in the Arab speaking world (e.g. Guessoum, 2010). All this suggests that even in cultures which place a high value on textual memorisation, there are limitations to acceptable copying, which we will now explore in more depth.

### **Acceptable copying**

In this section, we will examine acceptable practices of copying in terms of the rhetorical structure of texts, and in more detail, the re-use of academic phrases. We will start by exploring the four-part typology created by Pecorari and Shaw (2012) to analyse the kinds of intertextuality that may emerge in the academic writing of novice writers who are non-native speakers. They distinguish between legitimate and illegitimate forms based on the criteria presented in the table below:

---

(INSERT TABLE 1 ABOUT HERE)

---

Legitimate forms of intertextuality include indirect and conventional forms. Conventional forms adhere to the standardised ways of referring to sources in a written piece, whether through paraphrase or quotation, with appropriate referencing. As long as the correct conventions are followed, such forms of copying are acceptable, and they are even encouraged as a characteristic of scholarly writing.

Perhaps the most interesting category of legitimate copying is what they call indirect intertextuality. In addition to the commonalities of topic or purpose which determine the relationship between many texts, two other prominent forms of indirect intertextuality can be seen in rhetorical and phraseological commonalities. At the rhetorical level, for example, it is well known that research dissertations and articles typically share a similar structure (IMRaD). In many disciplines, writers are expected to reproduce this or a similar textual structure when they present their own work. The point here is that novice writers, if they are to be successful, can be advised to copy such frameworks for their own writing. Indeed, this type of rhetorical copying is encouraged on academic writing courses around the world.

Another prominent form of legitimate, indirect intertextuality which Pecorari and Shaw give as an example operates at the phraseological level; it is the copying and re-use of key ready-made phrasal constructions. It is likely that successful language learners have understood that language has an important phraseological dimension from the earliest times, but theoretical discussions about the importance of phraseology within the field of applied linguistics only emerged in the 1970s. Observations of this phenomenon were initially concerned with speech production. It was argued that although speakers have the potential to be creative with language, only a minority of the lexical sequences that a speaker produces are original creations, and that many have in fact been produced before; it is this use of memorised phraseological elements that permits rapid speech production (Bolinger, 1976;

Pawley & Syder, 1983). Empirical analysis of language corpora using specialised software permitted analysts to identify recurrent word sequences and supported these earlier claims. Various studies were able to show that a significant proportion of both written and spoken discourse is composed of phraseological elements of various kinds (Sinclair, 1991; Altenberg, 1998; Erman & Warren, 2000). From a psycholinguistic perspective, we can say that much of the language we use is acquired, stored and retrieved as pre-constructed wholes. Furthermore, there is evidence that language acquisition of this kind takes place from the earliest year of life (Peters, 1983), and in the section about why we copy above, we have seen how acquisition of new phrases continues into adulthood. The point here is that just as the words we use in communication are not our own, many of the phrases that we use in speech or writing have been produced previously by others.

It is now widely accepted that pre-constructed phrases play an important role in academic discourse, and that to write successfully, student writers, particularly those who are writing in a second language, need to acquire these standard collocations and phrases (e.g. Biber, Conrad & Cortes, 2004; Cowie, 1992; Hyland, 2008; Nattinger & DeCarrico, 1992; Wood, 2015). Many teachers of academic writing encourage students to reproduce these phrases in their writing. Swales and Feak (2004), for example, advise graduate students that: 'Borrowing the words and phrases of others can be a useful language learning strategy. Certainly, you would not be plagiarizing if you borrowed items that are commonly or frequently used in academic English' (p. 172). The assumption here is that such phrases (see Figure 2) are common property and that, once identified, they can be 'borrowed' and reused by other writers.

Whether the form of indirect intertextuality identified by Pecorari and Shaw (2012), and recommended by Swales and Feak (2004), is considered legitimate practice among academic writers was the question asked in a recent study undertaken by the current authors of this chapter (Davis & Morley, 2015). We asked 45 academics, representing a cross section of disciplines at two UK universities, whether they felt it would be legitimate for novice writers to reuse any of a set of 32 phrases presented to them in a questionnaire. Figure 1 shows some of the phrases that 95% of the respondents thought could be legitimately reused.

---

(INSERT FIGURE 1 ABOUT HERE)

---

Our research also endeavoured to find out, through in-depth interviews with a smaller group of the respondents ( $n = 8$ ), more about what the academics felt were the characteristics which rendered a phrase legitimately reusable. Through discussions with the respondents, the following conditions for legitimate re-use emerged:

- i) when the phrases were not unique or original constructions;
- ii) when they did not contain an opinion or an evaluative element;
- iii) when they did not contain topic-specific content words.

In addition, it was felt by some of the respondents that for phrasal re-use to be legitimate, the writer needed to understand the meaning fully and to use it in an effective way in the text. So a fourth condition might be added:

- iv) when it is reused in a correct way

Interestingly, for the interview respondents, the number of words in each phrase did not appear to be among the most important conditions for legitimate re-use. It is also notable that one of the phrases in Figure 1 contains 11 words and reusing this was viewed as acceptable by 95% of survey respondents. Even longer phrases appear to pass this legitimacy test: in an earlier publication, Pecorari (2009) reports coming across a re-used phrasal 'bundle' of 12 words identified in a small corpus of topic specific research papers. Nevertheless, when we examined the strength of the relationship between the proportion of survey respondents in our study who thought a phrase could be legitimately reused with the number of words in each phrase, we found that legitimacy correlated negatively with length of phrase, i.e. more academics found the recycling of the shorter phrases to be more legitimate, and longer phrases less so (Spearman's  $\rho = -0.616$ ,  $n = 28$ ,  $p = 0.004$ , two-tailed). Furthermore, notably fewer phrases of more than nine words were considered acceptable for re-use. Thus, based on these findings, as long as the above four conditions are met, a fifth condition for re-use of general academic phrases could be added:

- v) the phrase should not normally be longer than nine words.

This could serve as a useful cut off point to advise student writers about the acceptable length of reusable academic phrases. One further implication of this indication of length is that if tutors set up assignments on *Turnitin* using the function to match strings of a low number of words, such as 5 or 6, they need to realise that such matches are likely to include academic phrases considered generic and acceptable; in other words, text-matching could indicate good practice with the re-use of suitable phrases.

Another aspect of the phrases that our respondents did not mention as being a condition for legitimate phrasal re-use is frequency. This is interesting because in the field of applied linguistics, one of the defining criteria for identifying phrasal sequences is their high frequency count in large corpora of spoken or written language. However, among applied linguists, there is no clear agreement about which frequency parameters should be used (Wray, 2008), and different results are obtained depending on the nature of the corpus being searched, the frequency measurement used, and the number of words in the phrasal sequence being searched for. In addition, this criterion may work less well for longer phrasal sequences with one or more substitutable elements, especially where these may be content words relating to a specific discipline or topic. Having said that, for the non-expert, a search of a phrasal sequence in Google, which incidentally permits substitutable elements to be discounted, can offer a quick indication of the extent to which it has been used by other writers. Indeed, this approach is recommended by Swales and Feak (2012), building on their earlier advice about borrowing phrases: 'Skeletal phrases . . . are phrases that you can adapt for your own writing, taking care that you add your original content . . . . One way to find skeletal phrases is to use Google Scholar . . . you can search for strings of language' (p. 138). For example, applying the frequency criterion to some of the phrases listed in Figure 1, we found that a Google search produces 53.5 million instances of *the findings of this study*, and 17.6 million instances of *recent evidence suggests that*. Not all phrases will produce such high counts; nevertheless, even if frequency cannot be advocated as a categorical condition, with informed use, a frequency count using Google can provide writers with additional information to help them determine the legitimacy of reusing particular phrases.

To sum up, we have found that rhetorical structures of texts tend to be copied within academic norms, and with regard to phrases, copying is acceptable where they are not original, do not contain opinion or content, are used correctly – and generally have no more

than nine words. Checking frequency counts on Google can also help with decisions about the acceptable use of phrases.

### **Unacceptable copying**

Having examined acceptable copying, we will now discuss what constitutes unacceptable levels of copying. Referring back to Table 1 in the section above, unacceptable copying broadly comes under the categories defined by Pecorari and Shaw (2012) of 'deceptive' intertextuality, where there is a clear intention to cheat and 'unconventional' intertextuality, where the writer is not following academic regulations but does not intend to cheat. Thus, unacceptable copying can mean plagiarism, but could indicate other intertextual problems. We will firstly discuss unacceptable copying with regard to academic phrases. Then we will examine where copying becomes unacceptable in terms of authorship of text and quotations, and discuss decisions about copying related to text-matching tools.

In our earlier research (Davis & Morley, 2015), we found that a small proportion (0-29%) of respondents thought the following phrases were acceptable for re-use:

---

(INSERT FIGURE 2 ABOUT HERE)

---

We concluded that there were four elements that affected these decisions: increasing length (except in the first example), a higher number of content words (defined as nouns, verbs, adjectives and adverbs which carry lexical meaning), and a more original construction, with an opinion. In the above examples, it can be seen that more than half of the words were content-based. All could be seen as specific, original phrases created by their authors, which also indicate opinions. Thus they evidently breach the conditions for acceptable re-use of academic phrases set out in the section above (apart from the condition to be used accurately, which we cannot measure when the phrase is in isolation). It is notable, however, that the first phrase was relatively short (6 words), so did not breach the length guideline (up to 9 words), but was generally considered unacceptable because of its originality and use of opinion. At the same time, our interview respondents demonstrated different views about length, especially when highlighted by *Turnitin*. For example, some explained that they had a strategy to exclude

up to six words from text matches, while others cautioned students about plagiarism if any phrases were highlighted.

Unacceptable copying also tends to be noticed by markers when it is juxtaposed to a student's own discourse. Often the first trigger to suspect plagiarism is where there is a sudden change of discourse in a text, particularly to one more sophisticated and expert, therefore very likely to come from a published source (Bull, Coughlin, Collins & Sharp, 2001). Spotting this sudden change in discourse was described vividly by a tutor participant in our previous research:

Suddenly you spot a stretch of language which just seems out of kilter with what's gone before and you just think, that doesn't sound like her, or that doesn't sound like him,...(it) will cause that sort of clunk in the reader's brain which will raise their suspicions that something dodgy is going on. (Davis & Morley, 2015, p. 30)

The decision about acting on this, however, is within the hands of tutor-reader (Pecorari, 2008). Where the reader perceives that the writing is too different, in part or wholly, from the student's own, based on their previous submissions or in class writing, one explanation is that the student has copied from sources. Here it must be acknowledged that one of the issues related to anonymous marking is that markers are not able to identify students in this way and to make any kind of comparison with other work. Some argue this may mean more students plagiarise as they cannot be identified; indeed, Newton (2015) asks whether we should limit anonymous marking as a consequence.

Another area of concern with too much copying is that it demonstrates a lazy or disengaged approach to assignments. Some students find a text, copy a section for their assignment and employ a risky strategy to change words later (Pecorari, 2013), which could result in plagiarism if they forget; this is sometimes termed 'sloppy paraphrasing' (Schick, 2011, p. 96). Furthermore, students who produce assignments by copying and pasting large sections of text, without their own input, language or ideas, seem disengaged with learning. This behaviour may be connected with 'deceptive' copying practices (Pecorari & Shaw, 2012), such as buying ready-made essays, wholesale use of tools for paraphrasing or translating, instead of producing original work. One major element of these deceptive practices is that they involve the entire assignment, bought as a whole, or where a tool has been used to produce it as a whole. These practices are thus deemed intentional acts, and considered serious breaches of regulations. One of the main ways that policy makers,

teachers and markers distinguish between poor practice and plagiarism is whether a student's actions can be seen as deliberate (Carroll, 2007).

A further dimension to copying from source texts at an unacceptable level is that it is likely to result in students lacking an authorial voice; the words and views of other authors are foregrounded and those of students are missing or hidden. Where students rely on source text for information, ideas and words up to the point Ridley (2008, p. 132) calls 'over-dependence', the amount of copying is clearly problematic. This is the case even if each copied part is formatted correctly in quotations (discussed above as 'legitimate' intertextuality), because it means students are not engaging with source text or demonstrating their own thinking about what they read. Pittam, Elander, Lusher, Fox and Payne (2009) argue that establishing an authorial stance is very difficult and students often want to play safe with quotations.

Quotations require further discussion here, since they are a very frequently used copying strategy, especially where writers struggle to paraphrase (Shi, 2008). However, excessive use of quotations can be a form of unacceptable copying. We will consider here how much copying is acceptable both in terms of the length of quotations and the overall use of quotations in a written piece of work. While there is no standard regulation in terms of appropriate maximum length for quotations, many tutors attempt to give guidance about the quantity: for example, ten words, two lines, one sentence. Tutors often have individual preferences, which could be for a certain number of quotations per assignment, or none at all, as one student explained: 'The tutor of this module said he didn't like quotation...so I didn't use' (Davis, 2012, p. 25). The more common advice from study skills books is to use quotations sparingly and keep them short (Bailey, 2014; Cottrell, 2011; Godfrey, 2016; Sowton, 2012) or to use them judiciously (Pears & Shields, 2016). Regarding quantity, Godfrey (2016) suggests a maximum of one or two short quotations per page, while Sowton (2012) recommends using 15 words as a limit above which students should question whether all the words are needed and keep only the essential elements. This recommendation is helpful, as it offers both a suggested word limit and a useful strategy to employ with decisions about quotations beyond this limit. Other general advice is not to use quotations just to fill in the word count, as this is likely to limit a student's own learning and thinking. Returning to our conditions discussed above for legitimate re-use of academic phrases, we said that phrases should not be 'unique or original constructions'. This is one of the main criteria given for using a section

of text as a quotation, such as the following advice: 'only quote if you feel you have found a powerful or unique phrase' (Godfrey, 2016, p. 32).

Interesting research into quotations has been carried out by Petrić (2012). Petrić discovered that students using long quotations for much of their assignments tended to be worried about their own language use, about misrepresenting sources or plagiarising by not paraphrasing enough, so they concluded it was better to copy source text exactly in the form of quotations. These students gained lower marks, but passed because they did not breach regulations. Furthermore, students who gained high grades made use of three times as many quotes (where the mean number of quotes per dissertation was 71.6 instances vs. 19.6 among those with lower grades) – but they made use of them as sentence fragments, not whole sentences or sections. This was in contrast to those with lower marks who relied on the clause based quotes without modifying, which she terms 'legitimate but excessive' borrowing (p. 112) that could lead to over-reliance on source text, but not a breach of regulations. From this study, we could assume that it is more important to keep quotations short than to be concerned about the total number of quotations in a text.

Regarding the overall amount of copying in a text, some attempts to establish levels of acceptability have been made. For example, Bennett (2005) categorised copied text at the sentence level as minor plagiarism, and copied text at the paragraph or whole text level as major plagiarism. Other authors have argued that it is impossible to set levels as tutors cannot form a consensus on degrees of plagiarism (for example, Price, 2009; Sutherland-Smith, 2008). There are many factors involved; it depends often on what and how it is copied, and the student's intentions. For example, sometimes students may use verbatim text from sources with citation, but not formatted as a quotation. The reader assumes it is a paraphrase, but then discovers it is an unformatted quotation. This practice may be considered acceptable by students (Chanock, 2008; East, 2005), while tutors may not agree on whether it is minor plagiarism or a more serious breach (Price, 2009). Thus it is important for both staff and students to look at examples of source use together, and draw lines between acceptable and unacceptable practices (Swales & Feak, 2004; Williams & Davis, 2017).

Universities sometimes attempt to suggest what level of copying constitutes a breach of regulations. For example, plagiarism is defined by Leeds Beckett University (2015) as: 'The substantial, unacknowledged incorporation into a student's work of material derived from the work (published or unpublished) of another'. It is notable that this definition

includes the word 'substantial'; thus it can be inferred that copying a small amount is not plagiarism, according to this guidance. Using a word that refers to quantity like 'substantial' leads to further questions about how much copying constitutes an unacceptable amount. Institutions have tended to respond to student copying with very serious measures in recent years and have often set a low bar for unacceptable copying. Badge and Scott (2009) report that the tools used in universities for detecting copying have had a profound effect on teaching and the way markers look at how sources are used in texts. With the emphasis on matched strings of words highlighted in colour in the student text, the attention of markers has shifted to the words used, rather than the ideas. This trend is described by Angélie-Carter (2000, p. 93):

Most markers, though they understand plagiarism to include the use of the ideas of others without acknowledgement, are generally fairly lenient when this occurs in student essays. They are annoyed far more by word-for-word copying whether acknowledged or not.

Yet, it clearly needs to be recognised that the text-matching tools simply match whatever strings of words are the same, including mundane details such as the address of the university, or required elements such as the assignment title, instructions or a given table of contexts. Short standard academic phrases also tend to be matched, which, as we have argued above and elsewhere, should be seen as acceptable practice (Davis & Morley, 2015). Bloch (2012) also noted that 'given the large number of 'chunks' or lexical bundles that exist in English...which words should be counted in trying to determine plagiarism can add another level of complexity' (p. 8). The prevalence of these standard phrases in student writing should be taken into account in decisions about acceptable copying when considering text-matching reports.

However, practices vary widely, and it appears that in some contexts, tutors rely on *Turnitin* scores to determine plagiarism, and set arbitrary thresholds of overall matches: for example, student scripts with 15% overall text match were investigated for plagiarism in a study by Barrett and Malcolm (2006). Bentley (2012) has argued that the overall percentage match should be removed as it 'tempts people to infer more from the headline percentage than is appropriate'. Sutherland-Smith (2008) reports teachers only looking at red (75-100% matches), orange (50-74%) and yellow (25-49%) overall matches for possible plagiarism. However, the traffic light colours of *Turnitin* should not be relied upon (Cohen, 2010; Davis, 2007); for example, green is perceived to be safe,

but as it covers the range 0-24%, it may not be, and the yellow range may match to correctly cited text. One of the problems of relying on the overall results was reported by Cohen (2010, p. 5): markers 'were tempted to just check high percentage matches without considering the implications of plagiarism occurring with lower percentage matches'. Clearly, it is necessary for markers to interpret text-matching results carefully and also to examine other factors. Cohen (2010) found students understandably concerned that their work would be judged by a number. Thus, reliance on overall similarity results can also be confusing for students, who tend to be very pre-occupied with the question 'what percentage is good or bad?' (Davis, 2007). This question demonstrates the uncertainty of students about what text-matching means, and may also result in them losing sight of the actual objectives of assessment. Some conclude 'I need to make the colour go away' (Williams & Davis, 2017, p. 98), which might even lead to them removing all the sources from their work to avoid the text match. Clearly, it is extremely important to understand how to approach text-matching and to avoid making decisions based only on overall similarity.

In contrast, we would recommend tutors and students pay more attention to high individual matches to other texts, as these are more likely to reflect unacceptable copying or at least over-reliance on certain sources, although all results should be carefully interpreted (Davis, 2007). The colour coding, ranking and percentages of individual matches on *Turnitin* can be very helpful indicators. In further research (Davis & Carroll, 2009), in order to assess over-reliance on any copied text (including quotations), the highest ranked individual source match was assessed as severe at 5%, 2-4.9% as moderate, 0.1-1.99% as minor. Using this scale, a match of 5% to an individual source may indicate unacceptable copying, which would be over-reliance if it stems from an appropriately cited quotation, and plagiarism if it is source text that has not been appropriately attributed. Conversely, it is generally accepted that individual matches of 1-2% are considered acceptable (Davis, 2007).

Ironically, the most serious kind of breach in terms of unacceptable copying could occur in an assignment which produces a zero match on *Turnitin*, a phenomenon that is currently causing educators considerable concern (Plagiarism forum, 2015). While zero matches may occur simply from a lack of material on the subject on the text-matching database, a new concern is that essay writing sites offer 'plagiarism free' essays (Wallace & Newton, 2014), by which they mean the text will have a zero match. This can be achieved, for example, by formatting, spacing, photographing, replacing a common letter

with a similar symbol, using old sources, or adapting and fabricating sources in some way. Zero matches can then hide serious cases of unacceptable copying, often of the entire assignment. Therefore, it is evident that zero matches should not be seen as an indicator of acceptable copying, or any kind of goal for students to strive towards.

## **Conclusion**

In this essay, we set out to answer a simple question: 'how much can you copy?' We have not found a simple answer. It is not surprising that educators find it very difficult to agree on exactly what plagiarism is; by the same token, they may find it hard to stipulate what level of copying is acceptable. Furthermore, detecting deceptive practice seems far from straightforward. By embarking on a reflective journey to study the human practices of copying in different domains and contexts, we have considered the factors which influence decisions about copying, and we have tried to establish where tentative lines might be drawn. We have discussed the natural, developmental, historical, cultural contexts of copying and we have then set out acceptable practices with copying and contrasted them with unacceptable copying practices. Here we have argued that copying is a large part of learning and that it plays a key role in human development and socialisation; consequently, much of what it means to be educated is derived from copying. Thus, an insistence on 'no copying' is an uninformed and unrealistic requirement; furthermore, academic writing is based on building on what is already written, whether at the linguistic or ideational level, using certain accepted conventions.

Using Pecorari and Shaw's (2012) typology of intertextuality, we have discussed indirect intertextuality in relation to academic phrases in particular. What we have found, in terms of quantity, is that general academic phrases of up to nine words are likely to be regarded as reusable, if they are not unique constructions, do not contain opinions and do not have topic specific content; individual matches of 1-2% to source texts or copying of less than one sentence do not usually constitute problems of plagiarism; quotations should generally be short sentence fragments in order to be acceptable and where they extend to beyond 15 words, care should be taken to include only the essential elements and to avoid over-reliance on source texts. We intend these figures to be general guidelines, and we emphasise that both staff and students need to think carefully about their decisions regarding copying, including what is copied, the purpose of copying and how it impacts the rest of the text. Through our exploration of why and how we copy, and discussion of acceptable and unacceptable practices in this chapter, we hope the reader has gained a

greater awareness of copying in academia, and can make more informed decisions about how much can be copied.

## References

- Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word combinations. In A. Cowie (Ed.), *Phraseology: Theory, analysis and applications*, (pp. 101–122). Oxford: Oxford University Press.
- Angélil-Carter, S. (2000). *Stolen language? Plagiarism in writing*. Harlow: Pearson Education.
- Badge, J., & Scott, J. (2009). *Dealing with plagiarism in the digital age*. Synthesis Project, Final Report. Retrieved from <http://evidencenet.pbworks.com/Dealing-with-plagiarism-in-the-digital-age>
- Bailey, S. (2014). *Academic writing: A handbook for international students*. 4<sup>th</sup> Ed. London: Routledge.
- Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.
- Barrett, R., & Malcolm, J. (2006). Embedding plagiarism education in the assessment process. *International Journal for Educational Integrity*, 2, 38–45.
- Bennett, R. (2005). Factors associated with student plagiarism in a post-1992 university. *Assessment and Evaluation in Higher Education*, 30, 137-162.
- Bentley, S. (2012). Time for Turnitin to get rid of originality percentage scores? Retrieved from <https://stevebentleyhud.wordpress.com/2012/12/04/time-for-turnitin-to-get-rid-of-originality-percentage-scores/#comments>
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at. . . : Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25, 371-405.
- Bloch, J. (2012). *Plagiarism, intellectual property and the teaching of L2 writing*. Bristol: Multilingual Matters.
- Blum, S. (2009). *My word! Plagiarism and college culture*. New York: Cornell University Press.
- Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18, 355-387.
- Bolinger, D. (1976). Meaning and memory. *Forum Linguisticum*, 1, 1–14.
- Bull, J., Coughlin, E., Collins, C., & Sharp, D. (2001). *Technical review of plagiarism detection software report*. Luton, UK: University of Luton Computer Assisted

- Assessment Centre, for Joint Information Systems Committee, . Retrieved from [http://www.jisc.ac.uk/uploaded\\_documents/luton.pdf](http://www.jisc.ac.uk/uploaded_documents/luton.pdf)
- Cappella, J. N., & Planalp, S. (1981). Talk and silence sequences in informal conversations iii: interspeaker influence, *Human Communication Research*, 7, 117-132.
- Carroll, J. (2007). *A handbook for deterring plagiarism in higher education* (2<sup>nd</sup> ed). Oxford: Oxford Centre for Staff and learning Development.
- Carter Simmons, S. (1999). Competing nations of authorship: a historical look at students and textbooks on plagiarism and cheating. In L. Buranen & A. Roy (Eds.), *Perspectives on plagiarism*, (pp. 41-51). New York: State University of New York Press.
- Chanock, K. (2008). When students reference plagiarised materials – what can we learn (and what can we do) about their understanding of attribution? *International Journal for Educational Integrity*, 4(1), 3–16.
- Cohen, J. (2010). Using Turnitin as a formative writing tool. *Journal of Learning Development in Higher Education*, 2, 1-14.
- Cottrell, S. (2011). *Critical thinking skills: Developing effective analysis and argument*. (2<sup>nd</sup> ed). Basingstoke: Palgrave Macmillan.
- Cowie, A. (1992). Multiword lexical units and communicative language teaching. In P. Arnaud, & H. Béjoint (Eds.), *Vocabulary and applied linguistics*, (pp. 1-12). London: Macmillan.
- Davis, M. (2007). The role of Turnitin in the formative process of academic writing: A tool for learning and unlearning? *Brookes e-Journal of Learning and Teaching*, 2(1). Retrieved from: [http://bejlt.brookes.ac.uk/article/the\\_role\\_of\\_turnitin\\_within\\_the\\_formative\\_process\\_of\\_academic\\_writing/](http://bejlt.brookes.ac.uk/article/the_role_of_turnitin_within_the_formative_process_of_academic_writing/)
- Davis, M. (2012). International postgraduate students' experiences of plagiarism education in the UK: Student, tutor and expert perspectives. *International Journal of Educational Integrity*, 8(2), 21-33. Retrieved from: <http://www.ojs.unisa.edu.au/index.php/IJEI/article/view/807/601>
- Davis, M. (2013). The development of source use in international postgraduate writing. *Journal of English for Academic Purposes*, 12, 125-135.
- Davis, M. & Carroll, J. (2009). Formative feedback within plagiarism education: Is there a role for text-matching software? *International Journal of Educational Integrity* 5(2), 58-70. Retrieved from: <http://www.ojs.unisa.edu.au/index.php/IJEI/article/view/614>

- Davis, M., & Morley, J. (2015). Phrasal intertextuality: The responses of academics from different disciplines to students' re-use of phrases from sources in their writing. *Journal of Second Language Writing, 28*, 20-35.
- East, J. (2005). Proper acknowledgement? *Journal of University Teaching and Learning Practice, 2(3a)*, 1-11. Retrieved from:  
[http://jutlp.uow.edu.au/2005\\_v02\\_i03a/east005.html](http://jutlp.uow.edu.au/2005_v02_i03a/east005.html)
- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text, 20*, 29-62.
- Godfrey, J. (2016). *Writing for university* (2<sup>nd</sup> ed). London: Palgrave Macmillan.
- Guessoum, N. (2010). *Plagiarism in Arab-Muslim academia*. Retrieved from:  
<http://www.irtiqa-blog.com/2010/11/plagiarism-in-arab-muslim-academia.html>
- Howard, R. M. (1993). A plagiarism pentimento. *Journal of Teaching Writing, 11(3)*, 233-246.
- Howard, R. M. (1995). Plagiarisms, authorship, and the academic death penalty. *College English, 57*, 788–806.
- Howard, R. M. (1999). *Standing in the shadow of giants: Plagiarists, authors, collaborators*. Stanford, CT: Ablex.
- Hyland, K. (2008). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics, 18*, 41-62.
- Introna, L., & Hayes, N. (2004). Plagiarism, detection and intentionality: On the (un)construction of plagiarists. In A. Smith, & F. Duggan (Eds.) *Plagiarism: Prevention, Practice and Policy Conference Proceedings*, (pp. 83-95), Newcastle: Joint Information Systems Committee - Plagiarism Advisory Service, Newcastle, 28 - 30 June.
- Kuhl, P. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences, 97(22)*, 11850–11857.
- Leeds Beckett University (2015). Plagiarism definition. Retrieved from  
<https://www.leedsbeckett.ac.uk/about/academic-regulations.htm>
- Majali, A. S. (1976). *The development in higher education in the Arab world*. London: Longman.
- Mowrer, O. H. (1960). *Learning theory and the symbolic processes*. New York: Wiley.
- Nakosteen, M. (1964). *History of Islamic origin of western education*. Colorado: University of Colorado Press.

- Nasr, S. H. (1992). Oral transmission and the book in Islamic education: the spoken and the written word. *Journal of Islamic Studies*, 3, 1-14.
- Nattinger, J., & DeCarrico, J. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- Newton, P. (2015, June). Ghostwriting. In M. Davis and I. Glendinning (Chairs), *Current challenges to assessment from the field of academic integrity: Paid third parties, international students and institutional policy* Symposium conducted at *Assessment in Higher Education*, Birmingham, 24-26 June.
- Paukner, A., Suomi, S., Visalberghi, E., & Ferrari, P. (2009). Capuchin monkeys display affiliation toward humans who imitate them. *Science*, 325(5942), 880–883.
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191-226). New York: Longman.
- Pears, R. & Shields, G. (2016). *Cite them right* (10<sup>th</sup> ed). Basingstoke: Palgrave Macmillan.
- Pecorari, D. (2001). Plagiarism and international students: How the English-speaking university responds. In D. Belcher & A. Hirvela (Eds.) *Linking literacies: Perspectives on L2 reading-writing connections* (pp. 229-245). Ann Arbor, MI: University of Michigan Press.
- Pecorari, D. (2003). Good and original: Plagiarism and patchworking in academic second language writing. *Journal of Second Language Writing*, 12, 317-345.
- Pecorari, D. (2008). *Academic writing and plagiarism: A linguistic analysis*. London: Continuum.
- Pecorari, D. (2009). Formulaic language in biology: A topic-specific investigation. In M. Charles, S. Hunston & D. Pecorari (Eds.), *At the interface of corpus and discourse: Analysing academic discourses* (pp. 91-109). London: Continuum.
- Pecorari, D. (2013). *Teaching to avoid plagiarism: How to promote good source use*. Maidenhead: Open University Press.
- Pecorari, D., & Shaw, P. (2012). Types of student intertextuality and faculty attitudes. *Journal of Second Language Writing*, 21, 149-164.
- Peters, A. (1983). *The units of language acquisition*. Cambridge: Cambridge University Press.
- Petrić, B. (2012). Legitimate textual borrowing: Direct quotation in L2 student writing. *Journal of Second Language Writing*, 21, 102-117.
- Piaget, J. (1962). *Play, dreams and imitation in childhood*. New York: W. W. Norton.

- Pittam, G., Elander, J., Lusher, J., Fox, P., & Payne, N. (2009). Student beliefs and attitudes about authorial identity in academic writing. *Studies in Higher Education, 34*, 153-170.
- Plagiarism forum (2015). Zero Turnitin. Thread started by Chris Whyley, 26/10/2015.  
Retrieved from Plagiarism@jiscmail.co.uk
- Price, J. (2009). *Plagiarism in Higher Education: Consensus and consistency when punishing student cases*. (Unpublished doctoral thesis). University of Southampton, Southampton, UK.
- Ridley, D. (2008). *The literature review*. Los Angeles, CA: Sage.
- Schick, K. (2011). Citation obsession! Get over it! *Chronicle of Higher Education, 58(11)*, 96.
- Schoeler, G. (2006). *The oral and the written in early Islam*. (U. Vagelpol, Trans., J. Montgomery, Ed.). Abingdon: Routledge.
- Shi, L. (2008). Textual appropriation and citing behaviours of university undergraduates. *Applied Linguistics, 31(1)*, 1-24.
- Simner, M.L. (1971). Newborn's response to the cry of another infant. *Developmental Psychology, 5*, 136-150.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sowton, C. (2012). *50 steps to improving your academic writing*. Reading: Garnet Education.
- Stearns, L. (1999). Copy wrong: Plagiarism, process, property and the law. In L. Buranen & A. Roy (Eds.) *Perspectives on plagiarism* (pp. 5-17). Albany, NY: State University of New York Press.
- Stel, M., van Baaren, R. B., & Vonk, R. (2008). Effects of mimicking: Acting prosocially by being emotionally moved. *European Journal of Social Psychology, 38*, 965-976.
- Stel, M., & Vonk, R. (2010). Mimicry in social interaction: Benefits for mimickers, mimickees, and their interaction. *British Journal of Psychology, 101(2)*, 311-323.
- Sutherland-Smith, W. (2008). *Plagiarism, the internet and student learning: Improving academic integrity*. New York: Routledge.
- Swales, J., & Feak, C. (2004). *Academic writing for graduate students* (2<sup>nd</sup> ed). Ann Arbor, MI: University of Michigan Press.
- Swales, J., & Feak, C. (2012). *Academic writing for graduate students* (3<sup>rd</sup> ed). Ann Arbor, MI: University of Michigan Press.

- Wallace, M., & Newton, P. (2014). Turnaround time and market capacity in contract cheating. *Educational Studies* 40(2), 233-236. Retrieved from <https://cronfa.swan.ac.uk/Record/cronfa17023> doi:10.1080/03055698.2014.889597
- Williams, K., & Davis, M. (2017). *Referencing and understanding plagiarism* (2<sup>nd</sup> ed). London: Palgrave Macmillan.
- Wood, D. (2015). *Fundamentals of formulaic language: An introduction*. London: Bloomsbury.
- Wray, A. (2008) *Formulaic language: Pushing the boundaries*. Oxford: Oxford University Press.