

In language teaching and learning, contemporary research has also shown that code-switching offers many pedagogical benefits, including accelerating students' confidence, increasing their access to content, as well as improving their participation and engagement (Ahmad and Jusoff 2009; Carstens 2016; Daniel et al. 2019; Wang 2019). This has given rise to changes in policy which promote the use of plurilingual practices in education across many parts of the world (Faltis 2019; Lin 2013; Seed 2020)¹; language is no longer commonly seen as a static entity but rather as a fluid resource in the whole meaning-making process (Pennycook 2010).

Despite this 'multilingual turn', however, current educational technologies are not yet able to keep up and are partly responsible for the constraint of this practice to only classroom interactions. In fact, although multilingualism is and has always been the norm, natural language processing (NLP) tools capable of processing more than one language per input round as in (1) and (2) are still rather limited. This effectively creates a tension between ideologies and practice, with automated feedback, scoring and grading systems breaking more often than not when processing multilingual input. Not only does this frustrate learners and teachers alike, it also hinders our progress towards truly plurilingual practice.

In this paper, we tackle this problem by zooming into difficulties of existing educational NLP tools for CSW and suggesting possible directions for future improvements. Specifically, we start with a brief background of CSW in the context of teaching English as a second language (ESL) (Section 2), followed by a discussion of difficulties of current NLP technologies for ESL teaching and learning (Section 3). We focus on two cases of feedback and assessment technologies to demonstrate how the current state-of-the-art in these domains fails with CSW input, and finally propose possible technological solutions for each of these scenarios (Section 4).

2. CSW in an ESL Context

Investigation into classroom code-switching has been diverse, with studies ranging from local language use in English classes (Anderson and Lightfoot 2021; Kim and Tatar 2017; Mahboob and Lin 2016), first language (L1) use in English instructions (Brevik and Rindal 2020; Macaro et al. 2020; Sah and Li 2020), L1-influenced errors (Lapierre 2018; Qasem 2020; Yamashita and Jiang 2010), to bilingual content and assessment (Otto and José Luis 2019; Piccardo et al. 2021; Seed 2020). These studies broadly fall into two categories: (i) bilingual education classrooms and (ii) second language contexts. Although we restrict our focus to the latter only, it remains difficult to provide a comprehensive overview of the vast range of studies in this subfield. We thus aim to present, instead, a brief background of four key aspects that we deem most relevant to the present discussion: terminologies, reasons why learners code-switch, reasons why teachers code-switch, as well as the landscape of current practice and policies.²

2.1. Terminologies

2.1.1. Code-Switching vs. Code-Mixing

Some linguists attempted to distinguish code-switching from code-mixing in the early 2000s, stating that while code-switching refers to 'the rapid succession of several languages in a single speech event', code-mixing includes 'all cases where lexical items and grammatical features from two languages appear in one sentence' (Muysken 2000, p. 1). This means that while code-switching refers to intersentential alternation of languages (as in (1)), code-mixing refers to an intrasentential pattern (as in (2)). Most linguists, however, do not make this distinction. Code-switching is typically conceived in the broadest term to cover all kinds of language mixing, leaving it to individual researchers to specify which is applicable within the framework of their respective contributions. In ESL research in particular, the term 'code-mixing' rarely comes up; however, as we see in Section 3.2, it is this intrasentential mixing pattern that presents difficult problems for educational technologies involving parsing or error detection/correction.

2.1.2. Code-Switching vs. Borrowing

The next distinction to be made is that between code-switching and borrowing—the subject of a longstanding debate in code-switching research. Some researchers have proposed that code-switching and borrowing are essentially similar phenomena lying along the same continuum of language contact, evolving from code-switches to established borrowings (Gardner-Chloros 2009; Myers-Scotton 1993; Treffers-Daller 2005; Winford 2003, 2009), while others believe that they are distinct processes (Aaron 2015; Nguyen 2018; Poplack 1980; Torres Cacoullous and Travis 2018). Whether or not they need to be differentiated, and if so how, remains largely controversial. The criteria proposed and ways to apply them in treating single other-language items also vary, ranging from frequency, diffusion and dictionary attestation to integration and many more.

Setting aside this controversy, however, researchers have generally agreed that long-term borrowing is well-integrated into the community and thus forms part of monolingual speech. In the sentence ‘I had sushi for lunch’ for example, ‘sushi’ is widely used and understood by many monolingual English speakers without any Japanese knowledge, and hence would be considered a borrowing rather than a code-switch. For cases that are less straightforward, researchers typically rely on dictionary attestation: if a single word in language A can be found in a dictionary of language B, it is considered borrowing, and hence gives language B membership alongside language A. This method is not flawless, however, first because dictionaries often lag behind contemporary usage, and second, because the criteria for warranting a word entry in a dictionary are not always explicitly explained (and therefore poorly understood).

Linguists then turn to other means to consider the extent to which the item is integrated into the ‘grammar’ of the other language (Poplack et al. 1988), as well as the level of frequency and diffusion of such items in the community relative to the sample size. Poplack and colleagues (Poplack et al. 1988), for example, established four levels of frequency and diffusion of these single-word insertions: nonce, idiosyncratic, recurrent and widespread. A nonce item is a single other-language word that is used only one time in a given corpus, while items used more than once by just one speaker are idiosyncratic. Lexical items that occur more than 10 times are recurrent, and those that are used by more than 10 speakers are widespread. Only those that are both ‘recurrent’ and ‘widespread’ are clear-cut cases of established borrowing.

In the context of ESL, frequency and diffusion are nonetheless difficult metrics, as we do not have access to learners’ ‘community’ or a large enough sample of their individual repertoire. As learner data is often collected from exams and practice exercises, their linguistic behaviour may also vary significantly from their ‘natural’ usage. The fine-grained distinction between code-switching and all those various forms of borrowing is thus perhaps less pertinent; the key point to note is that not all ‘non-English’ items represent learner’s lack of English knowledge per se and should thus be treated with extra care.

2.1.3. Code-Switching vs. Translanguaging

The ultimate discrepancy in terminology on CSW in education, nonetheless, remains that between code-switching and translanguaging. While both concepts broadly refer to language practice across boundaries, they derive from markedly different theoretical premises. In particular, while code-switching implies an external linguistic concept with rigid boundaries between different ‘codes’, translanguaging signifies one coherent language system that allows speakers to draw on various features of their named languages (García and Kleyn 2016; García et al. 2017; Orellana and García 2014). In some areas of generative and educational linguistics, the code-switching approach is being heavily criticised for supposedly supporting such separation of linguistic systems (Otheguy et al. 2015). As the translanguaging approach argues, all languages are social constructs, and those in power construed languages as social properties that could be named. In this sense, there is no such thing as a ‘language’, rather just a ‘socially named language’; what matters is thus not what the languages are called but more what people do with their languages.

This called into question the code-based approach, promoting an alternative ideology that bilingualism must be understood from a noncode, monoglossic perspective. Bilinguals who translanguage tap into these features in innovative ways, ‘most often in ways that emerge through their interaction with other bilinguals, but sometimes as attempts to communicate with others through their evolving linguistic repertoires’ (Lin 2013).

It should be stressed at this point that it is not within our interest to settle the debate between code-switching and translanguaging, but rather to focus on how and when language mixing happens, as well as the extent to which it poses challenges for current educational technologies. Although there is often a clear preference among scholars in using one term over the other, ‘translanguaging’ has over time become ‘almost synonymous’ with ‘code-switching’, particularly in bilingual teacher education circles (Faltis 2019). In fact, despite translanguaging reflecting the current and dominant view of second language acquisition in educational linguistics (Duff and Byrnes 2019; Félix-Brasdefer and Shively 2021; García 2011; García and Kleyn 2016; García et al. 2017; Sun 2022; Walker 2021), the term ‘code-switching’ has become lexicalised in various subfields to an extent such that very few people, including researchers, fully grasp the nuanced theoretical implications that each term entails.³ We thus also want to make clear that while we fully acknowledge the epistemological incompatibility between the code-switching and translanguaging paradigms, we choose to use the term ‘code-switching’ here for the sake of our interdisciplinary audiences. Ultimately, ‘code-switching’ is used in keeping with the vast majority of studies on this phenomenon without implying any associated theoretical stance in this regard.

Having clarified the relevant terminologies, we next consider the reasons underlying this practice in an ESL context. We follow with a brief summary of why learners and teachers code-switch, before examining the current practice and what it means in a digital learning and teaching environment.

2.2. Why Do Learners Code-Switch?

Research has shown various reasons why learners might wish to code-switch (Carstens 2016; Masna 2020; Wang 2019), but one of the most common themes appears to be its ability to enable greater access to content. Specifically, students code-switch as an act of scaffolding to help them better grasp potentially complex concepts in the other language, thereby also increasing their self-confidence. Carstens (2016), for example, reported on a study in South Africa which investigated attitudes towards and the effectiveness of translanguaging between English, Afrikaans and other South African languages in an academic environment. Students were given a concept mapping task which required them to use both English and their L1 to become familiar with concepts within their field of study (waste management). Although this study was not strictly set in an ESL context, results demonstrate two important points: (i) learners tended to switch back into their L1 to make sense of new concepts being introduced, and (ii) they were highly satisfied that they were able to develop more skills and confidence in their weaker language. This latter point is of particular importance, as it showed how—contrary to the traditional monolingual belief—learners’ use of their L1 can in fact speed up their L2 progress, or at least their self-confidence in doing so.

Furthermore, several other studies have also reported that learners code-switch for a better affective experience. In particular, being able to use their L1 in the classroom helps students better connect with their friends and teachers. Wang (2019), for example, investigated learners’ perceptions of the code-switching pedagogy in Hong Kong, using questionnaires and focus groups. He reported a predominantly positive view of code-switching, including the fact that code-switching enhanced the learner–teacher and learner–learner relationship. Similarly, Ahmad and Jusoff (2009) noted an increasing level of comfort felt among lower-level English students in particular as soon as they were allowed to code-switch. Not only do students have positive attitudes towards code-switching in the classroom in particular (e.g., Carstens 2016; Wang 2019), but they are also more fond of teachers adopting this approach (e.g., Fareed et al. 2016; Yao 2011).

2.3. Why Do Teachers Code-Switch?

Given that many teachers' practices are driven by students' needs, it is no surprise that one of the most common reasons motivating teachers' code-switching practice is also to enable greater access to content. Johansson (2014), for example, found that although English teachers in Sweden only switch infrequently, when they switch, it is mainly for grammar instructions. In her semistructured interviews, four out of five teachers reported switching in order to clarify their teaching, especially when it comes to explaining vocabulary or structures. This is in line with previous and later research (e.g., Ahmad and Jusoff 2009; Lin 2013; Ngo and Phuong 2018; Shinga and Pillay 2021), which showed that both vocabulary introduction and grammar teaching are better conveyed in students' L1, even for those with a high L2 proficiency level (Lin 2013). Similarly, in a recent study on rural schools in South Africa, Shinga and Pillay (2021) found that ESL teachers consistently code-switch to clarify difficult concepts and enhance understanding of the content. This is a common approach in their bilingual teaching circles, despite parents preferring English instruction only.

Most importantly, teachers also reported code-switching as a means to enhance students' engagement. In Shinga and Pillay's (2021) interviews, teachers shared their experience that code-switching often helped bring students back to focus after they had zoned out and lost attention. This is particularly helpful in building an atmosphere more conducive to learning and teaching, or in other words, an 'enabling environment.' Many interviewees reported that they use code-switching to encourage learners' self-fulfilment and inclusivity, as 'language shouldn't be a barrier to learning'. This effect on students' participation is in line with findings in other studies such as Maluleke (2019) and Gumperz's (1982), who recognise the empowering nature of code-switching in a shared context. Furthermore, code-switching has also been traditionally used by teachers in many parts of the world for solidarity and personal identity (Liu 2010), as well as for establishing rapport with their students (García 2011).

2.4. Current Policies and Practices

Despite the fact that both teachers and learners use code-switching practices for these various benefits, there have been mixed views from other stakeholders. In fact, there is a vast range of how often teachers code-switch, and this is partly because of the differing amounts of support from above. Views from 'influencers' such as learners' parents or school management and policy makers generally vary according to whether family life/regional practice involves bilingualism, in which case code-switching in education is seen as a natural practice. Some examples include the Basque Country (Basque Government 2013), India (Tsimpli et al. 2020), or the Scoil Bhríde Cailini school in Ireland (Little and Kirwan 2019) which specifically involves L1 immigrant languages joining Irish, English and French in a code-switching environment. In almost everywhere else in the world, code-switching is much less the discourse mode, and the monolingual approach still 'lingers' (Wilson 2021). Those resisting this practice cited concerns including the norm of the monolingual approach, the institutional language policy, a lack of guidance on implementation, personal linguistic purism ideology, as well as assumptions and perceived dangers of the 'overuse' of the L1 on L2 progress (Liu and Fang 2020).

In Europe, nonetheless, the promotion of plurilingualism and code-switching in education has started to take off in recent years. Since 2001, the Common European Framework of Reference for languages (CEFR) (Council of Europe 2001) provides a framework of reference levels and descriptors of language ability based around an 'action-oriented approach', which focuses on how language is actually used in real life. Its most recent Companion Volume (Council of Europe 2020) sets out the importance of learners being able to utilise their full plurilingual repertoire by including specific descriptors about plurilingual comprehension and production. In some school environments, not merely is the practice of code-switching allowed to support L2 or other subject learning but also cross-linguistic mediation activities are specifically designed and carried out by teachers to develop their learners' ability to code-switch (North 2022).

The CEFR and its action-oriented approach to plurilingualism has had an impact on policy-making, especially in Europe. This is most notably evident by the European Commission's adoption of the 2019 Council Recommendation on a comprehensive approach to the teaching and learning of languages (European Commission 2019). Specifically, the recommendation encourages and promotes innovative multilingual competence practices, resulting in the increase in code-switching across language as appropriate within mainstream education (e.g., Le Pichon-Vorstman et al. 2020). Although the use of code-switching in assessment remains limited (Saville and Seed 2022), some examples of assessments which encourage code-switching can be found (Saville and Seed 2022; Seed and Holland 2020). The Austrian plurilingual oral exam, for example, requires test-takers to read a text in German and then code-switch between English and a third language in a given context (Piribauer et al. 2015). The assessment criteria not only recognise successful operation in the individual languages but also feature a 'language switch and interaction' set of descriptors, explicitly rewarding the ability to successfully code-switch.

2.5. Code-Switching in Digital Learning Environments

The aforementioned move towards plurilingualism has bred the need for technologies that are able to assist teaching and learning in a plurilingual mode. This is, however, currently a difficult task. Despite some evidence for code-switching in technologically mediated learning environments, such as in online courses (Adinolfi and Astruc 2017; Ho 2018), educational YouTube videos (Ho and Tai 2021) or machine translation tools (Séror 2022), most popular digital tools and environments 'are not designed with pedagogic objectives in mind' (Séror 2022, p. 459). Furthermore, the tools that are built for pedagogic reasons do so with a monolingual bias. As we shall show in Section 3.2, for example, English language learning tools often use NLP to correct and suggest improvements in an English learner's writing, yet detection of non-English words are treated as errors, or simply ignored without any recognition or feedback.

To complicate the matter further, there is currently no consensus on the overarching pedagogical goal when considering the role of code-switching for language education purposes. Despite some rising awareness and effort towards a plurilingual approach, it remains unclear as to whether the focus should be on providing scaffolds en route to a higher L2 proficiency with the aim of ultimately eliminating instances of code-switching, or to encourage the use of code-switching to achieve seamless communication across language boundaries in given contexts. This is ultimately the dilemma that has begun to be investigated in language education, then in language assessment, and is now making its way into language technologies.

Against this backdrop, our paper thus sets out to discuss the current difficulties and offer potential technological solutions for processing code-switching data in an educational context, with a particular focus on ESL. It is important to note that it is not within our capacity to determine one pedagogical goal over the other (i.e., to eventually eliminate instances of code-switching in L2 learning or to promote code-switching altogether for seamless communication); we rather focus on the practicalities of building educational tools that enable us to head in either or both of these directions.

3. Difficulties of Current NLP Technologies for ESL Teaching and Learning

3.1. Data

The first obstacle of building CSW-friendly NLP tools is the lack of appropriate training data. Code-switching is typically considered more prolific in speech than in writing (Gardner-Chloros 2020), but the costs and labour involved in building a speech corpus are far more significant (Caines et al. 2016). This means CSW data is generally scarce, and of those speech corpora available, most are community-based involving simultaneous/highly skilled bilinguals (e.g., Clyne 2003; Nguyen 2018; Torres Cacoullos and Travis 2018). Furthermore, the language-specific nature of code-switching and the labour-intensive process

of building these corpora also means that datasets are often small, niche and difficult to combine and compare.

Although learner data in written form is much more abundant than in speech, CSW data in this domain remains extremely limited. As an example, we used Google's Compact Language Detector library *pyclld3*⁴ to detect CSW in two large written learner corpora, namely Cambridge Write and Improve (Yannakoudakis et al. 2018),⁵ and the Teacher–Student Chatroom Corpus (Caines et al. 2020). We found no cases of code-switching out of 13.5K written conversational turns in the Teacher–Student Chatroom Corpus, and only 7.5K out of 17.26M instances of 'non-English' items in the Write & Improve database (0.04%). Most of these data, however, are simply Proper-Noun insertions rather than genuine code-switches.

Another issue related to CSW data in education is a serious lack of high-quality annotation. In fact, very few datasets have been explicitly annotated for code-switching. One exception we know of is the TLT-school corpus (Gretter et al. 2020), which collected non-native children's speech from students learning both English and German in schools of northern Italy. Although code-switching was identified as 'errors' in this work (p. 380), non-English tokens were all manually annotated. Quite a few of the labelled switches in this corpus, however, appear to be self-repairs or false starts upon inspection (e.g., 'Hello my name is (RETRACTED) *und* and I'm fourteen years old'), and so some degree of error is to be expected.⁶

Ultimately, it is fair to say that a large part of this lack of good annotation stems from the notorious lack of consensus in the field, where researchers still fundamentally disagree about the nature of the phenomenon (cf. Section 2.1). When experts still cannot settle on what constitutes a switch, annotation consistency at a macro level is difficult to achieve. In an ESL context particularly, annotation for CSW needs to go even a step beyond the surface level of identifying and labelling to classifying various types of code-switching: i.e., is a switch due to a lack of L2 knowledge, a lack of linguistic 'equivalents' in the other language, or a strategy to achieve certain pragmatic effects? This kind of fine-grained classification is crucial in improving the various parts of the pedagogical journey: from content creation and feedback to adaptive learning and assessment.

3.2. Difficulties of Current NLP Technologies for CSW

The difficulties with collecting CSW data also lead to the difficulties in improving relevant NLP models. Specifically, statistical and neural NLP models benefit from large amounts of high-quality training data. While this kind of resource exists for well-defined NLP tasks under monolingual settings (particularly English) including labelled and unlabelled data, as well as public benchmarks, this is not the case for CSW. Collecting CSW examples to construct this kind of annotated corpora is a slow and costly process, as we often need linguists and/or multilingual annotators to manually analyse the text.

Lack of CSW data also means lack of exposure to the noncanonical features of this kind of data, making existing models struggle in general. Downstream applications often use various NLP tools to help preprocess raw data (e.g., part-of-speech (PoS) taggers and parsers for extracting syntactic information from text for model training). The most recent systems adopt the neural approach and rely on large-scale pretrained language models (LMs) such as Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al. 2019). The majority of these models are, however, developed exclusively from monolingual text, making them ill-equipped to deal with the nonstandard CSW. Their limited exposure to the CSW environment also means that their performance is negatively affected by CSW examples present in the text.

3.3. Educational NLP Case Study

Another current limitation of computational approaches to CSW is the lack of end-to-end user-facing applications that interact directly with users in multilingual communities (Doğruöz et al. 2021). For example, we do not have any current technology that can both

understand and produce CSW, particularly in the way bilingual humans perform. In what follows, we contextualise this difficulty in language education by presenting two cases of widely used NLP applications for automated feedback and assessment.

3.3.1. Case 1: Feedback Technologies

The first two NLP applications to be discussed are grammatical error detection (GED) and grammatical error correction (GEC). GED is the task of automatically detecting grammatical errors in text, while GEC also suggests corrections. Both tasks have significant pedagogical benefits, such as offering proofreading tools that can help learners identify and correct their writing errors without human intervention, or educational software for automatically generating feedback comments on learners’ writing skills. An example of a learner sentence and its error feedback are shown in Table 1.

Table 1. An example sentence with error feedback. C: correct, I: incorrect, NN: noun number error, SVA: subject–verb agreement error.

Learner sentence	All	the	evidences	have	been	collected	.
GED feedback	C	C	I	I	C	C	C
GED feedback (detailed)	C	C	NN	SVA	C	C	C
GEC feedback	All	the	evidence	has	been	collected	.

Typically, GED is cast as a sequence labelling (or token classification) task, where each token is classified as either correct or incorrect (i.e., binary classification) (Bell et al. 2019; Rei and Yannakoudakis 2016) or according to error type at different levels of granularity (i.e., multiclass classification) (Yuan et al. 2021). Early approaches to GED focused on specific error types, particularly article errors (e.g., *I like playing in (team → a team) and deciding quickly what to do next*) and preposition errors (e.g., *our society is developing (in → at) high speed*), which are among the most frequent errors in non-native English learner writing (Han et al. 2004; Tetreault and Chodorow 2008). More general open-class GED systems were later developed using parse and text-based features extracted by NLP preprocessing tools (Gamon 2011). Recent work takes advantage of neural approaches and large-scale pretrained LMs (Bell et al. 2019; Kaneko and Komachi 2019; Rei and Yannakoudakis 2016). Yuan et al. (2021), for example, employed a pretrained language representation model ELECTRA (Clark et al. 2020), an extension of BERT with a different pretraining task which is a discriminator (rather than a generator) and aims to detect replaced tokens. A linear classification layer was added on top to perform error detection. After fine-tuning on annotated GED data for a small number of epochs, they reported state-of-the-art results on all GED benchmarks.

Similarly for GEC, beyond the development of feature-based machine learning classifiers for specific error types (De Felice and Pulman 2008; Rozovskaya and Roth 2011), a sequence-to-sequence translation task has been considered, where systems learn to ‘translate’ an ungrammatical sentence into a grammatical one (Yuan and Briscoe 2016). Both statistical machine translation (SMT) and neural machine translation (NMT) have been successfully applied to GEC with various task-specific adaptations (Felice et al. 2014; Yuan and Bryant 2021; Yuan and Briscoe 2016; Yuan et al. 2019). With recent advances in sequence-to-sequence modelling and the introduction of the Transformer encoder–decoder architecture (Vaswani et al. 2017), state-of-the-art GEC results have been reported (Lichtarge et al. 2020; Yuan et al. 2019).

All these existing feedback systems, however, have been mainly trained on monolingual data, i.e., annotated learner data and/or artificially generated data where errors have been injected into error-free native text (Felice and Yuan 2014; Rei et al. 2017).⁷ These systems have not been designed for CSW as they do not expect non-English words or phrases. Depending on the system training, non-English terms are either ignored—if they

are too distinct from the target language, without providing any feedback—or treated as errors. Consider the following examples:

- (3) 当下倾盆大雨时, we will stay home and our dad will tell us Halloween stories.
(Chinese–English)
 ‘When it is pouring, we will stay at home and our dad will tell us Halloween stories.’
(Write & Improve submission, 2021)

- (4) That was everything related to this *situación*.
(English–Spanish)
 ‘That was everything related to this *situation*.’
(Write & Improve submission, 2021)

In (3), the system ignored the Chinese characters. While in (4), the system flagged the Spanish word ‘*situación*’ as a learner error. We also observe that most of the CSW terms are treated as spelling mistakes. As a result, these feedback systems are not able to distinguish between genuine CSW, such as (3) and (4), and real errors, as in (5) and (6):

- (5) There we celebrated Halloween, *wearred* some *creapy* costumes, just like in America.
 (6) However, some people still *selebrate* it. (Write & Improve submission, 2021)

What is worse, systems fail to identify errors in their neighbouring words as soon as non-English words are involved, including cases where CSW is not that salient. This difference is demonstrated in examples (7) and (8). Specifically, while all GEC systems detected all three errors in (7), they were not able to highlight any in (8):

- (7) If I could be (anywher → **anywhere**) in the world right now, I would like to be in (Dubui → **Dubai**), one of the richest (country → **countries**) in the world.
(Write & Improve submission, 2022)
- (8) When the dancers are face to face and [*the*] music *stars*, their purpose is revealed: The Geommu or Sword Dance of Jinju, one of the most representative dances of Korea.
(Write & Improve submission, 2022)

This leads us to expect similar system failure for potentially more complicated interacting errors, such as code-switching subject–verb agreement, where the subject and the verb may be expressed in two different languages.

3.3.2. Case 2: Assessment Technologies

In terms of digital assessment, automated essay scoring (AES) is a case in point. AES is the task of employing computer technology to assign a score to a learner’s written text. Learning to write well in a foreign language requires a considerable amount of practice and appropriate feedback. On the one hand, AES systems provide a learning environment in which language learners can practise and improve their writing skills even when teachers are not available. On the other hand, AES reduces the workload of examiners and enables large-scale writing assessment. In fact, these technologies have already been deployed in standardised tests such as the TOEFL and GMAT (Chen et al. 2016; Chodorow and Burstein 2004) and also in a classroom setting (Wilson et al. 2019).

Generally speaking, AES systems exploit textual features in order to measure overall quality and assign a score on that basis. Research in this area has tended to prioritise feature-based models, and a lot of traditional work has focused on feature engineering. Early feature-based systems used superficial features, such as essay length, as proxies for understanding the text. As multiple factors influence the quality of texts, later systems have used more sophisticated NLP preprocessing tools to explore a large range of shallow and deep linguistic features, such as lexical and PoS *n*-grams, syntactic constructions, grammatical relations and measures of sentence complexity. More recent AES systems, however, exploit neural networks that learn a feature representation automatically (Alikaniotis et al. 2016; Mayfield and Black 2020; Taghipour and Ng 2016), with pretrained LMs (e.g., BERT) being applied to achieve state-of-the-art performance (Andersen et al. 2021; Mayfield and

Black 2020). These neural models nonetheless require a large amount of annotated data. Given the computational cost at both training and inference time, as well as the additional hardware requirements, end-to-end neural AES systems are still in the early stages.

Furthermore, it is worth noting that AES systems were originally used for summative purposes in standardised testing, and hence have been designed to mimic the judgement of examiners evaluating the quality of learner writing. Systems are therefore not expected to deal with CSW input. In fact, any use of non-standard English is penalised as it is often seen as a lack of second-language proficiency. It is common practice for online AES systems to reject submissions whose non-English content is above a certain threshold. For example, we observe that submissions on the Cambridge Write & Improve containing large chunks of non-English words are treated as noisy input, and therefore no marks are assigned. One of the reasons behind this design choice is that existing AES systems such as W&I are not able to provide reliable feedback and scores for CSW. The underlying models (including both NLP tools and pretrained LMs) employed by these AES systems often fail to process CSW data or extract useful features, as discussed in Section 3.2.

4. Future Directions

4.1. Data

At this point, it has become clear that the lack of appropriate training data lies at the heart of many difficulties we face. We showed in Section 3.1 that there is no established standard for collecting and labelling code-switching data, such that resources are often scattered, niche and difficult to combine and compare. The first step in resolving our problems is thus working towards a standardised data statement which underpins the collection and annotation for all pedagogical CSW data across the board. Basic demographic information of sampled code-switchers, such as the learner's L1 and level of proficiency, the teacher's general experience and attitude to code-switching, should be documented for each dataset.

Furthermore, successful data annotation requires a more developed guideline considering bona fide CSW cases in an ESL context, alongside those that are more likely genuine errors. For cases where CSW is identified, we also need to go beyond the surface to classify these cases by the learner's most likely motivations, e.g., whether it is a knowledge gap, a lack of linguistic 'equivalent', or a pragmatic strategy. Although this is not a straightforward task, we demonstrate some of the simplest cases of these scenarios using real learner data in (9), (10) and (11), respectively.

- (9) Suddenly, basketball in the ocean, one *haitun* help they catch the basketball. [sic]

'Suddenly, (the) basketball (fell) in the ocean, a *whale* helped them catch the basketball.'

(English–Chinese,
lack of L2 knowledge)

(Revised A2 Flyers Young Learners Reading and Writing test submission (trial), 2018)

- (10) Halloween tradition was also cerebrated in our country on full moon day of *tazaungdaing*. [sic]
'(The) Halloween tradition was also celebrated in our country on the full moon day of the *Myanmar calendar*.'

(English–Burmese,
lack of L2 linguistic 'equivalent')
(Write & Improve submission, 2022)

- (11) As the Arabs say; أهلا وسهلا بك يا عزيزي - Welcome my fellows! (English–Arabic,
pragmatic strategy)

(Write & Improve submission, 2022)

These examples collectively illustrate various reasons for why L2 learners code-switch. In (9), for instance, *haitun* is a Romanised Chinese word for 'whale', and it is reasonably

clear from the sentence that the learner did not know the word for ‘whale’ in English and hence reverted back into their (transliterated) L1 to achieve their communicative intent. This is a case of code-switching due to a lack of L2 knowledge. In (10), however, the scenario is quite different: *tazaungdaing* is an object specific to Burmese culture, which does not have an ‘equivalent’ per se in English. Similarly, the switch to Arabic in (11) is accompanied by a perfect English translation and is most likely a pragmatic strategy to mimic what a real Arab would say. All these types of code-switching in fact communicate quite different things about each learner, and therefore should be treated differently. The point of this classification is thus not to ‘prescribe’ what learners ought to perform but rather to better determine (i) when feedback might be needed, (ii) what type of feedback is needed and (iii) what these patterns tell us about each learner’s language development. Ultimately, these are crucial pieces of information to enhance individual learning experience, but more importantly, to inform adaptation of curriculum design and assessment that can truly embrace each learner’s plurilingual practice.

On a related note, the generation of artificial CSW data for training and testing purposes is another avenue worthy of attention. This draws on the fact that research has established certain cross-linguistic patterns that govern where switching is more or less likely to occur (e.g., the switch of nouns is the most common while that of ‘pronouns’ is much less, in-word switching is least likely and ‘language-neutral’ tokens such as discourse markers or proper nouns are more likely to facilitate a switch, etc.—see i.a. [Poplack 1980](#) et seq., [Bullock and Toribio 2009](#); [Clyne 2003](#)). Although artificial data is often not linguistically ideal, these ‘superficial’ probabilistic tendencies are a good starting point to generate artificial CSW that might mimic real-life behaviours of language users. If successful, this may ultimately pave a way into a future of more efficient and consistent CSW data collection, annotation and analysis.

Importantly, we also believe that it is necessary for future research to bring in an extended reference set for evaluation purposes. Since CSW data is so limited, until recently, the evaluation of CSW models was performed over individual tasks and language pairs ([Doğruöz et al. 2021](#)), primarily on annotated tweets. Although some new standard benchmarks such as GLUECoS⁸ ([Khanuja et al. 2020](#)) or LINCE⁹ ([Aguilar et al. 2020](#)) have been introduced in recent years, this purely ‘numerical’ approach is linguistically limiting and clearly does not represent the sophisticated spectrum of the CSW phenomenon. For ESL especially, there are many reasons why learners and/or teachers code-switch (Sections 2.2 and 2.3), and so any technological evaluation can only be comprehensive if it is repeatedly tested on different genres and domains (e.g., written essays, conversational practice, presentations, etc.) and on different types of learners (e.g., children/adults/adults with special learning needs, etc.) who have different L1s and proficiency levels. All of these should then be complemented by a certain level of human judgement. Since educational technology for CSW is still quite a young field and data remains limited, linguistic expertise is one of our best hopes (alongside language education) to explain patterns or types of switches that our automated systems struggle with, thereby enabling us to target improvements in a more sensible way.

4.2. Improving NLP Models for Educational CSW

It is clear that developing NLP technologies for CSW is still a nascent field, where many research questions remain unanswered. Although we do not have to start from scratch, the tasks and techniques required are significantly different from those for monolingual data. For future work, we propose some specific suggestions:

First, efficient exploitation of data must be of prime priority. As in low-resource scenarios, CSW research is likely to benefit from transfer learning across nearby languages. Access to CSW data is limited, while monolingual corpora are much easier to find. It is common that one language involved in CSW has significantly more resources (e.g., English), and exploiting those is likely to offer some advantages. Furthermore, we suggest adopting a multistep training strategy to make efficient use of unlabelled and labelled,

monolingual and CSW data that already exist. In fact, the rapid advancement of NLP technologies, especially the rise of massively multilingual LMs such as multilingual BERT and XLM-R (Conneau et al. 2020), has enabled the development of systems that can work with multiple languages. New CSW model training can include a pretraining stage and a fine-tuning stage, where unlabelled monolingual data and limited labelled CSW data can be used, respectively.

Second, serious efforts should also be made towards encoding language information to build (socio-)linguistically motivated CSW models. Computational approaches to CSW are not as developed as linguistic/educational studies, which are much more mature fields in comparison. While most work in computational CSW has focused on one particular language pair, it is clear that this does not represent the sophisticated spectrum of code-switching across the board. Social and cultural factors, speakers' relative fluency, their L1s, as well as the participating languages' grammatical features all have different effects on the CSW practice. Until we understand the intricate sociolinguistic mechanism underlining CSW as a phenomenon, we will not be able to offer targeted solutions that best serve bilinguals' interests.

Furthermore, building more end-to-end systems that can interact in CSW with multilingual speakers should be one of the top priorities. In fact, while there has been some recent work on individual NLP models for CSW such as PoS tagging and NER (Khanuja et al. 2020), end-to-end systems that can both understand and produce CSW are virtually non-existent. For educational technology in particular, we have not yet seen any user-facing computer-assisted language learning systems (CALL) for CSW. This is partly due to lack of data for such systems; however, the use of CSW in an ESL context and the aim of CSW applications should also inform the development of end-to-end user-facing educational systems for CSW.

Ultimately, a new design of **feedback** and **assessment** systems is required to resolve the tension between learner errors and CSW. Genuine CSW should not be treated as learner errors, which often demotivates ESL learners and negatively impacts their learning outcomes. For **feedback systems**, it is thus crucial that we shift the focus towards developing personalised systems that are capable of catering for different learner L1s and proficiency levels. In the mixed sentence 'My hobby is *fussball spielen*' (2), for example, an ideal feedback system should be able to process the non-English fragment *fussball spielen* and provide instant pop-up feedback to the learner, such as 'Did you mean "playing football"?''. Although we have shown that not all learner CSW is a result of knowledge gap, this kind of intelligent output has great potential in identifying genuine 'gaps' in learners' L2 knowledge, which could then be fed back into content creation and adaptive learning.¹⁰ Similarly, different **assessment systems** should be designed to address different ESL learning objectives: improving learners' English language skills or enhancing communication in a multilingual and multicultural society. Most existing assessment systems focus on the former, in which case CSW and the use of learners' L1 tend to be dismissed. As we move away from the traditional 'language purism' ideology, however, the aim has progressed more towards helping learners further their communicative skills. Assessment systems thus need to go beyond surface grammaticality judgement, thereby giving more weight to other aspects such as learner's organisation, content and overall communicative achievement.

5. Final Thoughts

In this paper, we discussed the phenomenon of code-switching in language education, its technological implications as well as possible directions for future work. We highlighted several issues in building technologies for plurilingual learners, including lack of training data, lack of robust evaluation benchmarks, and lack of end-to-end user-facing educational applications. We made some specific suggestions on how to tackle these problems moving forward, with a particular focus on generating, collecting and analysing learner's data that can be used to feed back into NLP models.

It is important to note, however, that while the technological solutions we proposed may have potential in encouraging plurilingual practice and assessment, the challenge does not stop there. It will still be very difficult to build a perfectly balanced system that can truly embrace learners' plurilingual repertoire without also inadvertently driving them towards conforming to these systems. Furthermore, each learner's plurilingual repertoire is unique to themselves, and so it is challenging to build automated plurilingual applications that can also be personalised to each user. There are some ways, however, that we can start carrying it out, hence what educational experts have been writing about in the last several years, as well as the increasing hope that the technology we are investing in today can be more personalised in the future.

What we fundamentally need, at this point, is thus ongoing efforts across the board, with incorporated insights from fields beyond NLP such as (socio-)linguistics, language education, anthropology and other social sciences that situate language technologies within the complex cultural and social hierarchies they operate in. This means technologists need to more actively engage with and give voice to not only researchers but also to learners, parents, teachers and others within the language teaching and learning community. On a broader scale, this requires us to first re-examine the dynamics between technologists and communities who make use of these systems, build an equal research relationship with them and finally to actively engage with their lived experiences. In an era of endless technology races to achieve the 'best' system, it can be easy to forget that technology is built to support people, and thus should always be developed with people at heart.

Ultimately, what we hope to achieve here is not to offer a perfect solution but rather to ascertain what might be technologically possible, thereby kick-starting a mutual dialogue with relevant stakeholders and other experts in related fields.

Author Contributions: Conceptualization, L.N. and Z.Y.; methodology, Z.Y.; software, Z.Y.; validation, L.N. and Z.Y.; formal analysis, L.N. and Z.Y.; investigation, L.N., Z.Y. and G.S.; resources, L.N., Z.Y. and G.S.; data curation, L.N. and Z.Y.; writing—original draft preparation, L.N., Z.Y. and G.S.; writing—review and editing, L.N., Z.Y. and G.S.; visualization, L.N. and Z.Y.; supervision, L.N.; project administration, L.N.; funding acquisition, L.N. and G.S. All authors have read and agreed to the published version of the manuscript.

Funding: The first and third authors are supported by Cambridge University Press and Assessment.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

- ¹ While this is seen as something new in the West, it is already considered the norm in other parts of the world, such as India or sub-Saharan Africa.
- ² It is worth making clear that there are other similar language contact phenomena in CSW research, such as L1 transfer and loan translation (or calques). These phenomena generally involve lexical items from only one language but the semantics and grammatical constructions from the other (see [Treffers-Daller 2009](#), [Backus and Dorleijn 2009](#) for a comprehensive overview). Since our focus in this work is on developing technologies that can process multilingual lexical input, these lie outside the remit of our discussion.
- ³ This is no surprise given the popularity of the term 'code-switching' in various fields over a long period of time.
- ⁴ <https://github.com/google/cld3/>, accessed on 31 May 2021.
- ⁵ Cambridge English Write & Improve (W&I) is an online web platform that assists non-native English learners with their writing. Learners from around the world submit letters, stories, articles and essays for automated assessment in response to various prompts. W&I then provides instant feedback and predicts the learners' proficiency level on the CEFR scale. The platform can be accessed at <https://writeandimprove.com/>, accessed on 30 April 2022.

- 6 As one of the reviewers pointed out, this finding is consistent with observations in recent work on child language acquisition. Specifically, intrasentential language mixing is extremely rare in the naturalistic speech of young children, particularly those raised in a Romance country with two languages from birth (Poeste and Müller 2020; Poeste et al. 2019).
- 7 While we focus on English, our analysis and methods presented in this paper can be applied to any language.
- 8 General Language Understanding Evaluation benchmark for Code-Switched NLP.
- 9 LINguistic benchmark for Code-switching Evaluation.
- 10 We are aware, nonetheless, that building such systems is a very difficult task, requiring integrated machine translation outputs, models that can capture complex grammatical and lexical features of all participating languages, as well as learners' relevant (socio-)linguistic background. This brings us back to our previous point, which emphasised the need to encode language and social information to build sociolinguistically motivated CSW models.

References

- Aaron, Jessi Elana. 2015. Lone English-origin nouns in Spanish: The precedence of community norms. *International Journal of Bilingualism* 19: 459–80. [CrossRef]
- Adinolfi, Lina, and Lluïsa Astruc. 2017. An exploratory study of translanguaging practices in an online beginners' foreign language classroom. *Language Learning in Higher Education. Journal of the European Confederation of Language Centres in Higher Education (CercleS)* 7: 185–204. [CrossRef]
- Aguilar, Gustavo, Sudipta Kar, and Tamar Solorio. 2020. LinCE: A centralized benchmark for linguistic code-switching evaluation. Paper presented at 12th Language Resources and Evaluation Conference, Marseille, France, May 11–16. Paris: European Language Resources Association, pp. 1803–13.
- Ahmad, Badrul, and Kamaruzaman Jusoff. 2009. Teachers' code-switching in classroom instructions for low english proficient learners. *English Language Teaching* 2: 49–55. [CrossRef]
- Alikaniotis, Dimitrios, Helen Yannakoudakis, and Marek Rei. 2016. Automatic text scoring using neural networks. Paper presented at 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Berlin, Germany, August 7–12. Stroudsburg: Association for Computational Linguistics, pp. 715–25. [CrossRef]
- Andersen, Øistein E., Rebecca Watson, Zheng Yuan, and Kevin Yet Fong Cheung. 2021. Benefits of alternative evaluation methods for automated essay scoring. Paper presented at 14th International Conference on Educational Data Mining (EDM 2021), Virtual, June 29–July 2.
- Anderson, Jason, and Amy Lightfoot. 2021. Translingual practices in english classrooms in india: Current perceptions and future possibilities. *International Journal of Bilingual Education and Bilingualism* 24: 1210–31. [CrossRef]
- Backus, Ad, and Margreet Dorleijn. 2009. *Loan Translations versus Code-Switching*. Cambridge Handbooks in Language and Linguistics. Cambridge: Cambridge University Press, pp. 75–94. [CrossRef]
- Bell, Samuel, Helen Yannakoudakis, and Marek Rei. 2019. Context is key: Grammatical error detection with contextual word representations. Paper presented at Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications, Florence, Italy, August 2. Stroudsburg: Association for Computational Linguistics, pp. 103–15. [CrossRef]
- Brevik, Lisbeth M., and Ulrikke Rindal. 2020. Language use in the classroom: Balancing target language exposure with the need for other languages. *TESOL Quarterly* 54: 925–53. [CrossRef]
- Bullock, Barbara E., and Almeida Jacqueline Ed Toribio. 2009. *The Cambridge Handbook of Linguistic Code-Switching*. Cambridge Handbooks in Language and Linguistics. Cambridge: Cambridge University Press. [CrossRef]
- Caines, Andrew, Christian Bentz, Calbert Graham, Tim Polzehl, and Paula Buttery. 2016. Crowdsourcing a Multi-lingual Speech Corpus: Recording, Transcription and Annotation of the CROWDED CORPUS. Paper presented at Tenth International Conference on Language Resources and Evaluation (LREC 2016), Paris, France, May 23–28. Paris: European Language Resources Association (ELRA).
- Caines, Andrew, Helen Yannakoudakis, Helena Edmondson, Helen Allen, Pascual Pérez-Paredes, Bill Byrne, and Paula Buttery. 2020. The teacher-student chatroom corpus. Paper presented at 9th Workshop on NLP for Computer Assisted Language Learning, Gothenburg, Sweden, November 20. Linköping: LiU Electronic Press, pp. 10–20.
- Carstens, Adelia. 2016. Translanguaging as a vehicle for l2 acquisition and l1 development: Students' perceptions. *Language Matters* 47: 203–22. [CrossRef]
- Chen, Jing, James H. Fife, Isaac I. Bejar, and Andre A. Rupp. 2016. Building e-rater® Scoring Models Using Machine Learning Methods. *ETS Research Report Series* 2016: 1–12. [CrossRef]
- Chodorow, Martin, and Jill Burstein. 2004. Beyond essay length: Evaluating e-rater®'s performance on toefl® essays. *ETS Research Report Series* 2004: 1–38. [CrossRef]
- Clark, Kevin, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. 2020. ELECTRA: Pre-training Text Encoders as Discriminators Rather than Generators. Paper presented at International Conference on Learning Representations, Addis Ababa, Ethiopia, April 26–30.
- Clyne, Michael. 2003. *Dynamics of Language Contact: English and Immigrant Languages*. Cambridge: Cambridge University Press. [CrossRef]

- Conneau, Alexis, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. Paper presented at 58th Annual Meeting of the Association for Computational Linguistics, Online, July 5–10. Stroudsburg: Association for Computational Linguistics, pp. 8440–51. [\[CrossRef\]](#)
- Council of Europe. 2001. *European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- Council of Europe. 2020. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume with New Descriptors*. London: Council of Europe.
- Creese, Angela, and Adrian Blackledge. 2010. Translanguaging in the bilingual classroom: A pedagogy for learning and teaching? *The Modern Language Journal* 94: 103–15. [\[CrossRef\]](#)
- Daniel, Shannon M., Robert T. Jiménez, Lisa Pray, and Mark B. Pacheco. 2019. Scaffolding to make translanguaging a classroom norm. *TESOL Journal* 10: e00361. [\[CrossRef\]](#)
- De Felice, Rachele, and Stephen G. Pulman. 2008. A classifier-based approach to preposition and determiner error correction in L2 English. Paper presented at 22nd International Conference on Computational Linguistics (Coling 2008), Manchester, UK, August 18–22. Manchester: Coling 2008 Organizing Committee, pp. 169–76.
- Department of Education, Language Policy, Culture of the Basque Government, and The Sociolinguistics Cluster. 2013. *Talking Pupils: The Arrue Project 2011: Research, Results and Contributions of Experts*. Vitoria-Gasteiz: Basque Government.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. Paper presented at 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, MN, USA, June 2–7. Stroudsburg: Association for Computational Linguistics, pp. 4171–86. [\[CrossRef\]](#)
- Doğruöz, A. Seza, Sunayana Sitaram, Barbara E. Bullock, and Almeida Jacqueline Toribio. 2021. A survey of code-switching: Linguistic and social perspectives for language technologies. Paper presented at 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Online, August 1–6. Stroudsburg: Association for Computational Linguistics, pp. 1654–66. [\[CrossRef\]](#)
- Duff, Patricia A., and Heidi Byrnes. 2019. Sla across disciplinary borders: Introduction to the special issue. *The Modern Language Journal* 103: 3–5. [\[CrossRef\]](#)
- European Commission. 2019. *Proposal for a Council Recommendation on a Comprehensive Approach to the Teaching and Learning of Languages*. Brussels: European Commission.
- Faltis, Christian. 2019. *Pedagogical Codeswitching and Translanguaging in Bilingual Schooling Contexts*. New York: Routledge, pp. 39–62. [\[CrossRef\]](#)
- Faltis, Christian, and Valdés. 2016. Preparing teachers for teaching in and advocating for linguistically diverse classrooms: A vade mecum for teacher educators. In *The Handbook of Research on Teaching*, 5th ed. Washington, DC: American Educational Research Association, pp. 549–92.
- Fareed, Muhammad, Samreen Humayun, and Huma Akhtar. 2016. English language teachers' code-switching in class: Esl learners' perceptions. *Journal of Education & Social Sciences* 4: 3–13. [\[CrossRef\]](#)
- Felice, Mariano, and Zheng Yuan. 2014. Generating artificial errors for grammatical error correction. Paper presented at Student Research Workshop at the 14th Conference of the European Chapter of the Association for Computational Linguistics, Gothenburg, Sweden, April 26–30. Stroudsburg: Association for Computational Linguistics, pp. 116–26. [\[CrossRef\]](#)
- Felice, Mariano, Zheng Yuan, Øistein E. Andersen, Helen Yannakoudakis, and Ekaterina Kochmar. 2014. Grammatical error correction using hybrid systems and type filtering. Paper presented at Eighteenth Conference on Computational Natural Language Learning: Shared Task, Baltimore, MD, USA, June 26–27. Stroudsburg: Association for Computational Linguistics, pp. 15–24. [\[CrossRef\]](#)
- Félix-Brasdefer, J. César, and Rachel L. Shively, eds. 2021. *New Directions in Second Language Pragmatics*. Berlin: De Gruyter Mouton. [\[CrossRef\]](#)
- Gamon, Michael. 2011. High-order sequence modeling for language learner error detection. Paper presented at Sixth Workshop on Innovative Use of NLP for Building Educational Applications, Portland, OR, USA, June 24. Stroudsburg: Association for Computational Linguistics, pp. 180–89.
- García, Ofelia. 2011. *Bilingual Education in the 21st Century: A Global Perspective*. West Sussex: Wiley-Blackwell.
- García, Ofelia, and Tatyana Kleyn. 2016. *Translanguaging Multilingual Students: Learning from Classroom Moments*. New York: Routledge.
- García, Ofelia, Susana Ibarra Johnson, and Kate Seltzer. 2017. *The Translanguaging Classroom: Leveraging Student Bilingualism for Learning*. Philadelphia: Carlson.
- Gardner-Chloros, Penelope. 2009. *Code-Switching*. Cambridge: Cambridge University Press.
- Gardner-Chloros, Penelope. 2020. *Contact and Code-Switching*. West Sussex: John Wiley & Sons, Ltd., chp. 9, pp. 181–99. [\[CrossRef\]](#)
- Gretter, Roberto, Marco Matassoni, Stefano Bannò, and Falavigna Daniele. 2020. TLT-school: A corpus of non native children speech. Paper presented at 12th Language Resources and Evaluation Conference, Marseille, France, May 11–16. Paris: European Language Resources Association, pp. 378–85.
- Gumperz, John J. 1982. *Discourse Strategies. Studies in Interactional Sociolinguistics*. Cambridge: Cambridge University Press. [\[CrossRef\]](#)

- Han, Na-Rae, Martin Chodorow, and Claudia Leacock. 2004. Detecting errors in English article usage with a maximum entropy classifier trained on a large, diverse corpus. Paper presented at Fourth International Conference on Language Resources and Evaluation (LREC'04), Lisbon, Portugal, May 26–28. Paris: European Language Resources Association (ELRA).
- Ho, Wing Yee Jenifer. 2018. Translanguaging in Online Language Learning: Case Studies of Self-Directed Chinese Learning of Multilingual Adults. Ph.D thesis, University College London, London, UK.
- Ho, Wing Yee Jenifer, and Kevin W. H. Tai. 2021. Translanguaging in digital learning: The making of translanguaging spaces in online english teaching videos. *International Journal of Bilingual Education and Bilingualism* 1–22. [CrossRef]
- Johansson, Sara. 2014. Code-Switching in the English Classroom: What Teachers Do and What Their Students Wish They Did. Available online: <http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A691914&dsid=5065> (accessed on 30 April 2022).
- Kaneko, Masahiro, and Mamoru Komachi. 2019. Multi-Head Multi-Layer Attention to Deep Language Representations for Grammatical Error Detection. *Computing Research Repository* 23: 883–91. [CrossRef]
- Khanuja, Simran, Sandipan Dandapat, Anirudh Srinivasan, Sunayana Sitaram, and Monojit Choudhury. 2020. GLUECoS: An evaluation benchmark for code-switched NLP. Paper presented at 58th Annual Meeting of the Association for Computational Linguistics, Online, July 5–10. Stroudsburg: Association for Computational Linguistics, pp. 3575–85. [CrossRef]
- Kim, Jeongyeon, and Bradley Tatar. 2017. Nonnative english-speaking professors' experiences of english-medium instruction and their perceived roles of the local language. *Journal of Language, Identity & Education* 16: 157–71. [CrossRef]
- Lapierre, Cynthia. 2018. The Role of First Language Influence in the Learning of Second Language Grammar: The Case of His/Her in English. Master's thesis, Concordia University, Montreal, QC, Canada. *unpublished*.
- Le Pichon-Vorstman, Emmanuelle, Hanna Siarova, and Eszter Szönyi. 2020. *The Future of Language Education in Europe: Case Studies of Innovative Practices: Analytical Report*. Luxembourg: European Commission and Directorate-General for Education, Youth, Sport and Culture, Publications Office. [CrossRef]
- Lichtarge, Jared, Chris Alberti, and Shankar Kumar. 2020. Data weighted training strategies for grammatical error correction. *Transactions of the Association for Computational Linguistics* 8: 634–46. [CrossRef]
- Lin, Angel. 2013. Classroom code-switching: Three decades of research. *Applied Linguistics Review* 4: 195–218. [CrossRef]
- Little, David, and Déirdre Kirwan. 2019. *Engaging with Linguistic Diversity: A Study of Educational Inclusion in an Irish Primary School*. Multilingualisms and Diversities in Education. London: Bloomsbury Publishing.
- Liu, Jingxia. 2010. Teachers' code-switching to the L1 in efl classroom. *The Open Applied Linguistics Journal* 3: 10–23. [CrossRef]
- Liu, Yang, and Fan Fang. 2020. Translanguaging theory and practice: How stakeholders perceive translanguaging as a practical theory of language. *RELC Journal* 1–9. [CrossRef]
- Macaro, Ernesto, Lili Tian, and Lingmin Chu. 2020. First and second language use in english medium instruction contexts. *Language Teaching Research* 24: 382–402. [CrossRef]
- Mahboob, Ahmar, and Angel M. Y. Lin. 2016. *Using Local Languages in English Language Classrooms*. Cham: Springer International Publishing, pp. 25–40. [CrossRef]
- Maluleke, Mzamani. 2019. Using code-switching as an empowerment strategy in teaching mathematics to learners with limited proficiency in english in south african schools. *South African Journal of Education* 39: 1–9. [CrossRef]
- Masna, Yuliar. 2020. Efl learners' code-switching: Why do they switch the language? *Englisia: Journal of Language, Education, and Humanities* 8: 93–101. [CrossRef]
- Mayfield, Elijah, and Alan W. Black. 2020. Should you fine-tune BERT for automated essay scoring? Paper presented at Fifteenth Workshop on Innovative Use of NLP for Building Educational Applications, Seattle, WA, USA, July 9–10. Stroudsburg: Association for Computational Linguistics, pp. 151–62. [CrossRef]
- Muysken, Pieter. 2000. *Bilingual Speech: A Typology of Code-Mixing*. Cambridge: Cambridge University Press.
- Myers-Scotton, Carol. 1993. *Duelling Languages: Grammatical Structure in Codeswitching*. Oxford: Clarendon.
- Ngo, Ngoc Bich, and Yen Hoang Phuong. 2018. The frequency and functions of teachers' use of mother tongue in EFL classrooms. *European Journal of English Language Teaching* 3. [CrossRef]
- Nguyen, Li. 2018. Borrowing or code-switching? Traces of community norms in Vietnamese-English speech. *The Australian Journal of Linguistics* 38: 443–66. [CrossRef]
- North, Brian. 2022. *Plurilingual Mediation in the Classroom Examples from Practice*. London: Routledge, chp. 17, pp. 319–36.
- Orellana, Marjorie Faulstich, and Ofelia García. 2014. Conversation currents: Language brokering and translanguaging in school. *Language Arts* 91: 386–92.
- Otheguy, Ricardo, Ofelia García, and Wallis Reid. 2015. Clarifying translanguaging and deconstructing named languages: A perspective from linguistics. *Applied Linguistics Review* 6: 281–307. [CrossRef]
- Otto, Ana, and Estrada José Luis. 2019. Towards an understanding of cil assessment practices in a european context: Main assessment tools and the role of language in content subjects. *CLIL. Journal of Innovation and Research in Plurilingual and Pluricultural Education* 2: 31–42. [CrossRef]
- Pennycook, Alastair. 2010. *Language as a Local Practice*. London and New York: Routledge.
- Piccardo, Enrica, Aline Germain-Rutherford, and Geoff Lawrence. 2021. *The Routledge Handbook of Plurilingual Language Education*, 1st ed. London: Routledge. [CrossRef]

- Piribauer, Gerda, Ute Atzlesberger, Irmgard Greinix, Thomas Ladstatter, Franz Mittendorfer, Helmut Renner, and Belinda Steinhuber. 2015. *Designing and Implementing Plurilingual Oral Exams: Framework for the Austrian Upper Secondary Level Oral Leaving Examination at Colleges for Higher Vocational Education*. Vienna: Center for Vocational Languages CEBS.
- Poeste, Meike, and Natascha Müller. 2020. Code-mixing in trilingual children: Domains and dimensions of language dominance. In *Contact, Variation and Change in Romance and Beyond*. Studies in Honor of Trudel Meisenburg. Berlin: Erich Schmidt Verlag, pp. 301–20.
- Poeste, Meike, Natascha Müller, and Laia Arnaus Gil. 2019. Code-mixing and language dominance: Bilingual, trilingual and multilingual children compared. *International Journal of Multilingualism* 16: 459–91. [CrossRef]
- Poplack, Shana. 1980. Sometimes i'll start a sentence in Spanish *y termino en español*: Toward a typology of codeswitching. *Linguistics* 18: 581–618. [CrossRef]
- Poplack, Shana, David Sankoff, and Christopher Miller. 1988. The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics* 26: 47–104. [CrossRef]
- Qasem, Fawaz. 2020. Crosslinguistic influence of the first language: Interlingual errors in the writing of esl saudi learners. *Macrolinguistics* 8: 105–20. [CrossRef]
- Rei, Marek, and Helen Yannakoudakis. 2016. Compositional sequence labeling models for error detection in learner writing. Paper presented at 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Berlin, Germany, August 7–12. Stroudsburg: Association for Computational Linguistics, pp. 1181–91. [CrossRef]
- Rei, Marek, Mariano Felice, Zheng Yuan, and Ted Briscoe. 2017. Artificial error generation with machine translation and syntactic patterns. Paper presented at 12th Workshop on Innovative Use of NLP for Building Educational Applications, Copenhagen, Denmark, September 8. Stroudsburg: Association for Computational Linguistics, pp. 287–92. [CrossRef]
- Rozovskaya, Alla, and Dan Roth. 2011. Algorithm selection and model adaptation for ESL correction tasks. Paper presented at 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, OR, USA, June 19–24. Stroudsburg: Association for Computational Linguistics, pp. 924–33.
- Sah, Pramod K., and Guofang Li. 2020. Translanguaging or unequal languaging? unfolding the plurilingual discourse of english medium instruction policy in nepal's public schools. *International Journal of Bilingual Education and Bilingualism* 1–20. [CrossRef]
- Saville, Nick, and Graham Seed. 2022. Language assessment in the context of plurilingualism. In *The Routledge Handbook of Plurilingual Education*. Edited by Geoff Lawrence, Enrica Piccardo, and Aline Germain-Rutherford. London: Routledge, chp. 19, pp. 360–76.
- Seed, Graham. 2020. *What Is Plurilingualism and What Does It Mean for Language Assessment?* Cambridge: Cambridge Assessment English: Research Notes.
- Seed, Graham, and Martine Holland. 2020. *Taking Account of Plurilingualism in Cambridge Assessment English Products and Services*. Cambridge: Cambridge Assessment English: Research Notes.
- Séror, Jérémie. 2022. Plurilingualism in digital spaces. In *The Routledge Handbook of Plurilingual Education*. Edited by Geoff Lawrence, Enrica Piccardo, and Aline Germain-Rutherford. London: Routledge, chp. 22, pp. 449–64.
- Shinga, Sibongile, and Ansurie Pillay. 2021. Why do teachers code-switch when teaching english as a second language? *South African Journal of Education* 41: 3–13. [CrossRef]
- Sun, Yachao. 2022. Implementation of translanguing pedagogies in eal writing: A systematic review. *Language Teaching Research* 1–25. [CrossRef]
- Taghipour, Kaveh, and Hwee Tou Ng. 2016. A neural approach to automated essay scoring. Paper presented at 2016 Conference on Empirical Methods in Natural Language Processing, Austin, TX, USA, November 1–5. Stroudsburg: Association for Computational Linguistics, pp. 1882–91. [CrossRef]
- Tetreault, Joel R., and Martin Chodorow. 2008. The ups and downs of preposition error detection in ESL writing. Paper presented at 22nd International Conference on Computational Linguistics (Coling 2008), Manchester, UK, August 18–22. Manchester: Coling 2008 Organizing Committee, pp. 865–72.
- Torres Cacoullous, Rena, and Catherine E. Travis. 2018. *Bilingualism in the Community: Code-Switching and Grammars in Contact*. Cambridge: Cambridge University Press. [CrossRef]
- Treffers-Daller, Jeanine. 2005. Evidence for insertional codemixing: Mixed compounds and French nominal groups in Brussels Dutch. *International Journal of Bilingualism* 9: 477–508. [CrossRef]
- Treffers-Daller, Jeanine. 2009. *Code-Switching and Transfer: An Exploration of Similarities and Differences*. Cambridge Handbooks in Language and Linguistics. Cambridge: Cambridge University Press, pp. 58–74. [CrossRef]
- Tsimpli, Ianthi Maria, Margreet Vogelzang, Anusha Balasubramanian, Theodoros Marinis, Suvarna Alladi, Abhigna Reddy, and Minati Panda. 2020. Linguistic diversity, multilingualism, and cognitive skills: A study of disadvantaged children in india. *Languages* 5: 10. [CrossRef]
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems 30*. Edited by Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna Wallach, Rob Fergus, S. V. N. Vishwanathan and Roman Garnett. Red Hook: Curran Associates, Inc., pp. 5998–6008.
- Walker, Ute. 2021. From target language to translanguing capabilities. harnessing plurilingual repertoires for language learning and teaching. *Language Education and Multilingualism—The Langscape Journal* 117–34. [CrossRef]
- Wang, Danping. 2019. *Multilingualism and Translanguaging in Chinese Language Classrooms*. Basingstoke: Palgrave Macmillan.

- Wilson, Joshua, Dandan Chen, Micheal P. Sandbank, and Michael Hebert. 2019. Generalizability of automated scores of writing quality in grades 3–5. *Journal of Educational Psychology* 111: 619–40. [[CrossRef](#)]
- Wilson, Sonia. 2021. To mix or not to mix: Parental attitudes towards translanguaging and language management choices. *International Journal of Bilingualism* 25: 58–76. [[CrossRef](#)]
- Winford, Donald. 2003. *An Introduction to Contact Linguistics*. Malden and Oxford: Blackwell.
- Winford, Donald. 2009. On the unity of contact phenomena and their underlying mechanisms. In *Multidisciplinary Approaches to Code-Switching*. Edited by Ludmila Isurin, Donald Winford, and Kees de Bot. Philadelphia: John Benjamins Publishing Company, pp. 279–306.
- Yamashita, Junko, and Nan Jiang. 2010. L1 influence on the acquisition of 12 collocations: Japanese esl users and efl learners acquiring english collocations. *TESOL Quarterly* 44: 647–68. [[CrossRef](#)]
- Yannakoudakis, Helen, Øistein E. Andersen, Ardeshir Geranpayeh, Ted Briscoe, and Diane Nicholls. 2018. Developing an automated writing placement system for esl learners. *Applied Measurement in Education* 31: 251–67. [[CrossRef](#)]
- Yao, Mingfa. 2011. On attitudes to teachers' code-switching in efl classes. *World Journal of English Language* 1: 19. [[CrossRef](#)]
- Yuan, Zheng, and Christopher Bryant. 2021. Document-level grammatical error correction. Paper presented at 16th Workshop on Innovative Use of NLP for Building Educational Applications, Online, April 20. Stroudsburg: Association for Computational Linguistics, pp. 75–84.
- Yuan, Zheng, and Ted Briscoe. 2016. Grammatical error correction using neural machine translation. Paper presented at 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, CA, USA, June 12–17. Stroudsburg: Association for Computational Linguistics, pp. 380–86. [[CrossRef](#)]
- Yuan, Zheng, Felix Stahlberg, Marek Rei, Bill Byrne, and Helen Yannakoudakis. 2019. Neural and FST-based approaches to grammatical error correction. Paper presented at Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications, Florence, Italy, August 2. Stroudsburg: Association for Computational Linguistics, pp. 228–39. [[CrossRef](#)]
- Yuan, Zheng, Shiva Taslimipoor, Christopher Davis, and Christopher Bryant. 2021. Multi-class grammatical error detection for correction: A tale of two systems. Paper presented at 2021 Conference on Empirical Methods in Natural Language Processing, Punta Cana, Dominican Republic, May 10. Stroudsburg: Association for Computational Linguistics, pp. 8722–36. [[CrossRef](#)]