The Institute of Materials,
Minerals and Mining

Routledge
Taylor & Francis Group

Check for updates

Visualizing the knowledge domain of code-switching: a bibliometric review

Xiaoping Liu [a], Lu Yu[b] and Jia Liu[c]

[a]School of English Studies, Tianjin Foreign Studies University, Tianjin, People's Republic of China; [b]School of Interpreting and Translation Studies, Guangdong University of Foreign Studies, Guangzhou, People's Republic of China; [c]Research Institute of Foreign Languages, Beijing Foreign Studies University, Beijing, People's Republic of China

ABSTRACT

Code-switching (CS) is prevalent in bilingual communication. With the bibliometric method CiteSpace, this paper gives an overview of the 1319 bibliometric records (1998–2018) on CS selected from the Web of Science. A co-occurring keyword network and a co-citation network of these publications were constructed. Analyses on the high-frequency keywords, document co-citation, co-cited reference clusters, publications with centrality betweenness and citation bursts have been conducted. The results indicate that the landscape of CS research has been expanded from the initial focus on the lexical/conceptual aspect to today's in-depth explorations on language contact, language control, systematic pedagogy, etc. Topics related to 'multimodality', 'translanguaging', 'switch costs' and 'automatic speech recognition' can represent the emerging trends of CS studies in various perspectives, such as linguistics, social science, cognitive science and computer science.

1. Introduction

Code-switching (CS) takes place when a person changes his/her speaking manner by moving back and forth between two languages. Most frequently, such an alternation takes the form of two subsequent sentences, as when a speaker uses a second language either to reiterate his message or to reply to someone else's statement (Gumperz 1982, 59). Because over half of the world population are bilingual, trilingual and even multilingual, CS abounds in daily communication and affects practically everyone who is in contact with more than one language or dialect (Gardner-Chloros 2009, 4). Against this background, CS, though a relatively young topic, has received more and more attention from scholars in and out of the linguistic field. It is necessary to consider CS with a multidisciplinary mind-set, instead of restricting it to a particular aspect (Gardner-Chloros 2009, 7). CS studies have sprung up; however, some inconsistencies or conflicting ideas of the past findings, like the thorny issues of terminological confusion, such as the distinction between CS and language borrowing, may baffle the researchers or the researchers-to-be in this field (Boztepe 2003, 4). Hence, this overview of the previous

---

research on CS, the first of its kind that uses the bibliometric method, aims to offer the status quo of CS studies and inspire researchers from interdisciplinary fields with diverse angles to join future research.

Recently, bibliometric methods have been more and more popularly used to provide a quantitative inspection on a certain field from a broader perspective (Ellegaard and Wallin 2015). They can not only bring more objectivity to the evaluation of the relevant literature, but also gather various opinions from the scholars working in the same field. This paper adopts CiteSpace (Chen et al. 2012; Synnestvedt, Chen, and Holmes 2005) to identify and summarize the relevant literature on CS in a systematic manner and to explore patterns in this research area. As a tool designed for visualizing knowledge domain, CiteSpace has been adopted in various fields (e.g. Niazi and Hussain 2011; Wang et al. 2019). This paper first introduces research methods and findings in light of CiteSpace 5.3.R4. The results on keyword network, document co-citation, co-cited reference clusters, betweenness centrality and citation bursts are herein reported. Based on this discussion as a starting point, the paper continues with the hot CS issues in diverse fields, namely, linguistics, social science, cognitive science and computer science. Eventually the paper recapitulates and challenges the results from CiteSpace and offers suggestions for future CS efforts.

2. Methods and results

The bibliographic records were extracted from the Web of Science (WoS) Core Collection of Thomson Reuter, with the core data set and expanded data set involved. The 'topic' search of 'code switch*' was performed, and a total of 1333 publications (1998–2018) were obtained. After filtering out the duplicates of these records by CiteSpace, we got 1319 records as the bibliographic data for generating visualizations. A keyword network and visual pictures of cited references from the data retrieved from the WoS were accordingly generated. The analyses based on the keyword network and the co-citation network were conducted. The keyword analysis was used to highlight the foci and hotspots covered in the knowledge domain of CS during 1998–2018. The co-citation analysis was adopted to weigh the impact of authors, studies and publications in the development of CS research. Along with the co-citation analysis, analyses related to cluster, betweenness centrality and citation bursts were also performed with an aim to present the development patterns and emerging trends of CS.

2.1. Keyword network analysis and results

Figure 1 is the keyword network generated from the 1319 bibliographic data. The top 10 high-frequency keywords include 'code-switching', 'language', 'bilingualism', 'English', 'multilingualism', 'Spanish', 'language contact', 'education', 'identity' and 'translanguaging'. The high-frequency keywords, such as 'language', 'bilingualism', 'multilingualism' and 'translanguaging', reflect the 'code' aspect. In particular, as the lingua franca, 'English' dominates the 6000 plus languages in the world communication and frequently appears in CS literature. Other keywords, like 'education' and 'identity', are more concerned with a pedagogical or sociological perspective of CS.

Figure 1. Keyword network.

2.2. Document co-citation analysis and results

The visualization of document co-cited references generated from the 1319 records is seen in Figure 2. Altogether 213 nodes emerged for a 3-year time slice, representing the most cited references from 1998 to 2018. In Figure 2, the recently cited references dominate the centre of the whole figure and the earlier cited ones mainly take up the upper part. The top 5 cited works are written by Gardner-Chloros (2009), Garcia (2009), Matras (2009), Garcia and Li (2014) and Creese and Blackledge (2010).

Another important message entailed in Figure 2 is the score a node gets in the network, representing the betweenness centrality of a node. The higher score a node gets, the more important the role it plays in the network. The nodes which enjoy the high betweeness centrality score can be viewed as the landmark works in the

Figure 2. The visualization of most cited references in CS.

development of CS research (Chen et al. 2012). As a complement to the foregoing top 5 cited works, the top 5 works with the highest betweenness centrality are also covered in this overview, written by Garcia (2009), Myers-Scotton (2002), Dorleijn and Nortier (2009), Edwards and Gardner-Chloros (2007) and Poplack and Meechan (1998). Obviously, Garcia (2009) takes a central position in terms of both co-citation analysis and betweenness centrality.

Another important function of CiteSpace lies in its power to detect the emerging trends in a certain knowledge domain. Based on the data collected, 45 citation bursts have been generated since 2012. Eleven publications with the strongest citation bursts (with strength ≥ 3.6) have been selected to be reported in this paper, as shown in Table 1.

Based on the co-citation analysis, CiteSpace can offer a cluster function, which shows the shared themes among those co-cited works. Figure 3 is a timeline view of the cluster for the knowledge domain of CS, as another way to portray the inter-connectivity of the clusters and the time span in a certain cluster. It can be seen that the clusters like 'multi-modality', 'translanguaging', 'switch costs' and 'automatic speech recognition' only appear more repeatedly in recent years' citations.

The foregoing CS references can be categorized into various fields, including linguistics (e.g. Dorleijn and Nortier 2009; Edwards and Gardner-Chloros 2007; MacSwan 2014; Matras 2009; Myers-Scotton 2002; Poplack and Meechan 1998), social science (e.g. Creese and Blackledge 2010; Garcia 2009; Garcia and Li 2014; Lin 2013), cognitive science (e.g. Gollan and Ferreira 2009; Green and Abutalebi 2013; Grosjean 2010; Muysken 2013) and computer science (e.g. Adel et al. 2013, 2014, 2015). A detailed discussion of the co-occurrence keywords coupled with the co-citation analysis in terms of different categories are introduced in the following section.

Table 1. Top 11 publications with the strongest citation bursts.

Title	Source	Author(s)	Strength	Begin
Language Contact	Cambridge University Press	Matras (2009)	4.3771	2012
Translanguaging in the Bilingual Classroom: A Pedagogy for Learning and Teaching?	*Modern Language Journal*	Creese and Blackledge (2010)	5.4523	2014
Translanguaging: Language, Bilingualism and Education	Palgrave Macmillan	Garcia and Li (2014)	8.3071	2015
Classroom Code-Switching: Three Decades of Research	*Applied Linguistics Review*	Lin (2013)	5.527	2016
Language Control in Bilinguals: The Adaptive Control Hypothesis	*Journal of Cognitive Psychology*	Green and Abutalebi (2013)	4.5951	2016
Language Contact Outcomes as the Result of Bilingual Optimization Strategies	*Bilingualism: Language and Cognition*	Muysken (2013)	4.5951	2016
Bilingual: Life and Reality	Harvard University Press	Grosjean (2010)	4.1307	2016
Code-Switching and the Internet	*Cambridge Handbook of Linguistic Code-Switching*	Dorleijn and Nortier (2009)	4.0972	2016
Grammatical Theory and Bilingual Codeswitching	MIT Press	MacSwan (2014)	3.6675	2016
Combining Recurrent Neural Networks and Factored Language Models During Decoding of Code-Switching Speech	*Proceedings of Interspeech*	Adel et al. (2014)	3.6675	2016
Recurrent Neural Network Language Modeling for Code Switching Conversational Speech	*Proceedings of ICASSP*	Adel et al. (2013)	3.6675	2016
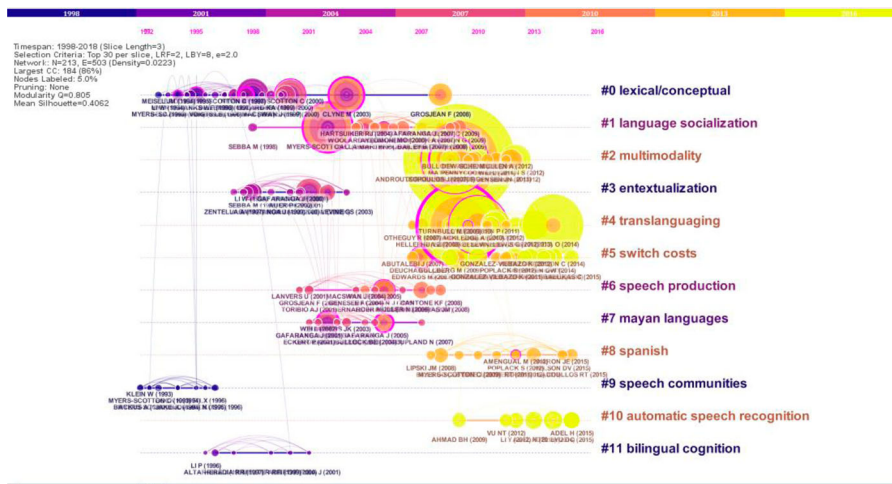
Figure 3. A timeline view of the knowledge domain of CS.

3. CS studies in different categories

As CS is a back-and-forth language shift in nature, it has been explored from diverse angles. Some linguistic researchers have noticed the necessity of investigating CS from diverse fields. For instance, Gardner-Chloros (2009) has written in his book *Code-switching* that a set of grammatical rules is not a one-size-fits-all approach to cope with the variation of CS; rather, contextualized, typological, or social perspectives are in great need. This book features its breadth, covering topics such as language contact, social factors in CS, approaches to CS, as well as CS acquisition. It influences the CS studies a lot and boasts its highest citation counts among all the references. Although to pigeonhole the previous works and the hot topics into categories like linguistics, social science, cognitive science and computer science might be cursory, it is meant to demonstrate these influential contributions from diverse perspectives with their different foci. The reason for this paper to divide CS in linguistics and CS in social science is that CS in linguistics is more related to language itself, to grammar in particular, while CS in the social science is more concerned with the relationships among individuals as members of the society.

3.1. CS in linguistics

CS in linguistic research has mainly been undertaken from the angles of multimodality, language contact and grammar, especially under the framework of the transformative grammar. Although 'language contact' is a social interaction in nature and can be elaborated in Section 3.2 of this paper, scholars are more concerned with the CS grammar in the process of language contact and is reported herein.

Matras (2009) introduces an integrated theory of language contact, covering various aspects, including linguistic socialization, second language acquisition, CS, etc. Language contact occurs when the speakers of different languages interact and their languages influence each other. Pidgins and creoles are the typical products of language contact.

Pidgins are formed between native speakers of different languages through communication, which feature a simplified grammar. Through closer language contact, a pidgin can become a creole which has a distinct grammar of its own and becomes the primary language of a speech community. Drawing on his first-hand observations of child and adult bilingualism, Matras (2009) believes that two main mechanisms operate in language contact, namely, the linguistic forms and the linguistic structures. In particular, he offers the specific parameters to address the distinction between CS and language borrowing, which has long been debated in linguistics (e.g. Myers-Scotton 1992; Poplack 1980; Treffers-Daller 1991): the degree of composition (utterance vs. single lexeme), the functionality of the items (stylistic vs. default use) and the structural integration (non-integrated vs. integrated), to name a few.

Similarly, Myers-Scotton (2002) has proposed the principle for language contact. In her opinion, behind various language contact phenomena (grammatical convergence, language attrition, lexical borrowing, mixed languages, CS, etc.), the same underlying principles are shared by all – the MLF (Matrix Language Frame) model and two auxiliary models. These models have been combined to explain the asymmetrical relationship between the two languages in the bilingual discourse. In CS production, the main language serves as the matrix language while the less dominant language acts as the embedded language, and it is the matrix language that decides the morphosyntactic frame of sentences in CS. The two auxiliary models give detailed explanations about lemmas' operation in the process of CS. The MLF model has opened up new avenues for the researchers to explore.

Another book which also gets a high betweenness centrality is the *Cambridge Handbook of Linguistic Code-switching* (Dorleijn and Nortier 2009). Like other handbooks which serve as reference works for follow-up studies, this one is concise; unlike other handbooks which may centre on one perspective, this one explores CS with rich and interdisciplinary angles, covering the structural, social, cognitive, psychological and phonetic approaches. The influential scholars, such as Gardner-Chloros and Myers-Scotton, mentioned earlier in this paper, have also opined in this comprehensive handbook. Being the first handbook of its kind and featuring its breadth of coverage, this handbook is a must-read for the CS researchers.

Apart from the books about CS in linguistics presented above, two articles are worth mentioning, both of which are published in the *International Journal of Bilingualism*, a top journal in language contact scholarship. In one article, Edwards and Gardner-Chloros (2007) take the bilingual compound verbs as examples to illustrate the verb-formation process, which touches upon grammaticalization, a long-standing theme in the CS research. The first-hand findings from this study pave the way for the fulfilment of Gardner-Chloros' frequently cited book *Code-switching*. In the other article, Poplack and Meechan (1998) argue that the status of lone other-language items should be appropriately determined with the avoidance of obscuring any patterns of true CS that exist. This article is helpful to address the conflict in classifying language mixture.

Besides language contact, CS research in linguistics also witnesses its explorations into CS grammar. For example, MacSwan (2014) contributes to this area by inviting the leading researchers to share their in-depth minimalist programme analyses based on the data gleaned from Spanish–English, Korean–English, German–Spanish, and Hindi–English bilinguals.

In addition to the foregoing CS investigations from the perspectives of language contact and grammar, scholars have also paid attention to the diverse modes of codes, which gives rise to the emergence and popularity of 'multimodality'. 'Multimodality' describes a phenomenon of human communication by means of different modes (such as textual, aural, visual, spatial, digital modes and so on). Researchers have attempted to develop relevant theories or analytical tools for the study of multimodality. Halliday's systemic functional linguistics, Scollon's mediated discourse analysis, Norris' multimodal interaction analysis and recently developed multimodal-type conversation analysis are some typical multimodal discourse analysis methods (Adami 2016, 451). Dorleijn and Nortier (2009) notice that written switching codes in e-mail or real-time chat groups collected from the computer mediated communication can provide insights into the communicative, stylistic, structural, pyscholinguistic, or even diachronic understanding of CS in that they combine the features of spoken and written language. Another interesting point mentioned in this article is that different language pairings demonstrate diverse CS patterns of usage, which is worth further exploring.

3.2. CS in social science

CS in social science is mainly conducted through the explorations about the motivation and pedagogy of CS.

As can be seen from Figure 3, the clusters like 'language socialization', 'entextualization', and 'speech communities' mainly fall into the social category, or the cultural category of CS studies. In a speech community where a group of people share a common language or dialect, people may use codes to socialize and vice versa. The study of language socialization aims to understand how people become competent members of social groups and how a language plays its role in this process. Schieffelin and Ochs (1986) propose that language socialization begins at the right moment when one initiates social contact. It can take place in a speech community, in a family, in school or even in today's pop songs. Apparently, 'language socialization' is more about the motivation or the function of CS, while 'entextualization' is more related to the research method. It means 'the process of rendering discourse extractable, of making a stretch of linguistic production into a unit – a text – that can be lifted out of its interactional setting' (Bauman and Briggs 1990, 73). It stresses a separation of the discourse from its original situational context and a recontextualization of a text in new settings of discourse. Instead of viewing the CS analysis as a static product analysis, this method can inject a dynamic view to the process per se.

One of the top 10 keywords in CS research over the past 2 decades is 'identity', which means the qualities, beliefs, personality, looks or expressions that make a person or a group. It is about negotiating new subject positions at the crossroads of the past, present and future (Block 2006, 39). In other words, an individual is shaped by his or her socio-history and vice versa. Over the past years, sociolinguists have shown particular interest in the relation between CS and identity. It is assumed that one of the most important reasons for people to switch codes lies in their necessity to fit in to certain social or professional settings. This may partially answer the motivation/function of 'language socialization' in certain 'speech communities'. In fact, to meet this 'identify' demand, landmark works about how to teach CS have sprung up over the previous twenty years.

Garcia (2009, 5) claims that 'bilingual education is the only way to educate children in the 21st century'. She maps out the concepts of 'translanguaging'. 'Translanguaging', quite related to 'languaging', was first coined by Williams from Welsh word 'trawsieithu' and later translated by Baker; it was Garcia that made this term more popular. 'Languaging' is a term to capture an ongoing process which is being constituted during interaction with the world. According to Garcia (2009, 140), 'translanguaging is the act performed by bilinguals of accessing different linguistic features or various modes of what are described as autonomous languages, in order to maximize communicative potential'. Baker (2011, 228) views 'translanguaging' as a type of languaging through the use of two languages. Garcia and Li (2014, 137) have developed its definition into the ways that bilinguals use their complex semiotic repertoire to act, to know and to be. Translanguaging can even take place when a person imitates others' accents, or switches among speech, writing and signing.

Based upon the earlier linear subtractive and additive models of bilingualism, Garcia (2009) also proposes the recursive and dynamic models for CS studies. In her view, social, economic, political and historical factors may all influence bilingual teaching practice. Built upon her previous studies, Garcia, collaborated with Wei Li, brings a comprehensive post-structural approach to translanguaging. Garcia and Li (2014) assume that it is the time for bilingual education to shift to translanguaging pedagogy. The translanguaging pedagogy features three innovative 'trans-': 'trans-system and trans-spaces' including meaning-making systems and subjectivities; 'trans-formation' covering criticality, critical pedagogy, social justice and linguistic human rights; 'trans-disciplinary' affordances embracing aspects of socialization, cognition, learning, etc. In their view, this fluid and dynamic way of learning and teaching not only benefits the language acquisition but also addresses the social injustice experienced by ethnic minorities. As noted above, translanguaging, though a relatively young theme of CS, has developed into a popular trend in this knowledge domain.

Besides the concept of 'translanguaging' frequently cited in sociolinguistic CS, some researchers also reported the classroom teaching experiences related to CS. According to Table 1, two articles enjoy their citation bursts in CS research. One of them, co-authored by Creese and Blackledge (2010), dwells on the necessity of bilingual pedagogy in such settings as Chinese and Gujarati community language schools in the UK. Their quantitative research reveals that by means of bilingual instructional strategies (e.g. to annotate texts in classroom teaching), in which two or more languages are used alongside each other, additional positive value to students' performance, lesson accomplishment, and participant confidence can be achieved. The other article written by Lin (2013) reviews a 30-year-long classroom CS research paradigm, approaches and challenges. In sum, these works offer readers the sociological or sociolinguistic lens in the understanding of how and why people switch codes against the multilingual global learning environment.

3.3. CS in cognitive science

Resting on different perspectives, sociology/sociolinguistics and cognitive science/cognitive linguistics/psycholinguistics have different foci on CS – the former is more concerned about the context in language use, while the latter focuses more on using

languaging as a tool, with which the speaker's thought is articulated and transformed into an artifactual form (Swain 2000, 97). CS studies in the cognitive science have been realized through experiments or theoretical models.

Among the clusters in Figure 3, 'switch costs' and 'speech production' are more related to a cognitive or psychological aspect of CS. From a psycholinguistic point of view, CS has mostly been explored in the area of language comprehension, while researchers have also attempted to understand the CS process of speech production (e.g. Filippi, Karaminis, and Thomas 2014; Macizo, Bajo, and Paolieri 2012). Most language comprehension experiments in CS research are lexical decision tasks (e.g. word vs. non-word), and speech production experiments in CS research are mainly in the form of numeral naming or picture naming. A major finding from the previous CS research is that there is an asymmetry of switch costs, where a larger cost can be found for the bilingual to switch into their more dominant language than vice versa (Filippi, Karaminis, and Thomas 2014). Switch costs can also occur when participants take longer to respond when switching from one task to another than repeating the same task from trial to trial (Gollan and Ferreira 2009, 641). Psychologists are particularly interested in how the phenomenon of switch costs is produced, and especially how bilinguals facilitate or inhibit the other language when they are engaged in one language. A lot of psychological models, such as the revised hierarchical model (Kroll and Stewart 1994), the inhibitory control model (Green 1998) and the bilingual interactive activation model (Dijkstra and van Heuven 1998) are quite influential in psychological CS studies.

Green and Abutalebi (2013) and Muysken (2013) are also quite influential for their cognitive explorations in CS. The former study concentrates on how the adaptive control processing imposed by interactional contexts (single language, dual language, and dense CS) may bring changes to neural regions. The latter is more concerned with the optimization strategies that bridge the gaps among the various types of language contact phenomena (CS, language development, language borrowing, and so forth). Grosjean (2010), a prolific psychologist who grew up bilingual (his mother is English and his father French), demystifies bilingualism and bilinguals in his book *Bilingual: Life and Reality*. Using a humorous tone and a few personal anecdotes, the author introduces bilingual thinking and dreaming, the effects of bilingualism on children, etc. Although it met with lukewarm support when it was just published, this book has gained increasing attraction in recent years.

3.4. CS in computer science

The recent appearance of the clusters like 'multimodality', 'translanguaging', 'switch costs' and 'automatic speech recognition' reflects that the CS studies have hitherto focused more on the dynamic CS process. In particular, 'automatic speech recognition' aims to design the algorithms which can convert the non-standard codes into an accurate text transcription. In today's digital-drenched society, the progress of CS research can never be isolated from the development of technology. As the intelligent technology gains momentum and human beings never stop pursuing convenience, it can be predicted that this area will enjoy booming growth.

In fact, the CS phenomenon poses a big challenge to today's automatic speech recognition in computer science. Traditionally, scholars tend to predict CS points by means of

linguistic features, such as word form, part-of-speech tags, or the position of the word (Adel et al. 2015). However, these traditional models yield high error rates. Adel, an often-cited name in natural language processing, has been devoted to the building of new CS models with an aim to help the development of artificial neural networks and human–computer interaction. Her articles in collaboration with other computer scientists (Adel et al. 2013, 2014, 2015) have proposed and developed the FLMs (factored language models). Similar to the traditional models like n-gram models, FLMs predict the probability for the next word in terms of the counts of the factor context occurrences in the training text. Different from the traditional models which use a word context as the predictor, FLMs use a pre-defined factor context (including the vectors of words, morphological classes, word stems or clusters) as the predictor. This attempt is more efficient than the traditional word context because the large vocabulary possessed by any language causes trouble in word context prediction. The results also show that the FLMs indeed outperform the traditional language models in terms of perplexity, mixed error rate and CS point recognition. It can be seen that these bursting areas have been of particular interest to the researchers in recent years, thanks to the development of the artificial intelligence.

4. Discussion and conclusion

CS has drawn upon intense scholarship in the past decades. In this paper, 1319 CS-related records retrieved from the WoS were analysed. Through the visualization of the keyword network, co-citation clusters, betweenness centrality and citation bursts generated by CiteSpace, both the hotspots and emerging trends of CS studies were detected. Basically, it can be seen that the CS research has undergone from the initial focus on the lexical/conceptual aspect to today's in-depth explorations from the perspectives of social science, cognitive science and computer science, etc. Yet it is worth noting that CiteSpace, just like any other biliometric tool, is subject to its limitations.

First, biliometric analysis is based on keyword frequency and citation counts, the principle of which is similar to the idea of 'word frequency' in reading research. Accordingly, to what extent the notion of 'frequency' can literally reflect the importance of a certain reference remains unknown. Another problem related to the counts is that the works published earlier may have a higher probability of being cited and thus enjoy more citation counts. Consequently, how to normalize the citation counts poses a challenge to the biliometric analysis.

Second, the big data provides a large pool for CiteSpace to explore, but the big data is actually a double-edged sword. What one gets from the data is the 'dots' scattered here and there. Although the betweenness centrality or the cluster analysis can present some themes or the interconnected links, it should be noted that the landscape constructed by CiteSpace is still sparse. Obviously, it is researchers that connect the dots to demystify the results obtained from CiteSpace.

Although there are still limitations to the biliometric tool, it is still popular among the academics. Different from the conventional approaches in which scholars have to subjectively utter their understanding towards certain studies, document co-citation analysis indeed offers the readers with the visualization of the structural and temporal patterns and trends in a certain knowledge domain (Trujillo and Long 2018, 1).

Fruitful publications on CS studies have been yielded in the past two decades, and future interdisciplinary studies on CS can be considered from the following aspects. First, how to explore the newly invented symbols, such as emoji icons, pictures, audio files, gestures, to name a few, in a thoroughly and systematic way can be a way-out to CS researchers. Second, it is worth noting that CS grammar, an old topic, has been reactivated in the field, which is reflected by the citation burst of MacSwan (2014). This is most probably due to the demand of the machine translation or the artificial intelligence. A better understanding of CS grammar, especially the processing models involved in CS, can offer the underlying mechanism for the computer programming. Third, CS is innately a human behaviour which involves human senses and emotions. Thus, how to get emotions involved in CS research may be a direction as well. Some computer scientists have already endeavoured to develop the models for human-technical collaboration in the sense-making of big data (e.g. Erwin, Bond, and Jain 2015), which may set an example for future CS studies. In addition, previous CS studies mostly focused on the European bilinguals or multilinguals whose language repertoire shares a lot in common. However, the 6000 plus languages in the world call for a broader view of CS research. Accordingly, it is quite valuable to explore typologically dissimilar language pairs in the global village to have a more general picture of the CS mechanisms.

In today's intimately interdependent world, studying CS can benefit both social and individual development. Also, in today's closely overlapping research fields, a cross-disciplinary study on CS is a must, and there is bound to be more fruitful CS-related outcome by combining the fields of linguistics, sociology, psychology and information technology. Admittedly, the visualization of CS research in the past two decades by the tool of CiteSpace may not offer a complete picture of the status quo in the field; however, it highlights the hotspots and the landmark publications in CS, which can complement or triangulate other reviews of the previous CS research to enrich the topics under discussion.

Disclosure statement

Funding

Notes on contributors

Xiaoping Liu is Associate Professor of English studies at Tianjin Foreign Studies University, China. She earned her PhD in Linguistics and Applied Linguistics. Her research interests include exploring translation processing and non-natives' sentence processing by using eye-trackers.

Lu Yu is a postgraduate student specializing in interpretation studies at School of Interpreting and Translation Studies, Guangdong University of Foreign Studies. She will soon graduate and be an interpreter with a cross-boarder business firm.

Jia Liu is now studying at Beijing Foreign Studies University for her doctorate in Linguistics and Applied Linguistics. She has been devoted to the investigation of emotion word processing by using event-related potential technique.

ORCID

Xiaoping Liu 🔟 http://orcid.org/0000-0003-2524-2957

References

Adami, E. 2016. "Multimodality." In *The Oxford Handbook of Language and Society*, edited by O. Garcia, N. Flores, and M. Spotti, 451–472. Oxford: Oxford University Press.

Adel, H., K. Kirchhoff, D. Telaar, N. T. Vu, and T. Schultz. 2014. "Features for Factored Language Models for Code-Switching Speech." In *Proceedings of International Workshop on Spoken Language Technologies for Under-Resourced Languages*, 32–38.

Adel, H., N. T. Vu, K. Kirchhoff, D. Telaar, and T. Schultz. 2015. "Syntactic and Semantic Features for Code-Switching Factored Language Models." *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 23: 431–440.

Adel, H., N. T. Vu, F. Kraus, T. Schlippe, H. Li, and T. Schultz. 2013. "Recurrent Neural Network Language Modeling for Code Switching Conversational Speech." In *Proceedings of ICASSP*, 8411–8415.

Baker, C. 2011. *Foundations of Bilingual Education and Bilingualism*. Bristol: Multilingual Matters.

Bauman, R., and C. L. Briggs. 1990. "Poetics and Performance as Critical Perspectives on Language and Social Life." *Annual Review of Anthropology* 19 (1): 59–88.

Block, D. 2006. "Identity in Applied Linguistics: Where Are We?" In *The Sociolinguistics of Identity*, edited by T. Omoniyi and G. White, 34–49. London: Continuum.

Boztepe, E. 2003. "Issues in Code-Switching: Competing Theories and Models." *Working Papers in TESOL and Applied Linguistics* 3 (2): 1–27.

Chen, C., Z. Hu, S. Liu, and H. Tseng. 2012. "Emerging Trends in Regenerative Medicine: A Scientometric Analysis in CiteSpace." *Expert Opinion on Biological Therapy* 12 (5): 593–608.

Creese, A., and A. Blackledge. 2010. "Translanguaging in the Bilingual Classroom: A Pedagogy for Learning and Teaching?" *Modern Language Journal* 94 (1): 103–115.

Dijkstra, T., and W. J. B. van Heuven. 1998. "The BIA Model and Bilingual Word Recognition." In *Localist Connectionist Approaches to Human Cognition*, edited by J. Grainger and A. M. Jacobs, 189–225. Mahwah, NJ: Erlbaum.

Dorleijn, M., and J. Nortier. 2009. "Code-switching and the Internet." In *The Cambridge Handbook of Linguistic Code-Switching*, edited by B. E. Bullock and A. J. Toribio, 127–141. Cambridge: Cambridge University Press.

Edwards, M., and P. Gardner-Chloros. 2007. "Compound Verbs in Code-Switching: Bilinguals Making Do?" *International Journal of Bilingualism* 11 (1): 73–91.

Ellegaard, O., and J. A. Wallin. 2015. "The Bibliometric Analysis of Scholarly Production: How Great Is the Impact?" *Scientometrics* 105 (3): 1809–1831.

Erwin, K., M. Bond, and A. Jain. 2015. "Discovering the Language of Data: Personal Pattern Languages and the Social Construction of Meaning from big Data." *Interdisciplinary Science Reviews* 40 (1): 44–60.

Filippi, R., T. Karaminis, and M. Thomas. 2014. "Language Switching in Bilingual Production: Empirical Data and Computational Modelling." *Bilingualism: Language and Cognition* 17 (2): 294–315.

Garcia, O. 2009. *Bilingual Education in the 21st Century: A Global Perspective*. West Sussex: Wiley-Blackwell.

Garcia, O., and W. Li. 2014. *Translanguaging: Language, Bilingualism and Education*. London: Palgrave Macmillan.

Gardner-Chloros, P. 2009. *Code-Switching*. Cambridge: Cambridge University Press.

Gollan, T. H., and V. S. Ferreira. 2009. "Should I Stay or Should I Switch? A Cost-Benefit Analysis of Voluntary Language Switching in Young and Aging Bilinguals." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 35 (3): 640–665.

Green, D. W. 1998. "Mental Control of the Bilingual Lexico-Semantic System." *Bilingualism: Language and Cognition* 1 (2): 67–81.

Green, D. W., and J. Abutalebi. 2013. "Language Control in Bilinguals: The Adaptive Control Hypothesis." *Journal of Cognitive Psychology* 25 (5): 515–530.

Grosjean, F. 2010. *Bilingual: Life and Reality*. Cambridge, MA: Harvard University Press.

Gumperz, J. 1982. *Discourse Strategies*. Cambridge: Cambridge University Press.

Kroll, J. F., and E. Stewart. 1994. "Category Interference in Translation and Picture Naming: Evidence for Asymmetric Connections Between Bilingual Memory Representations." *Journal of Memory and Language* 33 (2): 149–174.

Lin, A. 2013. "Classroom Code-Switching: Three Decades of Research." *Applied Linguistics Review* 4 (1): 195–218.

Macizo, P., T. Bajo, and D. Paolieri. 2012. "Language Switching and Language Competition." *Second Language Research* 28 (2): 131–149.

MacSwan, J. 2014. *Grammatical Theory and Bilingual Code-Switching*. Cambridge, MA: MIT Press.

Matras, Y. 2009. *Language Contact*. Cambridge: Cambridge University Press.

Muysken, P. 2013. "Language Contact Outcomes as the Result of Bilingual Optimization Strategies." *Bilingualism: Language and Cognition* 16 (04): 709–730.

Myers-Scotton, C. 1992. "Comparing Codeswitching and Borrowing." *Journal of Multilingual and Multicultural Development* 13 (1-2): 19–39.

Myers-Scotton, C. 2002. *Contact Linguistics: Bilingual Encounters and Grammatical Outcome*. Oxford: Oxford University Press.

Niazi, M., and A. Hussain. 2011. "Agent-Based Computing from Multi-Agent Systems to Agent-Based Models: A Visual Survey." *Scientometrics* 89 (2): 479–499.

Poplack, S. 1980. "Sometimes I'll Start a Sentence in Spanish Y TERMINO EN ESPANOL: Toward a Typology of Code-Switching." *Linguistics* 18 (7-8): 581–618.

Poplack, S., and M. Meechan. 1998. "Introduction: How Languages Fit Together in Codemixing." *International Journal of Bilingualism* 2 (2): 127–138.

Schieffelin, B., and E. Ochs. 1986. "Language Socialization." *Annual Review of Anthropology* 15: 163–191.

Swain, M. 2000. "The Output Hypothesis and Beyond: Mediating Acquisition Through Collaborative Dialogue." In *Sociocultural Theory and Second Language Learning*, edited by J. P. Lantolf, 97–114. Oxford: Oxford University Press.

Synnestvedt, M. B., C. Chen, and J. H. Holmes. 2005. "CiteSpace II: Visualization and Knowledge Discovery in Bibliographic Databases." In *AMIA Annual Symposium Proceedings*, 724–728.

Treffers-Daller, J. 1991. "Towards a Uniform Approach to Code-Switching and Borrowing." In *Papers for the Workshop on Constraints, Conditions and Models*, edited by ESF Network on Code-switching and Language Contact, 259–279. Strasbourg: European Science Foundation.

Trujillo, C. M., and T. M. Long. 2018. "Document Co-Citation Analysis to Enhance Transdisciplinary Research." *Science Advances* 4 (1): 1–9.

Wang, H., X. Yan, and H. Guo. 2019. "Visualizing the Knowledge Domain of Embodied Language Cognition: A Bibliometric Review." *Digital Scholarship in the Humanities* 34 (1): 21–31.