Management of Crowdsourcing in Language Teaching and Learning: The State-of-the-art and Future Directions

Fernando J. Garrigos-Simon[1]*, Yeamduan Narangajavana-Kaosiri[2]

[1]Universitat Politècnica de València, Valencia, Spain
[2]Universitat de València, Valencia, Spain

Abstract

The new environment of technological development has led to a transformation of teaching and learning methodologies, especially in the field of language teaching. In this field, the so-called Crowdsourcing, applied in other fields, is an essential element to be considered. This paper tries to identify the applications of Crowdsourcing for teaching and learning. The aim is to observe the state of the art and its trends in order to improve efficiency in learning processes. The paper analyses 192 documents on crowdsourcing and language teaching and learning, extracted from the Web of Science in the period 2012 to 2023. The article carries out a literature review and performs a bibliometric and visualisation analysis essentially on journals, authors and keywords. The results show the importance of crowdsourcing in language teaching and learning, and the particularities of aspects such as Crowdteaching and Crowdlearning associated with new methodologies and technological developments. New trends indicate the relevance of including aspects such as, apart from language and linguistic considerations technological developments such as machine learning, natural language processing, sentiment analysis, or classification models. The results offer guidance to researchers and teachers to plan their research and to improve language teaching and learning processes.

Keywords: *Crowdsourcing, Bibliometric, Language Learning, Learning and Teaching Methodologies, Learning Management*

Introduction

Information and communication technologies have had a significant impact on teaching and learning processes, and the area of language teaching and learning is no exception. The development of new technologies, together with the use of social networks by professionals and young people (Manca & Ranieri, 2016), the increasing accessibility of mobile devices, and their potential for improving learning (Lim & Churchill, 2016), has led to the need to transform the management and methodologies of teaching and learning.

In this framework, the so-called Crowdsourcing, with various typologies such as Crowdlearning, Crowdteaching, or Crowdvoting, have gained special relevance.

Crowdsourcing processes have a clear relevance in teaching and learning processes, especially with the development of e-learning. However, in order to improve the implementation of these processes, an understanding of Crowdsourcing, its potential in the area of teaching and learning, and specifically in its applications for language teaching, is necessary. To address this fact, this paper conducts a literature review of Crowdsourcing and language teaching and learning (CLT & L), mainly through the development of a bibliometric analysis. The aim is to understand and learn from the various practical applications of this methodology, its trends and potentials, in order to improve the management of Crowdsourcing in language teaching, and also to open up new areas of research in this field.

Crowdsourcing is a term coined by Howe (2006, p.1) in the area of management. He defined Crowdsourcing as "taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call". Several later authors emphasized the relevance of the participatory process in this methodology, whereby different individuals or organizations can obtain ideas, solutions, knowledge, behaviors useful for problem solving, task accomplishment or the creation of various innovations from the crowd. Looking at this literature, and mainly based on Garrigos et al. (2017), we could define Crowdsourcing in the area of teaching and learning, as "the action of taking a specific teaching or learning task, or a related problem, and proposing its voluntary performance, whether or not through an "open call", to a wide group of people (internal (e.g. students or teachers) or external to the educational entity (external agents or the general public via the Internet)), in exchange for compensation which need not be financial.

The analysis of Crowdsourcing in academia has experienced an explosive growth in recent years, and its study applied to the world of language teaching and learning is no stranger to it. However, we have not found any relevant study of the concept, state of the art, structure, trends, evolution, or potential of Crowdsourcing, neither in education and learning in general, nor specifically in the CLT & L literature. A popular and recognized quantitative method for this analysis, applied in academia in various fields, is the bibliometric study, which analyses, through statistical tools, bibliographic data, presenting them in a user-friendly way.

However, although there is some literature review on knowledge learning with Crowdsourcing (Zhang, 2022), some recent bibliometric analysis on Crowdsourcing research (Lian & He, 2017; Malik et al., 2019), or its development in areas related to science in general (Lenart-Gansiniec et al., 2022), and with some learning methodologies such as gamification in particular (Trinidad et al., 2021) and gamification in digital environments (Schöbel et al., 2021), our research has not found any bibliometric analysis that addresses the application of Crowdsourcing in teaching and learning processes in general, nor in language teaching or learning in particular.

Despite the lack of a review of the CLT & L literature, we believe that such a review is essential. We consider this analysis to be crucial for teachers, to observe new applications, strategies and teaching methodologies; it is vital for students, to visualise new and efficient

ways of learning; and it is essential for researchers, to observe the state of the art of the literature, its trends, and research perspectives and potentials (by detecting the main topics analyzed or not, their impact, or the shortcomings and evolution of research).

Given the relevance of CLT & L and the gaps in its study, the aim of this paper is to carry out an in-depth analysis of the study of CLT & L research, mainly through a bibliometric analysis. The results observe the state of the art in the literature, and the multiple uses of CLT & L, essentially in aspects related to Crowdteaching and Crowdlearning in general (Llorente & Morant, 2015; Llorente et al., 2015), or Crowdvoting, Gamification or particular aspects of e-learning. The results highlight the importance of observing, apart from classical considerations of language and linguistics, considerations associated with technological developments such as machine learning, natural language processing, sentiment analysis, or classification models. These results offer guidance to researchers, teachers and students for the improvement of language learning.

The paper begins by looking at Crowdsourcing applications and typologies, focusing on those relevant to CLT & L. The study then describes the data sources and methodology used. After presenting the results of the bibliometric analysis, the paper concludes by analysing the results and presenting the main conclusions.

Crowdsourcing and Language Teaching and Learning

The study of crowdsourcing in academia has been very relevant in recent years. Proof of this is its use in numerous academic fields (Estellés-Arolas et al., 2015), and through various typologies such as Co-creation (essentially with consumers), Crowd-creation (mainly of content in social networks), Crowdvoting (focused on the votes of the crowd), Crowdwisdom (basically focused on the generation of knowledge) (Brabham, 2013; Howe, 2008; Kleemann et al., 2008; Raineri & Elias Reno, 2022), Creative-Crowdsourcing, Microworking, Crowdsource-workforce- management (Garrigos et al., 2017) or Crowdfunding. The relevance of these processes is highlighted in literature, especially in organizational literature, for providing relatively lower cost, higher quality, and quicker solutions (Blohm et al., 2018), or for their indirect action in the development of aspects such as creativity, innovation or authenticity (Garrigos et al., 2017; Raineri & Elias Reno, 2022).

Within the field of teaching and learning the literature highlights the immense potential of Crowdsourcing, not only for the transformation of educational organizations (Llorente & Morant, 2015; Llorente et al., 2015), but also to transform the classical models of learning in the new digital era (Gómez-Prado et al., 2023). In addition, the literature highlights the relevance of Crowdsourcing for the generation of new ideas, projects, information and knowledge, or even for the financing of educational projects or students' careers. These aspects are highlighted by Llorente and Morant (2015), who explain the existence of 4 types of crowdsourcing techniques in Higher Education institutions:

a)"**Crowdteaching**", associated with teachers' search and management of reliable and high quality educational material (to be shared with other teachers), which enables them to improve the quality and efficiency of their classes, and to optimise teaching and learning time. For example, through Crowdsourcing (e.g. by crowdsourcing material from teachers,

students or the crowd), the teacher can improve, strengthen and update the quantity and quality of the material they provide and share with students, but also better manage their time and can promote students' learning and development outside the classroom (e.g. with the creation, sharing, and pooling of common resources, and possible activities, through repositories for sharing prerequisite courses, curricula, syllabi, videos, and various materials (course resources, assessments, activities, or outcomes) (Raineri & Elias Reno, 2022).

b)"**Crowdlearning**", focused on learning through real case studies, where students participate, teaching and learning from each other in a collaborative way (learning knowledge and execution skills) under the supervision of the teacher. As pointed out by Llorente & Morant, (2015, p.88): "superior knowledge is achieved when a large number of minds are connected (i.e. networked), via ad hoc tools and methods".

c) "**Crowdtuition**", for students to pay their tuition fees through Crowdsourcing methods, or

d) "**Crowdfunding**", used either by teachers (e.g. to finance laboratory and class materials) or students (to finance their stays in different institutions).

To these aspects, we could add, following mainly the Crowdsourcing literature, and Raineri and Elias Reno (2022) others such as: **Co-creation** of material for any topic in any class: for example, teachers, together with other teachers, students, and members of other universities or countries (or using the crowd in general, in the case of **Crowd-creation**), can contribute to the creation of textbooks, new concepts, interesting topics, diverse material such as videos in social networks, or in virtual encyclopedias. To this would be added **Crowdwisdom** (for the completion of tasks or to contribute to the solution of particular problems in a timely manner; for example, "versatile information can be obtained on-demand so that the wisdom of crowds is easily involved to facilitate the knowledge learning process" (Zhang, 2022, p.49), **Crowdvoting** (to select interesting topics, to develop skills such as problem solving, data analysis, and tool practice).

In this sense, the application of Crowdsourcing can offer immense opportunities for teachers and students to improve both teaching and learning processes. As it helps to exchange educational material and knowledge and skills, it optimizes costs and time, facilitates research, and supports students' projects (Gómez-Prado et al., 2023; Lenart-Gansiniec & Sułkowski, 2022) and their professional development. Furthermore, crowdsourcing helps to reinforce learning (inside and outside the classroom) and can be instrumental in improving motivation and incentive to learn (Abdi et al., 2020).

In addition to the initial learning in the classroom, and, in an environment where continuous lifelong learning is increasingly important, Crowdsourcing enables this continuous learning, by allowing to receive a more personalised education adapted to the peculiarities of the learner (specific skills, learning style) and to their specific needs at the specific moment, by allowing access to the best learning materials in a more playful way, and improve their learning efficiency.

Method

This bibliometric uses the WoS Core Collection database as a data source. We use this database because of its importance and popularity in the bibliometric literature, and because

it includes the most relevant journals with the highest standards (Merigó & Yang, 2017). For the selection of the data, all papers with "crowdsourcing" and "language learning" (or "language teaching") as topics are used simultaneously. The observed period includes all documents from 2012 (when the first article on CLT & L appeared) to 30 June 2023. From the initial 355 documents, we performed a filter by considering only articles, reviews and letters, obtaining a final sample of 192 results. The bibliometric indicators were selected for their popularity in the literature, and were: the number of papers and the number of citations to study productivity and the most relevant and influential papers and authors respectively (Merigó et al., 2015; Merigó & Yang, 2017); the h- index to analyse the quality of the papers (Blanco-Mesa et al., 2017; Hirsch, 2005); and the WoS impact factor to observe the dissemination power of the sources (Blanco-Mesa et al., 2017).

The work also uses the VOSviewer software (Van & Waltman, 2010) to map the data, and to be able to observe the structure, networks and nodes of the various analyses. The methodologies used are co-occurrence of all keywords (most common keywords below); co-citation (Small, 1973) (two documents are referenced together in a third), being a more dynamic analysis than bibliographic coupling (Kessler, 1963) (number of identical references between two documents); and co- authorship (main co-authors). The paper develops hierarchical cluster analyses.

Results

The paper conducts six main analyses. Firstly, it looks at the status, progress and structure of CLT & L research. Secondly, the most cited papers are studied. This is followed by an exploration of the main journals. Fourthly, the co-occurrence of author keywords is studied. The fifth section carries out an analysis of co-citations of references, journals and authors. Finally, the co- authorship of countries and institutions is studied.

Status and Evolution of Crowdsourcing, and Language Teaching and Learning

The first published article on CLT & L was Rossen and Lok (2012), entitled "A crowdsourcing method to develop virtual human conversational agents" and published in International Journal of Human-Computer Studies. Since that year, the growth was continuous and exponential, reaching 11 articles per year in 2015, 29 in 2020 and a maximum of 33 in 2022. The growth in citations also follows this positive trend, peaking at 968 citations in 2022, the last full year analysed (the data for 2023 correspond to the first half year). The evolution is shown in Figure 1.

Figure 1

Annual Publications and Citations in Web of Sciences (WoS) in Crowdsourcing, and Language Teaching & Learning
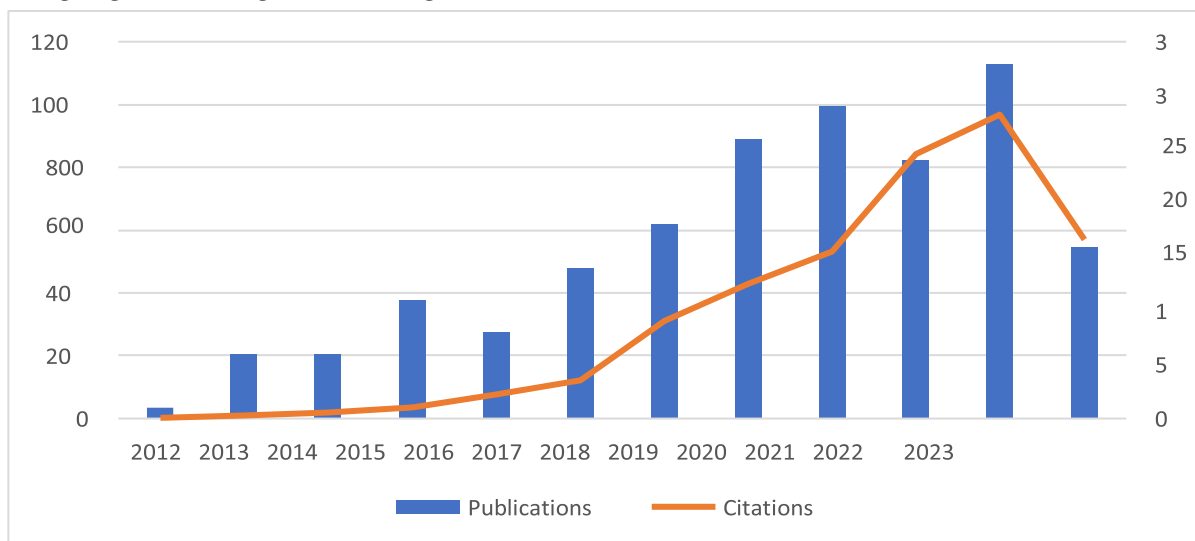
Source: Web of Science

Our sample of 192 papers on CLT & L receive 3924 citations in the period analysed, with an H- index of 25 (25 papers have 25 or more citations). Table 1 illustrates the total citation structure in CLT & L. Three documents receive more than 150 citations (1.56%), and 17.0771% exceed 10 citations.

Table 1

General Structure of Citations in Crowdsourcing, and Language Teaching and Learning

Crowdsourcing, and Language Teaching & Learning				
Number of citations	Number of articles	Accumulated n. of articles	% Articles	% Accumulated articles
≥150	3	3	1.56	1.56
≥50	4	7	2.08	3.65
≥25	19	26	9.90	13.54
≥10	34	60	17.71	31.25
<10	132	192	68.75	100.00
Total	192			

Source: Own elaboration based on WoS 2023

Top Cited Papers in CLT & L

Table 2 shows the top 10 most cited papers, in order to observe the quality, popularity and influence of the main papers in the field (Blanco-Mesa et al., 2017). The most cited paper in our CLT & L sample is Krishna et al. (2017), with over 1800 citations in just 5 years, followed by Castillo et al. (2013) and Brysbaert et al. (2016) with over 150 citations each. Krishna et al. (2017) also leads the annual citation ranking (303.67) in the CLT & L area. This paper, published in the International Journal of Computer vision, looks at a medical perspective on Crowdsourcing for language learning. The medical perspective is also observed by the work of Spasic and Nenadic (2020), noted for being second in annual citations (30.67), who review

clinical text data in machine learning. The article by Castillo et al. (2013) looks at the credibility of information in social media, and is published in Internet Research. Finally, the article by Brysbaert et al. (2016), published in Frontiers in Psychology, looks at a psychological perspective on language knowledge.

Table 2

Top 10 Papers with the Most Citations in Crowdsourcing, and Language Teaching & Learning

R	Journal	TC	Article	Authors	Year	CY
1	IJCV	1822	Visual genome: Connecting language and vision using crowdsourced dense image annotations	Krishna, R.; Zhu, Y.; (…) Fei-Fei, L	2017	303.67
2	IR	186	Predicting information credibility in time- sensitive social media	Castillo, C., Mendoza, M., & Poblete, B	2013	18.60
3	FP	150	How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age	Brysbaert, M.; Stevens, M.; (…) Keuleers, E	2016	21.43
4	JMIRMI	92	Clinical text data in machine learning: systematic review	Spasic, I.; Nenadic G.	2020	30.67
5	JAMIA	61	Identifying reports of randomized controlled trials (RCTs) via a hybrid machine learning and crowdsourcing approach	Wallace, B.C.; Noel-Storr, A. ;(…) Thomas, J.	2017	10.17
6	PO	53	THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images	Hebart, M.N.; Dickter, A.H.; (…) Baker, C.I.	2019	13.25
7	ISPRSJG	50	Extraction of pluvial flood relevant volunteered geographic information (VGI) by deep learning from user generated texts and photos	Feng, Y.; Sester, M.	2018	10.00
8	JMIR	48	Web 2.0-based crowdsourcing for high-quality gold standard development in clinical natural language processing	Zhai, H.; Lingren, T.; (…) Solti, I.	2013	4.80
9	LRE	48	Perspectives on crowdsourcing annotations for natural language processing	Wang, A.; Hoang, C.D.V.; Kan, M.Y.	2013	4.80
10	ESA	47	Query-oriented unsupervised multi-document summarization via deep learning model	Zhong, S.H.; Liu, Y.; (…) Long, J.	2015	5.88

Source: The author based on WoS 2023. R: Ranking; TC: Total Citations; CY: Citations per year. IJCV: International journal of computer vision; IR: Internet Research; FP: Frontiers in psychology; JMIRMI: JMIR Medical Informatics; JAMIA: Journal of the American Medical Informatics Association; PO: Plos One; ISPRSIJG: ISPRS International Journal of Geo-Information; JMIR: Journal of medical Internet research; LRE: Language resources and evaluation; ESA: Expert systems with applications

Leading Areas and Journals in CLT & L

The 192 CLT & L papers belong to 70 research areas. The main ones are Computer Science Information Systems (28.13%), Computer Science Artificial Intelligence (22.40%) and Engineering Electrical Electronic (13.02%), with Linguistics occupying the seventh place. The main sources are IEEE Access (3.12% of the total publications); Language Resources and

Evaluation (3.12%) and Journal of Medical Internet Research (2.60%). Only 5 journals publish 4 or more articles, and only 33 of the 148 sources publish 2 or more, so 77% of the journals publish only one article, which shows the dispersion, but also the potential of the CLT & L area. The H- index of CLLT is led by Language Resources and Evaluation (5), followed by Journal of Medical Internet Research (4). Table 3 shows the journals with 3 or more publications on CLT & L.

Table 3

The Top 9 Journals with Crowdsourcing, and Language Teaching & Learning Publications

R	Journal	APC	H-C	TAP	TCC	ACC	PCC	%APC	IF	≥20	≥10	≥5
1	IEEEA	6	3	73057	15	15	2.50	0.01	3.90			2
2	LRE	6	5	486	112	111	18.67	1.23	2.70	3	3	5
3	JMIR	5	4	6853	95	94	19.00	0.07	7.40	2	4	4
4	IS	4	3	10001	18	18	4.50	0.04	8.10		1	1
5	IJACSA	4	1	8070	3	3	0.75	0.05	0.90			0
6	ESA	3	3	12319	54	54	18.00	0.02	9.50	1	1	1
7	JSS	3	2	2299	33	33	11.00	0.13	3.50	1	1	1
8	WCMC	3	1	5844	10	10	3.33	0.05	2.15		1	1
9	WWWIWIS	3	2	969	15	15	5.00	0.31	3.70		1	1

Source: Own elaboration based on WoS 2023. R: Ranking; H-C: indicates the H index in the area of Crowdsourcing & Language Learning and Teaching; APC: Articles published in CLT & L.; TAP: Total Articles published (2012-2023); TCC: Total citations in CLT & L.: ACC: Articles in which is cited in CLT & L.; PCC. Average of cites by articles in CLT & L; %APC: Percentage of articles published in CLT & L (APC/TAP); IF: Impact Factor; ≥20. ≥10. ≥5: articles with more of 20.10. and 5 citations. IEEEA: IEEE Access; LRE: Language Resources and Evaluation; JMIR: Journal of Medical Internet Research; IS: Information Sciences; IJACSA: International Journal of Advanced Computer Science and Applications; ESA: Expert Systems with Applications; JSS: Journal of Systems and Software; WCMC: Wireless Communications & Mobile Computing; WWWIWIS: World Wide Web-Internet and Web Information Systems

Table 3 shows that only one journal devotes more than 1% of its publications to the area of CLT & L (with only 1.23%) in the period analyzed, namely Language Resources and Evaluation, which is indicative of the incipience of the area. Among the 9 journals with more than 3 publications, Language Resources and Evaluation also leads the number of total citations (112), although it is the Journal of Medical Internet Research, which leads the sources with more citations per article published in CLT & L with 19. This is closely followed by Language Resources and Evaluation, and Expert Systems with Applications with 18.67 and 18.00 citations on average respectively.

Keywords Analysis

In this section we develop the co-occurrence analysis of all keywords (the keywords that appear most frequently in the same works). This analysis helps us to observe the main topics, their distribution, and the research directions in the field of CLT & L. In the 292 publications in our sample, the VOS viewer software observes the existence of 950 co-occurring keywords. Figure 2 illustrates the main ones by their size, the frequency of their co-occurrences and the relative

strength of each relationship. The cluster analysis, shown by different colours, indicates the existence of 7 clusters when we observe a minimum of 3 co-occurrences.

Figure 2

Co-occurrence Network of All Keywords of CLT & L-related Publications

Note: The figure considers a threshold of three occurrences, which shows the 47 keywords with the most frequent co- occurrences, of the 950 keywords. Source: The author, based on WoS 2023

The biggest cluster is the red one (9 items), where there is no outstanding keyword among the top 7, although there are some such as "language", ("speech" or "children"), with a theme aimed at secondary education; also in the ranking of 18 are "model" and "topic modeling", indicating that in this area the Crowdsourcing analyses have been mainly directed towards language synthetics and modelling. The second cluster is green (7 items), which includes three of the 6 keywords with the most occurrences: "machine learning", "social media" and "deep learning", as well as "twitter" and "big data" in the ranking of 18. The dark blue cluster (7 items) includes "crowdsourcing", and "natural language processing" (first and third keywords by number of occurrences), as well as "human computation". The yellow cluster (7 items), includes "classification" and "systems" as the most prominent items. The purple cluster (6 items) is led by "sentiment analysis", the fifth keyword in our ranking, and also includes "annotation". The light blue cluster (6 items) looks at "data mining" and "task analysis". Finally, the orange group (4 items) only includes "artificial intelligence" among the top keywords in our ranking, with artificial intelligence and database (including outside words such as "dataset" and "data bae").

Table 4 shows the top 18 keywords with more than 5 co-occurrences, including also frequencies.

Table 4

The 18 Top All Keywords Co-occurrence of CLT & L Related Publications

R	Keyword (2008-2017)	Oc	Co
1	Crowdsourcing	75	101
2	machine learning	27	65
	natural language processing		
4	social media	14	31
5	sentiment analysis	12	30
6	deep learning	10	16
7	Classification	9	25
8	data mining	6	19
9	Language	6	9
10	task analysis	6	16
11	Twitter	6	15
12	Annotation	5	17
13	artificial intelligence	5	12
14	big data	5	3
15	human computation	5	13
16	Model	5	3
17	Systems	5	9
18	topic modeling	5	6

Source: Own elaboration based on WoS 2023. R: rank; Oc: all keyword occurrences; Co: all keyword co-occurrences link. The colours indicate the various clusters of keywords

Reference, Journal, and Author Co-Citation Analysis

The co-citation analysis of the CLT & L sample helps us to deepen the relationships in our literature. The co-citations of the references indicate that the most co-cited paper by our 192 papers is Snow et al. (2008), with 21 citations, followed by Blei et al. (2003) and Devlin et al. (2018) with 15 citations. The work of Snow et al. (2008), presented at a conference on empirical methods in natural language processing, focuses on Human linguistic annotation and natural language processing tasks. The paper uses Amazon's Mechanical Turk system, a Crowdsourcing platform, for collecting annotations, and combines linguistics with computer science fields to evaluate non-expert annotations for natural language tasks. Blei et al. (2003) in a paper published in Journal of Machine learning Research, describe a generative probabilistic model for collections of discrete data such as text corpora, using text modeling, to provide explicit representations of a document (the paper also looks at results in document modeling, text classification, and collaborative filtering). Devil et al. (2018), in a paper on computation and language, create a language representation model that can be used to pre-train deep bidirectional representations from unlabeled text, but can also be used to create new state-of-the-art models for a wide range of tasks, such as question answering or linguistic inference (they mention eleven natural language processing tasks).

Figure 3 shows the Co-citation of cited references on CLT & L. The paper by Devlin et al. (2018) leads the most important cluster (19 items, red). This cluster focuses

especially on machine learning, looking at areas such as deep learning, neural networks, generative models for machine learning, artificial intelligence, natural language processing and space representation of words (green and orange above). The second cluster (12 items, green), led by Papineni et al. (2002), the sixth most cited author, focuses mainly on areas of computational linguistics, and lexicographic information, including issues focused on machine translation and machine readable-dictionaries, oriented to the learning of languages such as English, or issues focused on use cases of Crowdsourcing for problem solving (Brabham, 2008). The third cluster (dark blue, 10 items), is led by Blei et al. (2003). With a combination of statistics & computer science (and machine learning) this cluster looks at issues related to computational linguistics and social media, but focusing on issues related to classifications and probabilistic topic models.) The fourth group (yellow, 10 items), led by Zheng et al. (2017) focuses on the characteristics of data provided by Crowdsourcing, and data management, specifically this work deals with the veracity of this data. The purple group (9 items), led by Snow et al. (2008), focuses mainly on empirical methods in natural language, and essentially on natural language processing. However, this group also includes questions related to Crowdsourcing as such, aspects related to data mining and informational retrieve, or medical imaging (dark blue and purple clusters previously). Finally, the sixth group is led by Howe (2006), and focuses essentially on the analysis of Crowdsourcing, since this work was the introductory of the concept of Crowdsourcing in the academic literature (although neither this work nor part of the previous ones belong to our sample of 192 works, as they do not refer to articles published in the WoS).

Figure 3

Co-citation of Cited References on CLT & L
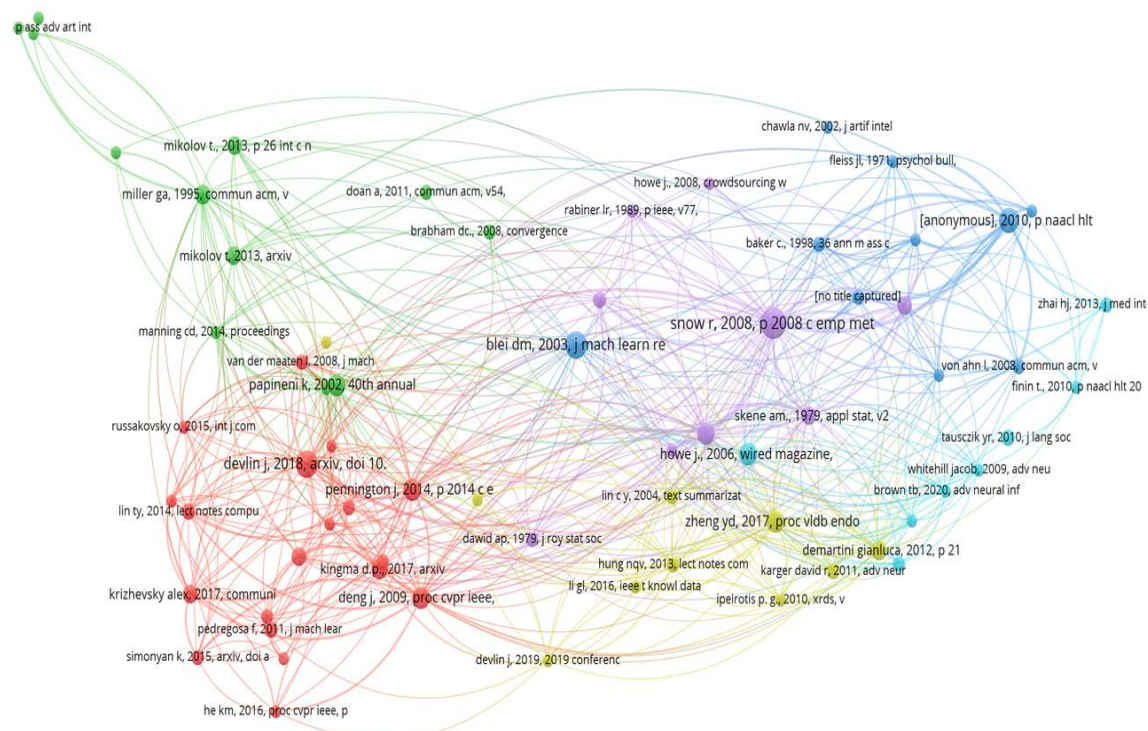
Note: 68 references, of the 8767 cited references that met the threshold of a minimum of 4 citations. Source: The author, based on WoS 2023

The analysis of the journal co-citation network on CLT & L is shown in Figure 4. The most cited sources are Lecturer Notes in Computer Science (148 citations), Arxiv (101 citations), Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (100 citations), AAAI Conference on Artificial Intelligence (69); Journal of Machine Learning Research (64 citations); Advances in Neural Information Processing Systems (62); Communications of the ACM (55); Journal of the American Medical Informatics Association (51); Proceedings of the VLDB Endowment (51) and IEEE Transactions on Pattern Analysis and Machine intelligence (50). The program indicates the existence of six clusters). The first cluster (red colour, 21 items) contains 5 top ranking members: Arxiv, AAAI Conference on Artificial Intelligence, Journal of Machine Learning Research, Advances in Neural Information Processing Systems, and Proceedings of the VLDB Endowment. The group has a computational perspective, focusing primarily on machine learning and artificial intelligence. The second (green colour, 13 items) has Plos one (37 citations) and Behavior Research Methods (34 citations) as top exponents, both outside the top 10. The predominant perspective is psychological. The third cluster (dark blue 11 items) contains three members of the top 10: Lecturer Notes in Computer Science, Proceedings of IEEE Conference on CVPR, and IEEE Transactions on Pattern Analysis and Machine Intelligence. These sources predominantly observe a multimedia and computer vision and pattern recognition perspective. The fourth cluster (yellow, 10 items), led by Computational linguistics (ranked 22nd with 30 citations), has a computer linguistics and generation of emotions in texts perspective. The fifth cluster (purple, 6 items) is led by the Journal of the American Medical Informatics Association, and observes a medical perspective on language learning. Finally the last cluster (light blue, 5 items), where Communications of the ACM is the main source, shows a mainly science perspective (and includes Science and Nature).

Figure 5 shows the results of the co-citation analysis of the leading authors. It illustrates five main clusters. The largest cluster (25 items, red colour) is led by Devlin, the second most cited author (26 citations), and Brysbaert (13 citations), already in the 10th position of the ranking. These authors deal with issues related to natural language processing, language modelling, psycholinguistics and cognitive psychology, studying English word recognition, computational linguistics, or computer vision. The second (19 items, green colour), includes Snow (23 citations), and Blei (22 citations), third and fourth most cited authors respectively. These authors work on issues related to statistics, natural language processing, or computational social science. The third (13 blue items) is led by Howe (18 citations), Von Ahn (17 citations) and Zhang J. (17 citations), fifth and sixth most cited authors. The most important areas are the concept and characteristics of Crowdsourcing, Crowdsourcing techniques, games and human computation. The fourth group (10 items, yellow colour) is composed of Zhang Y., sixth most cited author (17 citations) tied with the previous two, Zheng (14 citations), already in ninth position in the ranking, and Li (13 citations), in 10th position. These authors work on issues associated with machine learning with Crowdsourcing, datamining, crowdsourced data management, or trustworthy (of data). Finally, the fifth (9 items, violet colour) is led by the most cited author, Mikolov (29 citations), who works on aspects such as artificial intelligence, computational linguistics, neural information processing, natural language processing, and neuroscience,

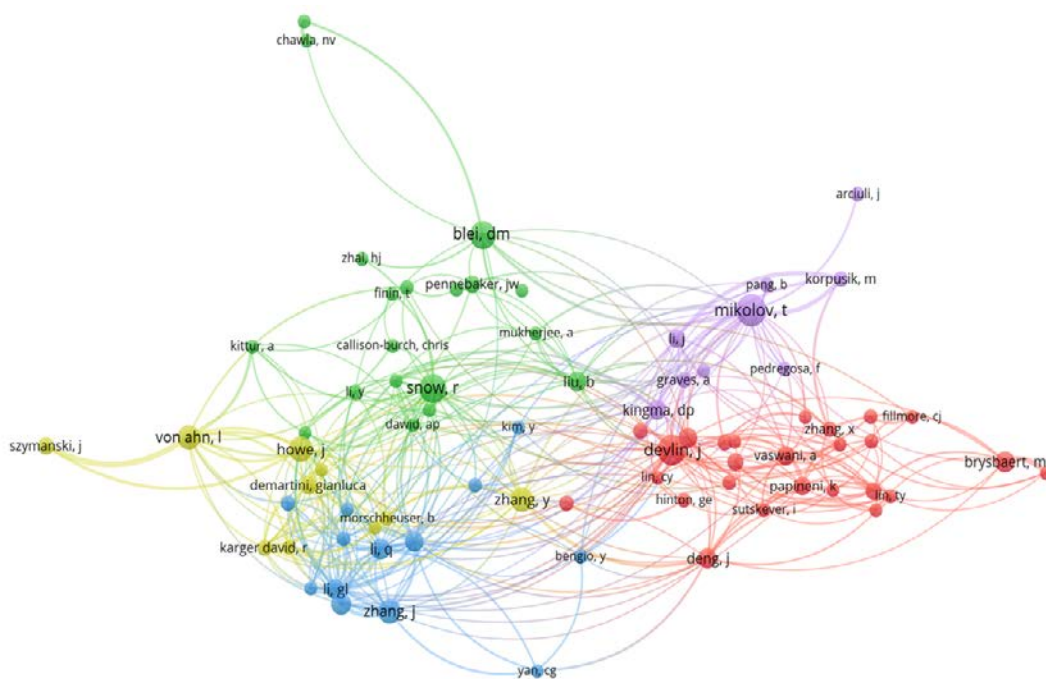with a medical perspective.

Figure 4

Journal Co-citation Network on CLT & L
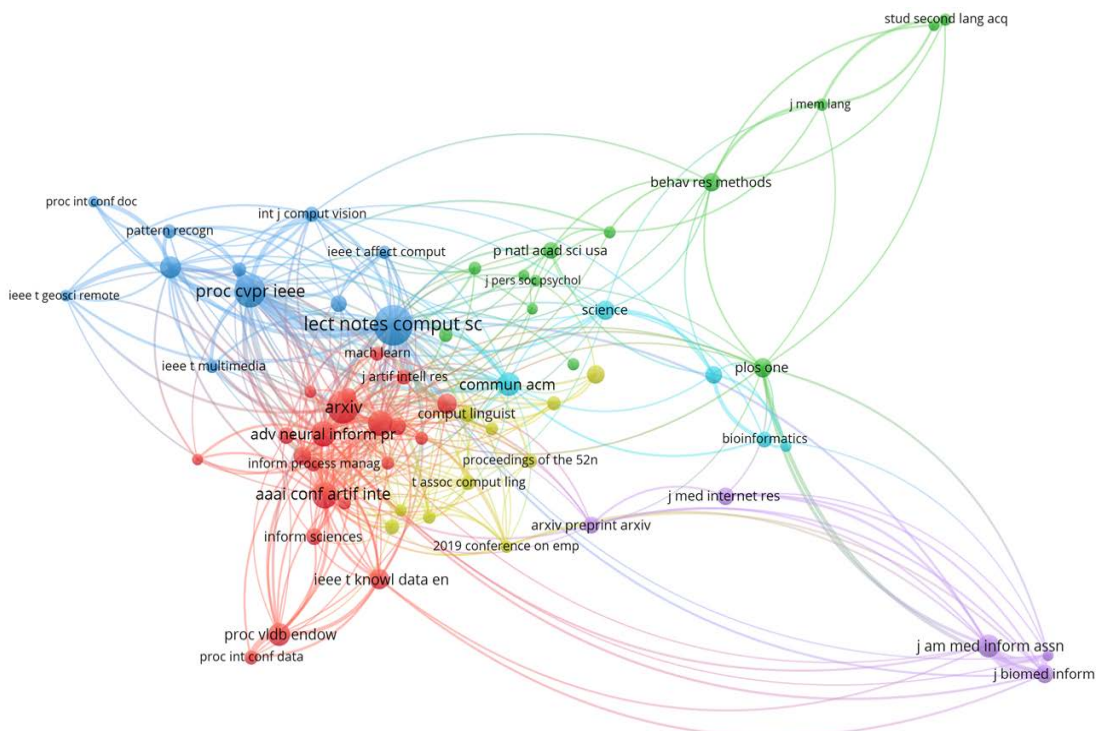
Note: 65 main journals, of the 4916 cited sources by the 192 documents regarding CLL&T, which met the threshold of a minimum number of 15 citations. Source: The author, based on WoS 2023

Figure 5

Author Co-citation Network on CLT & L

Note: 79 authors, of the 6736 cited authors, which met the threshold of a minimum number of citations of 6. Source: The author, based on WoS 2023

Country and University Co-author Analysis

To analyze research collaborations between countries and universities we developed co-authorship analyses. Figure 6 indicates the main countries that have studied the CLT & L literature, in which USA dominates (55 documents, 2632 citations), followed by China (38 doc., 207 citations), England (20 doc., 301 citations), Germany (12 doc., 1991 citations), and Italy (10 doc., 27 citations), as well as Switzerland (7, doc.), Netherland, Saudi Arabia and South Korea (6 doc.), and Poland, Australia, Singapore and India (5 doc.). The main cluster (6, red color) is led by Italy and includes mainly European Countries. China leads the second group (4 items, green), which includes Australia and Saudi Arabia. With the same number of components (4 countries, dark blue), USA and Germany lead the third cluster. Netherland and England lead the fourth and fifth cluster (3 countries, yellow and violet respectively). Finally, in the right corner Singapore and Portugal complete the last cluster (light blue). The graph shows a dispersion mainly between European countries (predominantly on the left side of the graph) and non-European countries (Asian countries on the right side).

Figure 6

Countries Co-authorship Network of CLT & L
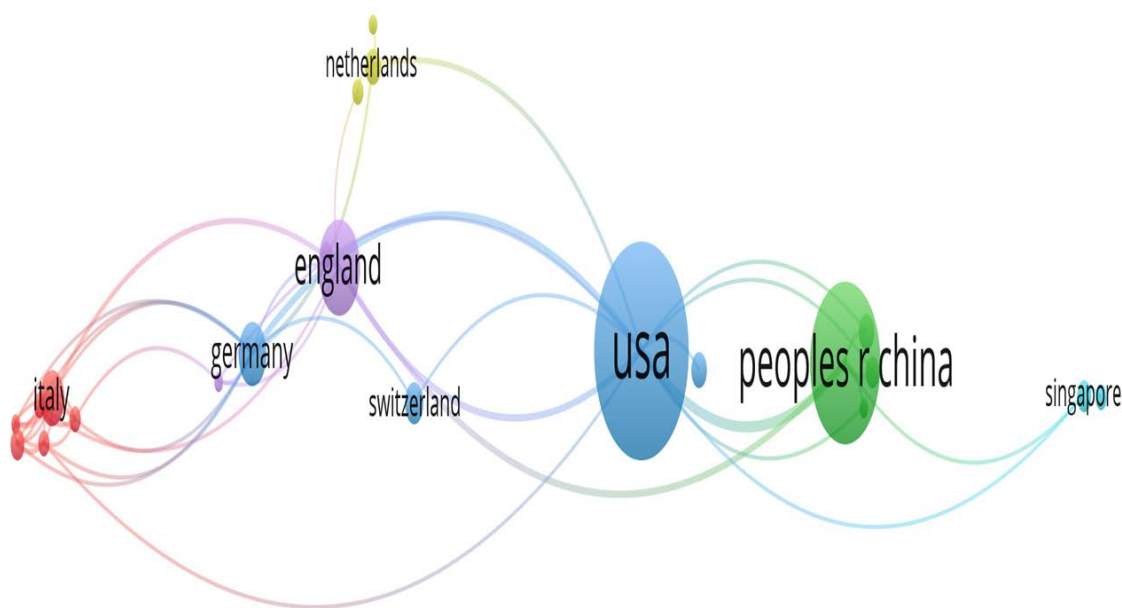
Note: 25 nations of the 58 countries, which met the threshold of a minimum number of three documents of a country. Source: The author, based on WoS 2023.

The leading institutions on CLT & L, by number of publications are The University of Washington (6 documents, 61 citations), Tsinghua University (5 doc., 35 citations), Cardiff University (4 doc., 143 citations), University of Illinois (4 doc., 95 citations), Nanjing University (4 doc., 79citations), City University of Hong Kong (4 doc., 64citations), and MIT (4 doc., 61 citations). The rest have less than 3 published papers. The largest set of connected items consist of 16 items. Most of them are US institutions, in the three main groups, led by the University of Washington (which includes firms such as Google and Microsoft), the Univesity of Illinois, and Carnegie Mellon University respectively. Figure 7 illustrates the density visualization of these organisations.

Figure 7

Institutions Co-authorship Network of CE

Note: 16 organizations connected. Source: The author, based on WoS 2023

Discussion

This paper has attempted to identify the relevance of Crowdsourcing for teaching and learning and specifically for language learning. After defining the concept of Crowdsourcing in general, and specifically in the area of language teaching and learning, the work has focused on its relevant typologies (Crowdteaching, Crowdlearning, Crowdtuition, Crowdfunding, Co-creation, Crowd- creation, Crowdwisdom and Crowdvoting) and its potential use in the area of language teaching and learning. The work has detected a lack of works that address the study of the state of the art, the structure of this research, and its trends and uses, so it has attempted to address a review of this literature, supported by the development of a bibliometric and visualisation analysis of CLT & L related documents. This analysis has visualised the wide relevance and rapid growth of the area, the multidisciplinary character of this literature, as well as the different approaches and perspectives that have addressed it, from those related to linguistics and concrete language learning, to other perspectives related to learning processes from areas such as medicine, psychology, or the development of information technologies. However, the work has detected the lack of both conceptual and methodological work, and the wide dispersion of the literature, with the lack of homogeneous sub-areas of development.

Research on Crowdsourcing and education started in 2009, but research on Crowdsourcing and language teaching and learning only started in 2012, although since then it has grown exponentially, reaching 33 papers just in 2022. However, this literature is very limited, if compared to the literature that has addressed Crowdsourcing in education, to the very extensive literature that has addressed Crowdsourcing in academia, and especially compared to the wide possibilities and potentialities detected in this work. This perception is supported not only by the exponential growth in the number of papers addressing the area, but also by the extensive citation of these papers, with continuous

growth and almost a thousand citations in 2022 alone, of the still small number of papers published up to that date. Krishna et al. (2017) is the paper that leads not only the number of citations, with more than 1800 total citations, but also the number of citations per year, with more than 300. This is followed by two other papers with more than 150 citations each, despite the recentness of these publications, which underlines the potential of the CLT & L area.

The main area of research on CLT & L, if we look at the number of publications, is surprisingly not linguistics (ranked seventh), but various areas related to new technologies, led by Computer Science Information Systems, and the journals IEEE Access and Language Resources and Evaluation (the latter also leading in terms of total citations). However, the dispersion of the research area shows that only 5 journals publish 4 or more articles, with 77% of the 148 sources that address the topic of CLT & L publishing only one article, an aspect that reaffirms the potential and broad growth of the area across very different sources and academic disciplines.

The preponderance of the area of new technologies is confirmed by the co-occurrence analysis of all keywords, which although it includes "Crowdsourcing" in first place, and "language" in ninth place, does not include "learning" or "teaching" as two of the most relevant keywords, and does highlight these technological aspects, with a prominent role for "machine learning" and "natural language processing" in the top three. However, the most numerous cluster, out of the 7 observed, is led by "language", with terms that indicate the relevance of language learning in formal education, but with a focus on language synthetics and modelling. The second cluster highlights the technological character of the area, including 6 of the 15 main keywords, led by "machine learning", and with a focus on "social media" as a source of information. Other important clusters highlight aspects such as "natural language processing" and "human computation", and specific areas around issues such as "classification", "sentiment analysis", "data mining" (and "task analysis") or "artificial intelligence".

The co-citation study illustrates that the most cited papers in the CLT & L literature are Snow et al. (2008) (focusing on natural language processing and human linguistic annotation; and leading a group of references dealing with empirical methods in natural language, and questions related to Crowdsourcing and information retrieving), Blei et al. (2003) (focusing on machine learning and text modelling; and leading the third cluster which also combines statistics and computer science), and Devlin et al. (2018) (on computation and language: and which leads the group of most co- cited references, focused on machine learning and deep learning). Apart from these groups, the second cluster, led by Papineni et al. (2002) looks at issues such as computational linguistics, machine translation (oriented to language learning) or Crowdsourcing for problem solving; the fourth, led by Zheng et al. (2017) focuses on data management (and data veracity); and the sixth cluster led by Howe (2006) (the originator of the term Crowdsourcing), focuses on the concept and analysis of crowdsourcing as such.

The analysis of co-citations of sources highlights the existence of 6 groups: the first, led by Arxiv (the second most cited source) addresses issues of machine learning and artificial intelligence; the second, led by Plos one, has a psychological perspective; the third, led by Lecturer Notes in Computer Science (the most cited source), focuses on issues

of multimedia and computer vision, and pattern recognition; other groups look at computer linguistics, or medical and science perspectives of language learning.

The main author co-citation analysis group is led by Devlin (second most cited author), and deals with natural language processing & modelling, psycholinguistics and cognitive psychology. The second, which includes Snow, and Blei (third and fourth most cited authors), includes authors working on statistics and natural language processing. The third, led by Howe, focuses on the concept and characteristics of crowdsourcing. Other groups analyse issues of machine learning (and trustworthy); artificial intelligence, computational linguistics, or neuroscience with a medical perspective, as does Mikolov, the most cited author.

The co-authorship analysis of CLT & L indicates the dominance of the USA, China, and England and other European countries in our literature. Visualisation of this research shows a dispersion between European countries at one end, and Asian countries (led by China) at the other, with the USA closest to the Asians in the middle. Most of the major institutions also belong to these countries, led by The University of Washington, Tsinghua University, and Cardiff University. However, the connections are only relevant among US institutions, with The University of Washington leading the most relevant group, which includes technology firms such as Google and Microsoft, an aspect that underlines the relevance of technology research in the CLT & L area.

In view of these results, the literature shows that the development of Crowdsourcing for language teaching and learning offers numerous possibilities to improve the processes developed by teachers and students, and thus the efficiency of teaching and learning. Crowdsourcing development offers a paradigm shift in teaching. The facilitation of the connection of diverse individuals and minds, and with it their ad hoc participation and exchange, facilitates the customization and personalization of the education and teaching process. In particular its use is essential, because it allows teachers and students to improve and continuously exchange not only material but also knowledge, tools, or methodologies. Moreover, these processes facilitate research, help to reduce costs, optimize time, reinforce learning, improve motivation and, in short, the efficiency of learning processes. In addition, Crowdsourcing enhances personalized support for students in their learning and professional development (Gómez-Prado et al., 2023; Lenart-Gansiniec & Sułkowski, 2022; Llorente & Morant, 2015), as well as allowing students to feel more engaged, relaxed, independent, and motivated (Ainoutdinova & Blagoveshchenskaya, 2017).

These aspects are essential in an environment where the development of lifelong learning processes is vital, obviously a crucial aspect in language learning. Given this, and as an open model, Crowdsourcing is essential to enable continuous, personalised, versatile learning adapted to the learner's peculiarities (learning style, specific needs, personal skills, specific environment) on demand (Zhang, 2022), allowing access to best practices and materials in a more playful way. In this sense Ainoutdinova and Blagoveshchenskaya (2017), in a paper analyzing the use of crowdsourcing for teaching foreign languages at Russian universities, indicate that Crowdsourcing in addition to catering to students' needs, helps to create better personalized foreign language learning spaces, enabling learners and teachers to create open communities, massive open online

courses (MOOCs) and virtual network platforms to share materials, ideas, or links.

This work has also highlighted the relevance, in the CLT & L literature, of combining Crowdsourcing and the development of new technological innovations, especially because of the great influence that modern information communication technologies in general, and the disruptive development that generative artificial intelligence in particular, is implying in today's society and specifically in foreign language education (Pokrivcakova, 2019). In this sense, several analyses indicate not only the relevance of the opportunities that Crowdsourcing generates for the development of Artificial Intelligence and language learning (Pokrivcakova, 2019; Vaughan, 2017; Zhang, 2022), but also the relevance of its joint use with Deep learning (specifically the importance of deep learning for text classification (Yang et al., 2021)), and Machine Learning (Alenezi, & Faisal, 2020; Raykar et al., 2010; Sheng & Zhang, 2019; Vaughan, 2017) and specifically Natural Language Processing, because the relevance of the ability of a computer to understand, analyze, manipulate, and potentially generate human language (Sabou et al., 2012). In this sense, Rong et al. (2020:1021) state that "Crowdsourcing is a hotspot research field which can facilitate machine learning". While Alenezi and Faisal (2020) stress that Crowdsourcing and Machine Learning can help improve some learning activities in education, such as research, student interactions in massive open online courses, personalized learning (e.g. with the use of online videos designed by non-professionals in the subject), assessment of assignments, exams, generation of tests and materials, detection of cheating or improvement of explanations/reviews, while the combination with Natural Language Processing can help to identify and manipulate multiple choice questions, to calculate the student's recommended grade based on the similarity between students' answers, or to identify correct and incorrect statements in a task. Moreover Gómez Prado et al. (2023) highlight the importance of using Crowdsourcing to create a bank of exam questions or classroom tools and obtain machine learning to improve academic performance.

Conclusion

This work has observed the rapid growth and broad potential of the CLT & L area. It has also looked at the broad perspectives that are integral to it, and that can shape new innovations and practices to improve CLT & L. The work has also analyzed various types of Crowdsourcing that can be applied in the area, including Crowdteaching, Crowdlearning, Crowdtuition, Crowdfunding, Co-creation, Crowd-creation, Crowdwisdom and Crowdvoting. Our work, based mainly on a bibliometric study, can be very relevant in order to visualize fields of application, possible uses, or serve as an exploratory analysis for the planning and future development of research in the area of CLT&T.

Practical Implications

The results have several practical implications, given that agents involved in language teaching and learning should try to understand the potentiality and possibilities of the use of Crowdsourcing, but also to understand its complexity in order to improve the effectiveness and efficiency of teaching and learning processes both inside and outside the

classroom.

Among the main practical applications, we point out that the use of Crowdsourcing is essential for the improvement of teaching processes, given that Crowdleaning is based on the fact that "A single individual cannot hold all available knowledge" (Llorente & Morant, 2015, p.88), and that new technological development, and spread of social networks, have created adaptive online environments which makes crowdsourcing much easier (allowing access to the contributions of a wide amount of people around the world in a short time and at a reduced cost), and then facilitating social learning (Gómez-Prado et al., 2023).

Technological development, currently promoted with generative artificial intelligence and the development of mobile applications, makes it impossible to ignore pedagogical innovations that do not integrate not only the virtual world, but also the participation and generation of multiple content by the crowd for the implementation of all teaching and learning processes. Given this, the very fact of considering Crowdsourcing implies forcing a paradigm shift in the planning and development of language teaching and learning processes by the teacher.

In addition, and more specifically, crowdsourced knowledge building opens up the possibility of collaborative projects where both the professors and the students, but also the crowd, can interact and exchange information, ideas, skills (Gómez-Prado et al., 2023; Llorente & Morant, 2015) or other essential aspects for language teaching and learning. For example, teachers can co-create and share learning resources with other teachers, with students and with the crowd, which can improve the language teaching and learning process.

The quality of lessons is improved as more diverse and innovative content, and more adapted to the personal characteristics and the particular environment and phase of the students' learning process is incorporated into the lessons. In addition, the change of concept, by incorporating more participation and new teaching resources, helps the teacher to free up time to prepare classes in a better and different way, being able to concentrate on: providing instructions in class, indicating what and where the most appropriate material is, suggesting learning tools, focusing their time on collecting and combining perspectives, opinions, ideas, material, votes or tools, or focusing their attention on the motivation of students and the application of active and novel methodologies such as gamification.

On the other hand, the change of teaching methodology, promoted by Crowdsourcing, is important in order to change the paradigm of learning paradigm on the part of the students. It can help them to consider the learning process differently and adapt it to their particular and personal needs.

Furthermore, the specific Crowdsourcing process itself can help these students, and the crowd in general, to obtain, share and complement different ideas, skills, content, or knowledge (e.g. blogs, courses, mobile applications) to increase their knowledge and skills and apply them to solve problems both inside and outside the classroom. In addition, students can be linked with other educational centers, private organizations, public sector or research institutes, allowing to provide innovative solutions to industries and communities, and also their further continuity and implementation in the labor market.

This benefits both, students and society in general, and ultimately educational institutions themselves in particular.

Theoretical Implications

On a theoretical level, this work is especially relevant for researchers interested in CLT & L. In this respect, the paper has looked at the diversity of application and conception of Crowdsourcing processes in language teaching and learning in different areas. It has also observed their state of the art and degree of development in each of them, or the applications in some areas that can be transferred to others. This knowledge is vital in the approach and planning of future research. In particular, the work has highlighted the existence of a wide literature that has developed the topic from technological (even from different sub-conceptions of technological perspectives), statistical, medical, and secondary psychological and educational perspectives.

However, not all academic fields, despite its application, have seen the same development. Our work, and especially the trends and gaps detected in the literature, as well as the degree of evolution of the different developments in CLT & L literature open the door to further work. The in-depth analysis of our results can help researchers from different scientific fields to discover new areas of research in CLT & L, complement research on topics where research is scarce, consider innovations observed in other areas, or deepen the analysis of the connections between the different areas of CLT & L. In this sense, following Garrigos et al. (2021), this work can serve as a basis to further analyze in depth three questions.

First of all, the trends in CLT & L in the different classical areas of research: Besides the areas and sub-areas of technology, and the areas of statistics or medicine, many fields have not been addressed, and in all of them the state of development is not equal. Thus, there is a gap in the literature in most of the classical fields of research in the sciences or social sciences. This opens the door to developments in each of these areas. We believe that particularly essential are new developments that conceive applications of psychology, education, and linguistics in the improvement of language learning, or the observed incipience of aspects related to other sciences (mathematics, physics, engineering, biology, sociology, economics, management, ethics...). This is because the literature has highlighted the almost testimonial and pioneering character of these studies.

Secondly, the development of new areas or topics where CLT & L research is still lacking, and the expansion of research towards fashionable issues that may be related to CLT & L. Thus, new areas that are widely developed in current research do not yet have a similar development in the CLT & L literature. For example, the study of generative artificial intelligence is currently in vogue, and this is observed in some studies on this topic in CLT & L. Other fashionable areas may open the field in aspects related essentially to e-learning and various new technological innovations already detected as fundamentals in our literature (machine learning, natural language processing, deep learning, big data) or not (virtual reality, robotics AI). To the technological developments in fashion, we add the necessary deepening in aspects linked to "social media", "citizen science" (observed in a more relevant way in the more general literature of crowdsourcing and education); and above all developments linked to "sentiment analysis" and "task analysis", which have

been configured among the 10 most relevant topics in the area according to the analysis of keywords. Other applicable fashionable aspects are related to gamification, training, communication, or learning source, learning analytics and assessment. In this sense, CLT & L could consider the most popular and fashionable areas in academia.

Thirdly, the analysis of the papers and sources with the most citations in CLT & L or the analysis of the most relevant keywords can help researchers to look at the research questions that are not only the most fashionable (in the various fields), but also have the greatest impact or are most accepted in the various journals.

Finally, the results show that, in addition to conceptual work, there is a marked lack of empirical work (much of which is reduced to the use of Amazon Mechanical Turk), methodological work, and development of possible uses, which should be attended.

Limitations

The paper has multiple limitations, which may open the door to future research. Firstly, lack of space has prevented us from an in-depth analysis of the CLT & L literature. The still scarce literature may have distorted the analysis, an issue that could be corrected by future work that collects the abundant literature that is currently emerging. The bibliometric and visualization methodology also has its limitations, as it is based on objective data, and should be complemented with more in-depth qualitative studies. The database and sample considered, although the most usual, also have obvious limitations, and could be complemented by extending the research to other types of documents (such as proceeding, doctoral theses, professional papers…), or by considering other databases with data in English or other languages. Subsequent studies could go deeper into the trends observed, into the development of these trends applied to particular theoretical fields, or into the deepening of the networks, clusters, or research themes observed. Finally, new methodologies (linked or not to bibliometric analysis), other software, or other multiple analyses could also be considered.

ORCID

https://orcid.org/0000-0001-6580-9124

https://orcid.org/0000-0002-6773-4493

Ethics Declarations

Competing Interests

No, there are no conflicting interests.

References

Ainoutdinova, I., & Blagoveshchenskaya, A. (2017). Crowdsourcing as an effective technology for teaching foreign languages at Russian universities. In *EDULEARN17 Proceedings* (pp. 1905-1913). IATED. https://doi.org/ 10.21125/edulearn.2017.1401

Abdi, S., Khosravi, H., & Sadiq, S. (2020). Modelling learners in crowdsourcing educational systems. In I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.), *Artificial intelligence in education. AIED 2020. Lecture Notes in Computer Science*, 12164 (pp.3-9). Springer. https://doi.org/10.1007/978-3-030-52240-7_1

Alenezi, H. S., & Faisal, M. H. (2020). Utilizing crowdsourcing and machine learning in education: Literature review. *Education and Information Technologies*, 25(4), 2971-2986. https://doi.org/10.1007/s10639-020-10102-w

Blanco-Mesa, F., Merigó, J. M., & Gil-Lafuente, A. M. (2017). Fuzzy decision making: A bibliometric-based review. *Journal of Intelligent & Fuzzy Systems, 32*(3), 2033-2050. https://doi.org/10.3233/JIFS-161640

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan), 993-1022. https://doi.org/10.1016/b978-0-12-411519-4.00006-9

Blohm, I., Zogaj, S., Bretschneider, U., & Leimeister, J. M. (2018). How to manage crowdsourcing platforms effectively? *California Management Review, 60*(2), 122-149. https://doi.org/10.1177/0008125617738255

Brabham, D. C. (2008). Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence*, 14(1), 75-90. https://doi.org/10.1177/1354856507084420

Brabham, D. C. (2013). *Crowdsourcing*. MIT Press. https://doi.org/10.7551/mitpress/9693.001.0001

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology*, 7, 1116. https://doi.org/10.3389/fpsyg.2016.01116

Castillo, C., Mendoza, M., & Poblete, B. (2013). Predicting information credibility in time-sensitive social media. *Internet Research*, 23(5), 560-588. https://doi.org/10.1108/IntR-05-2012-0095

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*. https://doi.org/10.48550/arXiv.1810.04805

Estellés-Arolas, E., Navarro-Giner, R., & González-Ladrón-De-Guevara, F. (2015). Crowdsourcing Fundamentals: Definition and Typology. In F. Garrigos, I., Gil and S. Estelles (Ed.), *Advances in Crowdsourcing* (pp. 33-48). Cham: Springer. https://doi.org/10.1007/978-3-319-18341-1_3

Garrigos-Simon, F. J., Galdon, J. L., & Sanz-Blas, S. (2017). Effects of crowdvoting on hotels: The Booking. com case. *International Journal of Contemporary Hospitality Management, 29*(1), 419-437. https://doi.org/10.1108/IJCHM-08-2015-0435

Garrigós-Simon, F., Sanz-Blas, S., Narangajavana, Y., & Buzova, D. (2021). The nexus between big data and sustainability: An analysis of current trends and developments. *Sustainability, 13*(12), 6632. https://doi.org/10.3390/su13126632

Gómez-Prado, R., Sánchez-Palomino, J., de los Ángeles Castillo-Butrón, L., Cuya-Velásquez, B. B., Alvarez-Risco, A., Del-Aguila-Arcentales, S., & de las Mercedes Anderson-Seminario, M. (2023). Crowdsourcing and Crowdfunding in Higher Education in Peru. In *Crowdfunding in Higher Education Institutions: Theory and Best Practices* (pp. 159-180). Cham: Springer. https://doi.org/10.1007/978-3-031-30069-1_10

Hirsch, J. (2005). An index to quantify an individual's scientific research output. *Procedings of the National Academy of Sciences. USA, 102*(46), 16569–16572. https://doi.org/10.1073/pnas.0507655102

Howe, J. (2006). The rise of crowdsourcing. *Wired Magazine, 14*(6), 1-4. http://www.wired.com/wired/archive/14.06/crowds.html

Howe, J. (2008). *Crowdsourcing: How the power of the crowd is driving the future of business*. Random House. https://dl.acm.org/doi/10.5555/1481457

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *Journal of the Association for Information Science and Technology, 14*(1), 10-25. https://doi.org/10.1002/ASI.5090140103

Kleemann, F., Voß, G. G., & Rieder, K. (2008). Un (der) paid innovators: The commercial utiliza-tion of consumer work through crowdsourcing. *Science, Technology & Innovation Studies*, 4(1), 5-26. https://doi.org/10.17877/DE290R-12790

Krishna, R., Zhu, Y., Groth, O., Johnson, J., Hata, K., Kravitz, J., ... & Fei-Fei, L. (2017). Visual genome: Connecting language and vision using crowdsourced dense image annotations. *International Journal of Computer Vision*, 123, 32-73. https://doi.org/10.1007/s11263-016-0981-7

Lenart-Gansiniec, R., & Sułkowski, Ł. (2022). *Crowdsourcing for innovation in higher education* (1st ed.). Routledge. https://doi.org/10.4324/9781003227175

Lenart-Gansiniec, R., Czakon, W., Sułkowski, Ł., & Pocek, J. (2022). Understanding crowdsourcing in science. *Review of Managerial Science*, 1-34. https://doi.org/ 10.1007/s11846-022-00602-z

Lim, C. P., & Churchill, D. (2016). Mobile learning. *Interactive Learning Environments*, *24*(2), 273–276. https://doi.org/10.1080/10494820.2015.1113705

Llorente, R., & Morant, M. (2015). Crowdsourcing in higher education. In F.J. Garrigos-Simon et al. (eds.), *Advances in Crowdsourcing*, (pp. 87–96). Springer International Publishing. https://doi.org/10.1007/978-3-319-18341-1_7

Llorente, R., Morant, M., & Garrigos-Simon, F. J. (2015). Crowdlearning, applying social collaboration in higher education. In *New Perspectives in Science Education* (pp. 256-261).

Malik, B. A., Aftab, A., & Ali, N. (2019). Mapping of Crowdsourcing Research: A Bibliometric Analysis. *DESIDOC Journal of Library & Information Technology*, *39*(1), 23-30. https://doi.org/10.14429/djlit.39.1.13630

Manca, S., & Ranieri, M. (2016). "Yes for sharing, no for teaching!": Social Media in academic practices. *The Internet and Higher Education*, *29*, 63-74. https://doi.org/10.1016/j.iheduc.2015.12.004

Merigó, J. M., Gil-Lafuente, A. M., & Yager, R. R. (2015). An overview of fuzzy research with bibliometric indicators. *Applied Soft Computing*, 27, 420-433. https://doi.org/10.1016/j.asoc.2014.10.035

Merigó, J. M., & Yang, J. B. (2017). Accounting research: A bibliometric analysis. *Australian Accounting Review*, 27,71-100. https://doi.org/ 0.1111/auar.12109

Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002, July). *Bleu: a method for automatic evaluation of machine translation*. Proceedings of the 40th annual meeting of the Association for Computational Linguistics (pp. 311-318). https://doi.org/10.3115/1073083.1073135

Pokrivcakova, S. (2019). Preparing teachers for the application of AI-powered technologies in foreign language education. *Journal of Language and Cultural Education*, *7*(3), 135-153. https://doi.org/10.2478/jolace-2019-0025

Raineri, E. M., & Elias Reno, L. S. E. (2022). A Study of Crowdsourcing in Undergraduate Entrepreneurship Education. *Journal of Higher Education Theory & Practice*, *22*(5),115-131. https://doi.org/10.33423/jhetp.v22i5.5206

Raykar, V. C., Yu, S., Zhao, L. H., Valadez, G. H., Florin, C., Bogoni, L., & Moy, L. (2010). Learning from crowds. *Journal of machine learning research*, *11*(4), 1297–132.

Rong, H., Sheng, V. S., Ma, T., Zhou, Y., & Al-Rodhaan, M. (2020). A self-play and sentiment-emphasized comment integration framework based on deep q-learning in a crowdsourcing scenario. *IEEE Transactions on Knowledge and Data Engineering*, *34*(3), 1021-1037. https://doi.org/10.1109/TKDE.2020.2993272

Rossen, B., & Lok, B. (2012). A crowdsourcing method to develop virtual human conversational agents. *International Journal of Human-Computer Studies*, *70*(4), 301-319. https://doi.org/10.1016/j.ijhcs.2011.11.004

Sabou, M., Bontcheva, K., & Scharl, A. (2012, September). *Crowdsourcing research opportunities: lessons from natural language processing*. In Proceedings of the 12th International Conference on Knowledge Management and Knowledge Technologies (pp. 1-8). https://doi.org/ 10.1145/2362456.2362479

Schöbel, S., Saqr, M., & Janson, A. (2021). Two decades of game concepts in digital learning environments–A bibliometric study and research agenda. *Computers & Education*, *173*, 104296. https://doi.org/10.1016/j.compedu.2021.104296

Sheng, V. S., & Zhang, J. (2019, July). *Machine learning with crowdsourcing: A brief summary of the past research and future directions*. In Proceedings of the AAAI conference on artificial intelligence (Vol. 33, No. 01, pp. 9837-9843). https://doi.org/ 10.1609/aaai.v33i01.33019837

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the Association for Information Science and Technology,* 24, 265-269. https://doi.org/10.1002/asi.4630240406

Snow, R., O'Connor, B., Jurafsky, D., & Ng, A. Y. (2008). *Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks*. In Proceedings of the conference on empirical methods in natural language processing, (pp, 254–263). Association for Computational Linguistics. https://doi.org/10.5555/1613715.1613751

Spasic, I., & Nenadic, G. (2020). Clinical text data in machine learning: systematic review. *JMIR Medical Informatics*, *8*(3), e17984. https://doi.org/10.2196/17984

Trinidad, M., Ruiz, M., & Calderon, A. (2021). A bibliometric analysis of gamification research. *IEEE Access*, *9*, 46505-46544. https://doi.org/10.1109/ACCESS.2021.3063986

Van, N. J. E., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*, 523-538. https://doi.org/10.1007/s11192-009-0146-3

Vaughan, J. W. (2017). Making better use of the crowd: How crowdsourcing can advance machine learning research. *The Journal of Machine Learning Research*, *18*(1), 7026-7071. https://doi.org/10.5555/3122009.3242050

Yang, K., Gao, Y., Liang, L., Bian, S., Chen, L., & Zheng, B. (2021). CrowdTC: Crowd-powered learning for text classification. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *16*(1), 1-23. https://doi.org/10.1145/3457216

Zhang, J. (2022). Knowledge learning with crowdsourcing: A brief review and systematic perspective. *IEEE/CAA Journal of Automatica Sinica*, *9*(5),749-762. https://doi.org/10.48550/arXiv.2206.09315

Zheng, Y., Li, G., Li, Y., Shan, C., & Cheng, R. (2017). Truth inference in crowdsourcing: Is the problem solved?. *Proceedings of the VLDB Endowment*, *10*(5), 541-552. https://doi.org/10.14778/3055540.3055547