The Open University

Open Research Online

Citation

Bárkányi, Zsuzsanna; Galindo Merino, M. Mar; Pérez-Bernabeu, Aarón and Solá Simón, Elena (2026). CORINES Corpus de la India en español (Corpus of Indian learners of Spanish). The Open University.

URL

https://oro.open.ac.uk/112737/

DOI

License

(CC-BY-NC-ND 4.0) Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

https://creativecommons.org/licenses/by-nc-nd/4.0/

Policy

This document has been downloaded from Open Research Online, The Open University's repository of research publications. This version is being made available in accordance with Open Research Online policies available from Open Research Online (ORO) Policies

Versions

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding

PROJECT REPORT

CORINES Corpus de la India en español (Corpus of Indian learners of Spanish)

Zsuzsanna Bárkányi (The Open University), María Mar Galindo Merino (University of Alicante), Aarón Pérez-Bernabeu (University of Alcalá)
and Elena Solá Simón (University of Glasgow)

October 2026

Summary

CORINES (Corpus de la India en español) is a reusable corpus of Spanish learner speech that addresses the under-representation of learners in India in research on Spanish language acquisition. Developed through collaboration between The Open University, the University of Alicante and the University of Delhi, the project documents learners' speech within a multilingual context.

Data were collected from 31 participants at the University of Delhi in April 2026. The corpus contains approximately 410 minutes of audio across five tasks: two image-based monologues, reading aloud, a prepared academic presentation and an academic interview. These tasks provide opportunities to examine variation across proficiency levels and communicative demands. Recordings have been orthographically transcribed in CHAT format, aligned with the audio and checked for consistency. Following ethical approval and participant consent for archiving and reuse, the pseudonymised corpus has been prepared for deposit in TalkBank's SLABank repository.

The project establishes a lasting resource for research on pronunciation, multilingual speech acquisition, grammar, vocabulary and academic discourse. It also provides a foundation for future publications, teaching materials and teacher training, supporting more inclusive, empirically informed Spanish language education.

Background

In recent decades, the teaching of Spanish as an additional language (ELE) has seen growing attention towards the oral component of communicative competence, particularly pronunciation. Although traditionally relegated to the background, pronunciation is now recognised as playing a fundamental role in effective communication, intelligibility, and the interlocutor's perception of linguistic competence. Within the scope of our interest in this dimension of language pedagogy

(Bárkányi, Galindo Merino, and Pérez-Bernabeu, 2024), we view learner oral corpora as key tools for empirical research, the design of teaching materials, and teacher training. This project contributes data from Hindi-speaking learners of Spanish in India, a large learner profile that remains under-represented in research on Spanish second and third language phonology and ELE teacher training. India is, since 2023, the most populated country in the world, being Hindi the second largest native community of speakers.

The Indian context also requires an explicitly plurilingual perspective. All the participants of CORINES are highly proficient in English and although Hindi is their dominant language, it does not necessarily mean that Hindi is a participant's only first, home or frequently used language.

Project aims

The project was designed to create a reusable corpus of Spanish learner speech produced by Hindi-speaking learners across A2-C1 proficiency levels to increase the representation of learners in India within documented research resources for Spanish language acquisition. The corpus can support analysis of variation associated with proficiency level and task demands. It provides reusable research infrastructure for researchers in second and multilingual language acquisition, including work on grammatical and lexical features beyond the corpus's primary phonological focus. Moreover, given that two of the tasks explicitly involve oral academic genre (an oral presentation of an academic topic and an academic interview), the corpus is especially suitable for research on academic discourse in an additional language.

Research partnership

The project was developed through collaboration between The Open University in the United Kingdom, the University of Alicante, Spain and the University of Delhi, India, with the supervision of the Head of Germanic and Romance Studies Dpt., Dr. Vijaya Venkataraman. The work was supported by the Open Societal Challenge Programme of The Open University and the project DAELA (*Discurso Académico en Español Lengua Adicional*) at the University of Alicante (Spain) (ref. CIAICO 2023/250, funded by Dirección General de Ciencia e Investigación, Generalitat Valenciana, 2024 – 2027).

Participants and sampling

The corpus contains data from 31 participants recruited through the Spanish programme of the University of Delhi. Participation was voluntary, learners received a 1000INR voucher as a compensation for their time. The sampling procedure is best described as purposive convenience sampling.

Participants' linguistic profile

Participant	**Gender**	**Languages spoken (self-assessed)**	**Level of Spanish (CEFR)**
S01	M	L1 Hindi, English (C1), Maithili (C1), Bengali (A1/2)	B1
S02	F	L1 Manipuri & English Hindi (B2/C1), Bengali (B2), Tripuri (A1/2)	A2

S03	F	L1 Hindi English (C1)	A1.2
S04	F	L1 Hindi English (C1)	A1.2
S05	M	L1 Hindi English (B2/C1), French (A1)	A2
S06	F	L1 Hindi English (C1), French (A1)	B1
S07	F	L1 Hindi, English & Malayalam Mandarin (A1/2), French (A1)	B2/C1
S08	F	L1 Hindi English (C1), Telugu (B1/2), Corean (A1), Japanese (A1), Mandarin (A1)	A2
S09	F	L1 Hindi English (C1)	B1/B2
S10	F	L1 Hindi English (C1), Portuguese (A1)	B2
S11	F	L1 Hindi English (C1), Panjabi (A1/2)	A2/B1
S12	F	L1 Hindi English (C1), Corean (B1/2), Urdu (A2)	B2
S13	F	L1 Hindi English (C1), Urdu (A2), German (A1)	B2/C1
S14	F	L1 Hindi English (B2/C1)	A1.2
S15	F	L1 Hindi English (B2/C1), Panjabi (B1/2)	A1.2
S16	M	L1 Hindi English (C1), Portuguese (A1)	B1/2
S17	F	L1 Hindi English (C1)	A1
S18	M	L1 Hindi English (B2/C1), Maithili (C1/2)	A1
S19	F	L1 Hindi English (C1), German (A1), Portuguese (A1)	B2/C1
S20	F	L1 Hindi English (B2/C1)	B1
S21	F	L1 Hindi English (B2/C1), Corean (A1)	A2/B1
S22	F	L1 Hindi English (C1), Punjabi (A1/2), Portuguese (A1)	B2
S23	F	L1 Hindi English (C1),	B2+
S24	F	L1 Hindi English (B2/C1)	A1.2
S25	M	L1 Hindi English (C1), Urdu (C1), Portuguese (A1)	B1
S26	F	L1 Hindi English (B2/C1), Urdu (A2)	B1
S27	F	L1 Hindi English (B2/C1), Punjabi (B2)	A2

S28	F	L1 Hindi English (C1)	B2+
S29	F	L1 Hindi & English Urdu (A2)	B2+
S30	F	L1 Punjabi English (C1), Hindi (C1), French (A1)	B2+
S31	M	L1 Hindi English (B2/C1), Urdu (A2), Portuguese (A1)	B1/B2

Course enrolment served as participants' Spanish proficiency measure and, if necessary, it was adjusted following the recording session. The aim was to include a variety of proficiency levels.

Data collection

Data were collected at the University of Delhi in April 2026. Each participant completed five short tasks selected to produce speech across different degrees of control, preparation and interaction.

Task	**Procedure**	**Analytical contribution**
1. Participant image	Monologue prompted by an image brought by the participant	Personally meaningful, self-directed, but less spontaneous speech
2. Researcher image	Monologue	Greater comparability across speakers, quasi-spontaneous speech
3. Reading aloud	Short story	More controlled phonological material
4. Academic monologue	Prepared talk prompted by a text selected in advance	Extended planned speech in academic register
5. Academic interview	Short interaction on academic-related topic	Spontaneous speech in academic register

The two image tasks balance individual engagement with cross-participant comparability. A participant-selected image may facilitate sustained speech through familiar content. A common image reduces variation in the prompt and supports comparison of how speakers describe the same visual information. Reading aloud controls selected segments, stress patterns and connected speech environments. The prepared academic task elicits longer stretches of planned speech, although preparation may differ between individuals, followed by a conversation that cannot be planned beforehand. These final two tasks reflect communicative situations familiar to university students.

Recording sessions lasted approximately 30 minutes per participant. The materials used for Tasks 2, 3 and 4 are included in the Appendix. For each participant, Tasks 1 and 2 are in Audio 1, Task 3 in Audio 2 and Tasks 4 and 5 are stored in Audio 3. Audio 3 is not included for four participants. In three cases because their limited Spanish proficiency meant that the recordings could not be transcribed reliably and in one case a technical

error was detected in Audio 3. For two of these participants Audio 1 is also missing as their speech was too fragmented; this means that only their read speech is included in the corpus. In total, the corpus contains approximately 410 minutes of audio.

Participants had received instructions before the recording session and were reminded of them before starting recording. Communication was carried out in in Spanish. If they required clarifications or felt that they could not express their thoughts in Spanish, they often switched to English.

Recordings were carried out in relatively quiet rooms at the University of Delhi using Praat software, dynamic headset microphones (Shure WH20 and Logitech) and USB pre-amplifier. The microphones' cardioid pickup pattern helped to reduce background noise. As a consequence, the researchers' speech was captured only faintly. This limits the corpus's suitability for analysing turn-taking and other interactional features; however, as these phenomena were not a focus of the present project, it was not considered problematic for the intended analyses.

Transcription

Original recordings were backed up and securely stored on password-protected institutional servers to which only the module team had access, working copies created for transcription and checking were also password-protected.

The recordings were transcribed orthographically in CHAT format for use in CLAN and preparation for deposit in TalkBank, SLABank. The transcription captured the participants' utterances and relevant speech phenomena, including false starts, repetitions, filled pauses and other disfluencies. It didn't include fine phonetic detail, intonation or background speech. These boundaries direct researchers back to the audio when the transcript does not contain the detail required for a particular analysis.

Transcript utterances were linked to the recordings using time alignment and files were checked in CLAN for media links and timing.

The transcribers received training in the transcription conventions and completed several standardisation exercises using the same sample recordings. Their transcriptions were compared at several stages, and disagreements were discussed and resolved to improve consistency. A final corpus check ensured that identifiers match across audio and transcripts, and that every public file falls within the participant's consent conditions.

Ethical considerations

The project received ethical approval from the Human Research Ethics Committee of The Open University (Reference number: 2026-1305-2). Participants received detailed information about the study and their right to withdraw at any time. They consented to the archiving and re-use of their pseudonymised audio recordings and the corresponding transcripts by other researchers. They understood that their data would be stored in a secure research repository such as TalkBank and accessed under appropriate conditions (registration and password required).

Outputs, dissemination and future research

The principal output is the corpus CORINES itself, prepared for deposit in TalkBank, SLABank. It will be a lasting open resource for researchers interested in second and plurilingual speech acquisition, interlanguage of Hindi speakers of Spanish and for Spanish language teachers and educators. Additional outputs include or may include in the future conference papers, articles, teaching and teacher training resources or workshops.

Acknowledgements

We are deeply grateful to Dr Venkatamaran for her invaluable support and for making us feel so welcome at the University of Delhi. Our heartfelt thanks also go to her PhD students for taking care of absolutely everything, even finding food that wasn't spicy, and their friends for showing us around in Delhi. We are equally grateful to all our participants, whose enthusiasm towards Spanish language and Hispanic studies truly touched our hearts.

References

Bárkányi, Zs., M. M. Galindo Merino & A. Pérez-Bernabeu (eds.) (2024) *La integración de la proncunciación en el aula de ELE*. John Benjamins Publishing, 310 pp, ISBN 978 90 272 1786 8

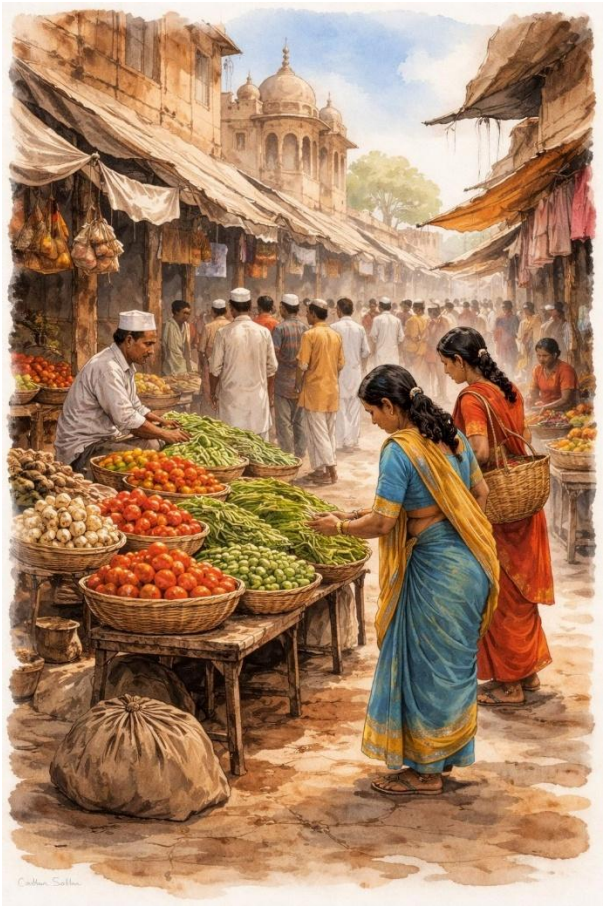
Appendix

Task 2

Opción 1



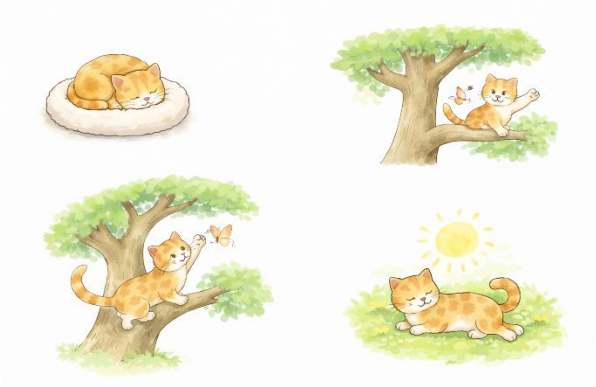
Opción 2

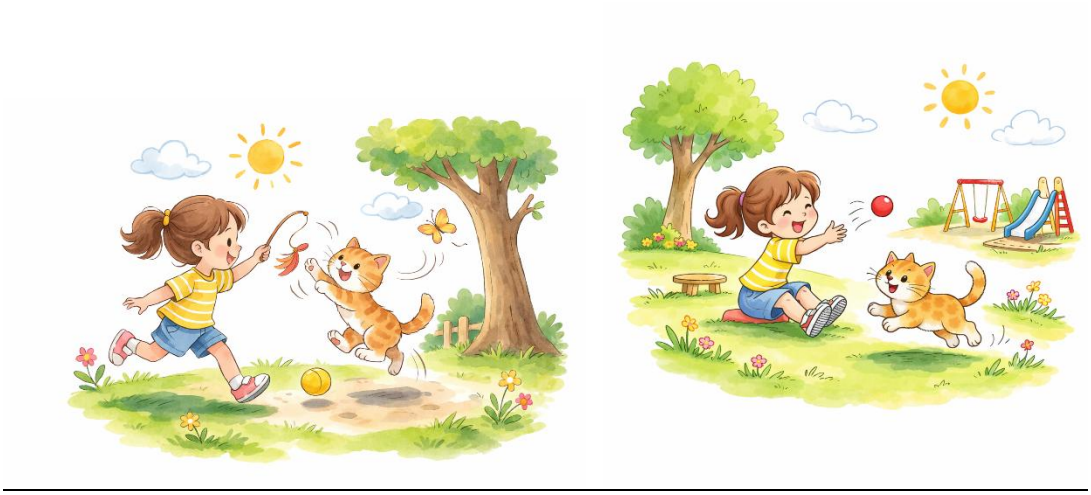




Opción 3











(The images have been generated by ChatGPT-4.1)

Task 2

(Beginners)

Cuenta una vieja leyenda guaraní que Yasí, la diosa luna, quiere conocer la tierra.

- ¿Qué hay en la tierra?

Un día baja del cielo con su amiga Araí, la diosa nube. Las dos llegan como jóvenes hermosas. Ya cansadas buscan un lugar para descansar. Entonces ven una isla y una cabaña.

De repente aparece un jaguar. El animal quiere atacarlas. Salta hacia ellas, pero en ese momento una flecha lo detiene.
—¡Qué susto! —dice Yasí.

Un cazador llamado Tupá las salva. A la mañana siguiente, vuelven al cielo y Yasí planta semillas mágicas. Al día siguiente crecen árboles con hojas verdes y flores blancas. Desde ese día, los guaraníes toman yerba mate.

(Intermediate and advanced)

Cuenta una vieja leyenda guaraní que Yasí, la diosa luna, quiso conocer la tierra y ver con sus propios ojos las maravillas que apenas distinguía entre la espesura de la selva. ¿Qué secretos guardaba aquel mundo que brillaba bajo su mismo cielo?

Un día con su amiga, Araí, la diosa nube, bajaron a la tierra en la forma de dos jóvenes bellísimas acompañadas de dos cisnes bonitos. Cansadas de recorrer, hallaron una cabaña; pero, al acercarse, descubrieron a un yaguareté acechándolas. El animal saltó con las zarpas listas... y al instante cayó atravesado por una flecha. ¡Qué susto! - gritó Yasí.

Se arriesgó y las liberó Tupá, el cazador. Las diosas agradecieron su valentía y a la mañana siguiente regresaron al cielo. Yasí no pudo olvidar aquella generosidad y empezó a sembrar semillas mágicas. Al alba crecieron árboles de hojas verdes y flores blancas. Desde entonces los guaraníes toman yerba mate.

Task 3

(Beginners)

Option 1. La inteligencia artificial en la educación

Read the text and explain to what extent you agree with what it says, and how you see the role of Artificial Intelligence in education.

Lee el texto y prepara una presentación corta. Puedes explicar hasta qué punto estás de acuerdo con lo que dice, y cómo ves tú el papel de la inteligencia artificial en la educación.

La inteligencia artificial, o IA, es una tecnología nueva e importante. Hoy muchas escuelas y universidades la usan. La IA puede ayudar en el aprendizaje.

La IA puede explicar temas y dar ejercicios. También puede responder preguntas. Así, los estudiantes pueden aprender a su propio ritmo. La IA también ayuda a los profesores. Por ejemplo, puede corregir trabajos o preparar actividades.

Pero la IA también tiene problemas. No todos los estudiantes tienen ordenador o internet en casa. Esto puede crear diferencias. Además, algunos alumnos pueden usar demasiado la IA y no pensar por sí mismos.

Los profesores deben aprender a usar esta tecnología. La IA no puede reemplazar al profesor. El profesor es muy importante porque enseña, explica y ayuda a los estudiantes.

También es importante cuidar la información personal y usar la tecnología con responsabilidad.

La gran pregunta es, ¿la inteligencia artificial es una buena ayuda en la educación o no?

Option 2. La globalización de la educación

Read the text and prepare a short presentation. To what extent has globalization changed the way teaching and learning take place? Do you think this is good or bad?

Lee el texto y prepara una presentación corta. ¿Hasta qué punto ha cambiado la globalización la forma en que se enseña y se aprende? ¿Crees que esto es bueno o malo?

Globalización y educación en el mundo actual

La globalización conecta a los países del mundo. También cambia la educación. Hoy, los estudiantes pueden aprender cosas de otros países. La educación ya no es solo local, también es global.

Gracias a la tecnología, los alumnos usan internet para estudiar. Pueden leer textos de otros lugares, ver vídeos y hacer cursos en línea. También pueden hablar con estudiantes de otros países. Así aprenden sobre otras culturas y formas de vida.

La globalización también cambia lo que se enseña en la escuela. En clase se habla de temas del mundo, como el medio ambiente o los derechos humanos. Además, aprender lenguas extranjeras es muy importante. El inglés es una lengua muy usada, pero también es bueno aprender otras lenguas.

(Intermediate)

Option 1. La inteligencia artificial en la educación

Read the text and explain to what extent you agree with what it says, and how you see the role of artificial intelligence in education.

Lee el texto y prepara una presentación corta. Puedes explicar hasta qué punto estás de acuerdo con lo que dice, y cómo ves tú el papel de la inteligencia artificial en la educación.

La inteligencia artificial en el sistema educativo

La inteligencia artificial (IA) es uno de los cambios tecnológicos más importantes de los últimos años. Con la llegada de nuevas herramientas digitales, la IA ha empezado a tener un papel importante en la educación. Hoy en día, muchas escuelas y universidades se preguntan cómo usar esta tecnología de forma correcta.

La IA ofrece varias ventajas. Una de las más importantes es la personalización del aprendizaje. Algunas herramientas pueden adaptarse al nivel y ritmo de cada estudiante y ofrecer ayuda extra cuando es necesario. Además, la IA puede ayudar a los profesores en tareas como corregir trabajos o preparar materiales, lo que les permite dedicar más tiempo a enseñar y apoyar a los alumnos.

Sin embargo, el uso de la IA también presenta algunos problemas. Uno de ellos es la desigualdad. No todos los estudiantes tienen el mismo acceso a la tecnología, y esto puede crear diferencias en el aprendizaje. También existe el riesgo de que los alumnos dependan demasiado de estas herramientas y practiquen menos habilidades como escribir o pensar de forma crítica.

Otro punto importante es la formación de los docentes. Los profesores necesitan aprender a usar la IA de manera adecuada. La tecnología no debe reemplazar al profesor, sino ayudarlo en su trabajo. La educación sigue siendo un proceso humano que necesita orientación y valores.

Por último, es importante pensar en temas como la protección de datos y el uso responsable de la información. ¿De qué dependerá su éxito?

Option 2. La globalización de la educación

Read the text and explain to what extent you agree with what it says, and how you see the role of artificial intelligence in education.

Lee el texto y prepara una presentación corta. Puedes explicar hasta qué punto estás de acuerdo con lo que dice, y cómo ves tú el papel de la inteligencia artificial en la educación.

Globalización y educación en el mundo actual

La globalización ha cambiado los sistemas educativos en todo el mundo. Hoy en día, la educación no se limita solo a un país, sino que forma parte de un contexto global. Las ideas, los conocimientos y la tecnología circulan entre países y afectan a la manera de

enseñar y aprender. Este proceso ofrece nuevas oportunidades, pero también algunos retos para escuelas, profesores y estudiantes.

Uno de los cambios más importantes es la creación de un aula global. Gracias a la tecnología y a las plataformas digitales, los estudiantes pueden acceder a materiales educativos de otros países, participar en cursos en línea y trabajar con personas de diferentes culturas. Estas experiencias ayudan a los alumnos a conocer mejor el mundo y a comprender otras formas de pensar. Además, los programas de intercambio permiten a muchos estudiantes estudiar en otros países y aprender en contextos educativos distintos.

La globalización también influye en los contenidos que se enseñan. Muchos programas educativos incluyen temas globales, como la ciudadanía mundial, los problemas medioambientales o los derechos humanos. El objetivo es preparar a los estudiantes para entender mejor los problemas actuales. Al mismo tiempo, el aprendizaje de lenguas extranjeras es cada vez más importante, especialmente el inglés, aunque también se valora el aprendizaje de otras lenguas.

(Advanced)

Option 1. La inteligencia artificial en la educación

Read the text and explain to what extent you agree with what it says, and how you see the role of artificial intelligence in education.

Lee el texto y prepara una presentación corta. Puedes explicar hasta qué punto estás de acuerdo con lo que dice, y cómo ves tú el papel de la inteligencia artificial en la educación.

La inteligencia artificial en el sistema educativo: oportunidades y desafíos

La inteligencia artificial (IA) se ha convertido en uno de los cambios tecnológicos más importantes de los últimos años. Desde la aparición de herramientas de IA generativa accesibles al público, su impacto en la educación ha sido inmediato. Escuelas, universidades y docentes se enfrentan ahora a una pregunta clave: ¿cómo integrar esta tecnología de manera eficaz y responsable en el aprendizaje?

La IA ofrece numerosas oportunidades. En primer lugar, permite la personalización del aprendizaje. Los sistemas inteligentes pueden adaptarse al ritmo y nivel de cada estudiante, ofreciendo explicaciones adicionales, ejercicios específicos o retroalimentación inmediata. Esto puede ayudar a mejorar el rendimiento académico y reducir algunas dificultades individuales. Además, la IA puede apoyar al profesorado en tareas administrativas, corrección de actividades o creación de materiales didácticos, liberando tiempo para centrarse en la enseñanza y el acompañamiento del alumnado.

Sin embargo, la incorporación de la IA también plantea desafíos importantes. Uno de los principales es el riesgo de aumentar las desigualdades. No todos los estudiantes tienen el mismo acceso a dispositivos digitales o a una buena conexión a internet. Si no se diseñan políticas inclusivas, la tecnología podría beneficiar solo a quienes ya disponen de más recursos. Además, existe el peligro de un uso inadecuado, como la dependencia excesiva de herramientas automáticas o la pérdida de habilidades básicas de escritura y pensamiento crítico.

Otro aspecto fundamental es la formación docente. Para que la IA sea realmente útil, el profesorado necesita comprender sus posibilidades y limitaciones. La tecnología no debe sustituir al docente, sino complementar su trabajo. La educación sigue siendo un proceso humano que implica valores, orientación ética y desarrollo social, elementos que ninguna máquina puede reemplazar completamente.

Finalmente, es necesario reflexionar sobre cuestiones éticas como la protección de datos, la transparencia de los algoritmos y la responsabilidad en el uso de la información. La integración de la IA en el sistema educativo requiere normas claras y una supervisión adecuada.

Option 2. La globalización de la educación

Read the text and prepare a short presentation. To what extent has globalization changed the way teaching and learning take place? Do you think this is good or bad?

Lee el texto y prepara una presentación corta. ¿Hasta qué punto ha cambiado la globalización la forma en que se enseña y se aprende? ¿Crees que esto es bueno o malo?

Globalización y educación en el mundo actual

La globalización ha transformado profundamente los sistemas educativos en todo el mundo. En un contexto cada vez más interconectado, la educación ya no se limita a fronteras nacionales, sino que se desarrolla en un entorno global en el que circulan ideas, conocimientos, personas y tecnologías. Este proceso ha generado importantes oportunidades, pero también nuevos desafíos para docentes, estudiantes e instituciones educativas.

Uno de los cambios más visibles es la creación de un aula global. Gracias al desarrollo de la tecnología y de las plataformas digitales, los estudiantes pueden acceder a recursos educativos internacionales, participar en cursos en línea y colaborar con personas de otros países. Estas experiencias amplían su visión del mundo y fomentan una mayor comprensión de otras culturas, lo que resulta esencial en una sociedad globalizada. Además, la movilidad académica y los programas de intercambio permiten a muchos estudiantes formarse en contextos educativos diversos y adquirir competencias interculturales.

La globalización también ha influido en los contenidos educativos. Los currículos incorporan cada vez más temas de alcance global, como la ciudadanía mundial, los problemas medioambientales, los derechos humanos o la economía internacional. Este enfoque busca preparar a los estudiantes para comprender y afrontar retos que afectan a toda la humanidad. Al mismo tiempo, el aprendizaje de lenguas extranjeras, especialmente del inglés como lengua de comunicación global, ha ganado importancia, aunque también se reconoce el valor del multilingüismo y de la diversidad lingüística.

Sin embargo, estos cambios no están exentos de dificultades. Uno de los principales problemas es la desigualdad en el acceso a la tecnología y a las oportunidades educativas. No todos los estudiantes disponen de los mismos recursos digitales, lo que puede aumentar la brecha entre distintos grupos sociales y países. Además, existe el riesgo de una cierta homogeneización cultural, en la que los modelos educativos dominantes reduzcan la presencia de enfoques locales y tradiciones propias.