54

Quaderns de Docència Universitària

La intersecció entre la IA i l'anàlisi lingüística

Experiències d'avaluació de la IA en la descripció i l'anàlisi lingüístiques de l'anglès i el català

COORDINACIÓ
Natalia Judith Laso Martín
Elisabet Comelles Pujadas

Universitat de Barcelona

Quaderns de Docència Universitària 54

Títol: *La intersecció entre la IA i l'anàlisi lingüística. Experiències d'avaluació de la IA en la descripció i l'anàlisi lingüístiques de l'anglès i el català*

Primera edició: març del 2026

Recepció de l'original: 03/06/25
Acceptació: 22/08/25

Ediciones OCTAEDRO
Bailèn, 5, pral. - 08010 Barcelona
Tel.: 93 246 40 02 - Fax: 93 231 18 68
www.octaedro.com - octaedro@octaedro.com

Universitat de Barcelona
Institut de Desenvolupament Professional
Passeig de la Vall d'Hebron, 171, 08035 Barcelona
Tel.: 934035175
www.ub.edu/idp/web/ - idp.ice@ub.edu

ISBN: 978-84-1079-195-4

Disseny i producció: Serveis Gràfics Octaedro

AUTORS

Natalia Judith Laso Martín i Elisabet Comelles Pujadas (coord.), Júlia Malet Raich, Ionela Valentina Vasilovici i Èlia Vilà Colomé

AGRAÏMENTS

Volem expressar el nostre sincer agraïment a la Dra. M. Àngels Massip per la seva valuosa codirecció del treball *Efectivitat de ChatGPT en la revisió i correcció de textos*. Igualment, agraïm al projecte *Making the CEFR/CV more user-friendly: fine-tuning descriptors with Learner Corpus Research (LCR) results* (PID2020-117041GA-I00) i al Servei de Tecnologia Lingüística (STeL) de la Universitat de Barcelona per haver-nos permès utilitzar el FineDesc Learner Corpus i el corpus GRERLI, respectivament. Ambdós recursos han estat eines fonamentals per al desenvolupament d'aquest estudi.

ÍNDEX

PRÒLEG

Darrerament, hem observat un creixement exponencial en l'ús de la intel·ligència artificial (IA) en diverses àrees del coneixement, fet que també és evident en la lingüística i l'ensenyament de llengües. Una de les conseqüències més notables d'aquesta expansió tecnològica és l'augment progressiu de propostes de treballs de final de grau (TFG) que aborden l'avaluació del llenguatge generat per IA. Aquesta tendència ens ha portat al grup d'innovació docent consolidat GraLiAng a considerar la incorporació d'aquesta àrea com una línia de recerca dels TFG que supervisem als ensenyaments de Filologia.

El present volum recull tres experiències desenvolupades en el marc d'aquests TFG als ensenyaments de Filologia, on es plantegen dos aspectes crucials: d'una banda, la potencialitat de diverses eines d'IA per proporcionar retroalimentació correctiva de la producció escrita d'aprenents de llengües, i, d'altra banda, la capacitat de diversos bots de conversa per detectar trets lingüístics tan complexos com la ironia.

Aquests dos eixos s'han seleccionat perquè han estat els més tractats pels estudiants en els TFG recentment presentats, dins una línia de recerca més àmplia impulsada per les coordinadores del volum, centrada en l'avaluació del llenguatge generat per la IA (*Evaluation of AI-generated language*). Tot i que, a priori, la detecció de la ironia i la retroalimentació correctiva de textos acadèmics poden semblar àmbits distants, ambdós responen a un mateix interès: analitzar fins a quin punt les eines d'IA poden comprendre, reproduir o avaluar fenòmens lingüístics complexos i, per tant, esdevenir un suport real en la descripció i l'anàlisi lingüístiques.

Aquest interès recent per l'avaluació de l'ús de la IA aplicada a la descripció i anàlisi lingüístiques es deu, en part, al fet que aquestes tecnologies han començat a ocupar un lloc cabdal com a eines de suport en infinitat de tasques acadèmiques dels alumnes de Filologia, com ara la redacció de textos acadèmics, la correcció d'errades ortogràfiques, gramaticals, la millora de l'estil d'escriptura acadèmica, així com la generació de continguts. En aquest sentit, el desenvolupament d'eines d'IA generatives que proporcionen resultats que sovint podrien passar per

produccions humanes planteja un conjunt de preguntes fonamentals en el camp de l'ensenyament i l'aprenentatge de llengües: és possible que una eina d'IA assoleixi el mateix nivell de comprensió i habilitat per identificar i analitzar elements lingüístics, com ara l'humor o la ironia, de la mateixa manera que un humà? Pot, una eina d'IA, proporcionar retroacció correctiva d'un text adaptada a diversos nivells de competència lingüística i a diferents tipologies textuals?

Els docents que hem impulsat aquesta línia de recerca no sols veiem en l'avaluació de l'ús de la IA aplicada a la descripció i l'anàlisi lingüístiques una oportunitat acadèmica, sinó també una necessitat. L'avenç tecnològic no és només una eina potencial per ajudar en l'aprenentatge, sinó que també genera nous reptes que cal afrontar amb rigor acadèmic. En particular, la IA aplicada a la descripció i anàlisi lingüístiques ofereix una eina poderosa per explorar fenòmens lingüístics de manera sistemàtica i amb un alt grau de precisió (Ettinger et al., 2023; Uchida, 2024; Llu et al., 2024). Tot i això, encara hi ha debats oberts sobre quins aspectes de la producció lingüística poden ser eficaçment reproduïts o identificats per una IA i sobre quins aspectes requereixen inevitablement la intervenció humana.

El potencial de la IA no acaba en la producció de textos. Una altra línia que es presenta amb força dins els treballs acadèmics dels estudiants de final de grau dels ensenyaments de Filologia és l'ús de la IA com a eina de retroalimentació correctiva. Aquest volum inclou exemples d'experiències acadèmiques en què els estudiants han utilitzat eines d'IA per rebre correccions automàtiques de produccions escrites per aprenents de llengües i, així, poder avaluar la naturalesa i la qualitat de les retroalimentacions proporcionades. L'avantatge d'aquestes eines és clar: ofereixen una resposta immediata, objectiva i, potencialment, poden ajudar a reduir la càrrega de treball dels professors de llengües. Tot i això, també plantegen un conjunt de qüestions importants. Poden aquestes eines detectar errors complexos de concordança, estructura discursiva i/o d'ús pragmàtic? Fins a quin punt podem garantir que aquestes eines són fiables, i adaptades a diferents nivells de competència lingüística i tipologies textuals?

Finalment, un altre aspecte que presentem en aquest volum és l'exploració de la capacitat de les IA per detectar la ironia en la producció lin-

güística. La ironia és un fenomen (extra)lingüístic subtil que requereix una comprensió profunda de les convencions socials i culturals, així com de les expectatives dels parlants. La detecció de la ironia en un text representa un repte fins i tot per als humans, ja que sovint es basa en elements extralingüístics, com el to, l'entonació o el coneixement compartit entre els interlocutors. L'ús de bots de conversa capaços de detectar ironia obre una porta fascinant cap a noves aplicacions de la IA, però alhora planteja la pregunta de fins a quin punt aquestes tecnologies poden realment comprendre els matisos de la comunicació humana.

Aquest quadern ofereix una recopilació d'experiències que exploren tots aquests aspectes, i aporten resultats que seran de gran utilitat per a futurs investigadors, estudiants i docents interessats en el paper de la IA en estudis sobre la descripció i l'anàlisi lingüístiques, així com en els processos d'ensenyament i aprenentatge de llengües. L'augment de les propostes de TFG que exploren aquestes qüestions reflecteix una realitat ineludible: la IA ja no és només una eina, sinó una àrea d'estudi amb entitat pròpia, que obre noves vies de recerca i ens obliga a replantejar molts dels supòsits fonamentals sobre com abordem l'anàlisi lingüística i com avaluem la producció escrita dels nostres estudiants.

Aquest treball col·lectiu posa en relleu el potencial de la IA no només com una eina de suport, sinó com un objecte d'estudi amb el potencial de transformar la manera com investiguem, ensenyem i entenem la llengua. Els resultats que es presenten aquí són només un punt de partida en una àrea que, sens dubte, continuarà evolucionant en els anys vinents.

1. INTRODUCCIÓ

En les últimes dècades, i especialment en els darrers anys, l'impacte creixent de la IA en el processament del llenguatge natural (PLN) ha causat una transformació radical en la manera com interactuem amb els textos escrits i amb les màquines. La capacitat de la IA per comprendre, corregir i generar llenguatge humà ha obert noves perspectives de recerca i aplicació pràctica en molts camps, com ara l'àmbit educatiu, l'ensenyament de llengües (Kristiawan et al., 2024; Almelhes, 2023; Schmidt i Strasser, 2022; Navarro-Carrascosa, 2024; Pérez-Núñez, 2024) i l'anàlisi lingüística (Machado i Ruiz, 2024; Chen et al., 2024; Hromei et al., 2024). En aquest escenari, volem destacar dues línies d'innovació que tenen un alt valor pedagògic i lingüístic: l'ús de la IA per oferir retroacció autocorrectiva de la producció escrita d'aprenents de llengües i l'aplicació per a la detecció automàtica de la ironia, un recurs comunicatiu molt complex que històricament ha estat resistent a la formulació computacional.

> *La incorporació d'eines d'escriptura assistida mitjançant IA cada cop és més habitual en entorns educatius, especialment en l'ensenyament de llengües i en l'adquisició de competències escrites. Eines com Grammarly, Writefull o MyEssai i Large Language Models (LLM) com ara ChatGPT (OpenAI, 2023) han demostrat una capacitat destacable per identificar errors ortogràfics, gramaticals, semàntics i estilístics, així com per proposar reformulacions per millorar el text, especialment quan es tracta de llengües amb molts recursos, com ara l'anglès o el castellà.*

L'ús d'aquestes eines presenta avantatges clars en l'àmbit educatiu, ja que permeten donar un retroacció immediata i personalitzada, la qual cosa pot ser especialment útil en contextos amb un alt nombre d'estudiants o en entorns d'autoaprenentatge. Així doncs, molts especialistes destaquen la quantitat de retroacció que aquestes eines poden donar (Guo i Wang, 2023), així com la capacitat que tenen de donar-lo de manera immediata, fet que possibilita que els aprenents puguin revisar la seva producció escrita diverses vegades (Taskiran i Goksel, 2022; Wei et al., 2023). No obstant això, l'ús d'aquesta tecnologia també planteja dubtes sobre la qualitat de la retroacció proporcionada, així com qüesti-

ons ètiques sobre l'ús que els estudiants fan d'aquestes eines. Així mateix, la seva aparició obre un debat molt interessant sobre quin paper han de tenir en l'ensenyament de llengües i com incorporar-les a l'aula. En particular, cal qüestionar-se en quina mesura aquests sistemes contribueixen realment a l'aprenentatge o, per contra, poden fomentar una dependència de l'assistència tecnològica o limitar la creativitat i la flexibilitat expressiva i espontània dels usuaris.

Un altre àmbit d'interès dins la intersecció entre la IA i la lingüística és la detecció de la ironia en textos escrits. Aquest fenomen discursiu, que s'emmarca en la pragmàtica, acostuma a implicar una discrepància entre el significat literal i la intenció comunicativa del parlant. Això fa que detectar automàticament aquest tret no sigui una tasca fàcil. A diferència d'altres tipus d'anàlisi textual, com ara l'anàlisi de sentiments o la classificació temàtica, la detecció de la ironia requereix unes habilitats que van més enllà de la forma superficial del text i que inclouen la intencionalitat, la dualitat del significat, juntament amb una comprensió profunda del context, de les expectatives culturals i de les relacions interpersonals implicades en la comunicació (Banasik, 2013; Reyes et al., 2013; Wallace et al., 2014).

Els sistemes de detecció automàtica de la ironia han anat evolucionant a la vegada que avançaven els mètodes aplicats al PLN. Els primers sistemes estaven basats en regles (Tang i Chen 2024; Frenda, 2016) i en enfocaments estadístics que utilitzaven diferents característiques lingüístiques (Reyes et al., 2013; VanHee et al., 2018a), però l'arribada de les xarxes neuronals en l'àrea del PLN va produir, com en altres tasques dins aquest camp, un salt qualitatiu (Wu et al., 2018; González et al., 2019; Ilić et al., 2018). Treballs més recents fan *fine-tuning*[1] de models com BERT i RoBERTa per a la detecció automàtica de la ironia (Xaing et al., 2020; Potamias et al., 2020). Aquests sistemes necessiten corpus per a ser entrenats i avaluats, per tant, l'aparició d'aquests sistemes ha comportat la compilació i anotació de corpus específicament dissenyats per a entrenar-los i avaluar-los, com ara el corpus Reddit Irony Dataset (Wallace et al., 2014) o el corpus presentat en la tasca de detecció de la ironia de SemEval-2018 (Van Hee et al., 2018b). L'aparició de LLM i de

1. En aprenentatge automàtic, procés d'adaptació d'un model prèviament entrenat per a la realització d'una tasca específica.

bots de conversa que permeten que l'humà pugui interactuar amb la màquina fent servir llenguatge natural ha fet que la detecció automàtica de la ironia sigui encara més necessària. Tanmateix, tot i els avenços tècnics, aquest continua sent un camp on encara queda molt per fer. L'avaluació de com aquests nous enfocaments fan aquesta tasca concreta és fonamental a l'hora d'avançar en la detecció automàtica de la ironia. Això planteja la necessitat d'un enfocament més interdisciplinari, que combini coneixement lingüístic, psicològic i computacional, i que consideri també les implicacions ètiques de la interpretació automatitzada de la intenció comunicativa.

Aquesta publicació pretén mostrar tres experiències emmarcades en aquests dos àmbits de l'aplicació de la IA i del PLN –la retroacció autocorrectiva en textos escrits i la detecció de la ironia– com a exemples de l'exploració que es pot fer d'aquestes tecnologies dins el context de grau.

Dues d'aquestes experiències analitzen l'ús de la IA per a oferir retroacció autocorrectiva de la producció escrita d'aprenents de llengües. En concret, una se centra en textos escrits per aprenents d'anglès com a segona llengua mitjançant la comparació de la retroacció proporcionada per dos sistemes que utilitzen IA, i l'altra s'emmarca dins l'ensenyament del català com a L1 i analitza la retroacció proposada per ChatGPT. Tot i que hi ha diverses eines que poden ajudar l'usuari a corregir i millorar el text, com ara Grammarly, Paperpal o arText, l'elecció de ChatGPT com a una de les eines a explorar es deu al fet que és l'eina usada de manera més habitual pels estudiants. Val a dir que la diferència entre la retroacció proporcionada per a una llengua amb molts recursos disponibles com l'anglès i la que es proporciona per a una altra llengua amb molts menys recursos, com el català, és molt significativa.

La tercera experiència que s'inclou en aquesta publicació estudia com tres LLM detecten la ironia en textos escrits en anglès, tot avaluant la capacitat d'aquests sistemes per detectar-la i classificar-la en una taxonomia proposada per l'autora.

Cal remarcar que el grau de profunditat de l'anàlisi varia entre els tres estudis per la naturalesa de les dades i els objectius de cada treball. El dedicat a la detecció d'ironia adopta un enfocament més quantitatiu,

mentre que el de la revisió i correcció de textos en català és qualitatiu i més detallat, atesa la incidència més alta d'errors en aquesta llengua amb menys recursos. Per la seva banda, l'anàlisi de la retroacció en anglès manté també un enfocament qualitatiu, però amb menys profunditat, ja que s'hi van detectar menys errors. Aquestes diferències metodològiques expliquen la variació en l'extensió i el nivell d'anàlisi presentat en cada cas.

Aquests estudis aspiren, així, a contribuir a la reflexió sobre el paper de la IA en la didàctica de la llengua i en la investigació sobre el llenguatge, en un moment de canvi accelerat en què la col·laboració entre disciplines esdevé essencial.

2. EFECTIVITAT DE CHATGPT EN LA REVISIÓ I CORRECCIÓ DE TEXTOS

› **Èlia Vila Colomé**

1. Introducció

En l'actualitat, el progrés accelerat de la tecnologia ha conduït a una proliferació sense precedents de les tecnologies de la informació i la comunicació (TIC) i de la IA, les quals s'han integrat cada vegada de manera més profunda en la nostra vida quotidiana. Aquesta expansió ha transformat la manera com interactuem amb el món que ens envolta i afecta des de les nostres rutines fins als processos industrials més complexos.

Les TIC han transformat profundament la societat facilitant la connexió global i impulsant noves formes d'organització, educació i consum. Paral·lelament, la IA ha avançat significativament amb tècniques com l'aprenentatge automàtic i les xarxes neuronals, cosa que permet als sistemes informàtics imitar la intel·ligència humana, aprendre de dades i prendre decisions. En aquest context emergeix ChatGPT, una implementació específica del model de llenguatge GPT (*Generative Pre-trained Transformer*), desenvolupat per OpenAI. Fa servir una xarxa neuronal de gran escala entrenada per processar i generar textos de manera coherent. ChatGPT està especialment dissenyat per mantenir converses lògiques i rellevants amb els usuaris, basant-se en el context de les interaccions. Aquest model pot ser usat en diverses aplicacions de PLN, i destaca per la seva capacitat d'entendre i generar textos en diversos idiomes, incloent-hi el català. La seva capacitat per comprendre i produir en català ofereix una oportunitat única per a la correcció automatitzada. L'ús de ChatGPT com a potencial instrument de correcció de redaccions en català pot contribuir a l'optimització del temps dels professors. Podria agilitzar el procés de correcció i permetria dedicar més temps a altres tasques. A més, podria proporcionar retroalimentació immediata als usuaris, i afavoriria la seva competència lingüística i la capacitat d'autocrítica. L'objectiu d'aquest treball és comprendre

com[2] la IA pot millorar aquest procés i potenciar les habilitats lingüístiques. Tot i que hi ha iniciatives com el projecte Aina, els recursos disponibles en català són limitats i inferiors als d'altres idiomes com l'anglès, evidenciant la necessitat de noves eines d'IA per a l'aprenentatge del català.

La proposta d'aquesta primera experiència se centra a explorar com aquesta tecnologia pot ser utilitzada de manera concreta per identificar i avaluar problemes en redaccions acadèmiques en català, incloent-hi aspectes com la gramàtica, el lèxic, la cohesió i el contingut.

També es busca determinar el *prompt* més adequat per aconseguir una correcció precisa i rellevant. Els *prompts* són les instruccions que es proporcionen a l'eina d'IA per orientar la generació de respostes. A través d'aquesta exploració, es busca comprendre fins a quin punt ChatGPT pot ser una eina útil per a la correcció i millora de textos escrits en català. També es busca respondre les preguntes de recerca següents:

1. Com pot ChatGPT ser utilitzat eficaçment per avaluar la gramàtica, el lèxic, la cohesió i el contingut en redaccions acadèmiques escrites en català?
2. Quins *prompts* són més efectius per a l'avaluació de redaccions acadèmiques mitjançant ChatGPT? Poden ser adaptats segons els diferents nivells educatius?
3. Quines són les limitacions i els reptes en l'ús de ChatGPT com a eina d'avaluació de redaccions acadèmiques en català, i com poden ser abordats per millorar-ne l'eficàcia i fiabilitat?

2. Metodologia

2.1. Eines

L'eina que s'ha utilitzat en aquest estudi ha sigut un model de llenguatge desenvolupat per OpenAI anomenat ChatGPT. Aquesta eina usa una arquitectura d'aprenentatge automàtic anomenada GPT (*Generative Pre-trained Transformer*) per generar respostes en text basades en les entrades que rep. El model és preentrenat amb una gran quantitat

2. Text o conjunt de dades que un usuari introdueix en un sistema d'intel·ligència artificial per fer-li una petició específica.

de text de diverses fonts i corpus lingüístics amb l'objectiu de dotar-lo d'una comprensió profunda del llenguatge humà. Concretament, s'ha usat GPT-3.5 (Brown et al., 2020), que és una versió millorada de l'arquitectura GPT que destaca per la capacitat per recordar missatges anteriors i utilitzar-los de context en les converses, i per crear respostes coherents i rellevants.

Quan ChatGPT rep un *prompt*, processa el text usant les seves capes d'atenció per entendre el context i els patrons lingüístics presents en l'entrada. A continuació, el model genera una resposta basada en aquesta comprensió, usant una estratègia probabilística per seleccionar les paraules següents de manera coherent i significativa. Això vol dir que el model assigna probabilitats a les paraules següents possibles basant-se en el context proporcionat i en el seu coneixement preexistent. A mesura que el model avança en la generació del text, utilitza el mecanisme d'atenció per donar més pes a certs elements de l'entrada, i així s'assegura que la resposta generada sigui adequada al context, amb una estructura coherent i natural.

L'eficàcia de ChatGPT depèn en gran manera dels *prompts* que s'usin. Per garantir una avaluació precisa en aquesta recerca, és necessari trobar-ne un que defineixi els criteris d'interès i proporcioni un marc de referència clar. Això implica dissenyar *prompts* pertinents per a l'objectiu de l'avaluació, abordant els aspectes de la redacció que es volen avaluar.

2.2. Corpus

Aquesta recerca és posada a prova mitjançant l'ús de textos autèntics elaborats per alumnes i professors, per tal de validar la precisió i la utilitat de ChatGPT en aquest context específic. A través d'aquesta metodologia, s'analitzen casos reals i variats de redaccions en català. Aquesta diversitat de mostres facilitarà una comprensió més profunda de les capacitats i les limitacions de ChatGPT en l'àmbit de la correcció de textos.

El treball es basa en l'anàlisi del corpus GRERLI compilat l'any 1998 i creat en el marc del projecte internacional «Developing literacy in different contexts and in different languages». El corpus va ser creat pel Grup de Recerca per a l'Estudi del Repertori Lingüístic, amb la investi-

gadora Liliana Tolchinsky al capdavant i Joan Perera com a encarregat del subcorpus GRERLI-CAT1. Aquest projecte tenia com a objectiu examinar el repertori lingüístic dels parlants de diverses llengües en una varietat de situacions comunicatives, i explorar les habilitats comunicatives dels participants, segons el seu nivell educatiu.

El corpus GRERLI consta de dos subcorpus: un en castellà, anomenat GRERLI-CAST1, i un en català, conegut com a GRERLI-CAT1. Els textos d'aquests subcorpus van ser obtinguts a través d'un mateix procediment. Cadascun dels participants va observar un vídeo sense text ni veu que mostrava situacions conflictives a l'entorn escolar, i després van produir quatre tipus de textos relacionats amb aquest tema: una exposició oral, una exposició escrita, una narració oral i una narració escrita. Es va sol·licitar als participants que escrivissin una redacció i es va destacar la importància de reflexionar a fons sobre el tema abans d'escriure. Molt probablement, aquesta estratègia d'utilitzar vídeos sense text ni veu pretenia estimular una resposta escrita que reflectís les diferents situacions de tensió i interacció a l'entorn escolar, per reduir les interpretacions variables que podrien sorgir de l'ús del llenguatge oral o escrit.

El GRERLI-CAT1 inclou textos produïts per 79 parlants bilingües català/castellà de Barcelona (Catalunya), amb el català com a llengua principal. Dins aquest subcorpus es troben textos d'Educació Primària, Educació Secundària, estudiants universitaris, i professors de Llengua Catalana i Literatura. En total, el GRERLI-CAT1 consisteix en 316 textos, 80 de cada nivell, llevat del grup de Secundària, que en té 76. Dins el corpus disponible, hi havia dos tipus de textos: *expository written* i *narrative written*. Tanmateix, per a aquest estudi només s'han escollit textos del primer tipus, seleccionant-ne vint en total, cinc de cada nivell educatiu. Els dos tipus de textos no presentaven grans diferències entre si, i les respostes obtingudes per part de ChatGPT eren similars, la qual cosa feia irrellevant diferenciar entre els dos.

Aquesta recerca ha utilitzat textos dels quatre nivells esmentats, per, així, poder comparar com varia la correcció depenent dels diferents nivells d'educació i edats.

Els textos seleccionats (taula 1) presenten longituds diferents a causa de la limitació de les opcions disponibles en el corpus. Malgrat això,

s'ha intentat garantir una comparació justa i precisa de les característiques lingüístiques i discursives entre els diferents grups.

Taula 1. Corpus de textos utilitzats

Nom	Nivell	Nombre de paraules
cg01fewa-net	PRIMÀRIA	107
cg02fewb-net	PRIMÀRIA	174
cg05fewd-net	PRIMÀRIA	150
cg08mewc-net	PRIMÀRIA	90
cg10fewc-net	PRIMÀRIA	143
cj01fewa-net	ESO	127
cj07mewb-net	ESO	94
cj08fewb-net	ESO	77
cj10mewb-net	ESO	175
cj12mewc-net	ESO	422
cu01fewa-net	UNIVERSITAT	241
cs07mewb-net	UNIVERSITAT	294
cu08mewb-net	UNIVERSITAT	221
cu16fewd-net	UNIVERSITAT	129
cu17mewd-net	UNIVERSITAT	433
ct01mewa-net	PROFESSORAT	260
ct02mewa-net	PROFESSORAT	134
ct08fewd-net	PROFESSORAT	147
ct09fewd-net	PROFESSORAT	191
ct10mewd-net	PROFESSORAT	115

2.3. Procediment

El primer pas va consistir en la selecció els textos que servirien com a corpus per a l'avaluació de l'eficàcia de ChatGPT. Amb l'objectiu de garantir una recopilació de textos que oferís el màxim de retroacció, es van escollir els que presentaven una quantitat significativa d'errors gramaticals, ortogràfics i de cohesió.

Per tant, en aquesta primera etapa del procés, es va dedicar una atenció especial a la tria dels textos, posant l'èmfasi en la seva idoneïtat per a l'objectiu de la recerca. Així, es va procurar una selecció que proporcionés una mostra representativa dels tipus de redaccions que es podrien trobar en un context educatiu, i que també fos una base rica en exemples concrets que permetés observar el comportament de ChatGPT davant diferents tipus d'errors lingüístics. Aquesta selecció va ser fonamental, ja que la presència d'errors en els textos permetria veure realment la capacitat de detecció i correcció de ChatGPT.

El pas següent va consistir a cercar els *prompts* més adequats per a una avaluació precisa, analitzant-ne l'eficàcia i l'adequació. En el cas específic de corregir textos, un *prompt* adequat per a ChatGPT hauria de proporcionar prou context sobre el text que es vol corregir. Aquest *prompt* ha de contenir el text original, indicacions de les àrees on es necessita correcció, quin tipus de correcció es desitja i un cert context per poder-se situar. A més, pot incloure una petició explícita per a la correcció, per exemple, «Corregeix els errors ortogràfics i gramaticals en el següent text». D'aquesta manera, ChatGPT pot utilitzar aquest *prompt* per comprendre el propòsit de la tasca i generar una resposta que ofereixi correccions coherents i rellevants per al text proporcionat.

Es buscava un *prompt* que proporcionés una correcció eficaç i que s'adaptés als criteris d'interès del treball. Entre aquests criteris, es destacaven la correcció gramatical, l'adequació del lèxic, la cohesió textual i la coherència del contingut. Molts dels *prompts* inicialment provats no van donar la resposta esperada, la qual cosa va requerir una continuació de l'experimentació fins a trobar-ne que s'ajustessin de manera òptima als objectius del treball. Progressivament, es va refinar el *prompt*, amb ordres com ara: «Fes un comentari esquemàtic sobre la gramàtica, el lèxic, la cohesió i el contingut d'aquest text», que va començar a proporcionar respostes més coherents. Seguidament, es va especificar que cada text havia estat escrit per una persona d'un nivell acadèmic concret i que es tractava d'un exercici de classe, esperant que ChatGPT corregís com ho faria un professor.

Cal esmentar que al model se li pot demanar que adopti el paper de corrector o professor, però respon que la informació que proporciona pot no ser adequada o completa, ja que no té la capacitat d'analitzar o

comprendre situacions en temps real com ho faria un ésser humà. També és important destacar que les respostes generades per ChatGPT són infinites, és una eina en constant desenvolupament, i les respostes poden variar amb el temps. Les proves d'aquest treball es van fer durant els mesos de març, abril i maig del 2024.

Després de detallar el nivell educatiu dels autors i el context en què s'emmarcaven, es va experimentar amb diverses formulacions de *prompts* noves. Aquesta cerca va culminar en el *prompt* que ha demostrat ser el més eficaç fins ara: «Soc alumne d'X/professor i aquest text és un exercici fet a classe. Fes-ne un comentari de tots els errors en la gramàtica, el lèxic, la cohesió i el contingut:». Aquesta ordre proporciona al model una orientació precisa sobre els aspectes rellevants per aquest estudi.

3. Resultats

3.1. Resultats per nivell

La correcció de ChatGPT ha donat com a resultat respostes similars per a cadascun dels nivells educatius quan els considerem individualment. Tanmateix, les diferències es fan més evidents quan agrupem els nivells en dos grups: Primària i ESO, i universitat i professorat.

3.1.1. Primària i ESO

En analitzar els textos dels alumnes de Primària i ESO, es pot observar que ChatGPT detecta i corregeix una gran quantitat d'errors gramaticals i ortogràfics, especialment els relacionats amb la concordança de nombre i gènere. Per exemple, corregeix «les dos noies» per «les dues noies», però no reconeix tots els casos. A més, s'han detectat problemes recurrents, com és la inconsistència a l'hora de corregir errors de majúscules, minúscules i accentuació.

Pel que fa a les correccions de textos d'ESO, ChatGPT segueix un patró similar. Es concentra principalment en la gramàtica i el lèxic, i en errors de concordança i errors comuns. Per exemple, corregeix errors típics com «hi han» per «hi ha», o també «en el pati» per «al pati». Tot i assenyalar gran part dels errors, les correccions dels textos de Primà-

ria i ESO destaquen negativament per les irregularitats que presenten. Sovint proporciona una solució idèntica a l'error detectat o una solució errònia. Per exemple, en una suposada correcció sobre l'ús dels temps verbals, suggereix canviar «un nen em va ficar el dit a l'ull» per «un nen em va ficar el dit a l'ull», que no modifica la frase original i tampoc conté cap error. En un altre cas, proposa canviar «t'insultin», que és correcte, per «et insultin», una construcció gramatical incorrecta. A més a més, no sempre detecta tots els errors presents en un text; així, en la frase «Ademés si ets un marginat», no suggereix corregir «ademés» per «a més» o una expressió equivalent adequada.

3.1.2. Universitat i professorat

En els textos de nivell universitari i del professorat, ChatGPT sovint se centra a recomanar maneres diferents de dir les coses, i ofereix retroacció estilística. Aquesta pretén millorar l'expressió, la fluïdesa i la claredat del text, i no corregir errors concrets. S'han donat casos on ChatGPT ha suggerit dividir una frase llarga per millorar-ne la comprensió. També ha proposat sinònims per evitar repeticions, i ha suggerit estructures de frases diferents per fer el text més elegant i cohesionat. Igualment, ha fet correccions adequades com l'ús correcte de l'accent a «què» en lloc de «que» quan no és una conjunció, la correcció d'errors de majúscules o la falta d'una coma normativa. Tanmateix, passa per alt moltes altres errades com l'absència de la primera preposició en l'expressió «a poc a poc» o «tant» escrita sense la «t» final de l'adverbi quantitatiu. A més, no corregeix «gimnàsia» per «gimnàstica», «a mida que» per «a mesura que» i «s'enreien» per «se'n reien». No és infal·lible. Proposa substitucions inadequades, com canviar «acudits» per «chistes», una correcció forassenyada.

En els textos redactats pel professorat, ChatGPT fa menys correccions d'errors bàsics, perquè presenten menys faltes. Així que proposa la reestructuració de frases per fer-les més fluides i precises. Són canvis subtils i subjectius, basats en matisos lingüístics o preferències.

Altres correccions inadequades i incorrectes han sigut l'eliminació d'«Hi» en «Hi ha hagut», que passa a «Ha hagut», on «hi» és una partícula pronominal necessària per donar sentit complet a la frase i omplir la funció de subjecte impersonal. La inconsistència es veu quan

ChatGPT, en la correcció d'un altre text, hi afegeix la partícula «hi» quan és necessària.

3.2. Resultats per aspecte

3.2.1. Gramàtica

L'apartat de gramàtica ha estat el més extens en les correccions de ChatGPT, amb ajustos adequats sobretot a Primària i ESO. Tot i això, el resultat sovint ha estat superficial i confús. Entenem per *correcció gramatical* el fet d'ajustar l'ús dels temps verbals, les preposicions i la concordança, entre altres aspectes. No obstant això, ChatGPT ha demostrat tenir dificultats per delimitar clarament aquest apartat. En nombroses ocasions, hi ha inclòs altres comentaris que no tenien relació directa amb la gramàtica estricta, com la puntuació i cohesió.

Malgrat que destaca en la correcció gramatical, cal tenir en compte que algunes de les expressions assenyalades no són estrictament errors gramaticals i que s'han «detectat» errors inventats. Aquests errors inventats són de dues classes: en primer lloc, s'assenyala un suposat error en un verb correcte i es proposa una forma alternativa errònia. Per exemple, proposa canviar «Hem pogut veure» per «Hem vist», inventant-se un error on no n'hi havia. En segon lloc, altera la versió original per poder aportar solucions. Per exemple, «Els pronoms febles s'han utilitzat incorrectament en alguns casos, com en "és situacions", on hauria de ser "són situacions"», ChatGPT s'ha inventat l'error i l'ha categoritzat malament. Al text original ja hi apareixia «són situacions», així que no hi ha cap errada. En tercer lloc, aquesta observació no té res a veure amb els pronoms febles; si l'error fos vertader, es tractaria d'un problema de concordança de nombre del verb, res a veure amb pronoms febles.

En general, la correcció gramatical no ha estat un fracàs absolut; de fet, també s'han produït correccions adequades com «hi han» per a «hi ha». Aquest error ha estat identificat de manera consistent per la màquina, com els errors de concordança, pronoms febles i pronoms relatius.

3.2.2. Lèxic

L'anàlisi del lèxic ha estat notablement inconsistent i insuficient, ja que s'ha centrat a suggerir millores, per exemple, proposant sinònims. Tan-

mateix, molts d'aquests sinònims no han resultat en una solució equivalent al sentit original de l'expressió. A més, ha corregit errors d'accentuació com a lèxic, cosa que reafirma la manca de comprensió de les seccions.

En termes d'adequació de registre, ChatGPT ha identificat correctament que expressions com «una putada» no són adients en un context acadèmic, i ha proposat sinònims més apropiats. Tot i això, no ha detectat altres errors, com l'ús de «choca» amb la «ch» castellana o l'expressió «tenir que». També ha suggerit substituir «ficar-se amb algú» per «no discutir amb ningú» o «barallar-se», alternatives que no recullen el significat original de l'expressió. Una descripció més precisa seria «posar-se amb algú» o «burlar-se'n».

Un altre aspecte controvertit ha estat la proposta de corregir «desequilibrats» per «nerviosos». Socialment, aquests termes no són equivalents, ja que el primer té una connotació negativa. Aquest suggeriment posa en relleu la manca de comprensió de ChatGPT del context social i cultural. A més, ChatGPT ha fet propostes innecessàries, com ara substituir «insolidària» per «manca de solidaritat», «ací» per «aquí» i «deshonesta» per «manca d'honor». En alguns casos, ha modificat paraules del text original, com en l'afirmació: «S'han utilitzat algunes paraules incorrectes o inusuals, com "fet" en lloc de "fetit"». També hi ha exemples on no ha entès el significat concret d'algunes paraules. Per exemple, en suggerir que «maquineta» es refereix a una «màquina petita» en lloc d'un objecte concret, o en proposar canviar el nom del joc «fet i amagar» per «amagar i buscar», fet que demostra una manca de comprensió del context.

3.2.3. Cohesió

Les respostes relacionades amb la cohesió han estat breus, superficials i generalitzades, limitant-se a suggerir una millora en la connexió entre frases per reforçar la claredat del discurs. En totes les correccions s'han repetit dues observacions principals: la manca de transicions clares i la necessitat d'introduir connectors per assegurar la fluïdesa del text.

Tot i que alguns textos presenten indubtablement una estructura fragmentada i una manca de cohesió, aquestes observacions s'han aplicat indiscriminadament, fins i tot en textos ben redactats. Això suggereix

una manca de criteri específic en l'avaluació, ja que les correccions no han identificat errors concrets ni han proporcionat propostes de millora detallades.

3.2.4. Contingut

Els comentaris sobre el contingut han seguit la mateixa línia que els de la cohesió: repetitius i amb explicacions vagues. ChatGPT no ha proporcionat una anàlisi gaire útil, perquè s'ha limitat a descriure el text sense identificar-hi errors. L'exercici no requeria proposar solucions ni recomanacions específiques, però les observacions podrien haver estat més detallades i constructives. Tot i ser adequades, no aporten una correcció real, sinó que assenyalen possibles millores sense aprofundir en com dur-les a terme. Alguns comentaris se centren en el contingut del text, mentre que altres s'enfoquen en la cohesió i l'estructura. En alguns casos, ChatGPT proposa explorar solucions més a fons, tot i que això no era necessari per a la tasca.

Es fa evident que ChatGPT no entén la tasca, ja que insisteix a suggerir el desenvolupament de contingut amb exemples concrets i mesures de millora, quan l'exercici només requeria comentar el que es veia al vídeo i les experiències personals, sense proposar-hi solucions.

4. Discussió

L'anàlisi té com a objectiu valorar les correccions fetes per ChatGPT. Aquesta proporciona una visió general dels aspectes analitzats, incloent-hi les discrepàncies en les correccions, la qualitat dels suggeriments i les dificultats trobades en la comprensió de la tasca encarregada.

4.1. Impacte del domini de la llengua i tasca encarregada

ChatGPT aplica un enfocament diferenciat en la correcció de textos segons el nivell educatiu. En textos d'estudiants d'ESO i Primària, se centra en errors gramaticals i lèxics, que són més comuns. En canvi, en textos de nivell universitari o del professorat, tendeix a oferir recomanacions estilístiques, ja que els errors lingüístics són menys freqüents.

Aquest enfocament, tot i ser adequat a les necessitats de cada nivell, no s'ajusta al que se li havia sol·licitat: limitar-se a corregir errors sense oferir suggeriments addicionals. Quan no hi ha errors, l'eina hauria de deixar el text com està. Proporcionar recomanacions estilístiques, inventar-se errors o ampliar temes complica el procés de revisió i mostra que l'eina no comprèn la tasca assignada. Per exemple, ha fet comentaris com: «No s'aborda completament el tema de com evitar aquests problemes més enllà de l'expressió d'un desig abstracte». ChatGPT considera que seria útil explorar més a fons altres possibles solucions, però aquest enfocament no és necessari per a la correcció. La tasca encomanada consistia únicament a comentar i corregir els errors presents en el text original, sense afegir informació ni suggerir ampliacions temàtiques. Aquest comportament de ChatGPT reflecteix que està programat per donar una resposta completa i detallada en qualsevol context, fins i tot quan no es requereix.

4.2. No detecció

Els errors de concordança de nombre i gènere han estat corregits amb gran regularitat, mentre que molts altres errors no. En alguns casos no s'han detectat ni corregit errors com «tenir que» ni s'ha assenyalat la manca d'accents a «teníem» i «òbviament». També hi ha hagut omissions en la detecció d'errors gramaticals, com l'expressió «a mida que», que hauria de ser «a mesura que», ni tampoc la de «choca» amb la «ch» castellana. Aquesta omissió d'errors demostra una limitació significativa en la capacitat de ChatGPT per complir plenament amb les necessitats de correcció, cosa que repercuteix directament en la qualitat del resultat. Això posa en dubte l'eficàcia de l'eina per a una correcció precisa, especialment quan es tracta de textos d'estudiants. Per tant, és crucial que els usuaris de ChatGPT assumeixin aquestes limitacions i complementin la revisió automàtica amb una revisió humana, en especial en contextos educatius, on la qualitat i la precisió són fonamentals.

4.3. Dificultats amb la combinació de pronoms

S'han detectat problemes en la correcció de ChatGPT pel que fa a la combinació de pronoms. En alguns casos, l'eina no ha identificat errors en l'ús i col·locació dels pronoms, mentre que en altres, ha fet correccions inadequades que no milloraven el text i que, de fet, hi introduïen nous

errors. Per exemple, ChatGPT no ha detectat l'error en la frase «me'n enterava». Aquí, el pronom «me'n» està mal col·locat i el verb «enterava» no és l'adequat. La correcció correcta seria «me n'assabentava», que mostra el verb correcte i la col·locació adequada del pronom. De manera similar, la frase «s'enreien» hauria de ser corregida a «se'n reien», però ChatGPT no ha assenyalat aquest error. En un altre exemple, ChatGPT ha proposat una correcció errònia a una frase que ja era correcta. La frase original: «Em vaig adonar que em manava» era clara i ben formada, però ChatGPT va suggerir la correcció: «Em vaig adonar que em estava manant». Aquesta proposta introdueix un error, ja que ignora la necessitat d'apostrofar el pronom «em», com en «m'estava manant».

4.4. Comentaris positius

S'ha apreciat una manca de comentaris positius en les respostes de ChatGPT en català. Tot i que l'usuari hagi referit que és un alumne de Primària, ChatGPT no ha fet comentaris encoratjadors ni d'ànims. Això és especialment rellevant en un context educatiu, on el reforç positiu pot ser molt beneficiós per a la motivació dels estudiants. Els pocs comentaris positius que s'han obtingut eren d'un estil força neutre, com ara: «És una paraula adient per descriure el comportament violent dels nens» o observacions com ara: «La idea central del text és clara» i: «Malgrat la claredat general del missatge, la idea podria ser més precisa». Aquests comentaris són funcionals i descriptius, però no transmeten uns suggeriments positius o motivadors.

Es poden demanar aquests tipus de comentaris encoratjadors explicitant-ho al *prompt*, però en aquest cas no es va fer. En anglès és força diferent, ja que, sense demanar-ho, ChatGPT sovint fa comentaris positius. Per exemple, en corregir un text en anglès, es troben comentaris com: «You did a great job with your text. I noticed a few mistakes, so let's fix them together».

Aquests comentaris mostren una clara diferència en l'enfocament entre les respostes en anglès i en català. Mentre que en anglès ChatGPT ofereix un suport més positiu i motivador, en català falta aquest component. Aquesta diferència posa en relleu la necessitat de millorar l'aproximació de ChatGPT en català per tal de proporcionar una retroacció més adequada.

4.5. Corpus i limitacions

Gran part dels problemes identificats poder ser deguts a la manca de corpus en català, cosa que limita el desenvolupament de ChatGPT. Una de les causes per les quals el model presenta dificultats per corregir textos de manera precisa pot ser el no tenir prou dades en aquesta llengua per aprendre-la adequadament. A més, ChatGPT pot no estar actualitzat amb les últimes reformes ortogràfiques o canvis normatius, ja que es basa en un conjunt limitat de textos i pot no tenir suficient informació. També és possible que no copsi matisos culturals o contextos específics importants per a una correcció precisa, com ara referències culturals, jocs de paraules o expressions idiomàtiques que no són directament traduïbles. Això també afecta la comprensió de les diferents variants dialectals, els errors tipogràfics complexos i les subtileses semàntiques.

Altres problemes amb la IA es deriven de la manca de sentit comú i de coneixement del món per part dels programes i les màquines. Això pot provocar que corregeixin erròniament coses en què no cal fer-ho, perquè no entenen el context, i també poden passar per alt errors que sí que requereixen correcció. Aquesta manca de coneixement del món es fa evident en algunes correccions; per exemple, en un text universitari, es menciona una «xuleta» amagada «dins una maquineta». ChatGPT suggereix: «L'ús de "maquineta" pot no ser del tot clar. Podries considerar "objecte" o "dispositiu"». Això revela que no reconeix que es refereix a una maquineta de fer punta al llapis i assumeix, equivocadament, que es tracta d'un petit dispositiu. També es manifesta aquesta manca de coneixement quan suggereix «desequilibrats» com a sinònim de «nerviosos», sense adonar-se de la connotació negativa que aquesta paraula té en la nostra societat. De la mateixa manera, sorprèn que no comenti res de l'ús de la paraula «tonta», que es podria canviar per un sinònim menys malsonant.

També s'han donat casos en els quals la màquina ha respost en castellà, i altres ocasions en les quals ha fet correccions en apartats on no tocava. Aquestes incidències mostren encara més la necessitat de millorar la comprensió contextual i lingüística de ChatGPT en català, ja que afecten directament la seva utilitat i fiabilitat com a corrector en català. Es pot veure una clara diferència respecte a quan se'n fa ús en anglès, perquè, en tenir un bagatge molt més ampli, les seves respostes

són molt més encertades. Això posa de manifest la importància d'ampliar el corpus de dades en català, si es vol arribar a una correcció més precisa i adequada a les necessitats dels parlants de la llengua.

5. Implicacions pedagògiques

La implementació de tecnologies avançades com ChatGPT en l'educació presenta oportunitats i desafiaments significatius. La integració de la IA pot transformar l'aprenentatge i l'ensenyament, però requereix abordar qüestions ètiques i pedagògiques, com el respecte a la privacitat i la prevenció de biaixos. La formació adequada del professorat i la personalització de l'eina són imprescindibles per assegurar-ne un ús efectiu. Abans de generalitzar-ne l'ús, seria recomanable implementar projectes pilot per avaluar-ne l'efectivitat en diversos contextos i obtenir dades per millorar-ne la funcionalitat.

A més, és crucial establir mecanismes d'avaluació contínua que analitzin l'impacte de ChatGPT en l'aprenentatge i permetin ajustar-ne la configuració a necessitats específiques. Futures investigacions podrien explorar l'ús de *prompts* específics, l'aplicació de rúbriques i la comparació amb altres eines de correcció, així com el seu efecte en el desenvolupament d'habilitats lingüístiques. Aquest enfocament contribuiria a optimitzar la integració de la IA en l'entorn educatiu en benefici de professors i estudiants.

6. Conclusions

L'ús de ChatGPT per a la correcció de textos en català pot oferir avantatges significatius, especialment pel que fa a la velocitat i precisió en la detecció d'errors bàsics. No obstant això, la seva efectivitat es veu limitada per la manca d'un corpus prou ampli i divers en aquesta llengua, la qual cosa pot afectar la capacitat de correcció i la qualitat dels suggeriments.

S'han pogut observar diverses limitacions i obstacles quan s'usa ChatGPT com a corrector de textos en català. Actualment, no és prou fiable ni consistent per ser utilitzat com a única eina de correcció. Sovint passa

per alt errors significatius i pot oferir correccions inadequades. A més, la correcció de textos implica la detecció d'errors gramaticals o ortogràfics, al costat d'una comprensió profunda del contingut, un àmbit en el qual ChatGPT encara presenta limitacions. Aquesta manca de precisió es deu, en part, a la insuficiència d'un corpus ampli i divers en català, que limita la seva capacitat per generar correccions contextualment adequades. Això posa de manifest la necessitat de millorar el corpus disponible per tal de millorar la qualitat de les correccions proporcionades per ChatGPT.

A més, és important tenir en compte que ChatGPT tendeix a prioritzar la complaença de l'usuari, fet que pot comprometre la qualitat de les seves correccions. Aquesta característica pot portar a suggeriments que, encara que semblin correctes, no sempre siguin els més adequats. Per tant, és crucial que els usuaris apliquin el seu sentit comú i els seus coneixements lingüístics per avaluar les correccions proposades per ChatGPT, especialment per a la detecció i correcció d'errors més complexos i contextuals o en casos que requereixen tenir coneixements del món.

Tot i aquestes limitacions, ChatGPT pot ser una eina complementària molt útil per a la revisió de textos, especialment com a primera passada per identificar errors bàsics. Tanmateix, la seva aportació no és equiparable a la revisió humana i no pot substituir-la completament. Per garantir una correcció precisa, adequada contextualment i de qualitat, és imprescindible combinar l'ús de ChatGPT amb la supervisió i experiència humanes, atès que aquesta col·laboració pot millorar notablement l'eficiència i l'exhaustivitat de la revisió de textos en català.

Amb vista a propers estudis, seria interessant comparar les correccions proporcionades per ChatGPT amb les facilitades per altres eines, com ara el corrector ortogràfic i gramatical de Softcatalà o el redactor assistit arText. D'aquesta manera es podria valorar quina eina és més efectiva i pot ser-li més útil a l'usuari.

Agraïments

Volem agrair la tutorització i guia de la Dra. M. Àngels Massip per a la realització d'aquest estudi.

3. RETROACCIÓ GENERADA PER IA EN PRODUCCIÓ ESCRITA EN ANGLÈS COM A LLENGUA ESTRANGERA D'APRENENTS AMB ESPANYOL COM A L1 A TRAVÉS DELS NIVELLS DEL MCER: UNA COMPARACIÓ ENTRE MYESSAI I CHATGPT

› **Júlia Malet Raich**

1. Introducció

En el camp de l'aprenentatge de llengües, la IA es fa servir cada vegada més, tot i que el seu paper i la seva eficàcia continuen sent incerts. La segona experiència d'aquest volum explora la retroacció generada per la IA per comprendre millor com aquesta pot contribuir a l'aprenentatge de llengües i oferir informació sobre els seus possibles avantatges i limitacions.

Aquest estudi analitza la retroacció de dues eines d'IA, ChatGPT i MyEssai, sobre 18 textos produïts per aprenents d'anglès com a llengua estrangera que es preparaven per a un examen oficial. Cada estudiant va redactar un text transaccional (és a dir, un correu electrònic) i un text creatiu (una història breu). Els participants es van classificar en tres nivells de competència segons el Marc Europeu Comú de Referència (MCER): B1, B2 i C1, amb tres aprenents representant cada nivell. L'objectiu principal de l'estudi és avaluar i analitzar el rendiment de MyEssai i ChatGPT en la seva avaluació dels textos dels aprenents.

Per aconseguir-ho, l'estudi aborda diverses preguntes clau:

1. Quines són les característiques de la retroacció proporcionada per aquestes dues eines d'IA?
2. Són aquestes eines d'IA capaces d'ajustar la retroacció al nivell de competència de l'aprenent?
3. Les eines d'IA diferencien en funció de la tipologia del text?

2. Metodologia

2.1. Eines d'IA: MyEssai i ChatGPT

La primera eina d'IA utilitzada en aquest estudi és MyEssai, una aplicació dissenyada específicament per proporcionar comentaris sobre els textos escrits pels estudiants. Usa IA per revisar el contingut del text i proporcionar una retroacció detallada i extensa.

Aquesta eina ha estat escollida perquè, a diferència d'altres opcions generals, com ara ChatGPT, no requereix un *prompt* molt específic per generar la retroalimentació, ja que està programada expressament per donar retroacció escrita d'una manera determinada. Tot i això, per oferir una retroacció acurada, és necessari introduir tant la tipologia textual com l'enunciat al qual respon el text. MyEssai ofereix diferents opcions perquè la IA es concentri en aspectes concrets, però aquest estudi ha treballat exclusivament amb l'enfocament general.

En aquest estudi s'ha treballat amb dues tipologies de text: *creative writing* (text narratiu), que encaixava perfectament amb els textos narratius del corpus, i *cover letter* (carta de presentació), que es va triar com a opció més propera per als correus electrònics, atès que són textos transaccionals. A més, es van facilitar a l'eina els enunciats originals als quals els estudiants havien respost, a fi que la retroacció fos al més precisa possible.

D'aquesta manera, MyEssai resultava la millor opció per a aquest estudi, ja que s'ajustava a les necessitats del corpus i donava una retroalimentació més acurada que altres eines d'ús més general.

La segona eina és ChatGPT, versió ChatGPT 3.5. A diferència de MyEssai, ChatGPT no ha estat dissenyat específicament per proporcionar retroacció. ChatGPT funciona com un bot de conversa i és un LLM, que significa que ha estat entrenat amb un corpus lingüístic extens. Tot i no ser exclusiu per a l'aprenentatge de llengües, és capaç d'analitzar els textos dels estudiants amb certa profunditat i, per tant, pot proporcionar retroacció sobre una redacció, si se li demana. Tanmateix, els usuaris han de ser específics a l'hora de donar instruccions i interaccionar amb el bot de conversa, atès que això afectarà la qualitat de la retroacció.

2.2. Dades del corpus

Aquest estudi ha treballat amb el corpus FineDesc, compilat per al projecte de recerca FineDesc. El corpus, que conté un total de 800.000 *tokens*,[3] consisteix en textos escrits en anglès produïts per aprenents d'anglès amb espanyol com a primera llengua. Concretament, l'estudi ha utilitzat 18 redaccions escrites per aquests aprenents, que es preparaven per a l'examen d'acreditació en anglès CertAcles. Els textos escrits estan dividits en tres nivells de competència segons el MCER: B1, B2 i C1. Per tant, hi ha sis textos per a cada nivell del MCER. A més, els textos també estan classificats segons la seva tipologia, de manera que hi ha tres textos transaccionals (correus electrònics) i tres textos creatius (textos narratius) per a cada nivell. Així, els escrits estan organitzats en grups de tres, i cada grup respon a un enunciat concret.

2.3. Procediment

2.3.1. Elaboració del prompt

MyEssai no necessita un *prompt* molt específic, perquè l'aplicació ja ha estat creada per proporcionar retroacció. Així i tot, és necessari introduir la tipologia del text, així com l'enunciat al qual el text respon. Com s'ha esmentat anteriorment, per aquest estudi s'han seleccionat les opcions de *creative writing* i *cover letter*. L'opció de *creative writing* encaixava perfectament amb els textos narratius del corpus, però no hi havia cap opció adequada per als correus electrònics, per la qual cosa es va seleccionar l'opció de *cover letter*, ja que era la més similar, en tractar-se també d'un text transaccional. A més, es van proporcionar a la IA els enunciats als quals els estudiants havien respost perquè la retroacció pogués ser més precisa. MyEssai no permet donar més informació (com el nivell MCER) ni demanar retroacció més específica.

En contrast, ChatGPT necessita un *prompt* molt específic i detallat perquè la IA analitzi els textos de la manera desitjada. En primer lloc, es va indicar a ChatGPT que havia de revisar una redacció escrita per un aprenent d'anglès amb el castellà com a L1 i que es preparava per a un examen oficial. A més, es va incloure en el *prompt* el nivell MCER

3. Unitat mínima (generalment a nivell de paraula) que s'obté al segmentar un text automàticament.

de cada estudiant i el tipus de redacció. Es va demanar a ChatGPT que proporcionés comentaris específics sobre les àrees següents: gramàtica, vocabulari, contingut, organització i assoliment de la tasca. Aquestes àrees es van especificar perquè, en demanar simplement un retorn general, les respostes eren inconsistents. Seguidament, es van indicar les instruccions que havien seguit els estudiants per a escriure les redaccions. Finalment, es va proporcionar a la IA la producció escrita que havia d'analitzar. En concret, el *prompt* utilitzat va ser el següent:

> *I am an EFL learner currently preparing for a high-stakes exam, and my level is a B1 according to the CEFR. Could you give me feedback on grammar, vocabulary, content, organisation and task achievement for my creative writing (100-120 words)? Please, give me feedback on my mistakes, you don't need to provide a revised version. The essay had to begin with the sentence: «He had never called the police before...» This is what I have written: [Text]*

2.3.2. Anàlisi de la retroacció

Les dues mostres de retroacció generada per IA es van comparar per identificar diferències i semblances entre el rendiment de MyEssai i ChatGPT. En primer lloc, l'estudi explora com cadascuna de les eines aborda les àrees objecte de revisió (gramàtica, vocabulari, contingut, organització i assoliment de la tasca). L'anàlisi també proporciona informació sobre els tipus de retroacció que cada eina ofereix amb més freqüència. Amb aquesta finalitat, l'estudi distingeix entre comentaris positius i correctius, incloent-hi els subtipus de retroacció correctiva: correccions explícites, reformulacions[4] i informació metalingüística.[5] A més, es comparen les retroaccions referides als textos narratius amb les dels correus electrònics. Finalment, s'avalua l'estructura general de la retroacció.

4. Tornar a formular una cosa d'una altra manera.

5. Ús del llenguatge per parlar del mateix llenguatge, analitzar-lo o explicar-ne el funcionament.

3. Resultats i discussió

En la subsecció següent, s'analitzarà la retroacció proporcionada per ChatGPT i MyEssai en relació amb tres àrees principals: llengua, organització i contingut, i assoliment de la tasca. A continuació, hi haurà una altra secció que revisarà les fortaleses i debilitats generals de MyEssai i ChatGPT, en la qual les eines seran analitzades des d'una perspectiva més general.

3.1. La retroacció de ChatGPT i MyEssai

3.1.1. Retroacció sobre aspectes lingüístics

Aquesta subsecció explora la retroacció específica relacionada amb els aspectes de llenguatge formal, centrant-se en els comentaris de les eines sobre la gramàtica i el vocabulari.

3.1.1.1. Retroacció sobre qüestions gramaticals

En el context de la retroacció correctiva, es poden observar diferències i semblances clares en la retroacció per a la gramàtica de cadascuna de les eines. Una diferència notable és que la retroacció de MyEssai és més indirecta que la de ChatGPT. MyEssai utilitza més reformulacions, amb un total de 9 errors, que informació metalingüística (6 errors) i correccions explícites (6). Per contra, ChatGPT tendeix a confiar més en les correccions explícites, oferint un total de 25 errors, amb menys reformulacions (11) i informació metalingüística (6).

Per tant, cap de les eines proporciona gaire informació metalingüística i, quan ho fan, és, generalment, breu i no elaborada. Per exemple, en la frase: «He was in his flat that night, watching the TV, as he always do», ChatGPT aconsella tenir en compte la concordança entre verb i subjecte: «Use 'does' instead of 'do' to match the singular subject 'he'». Per a aquest mateix estudiant, MyEssai destaca el canvi de temps verbal, passant de passat a present, i aconsella mantenir el passat: «To maintain clarity, ensure that you consistently use the same tense throughout your story».

ChatGPT acostuma a efectuar més correccions gramaticals que MyEssai. ChatGPT corregeix una mitjana de 4,7 errors gramaticals per

aprenent, mentre que MyEssai només en corregeix una mitjana de 2,3. Hi ha moltes ocasions en què MyEssai no corregeix errors gramaticals i fins i tot pot recomanar l'ús d'altres eines com ara Grammarly per fer-ho. Tanmateix, ChatGPT també passa per alt errors gramaticals, com en el cas vist en el paràgraf anterior, quan no corregeix el canvi de temps verbal. A més, hi ha ocasions que ChatGPT ressalta errors sense proporcionar correccions. També pot no especificar errors, com és el cas d'un text narratiu d'un dels aprenents de nivell C1: «sentences are overly complex or convoluted, which affects clarity», però no menciona quines frases cal revisar pel que fa a la complexitat.

MyEssai i ChatGPT sovint corregeixen els mateixos errors i de manera similar. Per exemple, en un text narratiu de B1, ambdues corregeixen la frase: «There isn't nobody» a: «There wasn't anybody». Així i tot, també hi ha moltes ocasions en les quals aborden errors diferents. En un mateix text narratiu de B2, ChatGPT detecta diverses inconsistències de temps verbal i suggereix canviar «make» per «made», «forget» per «forgot» i «see» per «saw». A més, rectifica «that feelings» i proposa «those feelings». En canvi, MyEssai només esmena els dos primers errors («make» i «forget») i no n'esmenta la resta. Tot i això, identifica un altre error que ChatGPT no assenyala: substitueix el plural incorrecte de «woman» per «women». Per tant, tots dos corregeixen inconsistències gramaticals, però no s'enfoquen en les mateixes. Finalment, també hi ha molts errors que no són esmentats per cap de les dues. Per exemple, en un text creatiu de B2, cap eina corregeix l'errada en «a cheap tickets».

Més enllà de la retroacció correctiva, també hi ha diferències en la retroacció positiva i les estratègies motivacionals de MyEssai i ChatGPT. MyEssai proporciona constantment raons per a les seves correccions: «Careful attention to spelling and grammar will go a long way in ensuring your readers are immersed in the story and not distracted by these hiccups». No obstant això, MyEssai no ofereix retroacció positiva específicament per a la gramàtica. En canvi, ChatGPT normalment evita explicar el raonament darrere de les seves correccions, tot i que ocasionalment proporciona suggeriments constructius, com ara: «Simplifying your sentences could improve readability». A més, ChatGPT sí que ofereix retroacció positiva en algunes ocasions, amb comentaris

com: «Overall, your grammar is quite good. You demonstrate a good command of sentence structure and verb conjugation».

Pel que fa als diferents nivells de competència, ambdues eines mostren poca variació en el tipus de retroacció correctiva. MyEssai dedica menys línies a la gramàtica amb els aprenents més avançats, possiblement a causa d'un nombre menor d'errors en nivells més alts. De manera similar, ChatGPT també fa menys correccions per als aprenents avançats. En termes de retroacció metalingüística, ChatGPT no proporciona informació metalingüística als aprenents de nivell C1. En canvi, l'enfocament de MyEssai es manté constant a tots els nivells, ja que les explicacions no s'adapten al nivell de competència dels aprenents. A més, sol oferir reformulacions, però la complexitat del llenguatge emprat pot ser massa elevada per als aprenents de nivell inferior.

En relació amb la retroacció positiva i la motivació, ChatGPT tendeix a proporcionar més comentaris positius en gramàtica per als aprenents més avançats. De la mateixa manera, tot i que ChatGPT no aprofundeix gaire en els aspectes motivacionals, per a aprenents dels nivells B2 i C1 fa comentaris encoratjadors com ara: «Overall, your grammar is good. You demonstrate a clear understanding of sentence structure and verb conjugation», mentre que, per als aprenents de nivell inferior, sovint es limita a proporcionar una llista de correccions. MyEssai no proporciona correcció positiva ni per als aprenents de nivell inferior ni per als més avançats. Les estratègies motivacionals utilitzades per MyEssai no s'adapten segons el nivell. Per il·lustrar-ho, comparem el comentari: «Language and grammar are the threads that stitch your story's fabric», adreçada a un aprenent de nivell B2, amb aquesta altra dirigida a un aprenent de nivell C1: «Misspellings and grammatical quirks can momentarily distract readers, so a quick proofreading session is always beneficial».

Finalment, MyEssai mostra una variació més gran en les correccions segons la tipologia textual en comparació amb ChatGPT. En el context dels correus electrònics, MyEssai pot no corregir explícitament nombrosos errors gramaticals, sinó centrar-se a mantenir la formalitat, principalment mitjançant reformulacions. A més, les raons de MyEssai per corregir la gramàtica varien segons el tipus de text: en els correus electrònics, l'objectiu és millorar la claredat i aconseguir un to formal,

mentre que, en els textos creatius, l'objectiu és mantenir la immersió del lector. En canvi, ChatGPT no presenta diferències en les correccions gramaticals entre els correus electrònics i els textos creatius. A més, les estratègies de correcció positiva i motivació emprades per a cada tipus de text són molt similars. Per exemple, comparem aquest consell que ChatGPT proporciona per a un correu electrònic: «This sentence is quite long and could be divided into two or more sentences for clarity», amb aquest altre adreçat a un text creatiu: «Pay attention to subject-verb agreement and word order to ensure clarity and accuracy in your sentences».

3.1.1.2. Retroacció sobre qüestions lèxiques

En la retroacció correctiva per al lèxic, MyEssai i ChatGPT fan servir estratègies diferents. Tot i que MyEssai utilitza correccions explícites per abordar problemes d'ortografia, també proporciona reformulacions ampliades, com ara la reformulació de la frase en el text narratiu d'un aprenent de nivell B2. En canvi, ChatGPT empra directament correccions explícites, com ara substituir «requeriments» per «requested amenities». A més, ChatGPT destaca la importància de la varietat lèxica i recomana als aprenents diversificar el seu vocabulari per millorar la implicació del lector, per exemple, suggerint sinònims per evitar repeticions. La informació metalingüística sobre vocabulari és molt limitada en la retroacció d'ambdues eines.

Una característica compartida per ChatGPT i MyEssai és que fomenten l'ús d'un llenguatge descriptiu. Això es pot observar en la retroacció al text narratiu d'un aprenent de nivell B2, a qui ambdues eines recomanen incloure més detalls en la seva narració. ChatGPT explica: «using more descriptive language to paint a vivid picture of the event and your experiences». De la mateixa manera, MyEssai aconsella: «Include sensory details such as the smell of burning rubber, the feel of the vibrations from the roaring engines, or the sight of dusk settling over the rally».

Encara que la retroacció positiva sobre vocabulari no és molt freqüent, ambdues plataformes ofereixen ocasionalment comentaris encoratjadors. ChatGPT felicita un aprenent: «Your choice of vocabulary is generally appropriate and helps to convey your experience vividly». MyEssai

destaca un punt fort en l'escriptura d'un aprenent: «You've chosen a conversational tone, which makes the story approachable, and that's a strength to build on». No obstant això, aquesta mena de comentaris per a animar són escassos en la retroacció de totes dues eines. En canvi, pel que fa a la motivació, les eines solen explicar per què és important seguir les seves recomanacions. En una ocasió, MyEssai argumenta: «Descriptive language is a powerful tool to transport readers into the scene». De manera similar, ChatGPT afirma: «Use a variety of vocabulary to make your writing more engaging».

Hi ha algunes diferències en l'enfocament depenent del nivell de competència de l'aprenent entre MyEssai i ChatGPT. MyEssai no adapta la retroacció al nivell de l'aprenent; de manera similar a la seva retroacció sobre gramàtica, les reformulacions que proporciona sovint inclouen vocabulari que pot ser complicat per als aprenents de nivell inferior. En canvi, ChatGPT adapta la seva retroacció en cert grau. Per als aprenents de nivell inferior, ChatGPT se centra a corregir errors d'ortografia i selecció de paraules, de manera que les correccions que proporciona són més bàsiques, com ara: «It would be more appropriate to refer to it as a 'post' or 'listing' rather than a 'publication'». En canvi, per als aprenents més avançats, ofereix suggeriments orientats a millorar la varietat lèxica: «Your vocabulary is appropriate for the context, but consider varying your word choices to avoid repetition. For example, instead of using 'materials' multiple times, you could use synonyms like 'resources' or 'content'».

ChatGPT i MyEssai semblen adaptar les observacions sobre vocabulari d'acord amb el tipus de text. Totes dues eines posen l'èmfasi en el llenguatge descriptiu en els textos narratius per millorar la immersió en la història, mentre que prioritzen la formalitat i el professionalisme en els correus electrònics. Per exemple, en un correu electrònic, ChatGPT suggereix canviar: «I could consider this a scam» per: «I believe this situation may constitute fraudulent misrepresentation», aconseguint un to molt més formal. Però, en els textos narratius, aconsella utilitzar un llenguatge més descriptiu per crear una imatge més clara per al lector. De manera similar, MyEssai recomana moderar el llenguatge contundent en els correus electrònics i reformular certs termes per transmetre més diplomàcia. Pels textos narratius, MyEssai fa recomanacions

més enfocades en les sensacions transmeses pel llenguatge: «More vivid and sensorial language could enhance the atmosphere».

3.1.2. Retroacció sobre l'organització del text

Tant MyEssai com ChatGPT fan suggeriments similars per organitzar els textos, destacant la importància de la claredat i la coherència. Ambdues eines sovint recomanen dividir els textos en més paràgrafs per millorar-ne la llegibilitat. Per exemple, MyEssai suggereix escriure paràgrafs més curts: «White space in writing allows the reader to absorb each thought before moving on to the next, which can be quite effective when dealing with abstract concepts». De manera similar, ChatGPT afirma: «The writing lacks clear paragraph breaks, which can make it challenging to read. Organize your ideas into paragraphs to improve clarity and coherence». De fet, ambdues eines aconsellen situar cada idea en un paràgraf separat. ChatGPT suggereix: «Each paragraph could focus on a specific aspect of how music breaks down boundaries or connects people, allowing for a smoother flow of ideas...». MyEssai recomana a l'aprenent: «Avoid jumping from one topic to another without a smooth transition. To improve flow, group related questions together and ensure each paragraph has a clear focal point».

A més, ambdues eines aborden la complexitat de les oracions. MyEssai assenyala que, tot i que presentar idees complexes és positiu, les oracions denses poden dificultar la comprensió del significat. Per això, recomana dividir-les en estructures més simples per guanyar claredat. ChatGPT coincideix en aquest punt i aconsella dividir les oracions llargues en altres més curtes per millorar la llegibilitat. Totes dues eines també suggereixen a diversos aprenents començar els correus electrònics amb una frase introductòria per oferir context.

Les retroaccions proporcionades per MyEssai i ChatGPT presenten tant coincidències com diferències en la retroacció per a un mateix aprenent. Per exemple, en un correu electrònic de nivell B1, les eines coincideixen en la importància d'una estructura clara i ofereixen retroacció positiva abans de fer suggeriments de millora. Totes dues posen èmfasi en la necessitat de reorganitzar els paràgrafs agrupant preguntes similars per facilitar la comprensió, encara que MyEssai és més detallat en la seva explicació i proporciona un exemple concret de

com estructurar el text. Malgrat això, en altres ocasions es poden observar discrepàncies, com és el cas del correu electrònic de l'aprenent 3 del mateix nivell. Mentre que ChatGPT suggereix únicament dividir les oracions més llargues per millorar la llegibilitat, MyEssai proposa una reorganització més profunda, incloent-hi la necessitat de presentar-se adequadament i explicar la motivació abans de formular preguntes. Per tant, es contradiuen, ja que ChatGPT considera que l'organització dels paràgrafs és adequada, mentre que MyEssai suggereix canviar-los radicalment.

A més, amb alguns textos, MyEssai no ofereix cap mena de retroacció sobre l'organització; per exemple, en el cas del text narratiu de l'aprenent 1 de C1. Es tracta d'un aprenent avançat, per la qual cosa es podria pensar que MyEssai no ofereix consells d'organització perquè considera que en aquest text no són necessaris. No obstant això, aquesta suposició contrasta amb la retroacció de ChatGPT per al mateix text de l'aprenent: «The essay lacks clear organisation and transitions between ideas. Consider structuring it into paragraphs to improve readability and coherence». Una altra possible explicació a la manca de retroacció sobre l'organització per part de MyEssai és que la IA decideix que l'aprenent ha de prioritzar altres àrees. En tot cas, no és fàcil discernir la raó per la qual MyEssai no fa cap comentari sobre l'organització.

Tant MyEssai com ChatGPT ofereixen motivació i retroacció positiva, però les aproximacions i la freqüència varien. Com passa amb altres àrees, MyEssai se centra a explicar les raons darrere dels suggeriments d'organització: «Structuring your complaints like this gives the letter recipient a quick, clear understanding of your concerns». ChatGPT no s'estén tant en les raons, tot i que pot esmentar-les breument en algunes ocasions: «Consider structuring it into paragraphs to improve readability and coherence». Respecte a la retroacció positiva, MyEssai tendeix a oferir-ne menys, però, quan ho fa, inclou comentaris detallats i específics, com ara: «The introduction is effective in establishing the universal resonance of music. The phrase 'It has the power to give us peace and the strength to make us jump' is striking and sets the duality of music's impact beautifully». ChatGPT, però, gairebé sempre proporciona retroacció positiva: «The organisation of ideas is clear, with a beginning, middle, and end».

No hi ha diferències significatives en els nivells de competència entre la retroacció proporcionada per ambdues eines. Per als aprenents de B1 i B2, MyEssai es preocupa per oferir un ordre més natural i lògic, aconsellant als aprenents que diferenciïn clarament entre introducció, cos i conclusió. També recomana habitualment l'ús d'oracions més curtes i paràgrafs més curts. Amb la majoria dels aprenents C1 no s'estén gaire en aspectes d'organització, però els escassos comentaris que fa segueixen la mateixa línia que els adreçats a nivells inferiors: «In rewriting, you might want to work with shorter paragraphs to create a more impactful pace». De manera similar, pel que fa a la retroacció correctiva, ChatGPT tendeix a oferir la recomanació d'organitzar el text en paràgrafs més curts i agrupar idees similars. Tanmateix, només ofereix retroacció positiva als aprenents de B2 i C1, amb comentaris com ara: «The email is reasonably well-organized, with each paragraph focusing on a different aspect of the complaint».

MyEssai adapta la retroacció correctiva al tipus de text. En els correus electrònics, la retroacció té l'objectiu de fer-los més comprensibles, suggerint maneres lògiques i efectives d'organitzar el text: «Avoid jumping from one topic to another without a smooth transition. To improve flow, group related questions together and ensure each paragraph has a clear focal point». La retroacció correctiva que MyEssai ofereix per als textos narratius no només té l'objectiu de fer la història més clara, sinó també més immersiva i fluida. Per exemple, ofereix aquest consell: «Regarding the structure, an engaging short story often has a beginning that hooks the reader, a middle that builds upon that interest, and an end that leaves them thinking».

En canvi, la retroacció correctiva que ChatGPT proporciona per a cada tipus de text no presenta contrastos notables. Per exemple, comparem aquests consells gairebé idèntics: «Consider breaking the text into smaller paragraphs for better readability» i: «Consider breaking the text into smaller paragraphs to improve readability and flow». Malgrat això, la retroacció positiva és més diferenciada, ja que reconeix els objectius propis de la tipologia textual. D'una banda, per a un correu afirma: «The email is well-organized, with each paragraph focusing on a different aspect of the problem». Així, destaca el fet que el text és clar i comprensible. D'altra banda, en els textos narratius dona prioritat a la fluïdesa: «The story flows logically from your anticipation before the

performance to the climax of experiencing the music and emotions during the opera». Aleshores, també es pot dir que ChatGPT ajusta fins a un punt la retroacció organitzativa segons el tipus de text.

3.1.3. Retroacció sobre el contingut i l'assoliment de la tasca

MyEssai sempre proporciona suggeriments extensos relacionats amb el contingut del text. Per exemple, per al text narratiu d'un aprenent de nivell B2, MyEssai fa múltiples recomanacions d'aquest tipus, com explorar les emocions de l'escriptor, proporcionar més detalls sobre els personatges i expandir la conclusió, dedicant un paràgraf complet a cadascuna. En canvi, ChatGPT tendeix a oferir menys suggeriments o a només incloure retroacció positiva. Per il·lustrar-ho, per a l'aprenent esmentat anteriorment, ChatGPT fa dos comentaris positius: «You effectively convey the excitement and anticipation of attending the opera for the first time» i: «You describe the impact of the music and performances on your emotions well, which enhances the storytelling». Aquesta comparació indica que MyEssai proporciona consells detallats i específics sobre el context, mentre que la retroacció de ChatGPT no entra en recomanacions detallades per a millorar.

En relació amb l'assoliment de la tasca, les dues eines d'IA de vegades proporcionen retroaccions contradictòries. Per exemple, pel que fa al text narratiu d'un aprenent de C1 que no respon correctament a l'enunciat, ChatGPT afirma: «The essay effectively addresses the prompt by discussing how music transcends boundaries and connects people emotionally». No obstant això, MyEssai sí que identifica que l'aprenent es desvia de la tasca: «The quote you are reacting to speaks specifically to music's ability to communicate in ways that surpass verbal and physical expression». A més, MyEssai inclou un suggeriment per ajustar l'escriptura més estretament a la tasca: «Consider focusing on a narrative that directly illustrates this unique ability of music to transcend those boundaries, instead of speaking broadly about its various applications». En canvi, amb un dels textos narratius de nivell B1, només ChatGPT identifica que la tasca no s'ajusta bé al text: «While the story begins with the given sentence, it deviates from the prompt's context, focusing more on the protagonist's experience with BTS tickets rather than an incident requiring a call to the police». Tanmateix, no hi ha cap consell per corregir aquest problema.

En analitzar la retroacció sobre el contingut per a cada nivell de competència, apareixen patrons diferents entre ChatGPT i MyEssai. ChatGPT proporciona tant retroacció correctiva com positiva per als aprenents B1, però per als nivells B2 i C1 només ofereix retroacció positiva. En contrast, MyEssai mostra poques diferències en el seu enfocament de retroacció segons els nivells. Per exemple, en el text narratiu d'un sol aprenent de nivell B1, MyEssai suggereix donar més detalls, desenvolupar els personatges, millorar el desenvolupament de l'acció per evitar un final brusc, etc. A més, cada recomanació ocupa un paràgraf. Així, els nombrosos suggeriments i la seva extensió prolongada poden resultar aclaparadores per a un aprenent d'aquest nivell. Això indica que, mentre que ChatGPT ajusta, fins a cert punt, la complexitat de la retroacció sobre el contingut segons el nivell de l'aprenent, MyEssai proporciona constantment retroacció detallada i completa, independentment del nivell de competència de l'aprenent.

Quant a la tipologia de text, MyEssai sembla adaptar la retroacció sobre el contingut per proporcionar les correccions i suggeriments més adequats. En els textos narratius, fa èmfasi en la creació d'una narrativa més atractiva i evocadora, proporcionant retroacció detallada i específica sobre el context, animant els aprenents a transmetre vívidament les emocions i les experiències. En la redacció de correus electrònics, MyEssai es concentra en el professionalisme i la formalitat, mantenint la claredat i l'eficàcia: «It's good to make the personal impact clear, but it might help to stick to the most pertinent details to keep the letter concise».

De manera similar, ChatGPT ofereix consells diferents segons el tipus de text. Per exemple, per a un text narratiu comenta: «There's room for further development of the character's emotions and reactions to the unfolding events. Adding more depth to the protagonist's thoughts and feelings can enhance the story's impact». També pretén que les narracions siguin coherents i clares: «It's not clear why the protagonist's friend didn't respond or why the police were called». En la redacció de correus electrònics, ChatGPT no fa tant èmfasi en la formalitat com MyEssai, sinó que se centra en la claredat en la comunicació: «Some questions are unclear or could be more specific. For example, you could ask about specific challenges Stephanie faced while living alone rather than just asking what she learned».

3.2. Fortaleses i debilitats generals de ChatGPT i MyEssai

Hi ha diferències significatives en la manera que MyEssai i ChatGPT han estat dissenyats, i, per tant, els comentaris que poden proporcionar són diferents. Un avantatge important de ChatGPT és la seva capacitat de ser personalitzat. Els usuaris poden sol·licitar comentaris específics o adaptar-los a les característiques o necessitats individuals. Tanmateix, això requereix un *prompt* precís, ja que les consultes ambigües poden causar respostes inexactes. A més, els seus comentaris tendeixen a ser més breus i, de vegades, superficials, com es demostra pel fet que ofereix els mateixos suggeriments exactes a aprenents diferents. En canvi, MyEssai ofereix comentaris extensos i específics del contingut, però li manca la capacitat de personalització de ChatGPT. No es pot ajustar al nivell de competència de l'usuari ni centrar-se en àrees específiques.

La manera que les eines d'IA proporcionen comentaris és distinta, amb cada eina utilitzant el seu propi mètode. MyEssai, normalment, té un to més amable: «Your plot takes an unexpected turn, which is great for surprise, but the resolution feels a bit abrupt. Perhaps you could build up to the twist more gradually». Així doncs, intenta fer comentaris positius per suavitzar l'impacte de la crítica. En canvi, ChatGPT a vegades és molt directe: «The writing lacks clear paragraph breaks, which affects readability and organisation». La crítica es presenta sense cap atenuació. A més, sovint destaca aspectes negatius sense suggerir maneres de millorar. Per exemple, diu a un aprenent de nivell B1: «The sequence of events could be clearer and more logically connected», però no aporta consells sobre com fer-ho.

MyEssai sempre comença i acaba els comentaris motivant l'aprenent. Comença lloant l'aprenent amb comentaris com: «Firstly, I want to commend you for the professional tone of your cover letter. You've presented your complaint in a measured and courteous manner, and that is definitely the right approach when addressing a contentious issue». A més, la conclusió dels comentaris intenta encoratjar l'aprenent perquè continuï progressant: «Keep writing, as your voice is as deserving of an audience as the opera singers who've inspired you. Your passion for the subject is the melody of your prose – let it sing with the same grace and strength you admired on stage». Com es mostra en aquest fragment, els comentaris motivacionals no semblen buits, ja que estan relacio-

nats amb el context del text. A més, MyEssai sovint inclou un resum de tots els comentaris a la conclusió, proporcionant una visió general completa per a l'aprenent.

ChatGPT no és tan consistent en la provisió de comentaris motivacionals al principi, pel fet que només ofereix lloances per als aprenents de B2 i C1 i solen ser comentaris més breus: «Your short story effectively captures the excitement and passion you have for the WRC-Catalonia event». A la conclusió, ofereix declaracions encoratjadores a tots els aprenents, que solen ser més llargues: «Overall, you've done a commendable job of crafting a short story about your experience at the opera. Paying attention to minor grammar errors and refining your vocabulary choices can help enhance the clarity and impact of your writing. Keep up the good work!». Encara que no siguin tan personalitzats com els de MyEssai, els comentaris també són individuals per a cada aprenent, ja que es refereixen al seu text. En lloc de resumir els punts importants, ChatGPT tendeix a esmentar les àrees que creu que necessiten més millora: «Paying attention to grammar and organisation can help enhance the clarity and impact of your writing».

4. Implicacions pedagògiques

Les implicacions pedagògiques d'aquest treball estan relacionades amb el paper que MyEssai i ChatGPT podrien tenir en l'àmbit de l'ensenyament i l'aprenentatge de llengües. L'estudi suggereix que els aprenents podrien utilitzar aquestes eines per facilitar el treball autònom, ja que proporcionen comentaris útils. A més, són capaces, fins a cert punt, d'adaptar el tipus de retroalimentació segons els errors que cometen els aprenents, a més d'incloure comentaris positius i motivacionals. Així i tot, alguns problemes detectats en l'anàlisi, com l'ajustament limitat al nivell de l'aprenent i la manca d'informació metalingüística, indiquen la necessitat persistent de la intervenció humana per ajudar els aprenents a comprendre i assimilar la retroacció.

5. Conclusions

Hi ha similituds interessants, així com diferències, entre la retroacció correctiva proporcionada per MyEssai i ChatGPT. Pel que fa al llenguatge, la retroacció de MyEssai és més indirecta que la de ChatGPT, que tendeix a utilitzar moltes correccions explícites. Cap de les dues eines proporciona gaire informació metalingüística. ChatGPT corregeix 89 errors formals, mentre que MyEssai només 50. Això, juntament amb el fet que recomana l'ús de Grammarly en diverses ocasions, indica que l'objectiu principal de MyEssai no és corregir aquests errors, sinó que es preocupa més pel contingut. De fet, la major part de la retroacció de MyEssai se centra en el contingut del text, dedicant-hi diversos paràgrafs que inclouen observacions estretament relacionades amb el text. Contràriament, els comentaris de ChatGPT sobre el contingut de vegades semblen superficials. Així i tot, el rendiment varia en l'assoliment de la tasca: de vegades és ChatGPT el que detecta desviacions de l'enunciat, mentre que altres vegades és només MyEssai qui ho fa. En termes d'organització, MyEssai tendeix a ser més específic, oferint exemples de com estructurar el text, mentre que ChatGPT sol donar consells breus.

Respecte a la retroacció positiva i la motivació, ChatGPT proporciona retroacció positiva en més àrees que MyEssai, i de vegades exclou la retroacció correctiva. En canvi, MyEssai busca establir una connexió amigable amb l'aprenent abans de proporcionar retroacció correctiva, mentre que ChatGPT pot ser força directe. A més, MyEssai fa un esforç més gran per ajudar l'aprenent a entendre la importància de millorar els aspectes que suggereix. Tot i que ChatGPT inclou comentaris similars, aquests són menys freqüents en comparació amb els de MyEssai. Totes dues eines d'IA acaben sistemàticament els comentaris animant els aprenents a continuar progressant, un element crucial per a la seva motivació.

Cap de les dues eines sembla capaç d'ajustar la retroacció al nivell de competència de l'aprenent. Tot i això, ChatGPT és el que mostra més diferències en la retroacció en funció del nivell de l'aprenent, mentre que MyEssai no fa cap mena de distinció i l'extensió de la seva retroacció pot ser contraproduent per a aprenents de nivell inicial. Per tant,

la retroacció detallada de MyEssai pot ser més útil per als aprenents avançats, que sabran millor com aplicar les seves correccions als textos. En canvi, la simplicitat i la claredat de ChatGPT poden ser més útils en els nivells inicials.

En canvi, MyEssai adapta la retroacció de manera més efectiva al tipus de text. Per exemple, en els correus electrònics, se centra especialment en aspectes de formalitat i en escriure de manera eficient perquè el destinatari pugui entendre clarament les intencions de l'autor. En l'escriptura creativa, ofereix nombrosos suggeriments per fer que les històries siguin més atractives i evocadores. ChatGPT també ajusta la retroacció suggerint històries més captivadores i correus electrònics més comprensibles, però s'adapta menys, ja que les seves recomanacions entre tipologies textuals no difereixen tant com les de MyEssai.

En conclusió, tot i que MyEssai i ChatGPT comparteixen algunes característiques, el seu rendiment varia significativament. Per tant, l'elecció de l'eina hauria de dependre de les circumstàncies individuals i de les necessitats específiques de l'aprenent. Encara cal més investigació per explorar l'impacte que aquestes eines poden tenir en la producció escrita dels aprenents. Tanmateix, l'estudi actual ha demostrat que, si bé totes dues eines d'IA presenten punts forts, també tenen problemes, de manera que cal complementar-les amb la retroacció proporcionada pels docents. Com a treball futur, es podria ampliar l'estudi comparant les correccions proporcionades per altres eines (per exemple, Grammarly o Paperpal) o altres LLM, com ara Claude o Gemini.

4. AVALUACIÓ DE LA CAPACITAT DELS BOTS DE CONVERSA AMB IA EN LA DETECCIÓ D'IRONIA: UN ESTUDI DE CHATGPT, COPILOT I GEMINI

› **Ionela Valentina Vasilovici**

1. Introducció

La ironia és un recurs lingüístic important en relació amb la parla humana, ja que es fa servir per transmetre idees complexes, emocions i sentiments, entre altres coses. A més, forma part de la creativitat del llenguatge humà i té diverses funcions. Així doncs, a mesura que avança la IA, va augmentant l'interès per fer que les màquines capturin el llenguatge no literal dels humans i vagin més enllà d'un simple conjunt de paraules, ja que la detecció de la ironia té moltes aplicacions (per exemple, anàlisi de sentiments o d'opinions). Els investigadors en IA tenen com a objectiu millorar la comunicació entre l'ésser humà i la màquina, i la ironia verbal és un element clau en el llenguatge no literal i en el camp de la pragmàtica.

En aquest context, l'aparició de ChatGPT i altres bots de conversa amb IA o LLM com ara Copilot (abans Bing) iGemini (abans Bard) representa un avanç significatiu, ja que poden entendre un *prompt* i proporcionar una resposta creativa. En altres paraules, amb l'aparició d'aquests bots de conversa, s'obre un ventall nou de possibilitats per explorar la seva comprensió de la producció (escrita i oral) humana i la seva capacitat per detectar ironia i classificar funcions comunicatives. A més, la comprensió de la ironia verbal permet als bots tenir una idea més clara del que es demana. Per tant, és interessant analitzar-ne el potencial i les limitacions per contribuir al camp de les tecnologies del llenguatge i al seu desenvolupament continu. L'última proposta presentada en aquest volum s'endinsa en la detecció de la ironia per avaluar fins a quin punt la IA pot comprendre i detectar fenòmens lingüístics complexos. Més específicament, l'objectiu de l'estudi és posar a prova la capacitat dels bots de conversa esmentats per entendre com abasten un recurs lingüístic complex com la ironia i com classifiquen les funcions comunicatives. Per explorar aquest tema, aquest estudi aborda les següents preguntes de recerca:

1. Quins són els punts forts i les limitacions dels bots de conversa amb IA com ChatGPT, Microsoft Copilot i Google Gemini en relació amb la detecció de la ironia?
2. Quin d'ells és més precís? Quines són les diferències?
3. Quin tipus de *prompt* (*zero-shot* o *few-shot*) funciona millor pel que fa a la classificació de la ironia?

2. Metodologia

2.1. Disseny del guió per a la detecció de la ironia

Per tal d'explorar si ChatGPT, Copilot i Gemini són capaços de detectar ironia, i per poder-ne valorar els resultats, s'ha creat un guió que consisteix en una conversa entre dos personatges:

> Personatge 1 (Ally): La seva funció principal és expressar la majoria d'enunciats irònics.

> Personatge 2 (Daniel): La seva funció principal és donar resposta a aquestes expressions i permetre que la conversa continuï.

Els parlants inicien una conversa mentre esperen el seu tren i el Personatge 1 es queixa constantment de la seva feina i de la seva companya de pis, tot utilitzant moltes expressions irròniques. El guió inclou disset exemples d'ironia que es poden classificar segons el seu recurs retòric (hipèrbole, metàfora, exclamació, pregunta retòrica i símil), ja que l'objectiu és tenir una varietat d'expressions irròniques. A més, les categories també s'utilitzaran en la interacció amb els bots de conversa. S'han inclòs tres hipèrboles, quatre metàfores, quatre exclamacions, tres preguntes retòriques i tres símils. El Personatge 1 és el que expressa la majoria dels exemples d'ironia (15). En canvi, el Personatge 2 només produeix dos exemples per explorar si hi ha alguna diferència rellevant pel que fa a la detecció de la ironia. La taula 1 conté tots els exemples d'ironia amb el recurs retòric corresponent, en el mateix ordre en què apareixen en el guió. No obstant això, també inclou una columna amb altres categories que serien acceptades, ja que algunes poden superposar-se.

Taula 1. Expressions irròniques del guió i les seves categories corresponents

Expressió irònica	Categoria principal	Altres categories acceptades
1- It's as cold as ice today, huh?	Símil	Hipèrbole
2- My job is a dream!	Exclamació	Hipèrbole, metàfora
3- You can imagine that my boss is such an angel.	Metàfora	Hipèrbole
4- Do you also have a boss who is as happy as a clam?	Símil	-
5- As his voice was music to my ears...	Metàfora	-
6- I love working here for 12 hours in a row!	Exclamació	Hipèrbole
7- I have just been waiting here for 5 hours.	Hipèrbole	-
8- I can see that calmness is your friend.	Metàfora	-
9- This is the best week of my life!	Exclamació	Hipèrbole
10- ...no one dares to complain about the best boss of all time.	Hipèrbole	-
11- She is always as busy as a bee.	Símil	-
12- Aren't you tired of studying all day?	Pregunta retòrica	-
13- Who wouldn't want to read twenty romance books in a month?	Pregunta retòrica	-
14- ...how could I be so absent-minded as to forget a book that I never picked up?	Pregunta retòrica	-
15- I have done this a million times!	Exclamació	Hipèrbole
16- Does it have something to do with the super-organised flatmate?	Hipèrbole	-
17-You are the worst kind of monster.	Metàfora	Hipèrbole

Aquests exemples d'ironia s'insereixen artificialment en la conversa perquè el bot els detecti. De cara a facilitar-ne la interpretació, inclouen contrastos forts dins l'entorn. Per exemple, hi ha una gran contradicció en aquesta afirmació: «She is always as busy as a bee: chilling on the couch all day!». Així doncs, el context i els contrastos són clau en aquest text. A més, la conversa inclou exemples artificials d'ironia en lloc d'una conversa natural, ja que ha d'incloure una quantitat re-

presentativa d'expressions irròniques de cada categoria (tres o quatre exemples d'ironia de cinc categories en una conversa natural no és realista). Finalment, el to humorístic general contribueix a la comprensió dels enunciats irònics.

2.2. Aplicació del guió per a la detecció de la ironia

Abans d'aprofundir en la interacció amb els bots de conversa, cal tenir en compte les característiques de cadascun i les propietats que puguin ser rellevants, com es mostra a la taula 2.

Taula 2. Bots de conversa utilitzats en aquest estudi i les seves característiques

Bot de conversa amb IA[6]	LLM	Preu	Inici de sessió	Propietats	Dates d'accés
ChatGPT	GPT-3.5	Gratuït (versió de pagament disponible basada en GPT-4)	Creant un compte al lloc web	-	4/5/2024 - 21/5/2024 (actualització el 16 de maig)
Copilot	GPT-4	Gratuït	Mitjançant un compte de Microsoft	Estil de conversa precís	4/5/2024 - 20/5/2024
Gemini	PALM-2	Gratuït (versió de pagament disponible anomenada Gemini Advanced)	Mitjançant un compte de Google	-	4/5/2024 - 20/5/2024 (actualització el 14 de maig)

Les dades de la taula 2 indiquen que tant GPT-4 com PaLM-2 són tecnologies més avançades que GPT-3.5. A més, Copilot ofereix tres estils de conversa diferents: creatiu, equilibrat i precís. L'estil creatiu és per a iniciar un xat original i imaginatiu, l'estil equilibrat és per a xats quotidians i l'estil precís és per a iniciar un xat concís (útil per a la recerca de fets). A l'efecte d'aquest estudi, i en relació amb la detecció d'ironia, l'estil conversacional precís és el que millor funciona.

Pel que fa a la interacció, la conversa amb els bots consisteix en dos passos i dos tipus de preguntes, com s'il·lustra a la taula 3.

6. Inspirada en la taula de la recerca de Morreel et al. (2024) https://doi.org/10.1371/journal.pdig.0000349.t001

Taula 3. Tipus de tècniques i prompts utilitzats en la fase d'interacció

	Prompt 1: Identificació	*Prompt* 2: Classificació
Zero-shot	Hi poden haver (1) alguns exemples d'ironia en el següent guió. Pots identificar si hi ha exemples d'ironia verbal (2)? Identifica'ls tots (3).	Classifica els exemples d'ironia que has identificat en el guió en les categories següents: hipèrbole, metàfora, exclamació, pregunta retòrica i símil. Hi pot haver més d'un exemple d'ironia en cada categoria.
Few-shot	Hi poden haver alguns exemples d'ironia en el següent guió. Pots identificar si hi ha exemples d'ironia verbal? Identifica'ls tots.	Classifica els exemples d'ironia que has identificat en el guió en les categories següents: hipèrbole, metàfora, exclamació, pregunta retòrica i símil. Hi pot haver més d'un exemple d'ironia en cada categoria. Aquí tens exemples de cada categoria: - Hipèrbole: «This suitcase weighs a ton» (The suitcase doesn't actually weigh a ton, but it feels very heavy.) - Metàfora: «He is a lion on the battlefield» (This metaphor compares a person to a lion, suggesting that they are fierce and powerful in a particular situation.) - Exclamació: «What an amazing view!» - Pregunta retòrica: «Do you want to lose weight without feeling hungry?» (This question is used to persuade the audience to consider a particular approach to weight loss, rather than expecting a direct answer.) - Símil: «As quick as a flash»

En la fase *zero-shot*, es proporciona un *prompt* al model que el guia per completar una tasca específica sense cap exemple, mentre que al *few-shot* es proporciona al model un petit nombre d'exemples (o només un) per a una tasca en particular, juntament amb un *prompt*. Pel que fa al primer *prompt*, que és exactament el mateix tant a *zero-shot* com a *few-shot*, hi ha algunes paraules o expressions que són rellevants per aconseguir que el bot de conversa amb IA faci el que es demana:

1. No s'assumeix que hi ha ironia verbal. El bot de conversa està forçat a detectar-la.
2. Cal especificar que ha de trobar ironia verbal. En cas contrari, el bot pot identificar altres tipus d'ironia, com ara la ironia situacional.
3. Es demana al bot que els identifiqui tots, ja que podria només identificar un nombre específic d'exemples (només cinc o sis).

De la mateixa manera, pel que fa al segon *prompt*, s'inclou aquesta frase: «Hi pot haver més d'un exemple d'ironia en cada categoria». Si no s'afegeix això, el bot de conversa pot classificar només un exemple en cada categoria. Per tant, la primera fase és la identificació o detecció de la ironia verbal, mentre que la segona fase consisteix en la classificació dels exemples identificats en les cinc categories que s'han esmentat prèviament. Això aporta informació sobre la detecció de la ironia i sobre la capacitat del bot per entendre i classificar el llenguatge en dos contextos diferents: amb exemples de cada categoria i sense. Així, aquests dos passos donen una visió més àmplia de la seva capacitat en relació amb el llenguatge.

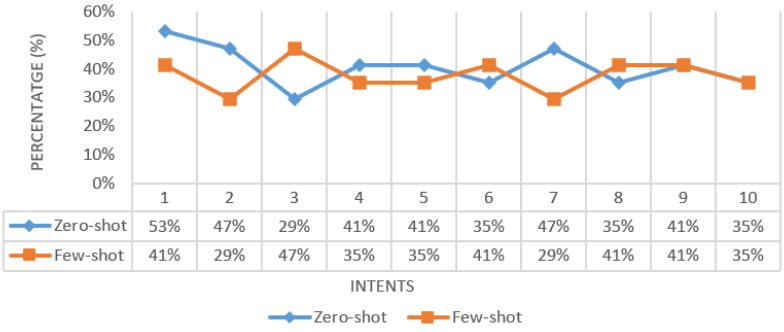
La interacció consisteix en deu intents per a cada bot de conversa i cada tipus de *prompt* (*zero-shot* i *few-shot*) en dies diferents, ja que els resultats varien. A més, s'utilitza un nou xat per a cada intent, per tal d'evitar la influència entre les interaccions. Tots els resultats s'han registrat en una taula que inclou informació sobre el nombre d'exemples irònics identificats en cada bot de conversa i de quina categoria són, el percentatge d'encerts, el nombre d'errors, el nombre de categories classificades correctament juntament amb el percentatge d'encerts i el nombre d'errors, la data d'accés i l'enllaç de la conversa. De vegades, el nombre d'exemples d'ironia identificats i el nombre d'exemples classificats correctament no coincideixen, ja que podria haver-hi un exemple que es classifiqués correctament, però no s'identifiqués en la primera fase. De la mateixa manera, algunes de les categories poden superposar-se.

3. Resultats

3.1. Resultats de la detecció de la ironia per part de ChatGPT

Com ja s'ha esmentat, ChatGPT utilitza la tecnologia GPT-3.5, que no està tan avançada com GPT-4. Pel que fa al *prompt* d'identificació, el bot reconeix, de mitjana, el 41% dels exemples d'ironia del guió (aproximadament, 7 de 17) en *zero-shot*, i el 38% en *few-shot* (aproximadament, 6 de 17), com s'il·lustra en el gràfic 1.

Gràfic 1. Fase 1 – Identificació per part de ChatGPT



	1	2	3	4	5	6	7	8	9	10
Zero-shot	53%	47%	29%	41%	41%	35%	47%	35%	41%	35%
Few-shot	41%	29%	47%	35%	35%	41%	29%	41%	41%	35%

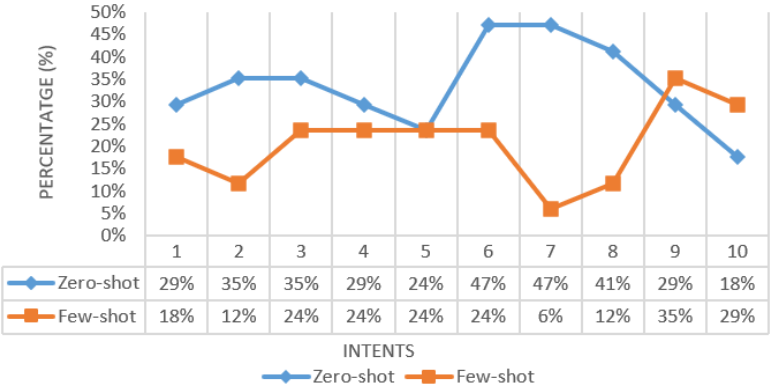
Els resultats són força similars i, després d'analitzar totes les dades, la conclusió és que la diferència entre *zero-shot* i *few-shot* (*zero-shot* identifica un 3% més d'exemples) no és significativa, ja que el nombre d'exemples identificats depèn exclusivament de l'intent, i aquesta primera tasca és exactament la mateixa en tots dos tipus de *prompt*. És a dir, el nombre d'expressions irròniques identificades en cada intent varia, però no hi ha cap raó específica per a aquesta variació de resultats. A més, el nombre d'errors d'excés d'identificacions fluctua de 0 a 3, amb una mitjana d'1 error en ambdós tipus de *prompt*.

Pel que fa als millors resultats en *zero-shot*, el primer intent va ser el més precís en relació amb la detecció d'ironia. Encara que va tenir un error d'excés d'identificació, ChatGPT va ser capaç d'identificar 9 enunciats irònics de 17 (53%). D'altra banda, el pitjor intent va ser el tercer, amb només 5 exemples d'ironia detectats (29%). A més, va identificar massa expressions irròniques i va cometre 3 errors. Pel que fa a *few-shot*, el millor intent va ser el tercer, amb un 47% dels exemples identificats, i només va tenir un error d'excés d'identificació. Els intents 2 i 7 van detectar només el 29% de les expressions irròniques, dada que coincideix amb el pitjor percentatge de *zero-shot*. A més, hi va haver una actualització a la pàgina web el 16 de maig de 2024, però no és rellevant en relació amb aquest estudi, ja que els resultats van continuar sent els esperats a partir dels primers intents.

De mitjana, a la fase de classificació, el 34% dels exemples d'ironia (anteriorment) identificats s'han classificat correctament en *zero-shot* (aproximadament, 6 exemples en cada intent), i el 21% en *few-shot*

(aproximadament, 3-4 exemples en cada intent), mostrant millors resultats en *zero-shot* (gràfic 2).

Gràfic 2. Fase 2 – Classificació per part de ChatGPT



	1	2	3	4	5	6	7	8	9	10
Zero-shot	29%	35%	35%	29%	24%	47%	47%	41%	29%	18%
Few-shot	18%	12%	24%	24%	24%	24%	6%	12%	35%	29%

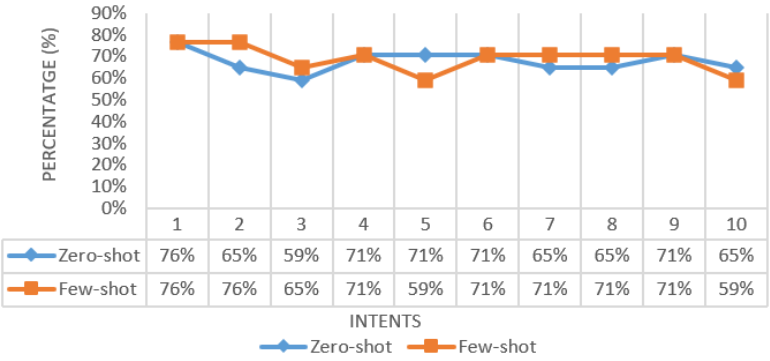
A més, hi ha, de mitjana, 3 errors d'excés d'identificació en *zero-shot* i 2 en *few-shot*. En aquesta fase, sembla que *few-shot* millora amb el temps, mentre que *zero-shot* fa el contrari. De vegades, els números no coincideixen entre les fases 1 i 2, ja que podria haver-hi algunes expressions irròniques classificades correctament, però que no es van identificar en el pas anterior, com en l'intent 6 en *zero-shot* (hi ha 2 símils que es classifiquen en la segona fase, però no es van detectar en la fase d'identificació). A més, com que algunes categories poden superposar-se, els números poden variar de la fase 1 a la fase 2.

En general, ChatGPT i la tecnologia GPT-3.5 han identificat menys de la meitat de les expressions irròniques presents en el guió. No obstant això, no tendeixen a cometre molts errors d'excés d'identificació, la qual cosa és positiva. D'altra banda, també han estat capaços de classificar menys de la meitat dels exemples d'ironia, i el bot tendeix a ser més precís quan no es proporcionen exemples (és a dir, amb *zero-shot*). També sol haver-hi més errors que en la primera fase (entre 0 i 6). Normalment, els errors són classificacions errònies de frases no irròniques que no s'han identificat prèviament, frases irròniques que no coincideixen amb aquella categoria, o una justificació incorrecta.

3.2. Resultats de la detecció de la ironia per part de Copilot

Microsoft Copilot utilitza la tecnologia GPT-4 i la cerca en línia. Per tant, s'esperaven millors resultats pel que fa a la detecció d'ironia, com es mostra en el gràfic 3.

Gràfic 3. Fase 1 – Identificació per part de Copilot



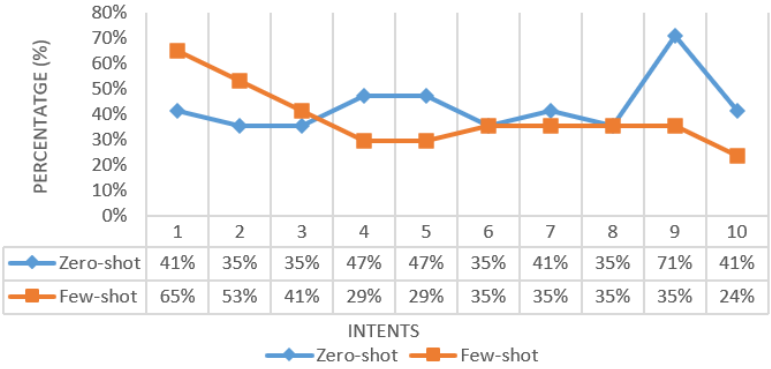
	1	2	3	4	5	6	7	8	9	10
Zero-shot	76%	65%	59%	71%	71%	71%	65%	65%	71%	65%
Few-shot	76%	76%	65%	71%	59%	71%	71%	71%	71%	59%

Copilot ha pogut detectar de mitjana un 68% dels exemples d'ironia en *zero-shot* (aproximadament, 11-12 de 17) i un 69% en *few-shot* (aproximadament, 12 de 17), mostrant resultats similars en la fase d'identificació. A més, el bot de conversa no ha comès cap error d'excés d'identificació en cap dels intents. L'intent 1 en *zero-shot* va mostrar els millors resultats pel que fa a la detecció d'ironia, atès que va identificar 13 enunciats irònics de 17 (76%). Pel que fa a *few-shot*, els millors intents són el primer i el segon, ja que també s'han detectat 13 exemples. D'altra banda, només es van reconèixer 10 expressions iròniques en el tercer intent de *zero-shot*. En *few-shot*, els intents 5 i 10 són els menys precisos, ja que Copilot va trobar 10 exemples en el guió. No obstant això, com que el nombre fluctua de 10 a 13, els resultats són bastant estables al llarg dels intents.

Pel que fa a la segona part de la interacció (gràfic 4), el 43% (de mitjana) de les expressions iròniques es van agrupar correctament en *zero-shot* (aproximadament, 7 exemples en cada intent) i el 38% en *few-shot* (aproximadament, 6 exemples en cada intent). Novament, alguns exemples que es classifiquen en aquesta fase poden no ser identificats

en l'anterior. A més, els *prompts* amb *zero-shot* mostren millors resultats pel que fa a la classificació de les expressions irròniques. Com en el cas de ChatGPT, les categories que més s'han identificat són la metàfora i l'exclamació.

Gràfic 4. Fase 2 – Classificació per part de Copilot



No obstant això, de mitjana, hi ha el mateix nombre d'errors de classificació errònia en els dos tipus de *prompts* (aproximadament, 3 errors en cada intent), que fluctua de l'1 al 5. En aquesta fase, a banda de classificar les categories, el bot també ofereix una breu definició del recurs lingüístic.

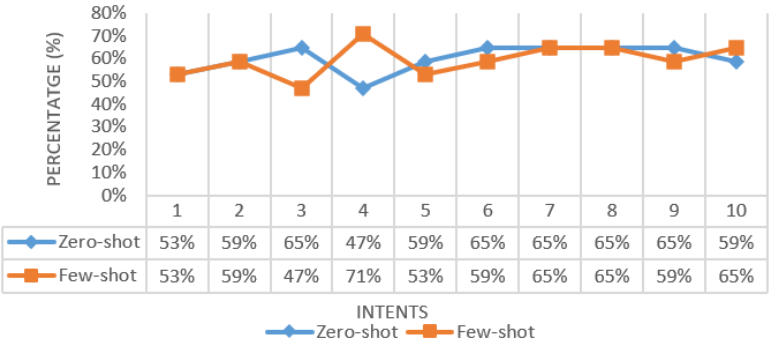
En general, Copilot i la tecnologia personalitzada GPT-4 són prometedors. Com era d'esperar, ha superat ChatGPT 3.5 en relació amb la detecció d'ironia i classificació del llenguatge. El bot de conversa ha pogut identificar entre 10 i 13 exemples d'ironia, que representa el 68-69% de totes les expressions irròniques del guió. A més, sol ser força més precís, ja que no comet errors d'excés d'identificació en la primera fase. Pel que fa a la segona fase, el bot ha classificat gairebé la meitat dels exemples d'ironia (anteriorment) identificats. En definitiva, en relació amb l'objectiu d'aquest estudi, Copilot ha mostrat els resultats més bons.

3.3. Resultats de la detecció de la ironia per part de Gemini

Gemini funciona amb tecnologia PaLM-2 i és un LLM diferent que no està relacionat amb la tecnologia GPT. No obstant això, utilitza un mo-

tor similar desenvolupat per Google i és el bot de conversa amb IA més avançat de l'empresa. Per tant, també s'esperaven bons resultats pel que fa a la detecció d'ironia, com es mostra en el gràfic 5.

Gràfic 5. Fase 1 – Identificació per part de Gemini



Gemini ha estat capaç d'identificar de mitjana el 60% dels exemples d'ironia en *zero-shot* (10 de 17, aproximadament) i el 59% en *few-shot* (també 10 de 17 exemples en cada intent, aproximadament). A més, no hi ha hagut errors d'excés d'identificació en la majoria dels intents. No obstant això, si un intent inclou algun error, només n'hi ha 1 o 2. Tot i que els resultats són bastant estables, el millor intent és el quart en *few-shot*, ja que el bot va detectar el 71% de les expressions irròniques en el guió. Pel que fa als pitjors resultats, es va trobar el 47% dels exemples d'ironia tant en *zero-shot* com en *few-shot* (intents 4 i 3, respectivament). Gemini es va actualitzar el 14 de maig, tot i que aquest fet no és rellevant en relació amb el propòsit d'aquesta recerca, ja que els resultats van continuar sent els esperats a partir dels primers intents.

Pel que fa a la classificació de categories, Gemini ha classificat, de mitjana, un 17% dels exemples d'ironia (aproximadament, 3 exemples en cada intent) en *zero-shot* i un 15% en *few-shot* (aproximadament, 3 en cada intent). En aquest cas, la diferència entre els dos tipus de *prompts* no és molt significativa, ja que mostren resultats similars. A més, com es mostra en el gràfic 6, són bastant inestables i tendeixen a canviar molt depenent de l'intent.

Gràfic 6. Fase 2 – Classificació per part de Gemini

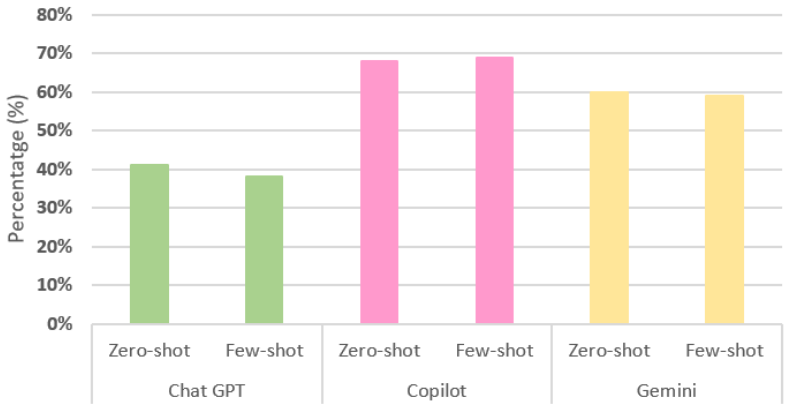
El nombre d'exemples no coincideix entre les fases 1 i 2, a causa de la superposició de categories i de la classificació d'expressions iròniques que no s'han identificat prèviament. A més, Gemini fa de mitjana 1 error de classificació errònia en ambdós tipus de *prompts*. Tot i ser un bot de conversa amb IA avançat i comparable a GPT-4, ha tingut problemes a l'hora de classificar els exemples d'ironia en les categories de l'estudi. És a dir, tot i haver-li especificat que ha de classificar els enunciats irònics en les cinc categories, el bot es va negar a fer-ho en gairebé tots els intents i, finalment, es va veure obligat a fer-ho després d'un parell d'interaccions. Per tant, en comptes de cenyir-se a les instruccions de l'usuari, Gemini desafia la classificació donada i ofereix la seva pròpia organització i classificació d'ironia.

En general, Gemini és un bot de conversa amb IA interessant, ja que és capaç de detectar més de la meitat dels exemples irònics presents en el guió. No obstant això, també desafia el *prompt* donat per a la classificació de categories i mostra la seva pròpia organització en relació amb els enunciats irònics, que normalment té relació amb el sarcasme, el context i les contradiccions.

4. Discussió

Si es comparen tots els bots de conversa amb IA, s'arriba a la conclusió que Copilot supera ChatGPT i Gemini (gràfic 7) en la detecció d'ironia. A més, ChatGPT ocupa la tercera posició i hi ha una diferència d'un 20% amb Gemini i d'un 30% amb Copilot, aproximadament. S'esperaven millors resultats de Copilot, ja que es tracta d'una versió personalitzada de GPT-4 i també inclou la cerca a internet. A més, és la més estable identificant ironia (sempre troba entre 10 i 13 exemples dels 17 presentats en aquest estudi).

Gràfic 7. Fase 1 – Detecció d'ironia

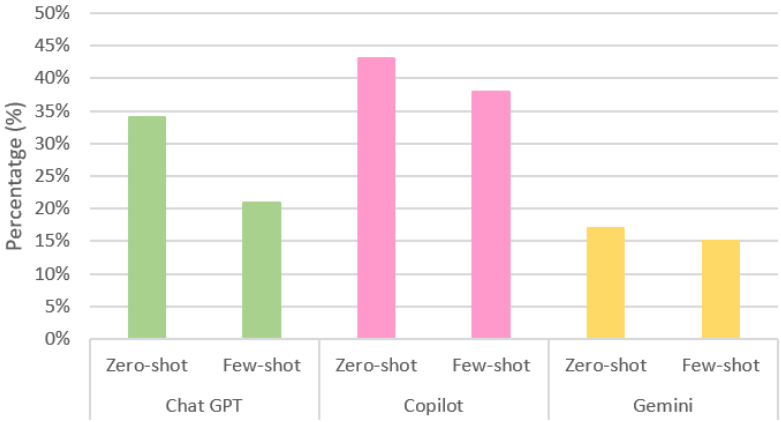


Les metàfores i les exclamacions són les categories que més s'han detectat en tots els bots de conversa. A més, la variació de resultats entre tots els intents de cada bot demostra que poden reconèixer més enunciats irònics dels que realment esmenten. Per exemple, si un exemple d'hipèrbole en particular s'esmenta almenys en un dels intents significa que el bot és capaç de categoritzar-lo com a exemple d'ironia en general, encara que no es detecti en altres intents.

Pel que fa a la classificació de les categories, Copilot també supera els altres bots. A més, hi ha diferències significatives entre els *prompts zero-shot* i *few-shot*. Les dades mostren que la classificació de categories funciona millor en ChatGPT i Copilot quan no se'n proporciona

cap exemple en el *prompt*, mentre que la diferència entre els tipus de *prompts* no és tan significativa en Gemini, com es representa en el gràfic 8. Per tant, en general, s'obtenen resultats més bons si no es donen exemples i els bots es basen només en les dades en què s'han entrenat.

Gràfic 8. Fase 2 – Classificació



En definitiva, Copilot és el bot que ha proporcionat resultats més bons en relació amb la detecció i la classificació d'ironia, i ChatGPT ha obtingut els pitjors resultats en la detecció d'ironia en comparació amb els altres bots de conversa amb IA, tot i que ha superat Gemini en la classificació. Com que GPT-3.5 és el predecessor de GPT-4, es pot observar una certa evolució i millora en el model. ChatGPT tracta amb dades emmagatzemades fins al 2021, mentre que Copilot es basa en GPT-4 i altres dades de la cerca a internet (Bing). En altres paraules, hi ha una millora d'un LLM a un altre en relació amb la detecció de la ironia, ja que Copilot està entrenat amb més dades i també es basa en informació actualitzada del cercador a internet. D'altra banda, Gemini, el bot més avançat de Google, també és interessant d'analitzar a causa de la seva tecnologia i un LLM diferent (PaLM-2). Per tant, encara que no superi Copilot, els seus resultats són força precisos i millors que els de ChatGPT. En particular, és important destacar que aquest bot va més enllà de les instruccions donades a través d'un *prompt* de l'usuari. És a dir, mentre que ChatGPT i Copilot s'adhereixen a les instruccions, Gemini és capaç

de desafiar les instruccions de l'usuari si pensa que el *prompt* és incorrecte, ja que es nega a classificar els exemples d'ironia en les categories donades i ofereix la seva pròpia classificació per identificar expressions irròniques.

5. Conclusió i limitacions de l'estudi

Aquesta investigació ha explorat el potencial de la IA per detectar la ironia a través d'alguns dels bots de conversa amb IA més comuns entre els usuaris (OpenAI ChatGPT, Microsoft Copilot i Google Gemini), ja que la ironia és un recurs lingüístic molt utilitzat en el llenguatge humà per transmetre idees complexes. Per tant, és interessant que els bots estiguin entrenats per poder copsar expressions irròniques i tenir una millor comprensió dels *prompts* donats per l'usuari. Després d'utilitzar el guió preparat per posar a prova les IA, s'ha trobat que Copilot supera els altres bots en relació amb aquest propòsit, i també és millor classificant expressions irròniques en comparació amb els altres. A més, pel que fa a la classificació en les cinc categories, el *prompt zero-shot* funciona millor que el de *few-shot* en ChatGPT i Copilot, mentre que la diferència no és tan significativa en Gemini. En altres paraules, els bots tenen millors resultats si no es proporcionen exemples de les categories en el *prompt*. A més, sembla que la raó per la qual Copilot és millor que ChatGPT en la detecció d'ironia és que GPT-3.5 és el predecessor de GPT-4. La tecnologia GPT-3.5 està entrenada només amb dades disponibles fins al 2021, mentre que la versió personalitzada de GPT-4 de Copilot s'ha entrenat amb quantitats més grans de dades i també es basa en la informació disponible en la cerca d'internet a Bing. Pel que fa a Gemini, que és el bot més avançat de Google i que es basa en PaLM-2, també té bons resultats en relació amb la identificació de la ironia. No obstant això, i en comparació amb els altres, aquest bot desafia les instruccions de l'usuari. És a dir, Gemini es nega a classificar els exemples en les categories especificades. Després d'analitzar totes les dades, es pot concloure que els resultats són positius i prometedors. Tot i que els bots tenen algunes limitacions (errors d'excés d'identificació, exemples que no s'identifiquen, etc.), el seu potencial en relació amb la ironia, el llenguatge no literal i la pragmàtica està millorant, ja que GPT-4 i PaLM 2, que estan entrenats amb més dades, superen GPT-3.5, que

és més antic. En altres paraules, sembla que els bots són cada vegada més precisos i estan entenent millor el llenguatge natural, incloent-hi enunciats irònics. Per consegüent, això pot implicar que futurs LLM i bots de conversa amb IA siguin encara superiors.

Aquest estudi ha contribuït als camps de la lingüística, la pragmàtica i les tecnologies del llenguatge. A més, podria ser d'interès per a l'anàlisi de sentiments i la recerca d'anàlisi d'opinions. Com que els LLM continuen evolucionant i apareixen nous bots, és difícil mantenir-se al dia i hi ha poca investigació sobre la detecció de la ironia i els bots de conversa amb IA. Aquest estudi pretén omplir un nou buit creat a partir de l'explosió de la IA, el PLN i els LLM: com aborden aquestes tecnologies el llenguatge complex (com, per exemple, les expressions no literals). Tot i això, la metodologia utilitzada té algunes limitacions. S'ha creat un guió amb una varietat d'exemples d'ironia, però s'implementen artificialment en el text i no representen una conversa natural o el *prompt* real d'un usuari. A més, el potencial dels bots de conversa es restringeix a aquest guió i als *prompts* utilitzats específicament per a aquest propòsit. Pel que fa a futurs treballs de recerca en aquesta àrea, seria interessant analitzar una mostra més representativa d'exemples d'ironia utilitzats en un ventall més ampli de contextos. Atès que el nombre d'exemples d'ironia en el guió és força baix, la quantitat d'expressions iròniques hauria d'augmentar. De la mateixa manera, un tipus diferent de *prompt* pot alterar els resultats. A més, s'hauria d'explorar més a fons altres bots, així com continuar explorant els d'aquest estudi per tal d'analitzar canvis i tendències i contribuir a l'evolució contínua dels bots, els LLM i les tecnologies del llenguatge.

5. CONCLUSIONS

Els tres treballs presentats en aquest volum han analitzat l'ús de la IA i, en concret, dels LLM, per a la retroacció autocorrectiva en textos i per a la detecció de la ironia.

En el cas dels treballs sobre la retroacció autocorrectiva, cal destacar la diferència que hi ha en la retroacció proporcionada depenent de la llengua dels textos. Tot i haver analitzat el mateix LLM (ChatGPT 3.5), en el cas de l'anglès les correccions fetes són, en general, encertades en tots els nivells explorats. Tanmateix, en el cas de les redaccions en català, moltes de les correccions són errònies, poc precises o, fins i tot, es proporcionen respostes en castellà. Aquí, per tant, es veu clarament el paper cabdal que tenen les dades amb les quals els sistemes han estat entrenats, no només en termes de qualitat, sinó també, especialment, de quantitat. El nombre de dades disponibles en anglès supera àmpliament el de les dades disponibles en català, i això té una clara incidència en els resultats obtinguts. També cal destacar que, en el cas de l'anglès, els sistemes analitzats són capaços de donar retroacció positiva a l'usuari, mentre que en el cas del català, això no és així. De nou, les dades amb els quals s'han entrenat aquests sistemes per a les dues llengües són, possiblement, responsables d'aquest tipus de resposta. El nombre de textos i de materials disponibles per a l'ensenyament de l'anglès com a L2 o com a llengua estrangera és àmpliament superior al nombre de materials per a l'ensenyament del català. Així doncs, el sistema està més familiaritzat amb la tasca de donar retroacció en anglès que no pas en català.

A partir d'aquests resultats, es poden suggerir algunes línies de millora per als LLM aplicats a llengües amb menys recursos, com el català. Per exemple, entrenar aquests models amb corpus normatius específics en català, incloent-hi textos i materials docents, podria millorar-ne la precisió. Així mateix, adaptar les respostes del sistema a diferents nivells educatius (per exemple, usuaris de primària, secundària o universitat) permetria oferir una retroacció més ajustada al context. També seria útil desenvolupar mòduls que incorporessin estratègies de retroacció positiva i constructiva, similar a com ja es fa per a l'anglès, per tal de fomentar la motivació i la millora progressiva en llengües minoritzades.

En el pla pedagògic, sembla que aquests sistemes, i en especial en el cas de llengües més minoritàries com ara el català, convé utilitzar-los per a una correcció inicial, però continua sent necessària la supervisió i intervenció humanes per garantir una correcció precisa i contextualment adequada. Això indica que cal formar l'alumnat i el professorat en l'ús d'aquestes eines, així com avaluar com integrar-les dins l'ensenyament.

Pel que fa a la detecció de la ironia, l'estudi que es presenta en aquest volum demostra que els sistemes d'IA encara han de recórrer molt de camí per poder detectar-la. Detectar la ironia té a veure clarament amb el coneixement del món i de les referències contextuals i culturals de l'usuari. Segons els resultats obtinguts en l'estudi realitzat, la IA encara no té aquest coneixement. Tanmateix, cal destacar que les versions més avançades (Copilot) duen a terme més bé aquesta tasca que els seus predecessors (ChatGPT 3.5), la qual cosa torna a indicar que la disponibilitat de grans quantitats de dades amb les quals entrenar els sistemes és fonamental. La detecció de la ironia és un aspecte clau que els sistemes d'IA han de millorar per excel·lir en la interacció entre l'ésser humà i la màquina.

Com a proposta de millora, seria necessari crear corpus específicament anotats amb exemples d'ironia i altres fenòmens pragmàtics (per exemple, pressuposicions, implicatures, eufemismes, hipèrboles), incorporant-hi també metadades culturals i contextuals. A més, un entrenament híbrid que combini dades lingüístiques amb informació sobre coneixement del món podria ajudar els models a capturar més bé la discrepància entre el significat literal i la intenció comunicativa. Igualment, la integració d'eines multimodals (per exemple, amb suport visual) podria enriquir la interpretació d'aquest tipus de contingut.

Les aportacions recollides en aquest volum mostren com l'ús de la IA en l'anàlisi lingüística pot ajudar a definir criteris clars per entendre i abordar qüestions complexes del llenguatge, alhora que ofereix exemples concrets d'aplicació didàctica en contextos universitaris. L'exploració d'aquesta temàtica en els treballs de final de grau dels estudis de Filologia obre noves vies de reflexió sobre l'ensenyament i l'avaluació de la producció escrita, i convida a repensar el paper de les eines digitals en la formació dels futurs professionals de la llengua. Es tracta d'un camp que comença a prendre forma i que, a partir de l'experiència acumulada en pràctiques docents, pot esdevenir un terreny fèrtil per a la innovació i el pensament crític.

REFERÈNCIES

Almelhes, S. A. (2023). Review of Artificial Intelligence Adoption in Second-Language Learning. *Theory and Practice in Language Studies*, *13*(5), 1259-1269. https://doi.org/10.17507/tpls.1305.21

Banasik, N. (2013). Non-literal speech comprehension in preschool children–An example from a study on verbal irony. *Psychology of Language and Communication*, *17*(3), 309-324.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Amodei, D. et al. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33. https://doi.org/10.48550/arXiv.2005.14165

Byron, C., Wallace, D. K. C., Kertz, L. i Charniak, E. (2014). Humans Require Context to Infer Ironic Intent (so Computers Probably do, too). Dins: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, p. 512-516. Baltimore, Maryland. Association for Computational Linguistics.

Chen, J., He, X. i Miyao, Y. (2024). Language Model Based Unsupervised Dependency Parsing with Conditional Mutual Information and Grammatical Constraints. Dins: *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, p. 6355-6366. Mexico City, Mèxic. Association for Computational Linguistics.

Ettinger, A., Hwang, J. D., Pyatkin, V., Bhagavatula, C. i Choi, Y. (2023). «You Are An Expert Linguistic Annotator»: Limits of LLMs as Analyzers of Abstract Meaning Representation. Dins: *Findings of the Association for Computational Linguistics: EMNLP 2023*, p. 8250-8263, Singapur. Association for Computational Linguistics. 10.18653/v1/2023.findings-emnlp.553

Frenda, S. (2016). Computational rule-based model for irony detection in italian tweets. Dins: *EVALITA 2016*, 1749, p. 1-6.

González, J. A., Hurtado, Ll. F. i Plà, F. (2019). Elirf-upv at irosva: Transformer encoders for spanish irony detection. Dins: *Proceeding of the Iberian Language Evaluation Forum (IberLEF 2019)*. https://ceur-ws.org/Vol-2421/IroSvA_paper_4.pdf

Google (2024). Gemini [Large Language Model]. https://gemini.google.com

Guo, K. i Wang, D. (2023). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*, 1-29. https://doi.org/10.1007/s10639-023-12146-0

Hromei, C. D., Croce, D. i Basili, R. (2025). U-DepPLLaMA: Universal Dependency Parsing via Auto-regressive Large Language Models. *IJCoL*, 10-11. https://doi.org/10.4000/125nm

Ilic, S. Marrese-Taylor, E., Balazs, J. i Matsuo, Y. (2018). Deep contextualized word representations for detecting sarcasm and irony. Dins: *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, p. 2-7. Brussel·less. Association for Computational Linguistics.

Kristiawan, D., Y., Bashar, K. i Pradana, D. A. (2024). Artificial intelligence in English language learning: A systematic review of AI tools, applications, and pedagogical outcomes. *The Art of Teaching English as a Foreign Language* (TATEFL), *5*(2), 207-218. https://doi.org/10.36663/tatefl.v5i2.912

Machado, M. i Ruiz, E. (2024). Evaluating large language models for the tasks of PoS tagging within the Universal Dependency framework. Dins: *Proceedings of the 16th International Conference on Computational Processing of Portuguese - vol. 1*, p. 454-460. Santiago de Compostela. Association for Computational Linguistics.

Microsoft (2023). *Copilot* [Large Language Model]. https://copilot.microsoft.com

Morreel, S., Verhoeven, V. i Mathysen, D. (2024). Microsoft Bing outperforms five other generative artificial intelligence chatbots in the Antwerp University multiple choice medical license exam. *PLOS Digital Health*, *3*(2), e0000349.

Navarro-Carrascosa, C. (coord.) (2024). La enseñanza del Español como Lengua Extranjera y la Inteligencia Artificial: perspectivas y retos. *Doblele: revista de lengua y literatura*, *10*.

OpenAI (2023). *ChatGPT* [Large Language Model]. https://chat.openai.com

Pérez-Núñez, A. (2024). ChatGPT in Spanish language instruction: exploring AI-driven task generation and its implications for teaching practices. *Journal of Spanish Language Teaching, 11*(1), 61-82. https://doi.org/10.1080/23247797.2024.2366053

Potamias, R. A., Siolas, G. i Georgios Stafylopatis, A. (2020). A transformer-based approach to irony and sarcasm detection. *Neural Computing and Applications*, 32(23):17309- 17320.

Reyes, A., Rosso, P. i Veale, T. A. (2013). multidimensional approach for detecting irony in Twitter. *Lang Resources & Evaluation, 47*, 239-268. https://doi.org/10.1007/s10579-012-9196-x

Schmidt, T. i Strasser, T. (2022). Artificial intelligence in foreign language learning and teaching: a CALL for intelligent practice. *Anglistik: International Journal of English Studies, 33*(1), 165-184. https://angl.winter-verlag.de/data/article/11071/pdf/92201014.pdf

Tang, Y. i Chen, H. (2014). Chinese irony corpus construction and ironic structure analysis. Dins: *Proceedings of COLING 2014, 25th International Conference on Computational Linguistics: Technical Papers*, p. 1269-1278.

Taskıran, A. i Goksel, N. (2022). Automated feedback and teacher feedback: Writing achievement in learning English as a foreign language at a distance. *Turkish Online Journal of Distance Education*, *23*(2), 120-139.

Uchida, S. (2024). Using early LLMs for corpus linguistics: Examining ChatGPT's potential and limitations. *Applied Corpus Linguistics*, *4*(1). https://doi.org/10.1016/j.acorp.2024.100089

Van Hee, C., Lefever, E. i Hoste, V. (2018a). We usually don't like going to the dentist: Using common sense to detect irony on Twitter. *Computational Linguistics*, *44*(4), 793-832.

Van Hee, C., Lefever, E. i Hoste, V. (2018b). Semeval-2018 task 3: Irony detection in English tweets. Dins: *Proceedings of The 12th International Workshop on Semantic Evaluation*, p. 39-50.

Wei, P., Wang, X. i Dong, H. (2023). The impact of automated writing evaluation on second language writing skills of Chinese EFL learners: a randomized controlled trial. *Front. Psychol.*, *14*, 1249991. DOI: 10.3389/fpsyg.2023.1249991

Xiang, R., Gao, X., Long, Y., Li, A., Chersoni, E., Lu, Q. i Huang, C. R. (2020). Ciron: a new benchmark dataset for Chinese irony detection. Dins: *Proceedings of the 12th Language Resources and Evaluation Conference*, p. 5714-5720, Marsella. European Language Resources Association.

Yu, D., Li, L., Su, H. i Fuoli, M. (2024). Assessing the potential of LLM-assisted annotation for corpus-based pragmatics and discourse analysis. The case of apology. *International Journal of Corpus Linguistics*, *29*(4), 534-561. https://doi.org/10.1075/ijcl.23087.yu

NORMES PER ALS COL·LABORADORS

https://bit.ly/3DKFESh

EXTENSIÓ

Les propostes de cada quadern no poden excedir **l'extensió de 50 pàgines (en Word)**, uns 105.000 caràcters, amb espais, referències, quadres, gràfiques i notes inclosos.

PRESENTACIÓ D'ORIGINALS

Els textos han de mostrar, en format electrònic, un **resum** d'unes deu línies i tres paraules clau, no incloses en el títol. Així mateix, han de contenir el **títol**, un ***abstract*** i tres ***keywords*** en anglès.

Per a les **formes de citar i les referències bibliogràfiques**, cal remetre's a les utilitzades en aquest *Quadern*.

AVALUACIÓ

L'acceptació d'originals es regeix pel **sistema d'avaluació externa per experts**.

Els originals són llegits, en primer lloc, pel **Consell de Redacció**, que valora l'adequació del text a les línies i objectius dels *Quaderns* i si compleix els requisits formals i els mínims de contingut científic exigit.

Els originals són sotmesos, en segon lloc, a **l'avaluació de dos experts**, especialistes en la temàtica de la qual tracta l'original i l'àmbit disciplinari corresponent. Els autors reben els comentaris i suggeriments dels avaluadors i la valoració final amb les esmenes i canvis que cal fer, si és el cas, abans de ser acceptat per a la seva publicació.

Si els canvis exigits són significatius o afecten bona part del text, el nou original és sotmès a l'avaluació de dos experts externs i d'un membre del Consell de Redacció. El procés es duu a terme com a «doble cec».

REVISORS

https://bit.ly/3oF4izw

L'Institut de Desenvolupamet Professional (IDP / ICE) de la Universitat de Barcelona inicià fa uns anys la publicació dels **QUADERNS DE DOCÈNCIA UNIVERSITÀRIA** amb l'objectiu de posar a l'abast del professorat universitari documents i materials de treball referits a temes relacionats amb la docència superior que facilitessin la seva formació, l'intercanvi d'experiències i la difusió de «bones pràctiques» docents. Amb aquests *Quaderns* pretenem estar atents als temes nous i emergents en l'actual conjuntura universitària, per tal de donar a conèixer i difondre iniciatives innovadores en el camp de la docència universitària que responguin a les línies següents:

- Propostes de marcs de referència rigorosos i generals que ajudin a clarificar conceptes clau.
- Estratègies docents i bones pràctiques de planificació, metodologia i avaluació de l'ensenyament i aprenentatge desenvolupades en contextos acadèmics específics i diversos.
- Tècniques i tàctiques, amb un marcat caràcter didàctic, presentades en materials i propostes concretes de treball i reflexió sobre la pràctica d'equips docents disciplinaris o interdisciplinaris.

IDP
Institut de Desenvolupament Professional
UNIVERSITAT DE BARCELONA