

# Quines faltes d'ortografia són més habituals entre l'alumnat de Llengua Catalana?

A Dictapp aquest trimestre hem passat del milió de faltes analitzades i hem pensat que era una quantitat prou representativa per tornar a fer una còmput de les dades recollides i compartir-les amb qualsevol persona interessada per l'aprenentatge de la llengua. El professorat hi trobarà dades per comparar amb les seves percepcions a l'aula i els no docents hi trobaran una visió de conjunt de per on falla més actualment l'alumnat de Llengua Catalana.

En aquest article només pretenem presentar les dades completes i aclarir alguns detalls sobre com funciona la plataforma i com les hem recollit per tal d'ajudar a entendre-les. I en tot cas, fer alguna observació a primer cop d'ull, però en cap cas trobar explicacions profundes.

Per tal d'entendre la primera taula (i les següents), la primera cosa que cal aclarir per a qui no conegui Dictapp és la classificació de faltes amb què treballem. La nostra és una plataforma creada per a treballar l'ortografia de manera personalitzada i poder-li dir a cada usuari/ària en què falla més i en què falla menys. És per això que el nostre motor de correcció detecta i identifica cada falta d'ortografia comesa i la classifica segons quinze categories.

Seguint aquestes categories, en aquesta primera taula hi teniu enumerat quin tant per cent faltes correspon a cada tipus de falta sobre el total de les 1.082.651:

Tipus de faltes més comeses	
Accentuació general	20,4%
Accents oberts/tancats	8,9%
Vocal neutra	8,3%
La B i la V	8,3%
Accents diacrítics	6,7%
Majúscules	6,5%
La G i la J	6,5%
La H	6,1%
Apòstrof i guionet	6,0%
Essa sonora i essa sorda	5,9%
Calaix de sastre	4,2%
O/U àtones	3,9%
Consonats finals	3,2%
La dièresi	3,2%
Juntes o separades?	1,9%

Un parell d'aclariments sobre les categories triades. *Calaix de sastre* és una categoria que recull tot de faltes (confusió entre q/c, i/y o r/rr...) que no tenen prou entitat per constituir una categoria pròpia. I *Juntes o separades?* inclou qüestions com perquè/per què, des de, etc.)

Com veieu, les faltes relacionades amb l'accentuació representen més d'una de cada tres faltes comeses.

## Comparació dels blocs dialectals

Tenim dues versions diferents de les locucions dels dictats a disposició dels centres, una per al català occidental i una per al català oriental. D'aquesta manera hem pogut separar les dades d'un i altre bloc dialectals per tal de poder-les comparar. En el bloc oriental s'hi inclouen la part oriental del Principat i les Balears. Al bloc occidental no s'hi inclou el País Valencià perquè encara no en tenim dades significatives.

BLOC OCCIDENTAL		BLOC ORIENTAL	
Accentuació general	19,23%	Accentuació general	21,08%
La B i la V	9,61%	Vocal neutra	9,90%
Accents oberts/tancats	8,48%	Accents oberts/tancats	9,06%
Accents diacrítics	7,11%	La B i la V	7,58%
Majúscules	7,06%	Accents diacrítics	6,51%
La G i la J	6,96%	La G i la J	6,22%
La H	6,78%	Majúscules	6,21%
Apòstrof i guionet	5,85%	Essa sonora i essa sorda	6,05%
Essa sonora i essa sorda	5,74%	Apòstrof i guionet	6,02%
Vocal neutra	5,49%	La H	5,67%
Calaix de sastre	4,79%	O/U àtones	4,47%
Consonants finals	4,20%	Calaix de sastre	3,88%
La dièresi	3,17%	La dièresi	3,17%
O/U àtones	2,79%	Consonants finals	2,70%
Juntes o separades?	2,73%	Juntes o separades?	1,49%
<b>Total de faltes analitzades</b>	<b>385.688</b>	<b>Total de faltes analitzades</b>	<b>696.963</b>

Com és d'esperar, si comparem percentatges la diferència més gran es troba en allò tocant a la neutralització vocàlica. La categoria de la vocal neutra, que en el bloc occidental passa a la desena posició de la llista i quasi té la meitat d'incidència, i la categoria de la O/U àtones, que també experimenta un descens en la part occidental, per bé que no tan acusat.

Malgrat això, cal dir també que en el bloc occidental hi ha algunes faltes particulars en vocals neutres a final de paraula que no trobem al bloc oriental, com per exemple el nom de la mateixa capital del Segrià escrit com *Lleide*.

## Influències entre llengües

Lògicament s'observen grafies errònies fruit de la contaminació (de vegades ortogràfica, però també lèxica o gramatical) entre català i castellà, tant en un sentit com en l'altre. En català veiem casos típics com accentuar erròniament paraules planes com si fossin *hiatos* castellans (tenía, coneixía, havia...).

En la versió castellana de Dictapp (utilitzada al Principat i les Balears) hi trobem també errades provocades per la influència del català. Aquests són només alguns dels casos recollits: d'arrel ortogràfica, en trobem casos com *pressión* per *presión*, *leies* per *leyes*, *interferències* per *interferències* o *hablar-le* per *hablarle*; d'arrel més lèxica, *obrió* per *abrió*, *acceptación* per *aceptación*, *judjarte* per *juzgarle*, *síntesi* per *síntesis* o *augmentan* per *aumentan*.

## Per zones geogràfiques

BARCELONA		GIRONA		LLEIDA	
ACCENT. GENERAL	23,3%	ACCENT. GENERAL	17,8%	ACCENT. GENERAL	19,9%
ACC OBERTS/TANCATS	9,6%	G / J / TG / TJ	12,5%	LA B I LA V	9,8%
LA VOCAL NEUTRA	9,6%	LA B I LA V	10,5%	ACC OBERTS/TANCATS	8,9%
ACCENTS DIACRÍTICS	7,5%	LA VOCAL NEUTRA	9,4%	ACCENTS DIACRÍTICS	7,2%
LA B I LA V	6,7%	S/C/Ç/SS/SC	7,6%	LA LLETRA H	7,1%
APÒSTROF i GUIONET	6,4%	ACC OBERTS/TANCATS	7,3%	LES MAJÚSCULES	6,9%
LES MAJÚSCULES	6,3%	LES MAJÚSCULES	6,0%	G / J / TG / TJ	6,6%
LA LLETRA H	5,6%	ACCENTS DIACRÍTICS	5,2%	APÒSTROF i GUIONET	6,0%
O/U ÀTONES	5,2%	LA LLETRA H	4,7%	LA VOCAL NEUTRA	5,8%
S/C/Ç/SS/SC	4,7%	APÒSTROF i GUIONET	4,3%	S/C/Ç/SS/SC	5,0%
G / J / TG / TJ	4,1%	CALAIX DE SASTRE	3,9%	CALAIX DE SASTRE	4,5%
CALAIX DE SASTRE	3,9%	LA DIÈRESI	3,8%	CONSONANTS FINALS	3,9%
LA DIÈRESI	3,1%	CONSONANTS FINALS	2,6%	O/U ÀTONES	3,1%
CONSONANTS FINALS	2,8%	O/U ÀTONES	2,3%	LA DIÈRESI	2,8%
JUNTES O SEPARADES?	1,2%	JUNTES O SEPARADES?	2,1%	JUNTES O SEPARADES?	2,7%

TARRAGONA		BALEARS	
ACCENT. GENERAL	35,3%	ACCENT. GENERAL	18,0%
LES MAJÚSCULES	15,6%	LES MAJÚSCULES	9,2%
CONSONANTS FINALS	7,0%	LA LLETRA H	8,5%
ACC OBERTS/TANCATS	5,6%	ACC OBERTS/TANCATS	8,3%
LA VOCAL NEUTRA	5,1%	LA B I LA V	7,6%
LA LLETRA H	4,5%	CALAIX DE SASTRE	7,1%
LA B I LA V	4,2%	LA VOCAL NEUTRA	6,3%
LA DIÈRESI	3,7%	G / J / TG / TJ	6,3%
CALAIX DE SASTRE	3,4%	ACCENTS DIACRÍTICS	5,8%
ACCENTS DIACRÍTICS	3,3%	APÒSTROF i GUIONET	5,0%
APÒSTROF i GUIONET	3,3%	CONSONANTS FINALS	4,5%
S/C/Ç/SS/SC	3,0%	S/C/Ç/SS/SC	4,3%
O/U ÀTONES	2,1%	O/U ÀTONES	3,9%
G / J / TG / TJ	1,9%	JUNTES O SEPARADES?	3,4%
JUNTES O SEPARADES?	1,9%	LA DIÈRESI	1,8%

Els alumnes de diferents zones del domini lingüístic català s'equivoquen en coses diferents? Sembla que no a grans trets però sí en algunes coses concretes, que poden ser consistents amb les serves variants dialectals, com per exemple la neutralització vocàlica, com ja hem vist

abans. A banda d'això, destaca el percentatge de faltes que tenen a veure amb la representació gràfica dels sons postalveolars (*g, j, etc.*) que hi ha a les terres gironines, que potser pot tenir a veure amb la diferent pronúncia que es fa en els parlars gironins en alguns casos. En els mateixos parlars, potser algú podria esperar més errades amb els accents oberts i tancats, però no és així. Pot sorprendre també la gran incidència que té la categoria Accentuació general a Tarragona, amb un percentatge força més elevat que a la resta malgrat que a totes les zones està en primera posició.

## Etapas educatives

La comparació de les diferents etapes no descobreix grans diferències, per bé que sí que hi ha alguns tipus de faltes que canvien significativament la seva posició a la llista i el seu percentatge d'incidència. Passa amb categories com La lletra H, Accents diacrítics o La vocal neutra. Cal recordar que el valor numèric no és absolut, sinó que correspon al percentatge d'aquella categoria sobre el total de faltes. Per exemple, si mirem el percentatge d'incidència de la categoria Accentuació general, al Batxillerat és bastant més alt. Això ens podria fer pensar que els alumnes de Batxillerat fan més faltes d'accentuació que els d'ESO, però potser la resposta és que fan menys faltes de les altres categories i per això puja el percentatge. I aleshores una possible explicació seria que han solucionat altres faltes però no encara l'accentuació. En aquest cas el lector pot està pensant que la dada útil seria la mitjana de faltes que fan, per veure si efectivament han baixat. Però aleshores cal recordar també que Dictapp té 5 nivells diferents de dificultat i és plausible pensar que els cursos de Primària (de 4t a 6è), ESO i Batxillerat treballin majoritàriament en nivells diferents.

PRIMÀRIA		ESO		BATXILLERAT	
POSAR ACCENTS	22,5%	POSAR ACCENTS	23,2%	POSAR ACCENTS	35,7%
LES MAJÚSCULES	9,3%	ACCENTS OBERTS/TANCATS	9,8%	ACCENTS OBERTS/TANCATS	12,5%
LA B I LA V	9,3%	LA VOCAL NEUTRA	9,4%	ACCENTS DIACRÍTICS	10,7%
ACCENTS OBERTS/TANCATS	8,1%	ACCENTS DIACRÍTICS	7,6%	LA B I LA V	6,9%
LA LLETRA H	7,2%	LA B I LA V	7,5%	LES MAJÚSCULES	5,1%
LA VOCAL NEUTRA	6,1%	LES MAJÚSCULES	6,5%	S/C/Ç/SS/SC	4,9%
APÒSTROF i GUIONET	5,8%	LA LLETRA H	5,8%	APÒSTROF i GUIONET	4,4%
G / J / TG / TJ	5,5%	G / J / TG / TJ	5,8%	LA VOCAL NEUTRA	4,1%
ACCENTS DIACRÍTICS	5,1%	APÒSTROF i GUIONET	5,6%	CALAIX DE SASTRE	3,9%
S/C/Ç/SS/SC	5,1%	S/C/Ç/SS/SC	4,1%	CONSONANTS FINALS	3,1%
CALAIX DE SASTRE	4,4%	CALAIX DE SASTRE	3,8%	G / J / TG / TJ	2,8%
CONSONANTS FINALS	3,9%	O/U ÀTONES	3,5%	LA LLETRA H	1,7%
O/U ÀTONES	3,1%	CONSONANTS FINALS	3,2%	JUNTES O SEPARADES?	1,6%
JUNTES O SEPARADES?	2,6%	LA DIÈRESI	2,9%	LA DIÈRESI	1,4%
LA DIÈRESI	2,1%	JUNTES O SEPARADES?	1,5%	O/U ÀTONES	1,1%

## L'aventura de predir

Si voleu posar-vos a la pell d'un docent quan corregeix textos, aquest és el vostre apartat. El motor de correcció de Dictapp funciona predictivament: intenta predir totes les possibles maneres en què es podria escriure una paraula per poder-ne categoritzar les errades, sigui en una de les 15 categories d'errades ortogràfiques o sigui com a errada no ortogràfica. Això pot semblar que són faves comptades ("en aquesta paraula podrien deixar-se l'accent, posar una s en lloc d'una c... i para de comptar"). Però els docents saben que de vegades la llista de les diferents maneres com un alumne pot escriure una paraula pot ser molt més llarga del que hom pot imaginar. De fet, en alguns casos la llista de possibles escriptures d'una mateixa paraula pot ajudar els filòlegs a entendre el concepte matemàtic de *tendir a infinit*. Mireu, si no, aquests dos casos: les 56 maneres diferents com el motor de correcció ha registrat una paraula de 4 lletres com *això* o les 136 de *barrejar* (ordenades alfabèticament):

acho	aio	aíxo	aixor	axo	ello	hajo
aicho	aisxo	aìxo	aixos	axó	eso	haxo
aigo	aixà`	áixo	ajo	axò	Exo	heixò
aiguxu	aixio	àixo	ajò	eijo	Exó	hexo
aiho	aixió	aixoa	ashó	eijò	Exò	hexò
aijo	aixo	aixóa	axio	eixo	haigo	iaixo
aijó	AIXO	aixòa	axió	eixò	haixo	sho
aijò	aixó	aixòn	axiò	ejo	haixò	so

avarregar	baregár	baretjar	barreijar	barretllar	beregar	varega	varrejar
bagreixar	baregiar	barjeijar	barreixar	barretxar	bereigar	varegar	varrellar
bairregar	bareguar,	barraga	barreixat	barrexar	bereijar	vareicha	varresar
baraga	bareigar	barraixar	barreja	barrexiar	bereixar	vareigar	varretjar
baragar	bareijant	barraja	barrejà	barrezar	berejar	vareijá	varrexar
baraia	bareixar	barrajà	barrejàr	barrgar	berrajar	vareijar	veirrajar
baraigar	barejà	barrajar	bárrejar	barrjar	berregar	vareixar	veragar
baraixa	barejar	barralla	bàrrejar	barrjear	berreigar	varejar	veraija
baraja	barejas,	barrallar	barrejat	barrregar	berreixar	varexar	verajar
barajar	barejat	barrasha	barrejer	barrreja	berrejà	variga	veregar
barajat	barejjar	barraxa	barrejr	barrrejar	berrejar	varragar	vereigar

barallar	barejjar	barreejar	barrelar	beraga	braxejar	varraixar	verejar
baratjar	barellar	barregar	barrellar	beragar	darrejar	varrajar	verèjar
baraxa	barellàr	barreges	barrenar	beraiga	varaga	varregar	verregar
baraxar	barellat	barreguer	barretgar	beraixar	varagar	varreigar	verrejar
baregar	baretga	barreigar	barretjar	beraja	varaje	varreijar	verrella
BAREGAR	baretgar	barreiguar	barretjat	berajar	varalla	varreixar	vreja

Abans de fer-ne qualsevol anàlisi, cal tenir present que això és un recull de faltes comeses per alumnat de totes les edats i amb tot tipus de casuística pel que fa a possibles dificultats d'aprenentatge i tot tipus de nivell de coneixement del català. Per tal de ser estrictes en la protecció de dades, registrem les paraules de manera anònima, així que no podem distingir etapes educatives, ni edats, ni nivells de dificultat ni zones, la qual cosa en alguns casos segurament ens donaria claus per explicar algunes escriptures.

Si mirem detingudament les llistes, amb l'ànim de fer una primera aproximació i no pas una anàlisi a fons ni encara menys establir explicacions, hi veurem tipologies d'errades diferents. D'entrada les més esperables, com ara bé treure o afegir un accent o posar-lo al revés (això, això / barrejàr), confondre l'escriptura d'una vocal neutra (*eixò / barrajar, berrejar, berrajar*) o equivocar-se en grafies "sospitoses habituals" com *ix (ajò, axò)* o *j (barreixar, barregar, barretjar)*. D'altres amb solucions més allunyades d'allò esperable (*aigo, aisxo / barreiguar, barretllar*). També algunes que denoten contaminacions d'altres llengües (*aicho, acho, ashó / vareicha, barrasha*). D'altres que proposen solucions més impensades i que no tenen cap explicació fonètica raonable (*aixóa, aixà, aixòn / varaje, braxejar*) i que en alguns casos potser no són sinó errades de picatge. Dins d'aquestes solucions fonèticament inviables, hi ha un subgrup interessant que són els accents posats en vocals que no són la tònica (*àixo, aixó, això / bårrejar*). De vegades també hi ha solucions gràfiques que semblen molt allunyades però que fonèticament encaixen (*aixor, heixò / barrejà, varrajar*).

Per acabar, una curiositat per a aquelles persones interessades en el funcionament intern del llenguatge. Una altra manera com alguns usuaris van escriure *barrejar és remenar*. L'àudio del dictat diu la paraula *barrejar*, però un cop processada pel cervell de l'usuari/a es va transformar en el sinònim *remenar*. Aquest és un fenomen que es repeteix en molts altres casos.

Sens dubte, una anàlisi més profunda podria trobar més conclusions i possibles explicacions. Aquestes dades i les que anem recollint tenen molt camp d'anàlisi per recórrer per davant i creuat amb món universitari. I a més en el futur es creuaran amb dades similars de gramàtica, etc.

### Com hem recollit aquestes dades? Aclariments metodològics

La nostra és una plataforma creada per a treballar l'ortografia de manera personalitzada a partir de dictats i activitats. El nostre motor de correcció detecta i identifica cada falta d'ortografia comesa i la classifica segons quinze categories.

No guardem cap dada personal, només guardem els totals de cada tipus de falta classificats per zones i per etapes educatives. No podem saber qui ha fet cada falta, ni tan sols de quina escola és. El que sabem és a quina etapa està assignat i a quina zona està el centre.

Accentuació general	G / J / TG / TJ	La lletra H
Accents oberts/tancats	La B i la V	Consonants finals
Accents diacrítics	S/C/Ç/SS/SC...	Juntes o separades?
Apòstrof i guionet	La dièresi	Les majúscules
La vocal neutra	O/U àtones	Calaix de sastre (un compendi d'altres faltes)

**Xavier Breil i Miret**

Director general de Dictapp

Gener de 2022