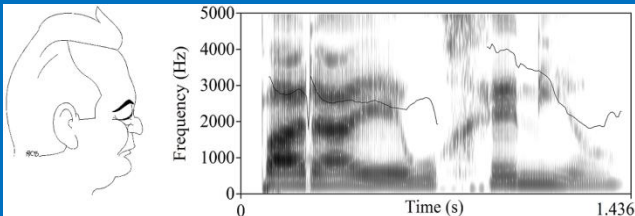


L'oralitat a la llengua escrita informal

M. Antònia Martí

Mariona Taulé

Irene Canudas



Fernández Planas, A. Ma. (ed.) (2016): *53 reflexiones sobre aspectos de la fonética y otros temas de lingüística*, Barcelona, págs. 483-491.

ISBN: 978-84-608-9830-6.

L'oralitat a la llengua escrita informal

M. Antònia Martí
Universitat de Barcelona
amarti@ub.edu

Mariona Taulé
Universitat de Barcelona
mtaule@ub.edu

Irene Canudas
Universitat de Barcelona
irenecanudas@gmail.com

Per a l'Eugenio, company, mestre i amic durant molts anys en el Departament de Lingüística General :-):=@.

1. INTRODUCCIÓ

Els textos que es produeixen en els nous mitjans de comunicació digital –xats, blogs, microblogs, fòrums, *reviews* i SMSs, entre d'altres– constitueixen un nou registre escrit que es caracteritza pel seu estil informal, no estàndard, que comparteix molts trets amb el registre oral col·loquial: l'escriptura és espontània, produïda en general de manera ràpida y sense un procés d'edició (Martí i Castellón, 1999; Crystal, 2006). Com a conseqüència, els textos que en resulten contenen errors ortogràfics i tipogràfics i una relaxació de les regles estàndards d'escriptura, especialment en l'ús dels signes de puntuació.

Aquests entorns de comunicació digital afavoreixen l'expressió d'estats interns, és a dir, l'expressió d'emocions, d'opinions, entre d'altres motius, perquè permeten interactuar de manera molt directa i personal i perquè permeten, també, la participació dels seus usuaris en espais de comunicació tradicionals, com la ràdio i la televisió per posar de manifest les seves idees i opinions.

Per tal d'aproximar els usuaris a una situació comunicativa propera a la interacció presencial, «cara-a-cara», els textos generats en aquests entorns contenen informació pragmàtica per copsar la modalitat, l'emoció, els sentiments i les actituds personals. Això explica la ràpida acceptació que ha tingut en Processament del Llenguatge Natural (PLN) el desenvolupament d'una línia d'investigació centrada en l'anàlisi de la subjectivitat y la detecció automàtica de la polaritat, és a dir, l'opinió positiva o negativa de productes i serveis a Internet.

Aquesta informació s'expressa sovint mitjançant marques paratextuals -com ara les emoticones- o bé mitjançant un ús creatiu de la llengua escrita, que es manifesta en un ús no convencional dels signes gràfics- les majúscules, els espais en blanc, els signes de puntuació-, les repeticions de determinats caràcters i l'ús d'escriptura fonètica, entre



d'altres. Com a resultat, aquest tipus de textos plantegen reptes de difícil solució per a les eines i aplicacions del PLN, que normalment s'han desenvolupat per a l'anàlisi automàtica de textos escrits de la llengua estàndard i formal. Noves eines i noves estratègies d'anàlisi constitueixen el focus d'atenció dels investigadors de l'àrea.

Paral·lelament, des d'una perspectiva lingüística i comunicativa, aquests textos constitueixen una font d'informació rica per a l'estudi del llenguatge en nous contextos d'ús que, ateses les seves característiques, ens permeten apropar-nos de manera indirecta a la llengua oral espontània (Torres, 1999; Mayans, 2002; Llisterrí, 2002; Pons, 2002; Blanco, 2002; Kobus et al, 2008).

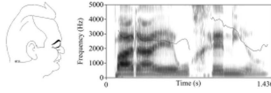
En conseqüència, actualment hi ha un interès creixent en l'anàlisi de textos escrits informals, i els corpus anotats amb aquestes característiques etiquetades de manera explícita han esdevingut recursos d'informació lingüística fonamentals. Actualment es disposa de molts pocs corpus anotats amb informació sobre els recursos expressius emprats a la llengua escrita informal; no existeix ni tan sols un conjunt comunament acceptat de criteris i etiquetes per a la seva anotació. Amb tot, han començat a aparèixer els primers corpus anotats com són el *English Web Treebank* (Bies et al, 2012), el *French Social Media Bank* (Seddah et al, 2012), el *No-Sta-D* de l'alemany i el *Latin American Spanish Discussion Forum Corpus* (Taulé et al, 2015) per a l'espanyol.

Aquest article és una reflexió feta per les autores sobre les característiques dels textos produïts en aquests entorns, a partir de l'anàlisi i anotació del *Latin American Spanish Discussion Forum Corpus*. El material anotat és un corpus de 50.291 paraules, que corresponen a 2,846 oracions, extret de fòrums de discussió informal, de temàtica variada i, en la seva majoria, escrit per hispanoamericans.

2. CARACTERÍSTIQUES GENERALS DE LA LLENGUA ESCRITA INFORMAL

Un dels factors que incideix de manera molt directa en la modalitat escrita informal és el condicionant de la immediatesa i l'espontaneïtat en el mode escrit. Una part important de les intervencions dels usuaris dels mitjans de comunicació social es dona en el marc del que s'anomena comunicació síncrona, és a dir, aquella en què la comunicació entre els interlocutors és immediata i on els sentiments i les emocions poden aparèixer sense filtres: és el cas dels xats, els wathsaps, els tweets, els sms i fins i tot podríem incloure-hi el correu electrònic i els blogs. La informalitat, l'espontaneïtat i la falta de revisió que caracteritza aquests textos són, en part, conseqüència d'aquesta immediatesa. Aquestes característiques, que tradicionalment s'han emprat per distingir la modalitat escrita de l'oral, deixen de tenir valor distintiu ja que són compartides per ambdues modalitats.

Una altra de les característiques de la comunicació digital és que els textos perden el seu caràcter lineal i esdevenen hipertextuals i multimodals, és a dir, no són necessàriament lineals i permeten la combinació de codi textual, visual i/o auditiu, amb la qual cosa, la frontera entre llengua escrita i oral perd també aquí el seu sentit.



En aquest escenari comunicatiu es genera una tensió, d'una banda, entre la necessitat d'interactuar de manera ràpida i eficient i, d'una altra, d'expressar el contingut emocional dels missatges de manera satisfactòria, que en la llengua escrita estàndard requereix una aportació important de recursos expressius. Aquesta tensió afecta de manera directa el codi emprat: la necessitat de respondre ràpidament afavoreix estratègies d'economia de recursos en el codi escrit (abreujaments, acrònims, escriptura fonètica, simplificació de l'ortografia, etc.), mentre que la necessitat de fer-se entendre i expressar emocions pot alentir el procés d'escriptura (repetició de caràcters, ús de negreta i majúscules, interjeccions, etc.). L'ús d'emoticones, sorgides arran d'aquest tipus d'escriptura, constitueixen una solució icònica a aquesta tensió: disposem d'emoticones per expressar tota mena de sentiments i emocions de manera ràpida i immediata. En general, es tracta de recursos fonamentats en l'ús creatiu dels signes emprats en l'escriptura que permeten així expressar el màxim amb el mínim de temps i espai.

Tot i que hi ha una gran llibertat en l'ús de recursos per reforçar l'expressivitat en els textos escrits, hi ha una sèrie de convencions que fan que els missatges siguin fàcilment decodificables. Es tracta de normes sorgides de l'ús, compartides per tots els usuaris, independents de la llengua, però generades en gran mesura des de l'anglès ja que va ser la primera llengua que les va emprar. Així «u2» ('you too') per a l'anglès, «desampara2» ('desamparados') per al castellà i «a10» (adéu') per al català són formes escrites basades en la mateixa estratègia en l'ús del codi.

A continuació presentem alguns dels recursos ortotipogràfics emprats per representar en la llengua escrita informal algunes de les característiques de la llengua oral, en especial aquells que tenen relació amb la forma fònica de les paraules, la prosòdia – la intensitat, la durada i el focus – i els recursos emprats per a l'expressió de les emocions.

3. EL CODI DE LA LLENGUA ESCRITA INFORMAL I LA SEVA RELACIÓ AMB L'ORALITAT

En aquest apartat ens centrem en aquells recursos de la llengua escrita que s'utilitzen per assolir els objectius de rapidesa i expressivitat propis de la llengua dels mitjans de comunicació digital i que tenen relació amb la fonètica de les paraules i la prosòdia dels enunciats.

3.1. Escriitura fonètica

L'escriitura fonètica (terme encunyat per Herring i Zelenkauskaitė, 2008) consisteix principalment en mecanismes de substitució o d'omissió que tenen com a objectiu reduir el nombre de caràcters i que afecten a diferents tipus d'unitats.

El mecanisme més simple consisteix en la no representació de certes convencions de la llengua escrita (ús de majúscules després de punt) i l'eliminació dels caràcters ortogràfics que no tenen correlat fònic (1), com per exemple la 'h' en català i castellà;



l'apòstrof, el guionet i el punt volat en català i els accents, entre altres. Tot i que aquestes simplificacions donen lloc a casos d'ambigüitat, el context permet desambiguar.

(1) Ola Eugenio, pots confirmarme q laula reservada es la 3.1!

Un altre mecanisme molt comú és la substitució d'un dígraf per un sol caràcter ortogràfic-fònic (2). És el cas de la substitució de «qu» per 'q' o 'k' en català i castellà i de la «ss» o «sc» per 's' en català. Podem incloure aquí també la substitució de caràcters amb diacrítics per les formes sense diacrític (p.e.: «n» per 'ñ'; «c» per 'ç'; «u» per 'ü', etc.), especialment quan les formes amb diacrític no apareixen al teclat com a primera opció.

(2) Q en penso? q es un estresat!!!

Altres vegades l'element que se simplifica és una síl·laba que se sol substituir per un caràcter numèric (3): «salva2» per 'salvados', aplicable a totes les paraules acabades en «-dos» en castellà.

Un pas més en aquesta direcció és la substitució de paraules completes per un signe específic manllevat moltes vegades dels símbols matemàtics (3): en anglès, la substitució de «and» per '&' o en català i castellà la substitució de «per» i «por» pel signe de la multiplicació 'x' o del «més» català i del «más» castellà pel signe '+', etc.

(3) Aproba2 x los pelos ...

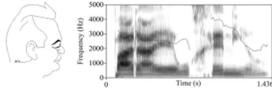
Una altra estratègia és l'omissió de síl·labes àtones dins d'una sola paraula o en expressions complexes, normalment molt freqüents (4): «nòstic be» per 'no estic bé'; «nem» per 'anem', «nanit» per 'bona nit', etc. Es pot donar també l'omissió de vocals ja que són fàcilment inferibles: «dmà» per 'demà', «tmpc» per 'tampoc' o «tb»/«tmb» per 'també'.

(4) Dma tho envio sensfalta, avui nòstic be :-(

De vegades, es combinen diferents mecanismes per donar lloc a expressions complexes: «&u2» per 'and you too' per a l'anglès o, en català, «xò» per 'per això', on la «x» té el doble valor de 'per' i 'x'.

L'ús d'acrònims¹ per a l'expressió de frases fetes d'ús molt freqüent és un recurs que redueix l'esforç en l'escriptura i li dona rapidesa. La llengua anglesa en fa un ús molt més abundant que el català o el castellà, tot i que, per contagi, en aquestes llengües s'està començant a utilitzar. En anglès tenim BTW ('by the way'), FYI ('for your interest'), OMG ('oh my God'), ASAP/asap ('as soon as possible'). En català i castellà s'opta més aviat per la reducció de la llargària de les paraules mitjançant l'omissió de vocals: «ptns» per 'petons'

¹ No incloem en aquest concepte els acrònims tradicionals com DNI, ONG, NY, USA, etc.



3.2. Escripura i prosòdia

Un dels paràmetres prosòdics de la veu, la intensitat, està vinculat en l'expressió d'estats emocionals. En la llengua oral, la intensitat determina de manera directa el volum, que es vincula a l'expressió de l'alegria, l'enuig, el dolor, estats eufòrics, etc. Així, els recursos de la llengua escrita per representar emocions i estats d'ànim relacionats amb la prosòdia tenen múltiples interpretacions i és el contingut del missatge el que ens donarà la clau interpretativa.

La intensitat es pot expressar en la forma escrita mitjançant diferents recursos (6): l'ús de majúscules, de signes d'admiració repetits, repetició de vocals, ús de negreta, marcatge de certes paraules amb asteriscos, etc. i combinacions de tots ells.

- (5) a. Estic **mooolt** cansaaaada!
- b. És un **PESAT!!!**
- c. Vine *****IMMEDIATAMENT*****
- d. Sóc **MOOOLT FELIÇ!!!!**

En la llengua oral, la intensitat es combina amb la durada i amb to per tal de marcar el focus del discurs, és a dir, allò que es vol emfasitzar. A la llengua escrita informal es representa de la mateixa manera que la intensitat:

- (6) a. **MATEMATIQUES**, es el qas destudiar
- b. *******matemàtiques******* es el qas destudiar

Una altra manera d'expressar l'èmfasi en la llengua oral és marcar les síl·labes separatament amb petites pauses. El seu correlat a la llengua escrita informal consisteix en separar les síl·labes amb guionet *alt* (7).

- (7) No-em-do-na-la-re-al-gana

4. EXPRESSIÓ DE LES EMOCIONS

En els seus inicis, els sistemes de comunicació social a Internet van desenvolupar mecanismes gràfics per a l'expressió d'emocions (Crystal, 2006) de manera ràpida i eficient mitjançant l'ús de combinacions de signes de puntuació i altres caràcters ASCII (*American Standard Code for Information Interchange*) com ara l'ampersand (&), el dòlar, l'arrova, etc. (8).

- (8) ;-) :-0 :-(:_(:-S :-/ :-@

Aquests sistemes de codificació han evolucionat i actualment es disposa ja d'emoticones amb una gran varietat d'opcions, no tan sols per representar l'estat d'ànim, sinó una gran varietat d'esdeveniments (aniversaris, naixements, casaments, celebracions, esports, etc.).



En alguns sistemes de comunicació social es disposa de mecanismes específics per contextualitzar comunicativament el missatge i indicar la intenció comunicativa. En el Twitter s'usen els *hashtags* (etiquetes precedides pel signe '#') per indicar aquestes contextualitzacions. Són habituals les etiquetes #ironia, #sarcasme i #broma, entre d'altres per facilitar la interpretació del missatge (9).

(9) Jamás diga una mentira que no se pueda probar #ironía

5. CONCLUSIONS

En aquest article hem tractat d'aproximar-nos a la naturalesa de la comunicació escrita duta a terme en entorns de comunicació digital. Ens trobem davant d'un espai on les categoritzacions tradicionals que defineixen la llengua (oral-escrita, formal-informal) i els diferents registres es difuminen. Aquests textos no segueixen la normativa de la llengua escrita i s'escriuen segons un codi fruit de la creativitat i imaginació dels participants. La comunicació digital escrita espontània, per tant, s'aproxima, com molts autors recalquen, a l'oralitat. Les característiques del codi són producte de la llibertat dels usuaris, que en fan un ús creatiu, obert a tota mena d'innovacions. Tot i això, els usuaris segueixen unes convencions tàcites que procedeixen fonamentalment de la llengua oral.

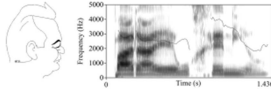
Estem parlant d'un codi que ha de facilitar l'escriptura per respondre al requeriment d'immediatesa que permeten els mitjans, que ha de ser expressiu per no perdre representativitat, i que ha de ser proper a la llengua oral perquè la comunicació sigui més efectiva. Aquestes tres premisses, rapidesa, expressivitat i oralitat, conformen tres eixos sobre els quals s'estructuren les desviacions intencionades respecte la llengua escrita estàndard.

AGRAÏMENTS

Aquest article ha estat possible gràcies a l'ajut del projecte TIN2012-38603-C02-02 i del projecte SGR-2014 632.

6. REFERÈNCIES BIBLIOGRÀFIQUES

- BIES, A.; J. MOTT; C. WARNER y S. KULICK (2012): *English Web Treebank*, Philadelphia, PA, Linguistic Data Consortium.
- BLANCO, M. J. (2002): «El chat: la conversación escrita», *Estudios de lingüística*, 16, pp. 1-90.
- CRYSTAL, D. (2006): *Language and the Internet*, Cambridge, Cambridge University Press.
- HERRING, S. C. y A. ZELENKAUSKAITE (2008): «Gender in Language: Classic Questions, New Contexts», dins J. F. Siegel, T. C. Nagel, A. Laurente-Lapole i J. Auger (eds.): *Gendered typography: abbreviation and insertion in Italian iTV SMS*, IULC Publications, pp. 73-92.
- KOBUS, C.; F. YVON y G. DAMNATI (2008): «Transcrire les sms comme on reconnaît la parole», *Actes de la Conference sur le Traitement Automatique des Langues (TALN'08)*, pp. 128-38.



- LLISTERRI, J. (2002): «Marcas fonéticas de la oralidad en la lengua de los chats: elisiones y epéntesis consonánticas», *Revista de Investigación Lingüística*, 5, 2, pp. 61-100.
- MARTÍ, M. A. i I. CASTELLÓN (1999): «Llengua i internet: text i pretext», *Homenatge a Jesús Tusón*. Empúries, pp. 210-214.
- MAYANS, J. (2002): «De la incorrección normativa en los chats», *Revista de Investigación Lingüística*, 5, 2, pp. 101-116.
- PONS, C. (2002): «Els xats: la ludoteca de la llengua», *Interlingüística*, 13, III, pp. 273-281.
- SEDDAH, D.; B. SAGOT; M. CANDITO; V. MOUILLERON i V. COMBET (2012): «The French Social Media Bank: A treebank of noisy user generated content», dins M. Kay i C. Boitet (eds.): *COLING 2012-24th International Conference on Computational Linguistics*, Mumbai. pp. 2441–2458.
- TAULÉ, M.; M. A. MARTÍ, A. BIES, A. GARÍ, M. NOFRE, Z. SONG, S. STRASSEL i J. ELLIS (2015): «Spanish Treebank Annotation of Informal Non-Standard Web Text», dins F. i O. Daniel Díaz (eds.): *ICWE-2015 Workshop, Lecture Notes in Computer Science*, Springer Verlag, pp. 3-14.
- TORRES, M. (1999). «Els xats: entre l'oralitat i l'escriptura», *Els Marges: revista de llengua i literatura*, 65, pp.113-126.