

Lexicometria e normalização

Dulce CARVALHO
(Universidade Aberta)

Não menos das figuras das letras nos manda Quintiliano ter muito cãrrego, porque elas são como instrumento, o qual, se for duvidoso, porá também em dúvida o efeito. E não imitemos os desvaios de tantas confusões, que assim lhe quero chamar, de letras, como se acostumam, mas sigamos uma certa regra de escrever, e a mais fácil.

Fernão de Oliveira¹

0. Introdução

O trabalho que apresentamos incide sobre a Crónica de D. Pedro e a Crónica de D. Fernando de Fernão Lopes. De entre as edições existentes, seleccionámos as duas únicas edições críticas de que dispomos, da responsabilidade de Giuliano Macchi. Ambas utilizaram o mesmo manuscrito de base que data dos fins do séc. XV, início do séc. XVI. Os textos situam-se no período histórico da língua portuguesa que José Joaquim Nunes² designa por período fonético e que José Leite de Vasconcelos e outros³ designam por período arcaico, cujo início remonta aos princípios do séc. XIII e o final à primeira metade do séc. XVI. O tempo de escrita (entre 1430 e 1450) não coincide com o tempo histórico de base (os acontecimentos narrados ocorreram durante os reinados de D. Pedro – 1357 a 1367 – e de seu filho D. Fernando – 1367 a 1383 – e os primeiros tempos de regência, por morte de D. Fernando, de sua esposa D. Leonor Teles).

Macchi fez um primeiro trabalho de regularização nas duas crónicas, mas a análise do índice alfabético destas crónicas mostra-nos que as variantes gráficas de uma mesma forma pululam. Tais “variantes puramente gráficas espelho das mutáveis modas em vigor no arco dos dois séculos e meio, cerca, ao longo dos quais foram copiados os numerosos testemunhos que constituem a tradição”⁴, não foram objecto de regularização. Vemo-nos assim confrontados perante diferentes representações gráficas de um mesmo item lexical, perante um número elevado de variantes, que temos de reduzir ao maior número possível de invariantes de modo a obtermos um *corpus* homogéneo do ponto de vista ortográfico e, só assim, susceptível de ser submetido a um estudo lexicométrico.

1. Constituição do *corpus*

As crónicas de D. Pedro e de D. Fernando foram introduzidas em computador. Constituímos em seguida o nosso *corpus* (conjunto representativo dos textos reunidos com o objectivo de serem comparados), eliminando as notas de rodapé e os títulos dos capítulos.

Para podermos fazer estudos comparativos, dividimos o *corpus* em quatro partes relativamente idênticas em extensão, isto é, em número de ocorrências, como se constata pelos resultados da segmentação automática das partes⁵:

*** Résultat de la segmentation par partie du fichier: Cts ***				
parties	occurrences	formes	hapax	fmax
1	34865	4342	2334	2455 e
2	38319	4264	2259	2732 e
3	39574	4799	2613	2588 e
4	4061	4347	2203	2739 e
*** Fin de la segmentation par partie du fichier: Cts ***				

A primeira parte é constituída pela Crónica de D. Pedro; a segunda pelo prólogo e pelos primeiros 58 capítulos da Crónica de D. Fernando, a terceira vai do capítulo 59 ao 112 e a quarta do capítulo 113 ao fim desta mesma crónica.

2. Normalização e Lexicometria

2.1. Normalização

O termo por nós utilizado, normalização, pressupõe a existência/criação de uma norma interna ao próprio *corpus* que leva à selecção da variante mais **normalmente** utilizada.

Consideramos na normalização duas vertentes. A primeira visa o tratamento da maiúscula, quer seja da maiúscula sistemática, quer seja da utilizada como forma de realce ou ainda da maiúscula do nome próprio. A outra vertente tem que ver com a constante variação gráfica detectada no interior de cada uma das duas crónicas. Trata-se, aqui, de uma normalização ortográfica que consiste em agrupar, sob uma mesma forma, as diferentes grafias do mesmo vocábulo. Criámos uma norma interna ao próprio *corpus*, a qual procurou ser a representação exacta e objectiva da ortografia nele dominante. Daí termos recorrido em primeiro lugar ao critério da frequência da variante da forma gráfica mais utilizada e não a uma ortografia etimológica nem a uma ortografia baseada na pretensa pronúncia da época. Também não se tratou de modernizar grafias antigas, mas tão somente de uniformizar grafias, evitando a "anarquia individualista"⁶ que caracteriza a época do português arcaico. Adoptámos uma ortografia regularizada em função do USO escrito no texto.

Não se tratou tão pouco de lematizar formas, dado que não fizemos qualquer agrupamento de variantes flexionais (reduzindo as diferentes formas de um mesmo verbo ao infinitivo, os plurais ao singular, os femininos ao masculino) nem procurámos desambiguar formas homógrafas, porque tal procedimento implicaria um mergulhar em todas as diferenças semânticas e "cela c'est s'ouvrir à toute la sémantique, à toutes les différences sémantiques"⁷. Ora tanto na literatura como na política ou em qualquer outro domínio (com excepção de alguns textos científicos em que as palavras são unívocas), os emissores jogam constantemente com as ambiguidades do sentido e o universo do contraditório e do "flou", do "indécidable"⁸ não pode ser sistematizado.

Situámo-nos exclusivamente a nível da superfície discursiva e dentro de um *corpus* bem determinado, pelo que a normalização feita é uma normalização de *corpus*.

Ao reduzirmos as variantes gráficas de um mesmo vocábulo a uma invariante textual pretendemos, respeitando a língua da época, tornar o nosso *corpus* homogéneo do ponto de vista ortográfico e susceptível de ser submetido a um estudo sistemático, a uma análise estatística. Se

não tivéssemos feito a normalização, teríamos problemas com o número de ocorrências da mesma forma, o que falsearia os resultados como demonstraremos com o método das especificidades, aplicado à Crónica de D. Pedro.

Resumindo, para normalizarmos o texto (o que só foi possível graças aos meios informáticos de que dispomos e aos programas de estatística lexical, "LEXICO 1", desenvolvidos por André Salem no "Laboratoire de Léxicométrie et Textes Politiques" do CNRS – InaLF) estabelecemos regras⁹ que assentam na distinção entre:

A – Pluralidade de referentes – Normalização da série: nome próprio (nomes de pessoa - "prénom"-) e todas as outras formas, quer sejam formas plenas (substantivos, adjectivos, verbos e alguns advérbios), quer sejam utensílios gramaticais (artigos, pronomes, preposições, conjunções, etc.).

B – Unicidade do referente – Normalização por caso (caso a caso): sobrenomes, topónimos, nomes de entidades e festividades religiosas, instituições.

Em **A**, a normalização foi feita tendo em vista

1º - a forma mais frequente ou, não sendo possível,

2º - a forma mais próxima da do português actual

Ao fazê-lo, pudemos constatar que havia dois tipos de variação na escrita: algumas formas funcionavam como variantes livres do mesmo vocábulo e outras como variantes condicionadas pelo contexto em que se inseriam. Entreviemos no primeiro caso, mantivemos a variação no segundo.

Em **B**, distinguimos os diferentes referentes e a normalização de cada um deles baseou-se, do mesmo modo, na frequência ou na variante mais próxima do português actual.

2.2. Lexicometrais textual – método das "especialidades"

Por lexicometria textual entendemos um conjunto de métodos que permitem operar, a partir de análises estatísticas, reorganizações formais do vocabulário de um *corpus* recolhido na sequência textual.

O método das "especificidades" julga em probabilidade a utilização de cada uma das formas e/ou segmentos em cada uma das partes em que se dividiu o *corpus*, partindo do princípio matemático da equiprobabilidade dos possíveis.

O computador ao avaliar cada sub-frequência¹⁰ de cada forma (ou de cada segmento) nas diferentes partes do *corpus*, classifica em probabilidade todas as formas e respectivas sub-frequências. Cada sub-frequência situa-se numa escala de probabilidade que vai do mais improvável ao mais provável. Basta apenas escolher, em seguida, um limiar (que varia em geral, em função das necessidades da pesquisa, entre 5% e 1%) que será o limite entre o provável e o improvável e abaixo ou acima do qual se escalonam os diversos empregos. Ao fazê-lo, três casos podem ocorrer:

- a) ou a forma analisada possui, na parte em estudo, uma sub-frequência muito superior à que seria de esperar (tendo em conta a totalidade das ocorrências do *corpus* bem como a totalidade das ocorrências na parte e a frequência absoluta da forma em todo o *corpus*) e então essa forma é considerada como **específica positiva (S+)**, uma vez que se caracteriza por um "sobre-emprego" determinado (o índice S+, ou simplesmente

- +, vem acompanhado de um coeficiente com dois algarismos que dão uma ideia do grau de especificidade da forma em questão);
- b) ou a forma analisada possui, na parte em estudo, uma frequência muito inferior à que seria de esperar e então é considerada específica negativa (**S-**), caracterizando-se por um "sub-emprego";
- c) ou ainda, quando a probabilidade da frequência do termo na parte é igual ou superior ao limiar estabelecido, a **forma** é considerada **banal (b)**, o seu emprego é banal. Quando a forma é banal em todas as partes em que se dividiu o corpus, então considera-se uma **forma de base**.

3. Normalização e Especificidades da *Crónica de D. Pedro*

3.1. Especificidades positivas (S+) antes e depois da normalização

Especificidades Positivas (S+)

Antes da Normalização

forma	F	f	esp
Purtugall	99	77	+E31
Aragom	234	119	+E21
galees	63	50	+E21
justiça	53	41	+E17
matar	92	57	+E16
Pedro	379	151	+E14
Sevilha	116	60	+E12
i	17	17	+E11
annos	32	24	+E10
tesouro	25	20	+E09
el	333	124	+E09
Enes	31	22	+E08
rrei	234	89	+E08
Branca	18	15	+E08
tio	36	23	+E07
seu	1118	329	+E07
mandou	283	102	+E07
virtude	19	15	+E07
Vermelho	13	12	+E07
Monvedro	10	10	+E07
Padilha	14	12	+E06
el-rrei	2752	725	+E06
des	51	27	+E06
h~ua	47	26	+E06
anno	34	20	+E06
Fradarique	21	15	+E06

Depois da Normalização

forma	F	f	esp
Aragom	234	119	+E21
justiça	53	41	+E17
matar	92	57	+E15
Pedro	476	177	+E13
Sevilha	117	60	+E11
mandou	331	129	+E11
tesouro	28	22	+E10
rrei	2990	816	+E09
Enes	31	22	+E09
Branca	18	15	+E08
seu	1121	331	+E07
senom	10	10	+E07
Vermelho	13	12	+E07
virtude	19	15	+E07
tio	36	23	+E07
Monvedro	10	10	+E07
entrou	39	22	+E06
Fradarique	21	15	+E06
Padilha	14	12	+E06
anos	84	38	+E06
semelhante	18	12	+E05
gaanhou	10	9	+E05
el	2767	726	+E05
rrecebera	12	10	+E05
tomou	53	26	+E05
tomar	92	38	+E05

Anrique	15	12	+E06
madre	44	22	+E05
mandou-ho	14	11	+E05
gaanhou	10	9	+E05
entrou	38	21	+E05
semelhante	18	12	+E05
galee	17	12	+E05
conselho	56	26	+E05
rrecebera	12	10	+E05
dizia	34	19	+E05
filhas	30	16	+E04
leis	19	11	+E04
tornou-sse	46	21	+E04
Tello	58	26	+E04
ali	18	11	+E04
tomar	89	36	+E04
queria	68	29	+E04
sabendo	29	15	+E04
necessaria	10	8	+E04
cuidando	34	18	+E04
tal	19	12	+E04
hordem	27	14	+E04
cabeça	13	9	+E04
rreinou	16	11	+E04
rrega	11	8	+E04
onde	178	61	+E04
sempre	98	39	+E04
amado	11	8	+E04
morte	131	47	+E04
paaço	23	13	+E04
achou	44	19	+E03
qual	12	7	+E03
Affonso	33	15	+E03
outro	192	59	+E03
seendo	113	39	+E03
nosso	32	15	+E03
dez	47	20	+E03
Albuquerque	14	9	+E03
Castro	55	21	+E03
rreino	331	94	+E03
tiinha	218	65	+E03
morto	43	18	+E03
alli	300	88	+E03
mar	69	25	+E03
dona	329	96	+E03

cuidando	34	18	+E04
alg~uua	54	23	+E04
cabeça	13	9	+E04
anno	64	28	+E04
paaço	23	13	+E04
hordem	32	17	+E04
madre	45	22	+E04
alli	318	99	+E04
morte	131	47	+E04
rreinou	16	11	+E04
rrega	11	8	+E04
sempre	98	39	+E04
filhas	30	16	+E04
amado	11	8	+E04
Tello	58	26	+E04
necessaria	10	8	+E04
queria	114	42	+E04
iffantes	15	9	+E03
memoria	11	7	+E03
rreal	15	8	+E03
sanha	23	12	+E03
trezentos	10	7	+E03
escripto	18	10	+E03
prougue	46	18	+E03
lhe	1438	367	+E03
dia	247	74	+E03
outro	192	59	+E03
Castro	55	21	+E03
nosso	33	15	+E03
achou	46	20	+E03
Maria	65	24	+E03
soube	80	30	+E03
Albuquerque	14	9	+E03
seendo	115	39	+E03
noveenta	12	7	+E03
mar	70	26	+E03
mandara	49	21	+E03
fiava	13	8	+E03
no	378	111	+E03
o	3152	781	+E03
dobras	48	19	+E03
sabendo	30	15	+E03
Nunez	24	12	+E03
aljofar	12	7	+E03
padre	109	39	+E03

3.2. Especificidades negativas (S-) antes e depois da normalização

Especificidades Negativas (S-)

Antes da Normalização

forma	F	f	esp
Ourem	20	0	-E03
minha	43	2	-E03
Cambrig	18	0	-E03
Andeiro	20	0	-E03
gallee	19	0	-E03
Beltram	39	2	-E03
senhor	235	35	-E03
soldos	38	2	-E03
fidallgos	122	15	-E03
meu	71	6	-E03
menag~ees	18	0	-E03
aaquella	19	0	-E03
avia	355	59	-E03
dessem	27	1	-E03
Elvas	18	0	-E03
papa	87	9	-E03
aviia	28	1	-E03
Rrodriguez	72	8	-E03
castellaãos	55	5	-E03
caso	26	0	-E03
Ellvas	25	0	-E03
sollido	42	2	-E03
dos	433	78	-E03
villas	63	5	-E03
certas	40	2	-E03
irmãa	38	2	-E03
os	1396	272	-E03
mim	72	6	-E03
Santarem	51	4	-E03
fezessem	49	4	-E03
leceña	23	0	-E03
Badalhouce	43	2	-E03
cerca	20	0	-E03
voz	73	8	-E03
rregnos	34	1	-E03
tall	319	54	-E03
arref~ees	30	0	-E04
hora	57	3	-E04

Depois da Normalização

forma	F	f	esp
Cambrig	18	0	-E03
aaquella	19	0	-E03
meu	71	6	-E03
capitollos	26	0	-E03
Badalhouce	43	2	-E03
el	18	0	-E03
Andeiro	20	0	-E03
Santarem	52	4	-E03
senhor	238	38	-E03
villas	63	5	-E03
llo	28	1	-E03
Beltram	39	2	-E03
sobre	232	38	-E03
dessem	27	1	-E03
Rrodriguez	76	8	-E03
fezessem	49	4	-E03
castellaãos	56	5	-E03
os	1440	281	-E03
mim	72	6	-E03
caso	26	0	-E03
paga	25	0	-E03
certas	40	2	-E03
papa	87	9	-E03
minha	43	2	-E03
Allancastro	18	0	-E03
prioll	47	3	-E03
Ourem	20	0	-E03
comarca	47	4	-E03
Gill	42	3	-E03
dos	433	78	-E03
cerca	20	0	-E03
defenssom	18	0	-E03
voz	73	8	-E03
menag~ees	19	0	-E03
arref~ees	31	0	-E04
senhora	43	1	-E04
lanças	44	1	-E04
avia	383	60	-E04

A observação destes quadros permite-nos verificar que antes da normalização há um total de 78 formas específicas negativas. O número de formas S-, após a normalização, fica reduzido a 65, portanto menos 13 formas. Só *capitollos*, *el*, *llo*, *sobre*, *paga*, *AllanCastro*, *prioll*, *comarca*, *Gill*, *defenssom* e *Nuno* é que não constam do vocabulário específico negativo, antes da normalização. 54 formas mantêm-se como específicas negativas antes e após a normalização.

Das 78 formas caracterizadas por um sub-emprego, antes da normalização, 24 deixam de constar como S- depois da normalização. Também antes da normalização coexistem como formas diferentes *avia* e *aviia*, *Elvas* e *Ellvas*, todas elas específicas negativas, tendo-se conservado como tal as formas normalizadas *avia* e *Ellvas*.

3.3. Especificidades de grafias, especificidades de vocabulário

Os casos observados tanto a nível das especificidades positivas: *anos*, *Affonso* e *alli*, como das especificidades negativas: *avia* e *Ellvas*, são simples, uma vez que estamos perante duas grafias que vão no mesmo sentido, o da especificidade positiva no primeiro caso, o da especificidade negativa, no segundo.

Mas que se passa quando as diferentes grafias não têm o mesmo estatuto quanto à especificidade na mesma parte do *corpus*?

Analisemos os quadros das formas S+ e das formas S-, antes da normalização na Crónica de D. Pedro. De imediato deparamos com as formas *Purtugall* (S+) e *Portugall* (S-); *galees* (S+), e *gallees* (S-); *galee* (S+) e *gallee* (S-); *conselho* (S+) e *consselho* (S-); *tal* (S+) e *tall* (S-). Estas variantes não têm o mesmo estatuto antes da normalização, caracterizando-se umas por um sobre-emprego e as outras, ao mesmo tempo, por um sub-emprego para o mesmo vocábulo. Tal situação repete-se ao longo de todo o *corpus* como se pode verificar pelo quadro síntese das variantes dos nomes "Portugal" e "conselho"

Variantes	Antes da Normalização					Depois da Normalização				
	F	1	2	3	4	F	1	2	3	4
Portugall	218	4 (-E13)	35 (-E04)	68 (b)	111 (+E14)	317	81 (b)	57 (-E03)	68 (b)	111 (+E04)
Purtugall	99	77 (+E31)	22 (b)	0 (-E13)	0 (-E14)					
Consselho	109	4 (-E08)	15 (-E03)	46 (+E04)	44 (+E03)	165	30 (b)	31 (b)	60 (+E03)	44 (b)
Conselho	56	26 (+E05)	16 (b)	14 (b)	0 (-E08)					

Se analisarmos agora o caso do nome "Portugal" verificamos que na Crónica de D. Pedro, parte 1, se misturam, *Purtugall* e *Portugall*. A variante *Purtugall* encabeça as especificidades positivas com uma frequência de 77 sobre um total de 99. É, digamos, ultra específica. Mas, ao mesmo tempo, o mesmo nome é altamente específico negativo: a variante *Portugall* ocorre apenas 4 vezes nesta parte quando a sua frequência total no *corpus* é de 218. Esperar-se-ia,

matematicamente que esta forma se encontrasse mais regularmente desseminalada pelo *corpus*. Acontece, porém, que ela encabeça de novo as especificidades negativas.

Constatamos também na análise feita sobre as formas normalizadas que o nome de "Portugal" deixa de ter qualquer relevância na Crónica de D. Pedro, é banal.

Que conclusão inferir? Na análise, antes da normalização a especificidade visa apenas uma das variantes gráficas sem ter em conta as outras. É a grafia, *Purtugall* que é específica positiva; é também a grafia *Portugall* que é específica negativa e não o nome de "Portugal". Inversamente, quando normalizamos as variantes, apercebemo-nos que este nome não era, na realidade, específico desta parte do *corpus*, mas que, aí, correspondia a um emprego banal. Assim, antes da normalização, as especificidades obtidas são especificidades de grafias e não especificidades de vocábulos. A especificidade do vocabulário do *corpus* só nos é fornecida após a normalização. Este exemplo demonstra que, ao normalizar, se evitou uma incorrecção grave a nível da especificidade, provocada exclusivamente por uma simples variante gráfica livre. A normalização evita que se emitam juízos de valor errados a nível do vocabulário específico, erros que provêm da especificidade de uma grafia em detrimento das outras variantes livres da mesma forma discursiva. Muitos outros exemplos podiam ser apontados, mas parece-nos que este constitui prova suficiente da importância da normalização.

A correcção deste erro, ao julgarmos as especificidades, visa também o conteúdo, mas na Crónica de D. Fernando. Com efeito, poderíamos pensar que, antes da normalização, as duas Crónicas se opunham no que concerne a designação do país "Portugal". Ora, após a normalização, constatamos efectivamente que as duas crónicas não se opõem uma vez que a designação do país é banal na Crónica de D. Pedro. É, no entanto, no interior da Crónica de D. Fernando que se passa algo à medida que os acontecimentos se desenrolam no eixo temporal. Aí deparamos com três momentos. No início, o coeficiente de especificidade da ordem do milésimo, denota que a designação é específica negativa, o seu emprego está abaixo do que seria de esperar. Na segunda parte, o seu uso passa a ser banal e é finalmente na terceira parte da crónica que o país ganha uma importância grande. Há assim uma progressão crescente no emprego da designação de Portugal com o desenrolar dos acontecimentos. Porquê? Tentará Fernão Lopes registar a irrupção de um verdadeiro sentimento nacional face à ameaça castelhana? Com efeito, uma forte especificidade positiva de *castellaños* (+E05) acompanha o sobre-emprego de "Portugal".

4. Conclusão

Para podermos trabalhar sobre um *corpus* homogéneo escrito (necessário em qualquer tipo de estudo científico e indispensável em lexicometria) do ponto de vista da escrita, tivemos que recorrer à lexicometria.

Para que a normalização fosse sistemática, houve necessidade de informatizar o texto e de recorrer a análises quantitativas. Os programas LEXICO forneceram-nos os índices alfabéticos e hierárquicos, as concordâncias, os contextos, os segmentos repetidos, sobre que trabalhamos para normalizar o texto. Forneceram-nos igualmente o "vocabulário específico", quer do *corpus* quer de cada uma das partes em que o dividimos. O método das especificidades provou a importância, da normalização das crónicas. Sem ela, jamais teríamos o "vocabulário específico" do nosso *corpus*. Apenas conseguiríamos obter "especificidades" de variações escritas. Em textos

medievais, e sempre que se pretenda trabalhar sobre superfícies textuais, lexicometria e normalização têm de caminhar lado a lado.

A lexicometria provou a necessidade da normalização. A normalização, por sua vez, não podia ter sido levada a bom termo sem a lexicometria.

Notas

¹ OLIVEIRA, Fernão (1536) – *Gramática da linguagem portuguesa*, Introdução, leitura actualizada e notas por Maria Leonor Carvalhão Buescu, Lisboa, Imprensa Nacional – Casa da Moeda, 1975, pág. 52.

² NUNES, José Joaquim (1960) – *Compêndio de gramática histórica portuguesa*, 6ª ed., Lisboa, Livraria Clássica Editora, pág. 192.

³ CASTRO, Ivo (1988) – *Sete ensaios sobre a obra de J. M. Piel*, Lisboa, Instituto de Linguística da Fac. de Letras, pág. 12.

⁴ MACCHI, Giuliano (1975) – “Introdução” à *Crónica de D. Fernando*, por Fernão Lopes, Lisboa, Imprensa Nacional – Casa da Moeda, pág. LXXXVI-LXXXVII.

⁵ **Número de ocorrências:** cada forma gráfica (palavra, *grosso modo*), constitui uma ocorrência; **número de formas:** duas sequências idênticas de caracteres não delimitadores constituem duas ocorrências da uma mesma forma; **número de hapax:** número de formas que têm a frequência um; **fmax:** frequência máxima e forma que a possui.

⁶ VAZQUEZ CUESTA P.; MENDES DA LUZ (1980) – *Gramática da língua portuguesa*, Lisboa, Edições 70, pág. 339.

⁷ TOURNIER, Maurice (1994) – *Stage de Lexicométrie*, notas pessoais.

⁸ *Ibidem*.

⁹ A apresentação pormenorizada e a exemplificação das regras aqui enunciadas pode ser consultada em CARVALHO, Dulce (1996) – *Normalização de corpus medieval: Crónicas de D. Pedro e de D. Fernando de Fernão Lopes*, Dissertação de mestrado em linguística, área de especialização em lexicologia e lexicografia, Universidade Nova de Lisboa, Faculdade de Ciências Sociais e Humanas.

¹⁰ Costuma designar-se por sub-frequência (f) a frequência da forma na parte e por frequência absoluta ou total (F) a frequência da forma na totalidade do *corpus*. Utilizaremos, de um modo geral, os termos de frequência na parte e de frequência no *corpus*.

¹¹ Colocámos o til imediatamente antes dos grafemas <<e>>, <<i>> e <<u>>, sempre que sobre eles incidia tal sinal de nasalização, uma vez que o computador apenas no-lo aceita sobre <<a>> e <<o>>.