

TRANSCRIÇÃO ORTOGRÁFICA DE TEXTOS ORAIS: problemas e perspectivas

Maria Celeste Ramilo e Tiago Freitas
Instituto de Linguística Teórica e Computacional

1. Introdução

Como investigadores responsáveis pela transcrição ortográfica no projecto REDIP, temos vindo a conviver directamente com os problemas inerentes à transcrição ortográfica de textos orais. Este projecto, que está a ser desenvolvido no ILTEC, em cooperação com o CLUL (Centro de Linguística da Universidade de Lisboa) e a Universidade Aberta, tem como objectivo estudar a linguagem dos meios de comunicação social portugueses. Está previsto um *corpus* com um total de trezentas e vinte e quatro mil palavras, dois terços do qual correspondem à língua oral.

Existem duas regras básicas no domínio da transcrição do oral, e estão enunciadas em FRENCH (1991):

- (i) Don't type anything that isn't there.
- (ii) Do include everything that is there.

Isto significa, por um lado, que o transcritor não deverá ser zeloso ao ponto de tentar resolver todas as lições, principalmente quando estas lhe são totalmente incompreensíveis. Por outro lado, não deverá subtrair elementos do texto, ainda que sejam geradores de confusão e dificultem a inteligibilidade. Deverá acautelar-se ao máximo no sentido de não impor sobre a transcrição quaisquer marcas pessoais, distinguindo o seu idiolecto daquilo que realmente está a ser dito.

É fácil compreender a boa vontade que normalmente acomete os transcritores e que é muitas vezes responsável pela má qualidade das transcrições: neste tipo de trabalho aparentemente mecânico, quanto maior for o número de lições resolvidas, maior será o sucesso (ou a ilusão de sucesso) do transcritor.

Não é, no entanto, da maior ou menor competência do transcritor que nos ocuparemos aqui. Falaremos sim de alguns dos problemas inerentes à prática da transcrição ortográfica e dos diferentes métodos usados para ultrapassá-los.

2. Problemas e perspectivas

Embora isso nem sempre seja claro para as pessoas que trabalham noutras áreas, existe uma série de questões relativas à transcrição ortográfica que, apesar de poderem parecer excessivamente minuciosas, assumem uma importância vital, uma vez que podem condicionar todo o trabalho de análise feito posteriormente. Na verdade, todas as decisões tomadas relativamente ao método de transcrição influem no resultado final do projecto e no seu posterior aproveitamento.

Nós vamos debruçar-nos, em primeiro lugar, sobre o problema das pausas e da pontuação. Esta questão tem levantado alguma controvérsia ao longo dos anos, sendo alvo de tratamentos muito diversos. Em seguida, consideraremos o problema da representação dos enunciados simultâneos. Como veremos, este fenómeno é sempre tratado de uma forma generalizada a não ser num dos sistemas estudados. Por fim, analisaremos a forma como os ideofones e as interjeições são representados nas transcrições, observando que também neste caso existem tratamentos generalizados e tratamentos específicos.

2.1. *As pausas e a pontuação*

Um dos aspectos mais marcantes da língua oral é, sem dúvida, o facto de as pausas não corresponderem de modo nenhum aos nossos hábitos de pontuação na escrita. Por isso, há muitos investigadores que pura e simplesmente abdicam da pontuação convencional, optando por sistemas de representação mais abstractos. Outros há, no entanto, que preferem pontuar os textos de uma maneira mais simples e intuitiva, conservando os símbolos usados na ortografia. Esta dualidade coloca muitas vezes os investigadores em posições antagónicas, como se pode observar nas seguintes citações, extraídas de BACELLAR DO NASCIMENTO (1987) e BLANCHE-BENVENISTE & JEANJEAN (1987), respectivamente:

- «Os textos não pontuados tornam-se-nos praticamente incompreensíveis;»
- «Les textes livrés sans ponctuation sont, moyennant une certaine accoutumance, assez faciles a lire.»

Na verdade, o favorecimento ou não do uso da pontuação depende sempre, em última análise, dos objectivos do projecto. Há certos casos em que a

pontuação convencional é seriamente desaconselhada, podendo mesmo tornar os textos inutilizáveis. Nos projectos especificamente destinados a análises prosódicas ou sintácticas, por exemplo, é natural que a pontuação seja posta de parte. Foi isso que aconteceu nas transcrições do GARS¹.

Nos projectos que não se destinam a utilizações tão específicas, contudo, a tendência é para adoptar a pontuação convencional, ainda que com algumas restrições. Não conhecemos, com efeito, nenhum projecto em que sejam usados todos os símbolos sancionados pela ortografia. Pelo contrário, em alguns, apenas é usada uma pequena parte desse conjunto de símbolos. É o que acontece no NURC²:

Frases interrogativas	Qualquer pausa
?	...

Este método de representação é o mais simples de todos aqueles que aqui contemplamos. Trata-se de um sistema que permite uma grande uniformidade gráfica, o que de certo modo o aproxima do contínuo sonoro da fala. No entanto, não podemos deixar de notar que tem algumas desvantagens ao nível da legibilidade, uma vez que não nos dá uma ideia muito clara da fragmentação do texto.

Observemos, então, um texto³ transcrito de acordo com o sistema do NURC:

- 1 Inf2. só se tu acreditares... que é possível mudar... é que tu vais mudar... se tu partires da atitude de que ah não a universidade é uma coisa VELHA CADUCA... não se pode fazer nada... então tu não vais fazer nada... eu nesse caso não estaria na universidade... portanto se
- 5 ESTOU... é porque acredito que é possível fazer coisas e que é possível inovar...

Um sistema igualmente simples, mas que apresenta diferenças substanciais face ao NURC, sendo porventura mais forte ao nível da legibilidade, é o do NERC-47⁴. A principal divergência consiste na introdução de um símbolo próprio para delimitar as fronteiras frásicas:

¹ Groupe Aixois de Recherches en Syntaxe, dirigido por Claire Blanche-Benveniste.

² Norma Urbana Culta, projecto de investigação brasileiro relacionado com a língua falada.

³ Programa Conversas Secretas, SIC, *corpus* REDIP.

⁴ Network of European Reference Corpora, projecto da Comissão Europeia para o estabelecimento de convenções gerais em *corpora* de referência. As referências NERC-47 e NERC-50 correspondem a diferentes versões do sistema proposto pelo projecto.

Frases frásicas	Frases interrogativas	Pausas inesperadas
.	?	[pausa]

Vejamos como é que isso resulta na prática:

<F01>: Só se tu acreditares que é possível mudar é que tu vais mudar. Se tu partires da atitude de que ah não a universidade é uma coisa velha caduca não se pode fazer nada então tu não vais fazer nada. Eu nesse caso não estaria na universidade. Portanto se estou é porque acredito que é possível fazer coisas e que é possível inovar.

É claro que esta diferença tem os seus custos, potenciando em muito o nível de subjectividade da transcrição, ao fazer apelo a uma noção sintáctica. Deste modo, fica sempre a cargo do transcritor a delimitação das fronteiras frásicas, o que poderá conduzir a resultados pouco consistentes. Por outro lado, é interessante verificar que este sistema permite a utilização de pontos de interrogação mas não de pontos de exclamação. Na verdade, cremos que se trata de um sistema que possibilita a subjectividade mas só até certo ponto.

De todos os projectos aqui considerados, o mais conservador em relação à ortografia é, sem dúvida, o REDIP. O seu sistema mantém grande parte dos sinais ortográficos, o que torna as transcrições mais subjectivas e graficamente mais complexas. Aqui está o quadro com os símbolos usados:

Pausa (sintáctica) longa		Pausa (sintáctica) breve	Outras pausas
Contexto declarativo	.	,	...
Contexto interrogativo	?		
Contexto exclamativo	!		

Neste caso, as principais vantagens verificam-se ao nível da inteligibilidade, dada a proximidade com a ortografia vigente. Somente as pausas sintacticamente motivadas podem ser representadas por ponto final ou vírgula, havendo em todos os outros casos recurso às reticências. Este sistema é bastante apelativo, apesar de se distanciar sobremaneira do contínuo sonoro do oral. O problema surge, contudo, nas sequências em que há fronteiras sintácticas que não são marcadas, no discurso, por nenhuma pausa. Nesse tipo de contextos, o sistema fica aquém do seu carácter eminentemente ortográfico, o que motiva alguma inconsistência. Este fenómeno é visível no texto em apreço, por exemplo na sequência <coisa velha caduca>, que, de acordo com o padrão ortográfico, deveria ser grafada como <coisa velha, caduca>:

L2: só se tu acreditares... que é possível mudar é que tu vais mudar. se tu partires da atitude de que ah não a universidade é uma coisa velha caduca, não se pode fazer nada, então tu não vais fazer nada, eu nesse caso não estaria na universidade. portanto se estou, é porque acredito que é possível fazer coisas e que é possível inovar.

Depois de termos analisado um sistema que usa uma pontuação muito próxima daquela que normalmente usamos na escrita, é altura de observarmos um método de transcrição que se distancia por completo da grafia convencional. Trata-se do sistema do GARS, grupo de que já falámos acima, cujas transcrições se destinavam a uma análise da sintaxe do francês oral. Consideremos, então, o conjunto de símbolos usados:

Pausa curta	Pausa média	Pausa longa	Interrupção bastante longa
+	--	---	////

Verifiquemo-lo na prática:

- L2 1 só se tu acreditares +que é possível mudar +é que tu vais
 2 mudar -- se tu partires da atitude
 3 de que ah não a universidade é uma coisa velha caduca +
 4 não se pode fazer nada +então tu
 5 não vais fazer nada eu nesse caso não estaria na
 6 universidade -- portanto se estou +é
 7 porque acredito que é possível fazer coisas e que é possível
 8 inovar ---

Uma das críticas que podem ser apontadas a este sistema prende-se com a ininteligibilidade destes símbolos. O seu uso requer um determinado grau de adaptação, facto que poderá afastar um leitor não especializado. BACELAR DO NASCIMENTO (1987) é bastante clara em relação a isso, justificando o sistema de pontuação adoptado nas transcrições feitas para o Português Fundamental:

«Se numa transcrição ortográfica a parcial inadequação dos símbolos gráficos ao sistema fonético empobrece já, irremediavelmente, o enunciado oral, a adopção de mais notações tão arbitrárias como as próprias convenções ortográficas, menos rigorosas do que as notações fonéticas e igualmente afastadas dos nossos hábitos de leitura e escrita, teria, ainda, acentuado o carácter artificial desta representação gráfica.»

Resta-nos ainda falar dos sistemas que dão prioridade à delimitação das fronteiras entoacionais do enunciado, em detrimento da marcação das pausas.

Tal é o caso do NERC-50⁴ e da versão italiana do CHAT⁵. Começaremos pelo mais simples dos dois, o NERC-50. Este é já um sistema multilinear, mantendo, no entanto, a simplicidade da versão anterior:

Fronteiras fráscas	Delimitação das fronteiras entoacionais
.	/

Como se vê, não existe nenhuma notação específica para as pausas. A delimitação das fronteiras fráscas é feita no primeiro nível, ao passo que a delimitação das fronteiras entoacionais é feita no terceiro. O quarto e último nível está reservado para informação de carácter prosódico, o que torna a introdução de símbolos como o ponto de interrogação (presente no NERC-47) redundante.

No entanto, num sistema tão simplista, e dada a prioridade à delimitação das fronteiras entoacionais, talvez fosse preferível abdicar de todo da pontuação, uma vez que, como os próprios autores reconhecem, esse é sempre um campo aberto à subjectividade. Observemos, então, como funciona o sistema do NERC-50:

<F01>: SÓ se tu acreditares que é possível muDAR/ é que tu vais muDAR./ se tu partires da atitude de que AI não/ a universidade é uma coisa VELha caDUca/ não se pode fazer NAda/ então tu não vais fazer NAda/ eu nesse caso não estaria na universiDAde./ Portanto se esTOU/ é porque acredito que é possível fazer COIsas e que é possível inoVAR./

A versão italiana do CHAT apresenta-nos aquele que é provavelmente o sistema mais criterioso em relação à questão das pausas e da fragmentação. Como dissemos, este é um sistema que dá prioridade à delimitação das fronteiras entoacionais. Fá-lo, no entanto, explicitamente, usando apenas os símbolos relativos a pausas quando estas correspondem, nas palavras de CRESTI (2000), a «una prominenzza particolare di silenzio nel flusso parlato». A pausa deverá corresponder a pelo menos quatrocentos milésimos de segundo de silêncio, sendo considerada longa ou muito longa a partir de um segundo. No seguinte quadro estão expostos os símbolos usados:

Esta é uma forma bastante elegante de resolver o problema da fragmentação do texto, se bem que restem algumas dúvidas relativas à delimitação das fronteiras entoacionais e à sua maior ou menor objectividade. Observemos, então, como funciona na prática:

⁵ Codes for the Human Analysis of Transcripts. Ao longo do texto, falaremos sempre do formato CHAT tal como foi implementado nos *corpora* do projecto italiano LABLITA, razão pela qual usamos a expressão *versão italiana do formato CHAT*.

Unidade entoacional terminal do enunciado		Unidade entoacional não terminal	Pausa	Pausa longa	Pausa muito longa
Contexto declarativo	//	/	[#]	[##]	[###]
Contexto interrogativo	?				
Contexto exclamativo	!				
Suspensão intencional	...				

*MFA: só se tu acreditares [#] que é possível mudar / é que tu vais mudar // se tu partires da atitude de que ah não / a universidade é uma coisa velha caduca / não se pode fazer nada / então tu não vais fazer nada / eu nesse caso não estaria na universidade // portanto se estou [#] é porque acredito que é possível fazer coisas e que é possível inovar //

Antes de terminar, gostaríamos ainda de falar um pouco sobre o Val.Es.Co⁶ e o Português Fundamental. Trata-se de projectos que usam sistemas completamente distintos, mas que não devem deixar de ser referidos. O Val.Es.Co emprega um sistema semelhante ao do GARS, mas com indicações mais precisas e símbolos mais intuitivos:

Pausa curta, inferior a meio segundo	Pausa entre meio segundo e um segundo	Pausa de um segundo ou mais	Indicação do número de segundos de silêncio
/	//	///	(5»)

O Português Falado, por sua vez, emprega um sistema que pura e simplesmente abdica de qualquer notação relativa a pausas ou a fronteiras entoacionais. Isso deve-se ao facto de as transcrições poderem ser sempre acompanhadas do respectivo alinhamento sonoro, o que torna a fragmentação do texto uma questão secundária⁷.

2.2. Os enunciados simultâneos

Como é que representamos a sobreposição das falas numa transcrição unilinear? Obviamente não podemos pôr as diferentes elocuições na mesma

⁶ Valencia, Español Coloquial, grupo de investigação dedicado ao estudo do espanhol falado. Agradecemos a Maria Helena Sereno a gentileza de nos ter dado a conhecer este grupo.

⁷ No entanto, é preciso ver que o alinhamento não beneficia directamente o *corpus* constituído, mas sim a relação com o utilizador final da transcrição.

linha, ainda que essa fosse uma estratégia defensável em termos de fidelidade de transcrição⁸:

L1: o senhor é independente?

L2: completamente!

L1: senhor doutor, perante todos os agentes, a lista que eu tenho é independente.

Tal como em relação ao problema das pausas, as soluções variam consoante o projecto ou grupo de trabalho. Na maior parte dos casos, porém, as diferenças manifestam-se a nível meramente gráfico.

Nas transcrições do grupo GARS, todas as falas sobrepostas são sublinhadas, não havendo nenhuma outra indicação. Este tipo de tratamento simplificado tem algumas vantagens. Uma delas reside no facto de nos permitir acompanhar de uma forma dinâmica a evolução do discurso, já que é possível observar súbitas mudanças de tópico (fenómeno muito frequente em enunciados simultâneos) por meio da translineação. Não é, no entanto, um sistema adequado a determinadas análises discursivas, uma vez que não prevê a existência de diferentes tipos de sobreposição.

Observemos, então, um texto com sobreposição de falas⁹ transcrito de acordo com o sistema do grupo GARS:

L4 1 ó senhor doutor tenha calma ~~que~~ o senhor doutor tem

2 tempo de ganhar o campeonato

L3 3 calmíssimo

L4 4 já anda há tantos anos

L3 5 estou calmíssimo

L4 6 à espera dum campeonato

Convém referir que este sistema não é totalmente preciso, uma vez que apenas tem em conta a sobreposição de palavras e não a de sílabas ou segmentos. Por isso, quando temos palavras com muitas sílabas, a discrepância entre a sobreposição real e a sobreposição representada na transcrição pode tornar-se óbvia. Este é um defeito comum a todos os sistemas observados.

Como nas transcrições do GARS, a representação dos enunciados simultâneos no REDIP é feita sublinhando as falas:

L4: ó senhor doutor tenha calma, que o senhor tem tempo de ganhar o campeonato.

L3: calmíssimo!

L4: já anda há tantos...

⁸ O trecho que se segue é do programa Livre e Directo, Antena 1, *corpus* REDIP.

⁹ Programa Os Donos da Bola, SIC, *corpus* REDIP.

L3: estou calmíssimo!

L4: anos à espera de um campeonato.

No NURC, porém, a representação do fenómeno é feita de outro modo. É usado um parêntese recto, ficando este no enquadramento da primeira palavra a partir da qual houve sobreposição:

- 1 Inf4. ó senhor doutor tenha calma... que o senhor tem tempo de ganhar o campeonato...
 [
 Inf3. calmíssimo
 Inf4. já anda há tantos
 [
 5 Inf3. estou calmíssimo
 Inf4. anos à espera dum campeonato.

O NERC-47 não prevê nenhuma notação especial para as falas sobrepostas. No entanto, os transcritores podem inserir comentários pessoais relativos a fenómenos desse tipo, desde que estejam delimitados pelos códigos pré-definidos no projecto. Vejamos em que é que isto se traduz:

- <M04> Ó senhor doutor tenha calma que o senhor tem tempo de ganhar o campeonato
 <ZZ1> sobreposição das três últimas palavras com a fala que se segue <ZZ0>
 <M03> Calmíssimo
 <M04> Já anda há tantos <ZZ1> sobreposição das últimas palavras com a fala que se segue <ZZ0>
 <M03> estou calmíssimo.
 <M04> anos à espera dum campeonato.

No NERC-50, os enunciados simultâneos recebem já um tratamento mais específico, existindo dois símbolos próprios para a sua representação. O momento em que se dá o início da sobreposição é marcado com um asterisco e o final com um cifrão. Esta informação faz parte do terceiro nível. Vejamos como funciona na prática:

- <M04> /Ó senhor doutor tenha CALma/ que o senhor tem TEMpo de *ganhar o campeonato\$./
 <M03> /* CalMÍssimo\$/
 <M04> /Já *anda há TANtos\$
 <M03> /* Estou calMÍssimo\$/
 <M04> anos à espera dum campeonato./

Os sistemas que temos vindo a observar até agora caracterizam-se pela sua simplicidade. Essa simplicidade traduz-se, por exemplo, numa maior compa-

tibilidade a nível informático. No entanto, pelo facto de serem muito simples, não se prestam a determinados tipos de utilização.

Na versão italiana do CHAT, pelo contrário, o caso dos enunciados simultâneos é tratado de uma forma especial. Em primeiro lugar, é feita a distinção entre sobreposição de falas e sobreposição de diálogos. A sobreposição de falas é aquela que ocorre dentro de um mesmo diálogo, havendo um fio condutor claramente definido. É mais frequente em diálogos com dois intervenientes. A sobreposição de diálogos, por seu lado, surge quando a conversa se dispersa simultaneamente por vários tópicos, criando pólos divergentes dentro da mesma situação conversacional. Esta sobreposição é típica nas conversas com muitos intervenientes (quatro ou mais).

Dentro da sobreposição de falas, são ainda assinalados dois casos diferentes: a sobreposição que é interna a uma determinada unidade de informação (que corta a fala precedente, criando um choque, originando muitas vezes interrupção) e a sobreposição que surge no seguimento da fala anterior, estando perfeitamente enquadrada no discurso (não gera interrupção, aproveita hesitações e tons descendentes). No primeiro caso, a situação é assinalada com um parêntese angular de fecho; no segundo, com um parêntese angular de abertura. Vamos, então, visualizar o tratamento dado pelo CHAT ao caso da sobreposição de falas, recorrendo, desta feita, a um outro trecho do nosso *corpus*¹⁰, de modo a poder visualizar melhor as diferenças:

*MF: antigamente não se fazia sobre os livros / como sabes / depois começou-se a fazer sobre os <livros> [>]

*BB: <isso> isso é umas coisas loucas da universidade / a universidade é <extremamente> [>]

*MF: <a universidade> tem coisas muito boas <também>

*BB: [<] <tem> tem mas também / quer dizer lá está / não sejas corporativista //

*MF: não sou <corporativista> [>]

*BB: <porque> a universidade é uma coisa conservadora tem sido, ouviste?

*MF: a universidade é uma instituição profundamente <conservadora>

*BB: [<] <de acordo>

Segue-se agora um caso de sobreposição de diálogos, em que a disposição do texto é feita em colunas:

*HA: o [/] o boavista-guim

*EB: fácil falar dos outros!

*DB: não há muitos dirigentes no

¹⁰ Programa Conversas Secretas, SIC, *corpus* REDIP.

*DB: futebol português que não protestem contra a arbitragem //

*HA: desculpe / o boavista-guimarães também

*HA: também neste [//] nesta [//]

*HA: no início

*EB: lembra-se de

Por fim, para concluir esta parte, apresentamos um quadro onde é possível comparar as diferentes soluções gráficas adoptadas nos sistemas em que se verifica um tratamento generalizado, por oposição àquilo que acontece na versão italiana do CHAT:

Projectos			CHAT (Versão Italiana)	GARS/REDIP	NURC	NERC-47	NERC-50 (Nível 3)
Sobre- posição	De falas	Discurso interrompido	*MF: texto <texto> [>] *BB: texto	L1: texto <u>texto</u> L2: <u>texto</u> texto	AL. texto [Inf. texto	<F01> texto <ZZ1> sobreposição com a fala seguinte <ZZ0>	<F01> texto *texto\$ <M01>*texto\$ texto
		Discurso seguido	*MF: texto <texto> *BB: [<] <texto> texto				
	De diálogos	*HA: texto *EB: texto					

2.3. Ideofones e interjeições

Os ideofones são sons que não desempenham um papel funcional na gramática da língua e que não são susceptíveis de receber uma representação ortográfica, sendo, no entanto, usados pelos locutores para veicular um determinado tipo de ideia ou reacção (agrado, desagrado, espanto, prazer, repulsa, etc.). É o caso do som normalmente usado pelo locutores portugueses para manifestar concordância, que, no nosso projecto, recebe a representação <hum>, mas que pode ser escrito de outras maneiras (<hmm>, <hm-hm>, <hum-hum>, etc).

Este tipo de sons, juntamente com as interjeições e onomatopeias, constitui um problema para os transcritores, que muitas vezes não sabem como reagir perante tais produções. Para as transcrições do grupo GARS, por exemplo, não existe nenhuma tabela definida, e as interjeições, onomatopeias e outros sons são grafados de acordo com os dicionários de língua. No projecto REDIP, também não temos nenhuma tabela, se bem que exista a grafia específica <hum> para representar os acordos, como já referimos.

No NURC, existe um conjunto de representações previstas para estes casos, mas está reduzido a oito elementos, e não vem acompanhado por nenhum tipo de descrição fonética, articulatória ou semântica:

Fáticos	ah	eh	éh	ahn	ehn	uhn	tá
---------	----	----	----	-----	-----	-----	----

Já os transcritores que estiverem a trabalhar de acordo com as estipulações do NERC¹¹ têm ao seu dispor uma tabela bastante detalhada, devendo reger-se por ela sempre que necessário. O quadro que apresentamos seguidamente dá conta dos ideofones e interjeições previstos para o inglês:

Tipo de som	Representações adoptadas	Descrição fonética, articulatória ou semântica
Sons de espanto	ah oh ooh	[ɑ] [əʊ] [u:]
Sons grunhidos	mm hm mhm uh huh ugh	uma sílaba, lábios fechados uma sílaba, lábios juntos, iniciando uma explosão de ar duas sílabas, lábios juntos duas sílabas, lábios afastados ruído de desagrado, muitas vezes só um grunhido
Outros	hey eh oi ah hah	para chamar a atenção ou expressar surpresa expressando perplexidade ou procurando acordo vocativo expressando surpresa

Na versão italiana do CHAT, também existe um conjunto de interjeições (do italiano e especificamente do dialecto toscano) pré-definido, correspondendo *grosso modo* à descrição apresentada na coluna da direita:

Interjeições gerais	Descrição semântica
ah	compreensão de um enunciado precedente
ah!	espanto, maravilha
oh! ih! uh!	espanto, contrariedade
ohiohi	dor, contrariedade
&che	apoio fonético numa indecisão
mh	chamada de atenção, aprovação, dúvida
eh	incompreensão, pedido de explicação, confirmação
hei	reclamação
bah	incredulidade
bè	pedido de explicação (começo)
boh	indecisão
mah	incredulidade

¹¹ O caso aplica-se tanto ao NERC-47 como ao NERC-50.

Interjeições exclamativas (toscano)	Descrição semântica
nananà	= <i>etc</i>
ecé	= <i>etc</i>
va'	comentário exclamativo
sie	concordar pela negativa (ironia)
zà	onomatopeia para altas velocidades
uah	onomatopeia para chorar

Este sistema contempla igualmente os sons que acompanham acções ou gestos e sons como os do choro, o riso, etc. Para transcrever estas realizações, é seleccionado um símbolo pré-definido, sendo este independente dos sons realmente produzidos. À direita do símbolo aparece sempre um comentário explicativo:

hhh [!= risos]

Este método tem uma vantagem que consiste em evitar a utilização de símbolos arbitrários para representar os ideofones, optando por uma caracterização explícita do sentido veiculado pelo sinal (ainda que despreze a componente fonética).

REFERÊNCIAS BIBLIOGRÁFICAS

- BACELAR DO NASCIMENTO, F. et aliae, 1987 *Português Fundamental – Métodos e Documentos*, Lisboa: INIC/CLUL.
- BACELAR DO NASCIMENTO, F. et aliae, 1989 *Como Escrever o Oral?*, in *RILP*, número 2, Lisboa: AULP.
- BLANCHE-BENVENISTE, C. e C. Jeanjean, 1987 *Le Français Parlé*, Paris: INALF/CNRS.
- BRIZ, A., 2000 *¿Cómo se Comenta un Texto Coloquial?*, Barcelona: Editorial Ariel.
- CALLOU, D. (coord.), 1991 *A Linguagem Falada Culta na Cidade do Rio de Janeiro: Materiais para o Seu Estudo*, Rio de Janeiro: UFRJ, 1991.
- CRESTI, E., 2000 *Corpus di Italiano Parlato*. Volume I, Firenze: Presso l'Accademia della Crusca.
- FRENCH, J. P., 1991 *Updated Notes for Soundprint Transcribers*. NERC-WP 4-47, Birmingham: JP French Associated, York e Cobuild.
- FRENCH, J. P., 1992 *Transcription Proposals: Multi-Level System*, NERC-WP 4-50. Birmingham: University of Birmingham.
- SINCLAIR, J. e J. Ball, 1995 *Spoken Language within a Written Language Corpus*, texto copiado.

