

José Teixeira
Universidade do Minho
jsteixeira@ilch.uminho.pt

Língua escrita e língua real: o que um corpus oral permite descobrir

Resumo:

Esta comunicação, baseando-se num extenso corpus oral que está a ser recolhido, procurará refletir de que modo a língua realizada oralmente nos permite perceber o que realmente são e como funcionam as línguas para além da dimensão formalizada da escrita. Pretende-se demonstrar que é a linguagem em uso que melhor evidencia muitos aspetos impossíveis de perceber por uma linguística designada *do sistema*. Assim, procurar-se-á verificar até que ponto a verdadeira realização linguística da oralidade respeita a noção de frase e de norma, assim como especificamente realiza várias dimensões lexicais e pragmáticas.

Palavras-chave: Língua oral, língua escrita, interjeições, norma, pragmática linguística.

Abstract:

The written language and the real language: what an oral corpus can reveal

Starting from an extensive oral corpus which is being collected, this paper focuses on how the oral language allows us to realize what languages really are and how they work beyond the formalized written dimension. It is our claim to

show that it is language in use that better highlights many aspects which are impossible to understand only by an analysis of the traditional “linguistic system”. So we will verify how oral speech really performs sentences and its lexical and pragmatic elements.

Keywords: oral language, written language, interjections, norm, pragmatics.

1. Como funcionam as línguas?

A simultânea simplicidade e complexidade da questão obviamente impedirá uma resposta satisfatória e completa. É a pergunta de um milhão de dólares para as ciências da linguagem. E como cada vez mais percebemos como os mecanismos cognitivos e os processamentos linguísticos perfazem um complexo de partes diferenciáveis mas interligadas, quanto melhor vamos sabendo responder à questão, mais adivinhamos o muito que precisaríamos de saber.

A problemática é, no fundo, aquela a que todas as áreas da Linguística e todas as teorias sobre a linguagem têm, cada uma a seu modo, procurado responder. E se a linguagem / pensamento constituem a componente humana mais espantosa e tida como misteriosa e em tempos até vista como divina, não é de admirar que o esclarecimento daquela questão não se fique apenas pelo âmbito da “curiosidade científica”, mas envolva quase todos os domínios do que é humano.

Quando surge a pergunta “O que é uma língua?”, habitualmente entende-se que é uma pergunta retórica, mecanismo habitual para se partir de um objeto de estudo que toda a gente sabe qual é. Mas não é bem assim. Olhando a história da Linguística, verifica-se que decidir qual o objeto de estudo sempre foi e tem sido a questão mais marcante e que mais consequências traz à investigação e aos resultados que se pretendem encontrar. Dizer que com Saussure o problema ficou resolvido é iludir a questão. Não é pacífico saber o que é verdadeiramente uma língua, por bem vincadas que sejam as dicotomias saussureanas relativas ao que é e ao que não é do âmbito do estudo da

“linguística propriamente dita”, usando a expressão do fundador da Linguística moderna:

(...) seria quimérico reunir sob um só ponto de vista a língua e a fala
(...) Podemos, em rigor, manter o nome de linguística para cada uma destas disciplinas e falar de uma linguística da fala. Mas será preciso não a confundir com a linguística propriamente dita – a da língua [Saussure, 1978: 50].

Saussure, na pegada do seu tempo e dos seus mestres comparativistas e Neogramáticos, via a língua através da escrita e a sua enorme influência nos estruturalismos europeus e americanos desviou a atenção das outras dimensões da linguagem. A fala (“parole”) era algo de secundário, algo que não deveria fazer parte das preocupações descritivas da ciência linguística. A escrita era a fala depurada das “imperfeições” e dos erros da língua-sistema e era esta a dimensão que deveria ser o objeto de estudo.

Por isso, é que só nos últimos tempos a “parole” (fala) saussureana ou o desempenho chomskyano (“performance”) se tornam matéria de investigação nas ciências da linguagem. Se resumíssemos a história da Linguística em quatro grandes fases, poderíamos ver que apenas na última, a partir das duas últimas duas décadas do séc. XX, a totalidade do fenómeno linguístico é tida como objeto válido de investigação:

Fase 1, ou a *Fase da Ingenuidade* (até ao fim do século XIX): Língua implica letras, implica o sistema de escrita. É a fase que vai até aos Neogramáticos, incluídos. Se bem que estes, por causa do valor que davam à componente fonética e ao “elemento psicológico”, defendessem que as ciências da linguagem deveriam estudar as línguas no “laboratório vivo” que é a linguagem em ação, o certo é que não o fizeram.

Fase 2, ou a *Fase da Catalogação*: A Geografia Linguística e a Dialectologia (sobretudo na primeira metade do século XX). Tal como fizeram a Zoologia e a Botânica, agora a Linguística procura elencar as línguas, dialetos, subdialetos existentes, descrevê-los e estabelecer as relações entre eles.

Fase 3, ou *Fase das Estruturações Definíveis*: sobretudo a partir da segunda década do século XX até à atualidade. Os estruturalismos

européus de influência saussureana e os americanos de origem bloomfieldiana e chomskyana são os grandes representantes. Procuram a sistematização, a descoberta da estrutura das línguas, tidas essencialmente como formais, lógicas, baseadas em estruturas de Condições Necessárias e Suficientes e para alguns (Chomsky) assentes em Universais comuns à espécie humana (se bem que nunca seja muito esclarecedor em que consistem estes Universais). É a Linguística da *langue* e da competência, a Linguística dos princípios formais e da estrutura.

Fase 4, ou *Fase da Experienciação e Cognição*: sobretudo a partir das duas últimas décadas do século XX até à atualidade. É a Linguística dos contextos, dos atos ilocutórios e perlocutórios, da pragmática linguística e a Linguística Cognitiva (de Lakoff, Talmy, Langacker, entre outros).

E é nesta fase que aparece valorizada a Linguística da *parole*, da *performance* a Linguística do uso e da oralidade, por oposição à anterior, à Linguística do sistema. Agora os usos “anormais” e os erros são tidos como fenómenos que evidenciam o funcionamento verdadeiro das línguas e da sua base biológica, cognitiva, cultural e pragmático-contextual.

2. A Linguística de estufa

A longa tradição de estudos linguísticos apresentada partia do modelo inquestionado de que a escrita é a parte “melhor” da língua. Note-se que o conceito popular mais referido que se tem sobre as palavras (“uma palavra é um conjunto de letras” é a definição mais frequente) pressupõe que é a oralidade a expressão (deficiente) da escrita e não o inverso.

Não apenas por estes motivos, mas também por razões práticas, os exemplos e *corpora* sobre os quais têm assentado os estudos linguísticos têm sido elaborados esmagadoramente a partir da escrita. As tecnologias informáticas, permitindo recolhas quase instantâneas de milhões de ocorrências escritas, exponenciaram a metodologia.

No entanto, a interface humana das línguas é maioritariamente oral e se quisermos que as máquinas simulem o mais adequadamente possível as línguas naturais, teremos que procurar perceber como é que a linguagem funciona nesta dimensão. Para tal é imprescindível a constituição de corpora da oralidade que permitam analisar e perceber as idiossincrasias desta vertente relativamente à escrita.

Costuma dizer-se que a Linguística tradicional para analisar a língua, pelos exemplos que inventa artificialmente, faz como um botânico que quisesse descrever as plantas, servindo-se de flores de plástico. Os exemplos construídos para análise também são língua e são parecidos com a língua em uso real. Tal como as flores de plástico, também são parecidas com as reais.

Talvez a metáfora das flores de plástico seja um pouco radical para ilustrar o afastamento entre o objeto tradicionalmente observado e o que deveria ser estudado pela Linguística. Mas se não tem sido uma Linguística de flores de plástico, pelo menos tem sido uma Linguística de flores... de estufa. Por mais naturais e bonitas, certinhas e formatadas que sejam as flores de estufa, elas nunca podem representar a realidade da natureza. Tal como não representam a verdadeira língua as frases construídas, os exemplos estereotipados e normativamente corretos da linguagem escrita. Isso é a estufa da língua, um pouco diferente da luta agreste, das variações, dos erros e desvios que acontecem na natureza. E também porque há flores que existem, mas que não se encontram em nenhuma estufa.

3. Sair da estufa e ver a realidade

3.1. O projeto Perfil Sociolinguístico da Fala Bracarense (PSFB)

Terá que ser obrigatoriamente através da constituição de *corpora* da oralidade que se poderá constituir a matéria para verdadeiramen-

te apreender as reais dimensões de como funciona uma língua na interação oral vivencial e contextualizada.

O projeto Perfil Sociolinguístico da Fala Bracarense (PSFB) é um projeto de investigação apoiado pela Fundação para a Ciência e Tecnologia (referência FCT PTDC/CLE-LIN/112939/2009) e desenvolvendo-se de 2011 a 2014 insere-se precisamente nesta dinâmica. Pretende a obtenção controlada de dados da fala natural e não apenas observações impressionistas sobre a realidade linguística de uma determinada comunidade, visando uma melhor compreensão dos aspetos ligados às formas e sistematicidade da variação linguística.

Recolhe-se um corpus oral a ser transcrito para posterior análise linguística. Neste texto, através da análise de partes do referido corpus tentaremos mostrar como a análise linguística que usualmente se serve de *corpora* escritos, só consegue ver o que a escrita lhe permite que veja e como muitíssimos marcadores de expressividade do discurso, ao não aparecerem escritos, acabam por ser tidos como não existentes.

3.2. O que um corpus oral permite descobrir (ou confirmar)

3.2.1. O discurso oral não é linear como o escrito, o que implica miríades de operações linguísticas

O óbvio desta primeira constatação costuma apenas fazer reparar como a oralidade é deficitária em relação à escrita. Subentende-se que é a escrita a língua real, definitiva.

Ora isto é ver a questão ao contrário. Passar a oralidade para um discurso escrito não está ao alcance de qualquer falante, mas apenas do falante treinado escolarmente. Um falante analfabeto que realize normalmente a oralidade, não consegue fazer a passagem para o discurso escrito e temos que admitir que ele realiza funcionalmente a sua língua. A realização oral típica (não a formal) não indicia se o falante sabe escrever ou não. Portanto, a realização oral tem de ser vista como a realização linguística normal. Isto implica admitir que, por norma,

o discurso não é linear, não cumpre com rigor as normas morfosintáticas e que os verdadeiros atos linguísticos normais (orais) estão muito longe da normatividade que a escrita aparenta. Qualquer gravação da oralidade (de registo não formal) prova isso à saciedade.

Ilustre-se o que se quer fazer ressaltar com este pequeno exemplo¹:

78M3D	..	132 [05:39.2]	133 [05:40.6]
Ent2 [v]		• • Brillhante!	
Fal78 [v]	vinte e dois no exame • • nacional.		• • Mas depois o que apa/ o que

	..	134 [05:45.1]
Fal78 [v]	eu aturei àquela turma, vocês nem possam passar pela cabeça.	• • Porque,

Como se transforma isto em frases “normais” da escrita?

E não se pense que o falante é iletrado: é uma professora do ensino secundário com muita prática – reformada – a falar, o que demonstra que a divergência entre a escrita e a oralidade não é apenas uma questão de escolaridade maior ou menor.

Para podermos perceber a frase, temos que a reconstruir mentalmente e terá que ficar com uma estrutura muito diferente:

- *apa* / terá que ser percebido como a eliminar;
- *vocês*, que parece que iria ser o sujeito, não é sujeito de nenhuma frase;
- a forma verbal “*possam passar*” não concorda com nada.

No entanto, o enunciado “absurdo e sintaticamente incongruente” *vocês nem possam passar pela cabeça* é interpretado com êxito:

¹ Elementos e significado: 78M3D = código da entrevista; ..132 [05:39.2] = momento 132 da entrevista que acontece aos 5 minutos 39 segundos e 2 décimos de segundo; Ent2 [v] = faixa das intervenções do entrevistador; Fal78[v] = faixa do falante/entrevistado.

- por causa do contexto precedente que indica uma coisa difícil de entender (*o que eu aturei àquela turma...*);
- porque aciona as combinatórias mais ou menos fixas *vocês nem podem imaginar, a vocês nem lhes passa pela cabeça*;
- o *vocês nem podem* transforma-se em *vocês nem possam* porque mentalmente a segunda opção, *passar*, ao ser acionada, altera a fonologia de *podem*. A forma “incompreensível” e desajustada *vocês nem possam* resultará do cruzamento de *podem* com *passa*.

Este ínfimo exemplo da normalidade dos processos da oralidade demonstra como os atos da oralidade implicam constantes processos cognitivos de equivalências em que o ALOC tem de reorganizar a fala do LOC de molde a construir um discurso que faça sentido, sem estar à espera de receber todos os dados corretamente ordenados. As implicações para a descrição do real funcionamento das línguas e para a linguística computacional que vise a adequação à oralidade são absolutamente essenciais, caso se queira construir um bom algoritmo sobre a percepção da oralidade.

3.2.2. *O discurso oral pressupõe por princípio que conhecemos o léxico*

Ao contrário do que muitas vezes se pensa, a realização da palavra não é a soma dos respetivos fonemas, mas apenas um esboço fonético. Pronunciar todas as sílabas é uma exceção que fazemos só quando o falante não tem acesso fácil ao mesmo léxico (com falantes não nativos). Partimos do pressuposto de que a unidade lexical é conhecida pelo interlocutor e por isso não precisamos de a realizar com grande nitidez. Acontece, por isso, que quando o interlocutor não conhece a palavra, dificilmente a constrói através dos sons ouvidos, embora o LOC pense que a pronunciou *toda* e que por isso será normal o ALOC percebê-la.

Veja-se nesta entrevista: o revisor 1 não conhece a novela *Gabriela* (baseada num romance de Jorge Amado) e por isso não consegue “ouvir” palavras que quem conhece a respetiva novela (revisor 2) “ouve nitidamente”.

78M3D	814 [39:07.4]	815 [39:13.4]
Fal78re- vis2	•• ele é uma/, portanto, é o capi/ é o governador da cidade de Ilhéus	•• e eu
revisor 1	•• ele é uma/, portanto, é o capi/ é o governador da cidade ((incompreensível))	•• e eu

78M3D	..	816 [39:17.5]	817 [39:18.5]	818 [39:20.0]
Ent2 [v]		•• Hum hum.		
Fal78re- vis 2	gostasse mais do Paulo Gracindo...		•• ((hesitação))	•• Porque o
revisor 1	gostasse mais de um ((incompreensível))...		•• ((hesitação))	•• Porque o

3.2.3. Na memória oral de curto prazo fazemos equivalências semânticas

Aparentemente, a análise linguística pressupõe que na interpretação de uma frase nos recordamos da totalidade das palavras pronunciadas. Ou seja, a interpretação da frase resulta da estrutura das palavras realmente emitidas.

Mas será assim? Se uma frase durar 10 segundos, temos na memória as palavras desde o início?

A psicologia defende que a nossa memória auditiva de curto prazo é de cerca de 4 segundos (a visual é de um quarto de segundo). Isto significa que não é fácil guardar na memória as palavras ouvidas, mas antes procuramos tratar de lhes dar uma equivalência semântica o mais rápido possível para ir ouvindo e fixando, nos segundos seguintes, as palavras que continuam a construir a frase.

Este processo não é o que se entende que preside à descodificação das estruturas linguísticas. A análise estrutural tradicional, ao

descrever a língua como uma estrutura, vê a interpretação semântica da frase como o resultado global da soma das estruturas parciais e estas como decorrendo dos valores das palavras utilizadas:

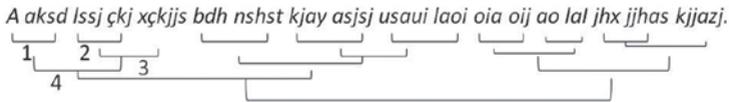


Figura 1

Assim, na frase da Figura 1, a estrutura 4 resulta da relação entre a estrutura 1 (*A aksd*) e a estrutura 3, que por sua vez resulta da relação entre a estrutura 2 (*lssj çkj*) e a palavra *xçkjjs*. Este total seria traduzido por uma estrutura semântica correspondente que iria relacionar-se com as restantes e no fim dotar a frase de uma interpretação semântica.

Ideal e teoricamente, isto seria assim. Na prática, há muitas dúvidas de que possa ser. Isto implicaria que conservássemos em memória as palavras que foram ditas na frase do princípio ao fim e que não houvesse falhas no processo de perceção. Na realidade, isto não acontece na maior parte dos atos da oralidade. Não é possível (nem útil) conservar em memória as palavras e por isso no processamento linguístico tratamos de, logo que possível, atribuir uma interpretação semântica à microestrutura apreendida: podemos “esquecer”, assim, as palavras dessa microestrutura e estar aptos para fixar as seguintes e proceder da mesma forma. Significa isto que a frase total acaba por ser um complexo construído realmente entre microestruturas semânticas que se servem de microestruturas morfossintáticas.

Este funcionamento permite colmatar as falhas no processo de perceção.

Imaginemos (Figura 2) que no exemplo atrás dado uma das palavras não é bem percecionada:

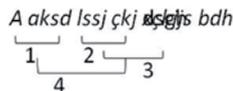


Figura 2

Isto não implica que o ouvinte bloqueie, mas que tente construir a estrutura que lhe parece compatível. Quando uma unidade não é (bem) percebida mas é possível construir um modelo admissível, o ALOC admite esse modelo e faz dele uma das etapas para a interpretação final. Não é a palavra em si, ou a sua não percepção, que é importante, mas a construção da microestrutura ou modelo mental de interpretação semântica em progresso.

Por isso é que nos recordamos de uma frase, de uma conversa, não recordando as exatas palavras, mas os modelos mentais semântico-cognitivos envolvidos.

Reexaminar o ouvido e o anotado pelos transcritores das entrevistas fornece exemplos interessantes sobre estes processos. Repare-se na seguinte passagem:

78M3D	..	1559 [72:22.0]	1560 [72:24.4]	1561 [72:26.2]
Fal78- revisor	s a b i a,	•• que era aquilo, que se- não , eu nunca teria dito,	•••	
1ª transcrição		que era aquilo, precisamen- te , eu nunca teria dito		

As diferenças entre *que senão* e *precisamente* são acentuadas a nível fónico, mas podem não o ser tanto a nível semântico. Para a transcritora (que podia ouvir quantas vezes quisesse aquele fragmento) essa imprecisão não era geradora de conflitos de interpretação da frase e por isso aceitou a sua suposta “audição”.

78M3D	123 [05:05.8]
Fal78-revisor	não foi um um teste surpresa que dei ou qualquer coisa assim do género e o ra/ vejo
1ª transcrição	não foi um um teste surpresa que dei ou alguma coisa assim do género e o ra/ vejo

O mesmo para a percepção de *qualquer coisa* como *alguma coisa*. A pouca diferença semântica não impede a construção do modelo que vai traduzindo a frase a ser processada.

Este funcionamento da oralidade demonstra como não “precisamos” de perceber todas as palavras. Os nossos mecanismos cognitivos do processamento da linguagem funcionam um pouco como os da visão. A fóvea da retina (única parte que percebe com nitidez o que varre) não tem tempo de secanear a totalidade do espaço e por isso o cérebro completa a imagem mesmo sem os dados sensoriais completos. Assim, também cada ouvinte completa com aquilo que acha que compõe o quadro global, tal como a vista (fóvea) faz para a visão, compõe os dados que faltam ou que não tem tempo de processar completamente.

3.2.4. *A interação emotiva é fundamental...*

A sua dimensão argumentativa, juntamente com a visão do *significado corporizado*² na perspectiva da Semântica Cognitiva, evidenciam a importância da dimensão emocional da linguagem.

Ora muito mais do que no registo escrito, só num corpus oral é possível verificar a realidade desta dimensão.

Não será aqui o espaço, nem é a finalidade deste texto, para uma reflexão suficientemente profunda sobre a importância e funcionamento dos elementos ligados à emotividade nas línguas. Apenas se pretende fazer ressaltar como é através de *corpora* da oralidade que a análise linguística conseguirá obter dados que possibilitem a referida análise de uma forma mais completa e adequada.

Restringindo-nos apenas ao plano das chamadas *partes do discurso*, embora o projeto de pesquisa ainda não tenha terminado, os elementos recolhidos permitem desde já fazer ressaltar a importância que deve ser dada aos elementos marcadores de expressividade (muito mais vastos que as interjeições tradicionais). A constatação disso mesmo comprova-se pelo facto de se ter sentido a necessidade

² Lakoff [1995]; Lakoff e Johnson [1999]; Gibbs [2003].

de alterar as regras de transcrição relativas à sua recolha, de tal modo que tais elementos expressivos não passassem despercebidos como usualmente acontece.

Até porque, na análise tradicional, há toda uma classe lexical globalmente menosprezada que costuma ser vista como de “não palavras”:

[interjections are] little words or ‘non-words’ which in terms of their distribution can constitute an independent non-elliptical utterance by themselves and do not normally enter into construction with other word classes, but which can be used conventionally as non-elliptical utterances by themselves to express a mental attitude or state [Ameka, 1994: 1713].

Não admira, por isso, que quando se procura transcrever por palavras “a oralidade” frequentemente se deixa “de fora” o que se entende não ser transcrevível por palavras. Na verdade, muita informação emotiva do discurso não é fácil de transcrever. E perante esta dificuldade tradicional em lidar com estas *não-palavras*, no projeto PSFB tentamos transcrever ou assinalar todos os elementos emotivos. E se muitos estão suficientemente lexicalizados para não apresentarem dificuldade de aceitação e transcrição como elementos do discurso (*pumba!*, *ah!*, *ui!*) outros não têm forma gráfica estável e por isso é costume nem sequer serem assinalados. Procuramos, depois de algum debate, uniformizar a transcrição destes elementos (primeiramente assinalados de variadíssimas formas diferentes) com o marcador ((onomatopeia)) de modo a que os investigadores que se queiram servir do corpus possam, com alguma facilidade, encontrar estes elementos.

Veja-se, como exemplo, o seguinte excerto:

78M3D	183 [08:22.3]	184 [08:24.1]	185 [08:25.9]	186 [08:26.9]	187 [08:29.4]
Ent2		•• O quê?			Ai, meu Deus!
Fal78	•• Ele bateu no pai. ••		((estalidos))	•••	E eu disse: – Agora
corretor			((onomatopeia))		

O que foi primeiramente assinalado como ((estalidos)) corresponde a um som produzido um pouco retroflexamente na zona alveolar (algo como *nthe-nthe*). Não há nenhuma unidade lexical que o traduza e por isso numa transcrição tradicional ele passa despercebido. No entanto, no enquadramento discursivo, este elemento possui um alto valor interacional na relação LOC-ALOC:

Falante 78: – *Ele bateu no pai.*

Entrevistadora: – *O quê?*

Falante 78: – *nthe-nthe-nthe...*

Entrevistadora: – *Ai, meu Deus!*

O *nthe-nthe-nthe...* segue-se a um *O quê?* bastante exaltado, mostrando muita admiração da parte da entrevistadora por lhe ter sido comunicado que o filho tinha batido no pai. Perante tal situação, o LOC quer acentuar esse dado invulgar, essa informação que anteriormente já tinha fornecido e simultaneamente dar tempo a que a sua interlocutora assimile a surpresa da informação. Por isso, aquele *nthe-nthe-nthe...* corresponde a “sim, senhor, foi mesmo isto que aconteceu, por incrível que pareça”. O espaço ocupado no diálogo, a altura e intensidade com que estas partes são realizadas, comprovam a importância de todos estes elementos discursivos.

E não se pense que os elementos de interação emotiva são raros ou pouco frequentes no discurso oral. Se se reparar na tabela em anexo 1, que apresenta apenas as palavras mais utilizadas (e não a totalidade) na entrevista 35H4A, comprova-se a bem alta frequência desses elementos tidos como menores.

A palavra(?) mais utilizada pela entrevistadora é mesmo aquele “Hum” de assentimento. Não será de admirar, poder-se-á dizer. Mas dada a sua frequência, seria de admirar que uma Linguística da oralidade achasse que ele simplesmente não existe. Outros elementos como os risos, os *Ah!*, *Oh!*, *Ai!* ocupam igualmente posições de destaque entre as palavras mais utilizadas na conversação, como no mesmo anexo se pode comprovar.

Por tudo isto, pensamos ser essencial que a análise linguística trate adequadamente e evidencie a importância das componentes de

reforço expressivo presentes nas línguas e mais evidenciadas num corpus da oralidade. Estas componentes, a nosso ver, não podem ser tratadas como um todo indiferenciado, nem sustentadas por critérios baseados na sua aceitação ou não através da grafia institucionalizada. Elas terão que ser subdivididas em dois grupos: um constituído pelas onomatopeias expressivas (destinadas sobretudo a ilustrar determinados aspetos fônicos na oralidade) e um outro grupo das exclamações/interjeições, englobando as unidades que possuem valores prioritária e eminentemente emotivos. A Figura 3 poderá ser um esquema muito simplificado do grupo.

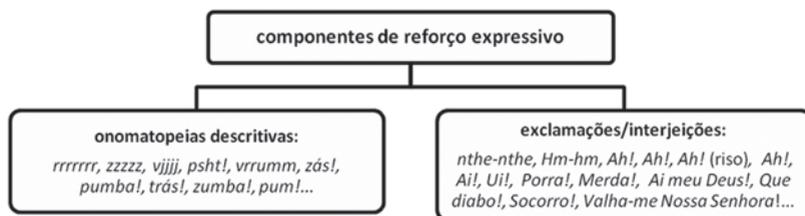


Figura 3

As subdivisões no interior de cada subgrupo, poderão ser variadas: onomatopeias de movimento rápido (*vjjjjt!*, *zás!*), de contacto violento (*trás!*, *tchan!*), de queda (*tumba!*), de discurso (*blá-blá*), etc. Por outro lado, nas exclamações, as tradicionais de dor, medo, alegria, admiração, ironia, desprezo, e outras que se considerarem pertinentes.

4. As conclusões a que um corpus oral deve levar

Poderão parecer lapalisseanas, mas há evidências que a Linguística tradicional ainda não destacou suficientemente, nem delas tirou as devidas consequências para a perceção e descrição dos reais processos de funcionamento das línguas.

Escrita e oral não são realizações iguais da língua. O registo oral utiliza variados elementos morfológicos dotados de grande carga emotiva na relação LOC-ALOC que a escrita e a análise gramatical tradicionais não representam nem valorizam.

As recolhas para a constituição de *corpora* da oralidade em Português permitirão construir bases de dados fundamentais para se perceber o funcionamento da língua real, de consequências variadas, desde uma melhor descrição linguística até ao fornecimento de dados e usos reais que permitam a construção de materiais didáticos para o seu ensino como L2.

Poderão, igualmente, possibilitar que o Português seja uma das línguas tecnologicamente aptas para ser incorporada nas futuras utilizações por máquina das línguas naturais e permitir que ocupe a sua posição no mercado global das indústrias das línguas.

Referências bibliográficas

- AMEKA, F. (1992), “Interjections: The universal yet neglected part of speech”, *Journal of Pragmatics*, 18, Amsterdam, pp. 101-118.
- AMEKA, F. (1994), “Interjections”, em: Asher, R. E. (ed.), *The Encyclopedia of Language and Linguistics*, vol. 4, Pergamon Press, Oxford, pp. 1712-1714.
- ARAÚJO CARREIRA, M. H. (ed.) (2012), *Les rapports entre l’oral e l’écrit dans les langues romanes*, Université Paris 8 Vincennes-Saint-Denis, Paris.
- GIBBS Jr., R. W. (2003), “Embodied experience and linguistic meaning”, *Brain and language*, 84, Edmonton, pp. 1-15.
- JOHNSON, D. J. (1993), “Relationships between oral and written language”, *School Psychology Review*, 22, 4, Bethesda, pp. 595-609.
- LAKOFF, G. (1995), “Embodied Minds and Meanings”, em: Baumgartner, P., Payr, S. (eds.), *Speaking Minds – Interviews with Twenty Eminent Cognitive Scientists*, Princeton University Press, Princeton.

- LAKOFF, G., JOHNSON, M. (1999), *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*, Basic Books, New York.
- NORRICK, N. R. (2009), “Interjections as pragmatic markers”, *Journal of Pragmatics*, 41, Amsterdam, pp. 866-891.
- SAUSSURE, F., de (1978), *Curso de Linguística Geral*, D. Quixote, Lisboa.
- WIERZBICKA, A. (1992), “The semantics of interjection”, *Journal of Pragmatics*, 18, 2-3, Amsterdam, pp. 159-192.
- WILKINS, D. P. (1992), “Interjections as deictics”, *Journal of Pragmatics*, 18, 2-3, Amsterdam, pp. 119-158.

ANEXO 1: palavras mais utilizadas na entrevista 35H4A

Entrevis-tadora	total palavras	falante 35H4A	total palavras
Hum	302	A	239
que	222	É	214
é	157	não	211
o	139	que	172
não	116	para	168
Mas	114	o	163
a	92	de	126
de	66	e	93
senhor	65	era	80
os	57	gente	73
e	51	já	70
em	44	bem	55
então	43	os	54
era	42	um	50

Entrevis-tadora	total palavras	falante 35H4A	total palavras
como	36	eu	49
risos	35	se	49
se	33	sei	47
mais	32	mais	45
assim	31	agora	44
há	31	em	43
Sim	31	Está	43
Olhe	30	isso	42
seus	28	casa	41
também	27	ela	41
ela	25	só	39
para	25	me	37
por	25	por	36
as	23	da	35
filhos	22	uma	31
acha	20	depois	30
Ah	20	mas	30
foi	20	tem	30
ao	19	Ai	29
Braga	19	do	28
do	19	nada	28
gente	19	com	26
aqui	18	foi	25
dia	18	ao	24
pessoas	18	sempre	24
sua	18	hesitação	23

Entrevis-tadora	total palavras	falante 35H4A	total palavras
exemplo	17	eles	22
tem	17	há	22
já	16	ainda	21
na	16	aqui	21
ou	16	Oh	21
quando	16	as	20
lá	15	((incompreen	20
no	15	tenho	20
Pois	15	tinha	20
porque	15	Braga	17
agora	14	ir	17
coisa	14	Isso	17
eu	14	na	17
fazer	14	nunca	17
tinha	14	tudo	17
muito	13	à	16
um	13	coisa	16
está	12	lembra	16
uma	12	no	16
alguma	11	são	16
bem	11	tempo	16
casa	11	ver	16
fazia	11	dos	15
hoje	11	havia	15
isso	11	lá	15
com	10	minha	15

Entrevis-tadora	total palavras	falante 35H4A	total palavras
da	10	muito	15
estão	10	nos	15
hesitação	10	trabalhava	15
lembra	10	vai	15
menos	10	fazer	14
pais	10	fazia	14
eles	9	Ah	13
gosta	9	ali	13
ir	9	nem	13
muita	9	onde	13
OK	9	resto	13
às	8	ele	12
depois	8	outro	12
dos	8	seiras	12
filha	8	às	11
forma	8	ossos	11
jovens	8	outros	11
lhe	8	pessoas	11
me	8	posso	11
mesmo	8	Quando	11
ter	8	Rua	11
Ai	7	trabalho	11