

## Artigos

### Fala e escrita: meio, canal, consequências pragmáticas e linguísticas Speech and writing: medium, channel, and pragmatic and linguistic consequences

Tommaso Raso \*

**RESUMO:** Este artigo debruça-se sobre a comparação entre as modalidades escrita e falada da linguagem com o objetivo de mostrar a necessidade de a linguística adotar uma abordagem nova, tanto na metodologia quanto nas categorias de análise, para compreender como a fala espontânea se estrutura. Através de exemplos de contexto natural extraídos de *corpora*, se colocam algumas questões relativas ao estatuto da fala e da escrita, antes de se examinar as principais diferenças pragmáticas nas duas modalidades, claras consequências das diferenças de meio e canal; uma seção dedicada à função da prosódia introduz uma discussão sobre as unidades de referência da fala e da escrita e seus diferentes valores comunicativos. Por fim, se avançam propostas para uma abordagem empiricamente embasada para o estudo da fala, com algumas observações sobre as características das novas modalidades tecnológicas aplicadas ao uso da linguagem.

**PALAVRAS-CHAVE:** fala e escrita; pragmática; ilocução.

**ABSTRACT:** This article relies on the comparison between speech and writing, aiming to show the necessity of a new approach in methodology and in analytical categories in order for spontaneous speech structure to be understood. Through examples from *corpora* of natural context texts, some aspects related to the status of speech and writing are discussed; afterwards the main pragmatic differences between the two modalities of language are examined, and are attributed to clear consequences of differences in medium and channel. A section dedicated to the function of prosody introduces a discussion of the different reference units in speech and writing and their different communicative status. Finally, we propose an innovative approach in order to study speech under an empirical perspective, with some observations on the characteristics of new technologies applied to language use.

**KEYWORDS:** speech and writing; pragmatics; illocution.

#### 1. Fala e escrita: uma distinção óbvia?

Se eu pudesse perguntar ao leitor desse artigo se o exemplo 1 constitui um texto escrito ou um texto falado, quase certamente a resposta da maioria seria que se trata de um texto falado.

Exemplo 1

*mas os filho também nũ são fácil também juntou os filho todo foram lá e trouxeram o corpo na força*

---

\* Professor Titular em Linguística na UFMG e bolsista do CNPq.

A resposta provavelmente seria argumentada assim: a falta de concordância entre *os, filho* e *fácil*, e em *os filho todo* é própria da fala, mas não da escrita; a forma *nũ* não é uma forma ortográfica apropriada; o uso de *também* repetido é típico da fala; a forma *juntou os filho todo*, ao invés de *se juntaram*, também é um sinal de que o texto é falado.

E se eu pudesse perguntar como segmentar a sequência, a resposta mais provável seria a seguinte:

Exemplo 1a

*mas os filho também nũ são fácil também. Juntou os filho todo, foram lá e trouxeram o corpo na força.*

Agora, se o que foi dito acima corresponder realmente ao que a maioria dos leitores do artigo consideraria, algumas observações interessantes podem ser feitas.

A primeira é relativa ao quê nos guia na decisão sobre se um texto é falado ou escrito. Nesse caso, a decisão seria guiada por fatores de ordem diafásica e diastrática. De fato, a falta de concordância, assim como a forma recíproca sem o pronome, são frequentes nos textos escritos de autores com controle limitado do registro (ainda que não seja raro entre alunos universitários) ou em textos escritos em situações comunicativas informais. Ou seja, não se trata de traços próprios da diamesia<sup>1</sup> falada, mas de traços amplamente observáveis tanto em textos falados quanto em textos escritos, como uma pesquisa em *corpora* poderia facilmente mostrar. Ainda poder-se-ia afirmar que a forma não ortográfica *nũ* deixaria claro que estamos observando um texto falado. Mas, mesmo nesse caso, poderíamos rebater que as paródias gráficas são muito frequentes em uma ampla variedade de textos escritos, como os textos de *chat* ou de *SMS*, ou textos de caráter privado, e até textos literários, como por exemplos a produção futurista do começo do século XX ou outras vanguardas ao longo do último século.

A segunda observação é ainda mais interessante. Se a resposta à pergunta for que se trata de um texto falado, essa conclusão se baseou na leitura. Isso significa que quem respondeu assim, o fez, mesmo que inconscientemente, achando que a atribuição de uma sequência linguística à diamesia falada possa ser feita através do canal visual. Poder-se-ia objetar que quem respondeu assim obviamente subentendia que o exemplo 1 é a *transcrição* de um texto falado. Apesar de termos mostrado que os traços que guiam a decisão são perfeitamente

---

<sup>1</sup> Berruto adota esse termo (em analogia com os mais conhecidos *diacronia*, *diatopia*, *diastatia*, *diafasia*) para designar a variável de meio e canal como fator que marca a variação sociolinguística. Veja-se Berruto (1987), (1993a), (1993b); Rossi (2001).

compatíveis com textos escritos, nesse caso se trata realmente da transcrição feita a partir de um texto falado. O texto falado pode ser escutado no áudio [bfamnm03-20]. Contudo, uma transcrição, por mais que possa ser útil para fixar e assim estudar alguns aspectos do texto produzido através da fala, constitui sem sombra de dúvida um texto escrito. O texto escrito pode ter como fonte um texto falado, mas se for percebido através do canal visual não deixa de ser um texto escrito, com todas as propriedades do meio e do canal da escrita. Uma primeira demonstração disso é confrontar a segmentação feita no exemplo 1a com a percepção do texto através da escuta do original falado. De fato, os segmentadores do *corpus* do qual esse exemplo foi extraído segmentaram o texto da maneira seguinte:

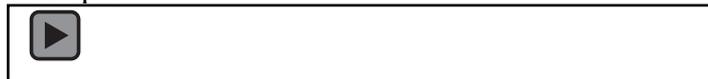
#### Exemplo 1b<sup>2</sup>



*mas os filho também / ãũ sãõ fácil também / juntou os filho todo / foram lá e trouxeram o corpo na força //*

Voltaremos a explicar e discutir melhor a segmentação da fala mais à frente. No momento observamos somente que a sequência que visualmente parecia facilmente segmentável em dois períodos, sendo o primeiro simples e o segundo composto de três coordenadas, à escuta foi segmentada em um único enunciado constituído por quatro unidades tonais (KARCEVSKY, 1931; CRYSTAL, 1975). A barra dupla indica a fronteira de enunciado, dada por uma quebra prosódica com valor conclusivo de enunciado, enquanto a barra simples indica a fronteira de unidade tonal interna ao enunciado. De fato, a escuta das partes da sequência não permite perceber como uma unidade completa nada antes do ponto em que foi colocada a quebra terminal. Isso pode ser verificado escutando-se o áudio dos exemplos seguintes:

#### Exemplo 1c



*mas os filho também / ãũ sãõ fácil também /*

---

<sup>2</sup> Todos os exemplos são extraídos do *corpus* C-ORAL-BRASIL (RASO & MELLO, 2012), com a exceção do último que é extraído do *corpus* C-ORAL-ROM (CRESTI & MONEGLIA, 2005) no seu sub-*corpus* português.

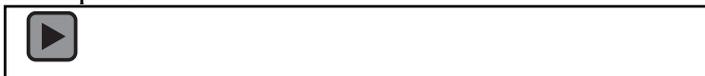
## Exemplo 1d



*mas os filho também / nũ são fácil também / juntou os filho todo /*

Os exemplos 2-4 mostram outros casos em que a segmentação que parece natural na escrita não necessariamente é aquela que corresponde à realização da fala.

## Exemplo 2



*\*BAL: tá saindo de uma garrafinha que tem um bico muito pequeno //*

## Exemplo 3



*\*BAL: cê tá com um jarro d'água // que tem uma espessura assim //*

Os exemplos 2 e 3 parecem ter a mesma estrutura sintática: uma sentença com um verbo principal, um objeto indireto e, dependente dele, uma relativa. Mas se ouvimos a sequência não podemos negar que a análise sintática não captura a evidente diferença estrutural entre os dois exemplos. Enquanto a análise permanece válida em 2, em 3 mostra duas unidades distintas, ambas interpretáveis autonomamente. Certamente não podemos dizer isso em 2, como mostram os exemplos seguintes:

## Exemplo 2a



*tá saindo de uma garrafinha*

## Exemplo 2b



*que tem um bico muito pequeno*

## Exemplo 3a



*cê tá com um jarro d'água*

## Exemplo 3b



*que tem uma espessura assim*

Ouvindo 2, temos que concluir que a parte que se sustenta sozinha é a segunda, ou seja, a parte que sintaticamente é analisada como dependente. A primeira parte, que é preenchida por uma estrutura sintática autônoma, certamente não resulta interpretável à escuta. Veremos como explicar isso mais à frente.

O exemplo 4 poderia ser analisado sintaticamente como um verbo seguido por um SP argumental e por seu adjunto, também um SP. Mas escutando-o, observamos que essa estrutura pode ser dividida em duas unidades autônomas, a primeira sendo representada pelo verbo com seu argumento, e a segunda pelo simples adjunto:

## Exemplo 4



\*SIL: *tava no jornal // de ontem //*

## Exemplo 4a



*tava no jornal*

## Exemplo 4b



*de ontem*

Antes de terminar a seção apresentamos ainda o exemplo 5

## Exemplo 5



\*PAU: *Não tá dando a altura daquele que a Isa marcou lá né*

Nesse caso, a leitura da transcrição nos permite interpretar o exemplo como um enunciado negativo. Podemos também imaginar que haja uma outra possibilidade, ou seja, que o *não* deva

ser considerado à parte e não um elemento composicional com o verbo. Mas somente através da escuta saberemos que a hipótese certa é

\*PAU: Não // tá dando a altura daquele que a Isa marcou lá / né //

## 2. Meio e canal

Todos os exemplos da seção 1, que poderiam facilmente ser multiplicados, mostram que a fala é interpretável corretamente somente através de algo que é veiculado pelo som. Ao contrário, a tentativa de analisar o texto falado com base na transcrição, ou seja, em uma forma de transposição para a escrita, gera enganos, o que é pior do que a falta de interpretabilidade.

Nessa seção começaremos a explicar o porquê disso, analisando as diferenças de meio e canal entre as duas modalidades: a fala e a escrita.

As principais diferenças de meio já apareceram claramente: o meio da fala são as ondas sonoras que são realizadas através do aparelho fonador e decodificadas através do aparelho auditivo humano. O meio da escrita é constituído por símbolos gráficos decodificados através dos olhos. A não ser em um uso metafórico da expressão, nós não podemos “ver” a fala.

O canal da fala é o ar, pelo qual se propagam as ondas sonoras. As partículas de ar são empurradas, com maior ou menor energia, para longe da fonte, ou seja, para longe do falante. O canal da escrita pode ser constituído por diferentes suportes mais ou menos estáveis: pedra, madeira, papel e outros.

Uma primeira diferença entre as duas modalidades é entre o caráter natural da fala e o caráter tecnológico da escrita. A fala é uma capacidade da nossa espécie, e talvez todas as línguas do mundo tenham a mesma origem (RUHLEN, 1994). Dependendo do que se entende exatamente por fala, a sua origem pode ser antecipada ou retardada, mas colocar a origem da fala, assim como a conhecemos, por volta de 100.000 anos atrás não constitui certamente uma datação exageradamente antiga (MACMAHON & MACMAHON, 2013). Como ela surgiu é amplamente debatido, mas as descobertas das neurociências, da paleontologia, da genética e de outras disciplinas, além da linguística, podem nos ajudar a formular e testar hipóteses (MACMAHON & MACMAHON, 2013; McNEILL, 2012; ARBIB, 2012). Contudo, este não é o objetivo desse trabalho. De fato, a capacidade da fala é uma dotação com a qual nascemos (juntamente com várias outras capacidades motoras e cognitivas) e que começa a se realizar muito cedo, desde que sejamos inseridos no meio social. O processo de aquisição completa da fala dura alguns anos, mas com poucos anos de vida a criança domina essa modalidade, sem

que isso impeça que ela seja enriquecida durante toda a vida. Trata-se enfim de uma característica da espécie, que só não se realiza plenamente em indivíduos com alguma patologia específica ou que, em casos muito excepcionais, são tirados muito cedo da convivência com outros seres humanos.

Ao contrário, a escrita surge apenas cerca de 5.500 anos atrás, em um contexto histórico completamente novo, em que as exigências próprias de um organismo social com características novas motiva a invenção de uma tecnologia que ajude a gerenciar uma complexidade social não experimentada antes (POWELL, 2009; FISHER, 2005; GOODY, 1986). O mesmo pode-se dizer da origem de qualquer outra tecnologia, seja ela um meio de comunicação ou não. De fato, a escrita pressupõe a existência da fala, assim como o desenvolvimento de tecnologias da comunicação mais complexas (imprensa e internet, por exemplo) pressupõem a existência da escrita.

Também não é objetivo desse trabalho aprofundar as razões sócio-históricas que originaram a escrita. Mas duas considerações devem ficar claras: a escrita não é diretamente um produto da *evolução*, como a fala, se tomamos o significado próprio dessa expressão, assim como é usado na biologia, e não em um sentido metafórico, como quando adaptamos o termo *evolução* a percursos de natureza histórico-cultural. A escrita de fato surge em uma época muito recente, inicialmente em pouquíssimos indivíduos (não em uma espécie) e ainda hoje nem todas as comunidades conhecem essa tecnologia, assim como muitos indivíduos em todas as comunidades, sem por isso serem indivíduos portadores de alguma patologia. Vale a pena resumir alguns passos que levaram à invenção da escrita para contextualizá-la como tecnologia; a invenção da escrita se explica com base em necessidades sociais, assim como outras tecnologias, seja a roda, o motor, o computador, as armas de fogo ou a bomba atômica. Todas têm em comum o fato de serem produtos da evolução cultural e de terem sido motivadas pelas exigências sociais da humanidade frente a problemas novos. Todas têm em comum o fato de não serem prerrogativa de todos os indivíduos não patológicos. A diferença está apenas no fato de a escrita ou a roda serem mais antigas e mais difundidas do que, progressivamente, o motor, a bomba atômica e o computador.

A escrita surgiu na Mesopotâmia, provavelmente com os Sumérios, onde se criou a primeira organização social tão complexa que mereceu o nome de estado. Os problemas da gestão de um estado gerou a necessidade, nunca sentida antes, de possuir uma tecnologia que permitisse à linguagem sobreviver ao tempo e ao espaço. Os estados, por sua vez, puderam

surgir graças a tecnologias anteriores, que permitiram que grandes quantidades de pessoas vivessem juntas em um mesmo lugar. O grande salto que permitiu isso foi a descoberta e a complexificação da agricultura, que se deram a partir de cerca de 13.000 anos atrás, sempre na Mesopotâmia e pouco depois na China, por razões ambientais casuais (DIAMOND, 1997). A agricultura, que surgiu autonomamente em vários lugares, foi progressivamente se estendendo, mas ainda hoje há comunidades pré-agrícolas em lugares isolados, e naturalmente sem escrita. A complexificação social trazida por agriculturas fortes gerou sociedades mais numerosas e complexas, que, por sua vez, sentiram a necessidade de desenvolver tecnologias para gerenciá-las, entre as quais a escrita. Várias formas de escrita surgiram autonomamente em lugares diferentes.

Uma outra importante característica da diferença entre uma dotação natural como a fala e uma tecnologia adquirida historicamente como a escrita (e tantas outras) é que nós não precisamos de treinamento específico para aprender a falar, mas precisamos de um forte treinamento para aprender a escrever. A fala, portanto, precede a escrita não somente do ponto de vista filogenético, mas também daquele ontogenético.

### **3. As consequências pragmáticas**

As diferenças de meio e de canal, que, como vimos, são radicais, produzem consequências pragmáticas muito diferentes. As principais são sem dúvida as diferenças relativas às categorias de tempo e espaço.

#### **3.1. Na fala**

Já que o meio da fala são as ondas sonoras e o canal é o ar, e já que as ondas sonoras possuem uma duração muito curta, a fala deve ser percebida pelo interlocutor ao mesmo tempo em que ela é realizada pelo falante. Aqui nos preocupamos com a descrição da modalidade primária da fala, ou seja, a fala espontânea, definida como “a fala realizada ao mesmo tempo em que é planejada” (NENCIONI, 1983). Na fala espontânea, o falante tem pouquíssimo tempo para planejar, assim como o ouvinte tem pouquíssimo tempo para prestar atenção no que ele ouviu. Isso porque a sucessão de enunciados é muito rápida e um enunciado não pode ser recuperado mais tarde. Por essas razões a fala possui uma forte vinculação ao *aqui e agora*. Essa vinculação se manifesta também na forma de vinculação recíproca do tempo do falante com o tempo do ouvinte. O primeiro efeito dessa vinculação recíproca é o fato de que o ouvinte

deve esvaziar rapidamente a memória de trabalho para podê-la utilizar novamente durante a interação. Isso faz com que:

- a) somente a unidade de significado básica seja estocada em uma memória mais duradoura, não a formulação linguística complexa com a qual a unidade de significado foi formulada;
- b) se se perdem um ou mais enunciados (por qualquer forma de ruído, interno ou externo ao ouvinte), eles não serão mais recuperáveis. Existem formas de controle parcial disso, como o feed-back, mas nunca será recuperável o mesmo enunciado perdido; o que pode ser recuperada é a unidade de significado, veiculada através de um outro enunciado;
- c) o silêncio na fala constitui um problema que coloca em risco a comunicação. De fato, um silêncio de apenas poucas dezenas de segundos (e frequentemente até menor) gera um mal estar nos ouvintes, o que produz reações que variam segundo as circunstâncias: ou a tentativa de interromper o silêncio e reestabelecer o fluxo de fala, ou o abandono do canal de comunicação em várias formas (os ouvintes começam a falar entre si, abandonam o lugar da comunicação, tomam iniciativas devidas à preocupação que o falante não esteja passando bem, etc.). Enfim, o falante deve constantemente se preocupar com a manutenção do canal comunicativo evitando o silêncio. Isso faz com que manter a fala seja mais importante do que elaborá-la de maneira considerada ideal ou satisfatória, já que o tempo necessário para isso pode levar à perda irreversível da comunicação;
- d) a necessidade de evitar o silêncio leva frequentemente à produção de enunciados interrompidos, *retractings*<sup>3</sup>, repetições, hesitações, estruturas defeituosas, enfim a vários tipos de disfluências, estruturando também a locução da maneira que Halliday (1989) descreve com muita propriedade e chama de *baixa densidade* (menor variação lexical, mais repetições dos mesmos lexemas, mais palavras funcionais, mais palavras de alta frequência, etc.). O falante não tem tempo para organizar a fala de outra maneira. A altíssima frequência desses fenômenos não nos permite considerá-los como acidentes de percurso. Eles são uma consequência própria das condições pragmáticas da fala espontânea. Lendo uma transcrição de um monólogo, até de um monólogo que à escuta

---

<sup>3</sup> Por *retracting* entende-se um tipo de disfluência que leva à correção ou repetição dentro do enunciado. Nos exemplos desse artigo o *retracting* é indicado com uma barra simples (já que constitui uma quebra) dentro de colchetes; a barra é seguida pelo número de palavras “retratadas”, como por ex. em \*ROG: [4] aqui já tá dando [/4] aqui já tá dando a altura //.

seja especialmente apreciável pela propriedade de linguagem, é impossível não ter uma sensação de inadequação e estranhamento que tornam a leitura extremamente difícil se não impossível. Mas o ouvinte, por sua vez, deve elaborar rapidamente o que escutou para poder estar imediatamente pronto a escutar mais. Por isso, ele não tem tempo para observar as interrupções, as hesitações e as estruturas defeituosas do falante, que, portanto, não o incomodam (dentro de certos limites) e frequentemente até lhe permitem ganhar mais tempo precioso para processar o conteúdo veiculado pelo falante.

O vínculo recíproco entre falante e ouvinte, e o vínculo ao *aqui e agora* é evidente também com relação ao espaço. As características de meio e de canal não permitem que a fala alcance distâncias significativas. Isso não torna necessário que falante e ouvinte estejam um na frente do outro, mas que eles ocupem um espaço de alguma maneira compartilhado no momento da fala. Entre as consequências disso podemos observar que a) falante e ouvinte compartilham o contexto situacional, e a interação é fortemente condicionada e guiada por ele: quem fala sabe para quem está falando e quem ouve sabe quem é o autor da fala; b) a distância entre os interlocutores condiciona o uso do meio; c) a presença/ausência de pessoas ou outras entidades condicionam o tipo de referência, tornando extremamente frequente o uso de expressões dêiticas cuja referência é dada imediatamente pelo compartilhamento do contexto situacional.

### **3.2. Na escrita**

Na escrita as categorias de tempo e espaço agem de maneira completamente diferente, como consequência da diferença no meio e no canal. O meio é um sistema simbólico que é tornado perceptível visualmente e o canal é qualquer suporte para o meio. Cada suporte possui suas próprias características, mas em princípio as características básicas do suporte são a durabilidade e, mesmo se não sempre, a transportabilidade. Essas duas razões tornam a tecnologia vantajosa. Assim, as informações (ou qualquer conteúdo linguístico) podem sobreviver ao momento de emissão e ser recuperadas em outros momentos e outros lugares. Alguns suportes chegam a durar séculos ou até milênios. Outros têm menor durabilidade, mas podem ser copiados (à mão, ou utilizando outras tecnologias). O conteúdo da escrita pode também ser transportado no espaço, sem que haja algum limite a não ser aquele devido a suportes específicos (como um muro).

Mas as vantagens evidentes da tecnologia tornam o comportamento da escrita muito diferente daquele da fala. A escrita não é uma atividade *online*, como a fala espontânea. É uma atividade cujo efeito é retardado. Se a fala se realiza no mesmo momento em que a concebemos, a escrita é o resultado de um processo. O processo é a própria ação de escrever, que pode ser mais lenta ou mais rápida dependendo dos instrumentos utilizados e da perícia do escritor, mas nunca é fruto imediato e natural da afetividade (no sentido pulsional) do autor. Existe sempre pelo menos uma mediação tecnológica entre o que o escritor quer transmitir e o produto da transmissão.

Além disso, o tempo do escritor e o tempo do leitor são independentes. Quem escreve o faz com o próprio tempo, e quem lê o faz com um tempo que não depende do tempo da escrita. Quem escreve não sabe quem lerá seu texto, mesmo se em algumas circunstâncias e dentro de certos limites pode ter uma previsão disso. Pode demorar horas ou anos na escrita, mas isso não aparecerá ao leitor. Pode apagar e reescrever o quanto quiser, mas o leitor não saberá nada desse processo. Por outro lado, o leitor, frequentemente desconhecido e imprevisível ao escritor, lerá em algum momento sucessivo àquele em que o produto textual inteiro foi licenciado. Ele poderá saber algo do escritor, mas normalmente porque foi informado e confia nas informações, não por ter durante o seu momento da leitura a percepção imediata do escritor. Ele poderá ler o texto de uma vez só ou ao longo de muitas sessões de leitura separadas até por anos. E, assim como o escritor pode voltar atrás e refazer o texto eliminando o que foi feito antes, o leitor também pode reler partes do texto já lidas ou até ler o texto de maneira não sequencial.

É importante entender que enquanto falante e ouvinte estão dentro de um *processo* em curso, escritor e leitor respectivamente constroem e usufruem de um *produto*, de maneira independente um do outro. A interação entre falante e ouvinte é dada pelo próprio processo da fala e da escuta no *aqui e agora*. A eventual interação através da escrita se dá por uma sucessão de produtos. A expressão *interatividade* acaba, portanto, tendo significados diferentes nas duas modalidades. Não se trata da mesma atividade. As novas tecnologias apresentam formas de trocas através da escrita que aumentam a interatividade. A típica forma interativa da escrita é o *chat*. Mas aqui também o destinatário é obrigado a aguardar que quem está escrevendo aperte a tecla *enter* do computador para que apareça, de uma vez só, o produto de quem fez o processo da escrita, seja um conjunto de frases, seja apenas uma interjeição. Antes desse momento o escritor pode desfazer, corrigir ou desistir do conteúdo que escreveu. Voltaremos a essa diferença entre a fala/processo e a escrita/produto mais à frente.

### 3.3. A prosódia

Nas seções anteriores deixamos de lado um fator muito importante de diferenciação entre a fala e a escrita: a presença da prosódia na fala. Sendo a prosódia um conjunto de parâmetros acústicos, é evidente que se trata de uma característica própria da fala. A prosódia, com seus componentes fundamentais (duração, intensidade, variação de F0, pausa, ruptura ou quebra, alinhamento silábico, velocidade de fala, ritmo) constitui o principal veículo de estruturação da fala. Já observamos na seção 1, de maneira apenas perceptual, como a prosódia de fato segmenta o fluxo da fala de maneira independente das estruturas sintáticas que observamos à leitura, e voltaremos a falar da prosódia na próxima seção. Mas isso por si só já deveria mostrar como é impossível estudar a fala sem acesso a esse componente, que se perde nas transcrições ou que é reduzido, na melhor das hipóteses, a um sistema de marcação insuficiente e enganador.

Na escrita, o suposto correspondente da prosódia é a pontuação. Contudo, o sistema de pontuação obedece a normas que diferem nas várias línguas/culturas, e historicamente têm mudado muito (CRESTI et al., 1992), com forte dependência de contextos e tradições específicos. Certamente não se trata de um recurso natural como a prosódia. Apesar disso, a pontuação desempenha um papel importante na estruturação do texto escrito. O ponto segmenta as sentenças e a vírgula marca uma separação lógica entre sintagmas que não possuem regência direta<sup>4</sup>. Os pontos de interrogação, de exclamação e os pontos de suspensão cobrem várias funções de caráter expressivo.

Escutemos os exemplos seguintes:

#### Exemplo 6a



cê divide comigo empurrar isso

#### Exemplo 6b



cê vai embora que dia Rena

---

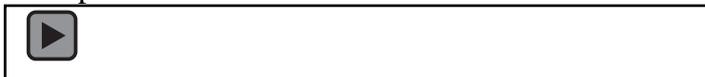
<sup>4</sup> Tanto que os docentes costumam corrigir um uso da pontuação que separa o sujeito do verbo, o que reflete uma interferência da marcação, possível na fala, como tópico de um sujeito.

Exemplo 6c



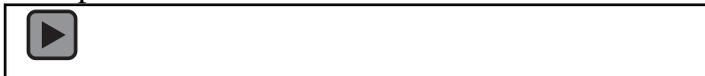
No' tá barato hein

Exemplo 6d



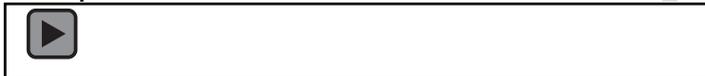
Ô nós nũ vamo levar alguma fruta

Exemplo 6e



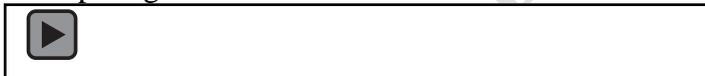
tá cheio mesmo viu

Exemplo 6f



então esse

Exemplo 6g



eh hoje cê tá faminta né filha

Exemplo 6h



que gracinha

Exemplo 6i



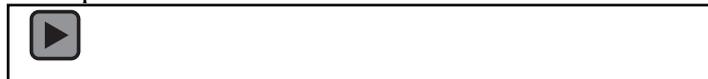
desinfetante não

Exemplo 6l



então desinfetante

## Exemplo 6m



Rena

## Exemplo 6n



brigada moço

É certo que, escutando o áudio, muitos exemplos seriam pontuados no final da mesma maneira, sendo o ponto de interrogação e de exclamação os mais frequentes. É provável que o ponto de interrogação seja o principal (ou o único) para os primeiros seis, enquanto o ponto de exclamação prevaleça nos outros seis. Portanto, os exemplos mostram que funções comunicativas diferentes (ou seja, ações realizadas através da fala ou *ilocuções* diferentes) são pontuadas com o mesmo recurso, e frequentemente fica a dúvida sobre a pontuação a ser preferida. Mesmo escutando o áudio, ou não existiria acordo na pontuação ou mais de uma ilocução, ou seja, mais de uma função comunicativa seria pontuada do mesmo jeito, neutralizando a principal diferença comunicativa. Mas o que é mais importante é que sem o áudio não é possível atribuir um valor comunicativo ao enunciado. Em alguns casos o léxico e a estrutura sintática induzem a hipóteses, mas seria sempre possível usar o mesmo léxico e a mesma estrutura para hipóteses diferentes.

O fato é que a escrita e a leitura não possuem a capacidade de transmitir um valor comunicativo através da sua unidade mínima (a sentença). A unidade comunicativa da escrita e da leitura é o texto como um todo, o *produto*. Para resolver esse problema, na literatura sobre as ilocuções frequentemente se introduz um conjunto de categorias que deveriam resolvê-lo: a modalidade interrogativa, a modalidade imperativa e a modalidade exclamativa (SWERTS & KRAHMER, 2005; MORAES & RILLIARD, no prelo). Essas categorias, contudo, não parecem ter nenhuma função se não aquela de agrupar categorias ilocucionárias de uma maneira que sejam compatíveis com a pontuação ou com a estruturação sintática que elas “normalmente” recebem na escrita. Ou seja, se trata de categorias que não parecem motivadas por outra razão senão pelo fato de que a escrita não tem recursos para fazer as diferenciações necessárias. A nossa posição é que a escrita não precisa desses recursos, pois a unidade mínima

na escrita não possui valor ilocucionário algum. Portanto, essa diferenciação entre as três modalidades nada mais seria que uma tentativa de interpretar a fala a partir de categorias próprias da escrita, ou seja, algo natural a partir de uma tecnologia. Isso também será melhor argumentado nas próximas seções.

Para concluir a seção sobre a prosódia, se escute o exemplo seguinte:

#### Exemplo 7



\*KAT: *o quê* //

\*SIL: *copos // copos de Urano / que tem aí* //

\*KAT: *copos de quê* //

\*SIL: *Urano* //

\*KAT: *Urano* //

\*SIL: *é // Urano // Urano* //

Concentramos a nossa atenção nos últimos 3 turnos. Neles aparece 4 vezes a mesma palavra (*Urano*), formando 4 enunciados autônomos. Trata-se, portanto, de quatro enunciados compostos pelo mesmo conteúdo semântico e pela mesma estrutura sintática: um SN. Se o valor comunicativo fosse em primeiro lugar função do conteúdo semântico e da estrutura sintática, os quatro enunciados deveriam ser puras tautologias. Mas é evidente à escuta que não é assim. Também é evidente que o diferente valor comunicativo que atribuímos a eles é veiculado unicamente pela prosódia. Isso parece nos dizer que a prosódia é a principal responsável por veicular o valor comunicativo na fala. Na próxima seção tentaremos explicar o porquê disso e mostraremos uma outra diferença fundamental entre a fala e a escrita.

#### 4. A unidade de referência

Uma questão importante é: como podemos segmentar a fala e a escrita? Naturalmente podemos segmentá-las de várias maneiras: fone e caractere, palavra, sintagma, etc. Mas como podemos segmentar fala e escrita nas suas unidades de referência, ou seja, nas unidades mínimas que permitam um sentido completo? Uma unidade que seja superior à palavra. A resposta a essa pergunta é muito importante, porque mostra uma outra grande diferença entre as duas modalidades da linguagem que decorre das diferenças mostradas acima e que gera consequências decisivas no estudo das duas diamesias.

Normalmente a sentença é considerada a unidade de referência da escrita. Para a fala a unidade de referência é considerada o enunciado. Uma das possibilidades que se encontram na literatura define o enunciado como sentença falada, ou seja, mantém para o enunciado a mesma concepção que é reconhecida para a sentença. Existem duas definições de sentença: a) aquela que se fundamenta na relação sujeito-predicado e vê uma dependência do primeiro do segundo, realizado normalmente através de um verbo. Essa é a visão pré-chomskiana de sentença<sup>5</sup>; b) aquela que se fundamenta diretamente na estrutura argumental do verbo, ou seja, identifica a sentença como a projeção máxima do constituinte V (CHOMSKY, 1970). Nos *corpora* de fala espontânea C-ORAL-ROM (para italiano, francês, PE e espanhol; CRESTI & MONEGLIA, 2005) e C-ORAL-BRASIL (RASO & MELLO, 2012), a frase, na sua primeira definição, aparece em menos de 5% dos casos, ou seja, abaixo da porcentagem quantitativamente relevante para que um fenômeno seja considerado como típico da fala. Além disso, essas sentenças são caracterizadas por um sujeito pronominal ou, quando o sujeito for mais pesado, por um predicado muito reduzido. O motivo disso é a necessidade de se fazer com que todo o material locutivo caiba em uma única unidade tonal<sup>6</sup>, como no caso do exemplo 1b. Enfim, nada que se compare com o que acontece na escrita, onde não pode existir uma dimensão comparável com a unidade tonal.

No caso da segunda definição de sentença, ou seja, um SV mais ou menos preenchido em todas as suas valências, dependendo da tipologia sociolinguística e diafásica do texto, temos uma frequência muito mais alta, entre 60 e 75%. Mas o fato de que a maioria das produções linguísticas tenha essa configuração não impede que em média mais de 35% delas não apresente verbo nenhum de modo finito, e cerca de 30% nenhum verbo em absoluto, como observado também para o inglês na Longman Grammar (BIBER et al., 1999). Isso significa que mais de um terço das produções linguísticas não corresponde a um SV, com uma frequência que impõe que esse fenômeno seja considerado típico da fala, e que se dê conta dele procurando um outro critério para a identificação da unidade de referência na fala.

---

<sup>5</sup> Para uma discussão sobre o conceito de predicação, veja-se Cresti (2001 e 2005a).

<sup>6</sup> Sobre a relação entre unidade tonal e unidade informacional, veja-se Cresti (2000), Moneglia (2005), Raso (2012). Sobre a relação entre unidade tonal/informacional e sintaxe, veja-se Cresti no prelo. O tamanho de uma unidade tonal é limitado por motivos fisiológicos.

As principais definições de enunciado que não se baseiem na transposição para a fala de uma das definições de sentença são: 1) a definição dialógica; 2) a definição temporal; 3) a definição pragmática<sup>7</sup>.

A definição dialógica se baseia na noção de turno, como intervalo entre um silêncio e outro do mesmo falante, seguido pela fala de outro falante ou pela conclusão da interação. O turno é sem dúvida uma unidade natural da fala, mas a dimensão dela é extremamente variável. Em diálogos com forte interatividade, o turno é extremamente curto, chegando a ser realizado com uma única unidade lexical, até mesmo uma interjeição. Em trocas pouco interativas, principalmente em tipologias monológicas, o turno pode durar até muitos minutos. Não é pensável, portanto, que não exista uma unidade da fala intermediária entre a palavra e o turno e que possa ser considerada a sua unidade de referência.

A definição temporal, utilizada, por exemplo, no *Dutch Corpus* (SCHUURMAN, 2003), propõe que o enunciado seja segmentado por pausas, ou seja, uma dimensão do silêncio diferente daquela do final de turno. Os problemas dessa definição de enunciado são vários. Em primeiro lugar, as pausas no fluxo da fala espontânea são raras. Em segundo lugar, o conceito de pausa é um conceito temporal que não possui uma quantificação objetiva. Podemos ir de um mínimo de silêncio que corresponde à fase pré-explosiva da articulação de uma oclusiva, e, portanto, um silêncio certamente interno ao processo de fala, até uma dimensão temporal muito grande. Também, nada impede que exista pausa dentro do enunciado. O fenômeno é muito dependente do tipo de interação e do estilo de cada falante. Como veremos, não existe uma dimensão de pausa que individualize as fronteiras de enunciado e não exclua silêncios internos ao enunciado.

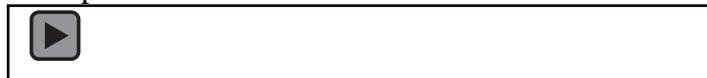
Antes disso, examinamos a definição pragmática de enunciado. Essa proposta (CRESTI, 2000) define o enunciado como "a menor unidade da fala interpretável pragmaticamente (e prosodicamente) em autonomia". Com base nessa definição, o enunciado corresponde a um ato de fala (AUSTIN, 1962). Ele não vai de silêncio a silêncio, mas de uma marca de conclusão a uma marca de conclusão. Essa marca é dada por uma quebra prosódica que veicula um valor percebível como terminal, o que é possível somente se a sequência veicula força ilocucionária. Contrariamente à definição sintática de enunciado, mas de acordo com a definição dialógica e

---

<sup>7</sup> Para uma discussão sobre as diferentes definições de enunciado, veja-se Cresti & Gramigni (2004), Miller & Weinert (1998), Chafe (1988), Moneglia (2006), Moneglia & Cresti (1997 e 2006).

temporal, essa definição é de natureza perceptual. Não se trata, contudo, de uma percepção mensurável, como seria aquela da noção de pausa. Observamos os exemplos 8 e 9:

#### Exemplo 8



\*BAL: tá saindo de uma garrafinha que tem um bico muito pequeno // então daquela coisa pequeninim ã vai encher rápido // agora imagina cê pega um balde e joga dentro //

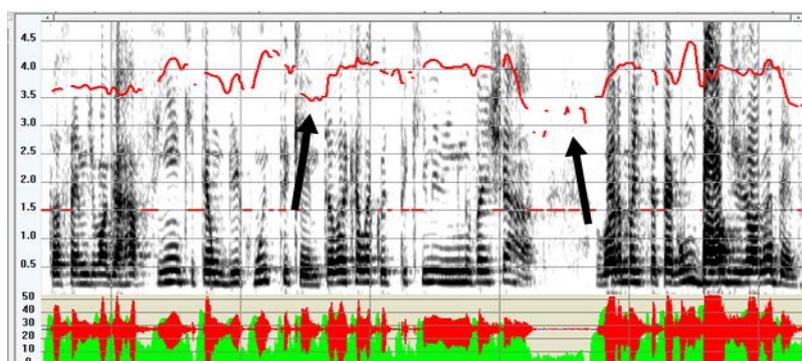
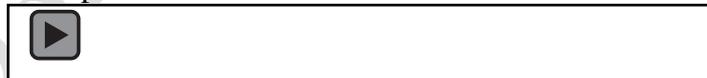


Figura 1. Tela do WinPitch para o Exemplo 8.

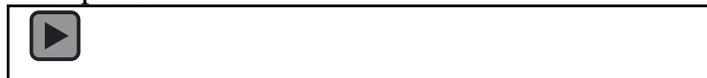
O exemplo 8 mostra uma sequência que foi segmentada em 3 enunciados. Entre o segundo e o terceiro há realmente uma pausa, no caso de 464 ms. Entre o primeiro e o segundo não há pausa alguma, como evidenciado pela figura 1, onde as setas mostram as fronteiras de enunciado. A escuta em isolamento do exemplo 8a e 8b mostra que ambas as sequências podem ser percebidas como autônomas prosódica e pragmaticamente (mas todos os exemplos anteriores se baseiam no mesmo critério).

#### Exemplo 8a



\*BAL: tá saindo de uma garrafinha que tem um bico muito pequeno //

#### Exemplo 8b



\*BAL: então daquela coisa pequeninim ã vai encher rápido //

## Exemplo 9



\*MAR: o ensino tá [tá] tá assim / difícil / mas tá mais fácil / né hhh //



Figura 2. Tela do WinPitch para o Exemplo 9.

O exemplo 9 e a figura 2 mostram um exemplo de uma pausa interna ao enunciado. A pausa, muito maior do que a pausa que separa o segundo e o terceiro enunciado do exemplo 8, não pode constituir fronteira de enunciado, como é evidente à oitiva da parte anterior à pausa.

Segundo Moneglia (2005), no C-ORAL-ROM Francês (que foi segmentado tanto em quebras quanto em pausas a partir de 200 ms) 63% das quebras com valor terminal são acompanhadas por pausas (enquanto 37% das quebras terminais não são acompanhadas por pausas ou são acompanhadas por pausas de duração inferior a 200 ms; por outro lado, 42% das pausas de 200 ms ou mais acontecem dentro de enunciado, em correspondência a quebras não terminais. Um trabalho estatisticamente confiável a esse respeito está sendo feito sobre o C-ORAL-BRASIL, que também foi segmentado com base na definição pragmática. Vale a pena ressaltar que a presença de uma pausa gera uma quebra, mas a quebra gerada pela pausa pode ser tanto de natureza terminal quanto não terminal. Por outro lado, não é necessário que haja uma pausa para que uma quebra se produza. Os fatores que intervêm na produção de uma quebra são vários: pausa e/ou reset da curva de F0 e/ou uma mudança brusca de intensidade e/ou uma mudança brusca de velocidade de fala. Contudo, não parece possível mensurar quando uma combinação desses fatores produz o fenômeno da quebra prosódica, que deve ser considerado (pelo menos no estado atual dos conhecimentos) apenas um fenômeno da percepção humana, não mensurável automaticamente. O mesmo critério de segmentação foi adotado nos *corpora* CoSIH (IZRE'EL, 2001; HARY, 2003) e AfroAs (METTOUCHI & CHANARD, 2010; METTOUCHI et al., 2010), e um critério muito parecido foi adotado no *corpus* Santa Barbara (DU BOIS et al., 2000-2005). O estudo sobre a relação entre pausa e quebra terminal no C-

ORAL-BRASIL, mesmo estando em andamento, já mostra que a margem de erro que se obteria segmentando por pausas seria alta demais, qualquer que seja o valor temporal atribuído à pausa. O erro seria de natureza dupla: deixaria de marcar fronteiras entre enunciados diferentes e atribuiria fronteira dentro do mesmo enunciado<sup>8</sup>.

O principal problema metodológico apresentado por uma segmentação de natureza perceptual, mas não mensurável, é como demonstrar a confiabilidade da segmentação. Todo o processo de treinamento, segmentação e validação quanto ao C-ORAL-BRASIL está descrito em Raso & Mittmann (2009), Moneglia et al. (2010) e em Mello et al. (2012); esse processo é uma evolução daquele adotado para o C-ORAL-ROM (MONEGLIA et al., 2005). Em extrema síntese, a confiabilidade é medida através de um teste de concordância estatística, o teste Kappa de Fleiss (1971). Esse teste mede as chances de que o acordo seja devido ao acaso ou à mesma percepção. A pontuação vai de 0 (porcentagem de acordo completamente casual) a 1 (acordo total). O resultado do teste para os segmentadores do C-ORAL-BRASIL foi de 0.86 (0.87 para as quebras com valor terminal). Um grau de acordo superior a 0.80 é considerado excelente. Considere-se também que o teste foi realizado durante a transcrição e as segmentações, as quais ocorriam em conjunto, ou seja, durante uma atividade com uma alta carga cognitiva. O teste de concordância do C-ORAL-ROM foi realizado *a posteriori*, ou seja, sobre textos já transcritos e segmentados, e ofereceu, portanto, um resultado bem mais alto, sempre claramente superior a 0.9.

Examinando agora comparativamente a unidade de referência da fala e aquela da escrita, podemos entender melhor porque a sentença não possui, por si, valor comunicativo, enquanto o enunciado sim. Escutando um enunciado, nós reconhecemos imediatamente, mesmo fora de contexto, qual é a ação que é veiculada por ele: pergunta, pedido de confirmação, chamamento, ordem, instrução, expressão de obviedade, asserção, etc. O tipo de força ilocucionária do enunciado é veiculado pela prosódia, como pode facilmente ser verificado no exemplo 7, prestando-se atenção à diferente força ilocucionária dos 4 enunciados iguais, todos preenchidos pelo mesmo conteúdo locutivo, e nos exemplos seguintes, também com igual conteúdo locutivo, mas produzidos em laboratório:

---

<sup>8</sup> Para algumas características estruturais da fala que emergem através de uma segmentação em enunciados definidos pragmaticamente, veja-se Cresti (2005b) e Raso & Mittmann (2012)

Exemplo 10



João vai pro Rio

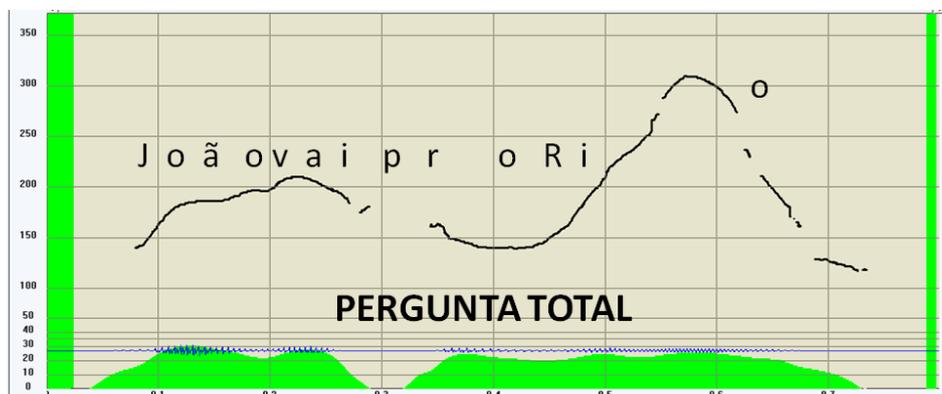


Figura 3. Tela do WinPitch para o Exemplo 10a.

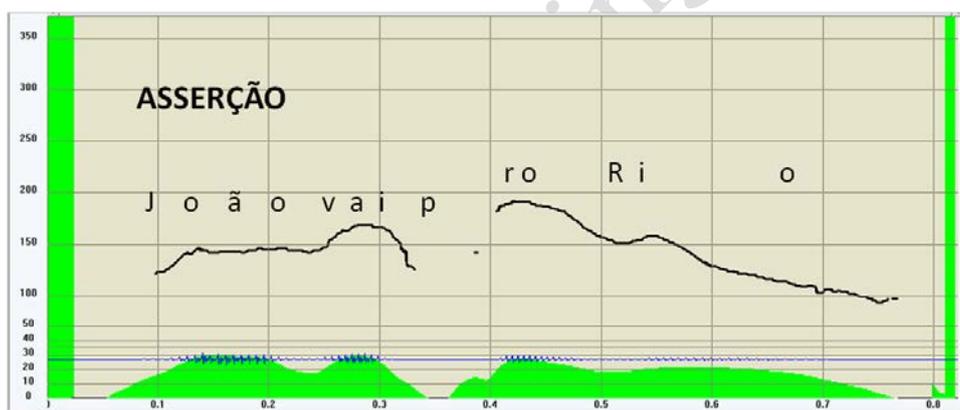


Figura 4. Tela do WinPitch para o Exemplo 10b.

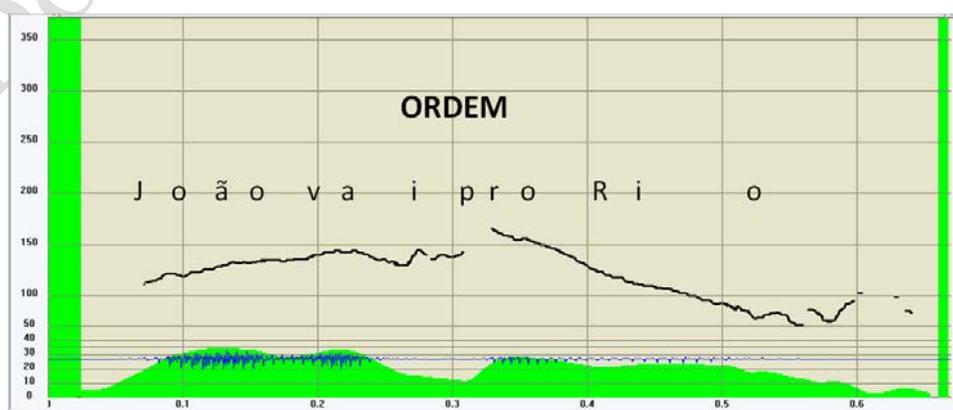
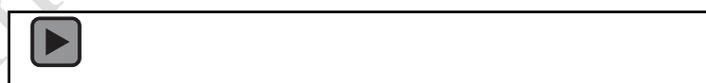


Figura 5. Tela do WinPitch para o Exemplo 10c.

Note-se que somente uma parte do enunciado (às vezes pouquíssimas sílabas) veicula a força ilocucionária. O resto do enunciado é formado por uma parte de preparação ou de ligação ou de coda, que é completamente dispensável aos fins de veicular o tipo de ação. Essa parte serve para hospedar as sílabas da locução, que, frequentemente, são em número maior do que aquelas necessárias para realizar o núcleo da ilocução, que na maior parte dos casos são apenas duas: as finais na asserção e na pergunta total, a inicial e a final na ordem. Isso explica o porquê do fenômeno observado no exemplo 2, repropsto a seguir:

\*BAL: tá saindo de uma garrafinha que tem um bico muito pequeno //

Quando analisamos esse exemplo, observamos que do ponto de vista prosódico a parte que poderia ser interpretável em autonomia não seria nunca a primeira, representada por uma construção sintática principal; ao contrário, a parte interpretável em autonomia seria a segunda, constituída por uma dependente, mais precisamente uma relativa. O fato é que o núcleo dessa específica ilocução cai na parte final do enunciado. Na figura 6, a parte nuclear da ilocução é circulada, tornando bem identificável o foco prosódico que veicula a asserção. Ela corresponde à locução *muit(o) p(e)quen(o)*. Essa parte (composta por apenas 2 sílabas fonéticas) é a única necessária e suficiente para veicular a ilocução, como pode ser apreciado na oitiva do áudio.



É importante observar que a forma e posição do núcleo de cada ilocução variam. A ordem, por exemplo, pode ser constituída por um núcleo descontínuo, com uma ligação (e eventualmente uma preparação ou uma coda) entre as duas sílabas necessárias. Essas se posicionam respectivamente na parte esquerda e na parte direita da locução. A maioria dos atos assertivos situa o núcleo na parte direita, enquanto muitos expressivos o situam na parte esquerda.

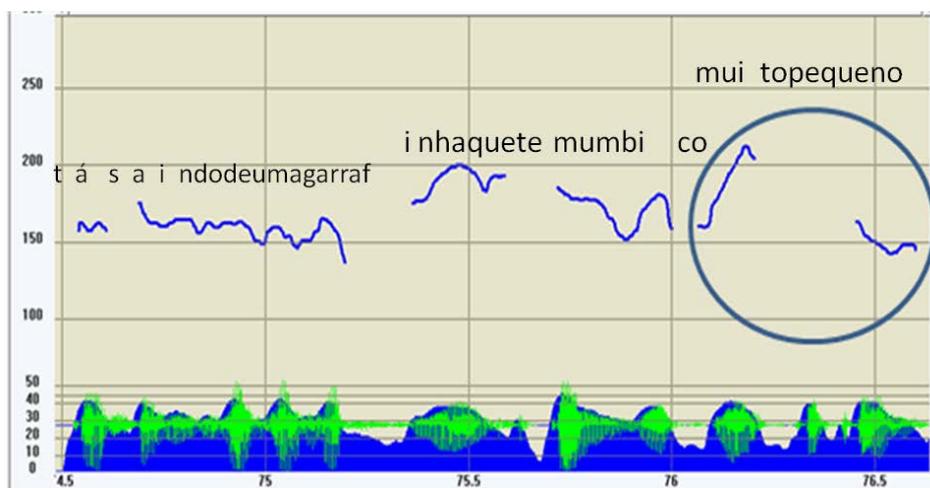


Figura 6. Tela do WinPitch para o Exemplo 2.

Ao contrário, uma sentença da escrita não possui essa capacidade ilocucionária. Uma sequência sintática e semanticamente bem formada não comunica nada por si só. Poder-se-ia objetar que um cartaz como *Proibido fumar!* ou *Silêncio*, ou a escrita *pare* em um cruzamento, ou ainda a escrita *Manter fora do alcance de crianças* na caixa de um remédio, ou muitos outros exemplos do gênero cumprem de fato um ato de fala. Mas, na verdade, o que consegue veicular a mensagem não é a locução em si, mas o texto. Esses exemplos são todos de textos formados por uma única estrutura sintática, e por isso podem dar a ilusão que a unidade de referência da escrita também veicule ilocuções. É evidente ao contrário que qualquer sentença dentro de um texto maior não veicula nenhuma ilocução e não é comunicativamente interpretável fora do texto como um todo. Mas uma reflexão melhor nos mostra que nem os exemplos acima podem ser interpretados fora do texto. Se a escrita *pare* não for colocada no chão em um cruzamento, e não for escrita com um aspecto gráfico convencionalizado, ela não cumpre a função de transmitir aos motoristas que, com base nas regras de trânsito determinadas pelo órgão competente, são obrigados a parar. A mesma expressão afixada na parede de um quarto ou escrita a mão em um bilhete deixado cair na calçada não possui nenhum valor comunicativo. O mesmo raciocínio poderia ser feito com todos os outros exemplos. A capacidade comunicativa de um texto é realizada por uma estrutura (ou, mais frequentemente, por um conjunto de estruturas) dentro de um contexto em que seja claro quem (seja um indivíduo ou uma instituição) se dirige a quem e com qual finalidade. Podemos dizer que um texto escrito como um todo possui uma dimensão perlocutiva, se com essa expressão entendemos a capacidade de gerar um efeito no mundo (AUSTIN, 1962; SEARLE, 1969); mas não podemos dizer que ele é composto por ilocuções como acontece com um texto falado. Isso porque a ilocução é veiculada

prosodicamente, e, em um texto falado, cada enunciado é interpretável ilocutivamente em isolamento, como mostram os enunciados isolados de todos os exemplos desse artigo.

Essa reflexão pode ser ulteriormente aprofundada. A fala é constituída pela imediata transposição em esquemas acionais convencionalizados e reconhecíveis de uma pulsão interna. A escrita ao contrário é a transposição em sequências coerentes e coesas de um pensamento através de uma mediação tecnológica que consiste na ação de escrever. A fala se transforma imediatamente em ação a partir de um estado interior (DAMÁSIO, 2010), a escrita constitui uma ação (a ação de escrever) que constrói um produto, o qual, como resultado, pode ter uma função comunicativa como um todo se caracteriza como texto, não como simples sequência sintática e semântica.

Compreender a diferença radical que existe entre o *produto* escrito e o *evento* da fala gera consequências importantíssimas para o estudo da linguagem entendida na sua dimensão natural: a fala.

## 5. O estudo da fala

Para estudar corretamente a fala é necessário compreender a fundo a necessidade de se capturar essa dimensão de *evento*. Observar como a fala se estrutura significa observar como um evento acontece. Ou seja, significa observar algo que desaparece enquanto se manifesta. As propriedades do evento "fala" não podem ser capturadas em uma transcrição, ou seja, na transposição da fala em uma dimensão escrita. Ao fazer isso, se transforma o evento em um objeto. É evidente que o objeto é muito mais fácil de ser estudado, porque ele é observável como algo substancialmente fixo e imutável. E é evidente que até tempos recentes o estudo da linguagem não possuía instrumentos para capturar a natural dimensão eventiva da fala e não havia de fato alternativa senão estudar a dimensão escrita.

Por muito tempo então a linguagem foi estudada como um produto, um objeto fixo e estável, resultado de um processo tecnológico, a escrita. E através da observação dessa modalidade foram elaboradas categorias interpretativas apropriadas para ela. Mas o preconceito existente, segundo o qual para o estudo da oralidade é suficiente aplicar as categorias elaboradas em séculos (ou milênios) de observação da escrita, eventualmente com algumas adaptações, é ainda muito forte. A ausência da dimensão acional na unidade de referência da escrita e, ao contrário, o fato que essa dimensão constitui o primeiro nível interpretativo na fala é apenas o exemplo mais gritante da impossibilidade de analisar uma modalidade com as categorias

elaboradas a partir da observação da outra modalidade. A linguística está ainda muito atrasada na elaboração de categorias que sejam fruto da observação da fala, sem querer olhar essa modalidade com a lente que uma longuíssima (e nobre) tradição elaborou a partir da observação da escrita.

Mas nas últimas décadas, mesmo se muitas vezes lutando contra preconceitos generalizados, os linguistas produziram também grandes avanços. Esses avanços foram fruto principalmente da disponibilidade de tecnologias novas que tornam possível capturar e apreciar melhor a dimensão de evento que a linguagem, na sua dimensão natural, desempenha.

Em primeiro lugar, as novas tecnologias permitiram a coleta e a análise de grandes quantidades de dados linguísticos e a sua organização em arquiteturas representativas de universos linguísticos definidos, exploráveis eletronicamente de maneira computável. A linguística de *corpora* permitiu uma compreensão nova de como a linguagem é realmente usada, tanto na sua modalidade escrita quanto naquela oral. Mas, se essa metodologia produziu resultados evidentes em primeiro lugar quanto à modalidade escrita, isso é devido às maiores dificuldades apresentadas pela compilação de recursos falados e pela análise da fala, principalmente em contexto natural. As razões disso são várias. Entre as principais estão as seguintes:

1. até tempos muito recentes, a tecnologia não permitia a gravação da fala em contexto natural com uma qualidade acústica apropriada a capturar eventos em situações variadas. Por essa razão, além da subestimação da importância crucial da variação diafásica para a estruturação da fala espontânea, quase todos os primeiros *corpora* de fala espontânea se limitavam a gravar as situações comunicativas que apresentavam menores problemas de planejamento e ausência quase total de ruído: narrativas, entrevistas ou em geral situações que podiam facilmente ser controladas (interação médico-paciente, interações telefônicas ou televisivas). Somente na última década surgiram *corpora* de fala espontânea que oferecem uma real variação situacional. Entre eles, os *corpora* C-ORAL-ROM (CRESTI & MONEGLIA, 2005), o *Corpus* de Santa Barbara (DU BOIS, et al. 2000-2005) e o *corpus* C-ORAL-BRASIL (RASO & MELLO, 2012).

2. As limitações tecnológicas e as diferentes tradições criaram poucas oportunidades de encontro entre as abordagens mais linguísticas e aquelas mais fonéticas do estudo da linguagem. As primeiras eram mais sensíveis ao estudo de dados extraídos em contexto

natural, mas sentiam uma menor exigência de uma qualidade acústica que permitisse a análise dos fenômenos prosódicos, mostrando maior preocupação com o estudo da sintaxe e do léxico, ou das dinâmicas interativas em uma ótica de análise da conversação. As segundas, preocupadas com a qualidade acústica dos dados, se concentravam no estudo da fala em laboratório, em um contexto, portanto, muito diferente daquele natural. Com o desenvolvimento de tecnologias de gravação *wireless* de alta qualidade, se tornou possível aplicar os avanços dos estudos prosódicos (e de fonética segmental) à observação de textos de diferentes diafásias coletados em situações não controladas.

3. Para se analisar a fala não é possível se basear na transcrição, como mostramos. Não é possível porque a transcrição reduz a fala a um texto escrito, na sua dimensão de *produto*, e não apresenta as características próprias do *evento* da fala. Concretamente, isso significa a perda do meio sonoro e de tudo o que é transmitido pela prosódia (marcação da estruturação em enunciados e unidades tonais, perfil prosódico que marca a função ilocucionária e a relação informacional entre as diferentes unidades tonais). Mas a pouco serve também fornecer a transcrição acompanhada da gravação sonora se elas são separadas. Isso continua equivalendo a ter somente a transcrição. De fato, não é possível procurar o ponto que se quer escutar todas as vezes que precisa (e precisaria sempre) e quantas vezes fosse necessário (e normalmente são necessárias muitas vezes). Sem contar que isso torna praticamente impossível uma análise fonética adequada. Para conseguir uma análise do *evento* falado de forma adequada foi necessário mais um salto tecnológico e mais um amadurecimento metodológico da disciplina que permitiram o alinhamento entre transcrição e áudio através de softwares que também possibilitam a análise do sinal acústico. Hoje existem vários softwares capazes de fazer isso com grande quantidade de fala. O software usado para os *corpora* C-ORAL-BRASIL e C-ORAL-ROM foi o WinPitch (MARTIN, 2005; [www.winpitch.com](http://www.winpitch.com)). Esse software, além de uma grande quantidade de análises acústicas, permite visualizar o texto e escutar o sinal sonoro enunciado por enunciado ao mesmo tempo e quantas vezes for preciso; em outra visualização, é possível ver o espectrograma e o texto do enunciado, e também escutar o sinal sonoro. Dessa forma, o processo da fala pode ser repetido enunciado por enunciado quantas vezes for necessário para apreciar os fenômenos que o *evento* apresenta. A operação pode ser realizada visualizando a seleção de trecho falado que

mais é adequado ao momento e ao objetivo da pesquisa. Podem-se selecionar vários enunciados, apreciando-se assim a relação dos parâmetros acústicos em trechos maiores, ou uma parte do enunciado, visualizando-se assim em detalhe os parâmetros e operando-se medições mais precisas. Isso é o que conhecemos de mais adequado para recuperarmos a dimensão *eventiva* da fala, juntando a transcrição, necessária pelos limites da nossa memória, com a possibilidade de repetir sem limite e sem necessidade de buscas complexas o processo acústico.

## 6. Conclusão (com observações sobre as outras tecnologias)

O objetivo principal das seções anteriores foi mostrar o quanto é profunda a diferença entre a fala e a escrita e como a razão principal dessa diferença é frequentemente ocultada nos estudos linguísticos devido à tradição e a carências metodológicas. Fizemos isso selecionando alguns aspectos da questão, que consideramos especialmente importantes. Naturalmente há outros aspectos em jogo, e mesmo aqueles que tratamos mereceriam um aprofundamento maior que não é possível aqui, e que, pelo menos em parte, pode ser feito recorrendo-se à bibliografia aqui indicada.

A abordagem ao estudo da fala sob o ponto de vista de várias subdisciplinas da linguística leva a analisá-la através de categorias elaboradas a partir da observação da escrita, que foi por milênios a única forma de linguagem observável para fins de estudo. Uma das principais consequências disso é a pouca atenção dada à ilocução como núcleo comunicativo da fala e à articulação informacional como princípio estruturante dessa modalidade<sup>9</sup>. E a razão fundamental é facilmente individualizável no fato que tanto a ilocução quanto a estrutura informacional são veiculadas pela prosódia, ou seja, um recurso ausente na escrita (e talvez por isso esse recurso seja ainda considerado como um elemento paralinguístico por parte dos linguistas). A escrita, ao contrário, funda a própria estruturação em princípios sintáticos e semânticos, que podem ser analisados a partir da visão. Enfim, tentamos mostrar como analisar uma modalidade, a fala, não através de categorias extraídas pela observação da modalidade que se quer observar (*iuxta propria principia*), mas através de categorias extraídas pela observação da modalidade mais conhecida e estudada, a escrita, leva à não compreensão da modalidade falada. Isso é ainda mais grave quando se trata da modalidade natural da linguagem, o que gera

---

<sup>9</sup> Nesse artigo não tivemos a possibilidade de abordar a articulação informacional da fala. A esse respeito, citamos Moneglia (2011), Raso (2012), Panunzi & Mittmann (no prelo).

consequências também na compreensão de questões biológicas e cognitivas relativas a essa capacidade da nossa espécie. A esse respeito, só para citar um exemplo, o reconhecimento de que os esquemas acionais são realizados na fala através de formas prosódicas que podem ser descritas (FIRENZUOLI, 2003; MONEGLIA, 2011; MORAES, no prelo) permitiria investigar a relação da ilocução com uma das mais importantes descobertas recentes dentro da neurociência, ou seja, os neurônios espelho (RIZZOLATI & CRAIGHERO, 2004; RAMACHANDRAN, 2000; ARBIB, 2012)<sup>10</sup>.

Essa tendência em analisar uma modalidade da comunicação linguística através de categorias extraídas de modalidades já conhecidas está se repetindo na análise das modalidades comunicativas surgidas com as novas tecnologias: o chat, o Skype, o email, o SMS, os blogs, etc. Em formas mais ou menos elaboradas, é muito frequente ler análises dessas modalidades de comunicação linguísticas que utilizam categorias extraídas das outras modalidades e as sintetizam como diferentes "misturas" de fala e escrita. Desse jeito se perde a capacidade de entender a fundo as novas modalidades e não se constroi uma metodologia adequada para isso. Sem querer aprofundar essa questão aqui, mas apenas para mostrar como está se repetindo para a análise dessas novas formas comunicativas o equívoco que ainda prejudica muito o estudo da fala, apresentaremos algumas observações sobre duas tecnologias, o SMS e o Skype, para indicar, sem aprofundá-los, os passos metodológicos que consideramos apropriados.

Trata-se de duas tecnologias que possuem meios diferentes. O SMS tem como meio um código de símbolos visuais da escrita. O Skype tem como meio a voz, mas esse meio pode ser integrado ou substituído com o meio visual da escrita, o que deve ser considerado; contudo, por comodidade expositiva, dados os nossos fins, consideraremos aqui o Skype somente em seu uso vocal, conscientes de que isso constitui uma simplificação. O canal das duas modalidades é muito complexo; a descrição a seguir é simplificada.

No caso do SMS, o canal é constituído por um aparelho complexo e móvel que, graças a um sistema de interface gráfica, permite que sejam digitados em número limitado símbolos em um pequeno teclado, para que esses símbolos sejam enviados para um sistema central que

---

<sup>10</sup> Em extrema síntese: a descoberta dos neurônios espelho mostra que nós ativamos a mesma configuração neuronal quando realizamos uma ação e quando assistimos à realização dessa ação por parte de outras pessoas; também, parece que essas configurações são holísticas, ou seja, que duas ações muito parecidas (como, por exemplo, pegar algum objeto no alto levantando o braço no caso em que o objeto seja visível e no caso em que seja não visível) ativam configurações completamente diferentes. Seria interessante verificar o que acontece com as ações linguísticas, que também parecem ser codificadas de maneira holística, ou seja por padrões prosódicos que devem ser interpretados como um todo, e não através de uma combinação de traços nas quais mudam somente um ou poucos traços relativos aos elementos que diferenciam uma ação da outra.

depois os encaminha para o aparelho móvel do destinatário final da mensagem. O canal é portanto constituído também por uma memória eletrônica e pela rede de telefonia móvel. No caso do Skype (uso vocal), o canal é constituído pelo ar integrado por um computador, que, através de um microfone, consegue converter as ondas sonoras (e a imagem) em sinais digitais, que, por sua vez, são empacotados e enviados para a rede internet que os distribui para o computador do destinatário, o qual decodifica o sinal e o reconverte em ondas sonoras. Em ambas as modalidades, em princípio, a transmissão deveria acontecer de maneira quase instantânea.

O cruzamento do meio e do canal produzem conseqüências pragmáticas específicas, entre as quais as seguintes. No caso do SMS, pode se produzir, por diferentes razões, um ruído no canal que retarda ou até bloqueia de maneira definitiva a chegada da mensagem, sem que o emissor ou o destinatário tenham conhecimento do problema. A dificuldade de digitação e o limite de caracteres obrigam a mensagens curtas. A rapidez da comunicação encoraja a interatividade, o que também induz a mensagens curtas. Isso tem como conseqüência linguística a formação de um código próprio dessa forma comunicativa (que eventualmente pode parcialmente ser transferido em outras formas de escrita), caracterizado, entre outras coisas, por sentenças muito curtas, pouca interpunção, abreviaturas e pela presença dos *emoticons*. A durabilidade do suporte é alta, mas de fato a vida de um celular é limitada e a necessidade de liberar memória reduz a possibilidade de guardar os textos. Eles podem ser transferidos em outros suportes, mas a rapidez e a quantidade de produção textual e o valor extremamente contingencial dela fazem com que essa produção seja normalmente muito volátil. Contudo, as companhias telefônicas possuem um sistema de armazenamento ao qual, em circunstâncias muito limitadas, pode-se recorrer.

No caso do Skype, as conseqüências pragmáticas são muito interessantes. Em primeiro lugar, a falta de uma simultaneidade total entre a emissão e a chegada do som (e da completa simultaneidade com a imagem, devido aos efeitos da distância física), aliado a um compartilhamento muito parcial da situação, geram problemas no gerenciamento dos turnos que induzem estratégias diferentes daquelas da fala natural. As sobreposições involuntárias podem ser frequentes, mesmo quando não há ruído devido ao canal. Contrariamente à comunicação face a face, o aviso de chamada pelo Skype pode não ser atendido sem conseqüências na interação, o que não seria pensável se a comunicação fosse presencial. O compartilhamento da imagem dos falantes é muito parcial. Normalmente, só a imagem do rosto

é compartilhada. O resto do corpo e principalmente o ambiente em volta, não são compartilhados (por exemplo não se sabe se há outras pessoas presentes no espaço do outro). A comunicação visual pode ser compartilhada ou não (uma fácil desculpa, e às vezes uma necessidade, para não se compartilhar a imagem é a baixa qualidade da conexão), o que gera efeitos muito diferentes. Essa decisão não é dependente da situação, como na fala natural, mas depende da vontade de um ou de ambos os interactantes. Sem o compartilhamento da imagem, a liberdade dos interlocutores para fazer coisas diferentes é ampla, desde que não fiquem longe demais do microfone. Mas mesmo com o compartilhamento da imagem, os interactantes podem usar o computador para outras tarefas sem serem percebidos. A mobilidade dos falantes é contudo muito limitada, já que é necessária a proximidade com o computador. Este, dependendo do tipo, pode ser transportado, e é até possível utilizar a mobilidade da câmara para mostrar o espaço em volta, mas com vínculos completamente diferentes daqueles que se determinam pelo olhar e o movimento na fala natural. Uma consequência evidente do ponto de vista linguístico é um uso particular das referências dêiticas, principalmente, mas não só, daquelas espaciais.

A análise dessas duas modalidades poderia ser muito mais aprofundada, mas o que nos interessa mostrar é a necessidade de analisar cada modalidade por si mesma, extraíndo categorias dessa análise, e não adotando, por analogia, as categorias das modalidades que já conhecemos. A metodologia apropriada para fazer isso deve seguir os passos realizados aqui: uma descrição do meio e do canal (que nas duas tecnologias usadas para a exemplificação são muito diferentes daqueles da fala natural ou da escrita tradicional, assim como são diferentes entre elas); uma análise aprofundada das consequências pragmáticas que decorrem da combinação específica de meio e canal, e que são extremamente originais. Somente a partir desses passos é possível começar a entender a estruturação linguística da interação, que é, naturalmente, uma consequência dos passos anteriores. A nossa impressão é que a linguística deveria ter maior consciência da necessidade desse procedimento e das consequências metodológicas que essa consciência impõe. A falta dessa consciência é extremamente prejudicial principalmente para o entendimento da modalidade natural da linguagem, aquela que usamos para fazer grande parte das nossas ações e que ciências diferentes procuram explicar com finalidades diferentes, seja a compreensão do nosso sistema neurológico e cognitivo, seja a causa e a cura de patologias, seja o processamento automático da linguagem.

## Referências bibliográficas

ARBIB, M. A. **How the brain got language: the Mirror System Hypothesis**. New York & Oxford: Oxford University Press, 2012.

AUSTIN, J. **How to do things with words**. Oxford: Oxford University Press, 1962.

BERRUTO, G. **Sociolinguistica dell'italiano contemporaneo**. Roma: La Nuova Italia Scientifica, 1987.

\_\_\_\_\_. Le varietà del repertorio. In: SOBRERO, A. A. **Introduzione all'italiano contemporaneo**. vol 2. Roma-Bari: Laterza, 1993. p. 3-36.

\_\_\_\_\_. Varietà diamesiche, diastratiche, diafasiche. In: SOBRERO, A. A. **Introduzione all'italiano contemporaneo**. vol 2. Roma-Bari: Laterza, 1993. p. 37-92.

BIBER, D., JOHANSSON, S., LEECH, G., CONRAD, S., & FINEGAN, E. **The Longman Grammar of Spoken and Written English**. London and New York: Longman, 1999.

CHAFE, W. Linking intonation units in spoken English. In: HAIMAN, J.; THOMPSON, S. (Eds). **Clause combining in Grammar and Discourse**. Amsterdam: Benjamins, 1988. p. 1-27.

CHOMSKY, N. Remarks on nominalization. In: R. JACOBS; ROSENBAUM, P. (eds.) **Reading in English Transformational Grammar**, 184-221. Waltham: Ginn, 1970.

CRESTI, E. **Corpus di Italiano Parlato**, Voll. I-II, CD-ROM. Firenze: Accademia della Crusca, 2000.

\_\_\_\_\_. Per una nuova definizione di frase. In: BONGRANI, P.; DARDI, A.; FANFANI, M.; TESI, R. (Org.). **Studi di storia della lingua italiana offerti a Ghino Ghinassi**. Firenze: Le Lettere, 2001. p. 511-550.

\_\_\_\_\_. Enunciato e frase: teoria e verifiche empiriche. In: BIFFI, M.; CALABRESE, O.; SALIBRA, L. (Ed.). **Italia linguistica: discorsi di scritto e di parlato – nuovi studi di linguistica italiana per Giovanni Nencioni**. Siena: Protagon, 2005a. p. 249-260.

\_\_\_\_\_. Notes on lexical strategy, structural strategies and surface clause indexes in the C-ORAL-ROM spoken corpora. In: CRESTI, E.; MONEGLIA, M. (Ed.). **C-ORAL-ROM: Integrated reference corpora for spoken Romance Languages**. Amsterdam/Philadelphia: John Benjamins, 2005b. p. 209-256.

\_\_\_\_\_. Syntactic properties of spontaneous speech in the Language into Act Theory: data on Italian complements and relative clauses. In: RASO, T.; MELLO, H. **Speech Corpora and Linguistic Studies**. Amsterdam/Philadelphia: John Benjamins, no prelo.

CRESTI, E. & GRAMIGNI, P. **Per una linguistica corpus based dell'italiano parlato: Le unità di riferimento.** In F. Albano Leoni, F. Cutugno, M. Pettorino, & R. Savy (Eds.), *Atti del Convegno Nazionale "Il Parlato Italiano"*, CD-ROM (pp. 1–26). Napoli: M. D'Auria, 2004

CRESTI, E.; MARASCHIO, N. TOSCHI, L. (Eds) **Storia e teoria dell'interpunzione.** Roma: Bulzoni, 1992.

CRESTI, E.; MONEGLIA, M. (Ed.). **C-ORAL-ROM:** integrated reference corpora for spoken Romance languages. Amsterdam: John Benjamins, 2005.

CRYSTAL, D. **The English Tone of Voice.** London: Edward Arnold. 1975.

DAMÁSIO, A. **Self Comes to Mind:** Constructing the Conscious Brain. New York: Pantheon, 2010.

DU BOIS, J. W.; CHAFE, W. L.; MEYER, C.; THOMPSON, S. A.; ENGLEBRETSON, R.; MARTEY, N.. **Santa Barbara corpus of spoken American English.** Parts 1-4. Philadelphia: Linguistic Data Consortium, 2000-2005.

DIAMOND, J. **Guns, Germs, and Steel: The Fates of Human Societies.** New York: W.W. Norton & Co., 1997.

FIRENZUOLI, V. **Le forme intonative di valore illocutivo dell'italiano parlato:** analisi sperimentale di un corpus di parlato spontaneo (LABLITA). PhD. Thesis. University of Florence, 2003.

FISHER, S. R. **A History of Writing.** London: Reaktion Books, 2005.

FLEISS, J. L. **Measuring nominal scale agreement among many raters.** Psychological Bulletin, 1971.

DOI: <http://dx.doi.org/10.1037/h0031619>

GOODY, J. **The Logic of Writing and the Organization of Society.** Cambridge University Press, 1986

DOI: <http://dx.doi.org/10.1017/CBO9780511621598>

HALLIDAY, M. A. K. **Spoken And Written Language.** Oxford University Press, 1989.

HARY, B. H. (ed.). **Corpus Linguistics and Modern Hebrew: Towards the Compilation of The Corpus of Spoken Israeli Hebrew (CoSIH).** Tel Aviv: Tel Aviv University, The Chaim Rosenberg School of Jewish Studies, 2003.

KARCEVSKY, S. **Sur la phonologie de la phrase.** Travaux du Cercle linguistique de Prague. IV, 1931. p. 188-228.

IZRE'EL, S.; HARY, B.; RAHAV, G. Designing CoSIH: The Corpus of Spoken Israeli Hebrew. In: **International Journal of Corpus Linguistics**, 6, 2001, p. 171-197.

MAC MAHON, A.; MAC MAHON, R. **Evolutionary Linguistics**. Cambridge: Cambridge University Press, 2013.

MARTIN, Ph. WinPitch Corpus: a text-to-speech analysis and alignment tool. In: CRESTI, E. & MONEGLIA, M. (Eds.). **C-ORAL-ROM**. Integrated Reference Corpora for Spoken Romance Languages. Amsterdam/Philadelphia: John Benjamins, 2005.

McNEILL, D. **How Language Began**. Cambridge: Cambridge University Press, 2012.

**DOI**: <http://dx.doi.org/10.1017/CBO9781139108669>

MELLO, H.; RASO, T.; MITTMANN, M.; VALE, H.; CÔRTEZ, P. Transcrição e segmentação prosódica do *corpus* C-ORAL-BRASIL: critérios de implementação e validação. In: RASO, T. & MELLO, H. (eds.) **C-ORAL – Brasil I: Corpus** de referência do português brasileiro falado informal. Belo Horizonte: Editora UFMG, 2012, p. 125-176.

METTOUCHI, A. & CHANARD, C. **From Fieldwork to Annotated Corpora: The CorpAfroAs project**. Cahiers de Faits de Langues n°2., 2010. p. 255-265.

METTOUCHI, A.; CAUBET, D.; VANHOVE, M.; TOSCO, M.; COMRIE, B.; IZRE'EL, S. CORPAFROAS, A *Corpus* for Spoken Afroasiatic Languages: Morphosyntactic and Prosodic analysis. In: FALES, M. F.; GRASSI, G. F. (eds). **Proceedings of the 13th Italian Meeting of Afro-Asiatic Linguistics**. Padova: SARGON, 2010, p.177-180.

MILLER, J.; WEINERT, R. **Spontaneous Spoken Language**. Oxford: Clarendon Press, 1998.

MONEGLIA, M. The C-ORAL-ROM Resource. In: CRESTI, E.; MONEGLIA, M. (Eds.). **C-ORAL-ROM**. Integrated Reference *Corpora* for Spoken Romance Languages. Amsterdam/Philadelphia: John Benjamins, 2005, p. 1-70.

\_\_\_\_\_. Units of analysis of spontaneous speech and speech variation in a cross-linguistic perspective. In: KAWAGUCHI, Y.; ZAIMA, S.; TAKAGAKI, T. (Eds.). **Spoken language corpus and linguistics informatics**. Amsterdam: John Benjamins, 2006. p. 153-179.

\_\_\_\_\_. Spoken corpora and pragmatics. In: **Revista Brasileira de Linguística Aplicada**, 11(2): 479–519.

MONEGLIA, M.; CRESTI, E. L'intonazione e i criteri di trascrizione del parlato adulto e infantile. In: BORTOLINI, U.; PIZZUTO, E. (ed.). **Il Progetto CHILDES Italia**. Pisa: Del Cerro, 1997. p. 57-90.

\_\_\_\_\_. C-ORAL-ROM: Prosodic boundaries for spontaneous speech analysis. In: KAWAGUCHI, Y.; ZAIMA, S.; TAKAGAKI, T. (Ed.). **Spoken language corpus and linguistics informatics**. Amsterdam: John Benjamins, 2006.

MONEGLIA, M.; FABBRI, M.; QUAZZA, S.; PANIZZA, A.; DANIELI, M.; GARRIDO, J. M.; SWERTS, M. Evaluation of consensus on the annotation of terminal and non-terminal prosodic breaks in the C-ORAL-ROM corpus. In: CRESTI, E.; MONEGLIA, M. (ed.). **C-**

**ORAL-ROM**: integrated reference *corpora* for spoken Romance languages. Amsterdam: John Benjamins, 2005. p. 257-276.

MONEGLIA, M.; RASO, T.; MITTMANN, M.; MELLO, H. Challenging the perceptual relevance of prosodic breaks in multilingual spontaneous speech *corpora*: C-ORAL-BRASIL/C-ORAL-ROM. Prosodic prominence perceptual and automatic identification. In: **Proceedings of the Speech Prosody International Conference Satellite Workshop**. Chicago: Université de Neuchâtel, 2010.

MORAES, J. & RILLIARD, R. Illocution, Attitude and Prosody. In: RASO, T. & MELLO, H. **Speech Corpora and Linguistic Studies**. Amsterdam/Philadelphia: John Benjamins, no prelo.

NENCIONI, G. **Di scritto e di parlato**: discorsi linguistici. Bologna: Zanichelli, 1983.

PANUNZI, A. & MITTMANN, M. The IPIC resource and a cross-linguistic analysis of information structure in Italian and Brazilian Portuguese. In: RASO, T. & MELLO, H. **Speech Corpora and Linguistic Studies**. Amsterdam/Philadelphia: John Benjamins, no prelo.

POWELL, B. B. **Writing**: Theory and History of the Technology of Civilization. Oxford: Blackwell, 2009.

QUIRK, R.; GREENBAUM, S.; LEECH, G.; SVARTVIK, J. **A Comprehensive Grammar of the English Language**. London and New York: Longman, 1985.

RAMACHANDRAN, V. S. **Mirror Neurons and imitation learning as the driving force behind "the great leap forward" in human evolution**. Edge, 69, 2000.

RASO, T. O C-ORAL-BRASIL e a Teoria da Língua em Ato. In: RASO, T. & MELLO, H. (eds.) **C-ORAL – Brasil I**: Corpus de referência do português brasileiro falado informal. Belo Horizonte: Editora UFMG, 2012b, p. 91-124.

RASO, T. & MELLO, H. (eds.) **C-ORAL – Brasil I**: Corpus de referência do português brasileiro falado informal. Belo Horizonte: Editora UFMG, 2012.

RASO, T. & MITTMANN, M. Validação estatística dos critérios de segmentação da fala espontânea no *corpus* C-ORAL-BRASIL. **Revista de Estudos da Linguagem**, 2009, p. 73-91.

\_\_\_\_\_. As principais medidas da fala. In: RASO, T. & MELLO, H. (eds.) **C-ORAL – Brasil I**: Corpus de referência do português brasileiro falado informal. Belo Horizonte: Editora UFMG, 2012

RIZZOLATTI, G.; CRAIGHERO, L. The Mirror-Neuron System. In: **Annual Review of Neuroscience**, 2004: 169-92.

**DOI** : <http://dx.doi.org/10.1146/annurev.neuro.27.070203.144230>

ROSSI, F. **Varietà diamesica**. In: Enciclopedia dell'italiano. Roma: Treccani, 2001. p. 1540-1542.

RUHLEN, M. **The Origin of Language**. New York: John Wiley, 1994.

SCHUURMAN, I.; SCHOUPPE, M., HOEKSTRA, H.; VAN DER WOUDE, T. CGN, An Annotated *Corpus* of Spoken Dutch. In: **Proceedings of the 4th International Workshop on Linguistically Interpreted Corpora (LINC-03)**. Budapest, 2003.

SEARLE, J. **Speech Acts. An Essay in the Philosophy of Language**. New York: Cambridge University Press. 1969.

**DOI** : <http://dx.doi.org/10.1017/CBO9781139173438>

SWERTS, M.; KRAHMER, E. Audiovisual prosody and feeling of knowing. In: **Journal of Memory and Language** 53(1), 2005. p. 81-94.

**DOI** : <http://dx.doi.org/10.1016/j.jml.2005.02.003>

Artigo recebido em: 15.09.2013

Artigo aprovado em: 26.11.2013