

LÍNGUA FALADA, LÍNGUA ESCRITA E ANÁLISE DE CORPORA ORAIS: BREVE REVISÃO*

Priscilla Tulipa da Costa – CEFET-MG, *Campus Contagem*
Dalmo Buzato – CEFET-MG, *Campus Contagem*

RESUMO: Os estudos da fala normalmente utilizam em suas pesquisas dados oriundos de transcrição da fala, entretanto, como apontam alguns autores, isso nada mais é do que língua escrita. É necessário, portanto, que se faça a distinção correta das modalidades da língua, para que haja a compreensão do fenômeno que está sendo estudado. Tendo em vista esse panorama, este estudo objetiva a realização de uma revisão teórica dos principais conceitos de língua falada e língua escrita no âmbito dos estudos de corpora orais, a fim de coletar informações sobre conceituação, classificações, implicações e outras questões iniciais relacionadas a essas modalidades linguísticas dentro da perspectiva dos estudos da fala.

PALAVRAS-CHAVE: Língua falada; Língua escrita; Sintaxe da fala espontânea; Corpora orais.

1. Língua escrita e língua falada: principais características

A fala e a escrita são modalidades da linguagem que apresentam características próprias. Há estruturas que são consideradas típicas da fala, como os inícios falsos, enquanto outras ocorrem predominantemente na comunicação escrita. A fala constitui a principal forma de comunicação da humanidade (SILBER-VAROD, 2011) e surgiu há aproximadamente 100 mil anos atrás. A escrita, por sua vez, surgiu há cerca de 5.500 anos e, derivada da fala, tem sua origem ligada à criação de uma tecnologia que pudesse auxiliar na gestão de uma complexidade social (RASO, 2013). Apesar da relação existente entre as duas, fala e escrita apresentam peculiaridades que as tornam distintas entre si, e que fazem com que o estudo de cada uma delas requeira a aplicação de metodologias e parâmetros específicos.

Em uma comparação entre as duas modalidades da língua em uso, Raso (2013) destaca alguns aspectos de distinção. O primeiro deles, de acordo com o autor, se baseia no caráter natural da língua falada, em oposição ao caráter tecnológico da escrita. Para Raso, o ser humano tem a fala como capacidade inata e desenvolve as habilidades orais ao longo do tempo de exposição à língua, por meio da interação social. Trata-se de uma atividade que não requer treinamento, mas que se aprende de forma natural. A capacidade de aprender a língua escrita, por outro lado, não é característica do ser humano, ao contrário, ainda hoje, muitas são as pessoas que não têm conhecimento dessa modalidade comunicacional. A escrita surgiu como uma tecnologia, uma ferramenta para atender às necessidades sociais, um produto da evolução cultural para resolver os problemas que surgiam na época de sua invenção e, diferentemente da língua oral, sua aquisição/aprendizagem requer prática e anos de treinamento.

Outro aspecto distintivo apontado por Raso (2013) entre as instâncias da fala e da escrita está relacionado ao meio e ao canal de efetivação da comunicação. Enquanto a fala é realizada por meio de sinais sonoros (ondas sonoras) que são propagados pelo ar (canal) e

* XVI Encontro Virtual de Documentação em Software Livre e XIII Congresso Internacional de Linguagem e Tecnologia Online

decodificadas pelo aparelho auditivo humano, a escrita é realizada por meio de sinais gráficos, que podem ser apresentados em diversificados canais (como papel, madeira, suporte digital etc.) e que são decodificados pelos olhos.

Da mesma forma, a modalidade de troca, ou seja, a maneira como se dá comunicação em cada uma das instâncias também é diferente. Por ser de curta duração (pouco tempo para planejar e ouvir a fala), a fala espontânea pressupõe interação face a face. Isso quer dizer que, no instante da troca, é necessário que falante e ouvinte estejam em interação, que haja compartilhamento do mesmo tempo e espaço. Já na escrita isso não é necessário. Os conteúdos escritos podem ser produzidos e recepcionados em momentos e espaços diferentes, não sendo necessária a interação entre o leitor e o autor para que a comunicação ocorra (RASO, 2013).

No que diz respeito a esse último aspecto, é interessante pontuar ainda que na linguagem escrita o indivíduo produtor tem a possibilidade de efetuar correções, alterações ou mesmo de apagar suas mensagens. Na fala espontânea, por outro lado, isso não acontece, pois a produção oral não é planejada/elaborada e o falante não dispõe de tempo para fazer preparações, alterações e correções, tampouco para “apagar” o que foi dito. Isso ressalta também o caráter imprevisível da fala, que conta com interlocutores ativos independente do controle do falante, enquanto o escritor tem a chance de controlar seu conteúdo e possíveis interações.

Outra importante diferença entre a fala e a escrita, conforme explica Raso (2013), é que a fala apresenta prosódia, ou seja, ela é dotada de um conjunto de parâmetros acústicos que a estruturam. Esses parâmetros, na escrita, são substituídos pelos sinais de pontuação que, apesar de serem essenciais para a organização da língua escrita, não constituem um sistema natural como é o da prosódia. Quando avaliamos um evento de fala por meio da transcrição em língua escrita, por exemplo, trechos prosódicos podem se perder ou ser omitidos no texto, acarretando em diversos problemas, como erros de segmentação, dificuldade de atribuição de valor comunicativo aos enunciados, entre outros.

Sintetizando as principais diferenças entre as modalidades língua fala e escrita, o Quadro 1 exibe um resumo das características e peculiaridades percebidas, conforme demonstrado por Raso (2013):

Quadro 1 – Língua falada x Língua escrita

| FALA | ESCRITA |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>Meio: ondas sonoras, decodificadas pelo aparelho auditivo humano.</p> <p>Canal: ar, que propaga as ondas sonoras.</p> <p>Origem: 100 mil anos.</p> <p>Características:</p> <ul style="list-style-type: none"> • Caráter natural (capacidade da espécie humana realizada pela inserção no meio social). • Aprendizagem natural. | <p>Meio: símbolos gráficos, decodificadas pelos olhos.</p> <p>Canal: diversos suportes (papel, madeira etc.)</p> <p>Origem: 5.500 anos.</p> <p>Características:</p> <ul style="list-style-type: none"> • Caráter tecnológico (uma ferramenta para atender às necessidades sociais). • Aprendizagem por treinamento. |

| | |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> • Possibilidade de elaboração restrita. • Troca: face a face (aqui e agora). • Espaço compartilhado no momento de produção da fala. • Presença de prosódia. | <ul style="list-style-type: none"> • Maior possibilidade de elaboração. • Troca: distância (sem interação). • Mensagens podem ser elaboradas e recuperadas em momentos diferentes. • Ausência de prosódia. |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Fonte: Elaborado pela autora com dados extraídos de RASO (2013).

2. A relação fala e enunciado x escrita e frase e suas implicações para o estudo da sintaxe da fala

Como visto no tópico anterior, assim como fala é diferente da escrita, também a forma de analisar cada uma dessas modalidades deve ser diferente. Uma distinção importante nesse sentido é a unidade de referência. Na literatura, a definição comumente encontrada atribui ao enunciado o papel de unidade de referência da fala, enquanto a frase corresponde à unidade de referência da escrita.

Conforme critérios propostos pela Teoria da Língua em Ato (Cresti, 2000), criada com base na Teoria dos Atos de Fala de Austin (1962), a frase é a entidade típica da diamesia escrita, na qual o texto é constituído seguindo princípios lógico-sintáticos. Já o enunciado é típico da diamesia falada e a organização da fala é regida por uma estrutura pragmático ilocutória que é baseada no enunciado.

Do ponto de vista da completude semântica e da interpretabilidade pragmática, outros dois critérios propostos pela teoria de Cresti (2000), a frase é considerada como uma estrutura que pode ser interpretada sintática e semanticamente sem que haja, necessariamente, ligação com um contexto, ou seja, ela pode ser analisada independentemente da realização contextualizada. Já o enunciado é uma sequência realizada na instância da entonação, e ele pode ser interpretado pragmaticamente de forma independente (ele é autônomo do ponto de vista da pragmática).

Considerando todos esses fatores, vemos que também a segmentação da fala implica uma nova abordagem, que considere o enunciado como unidade de referência e, logo, seguindo parâmetros diferentes. Conforme observa Bossaglia, “[Para a Teoria da Língua em Ato] a sintaxe no estudo da fala espontânea necessita ser estudada considerando a dimensão da articulação informacional, e, portanto, de forma diferente de como é tradicionalmente feita a análise sintática da escrita.” (BOSSAGLIA, 2014).

O Quadro 2 mostra as principais diferenças na relação fala e enunciado e escrita e frase, conforme proposta de Cresti (2000):

Quadro 2 – Fala e enunciado x Escrita e sentença

| FALA | ESCRITA |
|----------------------------------------|----------------------------------------------|
| - Unidade de referência: enunciado. | - Unidade de referência: sentença. |
| - Princípio de organização pragmático- | - Princípio de organização lógico-sintático, |

| | |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>ilocutório, com base no enunciado.</p> <ul style="list-style-type: none"> - Enunciado é interpretável pragmaticamente e se realiza pela entonação. - Na fala (espontânea) não há tempo para programação ou correção da mensagem. | <p>com base na frase.</p> <ul style="list-style-type: none"> - Frase é interpretável sintática e semanticamente e pode ser realizado independente do contexto. - Na escrita, há várias possibilidades de correção e estas podem acontecer em diversos momentos. |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Fonte: Elaborado pela autora com dados extraídos de RASO (2013).

Conforme é possível verificar, as diferenças entre fala e escrita são muitas e considerá-las de forma correta é imprescindível para o sucesso de análises baseadas em língua de fala espontânea. Considerando que cada modalidade é compreendida a partir de suas características próprias, o conhecimento das distinções, como a unidade de referência, ressalta a necessidade de se aplicarem metodologias específicas para o tratamento de dados da fala, a fim de que não haja a perda de dados importantes para compreensão dos eventos orais, como ocorre, por exemplo, nos estudos baseados apenas em transcrições de fala, tampouco que se faça análises errôneas de fenômenos em função da ausência de parâmetros sonoros ou de critérios sintáticos bem definidos.

3. Transcrição de fala: principais implicações

Conforme visto nos tópicos anteriores, os critérios da análise de dados da fala e da escrita devem ser diferentes, uma vez que ambas as modalidades da linguagem apresentam particularidades distintas entre si. No caso da fala, especificamente, a abundância de eventos típicos da língua oral (como as redundâncias, os ruídos paralinguísticos, as ligações não gramaticais etc.), torna a tarefa de análise bem mais complicada e, por isso, a cautela com a coleta e com o tratamento dos dados deve ser ainda maior.

Nas investigações baseadas em corpus, é comum vermos estudos da fala sendo realizados a partir de transcrições em língua escrita. Entretanto, como demonstrou Raso (2013), a representação da língua oral por meio da escrita não é o formato ideal para o estudo da fala espontânea, uma vez que a ausência de dados essenciais (como a prosódia) inviabiliza a visualização da estruturação desse tipo de fenômeno. Segundo o autor, “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.” (RASO, 2013, p. 6). Isso, porque, as transcrições não são capazes de evidenciar aspectos que são claramente interpretados por meio do som.

À parte dessa característica, podemos considerar também que a tarefa da transcrição da fala apresenta dificuldades que se relacionam à capacidade técnica e ao preparo do transcritor. Um primeiro problema diz respeito à percepção, ou seja, à forma como o transcritor percebe e interpreta aquilo que escuta. Como apontam Blanche-Beneviste e Jeanjean, “[...] escutar é uma atividade complexa [...] estamos sempre prontos a escutar o que acreditamos plausível” (BLANCHE-BENEVISTE; JEANJEAN, 1987). Assim, se o transcritor não tiver um bom treinamento, é bem provável que os dados transcritos apresentem problemas e, conseqüentemente, se tornem inválidos para representar um determinado fenômeno.

Outra dificuldade que pode ser apontada na tarefa da transcrição tem relação com a legibilidade, ou seja, se não houver critério e um trabalho cauteloso, a maneira como a

transcrição é realizada pode prejudicar a leitura do pesquisador e comprometer o trabalho. Conforme ressalta Goffman, “O transcritor ingênuo será vítima de sua ignorância e de todos os fenômenos ligados à reconstrução; ouvinte não advertido, ele arrisca entender mal, mesmo tendo boa vontade” (GOFFMAN, 1981). Logo, novamente, a tarefa de transcrição deve ser realizada com muita atenção e o profissional deve não somente estar ciente dos objetivos do corpus com o qual está trabalhando, mas também estar preparado tecnicamente para realizar a tarefa de transcrição. Conforme explicam Blanche-Beneviste e Jeanjean, “É preciso lhe [transcritor] dar uma formação mínima [...]. Ele deve ter uma ideia referente ao objetivo da transcrição [...]. Coletar uma quantidade de dados e identificar somente depois o que será utilizado não é uma boa maneira de começar o trabalho.” (BLANCHE-BENEVISTE; JEANJEAN, 1987).

Finalmente, por mais que seja feita com zelo, conforme já mencionado, por se tratar de uma representação escrita da fala, a transcrição suprime informações importantes da produção oral. Conforme explica Encrevé (1992)

Durante muito tempo os linguistas trabalharam a oralidade após o término da transcrição [dos dados coletados]. [...] Se transcrevemos o oral, fazemos dele escrita. É preciso preservar toda a extraordinária especificidade do oral, todas as marcas que não encontram correspondentes na escrita, mesmo com o auxílio dos alfabetos fonéticos mais completos” (ENCREVÉ, 1992, p. 104).

Essa visão também é compartilhada por Raso (2013), que explica que

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). (RASO, 2013, p. 26).

Dessa forma, o autor sugere que a maneira mais adequada de se analisar eventos da língua oral é por meio de uma metodologia que considere o alinhamento correto entre a transcrição e o sinal acústico da fala, pois esse procedimento permite ao pesquisador ter acesso aos dados de forma precisa e torna possível analisar os dois parâmetros ao mesmo tempo, aumentando a confiabilidade da tarefa e dos resultados.

4. Conclusão

A escrita e a fala são modalidades da língua que, apesar de serem relacionadas, apresentam profundas diferenças no que diz respeito à forma como devem ser compreendidas. Haja vista as peculiaridades de cada modalidade, é importante que o pesquisador tenha em mente que a análise de eventos da fala deve ser realizada com base em parâmetros específicos desse tipo de linguagem. Da coleta de material à análise dos dados, devem ser considerados os fatores que preservem as características do evento da fala espontânea. Não se trata de

verificação de transcrição bruta, mas do conjunto de suporte teórico, metodológico, etiquetagens e alinhamentos adequados.

Referências

BLANCHE-BENEVISTE, C.; JEANJEAN, C. *Le français parlé: édition et transcription*. Paris: Didier-Erudition, 1987.

BOSSAGLIA, G. Interface entre sintaxe e articulação informacional na fala espontânea: uma comparação baseada em corpus entre português e italiano. *Caligrama: Revista de Estudos Românicos*, n. 19, v. 2, 2014. p. 35–60.

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

ENCREVÉ, P. *Actualité de l'enquête et des études sur l'oral (Interventions à la Table ronde tenue à Paris, le 12 jan. 1991, avec J.-Cl. Chevalier comme modérateur)*. 1992.

GOFFMAN, E. *Forms of talk*. Filadélfia, University of Pennsylvania Press, 1991.

MARCUSCHI, L. A. *Da fala para a escrita: atividades de retextualização*. 4. ed. São Paulo: Cortez, 2010.

RASO, T. Fala e escrita: meio, canal, consequências pragmáticas e linguísticas. *Domínios de Linguagem*, n. 7, v. 2, 2013. p. 12–46.

SILBER-VAROD, V. *The SpeeCHain Perspective: Prosody-Syntax Interface in Spontaneous Spoken Hebrew*. PhD dissertation. Tel Aviv University, 2011.