



Universidade de Brasília

Instituto de Letras

Departamento de Linguística, Português e Línguas Clássicas

ALINE XAVIER DE ARAÚJO

**Uma análise sintática de fenômenos de
hipossegmentação e hipersegmentação**

**Brasília
2016**

ALINE XAVIER DE ARAÚJO

**Uma análise sintática de fenômenos de
hipossegmentação e hipersegmentação**

Monografia apresentada ao Departamento de Linguística, Português e Línguas Clássicas da Universidade de Brasília como pré-requisito parcial para obtenção de grau de Bacharelado em Letras, habilitação Língua Portuguesa e Respectiva Literatura.

Orientador: Prof. Dr. Marcus Vinicius Lunguinho

**Brasília
2016**

AGRADECIMENTOS

Agradeço primeiramente à todos os professores que fizeram parte da minha formação acadêmica pela maestria com a qual compartilharam seus conhecimentos, incentivando-me a sempre ir além.

Agradeço, especialmente, ao meu orientador - Professor Marcus Vinicius Lunguinho - por todo o apoio e orientações prestadas ao longo da elaboração trabalho, se mostrando sempre solícito e disposto a auxiliar-me.

Por último, agradeço aos meus pais por me incentivarem em todas as minhas escolhas e por estarem sempre ao meu lado.

RESUMO

Este trabalho tem o objetivo de estudar os desvios de segmentação ortográfica encontrados no ambiente virtual por meio de um viés sintático, levando em conta a influência das demais áreas de estudo da língua. Os dados foram retirados, em sua maioria, do site *yahoo respostas* e foram divididos de acordo com suas possíveis motivações - ortofônica, sintática e funcional. Foram evidenciados, ainda, desvios fronteiriços, ou seja, que obedeciam à mais de uma motivação. Esta classificação permitiu inferir a lógica por trás das ocorrências bem como entendê-las à luz de diferentes perspectivas. Dessa forma, procurou-se demonstrar a importância da perspectiva sintática na compreensão dos fenômenos de segmentação assim como o fato de que não são exclusivos às produções do período inicial da alfabetização. Ao longo da análise foi retificada a barreira tênue que existe entre a atuação da sintaxe e da morfologia, já apontada por MIOTO (2013). Ademais, foram levadas em conta as dificuldades que os falantes da língua possuem ao adquirir a modalidade escrita e como tais problemas podem se propagar pela escrita adulta na análise dos desvios. Por último, foi evidenciada - por meio dos desvios - a capacidade criativa dos autores e como são capazes de assimilar estruturas e regras da língua já conhecidas à novas construções.

Palavras-chave: Hipossegmentação. Hipersegmentação. Sintaxe. Escrita. Fronteiras vocabulares.

ABSTRACT

This paper aims to study the orthographic sectioning errors found in virtual interactions under a syntactic perspective, taking into account the influence of other language study fields. The data was found at the *yahoo respostas* website and then divided according to each possible motivation - orthographic, syntactic and functional. There were also border errors, that is, that belong to more than one motivation. This division has allowed seeing the logic behind such errors as well as understand them under different perspectives. Hence, it was sought to demonstrate the relevance of a syntactic view in understanding such errors and the fact that they are not exclusive to the initial literacy acquisition period. Throughout the analysis it was detrended the thin barrier that exists between the performance of syntax and morphology, as indicated by Mito (2013). Therefore, the difficulties faced by apprentices when learning to compose and how they can perpetuate through one's writing were taken into account when analysing the results. At last, the ability to create and assemble new structures to already known patterns was also identified.

Key-words: Hiposegmentation. Hipersegmentation. Syntax. Writing process. Vocabulary boundaries.

SUMÁRIO

INTRODUÇÃO.....	8
1. A modalidade escrita da língua.....	10
1.1 A aquisição da língua escrita.....	10
1.2 Escrita e oralidade.....	14
1.3 A questão da palavra.....	16
1.4 Hiposegmentação e hipersegmentação.....	20
1.5 Conclusões preliminares.....	23
2. METODOLOGIA DE PESQUISA.....	23
2.1 Introdução.....	24
2.2 Classificação dos dados.....	25
2.2.1 Desvios por similaridade ortofônica.....	28
2.2.2 Desvio por similaridade funcional.....	28
2.2.3 Desvio por assimilação sintática.....	29
2.3 Conclusões preliminares.....	31
3. ANÁLISE DOS DADOS.....	32
3.1 Os casos fronteiriços.....	34
3.1.1 Ortofônico + funcional.....	34
3.1.2 Ortofônico + sintático.....	34
3.1.3 Os casos que envolvem as três motivações.....	35
3.1.3.1 A crase em <i>o deio</i>	35
3.1.3.2 Segmentação da desinência.....	36

3.2 Casos interessantes.....	39
3.2.1 O caso do <i>aseguir</i>	39
3.2.2 A preposição <i>de</i>	40
Considerações finais.....	41
Referências Bibliográficas.....	43

Introdução

A hipossegmentação e hipersegmentação são fenômenos – ausência e inserção de espaço nas fronteiras vocabulares, respectivamente – frequentemente estudados devido aos diferentes contextos em que aparecem. Com efeito, majoritariamente estuda-se a questão em textos infantis pois este é o início da conquista de novos caminhos da língua - novas formas de expressão. Sendo assim, grande parte da literatura existente acerca do tema (cf. VELOSO 2003, CUNHA e MIRANDA 2009 e GUIMARÃES 2011) apresenta argumentos fonológicos ou morfossintáticos para explicar a dificuldade do aprendiz em identificar tais fronteiras vocabulares ainda no período da alfabetização.

Este trabalho, por outro lado, procura estudar as possíveis motivações sintáticas para a ocorrência da hipossegmentação e da hipersegmentação, sem desconsiderar a influência dos demais subsistemas da língua, uma vez que a sintaxe é a "disciplina linguística que estuda como combinamos palavras para formar sintagmas e como combinamos sintagmas para formar sentenças", ou seja, parte das palavras - produto da morfologia - para formar combinações, como afirmam Miotto, Silva e Lopes (2013). Sendo assim, parte-se de uma ótica da escrita em sua totalidade, levando em conta as grandes contribuições feitas a partir de textos da alfabetização. Esta ótica, contudo, não foi escolhida aleatoriamente uma vez que os exemplos de segmentação da escrita aqui analisados refletem a *língua em uso*, independente da individualidade do autor.

Para compreender melhor a origem destes “equivocos” gráficos é necessário refletir acerca da relação indivíduo-língua e posteriormente, entender a aquisição da escrita e da leitura. Os processos mentais envolvidos nestas aptidões são intrínsecos implicando que muitas das habilidades exigidas para escrever também se fazem necessárias para ler. Por outro lado, ambas são fases ulteriores e mais elaboradas da fala, que é o primeiro contato do indivíduo com a língua e por isso a questão é abordada também do ponto de vista cognitivo da aquisição por meio de estudos canônicos da aquisição da escrita.

Ademais, se discute a questão do *erro* na perspectiva escrita e na oral uma vez que o contraste entre as duas concepções elucida questões relacionadas ao valor social(?) que estas modalidades assumem. Por último, será discutida a terminologia aplicada aos desvios de segmentação ortográfica bem como a importância deles para o ambiente escolar, principalmente pelo o viés sintático.

Capítulo 1

A modalidade escrita da língua

1.1 A aquisição da língua escrita

A linguagem humana, o principal objeto de estudo da linguística, se configura de forma complexa pelo seu caráter abstrato e inerente às relações humanas. Sendo assim, a partir da metade do século XX, consideráveis avanços sobre a linguagem humana permitiram elucidar o papel da comunicação - e em quais parâmetros ela se dá - no desenvolvimento humano como todo. Para entender os desvios aqui discutidos serão consideradas duas concepções gerativistas acerca da língua: a Teoria dos Princípios e Parâmetros e a capacidade criativa do falante.

A Teoria dos Princípios e Parâmetros, por sua vez, se refere às regras gerais que se aplicam à todas as línguas naturais (princípios) e às propriedades exclusivas de cada língua responsáveis por diferenciá-la das demais (parâmetros). Este conceito está diretamente relacionado à **gramaticalidade** pois para as expressões linguísticas pertencerem a determinada língua precisam obedecer aos Princípios da faculdade da linguagem e aos Parâmetros da língua em questão. Inerente à tal concepção existe a noção de competência e desempenho do falante, ou seja, o indivíduo - baseado na gramática da língua - julga as estruturas como aceitáveis ou não a partir da **competência** que possui. Por outro lado, quando utiliza este conhecimento para construir (e compreender) sentenças, o **desempenho** do falante é ativado. Acerca da competência e do desempenho, Miotto, Lopes e Silva (2013) afirmam que

o que permite ao falante decidir, então, se uma sentença é gramatical ou não é o conhecimento que ele tem e que tem o nome técnico de **competência**. Quando o falante põe em uso a competência para produzir sentenças que ele fala, o resultado é o que chamamos de **performance** (ou **desempenho**).

De certo modo entender a aquisição da linguagem - e as concepções acerca dela - permite traçar um contraste entre a aquisição das demais formas de expressão da língua. Apesar de a aquisição da escrita exigir - além do domínio da

língua - um conjunto de habilidades cognitivas, os indivíduos se deparam com o mesmo paradigma de desempenho e competência, afinal a escrita também é uma forma de linguagem. Isto significa que nesta modalidade são colocadas em jogo questões como *gramaticalidade* e *aceitabilidade*¹. A aceitabilidade, porém, de maneira diferente uma vez que a escrita, socialmente, é tomada como a modalidade formal da língua, como afirma Marcuschi (1997:120): "A fala é adquirida naturalmente em contextos informais do dia-a-dia. A escrita, em sua faceta institucional, se adquire em contextos formais: a escola. Daí também seu caráter mais prestigioso como bem cultural desejável."

A gramaticalidade, com efeito, funciona da mesma forma para as mais variadas modalidades da língua, tendo em vista que contempla todas as sentenças pertencentes ou não (agramaticais) à uma dada língua, sendo que "quem sabe decidir se uma sentença pertence ou não a uma dada língua é o falante nativo daquela língua, escolarizado ou não." (Miotto, Silva e Lopes 2013:16). A sintaxe, por sua vez, parte desta capacidade de julgar uma sentença como gramatical ou não para entender como se dá a estrutura da língua e de que forma os elementos de uma sentença se combinam para formar uma frase gramatical.

A noção de aceitabilidade, ao contrário, depende muito mais de fatores de natureza individual do que propriamente da gramática internalizada do falante. Isso porque é individual, ou seja, para um grupo de falantes nativos de uma mesma língua existem construções que são aceitas por alguns e por outros, não. Miotto, Silva e Lopes (2013:19) ao compararem gramaticalidade e aceitabilidade, afirmam que:

(...) o falante nativo pode dizer se uma frase é aceitável ou inaceitável na sua língua materna, mas cabe ao linguista dizer se, no caso de uma sentença inaceitável para o falante, estamos diante de uma sentença realmente agramatical ou se a razão para a inaceitabilidade deve ser computada a outros fatores (como aqueles relativos à **performance**).

¹ Embora ambas trabalhem com a noção de "erro", gramaticalidade é entendida como a capacidade que o falante possui em reconhecer sentenças como pertencentes ou não à língua enquanto a aceitabilidade está relacionada à capacidade individual de julgá-las apropriadas independente de serem gramaticais ou não, ou seja, todas as sentenças agramaticais são inaceitáveis mas nem todas gramaticais são aceitáveis.

Sendo assim, pode-se pensar que o desempenho (*performance*) dos falantes na escrita se diferencia em relação ao da fala. Isto implica dizer que existem processos os quais transcendem a consciência sintática e fonológica do falante na escrita uma vez que tangem a estilística do texto. Mas até que ponto entender o funcionamento e aquisição de uma língua ajuda a esclarecer os processos da aquisição da escrita? Segundo Kato (1987:120),

O desenvolvimento da escrita parece ainda seguir alguns princípios já postulados para a aquisição da fala. A nível microestrutural, as falhas na leitura e os desvios na produção escrita podem ser explicados como decorrentes do uso automático das mesmas estratégias usadas na fala, o que nos leva a crer que exigem princípios que regem a produção e a compreensão, independentemente da modalidade.

O gerativismo, por sua vez, permitiu inovações no campo da aprendizagem e do ensino; mais especificamente na maneira de ensinar - a ler e escrever, por exemplo - pois, a partir da perspectiva gerativista, compreende-se que o indivíduo já possui conhecimentos acerca da língua que lhe será apresentada, afinal, ele a utiliza desde pequeno, como afirmam Ferreiro e Teberosky (1985:21):

"(...) é fácil mostrar que muitas das práticas habituais no ensino da língua escrita são tributárias do que se sabia (antes de 1960) sobre a aquisição da língua oral; a progressão clássica que consiste em começar pelas vogais, seguidas da combinação de consoantes labiais com vogais e a partir daí chegar à formação das primeiras palavras por duplicação dessas sílabas (*mamá, papá*), e, quando se trata de orações, começar pelas orações declarativas simples, é uma série que reproduz bastante bem a série de aquisições da língua oral, tal como ela se apresenta vista "do lado de fora" (isto é, vista desde as condutas observáveis, e não desde o processo que engendra essas condutas observáveis). Implicitamente, julgava-se ser necessário passar por essas mesmas etapas quando se trata de aprender a língua escrita, como se essas aprendizagem fosse uma aprendizagem da fala."

Por conseguinte, a educação sofreu grande impacto a partir do entendimento que o indivíduo não é uma "folha em branco" quando inicia a alfabetização, ou seja, já possui conhecimentos acerca da língua que estudará, como apontam as autoras. Seguindo o mesmo princípio - de que o aprendiz não é desprovido de conhecimentos acerca daquilo que lhe será ensinado - a teoria gerativa entende que a competência do falante lhe permite hipotetizar sobre a língua, aplicando regras e generalizações à novas construções devido à capacidade de criar novas estruturas a partir daquelas já familiares.

Paralelamente, a corrente piagetiana discute a questão da aquisição - mesmo que indiretamente - sob o ponto de vista cognitivo, ou seja, enquanto desenvolvimento da inteligência organizada por *esquemas de ação*:

Pegar uma vareta, para puxar um objeto distante, é assim um ato de inteligência (...). Neste ato, um meio, que é um verdadeiro instrumento, é coordenado a um objeto previsto; no exemplo da vareta, é preciso compreender, antecipadamente, **a relação entre ela e o objeto, para descobri-la como meio.** (...) De outro lado, os "esquemas" de ação, construídos desde o nível do estágio precedente e multiplicados graças a essas novas condutas experimentais, tornam-se suscetíveis de se coordenarem entre si, por assimilação recíproca, tal como farão **mais tarde as noções ou conceitos do pensamento.** Com efeito, uma ação apta a ser repetida e generalizada para situações novas é comparável a uma espécie de conceito senso-motor. É assim que, em presença de um novo objeto, ver-se-á o bebê incorporá-lo sucessivamente a cada um de seus esquemas de ação (agitar, esfregar ou balançar o objeto), como se se tratasse de compreendê-lo através do uso. (PIAGET, 1999:19-20) - grifos meus

Pensando acerca do exemplo fornecido pelo autor é possível entender a linguagem - por extensão, a escrita também - como o meio/instrumento pelo qual a fala (ou o texto) são conquistados. Consequentemente, a *assimilação* em Piaget - composta pela fase funcional, de reconhecimento e a generalização - corresponde, na aquisição da linguagem, às hipóteses que o aprendiz testa ao entrar em contato com um novo objeto (ou estrutura da língua). Dessa forma, justifica os equívocos da fala cometidos quando se começa a adquirir a linguagem como tentativas de aplicar o conhecimento já assimilado à novas construções (generalização); em termos práticos, a criança conjuga verbos irregulares como regulares, por exemplo, aplicando a mesma regra geral às exceções.

A escrita, por sua vez, apresenta em sua aquisição fenômenos de motivação semelhante, como os de ordem morfológico-gramatical que ocorrem, em parte, por conta da ideia de que a relação fonema-grafema é unívoca. Existem, ainda, alguns fenômenos na segmentação da escrita que também encontram respaldo nessa capacidade de utilizar estruturas já conhecidas em diferentes contextos, como será apresentado adiante.

1.2 Escrita e oralidade

De acordo com Ferreiro e Teberosky (1985), a aprendizagem da escrita não se limita aos conhecimentos específicos exigidos do aluno ou aos métodos de ensiná-la: deve ser encarada como um processo de construção de conhecimentos. Durante a alfabetização são exigidas do indivíduo diversas habilidades que nem sempre lhe vêm de maneira natural. Passado este momento, o aprendiz se depara com uma questão que problematiza ainda mais a habilidade de escrever: o domínio da capacidade de se expressar por meio da escrita tão bem quanto na fala, o que implica a eterna busca pelo texto coerente e adequado à norma padrão da língua, tornando escrita um objeto a ser conquistado todas as vezes que é utilizado.

Sendo assim, apesar de ser uma convenção, a escrita não se resume à memorizar palavras e transcrevê-las pois também se relaciona com a capacidade criativa do falante como forma de expressão do pensamento interiorizado, ou seja, evoca a habilidade de transformar ideias organizadas de forma complexa em um produto linear, coeso e conciso. Flower e Hayes (1981) - que estudam a escrita do ponto de vista cognitivo - postulam, entre outras questões, que "o processo da escrita é melhor compreendido enquanto um conjunto de processos mentais distintos que os autores orquestram ou organizam durante a composição de seu texto." (página 366)². Mesmo a transcrição não pode ser pensada puramente como a adequação de uma forma linguística à outra pois, segundo os autores, consiste no "processo de colocar ideias em linguagem visual" (página 373)³ que, por sua vez, exige o domínio de habilidades cognitivas que não se manifestam de maneira automática como a linguagem.

Uma segunda capacidade exigida para o desenvolvimento da escrita é compreender que o texto escrito se diferencia de diversas maneiras daquele falado, pois "fala e escrita são, portanto, duas modalidades da língua. Assim, embora se utilizem do mesmo sistema linguístico, cada uma delas possui características próprias. ou seja, a escrita não constitui mera transcrição da fala" (Koch e Elias, 2009:29). Além disso, a princípio, é esperado que o produto deste esforço obedeça às normas da língua padrão o que implica que seu autor tem que conhecê-las e

² No original: "The process of writing is best understood as a set of distinctive thinking processes which orchestrate or organize during the act of composing".

³ No original: "Translating - this is essentially the process of putting ideas into visible language."

dominá-las bem. Mas então só são considerados textos escritos aqueles que obedecem a tais princípios? Ou, sendo a escrita uma modalidade da língua, são esperadas produções que variem nestes aspectos?

Marcuschi (1997), em *Oralidade e Escrita*, se ocupa em responder tais questões, entendendo o lugar destas duas modalidades na sociedade atual ao estudar como esta relação é compreendida por correntes distintas e quais os efeitos destas interpretações para a educação. Com efeito, sustenta a fusão de duas visões acerca da dicotomia fala/escrita: a perspectiva interacional com a variacionista. Segundo o autor, a primeira defende um *continuum* tipológico no qual de um lado está a escrita e do outro, a oralidade; contudo, desconsidera alguns aspectos importantes da língua, como a fonologia e sintaxe. Ao se juntar à visão variacionista, por outro lado, obtém-se resultados mais produtivos nas correlações entre formas linguísticas pois essa visão

trata do papel da escrita e da fala sob o ponto de vista de processos educacionais e faz propostas específicas a respeito do tratamento da variação na relação entre padrão e não-padrão linguístico nos contextos de ensino formal (MARCUSCHI, 1997:131).

A fusão das duas perspectivas serve ao propósito do autor tendo em vista que ele busca desconstruir a superioridade da língua escrita em relação à língua falada. Ademais, defender a hipótese de que "As diferenças entre a fala e a escrita se dão dentro do *continuum* tipológico das práticas sociais de produção textual e não na relação dicotômica de dois polos opostos" (página 136) permite um olhar diferenciado para o texto e as práticas estabelecidas *a partir* dele.

O *continuum* tipológico, assim, se apresenta como uma interpretação produtiva da relação fala/escrita. Tendo em vista que as ocorrências estudadas neste trabalho foram retiradas do ambiente virtual, entendê-las dentro deste *continuum* ajuda a esclarecer dois pontos importantes: a questão do "erro" e a motivação linguística por trás destes fenômenos, uma vez que, como será apresentado adiante, o indivíduo utiliza de uma lógica respaldada na gramática da língua para acrescentar ou ignorar as fronteiras vocabulares da palavra. Além disso, este entendimento ajuda a desmistificar concepções acerca dos dois processos bem como abarca um grande leque de possibilidades textuais nas quais estão incluídas as interações virtuais.

1.3 A questão da palavra

Compreender o conceito de palavra - e que na modalidade escrita ela assume um formato diferente - é essencial para o início da aquisição da escrita. Contudo, a própria natureza da escrita torna este processo problemático: por ser uma convenção, a relação existente entre grafemas e fonemas é arbitrária, em outras palavras, não há nenhuma característica no grafema que remeta à sua realização sonora, quanto mais na palavra como um todo. É válido ressaltar que grafemas são a realização escrita dos fonemas, ou seja, unidades que permitem distinguir palavras na língua. Uma letra, por exemplo, é letra enquanto representa apenas o fonema mas se torna grafema ao exercer função distintiva na língua.

O processo da leitura, assim, seria a conversão de grafemas em fonemas e o da escrita, de fonemas em grafemas. Esta associação, entretanto, se levada a momentos posteriores na aquisição, causa conflitos à medida que o falante se depara com fonemas que correspondem à grafemas distintos dependendo da palavra (como o caso do 's' e do 'z', por exemplo). Com efeito, serão estudados os problemas existentes nas tentativas de conceitualizar a palavra e como esta definição pode afetar a aquisição da escrita.

Pensando na diferenciação entre palavras por meio de unidades significativas, Basílio (1987:13) afirma que as palavras são "uma dessas unidades linguísticas que são muito fáceis de reconhecer mas bastante difíceis de definir" pois a questão é complexa já que

Normalmente, a diferença entre palavras distintas e diferentes formas da mesma palavra é colocada a partir da diferença entre flexão e derivação. O problema é que não há uma distinção nítida e definitiva entre os conceitos de flexão e derivação, e muitas vezes a diferença entre os dois conceitos é colocada como decorrente da diferença entre "palavras distintas" e "formas de uma mesma palavra". Forma-se, portanto, um ciclo vicioso.

Sob o ponto de vista morfológico, a autora afirma que as palavras são formadas por um afixo mais uma raiz sendo que esta "é um morfema que pode, por si só, construir a base de uma palavra. [...]" enquanto "Os elementos que se acrescentam à raiz para formar uma palavra são chamados afixos" que podem, por sua vez, ser sufixos ou prefixos. Ocorre que a palavra pode ter sua raiz composta por outra base mais afixo, formando uma palavra morfológicamente complexa,

sendo assim, "[...] a palavra não é formada de uma sequência de morfemas, mas constituída estruturalmente de uma base acrescida de afixo."⁴ A complexidade da formação das palavras deveras dificulta o processo da aquisição da escrita, entretanto, em um primeiro momento, a maior dificuldade do aprendiz é encontrar uma relação lógica entre a realização sonora e sua grafia. Para tanto, o aluno necessita entender o caráter arbitrário da convenção escrita encontrando diferentes caminhos.

A respeito do desenvolvimento desta noção "grafema-fonema", Ferreiro e Teberosky, em *Psicogênese da língua escrita* (1985: 184-185), observaram por meio de amostras da escrita infantil uma tentativa por parte dos aprendizes em encontrarem alguma conexão entre a grafia da palavra e o signo que ela representa:

Estes dados e outros recolhidos nos mais diversos contextos **evidencia uma tendência da criança a tratar de refletir na escrita algumas das características do objeto.** (...) O notável é que, até agora, não encontramos exceção a esta regra: *a correspondência se estabelece entre aspectos quantificáveis do objeto e aspectos quantificáveis da escrita*, e não entre aspecto figural do objeto e aspecto figural do escrito. Isto é, não se buscam letras com ângulos marcados para escrever "casa", ou letras redondas para escrever "bola", mas sim um maior número de grafias, grafias maiores ou maior comprimento do traçado total se o objeto é maior, mais comprido, tem mais idade ou há maior número de objetos referidos. (grifos meus)

Ou seja, em um primeiro momento, existe uma tentativa por parte da criança em criar conexões lógicas entre aquilo que a palavra significa e sua grafia.

De todas as formas, os alfabetizando assimila conceitos gerais para posteriormente compreenderem as particularidades da convenção escrita. Tendo em vista que a palavra é a junção dos grafemas previamente apresentados, também serão entendidas como um correspondente gráfico de uma sequência sonora, entretanto, esta associação é complexa, como afirmam Ferreiro e Teberosky (1985:106):

Se tomamos a definição "intuitiva" de palavra como associação de um sentido a um pattern sonoro específico, nos encontramos de imediato com várias dificuldades: pelo lado sonoro do sentido, torna-se claro que as distintas classes de palavras (substantivos, verbos, artigos, preposições, etc.) veiculam sentidos muito diferentes. Por outro lado, unidades menores que as palavras contribuem para transmitir diferentes sentidos (como o -s do plural). Pelo lado do pattern sonoro, as dificuldades também

⁴ morfema é entendido como a menor unidade linguística provida de significado

abundam. Se considerarmos o problema dos verbos, torna-se claro que "cantamos" ou "cantarei" podem ser consideradas como duas palavras diferentes ou como duas ocorrências da mesma palavra, etc.

A questão da definição de palavra leva, ainda, ao questionamento das fronteiras vocabulares dado que, como será apresentado adiante, evidenciam a consciência por parte do autor ao grafar *começar-mos*, demonstrando reconhecer que a partícula *-mos* acrescenta um significado diferente à raiz, ou seja, elucida que a forma infinitiva deste verbo pode se juntar à diferentes afixos para formar palavras distintas com significados também distintos⁵.

Sendo assim, conceituar palavra é uma tarefa delicada na medida que ela se associa, ao mesmo tempo, a diferentes signos:

Na palavra, entendida como uma unidade lexical, uma sequência fônica se associa de modo relativamente estável a (a) um significado ou conjunto de significados; (b) um conjunto de propriedades sintáticas; (c) um conjunto de propriedades morfológicas e (d) um conjunto de determinações de uso. (BASÍLIO, 2009:10)

ou seja, o aprendiz da escrita leva em conta - mesmo que inconscientemente - todas essas associações quando produz um texto. Com efeito, estudar os desvios provenientes deste conceito implica em também investigar as propriedades morfológicas, sintáticas e fonéticas da palavra.

O conceito de palavra para Mattoso Câmara Jr., pai da linguística estruturalista no Brasil, sob o olhar de Margarida Basílio, põe em jogo uma outra problemática da conceptualização, chamando atenção em especial para a noção desenvolvida a partir da escrita:

A extrema preocupação de Mattoso Câmara em colocar os produtos da linguística a serviço do conhecimento da língua portuguesa leva ao primeiro problema fundamental encontrado em relação à palavra: o fato de que "toda e qualquer descrição da língua portuguesa leva em conta a existência do vocábulo", embora não se tenha cogitado, nas abordagens tradicionais, de "explicar e claramente definir em que consiste ele" (Mattoso Câmara 1969: 34). **A falha, naturalmente provém do fato de que nossas gramáticas tomam por base a língua escrita, na qual a palavra é definida pelos espaços em branco, conforme determinado pelo sistema gráfico.** (BASÍLIO, 2004:77 - grifos meus)

⁵ Ver tópico 4.2.3.1

Esta definição problemática pode ser um dos fatores-chave que leva os indivíduos a hesitarem enquanto escrevem sem saber ao certo como funcionam as fronteiras vocabulares. Em outras palavras, tomar a forma escrita como base para definir o conceito (inserção ou não de espaços brancos) implica em entender a palavra por meio de um sistema arbitrário, permitindo a ocorrência de hipossegmentações e hipersegmentações.

O trabalho desenvolvido por Guimarães (2011), em que são estudados os fenômenos da segmentação a partir da consciência morfossintática em alunos do 4o e 5o anos, parte da premissa de que como o sistema ortográfico possui aspectos fonológicos e morfossintáticos, "parece perfeitamente legítimo inferir que a aprendizagem da leitura e da escrita é influenciada tanto pela consciência fonológica como pela morfossintática." (página 24). Ao desenvolver testes com os sujeitos estudados, a pesquisadora encontrou em seus dados indicações de como é construída a noção de palavra e procurou comprovar a hipótese segundo a qual "nas línguas alfabéticas tanto a consciência fonológica quanto a morfossintática exercem papel fundamental na aprendizagem da leitura e da escrita." A capacidade morfossintática mencionada consiste em o aprendiz tem a ver com o conhecimentos dos mecanismos da "derivação e da flexão das palavras". Com efeito, a noção de palavra na escrita - para a autora - não necessariamente está bem estabelecida quando o indivíduo começa a escrever, pelo contrário, é desenvolvida ao logo de seu contato com outros textos:

"(...) os dados aqui analisados sugerem que os aprendizes primeiro constroem uma representação escrita das palavras gráficas e só depois de bem estabelecerem esta representação, com o aumento das experiências com a língua escrita, utilizam-na como referência para identificar as palavras escutadas. **Acredita-se que é justamente o desenvolvimento de habilidades morfossintáticas possibilitadas pelo aumento das experiências com a língua escrita que favorece o estabelecimento da noção convencional da palavra.**" (grifos meus)

Desta forma, pode-se afirmar que qualquer noção de palavra previamente adquirida é desconstruída ao passo que se entra em contato com a sua forma gráfica. Com efeito, o conceito de palavra que o indivíduo possuirá após os primeiros contatos com textos escritos será outro e a partir dali, relacionado não somente à realização fonética mas à gráfica também.

Oliveira (2013), em uma perspectiva neurocientífica da linguagem afirma que a palavra escrita diante do leitor

é facetada em milhares de fragmentos, estando perante um sistema de dupla rota: a fonológica e a lexical. A primeira converte a cadeia de letras em fonemas, a outra permite o acesso ao dicionário mental onde o sentido das palavras está armazenado.

Contudo, pelo caráter arbitrário⁶ da convenção, é comum que dificuldades quanto às fronteiras vocabulares apareçam a medida em que se evolui na escrita. Como, por exemplo, saber utilizar corretamente o *por que* e *porque* se, fonologicamente, a representação é a mesma? Argumentar que quanto maior o contato do indivíduo com textos que apresentem este paradigma mais fácil será utilizá-lo na escrita é problemático quando pensamos nas possibilidades que as expressões da língua podem assumir e que a gramática internalizada não funciona como o esquema "copia/cola".

1.4 Hipossegmentação e hipersegmentação

A pergunta "se escreve junto ou separado?" é recorrente quando pensada na construção de um texto pois, na fala, tem-se uma cadeia de fonemas pronunciados que dificilmente obedecem aos limites vocabulares estabelecidos na escrita, afinal, esta leva em conta outros aspectos que dificultam pensar na escrita como transcrição da fala. Sendo assim, a hipossegmentação - caracterizada pela junção de duas palavras que, na ortografia oficial, se escrevem separadamente - geralmente se dá diante de palavras menores que se juntam à maiores enquanto a hipersegmentação com palavras que também existem na ortografia oficial daquela forma (embora signifiquem diferente).

Acerca das motivações fonológicas por trás da hipossegmentação e da hipersegmentação, Cunha e Miranda (2009) estudam esses dois fenômenos em textos infantis, entretanto, com produções de alunos de níveis anteriores aos daqueles estudados em Guimarães (2011) na aquisição da escrita.

⁶ Arbitrariedade aqui, assim como em Saussure, quer dizer que não há relação imediata entre o signo e o significante. Em outras palavras, um *copo* não tem em seu nome nada que remeta ao seu formato, ou seja, surge a partir de uma convenção.

O objetivo das autoras - estudar os constituintes prosódicos nos fenômenos de segmentação de aquisição da escrita - foi alcançado por meio do estudo dos casos de hipossegmentação e hipersegmentação a partir do tipo de palavra, estrutura silábica, tonicidade e o tipo de escola e série. A variável *tipo de palavra*, por sua vez, foi subdividida em palavras gramaticais e palavras fonológicas, permitindo classificar as ocorrências entre as seguintes combinações: *palavra gramatical + palavra fonológica*, *palavra fonológica + palavra gramatical*, *palavra gramatical + palavra gramatical* e *palavra fonológica + palavra fonológica*.

É válido lembrar que, neste caso, palavras gramaticais se referem àquelas que "não possuem significado lexical" e as fonológicas, à "todas as palavras que possuem um acento primário e que, mesmo não tendo significado conhecido na língua, são candidatas potenciais para tal". A análise dos dados apresentou uma predominância da hipossegmentação em relação à hipersegmentação cujas ocorrências foram, em sua maioria, a junção de uma palavra gramatical à uma fonológica.

Pensando nos dados que serão analisados neste trabalho em um segundo momento, pode-se afirmar que a predominância dos casos de hipossegmentação em relação aos de hipersegmentação se confirma, contudo, se apresentam de maneira mais sutil do que na linguagem escrita infantil, por motivos que serão apresentados à frente. Uma segunda constatação pertinente feita pelas autoras é a de que

Nos dados de hipossegmentação, em geral, quanto à variável tipo de sílaba, pôde-se constatar que, ao suprimir os espaços entre as palavras, a criança tende a preservar as estruturas silábicas da língua, e somente em casos raríssimos isso não se verifica. Quando o contexto é favorável o aprendiz usa processos de ressilabação vocálica como a ditongação e a degeminação, conforme mostram os exemplos "siolharão" e "sesquecer". (CUNHA e MIRANDA, 2009:144)

O aprendiz da escrita prefere portanto obedecer ao padrão silábico da língua (CVC) e evita deixar vogais (artigos e pronomes) "sozinhas", produzindo estruturas como: *oque*, *ea*, *ocoelho*.

Pensando no recorte específico dos artigos, Ferreiro e Teberosky (1985) procuram entender a dificuldade das crianças em reconhecê-los na escrita,

primeiramente por não os considerarem uma palavra. Seguindo a linha do pensamento infantil a respeito da questão, as autoras afirmam:

Se o artigo não é uma palavra, não existem, então razões para escrevê-lo (como acabamos de ver); porém, além disso, ocorre que o fragmento de escrita que efetivamente corresponde ao artigo é demasiado pequeno, não tem o número suficiente de letras que a criança exige para que "isso possa ser lido". (1985:116)

Ou seja, além da questão da constituição silábica da língua portuguesa, existe a dificuldade de identificar os artigos como uma palavra, implicando nos equívocos de hipossegmentação: quando reconhecem a presença dos artigos têm a tendência de juntá-los à outra palavra ou sequer representá-los graficamente.

Outros testes feitos pelas autoras para entender como as crianças fazem a segmentação apresentaram resultados curiosos quanto à consciência sintática. Foi lhes apresentada a forma OURSOCOMEMEL, perguntando-lhes se seria necessário a inserção de espaços, sendo que grande parte deles disse que "não" (alguns justificando que se escreve "juntinho" fazendo alusão à letra cursiva). Entretanto, ao serem expostas à seguinte situação, tiveram reações variadas:

Quando propomos, a algumas das crianças que se negam a efetuar cortes no texto, uma situação forçada (se tivesse de separar, por onde?) os resultados são extremamente variados:

- alguns propõem fazer as "separações" acima e abaixo do texto, ou nos extremos, mas de nenhuma maneira dentro do texto;
- outros propõem um número arbitrário de separações que logo não sabem como interpretar;
- outros propõem separar letra por letra;
- **outros, finalmente - poucos - propõem uma separação em duas partes, ficando "urso" numa e "mel" na outra.** (1985:138) - grifos meus

No último caso, embora em menor quantidade, é possível detectar a ativação da consciência sintática tendo em vista que separam o trecho apresentado em sujeito e objeto, ou seja, parecem entender "urso" e "mel" enquanto os núcleos semânticos e principais elementos daquela sentença. Mas será que a ativação desta consciência também ocorre com formas da língua escrita *em uso*, ou seja, desconsiderando o nível de letramento dos falantes e apenas analisando suas produções?

1.5 Conclusões preliminares

É válido ressaltar que a exposição da teoria gerativa e da piagetiana não teve como objetivo a comparação ou a identificação de semelhanças, uma vez que seus respectivos objetos de estudo divergem, apenas foram expostas para esclarecer concepções acerca da linguagem e do conhecimento que servem como base para entender a aquisição da língua escrita.

Por último, dentro de toda a metodologia e considerações que serão feitas, os erros são entendidos como oportunidades de testar o caráter criativo da linguagem e entender como se dá a assimilação de certas características da língua.

Capítulo 2

2. Metodologia de pesquisa

2.1 Introdução

A inquietação em estudar a hipossegmentação e a hipersegmentação na língua escrita *em uso* se deu por frequentemente tais fenômenos também aparecerem fora do ambiente formal de aprendizagem (escola) não estando circunscritos ao processo de letramento. Em outras palavras, no ambiente informal de escrita que a internet se tornou, existe cada vez mais espaço para composições despreocupadas com a modalidade padrão da língua por conta da 1) rapidez com a qual a informação é trocada, 2) porquê os tópicos nos quais os fenômenos aparecem são em si informais além do 3) desconhecimento da forma prevista pela ortografia oficial.

Especificamente nos dados de hipersegmentação e hipossegmentação analisados aqui é possível notar o caráter coloquial uma vez que foram retirados, em sua maioria, do sítio *yahoo respostas* (<https://br.answers.yahoo.com/>), um espaço no qual os internautas têm a liberdade de fazer perguntas. Os fenômenos, contudo, também ocorreram nas respostas implicando que a dúvida em relação à grafia das palavras - *se escreve junto ou separado?* - é recorrente diante de várias palavras diferentes.

É válido ressaltar que muitas das ocorrências apresentadas aqui foram retiradas de perguntas feitas no site acerca da segmentação correta das palavras.⁷ O site possui uma ferramenta de pesquisa que, a partir destes dados, foi utilizada para buscar fenômenos cuja existência já era conhecida. Tal busca retornou outras construções que por vezes obedeciam a um mesmo padrão estrutural.

Foi feito um recorte sincrônico das construções encontradas, ou seja, desconsideram-se aspectos individuais dos autores que pudessem influenciar ou motivar os desvios. Este recorte, portanto, permitiu organizar mais claramente os

⁷ Exemplo extraído em 17/05/2016: "Em cima ou Encima?"
<https://br.answers.yahoo.com/question/index?qid=20080306062602AAzOkQA>

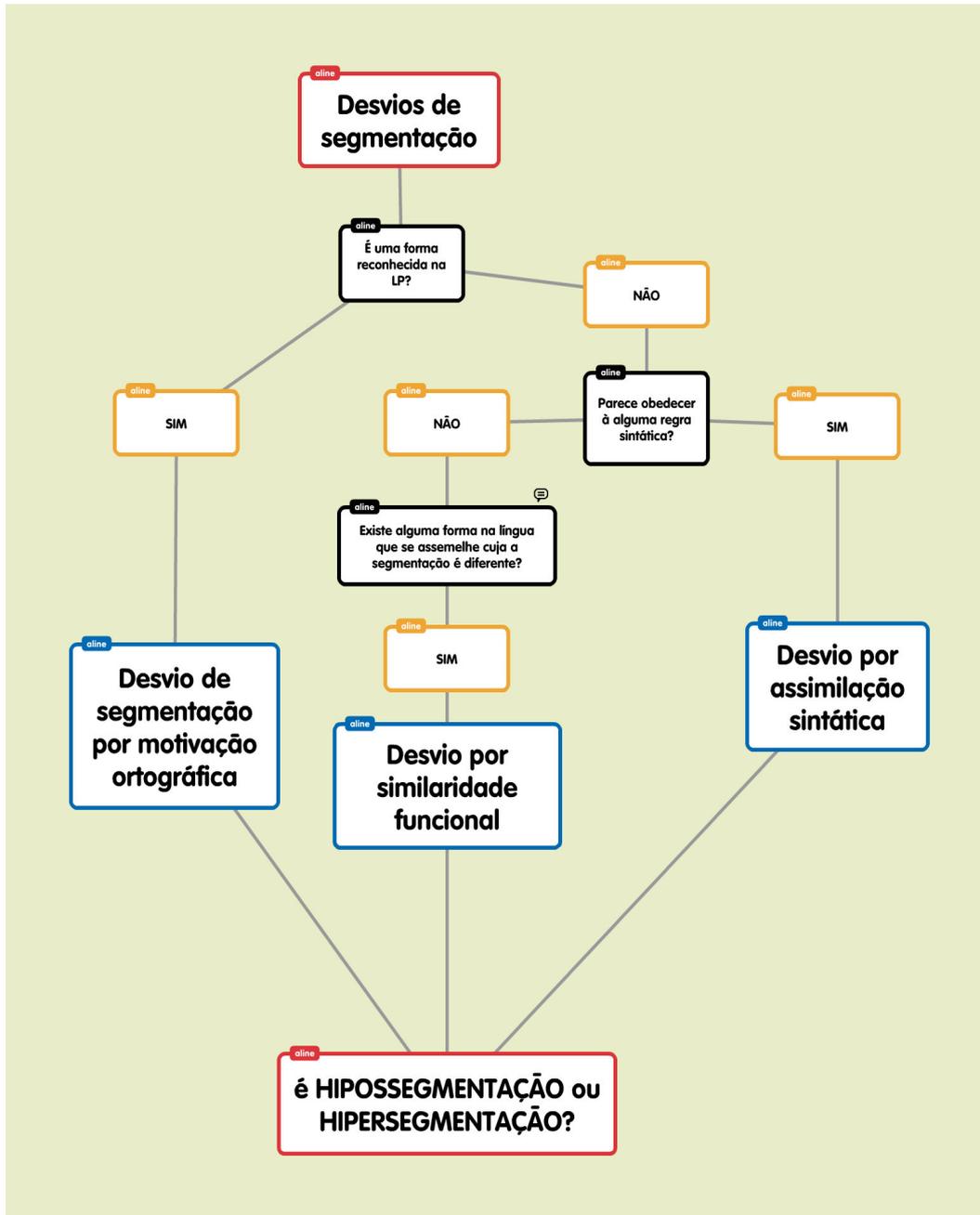
casos encontrados por meio de uma visão da língua em uso (vertente da linguagem que geralmente é palco de mudanças linguísticas).

Esta pesquisa é de cunho qualitativo uma vez que busca entender as ocorrências sozinhas independente de serem muito ou pouco recorrentes na escrita. Em outras palavras, trata da consciência sintática do indivíduo ao compor estas construções, buscando as possíveis associações feitas.

Um segundo aspecto observado tem a ver com o *continuum* proposto por Marcuschi (1997) pois foram observados comentários pejorativos que muitos usuários receberam por conta da escrita inadequada à norma padrão da língua (nos casos em que não se tratava de perguntas acerca da segmentação). Neste ponto foi possível notar que o preconceito linguístico está presente também no ambiente virtual. Será estudada, portanto, a noção de desvio dentro do *continuum* apresentado no capítulo anterior. Ao todo foram recolhidos 34 desvios de segmentação ortográfica dos quais 14 foram caracterizados como hipersegmentações e 20, como hipossegmentações.

2.2 Classificação dos dados

Para facilitar o entendimento acerca dos dados apresentados foi elaborado um fluxograma que permitiu classificar cada construção. Os grupos de ocorrências foram então divididos em desvios por **similaridade ortofônica**, por **similaridade funcional** e por **assimilação sintática**.



Em relação à esta divisão é válido ressaltar os seguintes pontos:

- 1) que a noção de erro aqui encontra-se desprendida de seu caráter pejorativo, e se configurando como *desvio à norma padrão*. Sendo assim, será empregado o termo *desvio* para designar os dados encontrados.
- 2) que uma ocorrência pertencente à um grupo pode também fazer parte de outros dependendo da abordagem da qual se parte. Os casos fronteiros, contudo, serão discutidos posteriormente.

3) as categorias designadas aqui são suficientes para descrever apenas a amostra de dados aqui analisada, contudo, podem não compreender outros casos.

4) ao contrário da maioria dos trabalhos acerca da segmentação a divisão hipersegmentação-hipossegmentação foi feita somente ao final uma vez que não afetaria a interpretação.

5) segundo Miotto, Lopes e Silva (2013:36)

(...) sintaxe é a disciplina que estuda como combinamos palavras para formar sintagmas e como combinamos sintagmas para formar sentenças. Esta concepção de sintaxe se apoia no que se chama a Hipótese Lexicalista, isto é, a sintaxe começa a atuar onde acaba a atuação da morfologia. **A sintaxe toma as palavras, que são o produto da morfologia, e realiza as combinações.** (grifos meus)

sendo assim, todas as classificações bem como a análise dos dados levaram em conta aspectos morfológicos pois, especificamente nos casos de segmentação das palavras, percebe-se uma linha ainda mais tênue entre a sintaxe e a morfologia.

Na tabela abaixo constam todos os desvios encontrados na amostra recolhida bem como o grupo ao qual pertencem - se são hipossegmentações ou hipersegmentações - além da quantidade que representam.

HIPERSEGMENTAÇÃO	HIPOSEGMENTAÇÃO	QUANTIDADE
o deio		1
se quintes		2
se quer		1
	aseguir	3
	denovo	4
	derrepente	3
	encima/enssima	3
	oque	2
	enfrente	2
	doque	1
	domeu	1
de oque		3
de mais		2
de vagar		3
a cima		2

	haver/aver	2
TOTAL		34

É válido ressaltar que em alguns exemplos extraídos foram identificados mais de um desvio e, portanto, contaram-se duas ocorrências nestes casos. Quanto aos casos fronteiriços, será necessária tanto a análise individual quanto a partir de um par pois o contraste entre duas formas elucida a lógica por trás destes desvios. Por outro lado, outras análises partem da construção em si, bem como a relação que estabelece com os demais elementos da frase dado que significam diferentemente apesar de terem sido grafadas da mesma forma⁸.

2.2.1 Desvio por similaridade ortofônica

Estão dentro deste conjunto todos aqueles casos cuja fronteira vocabular (ou sua ausência) modifica o significado da construção ou palavra, ou seja, ambas formas existem na língua portuguesa e suas respectivas realizações na fala são idênticas. São, portanto: *se quer* no sentido de *sequer*, *de vagar* no sentido de *devagar*, *encima* no sentido de *em cima*, *de mais* no sentido de *demais*, *enfrente* no sentido de *em frente*, e *haver* no sentido de *a ver*. Foram encontradas, ainda, variações na grafia de algumas destas construções que, embora não pertençam à língua escrita, estão no mesmo conjunto devido ao fato de dividirem as mesmas motivações e de serem avaliadas pelo viés da segmentação. São essas: *enssima* e *aver*.

Optou-se pela designação ortofônica - ortográfica + fonética - pelo português ser uma língua cuja grafia se aproxima da fala⁹ e porquê, como dito acima, não existem diferenças práticas na realização oral destas estruturas.

2.2.2 Desvio por similaridade funcional

Se encaixam nesta divisão todas as ocorrências cujo equívoco de segmentação encontra respaldo em construções similares quanto ao significado e

⁸ Ver os exemplos 1a e 1b no próximo tópico e 3a e 3b no capítulo 4.

⁹ Pinto (1998) em seu estudo com crianças falantes do português europeu chega à esta conclusão ao comparar a língua com o francês e o inglês.

posição na sentença. Para esclarecer tal classificação, consideremos o caso da grafia *a seguir* e sua similaridade com a forma *seguinte*:

Qual dos exemplos **aseguir** pode ser considerado uma rima rica?

Neste caso, portanto, é possível substituir *aseguir* por *seguintes* cuja grafia não apresenta fronteira vocabular. Ambos adjuntos designam o mesmo papel bem como ocupam a mesma posição na sentença e é possível inferir, assim, que a grafia de um pode influenciar a do outro.

A forma *derrepente/derepente*, por outro lado, exige uma análise mais profunda pois sua semelhança reside não tanto na grafia semelhante - pensando em *seguir* e *seguinte* - mas na posição que os advérbios de modo ocupam. Em 1a, poderia ser substituída por *talvez* e em 1b, por *rapidamente* sendo que ambos não possuem fronteiras vocabular:

1a

Ram 32Gb Armazenamento interno e querendo pegar um S3 com volta amigo, se **derrepente** interessar me avise.

1b

Pq a orelha esquenta **derrepente**?

Outro fator contribuinte para esta hipossegmentação é a problemática diante da preposição *de*, que será discutida adiante. Ademais, *repente* é uma forma usada em contexto medial de palavra, o que dificulta seu reconhecimento enquanto palavra. Pode-se notar, inclusive, que em dois dos três exemplos extraídos tem-se o dígrafo, implicando que o autor sabe fazer a diferenciação entre as pronúncias do *r* no português.

2.2.3 Desvio por assimilação sintática

Pertencem à essa categoria todos aqueles desvios que elucidam alguma motivação sintática por trás de sua realização. Todos os casos deste grupo são compostos por uma partícula mais uma palavra fonológica, na medida que o primeiro geralmente se apresenta como uma preposição ou um artigo e não excede

mais de dois grafemas. É válido ressaltar que o conceito de palavra fonológica foi utilizado a partir das considerações feitas por Cunha e Miranda (2009:134), o qual representa "todas as palavras que possuem um acento primário e que, mesmo não tendo significado conhecido na língua, são candidatas potenciais para tal". Esta definição, apesar de vaga, permite abranger as unidades que se formam a partir dos desvios de segmentação encontrados.

Reitera-se que alguns dos casos encontrados, quando isolados, dividem características com os outros grupos apresentados - por isso é relevante pensar nestas categorias como flexíveis.

Para o propósito deste trabalho o grupo de assimilação sintática representa um *corpus* interessante na medida que, ao longo de sua análise, será possível observar como as regras sintáticas da gramática internalizada também jogam com a ortografia. Observa-se aqui uma mesma tendência apresentada nos textos infantis de juntar as partículas à palavras fonológicas, possivelmente por conta da dificuldade em reconhecê-las enquanto palavras, como afirmam Ferreiro e Teberosky (1985). Contudo, os textos analisados se dão em um momento posterior à alfabetização - apesar de não se levar em consideração a escolaridade dos autores - e se diferenciam pois, em grande parte, os erros são mais sutis do que aqueles cometidos por recém-alfabetizados.

Sendo assim, existem duas outras questões que explicam estas construções: na árvore sintática as partículas estabelecem uma relação muito mais forte com a palavra à qual se juntaram do que com os demais elementos da sentença além de quase sempre serem compostas por palavras átonas que, no português brasileiro, têm a tendência a se juntar à palavra da direita, como afirma Nunes (1993). O autor, ao estudar os clíticos acusativos de terceira pessoa no português, conclui que

com a inexistência de aquisição dos clíticos acusativos de terceira pessoa por uma geração de falantes, construções com esses clíticos tornam-se menos frequentes no ambiente linguístico. Isso, por sua vez, forneceu combustível adicional para que as gerações subsequentes que já tinham adquirido um sistema com cliticização da esquerda para a direita também não adquirissem esses clíticos. (1993 : 220)

Em outras palavras, no português brasileiro moderno os pronomes clíticos acusativos de terceira pessoa se posicionam antes do verbo e "precisam, pelo

menos, de material fonológico que os preceda" (NUNES, 1993). Daí provém a tendência, na escrita, de juntá-los à palavras fonológicas que, na amostra recolhida aqui, se aplica também às demais partículas (*omeu, oque, aminha*).

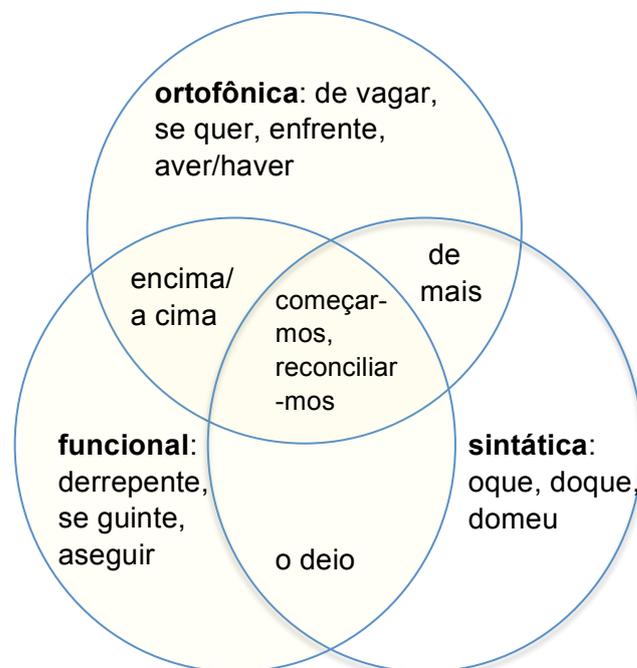
2.3 Conclusões preliminares

Reitera-se que os dados, apesar de não apresentarem grande quantidade, refletem uma variedade interessante de construções que serão analisadas nos próximos capítulos. Assim, são poucos os casos que apresentam o mesmo desvio sendo que alguns, quando o fazem, podem evocar motivações distintas devido à sua posição na estrutura. Com efeito, serão analisados dentro de suas respectivas sentenças uma vez que se encaixam em mais de uma classificação. De fato o número de casos fronteiros representa uma parcela considerável em relação à totalidade das ocorrências, fato que apenas reitera a necessidade de entender as classificações aqui apresentadas como flexíveis.

Capítulo 3

Análise dos dados

Tendo em vista a necessidade de entender as fronteiras classificatórias como flexíveis, muitos dos casos encontrados podem ter mais de uma motivação. Para tanto, foi desenvolvido o seguinte diagrama de Venn:



Pode-se notar, independente da classificação em que se encaixam, que a presença da partícula é problemática e influencia tanto os casos de hipersegmentação quanto os de hipossegmentação. Será considerada como partícula, portanto, toda parte que se separa ou é anexada de forma diferente em relação à grafia oficial do português, sendo que por vezes são encontradas "sozinhas" na língua.

Nesses casos há uma tendência, diante de uma palavra fonológica, à hipossegmentação (é o caso do *derrepente*, *doque* e *acima*, por exemplo) enquanto nas palavras formadas apenas por uma vogal mais palavra fonológica ocasionam

ocorrências de ambos tipos (*aseguir* e *o deio*, por exemplo). Assim, identificou-se uma dificuldade em reconhecer as fronteiras vocabulares quando a palavra é composta por dois fragmentos cujo sentido individual é conhecido fora daquele contexto. Contudo, isto não significa que diante de tais casos o desvio sempre acontecerá pois foram testadas - por meio da ferramenta de busca do *yahoo respostas* - possíveis segmentações que seguiriam esta lógica, como no caso da separação do *de* de *depois* mas os resultados se comportaram de maneira arbitrária, ou seja, parece que a tendência se aplica à algumas formas e à outras, não.

Outra característica recorrente nos dados encontrados foi que, por vezes, os autores cometiam desvios de segmentação diante de certas construções mas não diante de outras (2a) da mesma forma que também foram encontradas ocorrências com mais de um desvio de segmentação, como em 2b, ou seja, foi possível observar a variação na amostra coletada.

2a

Porque o corsa 1.0 se acelerar ele **demais** em uma marcha só ele vai diminuindo **de vagar**?

2b

Oque significa as **se quintes** carinhas?

Nota-se, por exemplo, que o autor de 2a grafou *demais* de acordo com a ortografia oficial e *devagar*, não, enquanto 2b exclui as fronteiras vocabulares entre as formas *o* e *que* constituintes do pronome interrogativo *o que* mas reconhece a partícula *se* de *seguinte* como uma palavra apesar de *quinte* não ser uma palavra conhecida na ortografia da língua.

3.1 Os casos fronteirios

3.1.1 Ortofônico + funcional

Pensando nas fronteiras vocabulares de **acima** e **em cima**, pode-se afirmar que os dois designam significados muito semelhantes bem como, por vezes, ocupam a mesma posição na sentença. Sendo assim, o autor poderia produzir *a cima* pensando na fronteira vocabular de *em cima* ou *encima* com base em *a cima*¹⁰. Esta ocorrência, portanto, ganha bastante força uma vez que as duas formas - *a cima* e *encima* - existem na língua padrão com outros significados. Observa-se aqui, portanto, um caso fronteiro uma vez que pode pertencer à duas classificações distintas: motivação ortográfica diante dos pares *a cima/acima* e *encima/em cima* e similaridade funcional quando tem-se o par *encima/a cima*.

3.1.2 Ortofônico + sintática

A ocorrência do *de vagar* se encaixa dentro do grupo ortofônico pelos mesmos motivos do caso fronteiro acima: encontra, na língua portuguesa, o par *devagar* (advérbio) e *de vagar* (preposição mais verbo) significando de forma diferente. Por outro lado, parece obedecer à alguma motivação sintática pois, analisando o caso abaixo, a preposição estabelece uma relação mais forte com *diminuindo* do que com *vagar* provavelmente porque nesta sentença faz parte de um constituinte diferente. Em outras palavras, o advérbio pode ser substituído por *de maneira lenta* ou *de maneira vagarosa*.

Porque o corsa 1.0 se acelerar ele demais em uma marcha só ele vai diminuindo **de vagar?**

Um outro caminho seria jogar com os elementos da interrogativa a fim de construir a seguinte sentença: *Como o corsa 1.0 vai diminuindo quando se acelera demais em uma marcha só?* cuja resposta seria "De (_____)." ou seja, é possível permutar as palavras que ocupam esta posição mas não a preposição.

¹⁰ O verbo *encimar*, na terceira pessoa do singular do indicativo ou na segunda pessoa do singular no imperativo, tem sentido de "colocar em cima" ou "coroar".

3.1.3 Os casos que envolvem as três motivações

3.1.3.1 A crase em o *deio*

O caso do *o deio* se apresenta de maneira interessante na medida em que permite levantar algumas hipóteses sobre a motivação por trás da segmentação. Primeiramente, de acordo com alguns dados da fala, o pronome *eu* que é realizado como 'ew tem o fim de sua pronúncia no início da do *o* de *odeio*, ou seja, compõem um fone só, justificando assim a crase com a primeira letra do verbo¹¹. Desta forma, sua separação da palavra fonológica elucida o reconhecimento do sujeito da frase por parte do autor. Em outras palavras, *o* estaria representando o pronome *eu*. Kato e Duarte (2014) estudam - a partir da comparação entre o Português europeu e o brasileiro - a possibilidade de *o* brasileiro estar se tornando ou uma língua de sujeito não-nulo ou para uma língua "parcialmente pro-drop". Segundo a análise das autoras, quanto mais referencial for o pronome maior é a probabilidade de ser expresso na fala:

Para uma língua que tem uma opção interna para variantes nulas ou não-nulas, um fator forte para a seleção de uma forma ou de outra ;e o estatuto referencial do referente. Segundo essa hierarquia, argumentos [+N, +humano] estão no extremo mais alto da hierarquia referencial, enquanto não-argumentos estão na posição mais baixa. Com relação aos pronomes, o falante (*eu*) e o interlocutor (*você*), sendo inerentemente argumentos humanos, primeira e segunda pessoas pronominais, estão no ponto mais alto da hierarquia [...] (2014 : 5)

Nunes (1993:211) por outro lado, já evidenciava que nos casos dos pronomes clíticos compostos por vogais que se apoiam fonologicamente em outras palavras "ocorre uma reestruturação de modo que a questão do licenciamento do *onset* torna-se irrelevante porque a sílaba do clítico funde-se à sílaba precedente." Ou seja, na fala, o pronome se funde à vogal que o antecede o que parece ser aplicável ao caso abaixo.

Sendo assim, é possível presumir que apesar do pronome não aparecer em sua forma padrão está realizado por meio da fusão com a primeira vogal do verbo e sua presença é notável a partir da fronteira vocabular inserida pelo autor. A

¹¹ A pronúncia de "eu odeio" seria, portanto, algo como ewo.d'ej.jʊ no qual os fones e, w e o seriam pronunciados de uma só vez.

separação, portanto, permite classificar o desvio como sintático - dado que a consciência sintática do autor exerceu papel fundamental na segmentação - e como fonético (ortofônico) - por conta da crase das vogais que não interfere na realização oral da palavra.

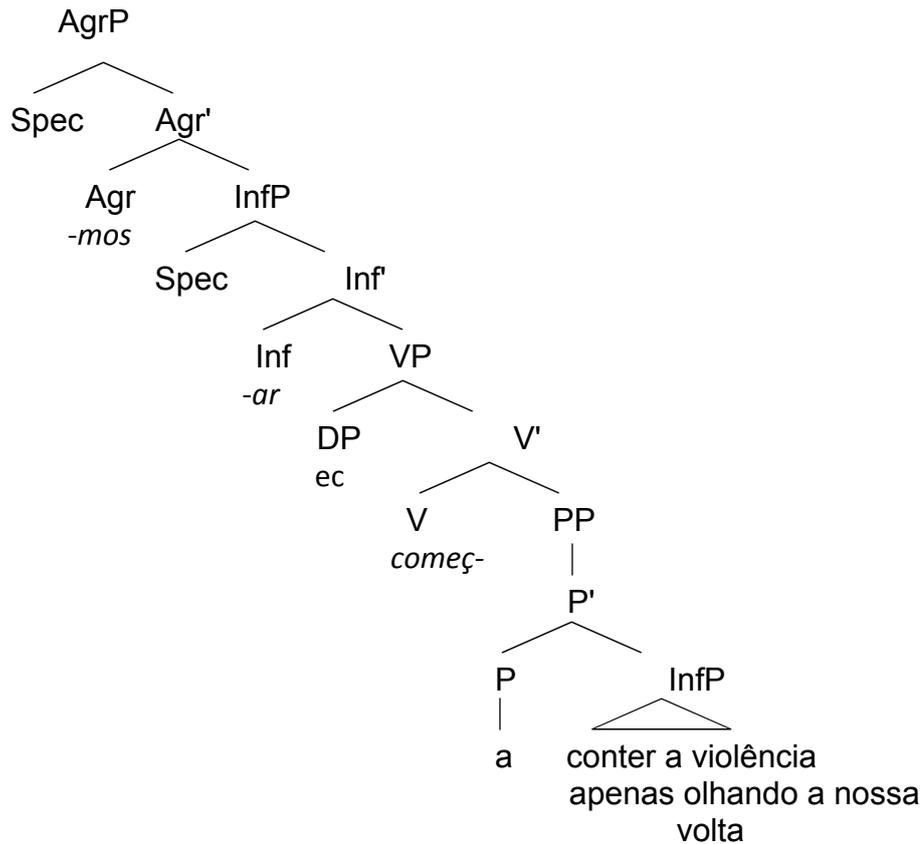
O deio o yahoo ! pq eu recebi o um email de infração?

À luz de uma segunda perspectiva, nota-se também uma motivação funcional pois estruturas com próclise (pronome relativo se posiciona à esquerda do verbo) são recorrentes no português.

3.1.3.2 Segmentação da desinência

Como apontado no capítulo 3, "a sintaxe começa onde acaba a atuação da morfologia" (MIOTO, 2013 : 36) e os casos fronteiros cuja motivação provém dos três tipos reforçam esta afirmação.

A princípio, classificar as formas abaixo como motivadas por características ortofônicas, apesar de não encontrarem palavras com grafia idêntica no português, parece plausível visto que o hífen não interfere na realização fonética da palavra - ou seja, tanto o desvio quanto a forma adequada à norma padrão são lidas do mesmo modo. Contudo, em 3a, pode-se inferir uma consciência morfossintática por parte do autor ao separar a partícula *-mos* do resto da palavra, deixando a forma infinitiva na esquerda. Não por acaso, na estrutura arbórea da sentença, a designação *-mos* do verbo *começarmos* (no futuro do subjuntivo) é representada da seguinte forma:



É válido ressaltar que, devido ao objetivo do trabalho ser estudar o infinitivo flexionado, a estrutura de InfP foi representada de maneira simplificada.

3a

Que tal **começar-mos** a conter a violência apenas olhando a nossa volta?

ou seja, separá-la da forma infinitiva faz com que sua leitura se aproxime da estrutura arbórea da sentença na qual tanto a desinência quanto o morfema de estruturas infinitivas e a vogal temática encontram-se separado da raiz. Mais uma vez, observa-se a linha tênue entre a morfologia e a sintaxe na análise dos desvios de segmentação ortográfica. A vogal temática e o morfema *-r*, por sua vez, estão junto à raiz no desvio porque formam o infinitivo *começar* e designam o tempo verbal da expressão, enquanto *-mos* representa apenas a desinência número-pessoal.

Ademais, os verbos formam núcleos lexicais e são capazes de fazer a seleção semântica de seus argumentos enquanto a designação de número e pessoa, além de selecionar, também escolhe seus argumentos de acordo com a categoria à qual pertencem. Isso significa que a separação dos dois mostra a capacidade do autor - mesmo que inconsciente - de reconhecer que a partícula *-mos*

irá evocar diferentes elementos do que a forma infinitiva e, seguindo tal lógica, deve ser separada da raiz verbal. No segundo exemplo o verbo separado também funciona como futuro do subjuntivo:

3b

Vivi separado com a minha mulher depois de **reconciliar-mos** a minha família
começou de novo com as brigas?

É válido notar que o autor grafou *minha* que, por um lado, pode ser um erro de digitação e, por outro, parece também obedecer ao princípio das palavras átonas que tendem a se comportar como sílabas fracas das palavras as quais se juntam. A partícula *-mos*, por sua vez, também funciona como argumento do verbo *reconciliar* - mas por um motivo distinto de 3a pois representa tanto o argumento externo quanto o interno do verbo - o qual deveria ter sido representado por uma próclise sem deixar de apresentar a desinência pessoal no final - algo como *depois de nos reconciliarmos*. Sendo assim, o autor fundiu o pronome pessoal com a desinência tempo-verbal e os representou separado pelo hífen. Mais uma vez é possível notar a consciência sintática jogando com os elementos da sentença dado que, na forma adequada, o argumento (pronome) encontra-se de fato separado do verbo.

Por último, pode-se afirmar que ambos possuem motivação por similaridade funcional na medida que a ênclise pronominal existe no português e sua estrutura é muito similar à esta empregada no exemplo 3¹².

¹² Pensando nas formas *levar-nos* ou *amar-nos*, por exemplo

3.2 Casos interessantes

3.2.1 O caso do *aseguir*

Apesar da amostra analisada não apresentar uma grande quantidade de ocorrências foi possível identificar uma diversidade no uso da expressão *aseguir*. Anteriormente foram apresentadas as classificações fronteiriças nas quais o caso abaixo poderia ser encaixado, contudo, as demais amostras desta grafia se comportam unicamente como funcionais. Comparando as duas estruturas a seguir, temos:

2a

Meu processo encontra-se no stj após o acórdão o que vem **aseguir**?

2b

Minha familia ta me obrigando **aseguir** o Exército?

Ou seja, embora ambas sentenças apresentem o mesmo desvio de hipossegmentação, eles não dividem traços semânticos ou sintáticos quando comparados: em 2a seria possível substituir *aseguir* por *depois* ou *seguinte* ao passo que em 2b, não. Isto se dá porque em 2b a preposição é uma exigência do verbo *obrigar*, embora esteja alocada junta do verbo *seguir*. Seguindo esta lógica se espera que a preposição se junte ao primeiro verbo ou então permaneça isolada, o que não acontece. Mas então por que são grafadas da mesma maneira se não significam de tal forma?

Pensando em um segundo fator que pode contribuir para a ausência de fronteiras vocabulares em *aseguir* - além da dificuldade que os indivíduos têm em reconhecer estas partículas enquanto palavras, como discutido acima - é possível constatar que se trata de uma palavra átona, ou seja, se comporta como uma sílaba fraca da palavra em que está apoiada (*seguir*) e por isso ambas ocorrências evidenciaram a hipossegmentação apesar de ocuparem diferentes posições - consequentemente, diferentes funções - na sentença.

3.2.2 A preposição *de*

No português algumas das preposições formam núcleos funcionais, ou seja, são capazes apenas de selecionar a categoria à qual o seu complemento deve pertencer, deixando o caráter semântico do complemento ser selecionado pelos núcleos lexicais. Apesar de alguns casos aqui evidenciados não derem compostos

Contudo, foram encontrados também casos cuja segmentação separou os componentes morfológicos da preposição para poder juntar o artigo à palavra lexical. O pertinente aqui é observar como essa estrutura joga com a representação arbórea da frase, em outras palavras, mostra a capacidade sintática do indivíduo de combinar elementos de acordo com sua proximidade, mesmo que inconscientemente:

Como faço para afastar o xixi do cachorro **de omeu** portão?

sabe-se que neste caso o artigo deveria ter se juntado à preposição, formando o *do* contudo sua a relação com o pronome possessivo parece ser mais forte. Ademais, a preposição *de* é exigida pelo verbo *afastar* ao passo que o artigo é exigência de *portão* e carrega traços semânticos dele (masculino).

Uma segunda questão pertinente é que o português permite tanto as formas com contração do artigo quanto sem, apesar de representarem significados distintos. Ou seja, a fala opta pela contração, portanto, tanto "como faço para afastar o xixi do cachorro de o meu portão" quanto "como faço para afastar o xixi do cachorro do meu portão" são aceitáveis¹³.

¹³ Ver o conceito de aceitabilidade no tópico 1.1

Considerações finais

O suporte teórico respaldado na aquisição da escrita infantil, quando aplicado aos fenômenos da língua em uso, permitiu estabelecer um contraste dos desafios encontrados pelos alfabetizandos que muitas vezes se propagam pelo seu caminho na escrita. Apesar de os desvios cometidos pelos alunos iniciantes parecerem mais evidentes do que aqueles encontrados no *yahoo respostas*, foi possível identificar que por vezes partem das mesmas motivações - como, por exemplo, a dificuldade de reconhecer as partículas com uma só letra como palavras. Neste sentido, pode-se considerar que o estudo destes desvios pelo viés sintático também contribui para entendê-los no meio da aquisição já que, desde muito cedo, a criança é capaz de combinar estruturas de sua L1, jogando com as possibilidades sintáticas da língua. Foi evidenciada, contudo, a forte influência da fonética e morfológica nos desvios de segmentação dado que, em sua maioria, podem ser explicados pelo viés sintático bem como pelo morfológico e fonético. Esta influência demonstrou-se mais forte do que esperado inicialmente comprovando a barreira tênue entre estes campos de estudo da língua.

Um segundo aspecto constatado foi em relação à terminologia utilizada para caracterizar tais desvios. Como discutido no capítulo 2, os desvios de segmentação aqui estudados foram classificados primeiramente de acordo com suas possíveis motivações para então se dividirem entre hipersegmentações e hipossegmentações. Isto se deu porque esta divisão, para fins metodológicos, não se fez útil na medida em que separar ou juntar uma palavra em desacordo com a ortografia oficial revela pouco a respeito da lógica por trás daquele desvio. Ademais, como foi apresentado, muitos desvios de hipersegmentação foram classificados juntamente aos de hipossegmentação já que tal divisão não afeta a interpretação da ocorrência. Por exemplo, desvios como *de vagar*, *se quer*, *enfrente*, *aver/haver* foram colocados dentro do grupo ortofônico apesar de variarem quanto à inserção e ausência de fronteiras vocabulares.

Pensando na ortografia oficial do português é válido levar em conta que suas recentes mudanças - o Novo Acordo Ortográfico estabelecido em 2009 cujo principal objetivo é unificar a escrita do português - alteraram algumas regras de

segmentação da língua escrita. Neste sentido, as regras de segmentação - que antes já confundiam os autores na hora de elaborar um texto - foram reestabelecidas tendo em vista obedecer ao acordo.

Tal alteração coloca em jogo a seguinte questão: se, *a priori*, a escrita é uma forma de transcrição da fala, como é possível padronizá-la em tantos países cuja língua varia em aspectos fonéticos, lexicais e semânticos? Como exemplo tem-se a grafia de *antinflamatório* que agora é *anti-inflamatório*: na fala não se pronuncia os dois i's tampouco uma vogal alongada, apenas um som - mais próximo da primeira forma apresentada. Sendo assim, em certos aspectos, o Novo Acordo reforça a arbitrariedade da convenção escrita dado que se trata de regras, ou seja, não há qualquer relação lógica que permita o indivíduo **entender** como funciona a nova ortografia. É válido lembrar, contudo, que muitas das palavras que sofreram alterações em suas fronteiras vocabulares após o Novo Acordo são de caráter erudito, ou seja, surgem em meios incomuns (vocábulo técnico). Assim, a apropriação e propagação das regras se dá de maneira mais eficiente visto que são menos utilizadas na linguagem coloquial.

Por último, foi possível identificar na amostra recolhida muitas ocorrências com mais de um desvio, fato que aponta para a hipótese de que o desconhecimento da ortografia padrão - no que concerne a segmentação - possui um importante papel no aparecimento destes desvios. Portanto pode-se concluir que quando o autor comete um desvio de segmentação há maior chance que faça o mesmo com outras palavras na composição.

Referências Bibliográficas

BARTHES, R. **Elementos de Semiologia**. 16. e.d. São Paulo: Cultrix, 2006

BASILIO, M. **Teoria Lexical**. São Paulo: Ática, 1987

BASILIO, M. "Em torno da palavra como unidade lexical: Palavras e composições." **Veredas, revista de estudos linguísticos Juiz de Fora**, v. 4, n. 2 p. 9 a 18

BASILIO, M. "O Conceito de Vocábulo na Obra de Mattoso Câmara". **D.E.L.T.A.**, 20: Especial, 2004 (71-84)

CUNHA, A.P.N; MIRANDA, A.R.M. "A hipo e a hipersegmentação nos dados da aquisição de escrita: a influência da prosódia". **Alfa**, São Paulo, 53 (1): 127-148, 2009

FERREIRO, E.; TEBEROSKY, A. **Psicogênese da língua escrita**. Tradução de Diana Myriam Lichtenstein, Liana di Marco e Nestor Jerusalinsky Porto Alegre: ARTMED, 1999.

FLOWER. L; HAYES. J. R. **A Cognitive Process Theory of Writing**. College Composition and Communication, Vol. 32, no. 4 (Dec., 1981), pp.365-387

GUIMARÃES, S.R.K. "Relações entre Capacidade de Segmentação Lexical. Consciência Morfossintática e Desempenho em Leitura e Escrita". **Psicologia: Teoria e Pesquisa**. Jan - Mar 2011, Vol.27 n.1, pp. 23-32

KATO, M. A. **No mundo da escrita: uma perspectiva psicolinguística**. São Paulo: Ática, 1986.

KATO, M. A; DUARTE, M. E. L. "Restrições na distribuição de sujeitos nulos no Português Brasileiro". **Veredas: Sintaxe das Línguas Brasileiras**, Vol.18/1, 2014

KOCH, INGEDORE VILLAÇA (2009), **Ler e escrever: estratégias de produção textual** / Ingedore Villaça Koch, Vanda Maria Elias. - São Paulo : Contexto, 2009

MARCUSCHI, L.A. "Oralidade e escrita". **Signótica**, 9:119-145, jan/dez. 1997

MIOTO, C; SILVA, M. C. F; LOPES, R. **Novo Manual de Sintaxe**. São Paulo: Contexto, 2013

NUNES, J. "Direção de cliticização, objeto nulo e pronome tônico na posição de objeto em português brasileiro". **Português Brasileiro: uma viagem diacrônica**. 2.e.d. São Paulo: Unicamp, 1996

OLIVEIRA, Ana Maria R. (2013), "As neurociências ao serviço da linguagem", p. 39-57 - último acesso em: 04/06/2015 <http://ler.letras.up.pt/uploads/ficheiros/12006.pdf>

PINTO, M. G. L. **Saber viver a Linguagem: Um desafio aos Problemas de Literacia**. Coleção Linguística. Porto: Porto Editora, 1999.

VEÇOSSI, Cristiano Egger; FERREIRA-GONÇALVES, Giovana. "A hipo e a hipersegmentação em dados de escrita de alunos da 8a série: influência exclusiva dos constituintes prosódicos?" **Verba Volant**, v . 2, no 1. Pelotas: Editora e Gráfica Universitária da UFPel, 2011.

VELOSO, J. 2003. "Da influência do conhecimento ortográfico sobre o conhecimento fonológico. Estudo longitudinal de um grupo de crianças falantes nativas do Português Europeu". **Dissertação de doutoramento**, Universidade do Porto.