

**SABE? E ENTENDEU? COMO MARCADORES DE BUSCA DE APROVAÇÃO DISCURSIVA: VARIANTES DE UMA MESMA VARIÁVEL?****SABE? AND ENTENDEU? AS DISCURSIVE APPROVAL SEARCH MARKERS: VARIANTS OF THE SAME VARIABLE?**

Isa Caroline Aguiar Zanin<sup>1</sup>  
Sebastião Carlos Leite Gonçalves<sup>2</sup>

**RESUMO:** Neste artigo, tratamos dos marcadores discursivos *sabe?* e *entendeu?* na função interacional de busca de aprovação discursiva. Assumindo que esses marcadores resultam de processo de gramaticalização de verbos plenos, nosso objetivo é mostrar como poderia ser dispensado a eles um tratamento variável, considerando que, uma vez gramaticalizados, eles passam a competir no mesmo domínio funcional de busca de aprovação discursiva. O quadro teórico que sustenta nossa proposta é o do *Sociofuncionalismo*, que procura conjugar aspectos da Teoria da Variação e Mudança Linguística (WEINREICH; LABOV; HERZOG, 1968) e da Teoria da Gramaticalização (HOPPER; TRAUGOTT, 2003; BYBEE, 2016, dentre outros), associado aos pressupostos da Gramática Textual-interativa (JUBRAN; KOCH, 2006). Em termos metodológicos, os dados de nossas análises provêm de entrevistas sociolinguísticas do Banco de dados Iboruna do Projeto “Amostra Linguística do Interior Paulista”, que registra a fala da região noroeste do Estado de São Paulo. Na proposição do envelope de variação, consideramos os dois marcadores como variantes da variável dependente *busca de aprovação discursiva*, e como independentes, as variáveis linguísticas *posição do marcador no enunciado e tipo textual*, e as variáveis sociais *sexo/gênero, faixa etária e escolaridade*. Os resultados nos permitem mostrar que a escolha das variantes mantem correlações estreitas com as variáveis *tipo textual e sexo/gênero*. Diante do nosso propósito e mesmo sem levar às últimas consequências a análise de regra variável, comprovamos, ao final, que a comparabilidade funcional dos marcadores discursivos permite tratá-los como variantes de uma mesma variável. Daí em diante, o pesquisador pode estreitar o foco de suas análises, lançando ao fenômeno dos marcadores discursivos um olhar ou mais variacionista ou mais funcionalista.

**PALAVRAS-CHAVE:** Sociofuncionalismo. Gramática Textual-interativa. Variação. Marcadores Discursivos.

**ABSTRACT:** In this paper, we deal with the discourse markers *sabe?* and *entendeu?* in the interactional function of seeking discourse approval. Assuming that these discourse markers result from the grammaticalization process of full verbs, our aim is to exemplify how a variable treatment could be given to these markers. The reasoning behind this objective is that when recategorizing *sabe?* and *entendeu?* as discourse markers, they compete in coding the same discourse approval-seeking function. The theoretical framework that supports our proposal is *Sociofunctionalism*, which seeks to combine aspects of the Language Variation and Change Theory (WEINREICH; LABOV; HERZOG, 1968) and the Grammaticalization Theory (HOPPER; TRAUGOTT, 2003; BYBEE, 2016, among others), associated with the assumptions of Textual-Interactive Grammar (JUBRAN; KOCH, 2006). In methodological terms, we analyzed data collected from sociolinguistic interviews from the Iboruna database of the Project “Amostra Linguística do Interior Paulista”, which records speech from the northwest region of the State of São Paulo. In the proposition of the variation envelope, the two markers are considered as variants of the dependent variable *seeking discourse approval*, and as independent, the linguistic variables *marker position in the utterance*, and *type of text*, and social variables *sex/gender, age and education*. The results allow us to show that the choice of variants maintains close correlations with the *textual type* and *sex/gender* variables. Considering our objective and even without taking the variable rule analysis to its ultimate consequences, in the end, we proved that the functional comparability of the discursive markers allows us to treat them as variants of the same variable. From then on, the researcher can narrow the focus of his analyses, giving the phenomenon of discourse markers a focus that is either more variationist or more functionalist.

<sup>1</sup> Doutoranda e Mestre em Estudos Linguísticos do Programa de Pós-graduação em Estudos Linguísticos da Universidade Estadual Paulista – Unesp, Câmpus de São José do Rio Preto. E-mail: [isa.zanin@unesp.br](mailto:isa.zanin@unesp.br). ORCID: <https://orcid.org/0000-0002-5785-7633>.

<sup>2</sup> Professor Assistente Doutor do Departamento de Estudos Linguísticos e Literários da Universidade Estadual Paulista – Unesp, Câmpus de São José do Rio Preto. Bolsista produtividade em pesquisa - CNPq. E-mail: [sebastiao.goncalves@unesp.br](mailto:sebastiao.goncalves@unesp.br). ORCID: <https://orcid.org/0000-0002-1798-729X>.

**KEYWORDS:** Sociofuncionalism. Textual-interactive Grammar. Variation. Discourse markers.

## 1 Introdução

Dentre os vários mecanismos envolvidos na organização textual-interativa dos textos de língua falada, destaca-se o grupo dos *marcadores discursivos* (MD, daqui em diante). Trata-se de um grupo amplo e diversificado de expressões linguísticas, envolvendo, no plano verbal, desde sons não lexicalizados a sintagmas mais desenvolvidos, e ao qual se pode atribuir a condição de uma categoria pragmática bem consolidada no funcionamento da língua. Por seu intermédio, a instância da enunciação marca forte presença no enunciado, ao mesmo tempo em que se manifestam importantes aspectos que definem sua relação com a construção textual-interativa (RISSO; OLIVEIRA E SILVA; URBANO, 2006).

Para o desenvolvimento deste trabalho, selecionamos os MD *sabe?* e *entendeu?* e procuramos investigar suas características nos contextos em que são empregados, sob a hipótese de que pode ser dispensado a eles um tratamento variável, em vista de se comportarem como formas variantes na codificação da mesma **função de busca de aprovação discursiva**. Considerada essa hipótese, o objetivo deste artigo é o de mostrar a viabilidade de se adotar uma abordagem variacionista, levantando fatores condicionantes (linguísticos e extralinguísticos), em interface com uma perspectiva funcionalista, de forma a ser possível caracterizar os contextos preferenciais de usos de MD que são funcionalmente equivalentes.

Para levar a cabo esse objetivo, adotamos a perspectiva do chamado *Sociofuncionalismo* (NARO; BRAGA, 2000; GÖRSKI; TAVARES, 2013; GONÇALVES, 2021), que alia pressupostos da Teoria da Variação e Mudança (WEINREICH; LABOV; HERZOG, 1968) e da Gramaticalização (HOPPER; TRAUOGOTT, 2003; BYBEE, 2016), aos quais acrescentamos também os pressupostos da Gramática Textual-interativa (JUBRAN; KOCH, 2006), uma teoria também baseada no uso da língua e que dispensa detalhada atenção aos MD. Assim, nosso trabalho procura oferecer uma contribuição a pesquisadores dessas áreas aqui colocadas em interface, na medida em que, por meio de estudo de casos dos MD *sabe?* e *entendeu?*, alcança generalizações que atendem interesses de estudiosos do texto, de funcionalistas e de variacionistas. Importa destacar, nesse passo, que não é nosso objetivo levar o tratamento variacionista às últimas consequências, com análises multivariadas, mas apenas mostrar, com tendências de uso, como sociolinguistas podem também incorporar em suas investigações fenômenos discursivos afeitos à variação, e como funcionalistas e estudiosos do texto podem adotar métodos sociolinguísticos em suas pesquisas.

Segundo a Gramática Textual-interativa, os MD, de um modo geral, podem desempenhar funções variadas, direcionadas para o gerenciamento tanto do fluxo textual quanto da interação (RISSO, 2006a, 2006b), com predomínio de um ou de outro tipo. Diante da diversidade de funções que resultam da combinação desses dois modos de gerenciamento do texto falado, focalizamos, neste trabalho, apenas a função de *busca de aprovação discursiva* (BAD, daqui em diante) ou de *checking* que *sabe?* e *entendeu?* podem exercer nos mesmos contextos, para assim, considerá-los como variantes de uma mesma variável, em que o predomínio do gerenciamento da interação prevalece sobre o do arranjo textual. Em perspectiva funcionalista, consideramos aqui que as funções de MD resultam de processos de mudança que atingem itens/construções originalmente de classes gramaticais variadas, como é o caso dos MD aqui enfocados, que se originam da classe de verbos, desenvolvem funções discursivas, via processo de gramaticalização, e passam a coexistir com outros MD de função equivalente.

Em levantamento prévio de MD em nosso cópulo de análise, identificamos também a mesma função BAD desempenhada pelos MD *né?* e *tá?*, o primeiro de frequência

extremamente alta e, por isso, considerado exemplar prototípico (BYBEE, 2016), e o segundo, contrariamente, de uso mais raro, e considerado exemplar mais marginal da função. Diante dessa constatação prévia, descartamos de nossas análises o MD *tá?* e adotamos *né?* apenas como modelo de comparação. Assim, para o tratamento variável, passamos a lidar com variável dependente binária, composta pelas variantes *sabe?* e *entendeu?*. O objetivo é então definir contextos de uso em que, a exemplo do MD *né?*, essas duas variantes podem ser intercambiáveis na mesma função BAD.

Feito esse preâmbulo, esclarecemos que o tema dessa proposta se justifica em razão de estudos de MD em abordagem sociofuncionalista, explorando as funções textual-interativas, serem ainda raros. Para argumentar em favor da viabilidade de nossa proposta, além desta introdução, o artigo segue estruturado em três seções principais: na seção 2, apresentamos os principais pressupostos da Gramática Textual-interativa, para o tratamento dos MD (2.1.), e da abordagem Sociofuncionalista, para a interface entre Sociolinguística e Gramaticalização, que justifica considerar os MD em análise como variantes de mesma variável (2.2.); na seção 3, descrevemos os procedimentos metodológicos adotados, compatíveis com a proposta sociofuncionalista, para, então, na seção 4, apresentarmos nosso modelo de análise. Em nossas conclusões, na última seção, apresentamos as generalizações alcançadas a partir da presente investigação, seguidas das referências bibliográficas.

## 2 Subsídios Teóricos

Os subsídios teóricos que embasam as discussões apresentadas neste artigo tomam por referência a Gramática Textual-interativa (GTI, daqui em diante), especificamente no que tange ao tratamento por ela dispensado aos MD, e o *Sociofuncionalismo*, uma via de análise que conjuga pressupostos da Sociolinguística Variacionista e do Funcionalismo, encarecendo, deste último, processos de mudança linguística identificados com a Gramaticalização.

### 2.1 O tratamento dos Marcadores Discursivos na Gramática Textual-interativa

A GTI assume a concepção de língua como interação social, uma forma de ação verbal, por meio da qual os interlocutores realizam tarefas comunicativas de troca de representações, executam metas, manipulam interesses, sempre no contexto de um espaço discursivo orientado por representações mútuas de seus papéis sociais e discursivos. Sob essa concepção de língua, um primeiro princípio teórico-metodológico da GTI é o de que os fatos nela considerados têm suas propriedades e funções definidas nas situações concretas de interlocução, coenvolvendo as circunstâncias enunciativas. Essa concepção de linguagem, que leva à eleição do produto da interação – o texto – como objeto de estudo, assenta-se em uma base teórica que congrega princípios da Pragmática, da Linguística Textual e da Análise da Conversação (JUBRAN, 2007).

Na organização do texto falado, dentre os vários processos investigados pela GTI (*Referenciação, Repetição, Correção, Parafraseamento, Parentetização, Tematização-Rematização*), destaca-se o papel dos MD, que assumem funções em duas direções: (i) no gerenciamento do texto, com marcadores de função *basicamente sequenciadora e secundariamente interacional*, e (ii) no gerenciamento da interação, com marcadores de função *basicamente interacional e secundariamente sequenciadora* (RISSO et al., 2006). Neste trabalho, assim como em Risso et al. (2006), MD são concebidos, portanto, como uma categoria discursiva, definida pelo contrabalanceamento de um conjunto de traços, dentre eles, a função que exercem. Para além de qualquer outro traço particularizador de MD, defendemos aqui que é a função o principal traço definidor da categoria.

O que deve ficar claro, neste ponto, é que as funções textual-interativas de MD devem ser entendidas em termos de predominância, assim estabelecida: quando a função de um MD é predominantemente textual, implica reconhecer que ela também comporta uma contraparte interacional menos saliente, e, contrariamente, quando a função predominante é interacional, a contraparte textual é a que fica minimizada (GUERRA, 2007). Dados os objetivos do presente artigo, passamos a tratar somente de MD que se enquadram nesse segundo tipo.

Urbano (2006), ao tratar de MD basicamente interacionais, com abstração à eventual função concomitante de sequenciadores tópicos, busca apurar funções, propriedades e comportamentos textual-interativos específicos dos seguintes grupos de MD: (i) *ah, ahn, ahn ahn, hem?, uhn uhn, uhn?*; (ii) *certo, claro, exato*; (iii) *é, é claro, é verdade*; (iv) *entende/entendeu?, sabe?, tá?, viu?*; (v) *não é verdade?, não é?, num é?, né?*; (vi) *olha/olhe, vamos ver, veja, vem cá*; (vii) *pois é, sei, sim*. Todos esses MD se particularizam por serem exteriores ao conteúdo proposicional, sintaticamente independentes e comunicativamente não-autônomos. Além desses traços, apresentam os seguintes aspectos, particularmente fortes e estáveis, registrados em suas matrizes básicas de traços: (i) são formas vazias de conteúdo (como é inclusive o caso das formas não-lexicalizadas *ah, ahn, heim, uhn*) ou apresentam perda total ou parcial de carga semântica; (ii) são prosodicamente demarcados; (iii) apresentam massa fônica reduzida, contendo até três sílabas tônicas. Em seu estudo, o autor analisa as formas *entende?/entendeu?, sabe?, tá?* e *viu?* em conjunto, dadas as propriedades que partilham, quais sejam: fonte gramatical verbal e função única de fáticos de natureza interrogativa. Em (1) e (2), seguem ocorrências exemplificativas de *entende?* e *sabe?*, oferecidas pelo autor, com essas características, e de interesse particular ao nosso estudo.

(1) Inf. - é engraçado que você saindo do Brasil ... a gente sente uma falta muito grande dessa parte de verduras ... eu falo muito em verduras porque justamente é a base da minha alimentação ... **entende?** então a salada pro ... pro pessoal de Buenos Aires ... a salada se resume a alface e tomate... aqui não ... você pede uma salada vêm outros legumes ... né? [DID RJ 328: 227-34] (URBANO, 1994, p. 513)

(2) Inf. - agora ... realmente ... eh ... a alimentação de outros estados é bem diferente daqui do Rio ... **sabe?** [DID RJ 328: 111-17] (URBANO, 1994, p. 513)

Outra característica dos MD em (1) e (2) é não estarem sujeitos à flexão gramatical, típica da categoria verbal de que se originam. Na função de MD, as formas verbais se fixam na terceira pessoa, no modo indicativo e no tempo de presente ou de pretérito perfeito. Características semelhantes às de *sabe?* e de *entendeu?* são apontadas também para o MD *BAD* prototípico *né?*, considerado variante reduzida de *não é?*, que, por sua vez, reduziu-se de *não é verdade?* (URBANO, 1994), sendo possível, por esse processo de redução, reconhecer a cristalização na forma de terceira pessoa do presente do indicativo. De acordo com Guerra (2007), *né?* apresenta as seguintes propriedades dos MD em geral: (i) independência sintática; (ii) autonomia prosódica, por estar envolto em curva entonacional ascendente distinta da do enunciado em que ocorre; (iii) esvaziamento morfossintático, semântico (total ou parcial) e, por vezes, ilocucional; e, (iv) função tipicamente interacional. É um dos MD mais produtivos no PB para a função basicamente orientadora da interação.

Enquanto Urbano (2006) denomina de *Busca de Aprovação Discursiva (BAD)* a função desempenhada por esse grupo de MD, Guerra (2007), alternativamente, a denomina de *Checking*. No corpus analisado por Guerra, também proveniente do Banco de dados Iboruna, são enquadrados nesse tipo os MD *né?, sabe?, viu?, ok? entendeu?* e *cê viu?*, os quais expressam nítida busca de uma aprovação discursiva por parte do falante em direção ao ouvinte. Complementarmente aos MD com função *checking*, Urbano (1995) reconhece aqueles com função de *feedback*, que comportam tipos que expressam nítida orientação por parte do ouvinte

em direção ao falante, através da manifestação de um acompanhamento atencioso da fala do outro, como é o caso de *ah*, *hum*, *huhum*, *ham* e *é* (GUERRA, 2007).

Apesar de particularidades que possam ser identificadas no uso de diferentes tipos de *checking*, todos eles compartilham a propriedade de constituírem perguntas retóricas, que, por seu conteúdo semântico-pragmático esvaziado, não pressupõem uma resposta do interlocutor. Isso se torna evidente pelo fato de o uso prototípico de *checking* ocorrer no interior de turno conversacional, em que o falante quer se sentir habilitado a dar prosseguimento a seu discurso, mas sem ceder o turno para a resposta do interlocutor, como mostram as ocorrências em (1) e (2). É certamente por essa razão que os falantes, normalmente, não desenvolvem um segmento de discurso muito longo sem utilizar algum tipo de *checking*. E, similarmente, o interlocutor, por sua vez, não passa um período muito longo sem emitir algum tipo de *feedback*, razão que leva Guerra (2007) a considerar que os MD com essas funções são responsáveis pela coesão interacional do/no discurso.

Como se pode observar, *sabe?* e *entendeu?*, foco de nossa análise, desempenham funções iguais ou semelhantes entre si, sendo, designados como MD de *BAD*, à semelhança do MD *né?*, aqui considerado o mais prototípico pela sua alta frequência em interações espontâneas ou semi-espontâneas, como veremos mais adiante na análise de resultados.

## 2.2 Sociolinguística Variacionista e o Funcionalismo em diálogo

Segundo Görski e Tavares (2013), *Sociofuncionalismo* é um rótulo que pode recobrir diferentes tipos de enfoque que envolvam pressupostos da Sociolinguística variacionista (WEINREICH; LABOV; HERZOG, 1968; LABOV, 1972) e do Funcionalismo Linguístico da Costa Oeste dos Estados Unidos (HOPPER, 1991; HOPPER; TRAUGOTT, 2003; GIVÓN, 1995, dentre outros), mais recentemente denominado *Modelos Baseados no Uso*, por conjugar aspectos funcionais e cognitivos na descrição e análise linguísticas (BYBEE, 2016).

Naro e Braga (2000) e Görski e Tavares (2013), ao darem relevo aos pontos de convergência dos pressupostos teórico-metodológicos do Funcionalismo Linguístico e da Sociolinguística Variacionista, trazem sustentação à aplicação da abordagem sociofuncionalista ao estudo de fenômenos variáveis instanciados por meio de processos de gramaticalização, construto teórico de forte impacto na abordagem funcionalista da mudança linguística. O primeiro desses pressupostos é o da variabilidade inerente ao sistema das línguas (LABOV, 1972).

A Sociolinguística variacionista assume que a língua apresenta variabilidade de usos em todos seus níveis de organização, na medida em que os falantes fazem escolhas entre duas ou mais formas alternantes de expressão linguística. Estudos variacionistas da década de 1970 fornecem fortes evidências de variação em larga escala e demonstram que sua manifestação na comunidade de fala é sistemática, regular, e segue padrões, não sendo, portanto, fruto de escolhas livres e aleatórias (LABOV, 1972). Assim, torna-se imperiosa a dissociação entre estrutura linguística e homogeneidade, porque a língua é, por natureza, heterogênea e inerentemente variável, sincrônica e diacronicamente, e a variação é passível de descrição sistemática, em função de restrições linguísticas e extralinguísticas. Além de ser uma característica essencial da língua, a variação é também um pré-requisito para a mudança linguística (WEINREICH; LABOV; HERZOG, 1968).

Como apontam Görski e Tavares (2013), a variabilidade inerente é o principal ponto de contato entre a teorias funcionalistas e a Sociolinguística Variacionista, porque assim como esta, aquelas também reconhecem que a variação está presente nos níveis mais profundos de representação gramatical, sendo, portanto, inerente à língua, uma vez que a gramática não é fixa e absoluta, mas variável e probabilística em sua essência (BYBEE; HOPPER, 2001).

Para Görski e Tavares (2013), também é digno de nota o papel central atribuído à mudança linguística pela Sociolinguística e pelo Funcionalismo, que compreendem que a mudança é disseminada gradualmente ao longo dos espectros linguístico e social, com incrementações contínuas em termos de frequência de uso, que, na perspectiva funcionalista, é fundamental para o estabelecimento e a manutenção da gramática, além de permitir também captar a difusão linguística e social da mudança. Na perspectiva variacionista, o aumento de frequência também é compreendido como índice de difusão linguística e social, e as variantes devem ter certa recorrência para que possam ser comparadas por meio de instrumental estatístico. É justamente na questão da quantificação estatística que Bybee (2016) identifica outro importante ponto de contato entre as abordagens sociolinguística e funcionalista:

Em uma teoria baseada no uso, os estudos quantitativos passam a ser extremamente importantes para a compreensão da amplitude da experiência com a língua. A tradição variacionista iniciada por Labov (1966, 1972), embora destinada à compreensão de como ocorre a interação de fatores sociais com a fonologia e a gramática, também fornece uma metodologia apropriada para o estudo da variação e da mudança gramatical. (BYBEE, 2016, p. 33)

É justamente a relação entre o processo de gramaticalização e o fenômeno de variação morfossintática que é apontada, por diversos pesquisadores, como um dos pilares que sustenta o duplo olhar funcionalista e sociolinguístico sobre os fenômenos de variação e mudança linguística. Anteriormente a Görski e Tavares (2013), também Naro e Braga (2000) já reconheciam que a gramaticalização, como processo de mudança, é capaz de produzir variabilidade na gramática por meio da estratificação de formas em um mesmo domínio funcional, com ao menos uma delas sendo resultado de processo de gramaticalização.

Em que pesem esses pontos de convergência para um trabalho de interface eficaz, o ponto principal de divergência entre as duas teorias fica por conta do modo como cada modelo teórico concebe a variação e a mudança, como enfatizam Naro e Braga (2000) e Görski e Tavares (2013). Enquanto para variacionistas a variação é considerada sempre um estágio anterior da mudança (WEINREICH; LABOV; HERZOG, 1968), para estudiosos da gramaticalização, essa proposição se reverte, uma vez que, ao se gramaticalizar, uma forma passa a coexistir, em um mesmo domínio funcional, com outras que cumprem a mesma função, instaurando um típico caso de variação por estratificação funcional (HOPPER, 1991).

Esse ponto de divergência, no entanto, não chega a comprometer a possibilidade de interface entre as duas teorias, se os métodos da pesquisa sociofuncionalista esclarecerem pontos cruciais das duas abordagens, como sugerem Görski e Tavares (2013): (i) identificação de situações de uso linguístico variável dentro de um domínio funcional; (ii) operacionalização da noção laboviana de variável; (iii) testagem de grupos de fatores diversos para identificar os contextos (linguísticos, discursivos, estilísticos e sociais) de uso das formas; (iv) detalhamento de cada grupo de fatores buscando captar variações e mudanças em curso ainda sutis (considerando inclusive sobreposição de funções), e posterior amalgamação de fatores em busca de generalizações; (v) interpretação da frequência das formas em determinados contextos como indício de: (a) perda de espaço de uma das variantes, (b) generalização de significado (os itens expandem seus contextos de uso), ou (c) especialização de uso (significados mais específicos restritos a certos contextos). Para as autoras, a abordagem sociofuncionalista traz, para o tratamento da variação linguística, um controle bastante refinado de grupos de fatores linguísticos, com a inclusão de restrições de natureza discursivo-pragmática que podem receber, de início, tratamento analítico escalar, sujeito a posteriores amalgamações.

Em defesa dessa proposta Sociofuncionalista, Gonçalves (2021), retomando os trabalhos de Naro e Braga (2000) e de Görski e Tavares (2013), mostra mais claramente como se implementaria esse trabalho de interface entre a Sociolinguística e a Teoria da

Gramaticalização, recorrendo aos *Princípios da estratificação e da divergência*, que, propostos por Hopper (1991), captam a natureza incipiente de processos de gramaticalização, quando uma forma ainda não se gramaticalizou completamente, a ponto de se tornar um clítico ou um afixo. Enquanto, por meio do primeiro, apreende-se a coexistência, em um mesmo recorte temporal, de diferentes formas que, em um mesmo domínio funcional, cumprem a mesma função, caso típico de variação, por meio do segundo, reconhece-se a coexistência de diferentes funções (ou significados) para a mesma forma que se encontra ainda em processo de mudança via gramaticalização. Parece claro que a mudança que atinge formas/construções rumo a funções que impactam a gramática das línguas é mais bem captada, na interface com a Teoria da Variação e Mudança, pelo Princípio da estratificação.

Considerados os pontos de convergência teórico-metodológica entre o construto funcionalista da Gramaticalização e a Teoria da Variação e Mudança Linguística, defendemos aqui que essa mesma proposta sociofuncionalista pode também ser estendida aos MD *entendeu?* e *sabe?*, com base no reconhecimento de que esses MD resultam de processos de gramaticalização que tornam construções de caráter lexical (com verbos plenos) em MD, com função interacional na gramática. Mais especificamente, construções interrogativas de base lexical se cristalizaram como perguntas retóricas até se esvaziarem de conteúdo e se tornarem MD, como ocorre com os MD em análise, para os quais se supõe a seguinte trajetória de mudança proposta em bases puramente sincrônicas: (i) *não é verdade? > não é? > né?* (URBANO, 1994; MARTELOTTA et al., 1996); (ii) *está bem/certo? > está? > tá?*; (iii) *você entendeu? > entendeu?*; (iv) *(você) sabe o que é? > sabe?* (MARTELOTTA et al., 1996).<sup>3</sup> Assumindo essas trajetórias semelhantes, neste artigo, daremos um tratamento variacionista, em perspectiva sincrônica, apenas aos MD *entendeu?* e *sabe?*, a fim de identificar em que contextos de uso uma das formas é escolhida de um domínio funcional mais amplo de *BAD*, e mostrar como é possível operar metodologicamente na interface entre variação e funcionalismo, conforme opções metodológicas apresentadas na próxima seção.

### 3 Aspectos metodológicos

O corpus utilizado na pesquisa é proveniente de uma subamostra do banco de dados Iboruna, constituído pelo Projeto ALIP (Amostra Linguística do Interior Paulista) e sediado na UNESP de São José do Rio Preto (GONÇALVES, 2007). Esse banco compõe-se de dois tipos de amostras do português falado no interior paulista, coletadas em sete cidades circunvizinhas da região de São José do Rio Preto, entre os anos de 2004 e 2006: (i) *Amostra Censo* (AC, daqui em diante), que reúne 151 entrevistas sociolinguísticas, socialmente estratificadas pelas variáveis censitárias *sexo/gênero* (masculino e feminino), *escolaridade* (até 4 anos, de 5 a 8 anos, de 9 a 11 anos e de 12 anos ou mais), *faixa etária* (7 a 15 anos, 16 a 25 anos, 26 a 35 anos, 36 a 55 anos e mais de 55 anos) e *renda familiar* em salários-mínimos (até 5, de 6 a 10, de 11 a 24 e de mais de 25); e (ii) *Amostra de Interação* (AI, daqui em diante), composta de 11 interações dialógicas, gravadas secretamente e sem controle de variáveis sociais, em diferentes situações de interação social envolvendo de dois a cinco participantes, que, posteriormente às gravações, concordaram livremente em ceder suas amostras de falas. Com base em um roteiro de entrevista, os informantes de AC produzem cinco tipos de texto: (i) *narrativa de experiência pessoal* (NE); (ii) *narrativa recontada* (NR); *descrição* (DE); *relato de procedimento* (RP); e (v) *relato de Opinião* (RO), todos de natureza predominantemente monológica e produzidos por estímulo de um documentador.

Para nossa pesquisa, utilizamos 32 entrevistas de AC, como se observa na tabela 1.

<sup>3</sup> Não é demais lembrar que, por se tratar de elementos gramaticais típicos da interação face a face, são raros trabalhos que investigam MD em perspectiva diacrônica.

**Tabela 1** - Subamostra de AC do Banco de dados Iboruna usada na pesquisa

Faixa etária/Sexo Escolaridade	15 a 25		26 a 35		36 a 55		+ de 55		Σ
	M	F	M	F	M	F	M	F	
Até 4 anos	1	1	1	1	1	1	1	1	8
5 a 8 anos	1	1	1	1	1	1	1	1	8
9 a 11 anos	1	1	1	1	1	1	1	1	8
+ 11 anos	1	1	1	1	1	1	1	1	8
Σ	4	4	4	4	4	4	4	4	32

Fonte: elaboração própria.

Como de início pretendíamos identificar propriedades da função *BAD* dos MD *né?*, *tá?*, *sabe?* e *entendeu?*, a investigação foi antecedida do levantamento de todos esses MD, tendo sido identificadas 7 ocorrências de *tá?*, 193 de *né?*, 87 de *entendeu?* e 145 de *sabe?*. Apesar de MD com função *checking e feedback* constituírem espécies de retorno do interlocutor na situação de interação, para a definição de nossa variável dependente consideramos que *sabe?*, *tá?*, *né?* e *entendeu?* cumprem a mesma função de *BAD* pertencente ao domínio funcional dos MD. No entanto, para operacionalização da proposta funcionalista, circunscrevemos nossa análise variacionista a *sabe?* e *entendeu?* e, por razões frequenciais, descartamos *tá?* e utilizamos *né?*, apenas como estratégia de delimitação do envelope de variação, por ser ele o protótipo da categoria da função *BAD*. Assim, nos concentramos no total dos 232 dados de *sabe?* e *entendeu?*, para operacionalizar a regra variável, que aqui será mostrada apenas em termos de tendência de uso, por recurso à proporção com que cada variante se correlaciona com as variáveis independentes. Uma análise sociolinguística mais sofisticada demandaria a ampliação do *corpus*, para a recolha de maior quantidade de dados de cada uma das variantes e para análises mais robustas. Como nosso intuito é averiguar o que motiva a escolha variável de *sabe?* e *entende?* na função *BAD*, propusemos fatores condicionadores, de forma a ser possível captar indícios de contextos de especialização de uso das variantes. É o que segue mostrado no quadro 1.

**Quadro 1** - Contextos de variação investigados

Variáveis		Variantes
Tipo	Especificação	
Dependente	Busca de aprovação discursiva	Sabe? / Entendeu? (sempre em contextos de uso semelhante a <i>né?</i> )
Independentes	Linguísticas	
	Posição do MD	Final de turno / Final de unidade comunicativa no meio do turno.
	Tipo de texto	Narrativa de experiência pessoal (NE) / Narrativa recontada (NR) / Relato de opinião (RO) / Relato de procedimento (RP) / Relato de descrição (DE)
	Sociais	
	Sexo/gênero	Masculino / Feminino
	Faixa etária	16 a 25 anos / 26 a 35 anos / 36 a 55 anos / + 55 anos
	Escolaridade	Até 4 anos (Ens. Fund. I) / 5 a 8 anos (Ens. Fund. II) / 9 a 11 anos (Ens. Médio) / Superior

Fonte: elaboração própria.

Como MD são de natureza fluida, não assumimos, *a priori*, qualquer hipótese que possa justificar o condicionamento das variáveis independentes propostas. Com a adoção de metodologia de Tavares (1999) no estudo de *aí*, *daí*, *e* e *então*, a expectativa é alcançar uma distribuição escalar dos resultados relacionada ao princípio meta-icônico da marcação, que prevê que categorias cognitivamente marcadas tendem a ser estruturalmente marcadas,

considerando-se três critérios básicos de marcação sugeridos por Givón (1995): (i) o de complexidade estrutural, sob o qual a estrutura marcada tende a ser mais complexa (ou maior) que a não marcada; (ii) o de distribuição de frequência, que prevê que a categoria marcada tende a ser menos frequente que a não marcada; (iii) e o de complexidade cognitiva, segundo o qual a categoria marcada tende a ser cognitivamente mais complexa, por demandar maior atenção, mais esforço mental e tempo de processamento, que a não marcada.<sup>4</sup> Seguindo, então, Tavares (1999), quanto à marcação, consideramos que *sabe?*, por sua constituição formal (2 sílabas), é menos marcado e, por isso, de uso privilegiado na fala de participantes com menor escolaridade e mais jovens, enquanto *entendeu?*, é mais marcado e de uso mais frequente por participantes com maior escolaridade e mais velhos, que tendem a usar expressões linguísticas mais complexas. Sobre a variável sexo/gênero, a expectativa é de que ela seja pouco reveladora de padrões de distribuição escalar de marcação. Esperamos, portanto, para a busca de correlações com as variáveis independentes, encontrar uma distribuição escalar das variantes *BAD* arranjada do seguinte modo: *sabe?* [- marcado] / *entendeu?* [+ marcado].

Realizamos, no programa *Goldvarb*, rodadas para o cruzamento da variável dependente com cada uma das independentes, de modo a monitorar os fatores condicionantes e, assim, entender quais favorecem mais cada uma das variantes. O programa permite quantificar os dados e analisá-los à luz dos mesmos grupos de fatores. Em função das variáveis, o programa fornece para cada variante frequência absoluta e proporção.

#### 4 Resultados e discussão

Dos contextos de variação, os resultados se mostraram categóricos para *posição do MD*, com todos eles aparecendo em meio de turno comunicativo, como em (3) e (4), onde representamos, entre colchetes, as paráfrases pela outra variante e pelo MD de controle *né?*.

- (3) Inf.: ah... a cama é:.... era de beliche **sabe?** [**né?/entendeu?**]... porque meus irmãos dormia tudo na beliche [BDI-AC-058; DE: L.220-223]
- (4) eu acho que as pessoas que tem MUIto... poderiam abrí(r) mão um po(u)co desse muito e começá(r) a ajudá(r) as pessoas... é um (inint.) que num depende só... deles... depende muito da gente também **entendeu?** [**né?/sabe?**]... tem pessoas que tem muito... e poderiam... doá(r) um po(u)quinho... pras pessoas que num tem nada [BDI-AC-035; RO: L.519-523]

Nas ocorrências acima, os informantes enunciam os MD em meio ao seu turno conversacional, sem esperar qualquer *feedback* do documentador, o que torna evidente a função *BAD* desses MD, a exemplo de função semelhante também cumprida por *né?*. Embora não esperássemos esse resultado categórico para a posição dos MD, a expectativa, ao propor essa variável, era a de que, em sua maioria, eles fossem de fato empregados no final de unidade comunicativa em meio de turno, já que, com função *BAD*, não seria de se esperar uma resposta ao MD e, portanto, sua ocorrência em posição final de turno.

Os tipos textuais das entrevistas sociolinguísticas são caracterizados e exemplificados, em (5), com a ocorrências de *sabe?* e *entendeu?*.

- (5) Caracterização e exemplificação de tipos de textos em que ocorrem os MD

**a. Narrativa de experiência (NE):** o informante narra evento vivenciado por ele.

<sup>4</sup> A título de esclarecimento, Tavares (1999) projeta a seguinte escala de marcação, do menos ao mais marcado, para os MD: *e > aí > daí > então*.

Inf.: o médico... ele serrô(u) o braço tirô(u) um pedaço do osso... pôs na mão... e colocô(u) pinos aqui no braço... só que agora eu tenho que fazê(r) O(u)tra cirurgia... [ou/] porque:... tá bonitinho **sabe?** no raio-x tudo... só que com o tempo... ele vai rejeitando... **entendeu?** vô(u) tê(r) que fazê(r) o(u)tra cirurgia pôr o(u)tro enxerto... [BDI-AC-066; NE: L. 114-120]

**b. Narrativa recontada (NR):** o informante narra eventos de que ele não participou.

Inf.: ah ele conta meio assim **sabe?** ((risos)) num dá pa entendê(r) muito bem o que ele conta [BDI-AC-058; NR: L. 154-157]

**c. Descrição (DE):** o informante descreve ambientes, objeto ou situações.

Inf.: ah... a cama é::... era de beliche **sabe?**... porque meus irmãos dormia tudo na beliche [BDI-AC-058; DE: L. 220-223]

**d. Relato de opinião (RO):** o informante emite opinião sobre algum tema.

eu acho que as pessoas que tem MUIta... poderiam abrí(r) mão um po(u)co desse muito e começá(r) a ajudá(r) as pessoas... é um (inint.) que num depende só... deles... depende muito da gente também **entendeu?**... tem pessoas que tem muito... e poderiam... doá(r) um po(u)quinho... pras pessoas que num tem nada [BDI-AC-035; RO: L.519-523]

**e. Relato de procedimento (RP):** descrição de etapas para realização de alguma tarefa.

Inf.: aí tipo nós fazia um relaTÓ::rio... falava das cenas... quem era o assassino do Vygotsky **entendeu?** tipo uma dinâmica... como que eu vô(u) te falá(r)? [BDI-AC-039; RP: L. 237-241]

Adaptando a proposta de Tavares (1999) sobre a atuação do princípio da marcação para tipos textuais aqui considerados, a expectativa seria a de que, em relação à função *BAD*, RO seria [+ marcado], em oposição à NE, tipo [- marcado], tendo como tipos intermediários NR, RP e DE. Assim, a hipótese final é de que *entendeu?* ([+ marcado]) seria de uso privilegiado em RO ([+marcado]), enquanto *sabe?* ([-marcado]) seria mais frequente em NE ([-marcado]). Os resultados para essa variável, única relevante, são mostrados na tabela 2.

**Tabela 2 - Total de aplicação das variantes *sabe?* e *entendeu?* versus tipo de texto**

Fatores	MD	<i>sabe?</i>		<i>entendeu?</i>		Σ	
		No	%	No	%	No	%
narrativa de experiência pessoal (NE)		59	76,6	18	23,4	77	33,2
narrativa recontada (NR)		49	86	8	14	57	24,6
relato de descrição (DE)		19	57,6	14	42,4	33	14,2
relato de opinião (RO)		10	21,3	37	78,7	47	20,3
relato de procedimento (RP)		8	44,4	10	55,6	18	7,8
	Σ	145	62,5	87	37,5	232	100

Fonte: elaboração própria.

Como se observa na tabela 2, os dois MD se concentram mais nos dois tipos narrativos, com destaque para a NE, tipo no qual se concentra a maior parte dos dados ( $77/232 = 33,2\%$ ) e que, de fato, condiciona a escolha de *sabe?* ( $59/76 = 76,6\%$ ), enquanto RO e RP condicionam o aparecimento de *entendeu?* ( $37/47 = 78,7\%$  e  $10/18 = 55,6\%$  respectivamente), resultado que confirma nossa hipótese de marcação acerca da influência do tipo de texto sobre o emprego dos MD: NE ([- marcado]) favorece o uso do MD *sabe?* ([- marcado]), enquanto RO ([+ marcado]) favorece o emprego de *entendeu?* ([+ marcado]).

Os tipos de textos NR, segundo a concentrar mais ocorrências dos MD ( $57/232 = 24,6\%$ ), e DE, com  $33/232 (= 14,2\%)$ , favorecem o uso *sabe?* ( $49/57 = 86\%$  e  $19/33 = 57,6\%$ , respectivamente), enquanto, em menores proporções, RO ( $47/232 = 20,3\%$ ) e RP ( $18/232 =$

7,8%) favorecem o uso de *entendeu?* (78,7% = 37/47 e 55,6% = 10/18, respectivamente). A predominância de *sabe?* em NE, NR e DE revela uma tendência de MD [-marcado] ocorrer em contextos também [- marcado], enquanto a predominância de *entendeu?* em RO e RP revela a tendência em direção oposta, por demandarem mais tempo de processamento e talvez manterem ainda traços que possam levar o ouvinte a interpretar que seu interlocutor espera dele uma confirmação. Em termos de comparação com *né?*, tipo ainda mais [- marcado] por sua menor complexidade estrutural e cognitiva e pela sua alta frequência, pode-se considerar então que *sabe?* é o mais próximo dele, quando empregado nos mesmos tipos textuais. Com base nesse resultado, é possível, então, chegarmos à escala de marcação para *sabe?* e *entendeu?* dada em (6).

(6) Escala de marcação de *sabe?* e *entendeu?* e tipos textuais

Contexto de marcação:	[-marcado]	-----	[+ marcado]
MD	:	<i>sabe?</i>	<i>entendeu?</i>
Tipo textual:	:	NE > NR > DE	RO > RP

Consoante os princípios de marcação estabelecidos por Givón (1995), a escala em (6) justifica então que, das variantes na função *BAD*, *sabe?* é a forma [- marcada] pelas seguintes evidências: (i) mantém forte correlação com tipos textuais também [- marcados], como é o caso das narrativas; (ii) é estruturalmente menos complexo, por ser forma dissilábica e cristalizada em 3<sup>a</sup>. pessoa singular e presente; (iii) em termos de proporção, é de frequência mais acentuada; (iv) revela menor complexidade cognitiva, porque, é pouco provável que seja processado como pergunta que espera do ouvinte uma manifestação de aprovação. Em direção oposta, *entendeu?* é considerado [+ marcado], por sua maior complexidade cognitiva e estrutural e por sua frequência de uso mais baixa e em tipos textuais [+ marcados], como é o caso de RO e RP, evidências que podem ainda favorecer a preservação de traços do contexto de que ele se origina, ou seja, de pergunta que aguarda do interlocutor a aprovação discursiva para a continuidade do fluxo conversacional, embora isso não se verifique nos dados.

Na verificação da influência de variáveis sociais sobre o uso dos MD com função *BAD*, retomando aqui nossas expectativas, nos valem os critérios de marcação propostos por Givón (1995) em lugar dos conceitos de *variantes padrão vs. não padrão, estigmatizada vs. prestigiada*, por estarmos lidando com um fenômeno discursivo variável, de caráter fluido, e, por isso, pouco sujeito à avaliação social. Assim, os critérios de marcação dariam conta de possíveis correlações entre aspectos cognitivos e sociais no uso dos MD em exame. A previsão então é de que *sabe?*, por ser cognitivamente menos complexo em termos de marcação, se manifestaria mais na fala de indivíduos mais novos e com menor escolaridade, ao passo que *entendeu?*, variante [+ marcada], estaria mais presente na fala de indivíduos mais velhos e de maior escolaridade. Em termos de marcação, previmos não haver diferença significativa no uso dos MD por homens e mulheres. Os resultados da tabela 3, para a variável *faixa etária*, confirmam apenas sutilmente nossa hipótese.

**Tabela 3** - Influência da faixa etária na escolha das variantes *sabe?* e *entendeu?*

Fatores	MD				Σ	
	No	%	No	%	No	%
16 a 25 anos	60	73,2	22	26,8	82	35,3
26 a 35 anos	17	34,7	32	65,3	49	21,1
36 a 55 anos	34	87,2	5	12,8	39	16,8
mais de 55 anos	34	54,8	28	45,2	62	26,7
Σ	145	62,5	87	37,5	232	100

Fonte: elaboração própria.

Como mostram os resultados da tabela 3, de fato, indivíduos da faixa etária mais nova usam mais frequentemente *sabe?* ( $60/82 = 73,2\%$ ), e da faixa etária imediatamente posterior (26 a 35 anos) preferem *entendeu?* ( $32/49 = 65,3\%$ ). Entretanto, esse comportamento não chega a ser um traço particular de indivíduos mais novos, visto que ele também se reproduz para indivíduos das duas faixas etárias mais velhas, de 36 a 55 anos ( $34/39 = 87,2\%$ ) e de + 55 anos ( $34/62 = 54,8\%$ ), mais acentuadamente para aquela do que para esta. A particularidade fica mesmo por conta de indivíduos da faixa de 26 a 35 anos, que empregam a variante *sabe?* em menor proporção ( $17/32 = 34,7\%$ ). Em que pese essas semelhanças e diferença detectadas nesses resultados, é possível afirmar que, nos termos da hipótese formulada, o efeito de faixa etária sobre a regra variável é praticamente nulo, visto que a distribuição escalar de marcação de MD não reflete correlações esperadas para as diferentes faixas etárias. Essa asserção se confirma, mesmo diante da amalgamação dos dados das duas faixas etárias mais novas e das duas mais velhas. Para essas duas faixas de idade agora consideradas, confirmamos que a tendência de escolha da variante [- marcada] *sabe?* anula a hipótese formulada, como se pode verificar nos resultados da tabela 4, que mostram um padrão de comportamento muito semelhante entre os indivíduos de faixa etária mais nova e dos de faixa etária mais velha, que preferem usar *sabe?* a *entendeu?* na busca de aprovação discursiva.

**Tabela 4** - Amalgamação de faixas etárias e escolha das variantes *sabe?* e *entendeu?*

Faixa etária \ MD	<i>Sabe?</i>		<i>Entendeu?</i>		$\Sigma$	
	No.	%	No.	%	No.	%
16 a 35 anos	77	<b>58,7</b>	54	41,3	131	56,4
Mais de 36 anos	68	<b>67,3</b>	33	33,6	101	43,6
$\Sigma$	145	62,5	87	37,5	232	100

Fonte: elaboração própria.

Apresentamos, na tabela 5, os resultados para a variável *escolaridade*, para a qual também não pudemos apreender um padrão de distribuição regular dos MD em função do aumento do nível escolar.

**Tabela 5** - Influência da escolaridade na escolha das variantes *sabe?* e *entendeu?*

Fatores \ MD	<i>sabe?</i>		<i>entendeu?</i>		$\Sigma$	
	No.	%	No.	%	No.	%
Até fundamental (até 4 anos)	31	<b>75,6</b>	10	24,4	41	17,7
Fundamental II (5 a 8 anos)	21	31,8	45	<b>68,2</b>	66	<b>28,4</b>
Médio (9 a 11 anos)	65	<b>74,7</b>	22	25,3	87	<b>37,5</b>
Superior	28	<b>73,7</b>	10	26,3	38	16,4
$\Sigma$	145	62,5	87	37,5	232	100

Fonte: elaboração própria.

Como se observa, na distribuição total dos dados, indivíduos dos níveis escolares intermediários empregam mais MD do que os dos níveis extremos, respondendo, juntos por cerca de 2/3 dos dados ( $66+87 = 153/232 = 66\%$ ). Entretanto, enquanto os do Fundamental II empregam mais o MD [+marcado] *entendeu?*, os do nível médio optam pelo [- marcado] *sabe?*, o que já indicia a nulidade de nossa hipótese, se associarmos esse resultado díspar ao padrão praticamente idêntico de comportamento dos indivíduos dos níveis extremos de escolaridade, que preferem empregar largamente *sabe?* em lugar de *entendeu?*. Assim, em termos de marcação, o correlato cognitivo que explicaria uma possível tendência de quanto maior a

escolaridade maior a tendência de emprego de MD [+ marcado] não se confirma, muito possivelmente por estarmos lidando com uma regra variável definida a partir de regras discursivas que orientam o discurso, e não a fonologia ou a morfossintaxe da língua.

Das variáveis sociais, *sexo/gênero* foi a que apresentou um padrão distintivo de comportamento entre indivíduos do sexo masculino e feminino. Diferentemente do que prevíamos, ao postular que essa seria uma variável com efeitos nulos sobre a regra variável, os resultados da tabela 6 deixam evidente que mulheres tendem mais ao emprego da variante [- marcada] *sabe?*, e os homens, ao da variante [+ marcada] *entendeu?*

**Tabela 6** - Influência da variável *sexo/gênero* na escolha das variantes *sabe?* e *entendeu?*

Fatores	MD	<i>sabe?</i>		<i>entendeu?</i>		$\Sigma$	
		No	%	No	%	No	%
Feminino		116	<b>83,5</b>	23	16,5	139	59,9
Masculino		29	31,2	64	<b>68,8</b>	93	40,1
	$\Sigma$	145	62,5	87	37,5	232	100,0

Fonte: elaboração própria.

Na base da hipótese postulada, que considera critérios de marcação em lugar do possível *status* social das variantes, parece-nos pouco provável que o padrão de comportamento de homens e mulheres possa ser explicado em termos de complexidade cognitiva e estrutural de *sabe?* e *entendeu?* e tão pouco em termos de marcação de frequência associada a um dos sexos. No entanto, o padrão de comportamento claramente diferenciado de homens e mulheres no uso dos dois MD talvez seja mais bem explicado na base da hipótese clássica da sociolinguística sobre inovações linguísticas implementadas por mulheres (LABOV, 1972), com a ressalva de que essa inovação não seria decorrente de avaliação social de uma ou de outra variante, mas simplesmente do reconhecimento do uso mais espreado de uma delas – a variante *sabe?* – na comunidade de fala, explicação que encarece a importância do papel da frequência de uso. Isso significa dizer que o MD [- marcado] *sabe?*, por estar, na comunidade de fala, mais difundido do que *entendeu?*, encontraria nas mulheres maior força de propagação, sem que isso seja revelador de qualquer indício de mudança em direção à obsolescência da variante *entendeu?*.

## 5 À guisa de conclusão

Para evidenciar a viabilidade de uma proposta sociofuncionalista na investigação de MD, lançamos mão de um estudo de caso, elegendo os MD *sabe?* e *entendeu?* e tratando-os, de uma perspectiva variacionista, como variantes da mesma variável, na medida em que, ao se gramaticalizarem como MD, eles se estratificam nesse novo domínio funcional e passam a disputar a mesma função de *BAD*, dadas suas equivalências funcionais apreendidas em contextos de uso efetivo da língua. A exemplo do que já se propusera para o tratamento de casos de variação sintática (LAVANDERA, 1978, 1984; ROMAINE, 1984), o conceito de regra variável da Sociolinguística requereu, em nossa investigação, um afrouxamento para dar conta do tratamento de formas diferentes resultantes de gramaticalização, mas que admitem comparabilidade funcional.

Propusemos, então, um conjunto de fatores condicionadores da regra variável, de forma a exemplificar a condução de um trabalho de interface entre Sociolinguística Variacionista e Gramaticalização para tratar dos MD com a função *BAD*. Em face da natureza inerentemente discursiva do fenômeno variável, nos apropriamos do Princípio cognitivo de marcação (GIVÓN, 1995), estabelecendo, então, diferenças salientes entre as duas variantes, em termos de complexidade estrutural, complexidade de processamento e frequência de uso, o que permitiu reconhecer *sabe?* como forma [- marcada] e [*entendeu?*], como [+ marcada]. Das

variáveis controladas, os resultados, para a regra variável, mostraram como relevantes a variável linguística *Tipo de texto* e a variável social *Sexo/gênero*, relevância por nós interpretada, dados os objetivos do trabalho, com base apenas em tendências de uso, apreendidas pela proporção com que as variantes se manifestam nesses dois contextos. A partir daí, é possível, então, prever como prosseguir com o trabalho de investigação, como sugerem Görski e Tavares (2013), lançando para o fenômeno um olhar ou mais variacionista, adotando-se métodos estatísticos mais apurados para confirmação das variáveis consideradas relevantes, ou mais funcionalista, propondo um conjunto maior de variáveis linguísticas que controlem diferenças funcionais mais sutis entre as variantes.

Os resultados aqui alcançados nos permitem algumas generalizações na investigação de MD em abordagem sociofuncionalista: (i) o cuidado na associação de variantes a valores sociais que elas possam não apresentar, a exemplo de outros fenômenos variáveis mais afeitos a julgamentos e a avaliações sociais, solução que nos levou a adotar o Princípio cognitivo de marcação, que, ainda assim, se mostrou relevante, pelo critério de frequência, apenas para a variável social *sexo/gênero*, não interpretada como um indício de distinção cognitiva entre homens e mulheres, mas como mulheres sendo uma força propagadora de usos de variantes mais frequentes na comunidade de fala; (ii) a necessidade de controle de tipo textual, visto que os MD são uma categoria pragmática com fortes efeitos na organização textual-interativa, o que, na análise empreendida, mostrou-se relevante por meio de correlações alcançadas entre tipos textuais considerados de uso [- marcado], como narrativas, e variante também considerada [- marcada], no caso, a variante *sabe?*; (iii) a atenção para a expectativa com que se propõe o controle de variáveis sociais na investigação de fenômeno discursivo variável, pois, a exemplo de variação de nível sintático (BENTIVOGLIO, 1987), variáveis sociais podem ter efeito nulo sobre a regra variável, em razão de a variação operar abaixo do nível de consciência dos usuários da língua e não estar, portanto, sujeita à avaliação social; isso não significa dizer que elas devem ser descartadas já na proposição do envelope de variação, mas que, antes de se descartá-las, é preciso provar sua nulidade; (iv) a necessidade de cruzamentos entre variáveis que possam revelar lócus sociais e linguísticos mais precisos de usos das variantes, tarefa não executada nesta investigação, dados seus propósitos mais programáticos.

Retomados nossos propósitos e mostradas algumas generalizações advindas de autocrítica à nossa própria empreitada, estamos convencidos da viabilidade de adoção da abordagem sociofuncionalista para estudo de MD, inclusive de outros MD fortemente orientados para a interação ou mesmo para a organização textual. Outra perspectiva promissora para a qual este estudo aponta é a necessidade de investigação diacrônica de MD, de forma a evidenciar tanto os contextos propícios de sua gramaticalização quanto o desenvolvimento de suas funções textual-interativas.

## Referências

- BENTIVOGLIO, P. A variação nos estudos sintáticos. **Estudos Linguísticos**, São Paulo, n. 15., p. 7-29, 1987.
- BYBEE, J.. **Língua, uso e cognição**. São Paulo: Cortez, 2016.
- BYBEE, J.; HOPPER, P. (Ed.). **Frequency and the emergence of linguistic structure**. Amsterdam: John Benjamins, 2001.
- GIVÓN, T. **Functionalism and grammar**. Amsterdam: John Benjamins, 1995.
- GOLDVARB X for windows. [s.l.]: [s.n.], 2005. Disponível em: <http://individual.utoronto.ca/tagliamonte/goldvarb.html>. Acesso em abr.2021.
- GONÇALVES, S.C.L. **Banco de dados Iboruna: amostras eletrônicas do português falado no interior paulista**. 2007. Disponível em: <http://www.alip.ibilce.unesp.br>. Acesso em out. 2020.

- GONÇALVES, S.C.L. Ainda a favor de uma interface entre Sociolinguística e Gramaticalização. **Working papers em Linguística**, Florianópolis, v. 22, n. esp., p. 111-136, 2021.
- GÖRSKI, E.M.; TAVARES, M.A. Reflexões teórico-metodológicas a respeito de uma interface sociofuncionalista. **Revista do GELNE**, v. 15, n. 1/2, p. 79-101, 2013.
- GUERRA, A.R. **Funções textual-interativas dos marcadores discursivos**. 233f. 2007. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2007.
- HOPPER, P. On some principles of grammaticization. *In*: TRAUGOTT, E.; HEINE, B. (Ed.). **Approaches to grammaticalization**. Philadelphia: John Benjamins Publishing Company, 1991. p. 17-35.
- HOPPER, P., TRAUGOTT, E. **Grammaticalization**. 2.ed. Cambridge: Cambridge University Press, 2003.
- JUBRAN, C.C.A.S. Uma gramática textual de orientação interacional. *In*: CASTILHO, A.T. et al. (Org.). **Descrição, história e aquisição do português brasileiro**. Campinas; São Paulo: Pontes; FAPESP, 2007. p. 313-327,
- JUBRAN, C.C.A.S; KOCH, I. G. V. (Org.). **Gramática do português culto falado no Brasil: a construção do texto falado**. Campinas: Editora da UNICAMP, 2006.
- LABOV, W. **Sociolinguistic patterns**. Pennsylvania: University of Pennsylvania Press, 1972.
- LAVANDERA, B. Where does the sociolinguistic variable stop? **Language in Society**, Cambridge, v. 7, n. 2, , p. 171-182, 1978.
- LAVANDERA, B. **Variación y significado**. Buenos Aires: Hachette, 1984.
- MARTELOTTA, M.E.; VOTRE, S.; CEZÁRIO, M.M. (Org.). **Gramaticalização no português do Brasil: uma abordagem funcional**. Rio de Janeiro: Tempo Brasileiro/UFRJ, 1996.
- NARO, A.J.; BRAGA, M.L. A interface sociolinguística/gramaticalização. **Gragoatá**, Niteroi, v.9, n. 1, p. 125-134, 2000.
- RISSO, M.S. Marcadores discursivos basicamente interacionais. *In*: JUBRAN, C.C.A.S; KOCH, I.G.V. (Org.). **Gramática do português culto falado no Brasil: a construção do texto falado**. Campinas: Editora da UNICAMP, 2006a. p. 497-527.
- RISSO, M.S. Marcadores Discursivos basicamente seqüenciadores. *In*: JUBRAN, C.C.A.S; KOCH, I.G.V. (Org.). **Gramática do português culto falado no Brasil: a construção do texto falado**. Campinas: Editora da UNICAMP, 2006b. p. 427-496.
- RISSO, M.S.; OLIVEIRA E SILVA, G.M. de; URBANO, H. Traços definidores dos Marcadores Discursivos. *In*: JUBRAN, C.C.A.S; KOCH, I.G.V. (Org.). **Gramática do português culto falado no Brasil: a construção do texto falado**. Campinas: Editora da UNICAMP, 2006. p. 403-425
- ROMAINE, S. On the problem of syntactic variation and pragmatic meaning in sociolinguistic theory. **Folia Linguistica**, Berlim, v. 18, n.3-4, p. 409-437, 1984.
- TAVARES, M.A. **Um Estudo variacionista de aí, daí, então e e como conectores seqüenciadores retroativo-propulsores**. 176f. 1999. Dissertação (Mestrado em Linguística) - Universidade Federal de Santa Catarina, Florianópolis, 1999.
- URBANO, H. Marcadores conversacionais: o caso do *né*? *In*: SEMINÁRIO DO GEL, XXIII, 1994. **Anais ...** São Paulo: GEL. p. 1430-1437.
- URBANO, H. Marcadores conversacionais: um novo aspecto do *né*? *In*: SEMINÁRIO DO GEL, XXIV, 1995. **Anais...** Ribeirão Preto: GEL. p. 660-666.
- URBANO, H. Marcadores discursivos basicamente interacionais. *In*: JUBRAN, C.C.A.S; KOCH, I.G.V. (Org.). **Gramática do português culto falado no Brasil: a construção do texto falado**. Campinas: Editora da UNICAMP, 2006. p. 497- 527.

WEINREICH, U.; LABOV, W.; HERZOG, M. Empirical Foundations for Theory of Language Change. *In*: LEHMANN, P.; MALKIEL, Y. (Ed.) *Directions for Historical Linguistics*. Austin: University of Texas Press, 1968. p. 95-188.

Submetido em 19/03/2022

Aceito em 24/05/2022