

O Conector Discursivo: uma unidade informacional?

Elias Oliveira

Faculdade de Letras

Universidade Federal de Minas Gerais (UFMG)

Belo Horizonte

elias.ptu@hotmail.com

Tommaso Raso

Faculdade de Letras

Universidade Federal de Minas Gerais (UFMG)

Belo Horizonte

tommaso.raso@gmail.com

Marcelo Vieira

School of Communication Sciences and Disorders

McGill University

marcelo.vieira@mail.mcgill.ca

Montreal

Abstract: O trabalho discute a unidade chamada *Conector Discursivo* (DCT), com base na *Language into Act Theory*. Trata-se de uma unidade prosódica preenchida lexicalmente quase unicamente por conjunções, para a qual foi proposta a etiqueta informacional de Marcador Discursivo com a função de conector genérico entre padrões informacionais. Dialogando com o trabalho de Cresti e Moneglia, este paper, a partir de dados de um corpus de fala espontânea do português brasileiro (PB), coloca em discussão a função informacional do DCT. Nós realizamos a etiquetagem de um minicorpus de 31.464 palavras, extraído do corpus C-ORAL-BRASIL e balanceado. Observamos que existem diversos aspectos que levam a pensar que o DCT não apresenta as propriedades que permitem considerá-lo como unidade informacional com base nos critérios que definem as outras unidades: principalmente (i) falta de uma forma prosódica própria; (ii) falta de uma função pragmática bem definida; e (iii) dificuldade em distinguir o seu estatuto pragmático de valores de natureza semântico-sintática. Neste trabalho mostramos (i) a variabilidade prosódica; (ii) a existência de um contínuo entre uma função claramente pragmática (em que a conjunção perde seu valor semântico-sintático) e a função de operador lógico; (iii) a dificuldade de distingui-lo de unidades escansionadas (unidades prosódicas que compõem uma unidade informacional maior). Em geral as unidades identificadas como DCT mostram uma tendência a funcionarem como estratégia por parte do falante para planejar a formulação de um padrão informacional que estabeleça continuidade discursiva, com graus de pragmaticização muito variáveis.

Keywords - Prosódia; Linguística de corpus; Estrutura informacional; Marcador discursivo

I. INTRODUÇÃO

Este trabalho discute a unidade chamada Conector Discursivo/DiscourseConnector (DCT) com base na *Language into Act Theory* (L-AcT) [1], [2], [3], dialogando com [4] e [5] sobre o mesmo objeto e dentro do mesmo arcabouço teórico. L-AcT é uma teoria corpus-driven que se constitui como extensão da teoria dos atos de fala de Austin [6] integrada com uma teoria sobre a estrutura informacional da fala espontânea. L-AcT parte da segmentação do fluxo da fala em unidades terminadas (UT), ou seja, marcadas por uma fronteira prosódica com valor conclusivo. As UTs podem ser de dois tipos: o enunciado e a *stanza*. O enunciado é constituído por um único padrão informacional terminado; a *stanza* é constituída pela justaposição de dois ou mais padrões, cuja transição é marcada por um sinal prosódico de continuidade. Um padrão informacional é formado por uma ou mais unidades prosódicas, substancialmente isomórficas com as unidades informacionais (UI): em princípio, a cada unidade prosódica corresponde uma UI. O núcleo do padrão informacional é a unidade que carrega a ilocução, o *comentário* (COM), a única suficiente e necessária para formar um padrão informacional. As UI podem ser de dois tipos: textual e dialógico.

As textuais constituem o conteúdo semântico do padrão¹; as dialógicas correspondem ao que em outros arcahouços são chamados Marcadores Discursivos (MD).

Contudo, uma UI textual pode ser realizada através de mais de uma unidade prosódica, chamadas unidades de *escansão* (SCA). Isso pode acontecer por razões articulatórias (a UI é muito grande), por dificuldades na realização (neste caso as SCA são frequentemente acompanhadas por formas de disfluências) ou por motivos enfáticos. As SCA não apresentam proeminência ou forma prosódica informacionalmente funcional; esta se encontra sempre na última unidade prosódica da mesma UI. Além disso, as unidades de SCA são sempre semântica e sintaticamente composicionais com as outras unidades prosódicas da mesma UI. O exemplo 1 mostra um COM formado por três unidades prosódicas. As duas primeiras são SCA. Os exemplos podem ser escutados no link <https://www.dropbox.com/sh/c95xbvs7gceus53/AABfnUjJt4x3THeB2HJ8IHtAa?dl=0>

Exemplo 1 – SCA em unidade de COM - bpubdl01[35]

PAU: [35] então tem que /=SCA= colocar mais uma /=SCA= carreira de pedra ai //COM=

Cada padrão informacional de enunciado apresenta um COM, que é pragmática e prosodicamente interpretável em isolamento; na *stanza*, as unidades ilocucionárias de cada padrão, com exceção da última, são chamadas *comentários ligados/boundcomments* (COB), para marcar o fato que elas são acompanhadas por um sinal prosódico de continuidade. Qualquer padrão pode apresentar outras unidades opcionais com funções diversas, não ilocucionárias. Estas unidades são prosódica e funcionalmente subordinadas à ilocução. Cada UI se define com base em uma função (por exemplo, o tópico é definido funcionalmente como a unidade que estabelece o âmbito cognitivo de aplicação da força ilocucionária), uma forma prosódica específica capaz de veicular a sua função, e uma distribuição com relação ao COM. Isso vale tanto para as unidades textuais quanto para as unidades dialógicas. Aqui não nos ocuparemos das unidades textuais.

¹Sobre o Tópico, veja-se em particular [7], [3] e [8]. Quanto à discussão sobre as diferenças entre o conceito de Tópico em L-AcT e na Alternative Semantics de [9], veja-se [10]. Sobre o Parentético, veja-se [11] e [12]. Sobre o Introdutor Locutivo [13].

II. AS UNIDADES DIALÓGICAS/MARCADORES DISCURSIVOS: A LITERATURA

As unidades dialógicas (ou MD) tem sido estudadas dentro L-AcT com o objetivo de responder a duas perguntas cruciais sobre elas: (i) como estabelecer quando um item lexical (ou locução curta) funciona como MD e quando realiza outra função? (ii) uma vez estabelecido que um lexema funciona como MD, como determinar a sua função específica? A literatura de fato não apresenta uma resposta satisfatória a essas perguntas². L-AcT propõe que a resposta para ambas as perguntas deva ser encontrada na prosódia e não no léxico, contrariamente ao que acontece em quase toda a literatura sobre o assunto. Apesar de não existir uma definição comumente aceita para os MD, alguns pontos parecem encontrar acordo na literatura: os MD seriam elementos não composicionais com o resto do enunciado, cujo significado semântico se perde em prol de funções pragmáticas. Contudo, a literatura não explica como seria marcada formalmente a não composicionalidade e a consequente pragmaticalização.

Inicialmente, dentro L-AcT, [1] propôs a existência de quatro unidades com funções diferentes, todas dirigidas ao interlocutor, e portanto externas ao conteúdo proposicional do enunciado e não composicionais. [17] acrescenta mais duas unidades, das quais uma com função interacional e o próprio DCT, cuja função seria coesiva e que constitui o objeto deste trabalho. [18] tenta analisar todas essas unidades em dois corporacomparáveis de PB e Italiano, extraídos respectivamente do C-ORAL-BRASIL [19] e do C-ORAL-ROM [20]. Retomando o trabalho de [1] e [17], [18] propõe que a resposta a pergunta (i) seria a seguinte: um MD é uma unidade informacional isolada em unidade prosódica delimitada por fronteiras, das quais pelo menos uma não terminal. A unidade não possui interpretabilidade em autonomia, e portanto não é ilocucionária; a fronteira prosódica marcaria a não composicionalidade com o resto do enunciado e a sua pragmaticalização. Se explicaria assim a diferença entre (a) e (b):

- (a) Olha a casa de vidro.
- (b) Olha, a casa de vidro.

Em (a) *a casa de vidro* funciona como objeto de *olha*, enquanto em (b) *olha* não é composicional, poderia ser substituído lexicalmente por diversos lexemas de diversas classes sem perder a sua função e constituiria um MD. Evidentemente, a não composicionalidade não pode ser devida ao léxico e só pode ser marcada pela fronteira prosódica e pela falta de autonomia do perfil prosódico do lexema. Em caso de interpretabilidade pragmática e prosódica em autonomia a unidade é ilocucionária; isso pode acontecer também com as interjeições, mostrando que uma interjeição não deve necessariamente ser um MD.

Contudo, ainda precisava-seresponder à pergunta (ii): encontrar uma descrição prosódica adequada das formas dos diferentes tipos de MD e uma melhor definição funcional. Os trabalhos de [21], [22] e [23]aperfeiçoaram a proposta inicial chegando a propor seis tipos de MD, dos quais cinco bem definidos funcionalmente e formalmente, enquanto o sexto ainda está sendo estudado. Todos são de natureza interacional e não consideram o DCT.

III. O CONECTOR DISCURSIVO

Como já dissemos, [17] propõe um MD de natureza coesiva que ele chama Conector Discursivo/DiscourseConnector (DCT) com as características seguintes: (a) funcionalmente conectaria enunciados ou padrões de *stanza*, marcando a continuidade discursiva; (b)prosodicamente apresentaria uma duração longa; (c)distribucionalmenteestaria sempre em posição inicial de enunciado ou de padrão de *stanza*; (d) lexicalmente seria preenchido por conjunções dessemantizadas

²A literatura sobre os Marcadores Discursivos é muito rica e diversificada. Citamos apenas [14], [15] e [16].

e que adquiririam valor genérico de conector pragmático. Pela sua própria característica funcional, é compreensível que o DCT seria o único MD mais presente em monólogos. De fato, a natureza funcional do DCT, o torna útil na conexão entre os padrões da *stanza*. Isso não exclui a sua presença em começo de enunciado e até de turno, como conector genérico com o enunciado ou o turno anterior.

Contudo, como distinguir um DCT de uma unidade de escansão? De fato, no caso das conjunções, não é fácil estabelecer se existe ou não composicionalidade entre uma conjunção isolada em unidade prosódica e o que segue. Como decidir se um *porque*, um *aí*, um *e* isolados em unidade entonacional são composicionais ou não com o que segue? [4] e [5] propõem dois critérios: a eliminabilidade ou a substituíbilidade. Se uma conjunção isolada em unidade prosódica pode ser eliminada ou substituída por outra conjunção que, do ponto de vista lógico-semântico e sintático, estabeleceria uma relação diferente entre os dois padrões, sem que isso cause prejuízo no valor semântico ou ilocucionário, então se trataria de um DCT; contrariamente, se trataria de SCA. Casos como (c)-(f), que traduzem exemplos de [5], exemplificariam conjunções que não podem ser eliminadas:

(c) *não sei **porque** fez isso. *não sei fez isso.*

(d) *é caro **mas** funciona. *é caro funciona.*

(e) *ensina matemática **e** ajuda os alunos. *ensina matemática ajuda os alunos.*

(f) *ele foi a pé, **então** vai chegar tarde. ele foi a pé, vai chegar tarde.*

Já esses exemplos apresentam, a nosso ver, um problema. Enquanto (c)-(e) são constituídos por uma única unidade prosódica, (f) é constituído por duas unidades prosódicas. Neste caso, nada impede que a relação consequencial seja mantida mesmo sem a conjunção. Ela seria inferível mesmo em ausência da conjunção e um conteúdo como *ele foi a pé, vai chegar tarde* seria perfeitamente interpretável com o mesmo valor consequencial. O mesmo aconteceria se a conjunção fosse isolada em uma unidade prosódica dedicada. Essa observação mostra como o princípio da eliminabilidade não permite uma distinção clara entre valores lógico-sintáticos e valores pragmáticos.

Passamos agora ao princípio da substituíbilidade. [4] fornece alguns exemplos, um dos quais é reportado a seguir em italiano e em tradução:

Exemplo 2 – um caso de substituíbilidade apresentado em [4]

*TIZ: <lei c’ha da tenergli l’amministrazione /=COB= **quindi**(perché / infatti)=DCT= dice /=INT= a casa /=TOP-r= tanto con Federico non fo’ niente /=COM-r=

*Ela precisa manter a administração para ele / **então** (porque / de fato) diz / em casa / com o Frederico eu não consigo nada //*

Contudo, o que observamos é que a substituíbilidade é possível em graus variados. Foram encontradas 129 ocorrências em nossos dados. Em cerca de 11% dos casos, não parece possível nenhuma substituição sem que se mude o significado semântico ou ilocucionário; em cerca de 21% dos casos a substituição é possível apenas parcialmente, ou seja, com poucos lexemas que preservam a mesma interpretação lógica ou alguma outra possível. Em cerca da metade dos casos o lexema pode ser substituído livremente, sem prejuízo da interpretação semântico-ilocucionária. Existiria, assim, uma gradualidade na substituíbilidade. Mencionamos também o caso interessante do *porque* epistêmico [26], em que o valor causal não é de natureza sintática; a causalidade é relacionada à ilocução que antecede. Os *porque* epistêmicos ocupam cerca de 16% dos casos. Mas outros casos podem ser interpretados tanto de maneira epistêmica quanto sintaticamente, e muitos *porque* não em unidade dedicada (e, portanto, não candidatos a DCT) também são de tipo epistêmico.

Exemplo 3– porque epistêmico não isolado em unidade prosódica

*DFL: que o meu avô /=TOP= era de uma família abastada /=COB= **porque** o professor ia em casa /=COB= nã ia po grupo não //COM=

Aqui, *porque* não pode ser interpretado como estritamente sintático sob o perfil proposicional. Não podemos dizer que *a família era abastada porque o professor ia em casa*; a relação causal se dá entre as ilocuções. O significado pode ser assim parafraseado: *o meu avó era de uma família abastada e digo isso porque o professor ia em casa*. Ou seja: realizei a ilocução anterior *porque*... Nem seria difícil substituir o *porque* com *tanto é que / de fato / e /...*

Exemplo 4– porque epistêmico isolado em unidade prosódica

*ALO: eu nã vou falar nome da cidade não /=COB= só panũ [1]=SCA= nãcompricar a coisa /=COB= **porque** /=DCT= a dona Elvira tá viva ainda hhh /=COB= depois ea fica sabendo disso /=TOP= e pode querer acertar comigo /=COB= então /=DCT= melhor ficar assim /=COB= do jeito que tá aí //COM=

Quanto à prosódia, [4] e [5] não individualizam uma forma prosódica definida, como é ao contrário o caso de todas as unidades, incluídas as unidades dialógicas. Não existiria, portanto, uma relação forte entre forma e função. [4] e [5] observam que os DCT são unidades longas, com uma intensidade média e um perfil que definem *modulado*. De fato, os exemplos apresentam diversas formas prosódicas, independentes do alinhamento: ascendentes, descendentes, niveladas e também com mais de um movimento.

Distribucionalmente, [4] e [5] notam que o DCT se encontra quase sempre em posição inicial de padrão. Em raríssimos casos ele pode ser encontrado em posição medial depois do que eles chamam *tópico pesado*, ou seja, um *tópico* iterado ou seguido por outras unidades textuais antes que seja realizada a ilocução. De fato, o exemplo indicado em [4] e os três exemplos identificados nos nossos dados mostram sempre a presença de parentético depois do tópico e antes da retomada com o DCT. A seguir um exemplo em PB, extraído de [25], com o padrão que contém o DCT evidenciado em negrito:

Exemplo 5– DCT em posição medial de padrão informacional – MED_19 [140]

*MAR: &he /=TMT= aí eu tive na casa da Patrícia /=COB= **a Eline** /=TOP= **tava na hora do almoço** /=PAR= **a filha dela iam chegar** /=PAR= **aí** /=DCT= **ela falou assim vão almoçar** /=COB_r= eu falei /=INT= não //COM_r=

IV. DADOS, METODOLOGIA E RESULTADOS

Analisamos um minicorpus de 31.464 palavras [24], extraído de [19] e etiquetado informacionalmente. [24] é balanceado por textos monológicos, dialógicos e conversacionais, e apresenta a máxima variação de falantes e situações comunicativas. Em [24], encontramos 129 unidades que poderiam ser consideradas DCT, com uma frequência compatível com os dados de [4] e [5] para o italiano. 56% delas seriam presentes em monólogos, os quais constituem menos de 1/3 do total de palavras do minicorpus. A presença dos DCT nos monólogos sobe para 66% se consideramos apenas aqueles que respeitam plenamente o critério de substituíbilidade.

Observamos que não é possível identificar uma forma prosódica para a unidade, nem mesmo se a definimos com base no princípio da eliminabilidade e/ou substituíbilidade. Prosodicamente, é possível dizer apenas que as unidades são (com pouquíssimas exceções) ou significativamente longas, com relação à média do COM (a duração foi calculada utilizando

o script SG_Detector para a normalização [27]), ou precedidos e/ou seguidos por pausa; e que sua intensidade não é muito inferior à média do comentário do padrão. Contudo, essas características são frequentes também nas unidades de escansão. A prosódia, portanto, não parece capaz de decidir na identificação da unidade. O alongamento e a frequência de pausa parecem aumentar em coincidência com a maior substituíbilidade do lexema.

Também observamos, em casos raros, que o DCT pode ser realizado por lexemas de classes diferentes, desde completamente dessemantizados. Foram certamente identificados DCT com os lexemas *praticamente* e *assim*.

A falta de uma forma prosódica dedicada para veicular uma específica função informacional constituiria uma exceção problemática para L-AcT, que postula que cada UI se define com base em função, forma prosódica e distribuição específicas. Também a definição funcional como conector genérico é vaga. A distribuição é quase sempre em início de padrão, mas não é impossível encontrar DCT em posição medial, em circunstâncias em que o falante parece ter perdido o programa planejado.

Diante destas observações, a pergunta não é apenas se a unidade chamada DCT deva ser considerado um MD, mas também se ela deve ser considerada uma verdadeira UI. De fato, diversos aspectos dessa unidade levam a pensar que se trate não de uma UI mas de um recurso para preparar o planejamento do padrão sucessivo em continuidade com o anterior, ou, em raros casos, a recuperar o programa depois de muitas unidades (em particular parentéticos) antes do núcleo ilocucionário do padrão. Esse planejamento naturalmente requer um certo tempo de elaboração e o falante pode ter uma ideia mais ou menos clara do tipo de relação entre o padrão anterior e o seguinte. Isso explicaria três características da unidade:

1. a duração longa e a frequente presença de pausa antes e/ou depois;
2. o fato que a substituíbilidade com lexemas que carregam tipos diferentes de relação seja gradual e não absoluta;
3. o fato que o preenchimento lexical seja quase sempre dado por uma conjunção, mas sem exclusão de outros lexemas dessemantizados.

Em conclusão, a nossa proposta é que o chamado DCT não deva ser considerado um MD, pois nem se trataria de uma UI, mas sim de um recurso que estabeleça uma ponte, mais ou menos vaga, entre padrões informacionais.

ACKNOWLEDGMENT

Os autores agradecem a FAPEMIG e o CNPq pelo suporte a essa pesquisa.

REFERENCES

- [1] Cresti E. *Corpus di Italiano parlato*. Firenze: Accademia della Crusca, 2000.
- [2] Moneglia M, Raso T. Notes on the Language into Act Theory. In: Raso T, Mello H (eds) *Spoken corpora and linguistics studies*. Amsterdam: John Benjamins, 2014, pp. 468–494.
- [3] Cavalcante F. *The information unit of topic: a crosslinguistic, statistical study based on spontaneous speech corpora*. Tese de Doutorado, Universidade Federal de Minas Gerais, 2020.
- [4] Cresti, Ee Moneglia, M. Il Connettore discorsivo secondo la Teoria della lingua in atto. Em: In A. De Meo & F. Dovetto (eds.), *Atti del Congresso GSCP “La comunicazione parlata”* Università degli Studi di Napoli

- "L'Orientale" - Università degli Studi di Napoli Federico II, Napoli: Aracne.2019.
- [5] Cresti, E.; Moneglia, M. The Discourse Connector according to the Language into Act Theory: data from IPIC Italiano. In: Bidese, E.; Casalicchio, J.; Moroni, M.C. (eds.). *La linguistica vista dalle Alpi Linguisticviews from the Alps*. Berlin: Peter Lang, 2020.
- [6] Austin J.L. *How to do things with words*. Oxford University Press, 1962.
- [7] Mittmann, M.M. O C-ORAL-BRASIL e o estudo da fala informal: um novo olhar sobre o tópico no português brasileiro. Tese de doutorado – Universidade Federal de Minas Gerais, Belo Horizonte, 2012.
- [8] Raso, T. Cavalcante, F. & Mittmann, M. Prosodic forms of the Topic information unit in a cross-linguistic perspective: a first survey. In: de Meo, A. & Dovetto, F.M. *Proceedings of the SLI-GSCP International Conference*. Rome: Aracne editrice, 2018.
- [9] Krifka, M. & Musan, R. (Eds) *The Expression of Information Structure*. De Gruyter Mouton, Berlin: 2012.
- [10] Cresti, E. The definition of focus in Language into Act Theory. Em: Mello, H., Panunzi, A., Raso, T (Eds.). *Pragmatics and Prosody. Illocution, Modality, Attitude, Information Patterning and Speech Annotation*. Firenze University Press, Firenze 2011, p. 39-82.
- [11] Tucci I. L'inciso: caratteristiche morfosintattiche e intonative in un corpus di riferimento. Em: *Atti del Convegno Nazionale*. Napoli: D'Auria, 2004, pp. 1–14.
- [12] Tucci I. Obiter dictum: La funzione informativa delle unità parentetiche. In: *Atti del GSCP*. Napoli: Università Orientale Press, 2009, pp. 635–654.
- [13] Maia Rocha B, Raso T. A unidade informacional de Introdutor Locutivo no português do Brasil: uma primeira descrição baseada em corpus. *Domínios de Linguagem*, 2011; 5: 327–343.
- [14] Fischer, K. (Ed.). *Approaches to discourse particles*. Oxford: Elsevier, 2006
- [15] Schiffrin, D. *Discourse Markers*. Cambridge University Press, 1987.
- [16] Bolden, G.B. *Discourse Markers*. In: Tracy, K., Sandel, T., Ilie, C. (eds.). *The International Encyclopedia of Language and Social Interaction*. Wiley, 2015. p. 1–7.
- [17] Frosali, F. Il lessico degli Ausili Dialogici. Em: Cresti, E. (Ed.). *Prospettive nello studio del lessico italiano: Atti del IX Congresso SILFI*. Firenze University Press, Firenze 2008, p. 417–424
- [18] Raso T. Prosodic constraints for discourse markers. In: Raso T, Mello H (eds) *Spoken Corpora and Linguistic Studies*. Amsterdam: John Benjamins 2014, pp. 411–467.
- [19] Raso T, Mello H. (eds). *C-ORAL-BRASIL: corpus de referência do português brasileiro falado informal*. Belo Horizonte: UFMG, 2012.
- [20] Cresti E, Moneglia M. *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam/Philadelphia: John Benjamins, 2005.
- [21] Raso T, Vieira M.A. Description of Dialogic Units/Discourse Markers in Spontaneous Speech Corpora Based on Phonetic Parameters. *CHIMERA* 2016; 3: 221–249
- [22] Raso, T., Ferrari, L. Uso dei Segnali Discorsivi in corpora di parlato spontaneo italiano e brasiliano. Em: Ferroni, R., Birello, M. (eds.). *La competenza discorsiva e interazionale: a lezione di lingua straniera*. Roma: Aracne, 2020. p. 61–107.
- [23] Raso T, Rilliard A, Santos S. Modeling the prosodic forms of Discourse Markers. Em: *Domínios de Linguagem*, no prelo.
- [24] www.c-oral-brasil.org> corpora > minicorpus português brasileiro 2018.
- [25] The C-ORAL-ESQ corpus. In preparation.
- [26] Sweetser, E. *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge: University Press 1990.
- [27] Barbosa P. Semi-automatic and automatic tools for generating prosodic descriptors for prosody research. In: Bigi B, Hirst D (eds) *Proceedings of the Speech Prosody conference*. Aix-en-Provence, 2013, pp. 86–90.