

## FRAMES E FALA ESPONTÂNEA<sup>1</sup>

HELIANA MELLO<sup>2</sup>  
TOMMASO RASO<sup>3</sup>

### RESUMO

Os estudos baseados em dados que tratam da noção de frame e do seu uso na análise da linguagem adotam unidades de análise que variam desde unidades lexicais até estruturas construcionais, sintáticas e textuais. Os mesmos princípios são utilizados em aplicações de processamento de linguagem natural. Neste artigo discutimos o porquê de se fazer necessária a inclusão das noções de unidade informacional e enunciado nas propostas de tratamento, via frames, da linguagem falada espontânea. A Teoria da Língua em Ato - TLA (CRESTI, 2000), que propõe o estudo da fala através da interface pragmática-prosódica e da sua segmentação em enunciados e unidades tonais, será brevemente apresentada, assim como sua aplicação e validação através dos estudos de corpora aplicados a línguas românicas europeias (C-ORAL-ROM) e ao português brasileiro (C-ORAL-BRASIL). Possíveis consequências para a incorporação do nível pragmático como a base analítica na constituição de frames para análise da fala espontânea serão discutidas.

**Palavras-chave:** Fala espontânea; português brasileiro; frames.

### ABSTRACT

Usage-based studies that deal with the concept of frame and its employment in language analysis take as their analytical object lexemes, constructions, sentences and texts. The same guiding principles are adopted in Natural Language Processing and other computationally-based studies. In this paper we discuss the necessary addition of utterance and tone unit as basic analytic constructs in the treatment of spontaneous speech via frames. In order to do so, we present the Language Into Act Theory - LAct (CRESTI, 2000) which proposes the study of speech through its pragmatic-prosodic interface through utterances and its tone units. The application and validation of the LAct is presented through exemplification from corpora-based studies undertaken by the C-ORAL-ROM (Romance European Languages) and C-ORAL-BRASIL (Brazilian Portuguese) projects. Some consequences of the incorporation of the pragmatic level of analysis into frame building will be discussed.

**Key-words:** Spontaneous speech; Brazilian Portuguese; frames.

## 1. O ESTADO DA ARTE: FRAMES E LÍNGUA FALADA

A noção de frame vem sendo utilizada nos estudos linguísticos e também naqueles voltados para a área de computação e inteligência artificial em sentidos

---

<sup>1</sup>. Os autores agradecem o apoio do CNPq e da FAPEMIG à pesquisa que subsidia este artigo.

<sup>2</sup>. UFMG, Belo Horizonte (MG), Brasil. heliana.mello@gmail.com

<sup>3</sup>. UFMG, Belo Horizonte (MG), Brasil. tommaso.raso@gmail.com

próximos (cf. Petruck, 1996). Neste artigo, a referência de frame adotada é aquela utilizada nos estudos da Semântica de Frames, como inicialmente desenvolvida por Fillmore (1977, 1982, 1985, 2008) e seus colaboradores e atualmente empregada nos diversos programas de pesquisa da rede FrameNet<sup>4</sup>. Assim, um frame, grosso modo, seria uma rede de elementos que, juntos, estruturam um domínio ou entidade conceptual.

A noção de frame se estende às construções, sentenças e estruturas textuais de nível mais alto (cf. Goldberg, 1995). Dessa forma, podemos inferir que a noção de frames, na verdade, busca uma explicação para o processamento da linguagem plenamente ancorada em estruturas conceptuais, que se cruzam, ou se projetam umas nas outras, ou se co-estruturam, desde um nível mais básico, lexical, por exemplo, até um nível de estruturação de ordem mais alta, como um texto.<sup>5</sup>

Toda a literatura fundadora da área da semântica de frames menciona a sentença e construtos sintáticos como as unidades de referência para a análise e determinação dos elementos constituintes de um frame (cf. Fillmore, 1977, 1985; Petruck, 1996). Mesmo os trabalhos que mencionam a contrapartida, ou interface, pragmática dos estudos semânticos (e sintáticos) que analisam a fala,<sup>6</sup> assumem tacitamente a sentença como a unidade fundamental de análise de onde se extrai os elementos que compõem um dado frame.

Entretanto, em nossos estudos da fala espontânea do português brasileiro, totalmente ancorados em metodologia empírica de análise de corpus, pudemos comprovar a precisão da proposta teórica de Cresti (2000) que instaura o enunciado como unidade de referência para a análise da fala. O enunciado é definido pela autora como “a unidade linguística mínima que permite uma interpretação pragmática” (Raso, 2012, p.77), o que distingue a sua proposta por assentá-la firmemente na interface prosódia-pragmática, deslocando a sintaxe e a semântica para um domínio muito mais restritivo, isto é, a unidade informacional. A proposta de Cresti (2000) amplamente testada e comprovada nos quatro corpora de línguas românicas europeias que compõem o C-ORAL-ROM (Cresti e Moneglia, 2005), é verificada com sucesso também no C-ORAL-BRASIL, corpus de fala espontânea do português brasileiro (Raso e Mello, 2012).

Partindo da premissa de que a fala é organizada em enunciados, que por sua vez, são estruturados através de unidades informacionais, passamos a discutir a organização desses níveis analíticos e as suas consequências para a utilização de frames na compreensão da fala.

---

<sup>4</sup> <https://framenet.icsi.berkeley.edu>

<sup>5</sup> Naturalmente, o estudo de um nível formal mais básico, como o fonológico, também pode, em princípio, ser estudado através da mesma lógica. Estudos seguindo uma tradição distinta, porém compatível com as propostas baseadas em frames e conceptualização são desenvolvidos, por exemplo, por Joan Bybee (cf. BYBEE, 2001, 2007, 2010).

<sup>6</sup> Por exemplo, toda a literatura que trata do processamento semântico da linguagem natural via frames (cf. Tur e De Mori, 2011).

## 2. FRAMES E ENUNCIADOS: a organização da fala espontânea

A organização da fala espontânea distingue-se em sua natureza da produção escrita. Esse fato, apesar de conhecido, não tem as repercussões teóricas que se esperaria, dada a sua importância, na maior parte dos quadros teóricos da linguística contemporânea. A fala na maior parte das vezes é tratada como texto escrito, uma vez que arquivos de som raramente são levados em consideração nas análises disponíveis. Ao contrário, as propostas analíticas baseiam-se nas transcrições da fala, que por sua vez, são segmentadas através de pausas que correspondem, grosso modo, a pausas sintáticas, com uma inerente correspondência às marcas de pontuação da produção linguística escrita.

Dentre as teorias linguísticas contemporâneas que tentam dar conta das particularidades da fala, aquela que se estruturou a partir da observação empírica ao longo de quase quarenta anos de estudo de *corpora* de fala espontânea, observando-se as regularidades que dali emergiam e que permitiram generalizações que conduziram a um quadro teórico formalizado foi a Teoria da Língua em Ato, proposta por Cresti (2000)<sup>7</sup>. Esta se trata, então, de uma teoria induzida por *corpora*, ou *corpus driven*, o que lhe confere uma sólida possibilidade de verificação empírica.

A TLA se estrutura como uma forma de análise pragmática da fala, isto é, essa teoria estabelece que há níveis hierárquicos definidos na expressão comunicativa da fala que devem ser considerados; a sua não-consideração levaria à impossibilidade de se conseguir, efetivamente, analisar os textos falados. Logo, esse tipo de análise pressupõe: i. que inicialmente se individualize a unidade ilocucionária; ii. que dentro dessa unidade seja estabelecida a sua estrutura informacional; e iii. que somente dentro da unidade informacional seja possível uma análise sintática.

Como dito acima, a TLA designa a unidade de referência da fala à dimensão do enunciado. A necessidade de se identificarem unidades de referência fica clara se analisarmos um trecho aleatório de uma transcrição de um texto falado. Vejamos o exemplo 1.<sup>8</sup>

(1) C-ORAL-BRASIL I (*bfammn02*)

(1.a) \*DFL: *e então tinha muito texto do tio Carlos então ele falava ah ele é tio da minha tia*

A fim de se interpretar essa curta sequência, precisaríamos fazer algumas escolhas de segmentação. Uma possibilidade intuitivamente aceitável para essa segmentação seria:

<sup>7</sup> Para boa parte da produção dentro deste paradigma teórico, vejam-se os sites <http://lablita.di.unifi.it> e [www.c-oral-brasil.org](http://www.c-oral-brasil.org).

<sup>8</sup> Todos os exemplos utilizados neste artigo foram retirados do C-ORAL-BRASIL (cf. Raso e Mello, 2012; Raso, 2012).

(1.b) \*DFL: *e então tinha muito texto do tio Carlos // então ele falava / ah / ele é tio da minha tia //*

Outra possibilidade poderia ser:

(1.c) \*DFL: *e então tinha muito texto do tio Carlos // então ele falava / ah / ele // é tio da minha tia //*

A barra dupla representa a quebra prosódica terminal, que sinaliza a conclusão de um enunciado, e a barra simples representa a quebra prosódica não terminal, que sinaliza a conclusão de uma unidade tonal dentro de um enunciado. Conduzindo a segmentação como mostrado em (1.b) e (1.c) acima, identificam-se dois ou três enunciados, ou seja, duas ou três unidades principais em que a sequência do exemplo (1.a) pode ser segmentada. Ambas as segmentações parecem aceitáveis, já que colocam a fronteira de enunciado depois do fim de uma sentença e antes de uma outra sentença<sup>9</sup>. Contudo, se escutarmos o áudio do exemplo, percebemos claramente que se trata de um único enunciado:<sup>10</sup>

(1.d) \*DFL: *e então / tinha muito texto do tio Carlos / então ele falava / ah / ele é tio da minha tia //*

Note-se que, ao levarmos o áudio em consideração, não só propusemos uma segmentação do fluxo textual diferente, mas alteramos, em consequência disso, também outros aspectos da interpretação da sequência.

Um outro exemplo da relevância de se conjugar som e transcrição no processamento do fluxo do texto falado pode ser visto em (2):

(2) C-ORAL-BRASIL I (*bpubdl01*)

(2.a) \*PAU: *não tá dando a altura daquele que a Isa marcou lá né*

Se partirmos apenas da transcrição, poderíamos pensar que (2.a) se trata de um único enunciado negativo, segmentável da seguinte maneira:

(2.b) \*PAU: *não tá dando a altura daquele que a Isa marcou lá / né //*

Mas, ao escutarmos o áudio não há dúvida de que a segmentação correta deva prever dois enunciados:

(2.c) \*PAU: *não // tá dando a altura daquele que a Isa marcou lá / né //*

As sequências textuais observadas em (3) e (4), abaixo, permitem uma análise sintática paralela: uma oração principal e uma dependente relativa. Mas,

<sup>9</sup> Para a diferença entre os conceitos de enunciado e de frase, veja-se Cresti (2005).

<sup>10</sup> Todos os arquivos de som, assim como os arquivos XML de alinhamento som-transcrição, referidos neste artigo estão disponíveis no DVD *dual layer* que acompanha Raso e Mello (2012).

ao se ouvir os seus arquivos de som, percebe-se que (3) e (4) têm estruturas distintas. Levando-se em consideração a informação prosódica, percebemos que os dois exemplos devem ser segmentados da seguinte forma: em (3) temos dois enunciados, ou seja, o falante gerou duas estruturas autônomas; em (4) temos um único enunciado, em uma única unidade tonal:

(3) C-ORAL-BRASIL I (*bfamdl02*)

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

(4) C-ORAL-BRASIL I (*bfamdl02*)

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

Os exemplos (5), (6) e (7) também possuem a mesma estrutura: um sintagma verbal principal com seus argumentos e um adjunto que modifica o último argumento. Mas enquanto no primeiro caso o sintagma que supostamente modificaria o argumento é entonacionalmente independente, nos outros dois casos ele se coloca na mesma unidade tonal do resto do enunciado:

(5) C-ORAL-BRASIL I (*bfamcv01*)

\*EVN: *eu te mandei uma lista // com uns dez lugares diferentes //*

(6) C-ORAL-BRASIL I (*bpubcv01*)

\*FLA: *a gente não sabe nada sobre elas //*

(7) C-ORAL-BRASIL I (*bpubcv01*)

\*FLA: *a gente é responsável por armazenar //*

Naturalmente, em outros contextos, seria possível encontrar os exemplos (3-7) realizados entonacionalmente de maneira diferente: (3) e (5) poderiam ser realizados em uma única unidade tonal, enquanto (4), (6) e (7) poderiam ser realizados em duas unidades tonais. Uma teoria capaz de analisar a fala deve dar conta dessas diferenças e explicar a função delas. Isso não é possível através de uma análise exclusivamente sintática ou semântica, mesmo se baseada em frames.

Esses sete exemplos, que poderiam facilmente ser multiplicados, nos dizem algumas coisas fundamentais:

1. que nós segmentamos naturalmente a fala com base em unidades que não podemos reconhecer a partir apenas da transcrição gráfica;
2. que a interpretabilidade da fala é determinada por essa segmentação, a ponto de o que parece ser negado na transcrição poder ter sido, na realidade, afirmado;
3. que as unidades que são consideradas completas na escrita, como uma sentença, podem na realidade constituir apenas um pedaço de um enunciado, o qual, em outros casos, pode ao contrário, ser constituído

- apenas por um advérbio de negação; inclusive, seria fácil achar exemplos de enunciados constituídos apenas por uma interjeição;
4. que o que determina a segmentação e a individualização de um enunciado não é a sua estrutura morfossintática, mas a sua forma prosódica, que nos diz onde e como segmentar a fala;
  5. que o que a prosódia nos diz é qual segmento da fala pode ser interpretado em autonomia; a autonomia prosódica reflete uma autonomia pragmática, e não semântica ou morfossintática;
  6. que se consideramos a informação prosódica, o que parece na transcrição como a mesma estrutura sintática pode, de fato, ser analisado em um número diferente de unidades e em um número diferente de estruturas autônomas; nesse caso, a análise sintática não pode se estender para fora das fronteiras da unidade que se estrutura autonomamente.

Mas por que um segmento de fala é percebido como algo que pode estar sozinho e constituir, portanto, um enunciado, ou seja, a menor unidade interpretável pragmaticamente no fluxo da fala, enquanto outros segmentos são percebidos como apenas parte de uma unidade autônoma, independente de razões de ordem morfossintática? A resposta da TLA a essa pergunta é que o enunciado corresponde ao ato de fala (Austin, 1962), ou melhor, constitui o ato locutivo, a contraparte linguística, da ação, da ilocução cumprida ao proferir o enunciado. Nos exemplos 1, 4, 6 e 7 nós identificamos, portanto, apenas uma ilocução, mesmo se organizada de maneira complexa; nos exemplos 2, 3 e 5, ao contrário, identificamos duas ilocuições, mesmo que mais simples do ponto de vista da estrutura morfossintática.

Através da fala nós, naturalmente, realizamos constantemente ilocuições muito diferentes, dependendo das ações que precisamos cumprir com relação ao interlocutor. Essas ilocuições não somente são segmentáveis com base na prosódia, mas são reconhecíveis e distinguíveis como ilocuições diferentes a partir da forma prosódica com a qual são executadas, não dependendo da estrutura morfossintática ou de seu conteúdo semântico. O exemplo 8 mostra isso com clareza extrema:

(8) C-ORALBRASIL I(*bfamd104*)

\*KAT: *o quê //*

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

\*KAT: *copos de quê //*

\*SIL: *Urano //*

\*KAT: *Urano //*

\*SIL: *é // Urano // Urano //*

Na pequena interação do exemplo 8, temos 9 enunciados, todos muito simples. Apenas um deles é constituído por mais de uma única unidade tonal. Quatro deles são absolutamente idênticos do ponto de vista morfossintático e semântico (*Urano*). Contudo, eles não são interpretados como tendo o mesmo sentido, o que geraria uma mera tautologia sem valor comunicativo. A diferença de sentido é devida ao fato de que os enunciados diferem sob um único aspecto:

o conteúdo ilocucionário que veiculam. O primeiro *Urano* constitui uma confirmação, o segundo uma expressão de incredulidade, enquanto os dois últimos são conclusões, realizadas com atitudes diferentes<sup>11</sup>. E o significativo que carrega os sentidos diferentes é a prosódia, como mostram também as diferentes formas prosódicas das três ilocuções aqui percebidas:

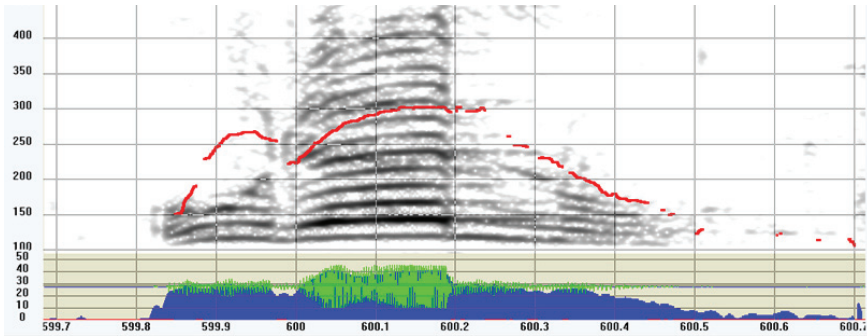


FIGURA 1. *Urano*. Ilocução de confirmação.

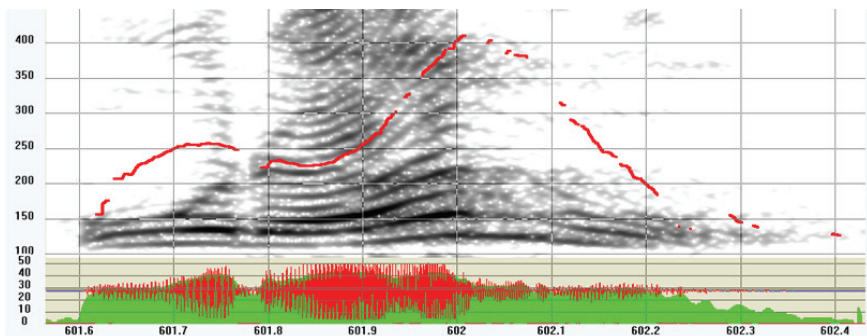


FIGURA 2. *Urano*. Ilocução de expressão de incredulidade.

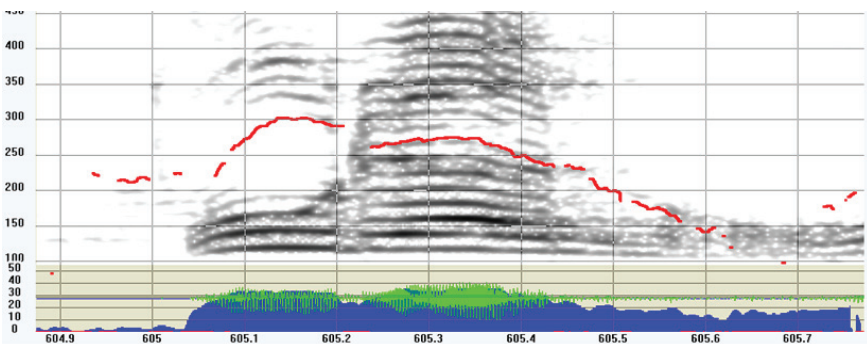
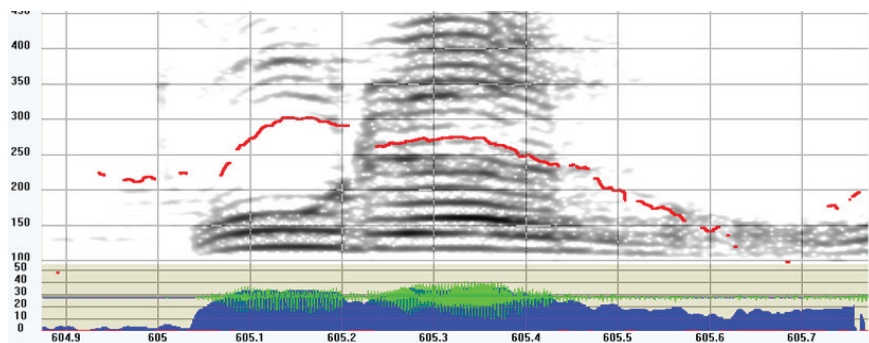


FIGURA 3. *Urano*. Ilocução de conclusão.

<sup>11</sup> Para a diferença entre atitude, ilocução e modalidade, e para uma proposta dos traços pertinentes às duas categorias, veja-se Mello e Raso (2012).

FIGURA 4. *Urano*. Ilocução de conclusão.

Note-se que as formas das figuras 3 e 4 são muito parecidas. A curva da figura 4 parece reproduzir em escala menor e com menor intensidade a curva da figura 3. O que muda parece envolver mais o nível do estado de espírito do falante e não o sentido, a ação que ele está realizando. Se essa análise for correta, não muda a ilocução, mas a atitude com a qual a ilocução é realizada. De fato, parece que nós podemos cumprir a mesma ação, por exemplo, uma ordem, com atitudes diferentes; isso é, podemos realizar uma ordem de maneira tranquila ou de maneira irritada, mas em ambos os casos trata-se de uma ordem. Como tanto a ilocução quanto a atitude são valores carregados pela prosódia, não é uma tarefa fácil distinguir entre os traços prosódicos pertinentes à análise da ilocução e aqueles pertinentes à análise da atitude (cf. Mello e Raso, 2012).

Mas o que interessa aqui é demonstrar que enunciados com o mesmo conteúdo locutivo (ou seja, idênticos do ponto de vista semântico e morfossintático) podem ter sentidos muito diferentes porque veiculam, através da prosódia, ilocuições diferentes. Naturalmente, podemos também realizar a mesma ilocução com conteúdos locutivos diferentes. O caso mais trivial é o chamamento: dependendo do nome da pessoa que estamos chamando, o conteúdo locutivo muda, mas isso não muda de forma alguma a ilocução e a forma prosódica de seu núcleo. Seguindo esse mesmo raciocínio, o exemplo 8 demonstra também que, para que a língua seja usada comunicativamente, a ilocução é mais importante para a decisão do sentido do que a estrutura morfossintática e o conteúdo semântico. De fato, podemos ter sentidos diferentes sem mudar nada do conteúdo locutivo, mas não podemos de maneira alguma ter um sentido comunicativo sem a expressão da ilocução. Se não conseguimos individualizar a ilocução veiculada, não podemos atribuir valor comunicativo a uma sequência verbal. E, como a ilocução é veiculada necessariamente pela prosódia, é por isso que a simples transcrição da fala não permite a sua segmentação e, conseqüentemente, nem o seu processamento computacional ótimo. Assim, a análise da linguagem pautada por frames, necessariamente deve levar em conta pelo menos dois níveis adicionais àqueles construcional e semântico, quais sejam, os níveis informacional e ilocucionário.



## CONCLUSÃO

Concluindo, podemos dizer que a prosódia constitui a interface entre o ato locutivo e o ato ilocutivo que cumprem, simultaneamente, o ato de fala. Através da prosódia segmentamos em unidades de referência o fluxo da fala (enunciados e unidades tonais), e essa segmentação é fundamental para individualizar as unidades que realizam a função comunicativa da língua falada, ou seja, as ilocuções. Através da prosódia atribuímos também a essas unidades um determinado valor comunicativo, ou seja, decidimos qual ação elas realizam, independentemente de sua forma semântica ou morfossintática. Dessa forma, dentro do escopo dos estudos lingüísticos baseados em análises via frames faz-se necessária a incorporação dos níveis informacional e ilocucionário à rede de níveis organizacionais de frames voltados para a língua falada. Sem esses dois níveis, o que se estuda, na verdade, é a escrita que tenta emular a fala sem, contudo, alcançar sucesso ao fazê-lo.

---

## REFERÊNCIAS BIBLIOGRÁFICAS

- AUSTIN, L. J. (1962). *How to do things with words*. Oxford, Oxford University Press.
- BYBEE, Joan. (2001). *Phonology and language use*. Cambridge, Cambridge University Press.
- \_\_\_\_\_. (2007). *Frequency of use and the organization of language*. Oxford, Oxford University Press.
- \_\_\_\_\_. (2010). *Language, usage and cognition*. Cambridge, Cambridge University Press.
- CRESTI, Emanuela. (2000). *Corpus di Italiano parlato*. Firenze, Accademia della Crusca, v. 1 e 2.
- \_\_\_\_\_. (2005). Enunciato e frase: teoria e verifique empiriche. In: BIFFI, M.; CALABRESE, O.; SALIBRA, L. (Orgs.). *Italia linguistica: discorsi di scritto e di parlato – nuovi studi di linguistica italiana per Giovanni Nebcioni*. Siena, Protagon, p.249-260.
- CRESTI, Emanuela e MONEGLIA, Massimo (Orgs.). (2005). *C-ORAL-ROM: integrated reference corpora for Romance languages*. Amsterdam, John Benjamins.
- FILLMORE, Charles J. (1977). Scenes-and-frames semantics. *Linguistic Structures Processing*. In: ZAMPOLLI, A. (Org.). *Fundamental Studies in Computer Science*. Amsterdam, North Holland Publishing, p. 55-82.
- \_\_\_\_\_. (1982). Frame semantics. *Linguistics in the Morning Calm*. Seul, Hanshin Publishing Co., p. 111-137.
- \_\_\_\_\_. (1985). Frames and the semantics of understanding. *Quaderni di Semantica*, 6, p. 222-254.
- \_\_\_\_\_. (2008). Border Conflicts: FrameNet Meets Construction Grammar. *Proceedings of the XIII {EURALEX} International Congress*. In: BERNAL, E. e DECESARIS, J. (Orgs.). Barcelona, Universitat Pompeu Fabra, p.49-68.
- GOLDBERG, Adele. (1995). *Constructions: a Construction Grammar approach to argument structure*. Chicago, University of Chicago Press.
- MELLO, Heliana e RASO, Tommaso. (2012). Illocution, Modality, Attitude: different names for different categories. In: MELLO, H., PANUNZI, A., RASO, T. (Orgs.). *Pragmatics and Prosody: Illocution, Modality, Attitude, Information Patterning and Speech Annotation*. Firenze, Firenze University Press, p. 1-18.
- PETRUCK, Miriam R. L. (1996). Frame Semantics. In: VERSCHUEREN, Jeff. , et al. (Orgs.). *Handbook of Pragmatics*. Amsterdam, John Benjamins.

MELLO e RASO – Frames e a fala espontânea

RASO, Tommaso. (2012). O C-ORAL-BRASIL e a Teoria da Língua em Ato. In: RASO, T. e MELLO, H. (Orgs.). *C-ORAL-BRASIL I: Corpus de referência do português brasileiro falado informal*. Belo Horizonte, Editora UFMG.

\_\_\_\_\_. (Orgs.). (2012). *C-ORAL-BRASIL I: Corpus de referência do português brasileiro falado informal*. Belo Horizonte, Editora UFMG.

TUR, Gokhan e DE MORI, Renato (Orgs.). (2011). *Spoken Language Understanding : systems for extracting semantic information from speech*. Oxford, Wiley-Blackwell.