

# LINGUÍSTICA DE TEXTO: o que é e como se faz?

Luiz Antônio Marcuschii

**T**  
Parápo

LUIZ ANTÔNIO MARCUSCHI

LINGUÍSTICA DE TEXTO: o que é e como se faz?

Luiz Antônio Marcuschii é formado em filosofia pela Pontifícia Universidade Católica do Rio Grande do Sul (1968), doutorado em Letras pela Universität Erlangen-Nürnberg (Friedrich-Alexander) (1976) e pós-doutorado pela Albert-Ludwigs-Universität Freiburg (1988). É professor titular da Universidade Federal de Pernambuco. Tem experiência na área de linguística, com ênfase em teoria e análise linguística. Atua principalmente nos seguintes temas: filosofia da linguagem, metodologia, epistemologia, lógica. É autor das seguintes obras: *Análise da conversação* (2007), *Cognição, linguagem e práticas interacionais* (2007), *Da fala para a escrita* (2010), *Fenômenos da linguagem* (2007), *Produção textual, análise de gêneros e compreensão* (2008), *Linguística de texto: o que é e como se faz?* (1988, 2007, 2012).



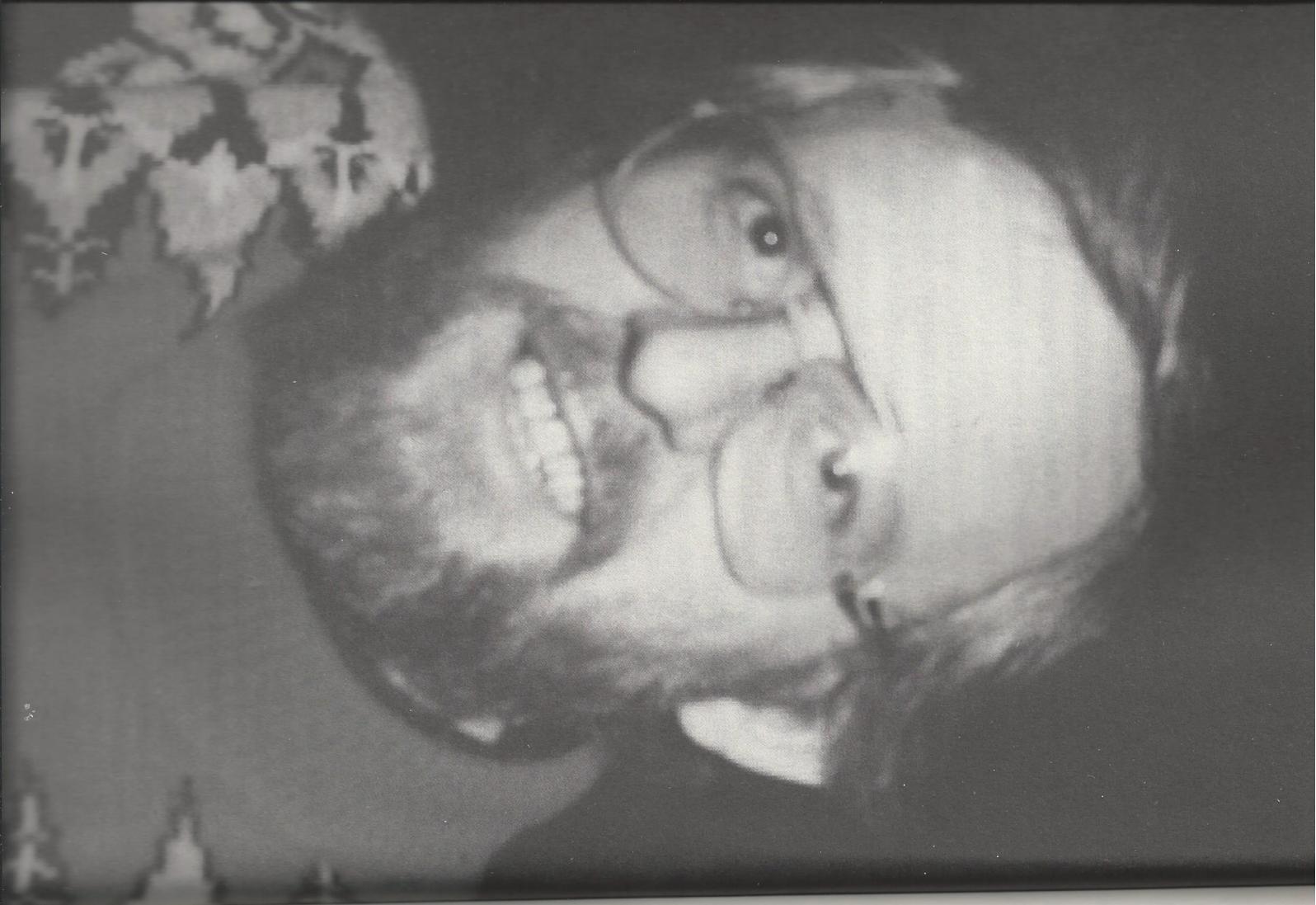
M.K.

## ALGUMAS DEFINIÇÕES DE TEXTO

... se os fatos de linguagem têm caráter social, o processo/produto pelo qual se dá a interação entre os interlocutores é o texto. O texto está, pois, ligado a uma situação material concreta, como igualmente a um contexto mais amplo, que configura as condições de vida de uma dada comunidade linguística.

HELENA HATHSUE NAGAMINE BRANDÃO

Todos nós sabemos, intuitivamente, distinguir entre um texto e um não texto. Também sabemos que a produção linguística geralmente se dá em textos e não em palavras isoladas. Ninguém se porá a ler um dicionário ou um catálogo telefônico assim como lê um roman-



ce, um artigo de jornal ou uma carta dum amigo. Mas, apesar desta noção intuitiva de texto, não saberíamos definir intuitivamente o que é que faz de uma sequência linguística um texto. Aliás, entre os próprios estudiosos do assunto ainda não se chegou a uma definição geralmente aceita.

A rigor, teríamos pelo menos duas alternativas básicas para definir o texto:

- (a) partindo de critérios internos ao texto (olhando-o do ponto de vista imanente ao sistema linguístico) e
- (b) partindo de critérios temáticos ou transcendentes ao sistema (considerando o texto como uma unidade de uso ou unidade comunicativa).

Vejam algumas das opções que surgem dessas duas posições:

## Definições de texto na imanência do sistema linguístico

Do ponto de vista da imanência ao sistema linguístico, o texto foi definido, de uma maneira geral, como "uma sequência coerente de sentenças. Nesta definição teríamos três termos técnicos a serem especificados: (a) "sequência", (b) "sentença" e (c) "coerência". Sequência é uma expressão que aponta para a necessidade de haver um conjunto linear, mas é problemática se indicar uma condição necessária, pois há textos que se compõem de uma só sentença, ou mesmo de uma só palavra, por ex.: "Fogo!", quando dito numa situação específica e permanecendo a única expressão na ocorrência. O termo *sentença*, sobejamente discutido em todas as gramáticas de frase, é deixado de lado pelos linguistas de texto que têm dele uma noção intuitiva e tácita. Veja-se, contudo, o caso de "Fogo!": é uma sentença ou não? Para tal decisão deve-se considerar tanto aspectos morfológicos como sintéticos, mas a distinção entre "estrutura de superfície" e "estrutura

profunda", por exemplo, traria outras perspectivas. Veja-se novamente o caso do texto (11), de Ricardo Ramos, que no primeiro capítulo só tem sequência de palavras; no segundo já são sintagmas e no terceiro aparecem frases e expressões.

Com a *coerência*, entra-se no aspecto da natureza fundamental da sequência e da relação entre as sentenças. Quanto a isso veremos que a noção de coerência envolveu aspectos demasiados e em vários níveis, resultando numa categoria difusa e pouco aproveitável. As teorias que trabalham ao nível do sistema linguístico procuram montar *gramáticas de texto*, fazendo predominar os aspectos sintáticos do texto, deixando o nível cognitivo-conceitual e o pragmático. Alguns dos autores enquadráveis nesta tendência seriam os seguintes (ressalvo que esta classificação contém simplificações imperdoáveis):

- (a) ZELUG S. HARRIS: "Um texto (discurso) compõe-se de uma sequência de expressões ou sentenças ligadas, podendo ir desde sentenças de uma só palavra até uma obra em vários volumes".

Z. Harris não dá uma definição explícita de texto, mas supõe essa que aí está. Parte de uma noção intuitiva de texto como sequência de morfemas ou sentenças ligados de alguma forma num todo. Com isto pode admitir como suficiente a gramática de frase já existente, sendo que a análise de um texto não passaria, uma vez que não nos comunicamos por frases ou morfemas isolados, mas por textos. Apesar da importância do estudo distribucionalista de Harris para a LT, ele ficou sem ressonância dentro da LT posterior. Na verdade, as classes de equivalência de Harris aplicam-se a qualquer sequência arbitrária de sentenças e não necessariamente entram como condição da estrutura textual, como observa van Dijk [cf. T. A. van Dijk (1977, pp. 270 e ss.]. Harris é talvez o mais radical dos imanentistas.

- (b) ROLAND HARWEG: "Texto é uma sucessão de unidades linguísticas constituída por uma cadeia pronominal ininterrupta" (1968, p. 148.).

Nesta definição do linguista alemão Harweg, há dois aspectos fundamentais: (1) a sucessão sintagmática (pela qual as palavras formam sentenças e sentenças formam textos) e (2) a sucessão paradigâmica (também chamada de substituição sintagmática, pela qual se realiza a cadeia pronominal dentro da sequência). Nessa concepção, uma das características fundamentais constitutivas do texto é o fenômeno do *múltiplo referenciamento*<sup>1</sup>, ou seja, os mesmos objetos, lugares, pessoas, grandezas etc., são várias vezes retomados e referidos num texto, nem sempre com as mesmas expressões. Isso é o que dá ao texto a característica de cadeia pronominal, pois as retomadas são vistas como formas de substituição pronominal (tomando-se aqui o termo "pronominal" no sentido de pró-forma, como ainda veremos).

Uma noção muito importante desenvolvida por Harweg é a distinção entre *texto ênico* e *texto ético*, termos cunhados a partir da distinção do tipo fonético e fonêmico, já proposta por Pike. O *texto ênico* é aquele que se realiza na sua relação de imanência ao sistema do texto em si (que Petöfi caracterizará como sendo o nível da *cotextualidade*); o *texto ético* é o que se realiza situativamente e se define na contextualidade, envolvendo por exemplo o título, o nome do autor, a data, o local e outros elementos. Posteriormente, Harweg mudará sensivelmente sua posição de modo a ampliar esta visão<sup>2</sup>.

(c) IRENA BELLERT: "Um texto é uma sequência de sentenças  $S_1, S_2, \dots, S_n$  de tal modo que a interpretação semântica de cada sentença  $S_i$  (para  $2 \geq i \geq n$ ) depende da interpretação da sequência  $S_1 \dots S_{i-1}$ " [Cf. I. Bellert (1970, p. 335)].

Em tese, I. Bellert baseia-se na noção de substituição como condição necessária (mas não suficiente) para a coerência de um texto. Faz da coerência o princípio constitutivo fundamental do texto, transfor-

mando o conhecimento do contexto numa condição geral necessária para a interpretação adequada. Apesar de introduzir aí o conhecimento do mundo como parte da textualidade, I. Bellert mantém uma posição imanente à estrutura da língua, pois a contextualidade é introduzida apenas para permitir a montagem de um sistema inferencial na compreensão. Poderíamos dizer que esta definição determina a estrutura profunda do texto por recursos apenas lógico-semânticos e não atinge a textualidade como tal, porquanto faz o texto ser uma sequência de sentenças adicionadas umas às outras.

(d) HARALD WEINRICH: "Texto é uma sequência ordenada de signos lingüísticos entre duas interrupções comunicativas importantes" [cf. H. Weinrich (1976, p. 186-187)].

H. Weinrich evita os termos "sentença" e "coerência", mas dificulta a delimitação de texto, ao deixar pouco claro o que seja uma "interrupção comunicativa marcante". Caso fôssemos bastante elásticos, poderíamos chegar a situação de que uma pessoa realizaria um texto só durante sua vida, o que seria um absurdo, embora já tenha sido proposto. Em seu estudo programático sobre a partitura textual como método heurístico<sup>3</sup>, Weinrich propõe como método da análise de texto uma teoria que visa a unir a análise frasal por tipo de palavra e estrutura de frase num só modelo tal como uma "partitura musical a duas vozes". Este método é apresentado como alternativa para as conhecidas análises de frase existentes. Mais do que tudo, temos aí uma inovação metodológica na análise de texto, mas ao nível do sistema lingüístico. Como para Weinrich toda lingüística é lingüística de texto, todo o tratamento de categorias gramaticais é um tratamento textual dos artigos, dos pronomes, dos verbos e todos os fenômenos gramaticais<sup>4</sup>.

<sup>1</sup> R. Harweg (1974, pp. 99-101) e também R. Harweg (1978).

<sup>2</sup> Refiro-me aqui aos estudos posteriores ao de (1968), onde se encontra a posição inicial padrão de Harweg. Para uma visão clara e minuciosa das posições de Harweg, cf. E. Güllich e W. Raible (1977, pp. 115-127).

<sup>3</sup> Para tanto, remeto ao estudo de H. Weinrich (1976, pp. 145-162), intitulado *Die Textpartitur als heuristische Methode*, onde o autor monta seu método de análise de texto partindo de um sistema semelhante a uma partitura musical e tomando como foco organizador do texto o verbo, seguindo de perto um sistema como o da sintaxe de Tesnière.

<sup>4</sup> Cf. os estudos de H. Weinrich (1976), especialmente estudos sobre os artigos em alemão (pp. 163-176) e sobre os artigos em francês (pp. 177-185) e (pp. 186-198).

## Definições de texto com critérios temáticos e transcendententes ao texto

As definições de texto que se propõem critérios mais amplos que os puramente linguísticos, tomam-no como uma *unidade comunicativa* e não como uma simples unidade linguística. Se no caso das definições anteriores tínhamos a ver com textos definidos *emicamente*, na terminologia de Harweg, aqui temos definições *éticas* de textos, ou seja, que consideram o arranjo de sentenças em seu funcionamento mais amplo no processo de comunicação e não apenas no âmbito da estrutura linguística. Neste caso, a *LT* é a descrição da correlação entre a produção, a constituição e a recepção de textos.

- (a) JANOS S. PETÖFI: "Uma sequência de elementos linguísticos escritos ou falados organizada como um todo, com base em algum critério qualquer (geralmente extralinguístico), resultando num texto"<sup>5</sup>.

Nesta definição de Petöfi encontramos uma posição extremada de pretensões teóricas profundamente ambiciosas, em que dois tipos de elementos estão implicados: *contextuais* (externos ao texto) e *cotextuais* (internos ao texto), fenômenos estes bem definidos por Petöfi<sup>6</sup>. A estrutura do texto corresponderia à estrutura do mundo, devendo obedecer a uma semântica contextual (intencional e extensionalmente visto) da dimensão do mundo e a uma gramática especial não linearmente fixada, mas de características gerativo-transformacionais. A definição acima é pouco operacional e o próprio Petöfi e seu grupo não dispõem ainda de bases para a realização de uma tal *LT*.

<sup>5</sup> Cf. J. Petöfi (1972, p. 31). Para outros detalhes, cf. também de Petöfi (1971); (1978).  
<sup>6</sup> Veja-se a definição destes termos proposta por Petöfi (1976) (p. 115): "Cotextualidade quer dizer, com respeito à 'gramática', que ela opera na análise apenas com suas próprias informações internas... e informações que podem ser obtidas do texto a ser analisado". "Contextualidade quer dizer, com respeito à semântica, que a informação referida acima fornece apenas as condições necessárias e suficientes que serão fornecidas extratextualmente."

- (b) TEUN VAN DIJK: O texto "é uma estrutura superficial governada por uma estrutura semântica profunda motivada", ou seja, "um conjunto ordenado de sentenças da estrutura profunda" (cf. van Dijk, 1978 e 1977).

Esta definição de van Dijk, um dos mais produtivos linguistas de texto, tenta aplicar duas noções básicas da gramática gerativa para explicar o processo de geração de sentidos textuais e estruturas textuais: "estrutura profunda" e "estrutura superficial". Para van Dijk, as relações textuais deveriam ser descritas com base no modelo das relações lógico-semânticas estabelecidas na estrutura profunda. São estas estruturas que possibilitam a coerência do texto, permitem que se consiga resumir um texto, que se memorizem conteúdos de textos longos e que se escrevam textos superficialmente diversos com o mesmo conteúdo. O texto torna-se a unidade linguística por excelência, pois para van Dijk é por textos e não por sentenças que nos comunicamos<sup>7</sup>.

- (c) SIGFRIED SCHMIDT: "Texto é qualquer expressão de um conjunto linguístico num ato de comunicação (no âmbito de um jogo-de-ação comunicativo), sendo tematicamente orientado e preenchendo uma função comunicativa reconhecível, ou seja, realizando um potencial ilocutivo reconhecível"<sup>8</sup>.

Para Schmidt, o texto define-se como uma unidade comunicativa tematicamente orientada que realiza um potencial ilocutivo na ação comunicativa. Com isso evita a noção de frase e de coerência, introduzindo como essencial o elemento pragmático. A coerência não é determinada lógico-sintaticamente e sim tematicamente, ou seja, na estrutura comunicativa profunda. Adotando a posição da semântica

<sup>7</sup> Cf. van Dijk (1978, p. 269). De resto, quando van Dijk mostra que uma sequência de sentenças ainda não é um texto, e, portanto, uma gramática de frase não é suficiente para uma explicação geral do texto (1972) e (1980, p. 22ss.), defende que descrever sentenças compostas ou complexas ainda não é descrever sequências de sentenças que formam um texto. Volta-se, neste caso, contra as críticas de Dascal e Margalit (1974) e (1974a).  
<sup>8</sup> Cf. S. J. Schmidt (1973, p. 237), e também Schmidt (1974, p. 45). Para uma visão mais ampla da teoria de Schmidt, cf. o livro traduzido para o português Schmidt (1978).

interpretativa (da posição inicial da GGT), Schmidt toma as palavras e as sentenças com *instruções* semânticas, fazendo com que a coerência não seja um fenômeno da superfície. Neste caso, a *textualidade* seria o modo de toda e qualquer comunicação transmitida por sinais, inclusive os linguísticos [cf. Schmidt (1974, p. 43 ss.)]. É esta concepção geral que faz com que Schmidt evite inclusive chamar de "linguística de texto" o seu tipo de estudo textual, preferindo a expressão "teoria do texto" [cf. Schmidt (1973, p. 233) e (1978) *passim*].

(d) M. A. K. Halliday e R. Hasan: "Um texto é uma unidade em uso. NÃO é uma unidade gramatical, tal como uma frase ou uma sentença; e não é definido por sua extensão. (...) Um texto é, melhor dizendo, uma unidade semântica: não uma unidade de forma e sim de sentido"<sup>9</sup>.

Para Halliday e Hasan o texto não consiste em sentenças; ele apenas se *realiza* nas sentenças, de modo que as partes do texto não se integram como as partes de uma sentença se unem entre si. A unidade de texto é de outra natureza que a da sentença. É a *textura* que distingue um texto de um não texto, sendo que a *textura* é formada pela relação semântica de *coesão*. Para Halliday e Hasan, a *coesão* não é uma relação sintática e sim semântica, determinada pela interpretação e pela pressuposição. O texto passa a ser uma unidade semântica e não gramatical.

## Texto como processo de mapeamento cognitivo

Das diversas definições aqui revistas fica um *desideratum*: o texto não é uma unidade virtual e sim concreta e atual; não é uma simples

<sup>9</sup> Cf. M. A. K. Halliday e R. Hasan (1976, p. 1-2), inclusive para as observações a seguir.

sequência coerente de sentenças e sim uma ocorrência comunicativa. Portanto: (a) embora desejável e aconselhável do ponto de vista do procedimento de elaboração de teorias científicas, parece impossível uma teoria formal abstrata e geral que permita a geração ou explicação de *todos* os textos possíveis de uma dada língua; e (b) não é possível aplicar ao texto as mesmas categorias gramaticais que possuímos para o estudo da frase. Isto, obviamente, leva-me ao reino da intuidividade por sacrificar alguns dos mais caros princípios de cientificidade propostos por Chomsky e que em hipótese alguma deveriam ser esquecidos. Contudo, uma observação mais acurada do que se fez até hoje em *LT* nos mostra que não se foi além da intuição, mesmo nos casos em que a autodeclaração de princípios de boa metodologia foram defendidos a todo custo. Veja-se o caso de Teun A. van Dijk em seu *Some Aspects of Text Grammars*<sup>10</sup>, onde, após feliz, clara e propositiva defesa do bom proceder em ciência, comporta-se exatamente no estilo que P. Feyerabend define como o início próprio da ciência: o método do *tudo vale*<sup>11</sup>. Não me parece que o método do "tudo vale" seja justificável em sua radicalidade, mas creio ser ele o mais *usado*, apesar de não ser o mais *defendido*.

Valha isto como posição pessoal para o procedimento "observacional" aqui adotado na elaboração da proposta apresentada.

Apoiados em Beaugrande e Dressler (cf. 1981, p. 34-37), poderíamos dizer que o texto é o resultado atual das operações que controlam e regulam as unidades morfológicas, as sentenças e os sentidos durante o emprego do sistema linguístico numa ocorrência comunicativa. Não é uma configuração produzida pela simples união de morfemas, lexemas e sentenças, mas o resultado de operações comunicativas e processos linguísticos em situações comunicativas. Um texto está submetido tanto a controles e estabilizadores *internos* como *externos*, de modo que uma *LT* razoável não deve considerar a estrutura linguística

<sup>10</sup> Cf. van Dijk (1972) e também (1977a).

<sup>11</sup> Cf. P. Feyerabend (1975, p. 23): "O único princípio que não inibe o progresso é: tudo vale".

como fator único para a produção, estabilidade e funcionamento do texto. Nem se pode tratar o texto simplesmente como uma unidade maior que a sentença, pois ele é uma entidade de outra ordem na medida em que é uma *ocorrência na comunicação*. Caso analisemos um texto como foi feito na perspectiva de Harris, por exemplo, corremos o risco de não estar analisando o texto e sim sempre agrupando elementos textuais em torno de categorias gramaticais como morfema, sintagma etc. O texto forma uma rede em várias dimensões e se dá como um complexo processo de mapeamento cognitivo de fatores a serem considerados na sua produção e recepção. O texto não é o resultado automático de uma série finita de passos em que se usaram algumas regras recorrentes observando a boa formação frasal de todas as relações na sequência, ao qual se aplicaria algum componente interpretativo. Em suma, o texto é algo essencialmente diverso de uma sentença muito longa.

## DEFINIÇÃO GERAL PROVISÓRIA DE LINGUÍSTICA DE TEXTO

*Os homens são animais que se sustentam em redes  
de significação que eles próprios tecem.*

MAX WEBER

**P**artindo das concepções de texto aqui expostas, não faz muito sentido discutir se o texto é uma unidade da *langue* ou da *parole*. Trata-se de uma unidade comunicativa atual realizada tanto ao nível do *uso* como ao nível do sistema. Tanto o sistema como o uso têm suas funções essenciais. Neste ponto podemos concordar com Bierwisch, ao con-