

UM ESTUDO DE *CORPUS* PARA A CARACTERIZAÇÃO DO USO DO  
PRONOME “ELES” EM CASOS DE ANÁFORA CONCEITUAL

Beatriz de Oliveira SALGADO  
Orientador: Edson Françaço

**RESUMO:** Esta pesquisa propõe investigar as anáforas conceituais em seus contextos de ocorrência real no português brasileiro (PB). Para isso, foi realizado um trabalho de levantamento de *corpus* em que buscamos, com ajuda de softwares, o pronome “eles” em mais de oitenta entrevistas do projeto NURC e em conteúdo textual extraído do jornal Folha de S. Paulo (*corpus* CETEN-Folha). Em seguida, classificamos as ocorrências desse pronome, com a elaboração de testes semânticos para distinguir anáforas conceituais de casos de indeterminação ou anáforas correferenciais. Um dos maiores impactos dessa pesquisa é a contribuição para os estudos recentes no campo, incluindo para a elaboração de experimentos futuros, em que poderá ser mais bem controlado o contexto das anáforas conceituais no português brasileiro, agora com base em ocorrências reais.

**Palavras-chave:** Psicolinguística, Anáfora conceitual, Resolução pronominal.

## INTRODUÇÃO

Este trabalho<sup>1</sup> teve como principal objetivo o levantamento de *corpus* para a complementação de recentes estudos sobre o processamento da anáfora conceitual. Segundo Godoy (2010), baseando-se em Gernsbacher (1991), anáfora conceitual seria o “nome dado aos casos em que um pronome plural *eles* é saturado a partir de uma expressão de desinência singular”.

As anáforas que estudamos são aquelas em que o pronome “eles” é interpretado a partir de expressões que denominaremos “termos de grupo”, que normalmente seriam definidos pela gramática normativa como coletivos, tais como “equipe” ou “batalhão”. O que nos interessa é a retomada desses termos no singular pelo pronome no plural, consistindo, então, de uma inferência que estabelece um vínculo entre o pronome a uma expressão nominal anterior. Na vertente da Linguística Textual, esses processos são chamados de anáforas conceituais (Cf. Koch, 2002 e Marcuschi, 2005).

---

<sup>1</sup> Pesquisa de Iniciação Científica financiada pela Fapesp (processo nº 2010/20135-0).

## OBJETIVOS

O primeiro ponto de nossa pesquisa é, portanto, um levantamento dos termos de grupos que licenciariam esse anafórico plural. Em Joosten (2007), os experimentos com termos coletivos em holandês são evidências a favor da hipótese de que diferentes termos de grupo (como “família”, “dupla”, “time”, “orquestra”) poderiam propiciar uma leitura distributiva ou uma leitura coletiva para o pronome, isto é, leituras em que o pronome anafórico ocorreria em função dos membros individuais do grupo (leitura distributiva) ou em função do grupo como um todo (leitura coletiva). Isso ocorre porque, dependendo do tipo de nome de grupo que usamos, lidamos com termos que apresentam diferentes traços (como, por exemplo, [+ animado] ou [- animado]).

Além dos próprios termos de grupos, outros elementos podem também apontar para uma leitura coletiva ou distributiva. Segundo Kaup et al (2002) [apud Godoy (2010)], “diferentes expressões linguísticas podem levar a uma ou outra denotação dos termos plurais”. Por exemplo, o quantificador universal “todo o” agiria sobre o termo que ele quantifica provocando uma leitura distributiva ao salientar os membros (ou átomos) do termo de grupo em questão. Prevemos, então, que “todo o time” teria uma leitura diferente de “o time”.

Outra questão levada em conta em nossa análise é a predicação nos contextos em que ocorrem as anáforas conceituais com o pronome “eles”. Atentamos para os verbos cuja maneira de predicar seus argumentos pudesse ser caracterizada como o fez Landmann (1989) [apud Godoy (2010)]:

- Predicações coletivas: agem sobre grupos predicando-os como uma instituição singular;
- Predicações distributivas: agem sobre grupos predicando-os como soma de indivíduos.

Desse modo, a partir do *corpus* levantado pudemos observar casos reais de anáforas conceituais para tentarmos compreender quais características instauram uma leitura distributiva do termo de grupo.

Propomos, também, que nossa análise dos predicados em que as anáforas ocorrem é um bom modo de avaliar se a classificação de Godoy (2010) é suficiente para definir o que é e o que não é um predicado coletivo, já que para a autora “a denotação dos termos designadores de grupos é [...] fruto da interação desses termos com os predicados que os acompanham”.

## ORGANIZAÇÃO E CLASSIFICAÇÃO DOS DADOS

Nesta pesquisa, nossos dados são os contextos (ou “vizinhanças”) em que as anáforas conceituais aparecem, em dois *corpora* diferentes. O primeiro *corpus* que analisamos apresentava 77 arquivos de entrevistas transcritas do projeto NURC – disponíveis no Centro de Documentação Cultural Alexandre Eulálio (CEDAE-IEL/ Unicamp). Essas entrevistas transcritas são da década de 1970, com falantes das seguintes cidades

brasileiras: São Paulo (SP), Rio de Janeiro (RJ), Porto Alegre (RS), Recife (PE) e Salvador (BA). O segundo *corpus* analisado é o CETEN-Folha, um arquivo único com todo o conteúdo textual do jornal Folha de S. Paulo do ano de 1994, com quase 24 milhões de palavras no total, que inclui o texto dos 365 dias daquele ano, já divididos em extratos pela equipe organizadora do *corpus*.

Uma vez processado o *corpus* com o software *Wordsmith*, buscamos o pronome “eles” e em seguida classificamos suas ocorrências, distinguindo-as dos casos de anáfora conceitual, com que trabalharemos mais detidamente na segunda etapa do projeto. Inicialmente, trabalhamos com o *corpus* do NURC, e os dados aqui apresentados referem-se a esse material.

Ao buscarmos o pronome “eles”, os dados que colhemos apresentam a seguinte estrutura, em um primeiro momento:

*Nome do arquivo – Dado – Classificação*

A razão de tal estrutura é porque de todas as ocorrências que obtivermos, haverá somente algumas de anáfora conceitual e nos propusemos a classificar a ocorrência de todas elas, como na tabela abaixo, em que os exemplos foram retirados dos inquéritos NURC/SP nº 62 D2, 78 D2, 90 DID e NURC/RE nº 151 D2, respectivamente.

Nome do arquivo	Dado	Classificação
nurc62_sp.txt	[...] uo paga entende?...a::o governo alega que precisa do...do técnico hoje em dia...eles precisam mais do técnico do que o indivíduo de formação...univer-sitária pr [...]	Anáfora conceitual
nurc78_sp.txt	[...] disso eu acho que:... ela foi:: sensacional o que ela fez com o Rui Guerra... eles se davam bem isso aquilo resolveram ter filho não deixaram de ter filho por [...]	Correferencialidade
nurc90_sp.txt	[...] ndo restaurar... mas é nesse mesmo estilo de sala completa a sala que era o que eles chamam... que se chama de jardineira... é um espelho muito grande e tem uma [...]	Indeterminação
nurc151_re.txt	[...] arde o carteiro bate lá em casa e quando bate palma já sabe que é o carteiro... eles sao pontualíssimos... os selos que às vezes vinham faltando nas cartas... a [...]	Genérico

Tabela 1. Classificação dos diferentes usos do pronome “eles”

É preciso destacar que, nesta pesquisa, foi necessário excluir alguns dados do *corpus* NURC, que é constituído de entrevistas transcritas. Nosso critério para tal exclusão era observar se o dado apresentava interrupções na fala, que são muito comuns. Com essas interrupções, nossos dados ficavam comprometidos quanto à análise para a classificação da retomada do pronome “eles”, muitas vezes tornando essa retomada impossível, seja porque o falante desistiu de continuar articulando sobre algum assunto, seja porque ele foi interrompido pelo outro falante da entrevista etc. Um dos exemplos atestados de interrupção em nosso *corpus* e, portanto, excluído de nossa análise foi:

(1) [...] les queriam nivelar todo mundo mas... éh:: nivelar por baixo não é?... que se eles... o ideal é nivelar você::... fazendo o:: negócio progredir pra:: turma ir [...] (NURC/SP nº 109)

Enfim, é válido comentar que nossa classificação ainda não é definitiva, pois conforme são encontrados possíveis novos casos de retomadas anafóricas com “eles”, faz-se necessário rever a classificação existente.

### **CASOS DE ANÁFORAS CONCEITUAIS**

Primeiramente, é importante observar que as anáforas conceituais com que trabalhamos aqui são de um tipo específico – quando os nomes de grupo são retomados pelo pronome “eles”, no plural. Um dos exemplos desses termos de grupo, atestados em nosso *corpus*, é:

(2) “[...] espécie de uma cantina que explora...essa parte de merenda para alunos...entao eles fornecem sanduiche...e tem as serventes...elas procuram fazer essa merenda.” (NURC/RJ nº 328 DID)

Nomes de grupo como esses, no singular, quando retomados por um pronome no plural (“eles”) representam um caso de anáfora conceitual, pois, segundo Godoy (2010):

“a possibilidade de termos que a gramática normativa chama de *coletivos* serem antecedentes de um pronome plural pode ser explicada se considerarmos que, nestes casos, associam-se a eles um grupo de pessoas às quais o pronome pode remeter.”

### **CASOS DE CORREFERENCIALIDADE ESTRITA**

Já os casos de correferencialidade serão classificados como tal quando o pronome “eles” retoma um antecedente correferencialmente e isso se dá quando o (núcleo do) antecedente concorda em gênero e número com o termo que o retoma (no caso, “eles”). Apresentamos abaixo um dos nossos dados cuja ocorrência foi classificada como um caso de correferencialidade, em que “eles” está retomando “os meninos (em geral)”:

(3) [...] so primário...o que é que os meninos em geral...levam além de livros? L1 ahn::eles levam geralmente seus livros e sua pasta éh::classificadores...ahn::...é ca [...] (NURC/SA nº 231 DID)

## CASOS DE INDETERMINAÇÃO

Em nossa classificação, estamos de acordo com os conceitos de outros autores quanto à indeterminação do sujeito, como a não especificação do agente da ação verbal quer pelo fato de o falante não o conhecer, quer pelo fato de não querer apontar o agente (Bechara, 1988) e a impossibilidade de identificação de seu referente no contexto em que se insere (Micheletti e Franchetti, 1996 *apud* Souza, 2007).

Por exemplo, na tabela 1 apresentada na seção 3 deste trabalho, nota-se que o dado indeterminado “eles chamam” é retomado por outra indeterminação, uma construção na voz passiva sintética (“que se chama”):

(4) [...] ndo restaurar... mas é nesse mesmo estilo de sala completa a sala que era o que eles chamam... que se chama de jardineira... é um espelho muito grande e tem uma [...] (NURC/SP nº 90 DID)

Outro exemplo de indeterminação, como (5) abaixo, apresenta um caso em que nem mesmo observando um contexto maior para o dado conseguiríamos obter uma referência para o “eles” da sentença:

(5) [...] CHEga...depois do maternal...( ) L1 é o primário é::o primário né? (que) eles não dizem mais primário agora né::...é curso série como?... Doc primeiro [...] (NURC/AS nº 231 DID)

## CASOS GENÉRICOS

É válido notar que há alguns casos que antes poderiam ter sido considerados ambíguos, em uma análise superficial, como o último exemplo da tabela 1 acima. Casos como esse foram classificados como casos “genéricos”, pois no exemplo em questão, o termo de grupo “o carteiro” refere-se não a um indivíduo específico ou a um conjunto específico de indivíduos, mas a toda a classe de carteiros.

Desse modo, todos os termos que designavam profissões, nacionalidades, classes de animais e termos semelhantes, que sempre se referiam a uma classe toda de pessoas, encaixaram-se nessa classificação (exemplos atestados no *corpus* do NURC: biólogo, filósofo, morcego, beija-flor, domador, aluno, operário, uruguaio etc).

Depois de rotuladas todas as ocorrências do pronome “eles”, selecionamos somente os dados de anáfora conceitual para classificá-los de acordo com os tipos de verbos e como os nomes de grupo são apresentados (com ou sem determinantes etc). Descrevemos, assim, a ocorrência da anáfora conceitual aqui estudada, que é a retomada de um termo de grupo através do pronome “eles”.

## DETALHANDO A ANÁFORA CONCEITUAL

Obtivemos até o momento, com o *corpus* NURC, um total de 195 dados de anáfora conceitual, relativo a 12% do total de ocorrências do pronome “eles” em nossos dados. Estima-se que no *corpus* CETEN-Folha, cuja análise ainda está em andamento, em torno

de 8% dos dados de “eles” sejam anáforas conceituais.

Nossos dados foram classificados quanto:

- ✓ ao termo de grupo a que o “eles” se refere;
- ✓ à animacidade do termo de grupo e como ele se apresenta;
- ✓ à apresentação do termo de grupo;
- ✓ ao predicado do contexto antecedente.

As categorias acima foram determinadas visando a uma caracterização mais detalhada das anáforas conceituais que ocorrem com nomes de grupo retomados por “eles”. O objetivo é atentar para quais características podem contribuir para que um termo de grupo tenha uma leitura distributiva ou coletiva, o que poderia ser observado através das características listadas acima.

Adaptamos uma tabela de Joosten (2007, p.92) para exemplificar nossa classificação, apresentada a seguir:

	“Fundar/instituir”	“Velho”	“Sentar”
A (Type 1)	c	-	-
B (Type 2)	c	m	m
C (Type 3)	-	m	m
N	-	-	-

Tabela 2. Variação na distribuição das propriedades de nomes de grupo em PB

Na versão de Joosten (2007) para esta tabela, são apresentadas mais propriedades (incluindo adjetivos como “loiro”, “bêbado” e “jovem”) e todas são classificadas quanto à sua *acessibilidade de membros* (‘member level accessibility’). Ou seja, uma propriedade com baixa acessibilidade de membros receberia a classificação *c* [+ coletiva], enquanto uma com alta acessibilidade de membros seria classificada como *m* [+ distributiva].

Quanto à classificação das propriedades relativas a um termo de grupo, é preciso dizer que Joosten (2007) defende que essa classificação, na realidade, é um *continuum*. Desse modo, é muito difícil apontar qual propriedade é *c* ou *m*, o que seria um ponto a favor da classificação das propriedades com traços como [+ coletivo], [- coletivo].

Em nossa tabela, assim como na de Joosten, propriedades classificadas como A (ou Type 1, segundo o autor), como os verbos “fundar” ou “instituir”, propiciariam uma leitura [+ coletiva] para o termo de grupo de sua sentença. Enquanto isso, propriedades classificadas como C (Type 3), representadas na tabela pelo verbo “sentar”, propiciariam uma leitura [+ distributiva], ou seja, ocorrendo somente com nomes de grupo cujos membros estejam mais acessíveis, segundo o autor. As propriedades classificadas como B (Type 2), assim como determina Joosten, são as que não se encaixam nem em A nem em C, localizando-se entre propriedades A e C no *continuum* mencionado acima.

No entanto, em nossa pesquisa, acreditamos haver a necessidade de repensar a tabela original de Joosten (2007), que até então não previa casos que não se adequavam às classificações acima, por mais que representasse o *continuum* sugerido pelo autor. Desse modo, nomes de grupos atestados em *corpus* como “desenho infantil” ou “crônica esportiva” não se encaixavam nas classificações A, B ou C, pois o que propomos é que

eles tenham alguma característica diferente dos outros nomes de grupo até então classificáveis, sendo classificados, então, em “N”.

Para classificar esses dados como A, por exemplo, testávamos se poderiam funcionar como uma entidade, assim como os nomes “banco”, “universidade” etc. Por outro lado, se funcionassem como os nomes “criança” e “família”, então teriam uma alta acessibilidade de membros. Desse modo, seja por poderem apresentar múltiplos sentidos (“crônica esportiva” como o artigo jornalístico, como todos que escrevem crônicas esportivas, como aquele autor específico?) ou uma dificuldade de especificação dos membros do grupo (em “desenho infantil”, seriam os autores do desenho os retomados por “eles” ou os personagens?), optamos por classificar tais dados em outra categoria.

Voltando à nossa lista com a classificação dos dados de anáfora conceitual, retomamos o exemplo (2) da seção 3.1 deste trabalho:

(2) “[...] espécie de uma cantina que explora...essa parte de merenda para alunos...entao eles fornecem sanduiche...e tem as serventes...elas procuram fazer essa merenda.” (NURC/RJ nº 328 DID)

Neste caso, classificamos o termo de grupo “cantina” segundo a tabela 3 abaixo:

Dado	Classif. dado	Predicado	Classif. predicado	"Pessoal"	Animacidade	Apresentação do dado
uma cantina	A	que explora	coletivo	não	não	indefinido

Tabela 3. Classificação dos nomes de grupo e seus predicados

A Tabela 3 acima sintetiza a etapa de nossa pesquisa que foi dedicada à classificação de cada ocorrência de anáfora conceitual atestada em *corpus*. Nela, temos o termo de grupo (chamado “dado” por ser nosso objeto de pesquisa) apresentado junto de determinantes ou modificadores que o acompanhem e que são, por sua vez, classificados na última coluna da tabela (“apresentação do dado”). O termo de grupo, como já explicado, foi classificado como A por apresentar baixa acessibilidade de membros, sendo assim [+ coletivo]. Seu predicado, “que explora”, foi na coluna seguinte classificado como “coletivo”, pois é um exemplo de predicado com características que o colocam mais próximo de A do que de C, se voltarmos à Tabela 2, baseada em Joosten (2007).

Optamos por apresentar também uma categoria a respeito da animacidade do termo de grupo, pois essa característica pode ter um papel importante ao observarmos elementos que possam contribuir para que seja licenciada uma leitura distributiva ou coletiva ao termo de grupo. No caso de “uma cantina”, classificamos como [- animado], ou seja, como sendo um termo de grupo não só com baixa acessibilidade de membros, mas podendo “viver uma vida própria, independente de seus membros individuais” (Joosten, *ibidem*, p 93). Dölling (1991, p. 164 *apud* Joosten, 2007, p. 93) ainda observa que alguns nomes de grupo

“may, within certain limits, gain or lose members without detriments to their identity and continued existence. For example, an orchestra can continue to exist even though in the course of time it undergoes a complete change of membership.”

Joosten completa afirmando que “the opposite is also true: an orchestra can cease to exist without members having died.”

Como até o momento obtivemos diversos dados (cerca de 40) com a palavra “pessoal”, julgamos relevante anotá-la como item de nossa classificação para podermos continuar observando se esta seria uma tendência: a construção de termos de grupo com a palavra “pessoal”, em textos orais. Além disso, atentamos para o fato da difícil classificação do termo “pessoal” segundo os critérios da Tabela 2, ou seja, em nossa pesquisa, os dados que contêm o termo “pessoal” foram classificados como B (Type 2).

## CONSIDERAÇÕES FINAIS

Enfim, é válido comentar que nossa classificação ainda não é definitiva, pois conforme são encontrados possíveis novos casos de retomadas anafóricas com “eles”, faz-se necessário rever a classificação existente. Podemos também dizer que os resultados de nossa busca pelos *corpora* foram apreciáveis, pois as anáforas conceituais em que o pronome plural (no caso, “eles”) não tem um antecedente correferencial mostraram-se bastante produtivas, o que nos possibilita dizer que o fenômeno também é relevante no português brasileiro, assim como no holandês (Joosten, 2007). Até o momento, os resultados desta pesquisa estão de acordo com Godoy (2010), ao descrever as anáforas conceituais como “bastante frequentes em português, tanto na oralidade quanto na escrita”.

## REFERÊNCIAS

- CETEN-Folha é um dos *corpora* acessíveis através do projeto AC/DC, parte do *Corpus* NILC, disponível em: <http://www.linguateca.pt/acesso/NILCsaocarlos.html>.
- GODOY, M. C. Resolvendo a anáfora conceitual: um olhar para além da relação antecedente/anafórico. Dissertação de mestrado, Campinas, 2010.
- JOOSTEN, F. G. et al. “Dutch collective nouns and conceptual profiling”. *Linguistics*, vol. 45, 2007, p. 85-132.

## BIBLIOGRAFIA

- BECHARA, E. **Moderna Gramática da Língua Portuguesa**. 31ª. ed. São Paulo: Nacional, 1988.
- GERNSBACHER, M. A. Comprehending conceptual anaphors. **Language and Cognitive Processes**, vol. 6, 1991, p. 81-105 *apud* GODOY, M. C. Resolvendo a anáfora conceitual: um olhar para além da relação antecedente/anafórico. Dissertação de mestrado, Campinas, 2010.
- DÖLING, J. Group expressions and the semantics of plurals and collective nouns. ZIMMERMANN, I. (ed.) **Syntax und Semantik der Substantivgruppe**, 1991, p. 147-183. Berlin: Akademie Verlag *apud* JOOSTEN, F. G. et al. Dutch collective nouns and conceptual profiling. *Linguistics*, vol. 45, 2007, p. 85-132.

- KAUP, B.; KELTER, S.; HABEL, C. Representing Referents of Plural Expressions and Resolving Plural Anaphors. **Language and Cognitive Processes**, n. 17, 2002, p. 405-450 *apud* GODOY, M. C. Resolvendo a anáfora conceitual: um olhar para além da relação antecedente/anafórico. Dissertação de mestrado, Campinas, 2010.
- KOCH, I. V. **Desvendando os segredos do texto**. São Paulo: Cortez, 2002.
- LANDMANN, F. Groups I. **Linguistics and Philosophy**, n. 12, 1989, p. 559-605 *apud* GODOY, M. C. Resolvendo a anáfora conceitual: um olhar para além da relação antecedente/anafórico. Dissertação de mestrado, Campinas, 2010.
- MARCUSCHI, L. A. Anáfora indireta: o barco textual e suas âncoras. In: Koch, I.V.; Morato, E. M.; Bentes, A. C. (orgs.) **Referenciação e discurso**. São Paulo: Contexto, 2005, p. 53-101.
- MICHELETTI, H.; FRANCHETTI, S. A Indeterminação do Sujeito: um estudo Sociolingüístico. **Anais do XXV Seminário do GEL**. Taubaté, São Paulo, 1996, p. 629-635 *apud* SOUZA, E. M. **O uso do pronome ‘eles’ como recurso de indeterminação do sujeito**. Dissertação de mestrado, Belo Horizonte, 2007.
- SOUZA, E. M. **O uso do pronome ‘eles’ como recurso de indeterminação do sujeito**. Dissertação de mestrado, Belo Horizonte, 2007.