

PARA ALÉM DA ESCOLHA LEXICAL NO PB E NO PE: DA PARADA DE ÔNIBUS À PARAGEM DE AUTOCARRO

Milena de Uzeda GARRÃO¹; Roberto Botelho RONDININI²
(Universidade Federal Rural do Rio de Janeiro)

Resumo: *Esse trabalho apresenta um estudo contrastivo sobre a formação de nomes deverbais no Português do Brasil (PB) e no Português Europeu (PE). Com base em análise de dados obtidos em corpora morfossintaticamente etiquetados, confirmamos a hipótese inicial de que em PB o sufixo -da tem uma função categorial mais abrangente do que em PE, derivando adjetivos e substantivos, nos casos dos verbos de movimento parar e virar.*

Palavras-chave: *Linguística de Corpus, Morfologia Contrastiva do PB e do PE, Nominalização, Processo de Formação Deverbal, Processamento de Linguagem Natural.*

INTRODUÇÃO

O presente trabalho trata da etapa inicial de um projeto de pesquisa que tem como objetivo a descrição contrastiva do processo de formação sufixal do PB e do PE, uma vez que podemos observar empiricamente uma distinção bastante acentuada em nível morfossintático entre essas variantes. Para a realização dessa primeira etapa do trabalho, concentramos a nossa análise na derivação por formação de nomes, mais precisamente sobre o processo de formação de nomes deverbais.

Sabe-se que os nomes ou substantivos têm alto poder discriminatório (função denominativa) em comparação às outras classes de palavras. Sua carga informacional traz um impacto imediato tanto para a descrição da língua portuguesa como para alguns campos de aplicação dos estudos sobre o léxico, como, por exemplo, para as áreas de Recuperação de Informações (RI) e de Tradução Automática, no domínio do Processamento de Linguagem Natural (PLN, doravante), que dependem prioritariamente da descrição de substantivos de uma determinada língua. E uma vez que muitos casos de nominalização no PB e no PE apresentam variações na sua formação sufixal (como, por exemplo, *sujeira* no PB/ *sujidade* no PE), essas preferências devem ser fornecidas aos sistemas de busca de informações ou ao programa tradutor.

Sob uma ótica descritivista, a falta de tratamento sistemático dos contrastes de formação de sufixos nominais em PB e PE pode vir a acarretar uma avaliação maquiada do uso de ambas as modalidades, comprometendo diretamente domínios da linguística aplicada, como, por exemplo, o ensino de português como segunda língua e de tradução, além de outros campos multidisciplinares, como a área de PLN, citada anteriormente.

Dentre as implicações práticas, destaca-se o campo da lexicografia computacional, mais especificamente os domínios de RI (campo em que os substantivos são considerados os mais importantes geradores de informações) e de tradução automática, domínio onde se evidencia a importância da descrição das distinções constitutivas dos sufixos nominais dessas duas modalidades do português. Veja-se, por exemplo, a colocação da língua inglesa “*bus stop*”, que

¹ Doutora em Linguística (PUC-Rio), Professora Adjunta de Linguística da UFRRJ, migarrao@terra.com.br

² Doutor em Letras Vernáculas (UFRJ), Professor Adjunto de Língua Portuguesa da UFRRJ, robertorondinini@hotmail.com.

seria traduzida para o PB como *parada de ônibus* e para o PE como *paragem de autocarro*. Casos como esse demonstram que o contraste lexicográfico vai além da descrição de um léxico acidentalmente herdado ou privilegiado pelas duas culturas, como *ônibus* e *autocarro*, mas também invade o domínio da morfologia derivacional. Portanto, há de se estabelecer pressupostos teóricos e metodológicos que possam apoiar uma descrição de uso de uma base + sufixo nas duas variantes.

Um exemplo bastante ilustrativo desse contraste pode ser evidenciado nas sentenças que seguem: um falante de PB está propenso a dizer “A escova**ção** e o fio dental garantem que seus dentes fiquem livres de suj**eira**”, enquanto um falante do PE tenderá a optar pela frase “A escova**gem** e o fio dent**ário** garantem que seus dentes fiquem livres de suj**idade**.” Como esses, há diversos outros casos de preferência de uso de sufixos que evidenciam um texto ou fala como característicos do PB ou do PE, como estabeleceremos aqui, através dos exemplos de nominalização dos verbos de movimento *parar* e *virar*, retirados de *corpora* morfossintaticamente etiquetados de ambas as variantes.

1. OBJETIVOS

Os processos de nominalização em PB e PE, como referidos anteriormente, merecem um estudo teórico mais profundo, uma vez que, em muitos casos, apresentam variações na sua formação sufixal. Note-se, contudo, que consideraremos uma base comum às duas variantes, muito embora estejamos partindo de uma análise contrastiva que implica o estabelecimento de possíveis regras para aplicações computacionais futuras.

Pretendemos, desse modo, não somente contribuir para os estudos da morfologia derivacional, mas também para aqueles que norteiam a lexicografia contrastiva, através da descrição de tais processos de nominalização deverbal. Com base nas (ir)regularidades detectadas, pretendemos, ainda, avaliar no âmbito de PLN, a real necessidade de conceber uma ferramenta computacional para produção automática de padrões de nominalização a serem analisados ou somente listagem dos contrastes derivacionais.

Por fim, buscamos demonstrar que o domínio da Linguística de *Corpus* revela de forma bastante realista a escolha dos falantes nativos do PB e do PE e traz dados estatísticos robustos que viabilizam a descrição das variantes em questão. Em função dessa solidez descritiva, sempre que possível, a Linguística de *Corpus* também viabiliza a formalização de alguns fenômenos linguísticos. Para o caso específico em análise no presente trabalho, pretendemos investigar os processos de nominalização dos verbos de movimento *parar* e *virar*, considerados exceção à regra geral dos verbos de movimento.

1.1. Um pouco sobre implicações práticas

No domínio de Recuperação de Informações, o problema de armazenamento de informações pôde ser razoavelmente amenizado com o uso de estocagem digital. No entanto, a maneira como os dados armazenados podem ser recuperados não tem uma resposta trivial e vem sendo trabalhada desde a década de 1940. Sistemas computacionais hoje têm uma capacidade virtualmente infinita de armazenagem de documentos, além de velocidade compatível para a manipulação de milhões de dados. Contudo, o que parece ser uma vantagem pode acabar por apresentar dois tipos de problemas. Em primeiro lugar, a quantidade de documentos apresentada ao usuário em busca de informação costuma ser tão extensa que fica quase impossível que ele

Para além da escolha lexical no PB e no PE: da parada de ônibus à paragem de autocarro

consiga verificar todos os documentos recuperados. Por outro lado, em uma quantidade tão grande de documentos, o fator relevância é crucial e nem todos os documentos são igualmente relevantes para aquilo o que o usuário pretende.

Para que um documento seja passível de recuperação, ele é previamente processado, isto é, são indexadas a ele as palavras que indicam seu conteúdo, ficando de fora todas aquelas que não têm nenhum valor informativo, tais como preposições, conjunções e artigos. Antes do aparecimento de computadores de alta velocidade, essa tarefa era feita manualmente. Contudo, hoje há técnicas de realizar a indexação automaticamente. Uma dessas técnicas consiste na tentativa de se estabelecerem as palavras e expressões que representam o conteúdo de cada documento, com maior precisão.

As palavras ou expressões usadas na indexação são chamadas de “termos” ou “palavras-chave”. Os termos são os responsáveis pela apresentação dos documentos, sempre levando em conta medidas convencionadas para a avaliação de documentos recuperados. Tais medidas são conhecidas como precisão e *recall*. Precisão é a relação entre o número de documentos relevantes recuperados e o número total de documentos recuperados; e *recall* é a relação entre o número de documentos relevantes recuperados e o total de documentos relevantes (recuperados ou não) (GARRÃO; DIAS, 2003). Em geral, as palavras utilizadas para a indexação e, conseqüentemente, para a recuperação de documentos, ou seja, os termos, são substantivos ou, mais precisamente, sintagmas nominais.

Portanto, tomando como pressuposto a relevância do substantivo para a Teoria da Informação e, mais precisamente, para o domínio de RI, iniciamos nossa investigação com os processos de nominalização entre o PB e PE. Uma das contribuições imediatas do estudo do processo de formação de nomes é a identificação da própria origem do documento. Isto é, um texto com substantivos como *paragem*, *sujidade*, *patinagem* denotaria a origem da variante europeia; enquanto um texto contendo substantivos como *parada*, *sujeira* e *patinação* implicaria uma origem brasileira. Isso já seria um ganho informacional bastante relevante do ponto de vista informacional.

2. METODOLOGIA

Esta investigação considera o fenômeno de formação de nomes deverbais com nítida ênfase nos dados empíricos disponíveis em PB e PE. Por isso, dependemos inevitavelmente da utilização de *corpora* para dar conta das falhas de intuições do pesquisador quanto ao que vem a ser a formação nominal utilizada e ou recorrente. Para esse estudo, utilizaremos o *corpus* NILC/São Carlos (AIRES; ALUÍSIO, 2001), que contém 42 milhões de palavras do PB com texto jornalístico da Folha de São Paulo, mas também cartas comerciais e textos didáticos; já para o PE, utilizaremos o *corpus* CETEMPúblico (Corpus de Extractos de Textos Electrónicos MCT/Público) (ROCHA; SANTOS, 2000), que contém aproximadamente 180 milhões de palavras.

O foco da análise nessa primeira etapa da pesquisa é a formação de nomes através do processo sufixal deverbal, que pode trazer evidências em relação ao uso de um dado sufixo em detrimento de outro. Apresentaremos aqui um estudo de caso que contempla a preferência de uso dos sufixos –AGEM e -DA para a nominalização sufixal deverbal de *virar* e *parar* em ambas as variantes: PB e PE.

3. DADOS SOBRE A FORMAÇÃO E USO DA NOMINALIZAÇÃO DEVERBAL

Em estudo para a descrição do PB na modalidade falada, Basílio (1996:28) foca a formação e uso da nominalização deverbal sufixal. Em uma análise de *corpus* que englobava tanto discurso formal, discurso dialogado e entrevista formal, a autora encontrou os seguintes resultados: de um total de 362 formas nominalizadas regulares, constataram-se 218 ocorrências de -ção (60,2%), 73 ocorrências de -mento (20,1%), 48 ocorrências de -nc(i)a (13,2%), 18 ocorrências de -da (4,9%) e 5 ocorrências de -agem (1,4%).

Em virtude de haver em -ção um caráter semântico generalizante, a autora não se surpreende com sua hegemonia na língua falada. Contudo, ratifica que a oposição -ção/ -mento deve ser mais aprofundadamente estudada tendo em vista a relativa analogia na sua função (caráter processual). Basílio argumenta que, em relação aos outros sufixos deverbais, uma proporção minoritária já era esperada, em virtude das especificações semânticas. Ela explica que a ocorrência de -da é condicionada a verbos de movimento e a ocorrência de -agem está atrelada a verbos indicadores de operações, nos quais compete com -ção e -mento. Note-se, contudo, que a análise de Basílio (1996) foi especificamente focada no discurso do PB.

Já Rio-Torto (1998:119), em um estudo não quantitativo mas descritivo da morfologia derivacional do PE, utiliza a denominação *Nomina actionis* para abranger os nomes deverbais parafraseáveis por “o fato de V” ou ação/processo e/ou resultado da ação/processo de V. Neste grupo, a autora cita os sufixos -mento, -ção, -gem, -ão, -nça. Interessante notar que não há aqui nenhuma menção direta ao sufixo -da. A autora analisa esse sufixo quando descreve nomes de eventos não deverbais mas sim denominais (como *facada*, *alfinetada*), algo compartilhado pelo PB, mas não exclusivamente.

Concordamos, portanto, com Basílio e sua distinção de -da para verbos de movimento e -agem para verbos indicadores de operações para o PB. A nossa hipótese, contudo, é de que em PE, -agem tende a exercer a função de -da em relação a alguns verbos de movimento, como *virar* e *parar*.

4. DADOS CONTRASTIVOS DO PB E DO PE

Voltemos nossa atenção aos sufixos -mento, -ção, -agem. Segundo Basílio (1996:28), eles tendem a ser intercambiáveis (usados com verbos indicadores de operações) se analisarmos os sufixos somente do ponto de vista semântico. Aqui há, portanto, uma explicação teórica que justificaria o fato de em PB o sufixo preferencial para adjunção ao verbo *deslocar* ser *deslocamento* e em PE, *deslocação*; o contrário ocorre com *apurar*, cujo uso preferencial em PB é *apuração* e em PE, *apuramento*. Em relação a concorrência -ção e -agem, podemos citar *patinação* e *rotulação* em PB e *patinagem* e *rotulagem* em PE. Contudo, em relação a -da e -agem percebemos um fenômeno particular e merecedor de destaque, conforme descreveremos na seção a seguir.

4.1. Estudo de caso da nominalização de *virar* e *parar* em PB e PE

Análises preliminares nos *corpora* indicam que os verbos de movimento seguem uma regularidade. Isto é, verbos como *andar*, *correr*, *ir*, *chegar*, *descer*, *subir* produzem formas nominais por meio do acréscimo do sufixo ‘-da’, que podem ser formalizadas tanto em PB quanto em PE por uma RFP (cf. BASÍLIO, 1980) do tipo

Para além da escolha lexical no PB e no PE: da parada de ônibus à paragem de autocarro

$$[X]_{V_{mov}} \rightarrow [X-da]_N$$

“ato de X”

Pudemos notar conforme análise dos *corpora* de PB e PE utilizados nesse estudo de caso que há uma diferença clara de uso dos sufixos *-da* e *-agem* em relação às duas variantes da língua portuguesa. Durante a análise dos dados, observamos que os verbos *virar* e *parar* apresentaram comportamento diferenciado: em PB, a nominalização deverbal ocorreu unicamente com a adjunção do sufixo ‘-da’, formando produtos de categorias nominais e adjetivais. Já em PE, verificamos a adjunção de ‘-da’ para a formação de adjetivos e de ‘-agem’ para nomes. Em outras palavras, em PE parece haver uma restrição morfosintática maior em relação à derivação que tem como base esses verbos, onde a regra de formação $[V_{mov}]$ -da formaria adjetivos e $[V_{mov}]$ -agem formaria substantivos. Vamos aos exemplos dos *corpora*:

(a.1) Exemplo de *virada* em PE (corpus CETEM Público): total de 999 ocorrências

par=ext23161-soc-92b-1: “A sensibilidade do mundo ocidental está **virada** para o êxito individual e para a curtição”.

par=ext184748-pol-97a-2: “Isto exige uma política **virada** para a constante audição dos cidadãos, para a prática democrática efectiva”.

par=ext66136-nd-94a-3: “Numa delas, a varanda **virada** para o mar” .

par=ext10769-soc-96b-2: “Todavia, a ajuda internacional está toda **virada** neste sentido, as palavras de ordem são repatriação e reconstrução.”

par=ext23868-eco-98a-2: “Inapa **virada** para o exterior”

(a.2) Exemplos de *viragem* em PE (corpus CETEM Público): total de 2712 ocorrências

par=ext3817-clt-93a-2: “Este livro corresponde a uma **viragem** na obra de Luísa Dacosta”.

par=ext24866-soc-98b-2: “Os jovens e a **viragem** do século”.

par=ext20924-clt-92a-2: “É o primeiro momento de grande clímax e com essa **viragem** dá-se um novo sentido à história” .

par=ext8935-nd-93b-2: 3. O Vaticano II operou ainda uma outra **viragem** decisiva .

par=ext22682-nd-95a-1: O ponto de **viragem** foi o massacre de Santa Cruz em Novembro de 1991 .

Pudemos observar, através de uma análise exaustiva dos *corpora*, que *virada* em PE exerce função adjetival, enquanto *viragem* exerce função nominal. Comparemos esses resultados com os do PB:

(a.3) Exemplos de *virada* in PB (corpus NILC/São Carlos): total de 809 ocorrências

par=19967: “Tem o nome pomposo de Brasil Real -- Desafios da pós-estabilização na **virada** do milênio”.

par=25733: “A principal razão da **virada** foi a retração de 40 % na indústria açucareira .”

Milena de Uzeda Garrão; Roberto Botelho Rondinini

par=26407: “Aliviado com a **virada**, Carlos Germano não se conteve.”

par=33435: “Chacina é página **virada**, prometeu.”

par=57478: “Sugerimos ao colega que transmitisse com a antena **virada** a 180, da posição atual, isto é, que transmitisse pela via longa.”

par=101922: “Mais tarde, para posicionar-se para a entrada em órbita lunar, a Apollo 8 foi **virada** de maneira que o motor ao final do módulo de serviço apontasse para a Lua.”

(a.4) Exemplos de *viragem* em PB (corpus NILC/São Carlos): total de 5 ocorrências

par=61576: TG: “O ponto de **viragem** foi, a meu ver, a Revolução.”

par=Mais-94b-nd-1: “Penso no século 4 a.C. (com a formação da filosofia grega clássica) e na **viragem** do século 19 ao 20 (com diversas tentativas de devolver à filosofia seu fundamentum absolutum).”

par=61298: “Na realidade, o autor iniciou-se como neo-realista, orientado pelo materialismo dialético..., mas num livro sintomaticamente intitulado Mudança dá-se a grande **viragem**.”

par=61473: “Dos meus tempos de liceu ficou-me a idéia de que a novela se limita a contar momentos de **viragem** de uma vida...”

par=Mais-94b-nd-1: “Da mesma maneira, na **viragem** do século 19 ao 20, filósofos como E. Husserl, H. Bergson e B. Russell, cada um à sua maneira, retoma a luta contra diversas formas de relativismo...”

Como pudemos observar no corpus do PB, *virada* exerce tanto função nominal (675 casos: três primeiros exemplos em a.3) quanto adjetival (134 casos: três últimos exemplos em a.3). Se levarmos em consideração que houve somente 5 ocorrências de *viragem* em PB, poderíamos inferir que esse número representa produção de falantes nativos do PE (ou em contato com falantes de PE) em um corpus brasileiro.

(b.1) Exemplos de *parada* em PE (corpus CETEM Público): total de 2712 ocorrências

par=ext7003-clt-98b-2: “A Deco também não está **parada**.”

par=ext10416-soc-94a-2: “A câmara está **parada** e ainda não realizou sequer dez tostões de obras, disse, porque ninguém sabe quem manda e cada um puxa para seu lado.”

par=ext22331-soc-96a-2: “O José já não via só a vida mal **parada**: já a via a andar para trás e ainda a procissão ia no adro...”

par=ext35046-des-96b-1: Foi apenas para a equipa não ficar **parada**, disse Carlos Brito, que já terá sido avisado de que, na pior das hipóteses, ficará no comando até ao próximo jogo, frente ao Gil Vicente .

(b.2) Exemplos de *paragem* em PE (corpus CETEM Público): total de 3132 ocorrências

par=ext1245-des-96b-1: “Luís Jesus era o favorito no sector masculino e confirmou as previsões, neste seu regresso à competição após **paragem** por lesão.”

par=ext5428-des-96b-2: “Três meses depois, foi a vez de o frio provocar fatal **paragem** cardíaca a três pessoas que aguardavam a passagem dos carros.”

Para além da escolha lexical no PB e no PE: da parada de ônibus à paragem de autocarro

par=ext6546-soc-92b-2: 5. “Está a chover a cântaros; você passa de automóvel pela **paragem** do autocarro e vê uma bicha de pessoas sujas de lama.”

par=ext4839-nd-97b-3: Atletismo: Fernanda Ribeiro antecipa **paragem**.

Vemos aqui o mesmo padrão daquele observado no verbo *virar* em PE. *Parada* é utilizada geralmente em função adjetival, enquanto *paragem* tem claramente uma função nominal. Um outro fator semântico relacionado à ocorrência significativa de *parada* nesse corpus é que, além de exercer uma função adjetival, também poderia ser identificada como uma sinonímia de “desfile”. Isso justificaria as 1165 ocorrências de “parada” como termo substantivo em PE.

(b.3) Exemplos de *parada* em PB (corpus NILC/São Carlos): total de 755 ocorrências

par=2854: “praticar a massagem cardíaca, se houver **parada** do coração.”

par=10283: “A última **parada** de Michael, e a mais demorada, foi na loja de brinquedos e roupas infantis La Look, onde ficou por mais de 40 minutos.”

par=34779: “Em frente à pirâmide da Federação das Indústrias do Estado de São Paulo (Fiesp), nova **parada**.”

par=62397: “Estive **parada** desde 1994.”

par=Brasil-94a-pol-1: “Agora vão dizer que tem muita obra **parada**.”

par=Brasil-94a-pol-2: “Na cidade de São Paulo, continua **parada** a Glaspac, e em Salto, a Alcoa.”

(b.4) Exemplos de *paragem* em PB (corpus NILC/São Carlos): total de 2 ocorrências

par=118382: “O guerreiro cristão percorrendo essa **paragem**, começou de cismar.”

par=Folhateen-94a-soc-1: “Nesta **paragem**, ninguém tem mais o que inventar.”

Como podemos identificar na análise de corpus do PB, *parada* exerce tanto função nominal (544 casos: três primeiros exemplos em b.3) quanto adjetival (211 casos: três últimos exemplos em b.3). Se levarmos em consideração que houve somente 2 ocorrências de *paragem* em PB, poderíamos também supor que esse número representa produção de falantes nativos de PE (ou em contato frequente com PE) em um corpus brasileiro.

5. IMPLICAÇÕES COMPUTACIONAIS E DESCRITIVAS

Os dados obtidos nos *corpora* corroboram a hipótese de que, de fato, em PB o sufixo *-da* tem uma função categorial mais abrangente do que em PE, derivando adjetivos e substantivos. Houve apenas 5 casos de *viragem* e 2 de *paragem* no corpus do PB, o que pode indicar dados de falantes do PE em entrevistas ou comentários para jornais brasileiros.

Durante a análise dos dados, observamos que os verbos *virar* e *parar* apresentaram comportamento diferenciado: em PB, a nominalização deverbal ocorreu unicamente com a adjunção do sufixo ‘-da’, formando produtos de categorias nominais e adjetivais. Já em PE, verificamos a adjunção de ‘-da’ para a formação de adjetivos e de ‘-agem’ para nomes. Em outras palavras, em PE parece haver uma restrição morfosintática maior em relação à derivação que tem como base esses verbos.

Tal fato é ratificado por Basílio (1996:28) que, em um estudo do discurso falado no PB, afirma ser a ocorrência do sufixo ‘-da’ condicionada a verbos de movimento e a ocorrência de ‘-agem’ atrelada a verbos indicadores de operações.

A partir desses dados, para verbos de movimento como *parar* e *virar*, poderíamos supor um padrão de Regra de Formação de Palavras (RFP) inspirado pela visão de Basílio (1976), que em PE tenderia a ser $[V_{\text{mov}}]\text{-da} \rightarrow \text{Adj}$ e $[V_{\text{mov}}]\text{-agem} \rightarrow \text{N}$. Já em PB, a RFP $[V_{\text{mov}}]\text{-da}$ tenderia a formar tanto adjetivos como nomes.

Contudo, essa regra não parece tão produtiva assim em PE. Isto é, em relação a outros verbos de movimento como *correr*, *caminhar*, *descer*, *subir*, em PE seria mantida a mesma RFP formulada para PB³:

$$[V_{\text{mov}}]\text{-da} \rightarrow \begin{matrix} \text{Adj} \\ \text{N} \end{matrix}$$

Conclui-se, portanto, que a implementação da regra apresentada para verbos como *virar* e *parar* não teria um poder de generalização necessário para dar conta de todos os verbos de movimento em PE e em PB. Em outras palavras, a suposta regra parece estar exclusivamente atrelada a esses dois verbos. Assim sendo, do ponto de vista computacional, ao invés de uma regra generalizante para formação de nomes deverbiais a partir de verbos de movimento para o PE e para o PB, é mais econômica a descrição da exceção de ambas as formações denominais (*parar* e *virar*) para o PE.

Rio-Torto (1998:109) considera:

“Um dos aspectos do português cujo estudo sistemático importa empreender é o da descrição de seu sistema da formação de palavras. Se razões de outra natureza não houvesse, convir-se-á que o tratamento computadorizado do léxico, que atualmente mobiliza tantas atenções, requer um sólido conhecimento do sistema derivacional do português, ainda por levar a cabo.”

Embora seja bastante perspicaz a análise da autora sobre a importância do ponto de vista computacional em relação à descrição do processo de formação de palavras, um fator a ser considerado, tendo em vista a análise aqui proposta, é que o sistema derivacional do português não é homogêneo e aplicável de uma forma generalizante.

Portanto, computacionalmente, é importante que o profissional se cerque de cuidados ao pensar em implementar regras morfológicas do português em um sistema de PLN. Do ponto de vista teórico e descritivo, o mesmo cuidado se faz necessário em relação à formação de palavras do português, uma vez que a análise de *corpora* nos permite pontuar diferenças cruciais nessas regras de formação para modalidades distintas do português.

Fica patente, então, que o domínio relativamente novo da Linguística de *Corpus* contribui para um olhar mais realista sobre o uso de variantes do português e traz dados contundentes e sólidos que viabilizam a sua descrição e, sempre que possível, a formalização de alguns

³ Já verbos como *nadar*, *retornar*, *contornar*, em ambas as modalidades do português, seguiriam a derivação regressiva para formação de substantivo (*nado*, *retorno*, *contorno*) e a regra $[V_{\text{mov}}]\text{-da} \rightarrow \text{Adj}$ para formação de adjetivos. Portanto, mesmo a RFP descrita acima teria uma aplicação relativamente restrita também em PB.

Para além da escolha lexical no PB e no PE: da parada de ônibus à paragem de autocarro

fenômenos linguísticos. Para o nosso presente propósito, um estudo-piloto com objetivo de dar início a um projeto de pesquisa descritivo envolvendo o uso do português no Brasil e em Portugal, a Linguística de *Corpus* se mostra imprescindível para uma caracterização contrastiva de formação de nomes deverbais em língua portuguesa, visto termos coletado mais de 10.000 ocorrências das formas nominalizadas de *parar* e *virar*, consideradas exceção à regra geral descrita.

6. DISCUSSÃO E TRABALHOS FUTUROS

Esse foi um trabalho-piloto que dará origem a um projeto mais minucioso sobre a formação de nomes por processo de sufixação deverbal do PB e o PE para fins descritivos e computacionais. Conforme expusemos anteriormente, pretendemos contribuir de forma efetiva a) para a lexicografia contrastiva do PB e do PE na descrição da formação das unidades nominais e, mais adiante, de outras unidades como adjetivos e verbos; b) para a elaboração de padrões de formações contrastivos em ambas as variantes do português; e c) para avaliação da aplicabilidade de uma ferramenta computacional com o objetivo de produzir automaticamente os padrões de sufixação a serem analisados.

Numa primeira apreciação dos dados do nosso estudo de caso da formação de nomes a partir de bases verbais *parar* e *virar*, podemos dizer que:

- (1) Em PB o sufixo *-da* atrelado aos verbos *parar* e *virar* exerce uma função categorial mais abrangente do que em PE, derivando adjetivos e substantivos;
- (2) não é computacionalmente rentável gerar uma regra do tipo $[V_{mov.}]_{-da} \rightarrow Adj$ e $[V_{mov.}]_{-agem} \rightarrow N$ para o PE, uma vez que essa regra parece estar atrelada exclusivamente aos verbos *parar* e *virar*.

Além disso, observamos também de uma forma mais geral, que em PE parece haver uma flutuação de uso sufixal mais evidente em relação a outros tipos de processo de formação de palavras (como *aposentaçã*o / *aposenta*doria, *deslocaçã*o / *desloca*mento) enquanto que o PB parece tender a uma rigidez maior (*aposenta*doria e *desloca*mento). Com isso, fica patente a necessidade de identificação de preferências na formação de um ou outro sufixo no PE e de uma lista das nominalizações preferenciais e algumas restrições semânticas.

Portanto, nesse futuro projeto intitulado *Lexicografia contrastiva e tratamento automático de sufixos nominais do PE e do PB*, pretendemos desvendar contrastivamente o comportamento de nominalizações deverbais sufixais e, mais adiante, outros processos sufixais formadores, por exemplo, de adjetivos (como *promissor* (PB) / *prometedor* (PE)) e de verbos (*potencializar* (PB) / *potenciar* e *potencia*lizar (PE)).

Também temos como objetivo fomentar tal análise contrastiva levando em consideração outros países lusofônicos, com o intuito de mapear e investigar os processos de formação sufixal lusófona. Isso poderia contribuir para um entendimento mais global do fenômeno de diferenças nos processos de formações sufixais.

Até o presente momento este é um trabalho que parece promissor, ou mesmo, “prometedor”.

REFERÊNCIAS

- AIRES, R.; ALUÍSIO, S. Criação de um Corpus com 1000.000 de Palavras Etiquetado Morfosintaticamente. Technical Report NILC-TR-01-8, NILC, Campinas. 2001.
- BASÍLIO, M. *Estruturas Lexicais do Português: uma abordagem gerativa*. Petrópolis: Vozes. 1980.
- _____. Formação e Uso da Nominalização Deverbal Sufixal no Português Falado. In: _____; CASTILHO, A. T. (Org.). *Gramática do Português Falado* v.IV. Campinas: UNICAMP/FAPEESP, v. 4, p. 23-30. 1996.
- GARRÃO, M. U. ; DIAS, M. C. P. A detecção de pseudo-termos no domínio de recuperação de informação: o caso das expressões V+N. *Anais da ABRALIN*. 2003.
- RIO-TORTO, M. G. *Morfologia Derivacional: Teoria e Aplicação ao Português*. Coleção Linguística, Porto Editora, 1998.
- ROCHA, P., SANTOS, D. "CETEMPúblico: Um corpus de grandes dimensões de linguagem jornalística portuguesa". In: NUNES, M. das G. V. (ed.), *Actas do V Encontro para o processamento computacional da língua portuguesa escrita e falada (PROPOR'2000)*, Atibaia, São Paulo: 131-140. 2000

BEYOND LEXICAL CHOICE IN PB AND PE: FROM THE “PARADA DE ÔNIBUS” TO THE “PARAGEM DE AUTOCARRO”

Abstract: *This paper presents a preliminary contrastive study on deverbal nouns suffixation in Brazilian and European Portuguese (BP and EP). Through tagged corpora analysis, we confirmed our previous hypothesis that in BP, deverbal suffix *_da* has a more comprehensive categorial function (namely noun and adjectival function) than in EP, in cases of movement verbs *virar* and *parar*.*

Keywords: *Contrastive Morphology; nominalization; Natural language processing.*