

DESCRIÇÃO DA FLEXÃO VERBAL REGULAR EM PORTUGUÊS: APLICAÇÃO DE UMA FERRAMENTA COMPUTACIONAL¹

LUIZ ARTHUR PAGANI²
(Univ. Est. de Londrina/UNICAMP)

0. INTRODUÇÃO

Através da utilização de recursos computacionais, a tarefa de formalização das teorias lingüísticas pode se tornar uma atividade menos árdua, devido à automatização do trabalho (que em sua maior parte passa a ser feito pelo computador) e mais precisa (pois a execução dos programas exige uma realização material), diferentemente da comprovação abstrata (*Gedanken-processing* - Shieber 1985: 192) que as formalizações lingüísticas anteriores exigiam.

Além disso, e além também de ampliar as possibilidades do processamento da linguagem natural, a implementação computacional das teorias lingüísticas permite ainda a comparação entre teorias lingüísticas rivais, não só em relação ao grau de “computabilidade” de cada uma, mas principalmente quanto à quantidade de recursos computacionais que cada teoria exige para sua implementação. Assim, do ponto de vista computacional, uma teoria lingüística seria mais econômica à medida que fosse mais “computável” e que exigisse métodos computacionais menos complexos.

Com a intenção de ilustrar essa idéia, será apresentada aqui uma implementação da morfologia flexional regular para os verbos do português, feita através de uma ferramenta computacional conhecida como “análise em dois níveis”, desenvolvida no início dos anos 80 por Kimmo Koskenniemi - um lingüista computacional finlandês. Essa tarefa foi facilitada pelo uso do PC-Kimmo, um programa elaborado pelo *Summer Institute of Linguistics*, que é uma interface para o desenvolvimento de descrições com análise em dois níveis, e que está descrito por Antworth 1990.

¹ As pesquisas e a redação desse texto foram desenvolvidas com o apoio da CAPES, através de uma bolsa de doutoramento. Devo agradecimentos ainda ao prof. dr. Angel Corbera Mori, que ministrou o curso em que estudei a Morfofonologia Lexical pela primeira vez e escrevi a primeira versão deste texto; agradeço também aos profs. drs. Edson Françaço (meu orientador) e Eleonora Cavalcante Albano, que leram e comentaram essa primeira versão. No entanto, nenhum deles é culpado dos erros que eu tenha insistido em cometer.

² Professor Assistente do Departamento de Letras Vernáculas e Clássicas da Universidade Estadual de Londrina e doutorando no programa de pós-graduação em Lingüística do Instituto de Estudos da Linguagem da Universidade Estadual de Campinas.

Essa ferramenta serve essencialmente para relacionar as etapas morfológica e fonológica da descrição lexical (os dois níveis). No nível subjacente, são definidos os encadeamentos morfológicos possíveis, por um processo bastante semelhante ao das antigas regras de estruturação sintagmática; já no nível superficial, as palavras são estabelecidas na forma efetiva com que elas aparecem nas manifestações lingüísticas concretas, através de regras parecidas com as velhas regras de alteração fonológica morfológicamente condicionadas.

Apesar de não apresentar nenhuma filiação com a morfofonologia lexical,³ ao oferecer um ambiente computacional para o estabelecimento de relações morfofonológicas, o PC-Kimmo pode funcionar como plataforma de teste principalmente para a questão da estratificação proposta pela morfofonologia lexical. O PC-Kimmo, do modo como está implementado enquanto ferramenta computacional, só permite a descrição de uma única camada; no entanto, a estratificação pode ser simulada pelo encadeamento seqüencial de descrições elaboradas em PC-Kimmo para cada estrato necessário. Ou seja, ainda que não seja possível representar diretamente com o PC-Kimmo toda a morfofonologia lexical, o PC-Kimmo oferece a possibilidade de representação individual de cada estrato.

Assim, questões controversas dentro da morfofonologia lexical, como a da quantidade de camadas de estratificação (Archangeli 1984: 9), encontram no PC-Kimmo um ambiente propício para sua avaliação. Com o presente trabalho, então, buscam-se evidências para a necessidade de uma estratificação complexa para a flexão verbal regular do português através da aplicação de uma ferramenta computacional: como com o PC-Kimmo só é possível representar uma camada de cada vez, se conseguirmos representar toda essa flexão num único algoritmo, isso significa que ela não exige mais do que apenas uma camada.

É claro, porém, que isso não é o suficiente ainda para invalidar uma teoria como a morfofonologia lexical; por outro lado, à medida que for possível expandir essa análise em dois níveis aos vários fenômenos morfofonológicos sem precisar postular qualquer estratificação, estaríamos nos aproximando gradualmente da desconfirmação de sua principal hipótese.

Para a morfologia verbal do português,⁴ o caminho mais simples pareceu ser o de começar pela flexão regular; a seguir deve-se passar para a descrição da flexão irregular

³ Não seria nada improvável que os pesquisadores de cada uma dessas áreas se encarassem até mesmo como opositores.

⁴ Preferi iniciar pela morfologia verbal, ao invés da morfologia nominal, devido principalmente à produtividade da flexão verbal em relação à nominal: com apenas um radical, enquanto a flexão verbal pode gerar até 73 formas (10 tempos com 6 pessoas: indicativo (presente, pretéritos imperfeito, perfeito e mais-que-perfeito, e futuros do presente e do pretérito), subjuntivo (presente, pretérito e futuro) e infinitivo pessoal; 2 tempos com 5 pessoas: imperativos afirmativo e negativo; e 3 formas impessoais: infinitivo impessoal, gerúndio e particípio passado - $10 \times 6 + 5 \times 2 + 3 = 73$), a flexão nominal só gera 12 formas (3 graus: normal, aumentativo e diminutivo; 2 gêneros: masculino e feminino; e 2 números: singular e plural - $3 \times 2 \times 2 = 12$). Além disso, do ponto de vista lingüístico, a flexão nominal exigiria ainda uma consideração sobre a natureza morfológica de um de seus componentes (para decidir se o grau é flexão ou derivação); já a flexão verbal parece não oferecer nenhuma grande controvérsia lingüística (mesmo a discussão sobre a posição estrutural da vogal temática, deixa de ter sentido numa implementação que só reconhece dois níveis).

e só depois para a da morfologia derivacional. Depois ainda, seria preciso repetir os mesmos passos para a morfologia nominal. Neste artigo, porém, apresenta-se apenas a primeira dentre todas essas etapas.⁵

Com o presente trabalho, então, dá-se início a este programa de “desconfirmação”.⁶

1. ANÁLISE EM DOIS NÍVEIS

O sistema de análise em dois níveis proposto por Koskeniemi consiste basicamente de dois módulos: um lexical, em que se lista os morfemas e se descreve sua morfotática, e um morfofonológico, onde as regras de adequação fonológica atuam todas ao mesmo tempo.

O módulo lexical é composto por um autômato de estado finito (*finite state automaton* - *FSA*), responsável por estabelecer uma estrutura seqüencial. Normalmente, esse tipo de autômato é representado através de uma rede de transição de estado finito (*finite state transition network* - *FSTN*), como a que aparece na Figura 1.

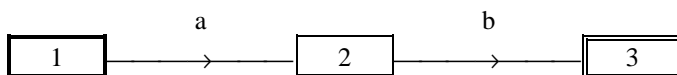


Figura 1: Rede de transição de estado finito para a seqüência *ab*

Nessas *FSTNs*, os quadrados numerados representam os estados pelos quais o autômato pode passar, enquanto os arcos rotulados determinam o sentido das transições de um estado para outro e as operações que devem ser executadas nessas transições. No exemplo da Figura 1, o estado 1 (marcado com uma borda larga) e 3 (marcado com a borda dupla) são, respectivamente os estados inicial e final do autômato; através desses dois estados, as aplicações apropriadas de operações são definidas como apenas aquelas que comecem no estado inicial e terminem no final. Através da *FSTN* da Figura 1, poderíamos gerar ou reconhecer apenas a seqüência *ab*, se considerarmos que a operação envolvida nas transições de estados é a de escrever ou ler seqüencialmente os símbolos contidos nos seus arcos.

O PC-Kimmo, no entanto, não tem como utilizar essas representações gráficas de *FSTNs*; seu módulo lexical só pode ser descrito através de uma linguagem especial. Nessa notação, definem-se as transições pela expressão *ALTERNATION* e as operações

⁵ Para evitar ainda a discussão desnecessária em relação principalmente à adequação da representação fonológica, ou mesmo de sua precisão fonética, optei por tratar as unidades ortograficamente; isso, no entanto, não deveria invalidar o principal objetivo deste artigo, que é o de mostrar como a análise em dois níveis é capaz de dar conta da flexão verbal regular, sem a postulação de qualquer estratificação entre eles.

⁶ O uso do termo “desconfirmação” não pretende sugerir que a refutação da morfofonologia lexical seja um objetivo desse texto; com ele, pretendia-se explicitar uma característica popperiana da metodologia empregada aqui, que consiste em buscar a evidência que refute a teoria - o que não significa necessariamente encontrá-la. Como foi dito acima, o objetivo é demonstrar que, pelo menos para a flexão verbal do português, a estratificação é desnecessária.

exigidas para as transições pela expressão LEXICON; o estado inicial é definido pela expressão INITIAL (que deve ser a primeira especificação de LEXICON), e o final é determinado por “#” . A linguagem exige ainda uma terceira expressão (END) que indica o término da descrição do autômato. Assim, o autômato da Figura 1 é descrito no PC-Kimmo pelo algoritmo da Figura 2.

```

ALTERNATION 2      B

LEXICON INITIAL
  a      2

LEXICON B
  b      #

END

```

Figura 2: Representação de uma FSTN⁷ no PC-Kimmo

Como o próprio Antworth (1990: 13) reconhece, no entanto, esse módulo lexical não é capaz de operar economicamente com alguns tipos de restrições morfológicas, tais como o fato da inexistência de **richment*, apesar da existência de *enrich* e *enrichment*; esse fato só poderia ser expresso indiretamente, através de um percurso que passasse primeiro por *en-* e depois pelos radicais adjetivais que, juntos, permitem a continuação com *-ment*, além de um outro percurso iniciado nos adjetivos, seguido por *-ment* - mas isso causaria uma duplicação dos radicais adjetivais.

Já o módulo morfofonológico, mais consistente que o anterior, é composto por regras de adaptação fonológica condicionadas morfofonologicamente. No entanto, ao contrário da fonologia gerativa, essas regras atuam em paralelo⁸ (sem a exigência de nenhum ordenamento) e têm acesso tanto ao nível subjacente quanto ao nível lexical.⁹

⁷ Essa notação, oferece uma espécie de abreviação para a representação de FSTNs mais complexos (uma notação bastante parecida com esta é apresentada em Gazdar & Mellish 1989: 27-36). Para que o FSA da Figura 2 opere com as seqüências *ab*, *ad*, *cb* e *cd*, bastaria incluirmos duas expressões:

```

ALTERNATION 2      B

LEXICON INITIAL
  a      2
  c      2

LEXICON B
  b      #
  d      #

END

```

⁸ Na verdade, devido à configuração serial dos computadores, o PC-Kimmo apenas simula uma operação paralela; contudo, como esta é uma restrição técnica, ela não invalida a postulação da análise em dois níveis como uma máquina paralela.

⁹ No PC-Kimmo, o termo *subjacente* designa as formas através das quais as palavras são constituídas, mas que nem sempre se realizam da mesma maneira. Já o termo *lexical* designa as realizações das formas

Ainda que diferentemente, esse módulo opera com regras muito parecidas com as da fonologia gerativa, do tipo “ $a \rightarrow b / c _ d$ ” (que poderia ser lida como: a se torna b quando vier imediatamente depois de c e antes de d). Mas como na análise de dois níveis qualquer regra tem acesso a ambos os níveis, a adaptação contextual é representada da seguinte maneira: “ $a:b \Rightarrow c:d _ e:f$ ”; essa regra estabelece a correspondência entre um a subjacente e um b superficial, quando eles ocorrem entre um c e um e subjacentes que correspondem, respectivamente, a um d e a um f superficiais. Uma outra distinção entre a análise de dois níveis e a fonologia gerativa é que nesta as regras eram apenas de reescrita; já a primeira reconhece quatro operadores que podem relacionar diferentemente o corpo e a cabeça das regras: 1) “ \Rightarrow ” (a correspondência apenas se realiza no contexto, ou seja, a correspondência não é permitida a não ser no contexto), 2) “ \Leftarrow ” (a correspondência sempre se realiza no contexto, ou seja, não se permite outra correspondência no contexto), 3) “ \Leftrightarrow ” (uma combinação dos dois operadores anteriores), e 4) “ \nrightarrow ” (a correspondência jamais pode se realizar no contexto).

No PC-Kimmo, essas regras de correspondência também são descritas através de FSA, mas como esses autômatos de correspondência são na verdade transdutores de estado finito (*finite state transducer - FST*), elas não são descritas através de uma linguagem especial, mas sim através de tabelas de estado finito.

Numa tabela de estado finito determina-se todos os passos que uma operação pode realizar; assim, a regra “ $a:b \Rightarrow c:d _ e:f$ ” é representada pela tabela de estado finito da Figura 3.¹⁰

| | c | a | e | @ |
|----|---|---|---|---|
| | d | b | f | @ |
| 1: | 2 | 0 | 1 | 1 |
| 2: | 0 | 3 | 1 | 1 |
| 3: | 0 | 0 | 1 | 1 |

Figura 3: Tabela de estado finito para " $a:b \Rightarrow c:d _ e:f$ "

subjacentes. Poderíamos dizer que as formas subjacentes são abstratas, enquanto que as formas lexicais correspondem às palavras efetivamente realizadas.

¹⁰ Não nos deteremos muito em explicações sobre o funcionamento das tabelas de estado finito, já que o PC-Kimmo vem com um utilitário (Kgen) que monta automaticamente essas tabelas a partir de regras escritas no formato “ $a:b \Rightarrow c:d _ e:f$ ”.

2. FLEXÃO REGULAR DO VERBO EM PORTUGUÊS

Adaptando a descrição de Câmara Jr. (1970: 104) para a formação dos verbos em português, já que a análise em dois níveis não exige uma estruturação sintagmática além da ordem linear,¹¹ podemos dizer que um verbo se compõe a partir de um radical que adquire primeiro uma vogal temática, depois um sufixo modo-temporal e finalmente um sufixo número-pessoal. Desse modo, poderíamos dizer que uma forma verbal como “cantavas” seria composta de um radical “cant”, de uma vogal temática “a”, de um sufixo modo-temporal “va” e de um sufixo número-pessoal “s”.

Para esta implementação, assumimos que a vogal temática poderia se realizar como “a”, “e” ou “i”.¹² Como sufixos modo-temporais, assumimos: “o” (para o presente e para o perfeito do indicativo), “va” (para o pretérito imperfeito do indicativo), “ra” (para o pretérito mais-que-perfeito do indicativo), “rá” (para o futuro do presente do indicativo), “ria” (para o futuro do pretérito do indicativo), “e” (para o presente do subjuntivo e para os imperativos), “sse” (para o pretérito do subjuntivo), “r” (para o futuro do subjuntivo e para os dois infinitivos), “do” (para o particípio) e “ndo” (para o gerúndio). Como sufixos de número e pessoa, foram considerados: “o” (para a primeira pessoa do singular), “s” (para a segunda pessoa do singular), “0” (para a terceira pessoa do singular), “mos” (para a primeira pessoa do plural), “is” (para a segunda pessoa do plural) e “m” (para a terceira pessoa do plural); como a flexão do pretérito imperfeito do indicativo não segue este padrão, são necessários ainda mais seis sufixos exclusivamente para a formação deste tempo: “i” (para a primeira pessoa do singular), “ste” (para a segunda pessoa do singular), “u” (para a terceira pessoa do singular), “mos” (para a primeira pessoa do plural), “stes” (para a segunda pessoa do plural) e “ram” (para a terceira pessoa do plural).¹³

No entanto, apenas isso não é suficiente para dar conta de toda a flexão verbal, pois certas formas estão sujeitas a alguma adaptação morfofonológica. Da maneira como estão postulados os morfemas, se quiséssemos gerar a primeira pessoa do pretérito do subjuntivo, obteríamos “cantassemos”, e não “cantássemos”. Por isso, precisamos ainda de regras que adequem as concatenações feitas pelo módulo morfológico. No nosso exemplo, a regra precisaria dizer que a vogal temática “a” é substituída por “á” (a:á) diante do morfema do pretérito do subjuntivo (“sse”) seguido pelo morfema da primeira pessoa do plural (“mos”).

¹¹ Em sua descrição, Câmara Jr. considera o verbo constituído por um tema, formado pelo radical mais uma vogal temática, e por um sufixo flexional, formado pelo sufixo modo-temporal e pelo sufixo número-pessoal; assim, obtém-se uma estrutura mais complexa, do tipo “R (T + VT) + SF (SMT + SNP)”. Como o módulo morfológico do PC-Kimmo suporta apenas estruturação seqüencial, essa adaptação visa a eliminar a hierarquização proposta por Câmara Jr.

¹² Felizmente, apesar dos poucos recursos morfofônicos, o PC-Kimmo dispõem de um mecanismo simples de subcategorização que permite captar a dependência que existe entre o radical e a vogal temática, de forma a impedir que um radical como “cant” adquirisse outra vogal temática que não o “a”.

¹³ Para ser exaustivo, seria ainda preciso mencionar o imperativo, mas para evitar muita repetição passaremos às considerações sobre a necessidade das adaptações morfofonológicas, deixando a questão dos imperativos mais para diante.

Para conseguir recorrer a essas informações morfológicas, é preciso incluir as marcas de fronteira nos morfemas. No PC-Kimmo isso é feito incluindo-se, antes de cada morfema, o sinal “+”. Ainda no nosso exemplo, os morfemas envolvidos seriam representados como: “cant”, “+a”, “+sse” e “+mos”; assim, sua concatenação resultaria na forma subjacente “cant+a+sse+mos”, que seria transformada em “cantássemos” por duas regras: “+:0” (todo “+” subjacente desaparece no nível lexical) e “a:á <=> +:0 __ +:0 sse +:0 mos #” (que poderia ser lida da seguinte maneira: um “a” subjacente se torna um “á” exclusivamente quando ele é um morfema, seguido pelo morfema “sse”, por sua vez também seguido pelo morfema “mos”, e pela marca de final de palavra - “#”).

3. FORMALIZAÇÃO

Para a implementação de uma descrição através do PC-Kimmo, são necessários dois arquivos: um com as regras de condicionamento morfofonológico e outro com as unidades morfológicas e sua organização seqüencial.

3.1. Desenvolvendo o arquivo lexical

Ao contrário do que é sugerido no manual do PC-Kimmo,¹⁴ iniciaremos pelo arquivo de itens lexicais. Para essa implementação, na verdade, o módulo morfológico é composto por dois arquivos (VERB_RAD.LEX e VERB_MRF.LEX) concatenados a um principal (VERB_ALT.LEX), através de uma facilidade do PC-Kimmo (o operador INCLUDE).

O primeiro passo é estabelecer, através do operador ALTERNATION, as possibilidades de seqüenciação que os morfemas apresentam entre si. Nossa implementação aparece na Figura 4, onde se estipula que a concatenação começa pela escolha do radical (ALTERNATION Início RADICAL).¹⁵

A partir do radical, pode-se chegar à vogal temática (ALTERNATION Radical_x VTx).¹⁶ Como as vogais temáticas apresentam uma distribuição complementar (ou seja, cada radical só pode ser seguido por apenas uma das vogais

¹⁴ Antworth 1990 sugere iniciar qualquer implementação pelo chamado arquivo de regras, onde se registram as regras de condicionamento morfofonológico, pois com este arquivo apenas já é possível gerar as formas lexicais da língua que se pretende descrever. Contudo, como a motivação inicial desse trabalho era o reconhecimento, optou-se por iniciar com o arquivo lexical.

¹⁵ O procedimento começa efetivamente pelo operador LEXICON INITIAL, que só então invoca o operador ALTERNATION Início RADICAL. Preferi este padrão, ao invés de iniciar diretamente com o radical, a fim de facilitar eventuais expansões da implementação: para incluirmos a morfologia derivacional, por exemplo, basta inserir as declarações adequadas entre Início e RADICAL e as unidades lexicais apropriadas. O operador LEXICON para o argumento FIM serve apenas para inserir o marcador de limite de palavra; o final do procedimento é determinado pelo operador END.

¹⁶ A variável x, introduzida aqui, não tem qualquer utilização dentro do PC-Kimmo, e só está sendo usada como metavariável - para a abreviação das três regras específicas a cada conjugação.

temáticas), elas são selecionadas individualmente pela categoria de continuação do radical escolhido (isso é determinado pelas linhas 2-4 do algoritmo).¹⁷

```

ALTERNATION Início          RADICAL
ALTERNATION Radical_1      VT1
ALTERNATION Radical_2      VT2
ALTERNATION Radical_3      VT3
ALTERNATION Vogal_Temática SMT
ALTERNATION Tempo-Modo     SNP
ALTERNATION Perfeito       SNP ( PERF )
ALTERNATION Imperativo     SNP ( IMP_AF ) SNP ( IMP_NG )
ALTERNATION Número-Pessoa  FIM

LEXICON INITIAL
0      Início      "[ "

INCLUDE verb_rad.lex

INCLUDE verb_mrf.lex

LEXICON FIM
0      #           "]" "

END

```

Figura 4: Seqüencição dos morfemas da flexão verbal (VERB_ALT.LEX)¹⁸

Logo após a vogal temática, podemos obter o morfema de tempo e modo (ALTERNATION Vogal_Temática SMT). Aqui também encontramos alguma complexidade para as categorias de continuação: se a maioria dos tempos admitem 6 pessoas, os imperativos só admitem 5; além disso os morfemas número-pessoais dos imperativos são diferentes não só entre si, mas completamente distintos (com exceção da primeira pessoa do plural) dos outros tempos. Por outro lado, os morfemas de número e pessoa do pretérito perfeito do indicativo também são distintos dos demais. Esse fato está declarado nas linhas 6-8 do algoritmo. Finalmente, depois de adquirir o sufixo número-pessoal, o procedimento é encerrado.

Apenas a título de exemplificação, o arquivo de radicais (VERB_RAD.LEX) só contém três verbos regulares (“cantar”, “vender” e “partir”), um de cada conjugação. O conteúdo desse arquivo pode ser visto na Figura 5; cada um dos radicais apresenta em sua categoria de continuação (segunda coluna) o indicador adequado para a seleção de sua respectiva vogal temática (Radical_1 seleciona VT1; Radical_2 seleciona VT2; e Radical_3 seleciona VT3), conforme já havíamos mencionado acima. A terceira coluna se destina apenas a uma descrição (chamada de glosa), do elemento a

¹⁷ A “categoria de continuação” é definida na segunda coluna dos operadores de tipo LEXICON, como veremos mais adiante.

¹⁸ A expressão INCLUDE, mencionada aqui, serve apenas para concatenar arquivos.

ser inserido; devido a sua importância secundária neste artigo, não nos deteremos sobre ela.

| LEXICON RADICAL | | |
|-----------------|-----------|------------------|
| cant | Radical_1 | "VERBO(cantar) " |
| vend | Radical_2 | "VERBO(vender) " |
| part | Radical_3 | "VERBO(partir) " |

Figura 5: Arquivo de radicais verbais (VERB_RAD.LEX)

O arquivo mais importante no módulo morfológico é o das unidades morfológicas (VERB_MRF.LEX), apresentado na Figura 6. Depois de selecionado o radical, através de sua categoria de continuação e da regra de alternância adequada (por exemplo, ALTERNATION Radical_1 VT1, na Figura 4), o autômato escolhe a vogal temática apropriada (nesse caso, “+a”). Depois da vogal temática, o autômato está habilitado a procurar um sufixo modo-temporal (ALTERNATION Vogal_Temática SMT, na Figura 4); dessa forma, ele pode escolher qualquer morfema da lista sob LEXICON SMT (na Figura 6). Para a maior parte dos morfemas, a categoria de continuação é Tempo-Modo, o que significa que o sufixo de tempo e modo já foi incorporado e o autômato pode buscar um morfema número-pessoal (ALTERNATION Tempo-Modo SNP, na Figura 4); sob LEXICON SNP (na Figura 6), estão os morfemas de número e pessoa considerados regulares.

Para o morfema do pretérito perfeito do indicativo, no entanto, a categoria de continuação é Perfeito; junto com a regra ALTERNATION Perfeito SNP(PERF) (na Figura 4), podemos obter os morfemas número-pessoais característicos desse tempo, completamente distintos dos outros (mais uma vez excetuando-se a primeira pessoa do plural). Esses morfemas estão discriminados sob LEXICON SNP (PERF) (na Figura 6).

| | | |
|-------------|-------------------|----------------|
| LEXICON VT1 | | |
| +a | Vogal_Temática "+ | 1ª CONJ " |
| LEXICON VT2 | | |
| +e | Vogal_Temática "+ | 2ª CONJ " |
| LEXICON VT3 | | |
| +i | Vogal_Temática "+ | 3ª CONJ " |
| LEXICON SMT | | |
| +0 | Tempo-Modo | " + PRES IND " |
| +0 | Perfeito | " + PERF IND " |
| +va | Tempo-Modo | " + IMPF IND " |
| +ra | Tempo-Modo | " + MQPF IND " |
| +rá | Tempo-Modo | " + FTPR IND " |
| +ria | Tempo-Modo | " + FTPT IND " |
| +e | Tempo-Modo | " + PRES SUB " |
| +sse | Tempo-Modo | " + PRET SUB " |
| +r | Tempo-Modo | " + FUTR SUB " |
| +r | Tempo-Modo | " + INF FLEX " |
| +e | Imperativo | " + IMP " |

| | | | | |
|---------|----------------|-----|-----------------------|--------|
| +r | Número-Pessoa | " + | INF | " |
| +do | Número-Pessoa | " + | PTC | " |
| +ndo | Número-Pessoa | " + | GRD | " |
| LEXICON | SNP | | | |
| +o | Número-Pessoa | " + | 1 ^a P | SING " |
| +s | Número-Pessoa | " + | 2 ^a P | SING " |
| +0 | Número-Pessoa | " + | 3 ^a P | SING " |
| +mos | Número-Pessoa | " + | 1 ^a P | PLUR " |
| +is | Número-Pessoa | " + | 2 ^a P | PLUR " |
| +m | Número-Pessoa | " + | 3 ^a P | PLUR " |
| LEXICON | SNP (PERF) | | | |
| +i | Número-Pessoa | " + | 1 ^a P | SING " |
| +ste | Número-Pessoa | " + | 2 ^a P | SING " |
| +u | Número-Pessoa | " + | 3 ^a P | SING " |
| +mos | Número-Pessoa | " + | 1 ^a P | PLUR " |
| +stes | Número-Pessoa | " + | 2 ^a P | PLUR " |
| +ram | Número-Pessoa | " + | 3 ^a P | PLUR " |
| LEXICON | SNP (IMP_AF) | | | |
| +a | Número-Pessoa | " + | AFIR 2 ^a P | SING " |
| +0 | Número-Pessoa | " + | AFIR 3 ^a P | SING " |
| +mos | Número-Pessoa | " + | AFIR 1 ^a P | PLUR " |
| +ai | Número-Pessoa | " + | AFIR 2 ^a P | PLUR " |
| +m | Número-Pessoa | " + | AFIR 3 ^a P | PLUR " |
| LEXICON | SNP (IMP_NG) | | | |
| +s | Número-Pessoa | " + | NEGT 2 ^a P | SING " |
| +0 | Número-Pessoa | " + | NEGT 3 ^a P | SING " |
| +mos | Número-Pessoa | " + | NEGT 1 ^a P | PLUR " |
| +is | Número-Pessoa | " + | NEGT 2 ^a P | PLUR " |
| +m | Número-Pessoa | " + | NEGT 3 ^a P | PLUR " |

Figura 6: Arquivo de unidades morfológicas (VERB_MRF.LEX)

Também o modo imperativo exige um tratamento especial. Tendo escolhido “+e” como seu morfema, precisamos de mais duas listas de morfemas número-pessoais: uma para o afirmativo e outra para o negativo (ambas com cinco pessoas apenas, já que não existe imperativo para a primeira pessoa).¹⁹ A regra que garante a escolha dessas listas é ALTERNATION Imperativo SNP (IMP_AF) SNP (IMP_NG) (na Figura 4); as duas categorias SNP (IMP_AF) e SNP (IMP_NG) são caminhos opcionais que o autômato pode seguir.²⁰ As duas listas de morfemas estão, respectivamente, sob LEXICON SNP (IMP_AF) e sob SNP (IMP_NG), na Figura 6.

¹⁹ No momento da redação desse trabalho, me pareceu de alguma arbitrariedade a escolha de “+e” como morfema para o imperativo, principalmente porque suas listas de morfemas número-pessoais são bastante diferentes, tanto entre si como em relação aos outros tempos. Uma alternativa seria escolher o morfema zero para os imperativos e modificar as regras de condicionamento morfofonológico para essa nova situação; contudo, essa solução não parece menos arbitrária. Porém, como a revisão da morfologia flexional do verbo em português não faz parte dos objetivos desse artigo, essa questão não será discutida aqui.

²⁰ No PC-Kimmo, elementos separados por pelo menos um espaço na posição de segundo argumento do operador ALTERNATION designam estados possíveis de serem percorridos numa *FSTN*.

Finalmente, as três formas impessoais, que não exigem morfemas de número e pessoa, apresentam como categoria de continuação Número-Pessoa; exatamente como para as outras formas depois de agregarem seu morfema número-pessoal, a regra de alternância para essa categoria de continuação (ALTERNATION Número-Pessoa FIM, na Figura 4) é a última operação do autômato. Nesse ponto, a determinação de qualquer forma subjacente está concluída. A seguir essa forma precisa ser submetida às regras de condicionamento morfofonológico.

3.2. Desenvolvendo o arquivo de regras

Um arquivo de regras do PC-Kimmo, responsável pelas operações de adequação condicionadas pelo contexto morfofonológico, precisa conter necessariamente algumas especificações iniciais. A primeira especificação (ALPHABET) define os caracteres que poderão ser operados pelo programa; a seguir é preciso declarar o elemento nulo (NULL 0), o “curinga” (ANY @)²¹ e o marcador de fronteira (BOUNDARY #). Logo a seguir, definem-se as regras de correspondência mais genéricas; ou seja, as que dizem que, na falta de alguma outra especificação, um caracter no nível subjacente normalmente corresponde ao mesmo caracter no nível lexical.

As regras de condicionamento começam, efetivamente, depois dessas especificações iniciais; para essa implementação da morfologia flexional regular dos verbos do português, foram necessárias apenas 20 regras, como podemos ver na Figura 7.²² A função da primeira destas regras (regra 4)²³ é de apagar a vogal “a”: 1) quando ela for vogal temática, e estiver ou diante do “e” do presente do subjuntivo ou diante da primeira pessoa do singular do presente do indicativo, e 2) do início dos morfemas número-pessoais, nas segundas pessoas do imperativo afirmativo, para a segunda e a terceira conjugações. (As outras adaptações - apagamento das vogais temáticas “e” e “i” - são feitas por outras regras - regras 17 e 23, respectivamente.)

A regra 5 visa a apagar o morfema “o”, de primeira pessoa, para todos os tempos do subjuntivo (“e”, “sse” e “r”), para os pretéritos imperfeito (“va”) e mais-que-perfeito (“ra”), e futuro do pretérito (“ria”), todos do indicativo, e finalmente para o infinitivo flexionado (“r”). Ela ainda depende das regras 18 (que transforma o “v” em “i”, como veremos mais adiante) e 21 (que faz o “e”, do subjuntivo, virar “a”, como também veremos adiante).

²¹ Esse elemento funciona como uma variável; ele serve para designar o conjunto de caracteres e através dele é possível garantir a aplicação simultânea de todas as regras.

²² As regras específicas de condicionamento começam na regra 4 - as três primeiras regras declaram as correspondências genéricas para as consoantes (regra 1), para as vogais (regra 2) e para os caracteres especiais (regra 3) - e terminam na regra 23.

²³ Nas regras do PC-Kimmo, os parêntesis indicam opcionalidade e a barra vertical denota disjunção; assim, uma instrução como “a (b)” designa as seqüências “a e ab”, já “a | b” significa ou “a” ou “b”. Os colchetes, por fim, determinam o escopo de uma operação; desse modo, uma instrução como “[a b] | c” é distinta de “a [b | c]”, pois a primeira significa ou “ab” ou “c”, enquanto a segunda denota ou “ab” ou “ac”.

```

RULE "4 a:o <=> ___ +:0 [ +:0 o # | e:@ +:0
] | [ e:@ | i:@ ] +:0 e:@ +:0 ___ ( i ) #"

RULE "5 o:0 <=> +:0 [ [ v:@ | r ( i ) ] a |
( s s ) e:@ | r ] +:0 ___ #"

RULE "6 a:á <=> ___ +:0 [ [ v | r ] a:@ | s
s e ] +:0 [ m o | i ] s #"

RULE "7 a:e <=> +:0 [ v:@ | r ( i:@ ) ] ___
+:0 i s # | ___ +:0 +:0 i # | e:@ +:0 e:@
+:0 ___ ( i ) #"

RULE "8 a:o <=> ___ +:0 +:0 u #"

RULE "9 á:e <=> +:0 r ___ +:0 [ o:@ | [ m o
| i ] s ] #"

RULE "10 o:i <=> +:0 r á:@ +:0 ___ #"

RULE "11 á:ã <=> +:0 r ___ +:0 m:@ #"

RULE "12 m:o <=> +:0 r á:@ +:0 ___ #"

RULE "13 i:í <=> +:0 r ___ a:@ +:0 [ m o |
i ] s # | ___ +:0 [ r a:@ | s s e ] +:0 [ m
o | i ] s #"

RULE "14 0:e <=> +:0 r ___ +:0 [ s | m ] #"

RULE "15 0:d <=> +:0 r +:0 ___ i:@ s #"

RULE "16 i:e <=> +:0 r +:0 0:@ ___ s # |
___ +:0 +:0 ( [ s | m ] ) #"

RULE "17 e:0 <=> a:@ +:0 ___ +:0 a ( i ) #
| ___ +:0 [ +:0 [ o | i ] # | [ v:@ a | e:@
] +:0 ] | i:@ +:0 ___ +:0 a:@ i #"

RULE "18 v:i <=> [ e:@ | i:@ ] +:0 ___ a
+:0 ( [ o:@ | s | m ] ) #"

RULE "19 v:í <=> [ e:@ | i:@ ] +:0 ___ a:@
+:0 [ m o | i ] s #"

RULE "20 e:ê <=> ___ +:0 [ r a:@ | s s e ]
+:0 [ m o | i ] s #"

RULE "21 e:a <=> [ e:@ | i:@ ] +:0 ___ +:0
( [ o:@ | ( [ m o | i ] ) s | m ] ) #"

RULE "22 e:i <=> ___ +:0 d o #"

RULE "23 i:0 <=> ___ +:0 [ +:0 [ o | i ( s
) ] # | [ v:@ a:@ | e:@ ] +:0 ]"

```

Figura 7: Regras de condicionamento morfofonológico (VERBO.RUL - parte)

A regra 6 troca a vogal temática “a” por “á”, na primeira e na segunda pessoa (“mos” e “is”) dos pretéritos imperfeito (“va”) e mais-que-perfeito (“ra”) do indicativo, e do pretérito do subjuntivo (“sse”). Essa regra está relacionada ainda com a regra 7, que transforma em “e” o “a” final dos pretéritos imperfeito (“va”) e mais-que-perfeito (“ra”), e do futuro do pretérito (“ria”), para a segunda pessoa do plural desses tempos, e a vogal temática “a” para a primeira pessoa do singular (“i”) do pretérito perfeito do indicativo (“0”). A regra 7, por sua vez, depende ainda das regras 13 (i:i) e 19 (v:i).

A regra 8 se destina exclusivamente a adequar a terceira pessoa do singular (“u”) do pretérito perfeito do indicativo (“0”). Já as regras 9-12, destinam-se a adequar o futuro do presente do indicativo (“rá”): a 9 muda o “á” para “e”, nas primeiras pessoas (“o” e “mos”) e na segunda pessoa do plural (“is”); a 10 passa o “o” de primeira pessoa para “i”; 11 e 12, finalmente, atuam juntas para formar a terceira pessoa do plural (“m”): a primeira troca o “á” por “ã” e a segunda o “m” por “o”.

Ainda para a primeira e para a segunda pessoa do plural, a regra 13 transforma em “f” a vogal temática da terceira conjugação (“i”), no pretérito mais-que-perfeito do indicativo e no pretérito do subjuntivo, e o “i” do futuro do pretérito do indicativo (“ria”).

As regras 14 e 15 inserem no nível lexical elementos de ligação que não existiam no nível subjacente: a primeira coloca um “e” antes da segunda pessoa do singular (“s”) ou da terceira do plural (“m”), no futuro do subjuntivo e no infinitivo flexionado (“r” em ambos); a segunda agrega um “d” antes da segunda pessoa do plural (“is”) para esses mesmos tempos. Além disso, as regras 15 e 16 estão relacionadas entre si, pois esta ainda precisa mudar de “i” para “e” a primeira vogal do morfema de segunda pessoa do plural.

Pela regra 17, apaga-se a vogal temática “e” diante da primeira pessoa do singular do presente e do pretérito perfeito do indicativo, e diante de todo o pretérito imperfeito do indicativo, junto com as regras 18 e 19; essa mesma vogal é eliminada, em conjunto com a regra 21, para todo o presente do subjuntivo. Também o morfema “e” de imperativo é apagado na segunda pessoa do plural, para a primeira e para a terceira conjugação, e na segunda pessoa do singular apenas, para a primeira conjugação; para completar essas formas, o autômato ainda precisa das regras 4 e 23.

Com as regras 18 e 19, prepara-se o pretérito imperfeito do indicativo (“va”) para a segunda e para a terceira conjugação: pela primeira, o “v” inicial torna-se “i”, em todas as pessoas do singular (para a primeira ainda é preciso apagar o “o”, através da regra 5) e na terceira do plural; pela segunda, o “v” muda para “f”, na primeira e na segunda pessoa do plural. Em todas as situações, ainda é necessário cancelar as vogais temáticas “e” (com a regra 17) ou o “i” (com a regra 23).

Pela regra 20, acentua-se a vogal temática “e”, para a primeira e para a segunda pessoa do plural (“mos” e “is”), no pretérito mais-que-perfeito do indicativo (“ra”) e no pretérito do subjuntivo (“sse”). Já com a regra 21, forma-se o presente do subjuntivo da segunda e da terceira conjugação: o morfema “e” vira “a” em todas as pessoas; além disso, o “o” de primeira pessoa (regra 5) e as vogais temáticas “e” e “i” (regras 17 e 23) são apagados.

A única função da regra 22 é garantir que, no particípio da segunda conjugação, a vogal temática “e” vire um “i”. E a regra 23, finalmente, cancela a vogal temática “i” para a primeira pessoa do singular e para a segunda do plural no presente do indicativo, e também para a primeira pessoa do singular do pretérito perfeito do indicativo; ela é apagada ainda em todo o pretérito imperfeito (junto com a atuação das regras 7, 18 e 19) e em todo o presente do subjuntivo (em conjunto com a regra 21); ela desaparece ainda nas segundas pessoas do imperativo afirmativo (em atuação conjunta com as regras 4 e 17).

4. CONCLUSÃO

A principal conclusão desse trabalho é a de que a flexão verbal regular, como sugere a morfofonologia lexical, pode ser completamente definida em um único estrato; isso se constata pois todas as formas verbais regulares do português puderam ser reconhecidas (e geradas) nessa implementação através do PC-Kimmo, que não comporta mais do que um estrato morfofonológico.

As próximas etapas no sentido de testar a necessidade de estratificação, conforme postulada pela morfofonologia lexical, exigem a inclusão gradativa de fenômenos de irregularidade cada vez mais acentuada à descrição apresentada. Assim, primeiro precisaríamos incluir paradigmas verbais que poderiam ser chamados de regularmente irregulares, como o caso dos verbos cujos radicais terminem em “c”, que diante de “e” e “i” vira “qu”, ou em “ç”, que diante de “e” vira “c”; em seguida, precisaríamos lidar com modificações menos regulares, como a de “ferir” ou “divergir”, que apresentam dois radicais: “fer” e “fir”, para o primeiro, e “divirj” e “diverg” para o segundo.²⁴ Casos cada vez mais complexos podem ser representados ainda pelos verbos auxiliares, como “ser” e “estar”, que apresentam até mais de dois radicais e formas distintas para o futuro do subjuntivo e para o infinitivo flexionado,²⁵ e pelos verbos defectivos.

Para os regularmente irregulares, ainda não parece ser necessário postular qualquer estratificação; e, à primeira vista, parece ser possível incluir na descrição regras que operariam facilmente com essas formas. Para o primeiro exemplo (“ficar”), bastaria acrescentar duas regras para garantir a mudança do “c”, no final do radical, para “q” e a inclusão do “u”, após o “q”, como: 1) c:q <=> __ 0:u +:0 [a:e | a:0 +:0 e +:0], e 2) 0:u <=> c:q __.²⁶ Mesmo para os casos de irregularidade de segundo grau

²⁴ A dificuldade a que nos referimos no verbo “divergir” está na oscilação entre “i” e “e”, pois a alternância entre “g” e “j” poderia ser considerada regularmente irregular. Por outro lado, parece haver uma ligação entre os dois fenômenos: o “i” só co-ocorre com o “j”, enquanto que o “e” aparece apenas junto com o “g”.

²⁵ Apenas os quatro verbos chamados de auxiliares (“ser”, “estar”, “ter” e “haver”) apresentam formas diferentes para esses dois tempos; para todos os outros verbos, até os mais irregulares, essas formas são sempre as mesmas.

²⁶ Assim como está, essa regra provavelmente não funcionaria perfeitamente na descrição proposta, pois ela foi elaborada exclusivamente a partir de um caso específico; como as interações entre as regras são abundantes no PC-Kimmo, a inclusão de qualquer regra exige um trabalho de casamento dos efeitos das regras que interagem entre si.

(“ferir” e “divergir”), ainda que não apresentemos aqui nem sequer um esboço, ainda parece viável imaginar que seja possível elaborar regras que relacionem e:i, para o primeiro verbo, e g:j (além de correlacioná-las entre si), para o segundo.

Já as irregularidades mais profundas não despertam o mesmo otimismo. Para esses casos ainda não conseguimos entrever qualquer possibilidade de solução através da análise de dois níveis que o PC-Kimmo oferece.²⁷

Contudo, as tarefas foram apenas iniciadas e muita coisa ainda precisa ser feita. Cada uma das etapas mencionadas exige um grande esforço, para o cumprimento das quais são necessários longos períodos de trabalho quase exclusivamente artesanal.

REFERÊNCIAS

- ANTWORTH, Evan L. **PC-Kimmo: A Two-Level Processor for Morphological Analysis**. Dallas, Texas: Summer Institute of Linguistics. *Occasional Publications in Academic Computing*, 16; 1990.
- ARCHANGELI, Diana. An overview of the theory of lexical phonology and morphology. In Margaret Speas & Richard Sproat (eds.), **Papers from the January 1984 MIT Workshop in Morphology**, Cambridge, Massachusetts: Department of Linguistics and Philosophy. *MIT Working Papers in Linguistics*, 7; 1984.
- CÂMARA Jr., Joaquim Mattoso. **Estrutura da Língua Portuguesa**. Petrópolis: Vozes, 1970.
- GAZDAR, Gerald & MELLISH, C. S. **Natural Language Processing in Prolog: An Introduction to Computational Linguistics**. Wokingham: Addison-Wesley, 1989.
- SHIEBER, Stuart M. Criteria for designing computer facilities for linguistic analysis. **Linguistics**, 23 (2); 1985:189-211.

²⁷ Porém, mesmo que ela envolva estratificação, isso não implica que o PC-Kimmo deixe ser usado como ferramenta de formalização computacional: à medida que forem sendo confirmadas as necessidades de outros estratos, podemos simulá-los como várias análises de dois níveis operando seqüencialmente; ou seja, cada nível poderia ser representado por uma descrição diferente.