

## TRATAMENTO DE TEXTOS EM COMPUTADOR

### ANÁLISE LINGÜÍSTICA

O aproveitamento do computador para análise linguística de um texto pode processar-se em duas fases distintas segundo haja ou não intervenção humana interpretando, e codificando, a cadeia de significantes.

A exploração do ficheiro, na sua constituição inicial de unidades ou conjuntos de caracteres separados por brancos permite obter imediatamente um índice apreciável de informações. Por simples indexação, e contanto que os dados estejam convenientemente referenciados, o computador poderá, com meros programas de rotina, responder a questões como: número de palavras por frase (período) do texto; distribuição dos resultados assim obtidos por sequência natural, por linha ou verso (em textos versificados), por ordem de capítulo ou de página; frequências de palavras no conjunto do texto ou por secções; frequências de caracteres por palavra (com distribuição ou não por capítulo, secção do texto ou por localização na frase — posição inicial, final, ou média), etc..

Não se pense que tais dados quantitativos são irrelevantes como informação linguística. Sabe-se, efectivamente que há uma relação directa entre frequência e estrutura fónica (as palavras mais frequentes são as menos longas; uma série de ocorrências marcadamente longas — por ex., nos versos alexandrinos — não pode deixar de ser significativa), entre frequência e natureza fónica (nem todas as línguas apresentam a mesma frequência para os mesmos fonemas; a ocorrência anormal de determinado fonema num texto ou numa secção deste constitui dado significativo), entre frequência e etimologia ou diacronia (as palavras mais antigas na língua são as mais frequentes), entre frequência e classe gramatical (as chamadas palavras gramaticais perfazem cerca de metade de um texto), etc.... A frequência faz parte da natureza do significante linguístico como dele fazem parte as suas propriedades diacríticas (distintivas) ou semânticas. Afinal, a convencionalidade do signo e a intercomunicabilidade lingüís-

tica supõem e implicam fenómenos de recorrência na cadeia significativa e de reconhecimento por parte dos utentes.

A análise, acessível ao computador a este nível de quantificação, não o é logo que a delimitação de unidades passa pela determinação de oposições de ordem funcional. Aí terá de intervir o linguista, embora, após uma definição de critérios e uma codificação de dados, a máquina possa continuar e desenvolver, por comparação, a análise encetada, ou proceder a recolhas em condições optimizadas (tempo, contexto, fiabilidade...) que facilitam a determinação de constantes e a resolução de ambiguidades. Exigências prévias para a colaboração: uniformidade de critérios e transposição de análise para código acessível à máquina. Uma metodologia aconselhável: delimitação formal (não semântica), progressiva e sistemática das unidades linguísticas.

Princípios linguísticos diferentes podem gerar grelhas diversas de análise e consequentemente resultados finais não coincidentes. Propomo-nos hoje apresentar tanto os princípios como a grelha que presidiram à exploração de alguns textos latino-medievais (cujos primeiros resultados acabam de ser publicados: *De Ministerio Armorum; Vita S. Fructuosi; Vita S. Martini Saurensis*). Uma entre muitas propostas possíveis, não constitui mais que uma hipótese de trabalho. Intencionalmente ampla, embora talvez não exaustiva, poderá parecer demasiado complexa e trabalhosa. Nada obriga a que tal grelha seja utilizada integralmente de cada vez. Além disso há posições de análise em que a exploração semiautomática é possível. Por último, a aplicação poderá não se confinar ao tipo de textos por nós trabalhado; haverá sempre a possibilidade de manter a programação de base e os índices de código, mas alterar a referência ou a denominação.

## I – PRINCÍPIOS ORIENTADORES

1. A construção de um esquema para análise linguística de determinado *corpus* supõe, não apenas a integração em dado modelo teórico de língua (susceptível de conjugar as várias ocorrências previsíveis numa coerência de compreensão global), mas igualmente uma adequação correspondente quer à natureza desse *corpus* quer à metodologia de aplicação e aos objectivos em vista ou mesmo aos sujeitos intervenientes na análise.

Para julgar do seu grau de adequação e operacionalidade não se poderá subestimar, entre outros dados, que o *corpus* não seja constituído por um texto escrito em língua natural, não inscrito, por isso mesmo, na prática de comunicação quotidiana do operador do esquema. Esta circunstância retira, pelo menos, a possibilidade de recurso imediato a uma intuição ou criatividade linguística que controle, por parte do investigador, a formulação correcta de uma regra. O único controlo possível é o que resulta do confronto de seqüências até um apu-

ramento de constantes.

A valorização do texto em si, como um todo de sequências organizadas, é assim a consequência mais imediata. Daqui derivam seguidamente tanto as perspectivas de método a utilizar como as restrições a impor na construção do esquema de análise. A obtenção de constantes relativas aos processos linguísticos presentes no texto será um dado final e por isso não faria sentido que se introduzisse qualquer tentativa de explicação ou de juízo valorativo. A perspectiva em aberto é, pois, a via descritiva da organização das sequências linguísticas.

Para que a obtenção desses dados finais não seja invalidada, há que proceder uniformemente ao longo da análise e por isso excluir qualquer dado subjectivo que possa sujeitar a grelha previamente fixada a interpretações variáveis. Não seria de admitir, por exemplo, uma hipótese de superação de ambiguidade semântica, já que ela envolveria pelo menos algumas operações adicionais com factores não totalmente ou nem sempre idênticos, e um manejo de língua que poderia não coincidir com o praticado com o texto em análise.

Esta precaução justifica-se tanto mais quanto o distanciamento relativamente ao tipo de textos que pretendemos analisar é grande e tanto mais também quanto a eles não podemos impor um tipo de explicação totalmente uniforme. É o caso do tipo linguístico conhecido por latim medieval, de que nos propomos ocupar.

Qualquer que seja o juízo que se formule relativamente a ele (desde latim de transição ao de latim escolar mais ou menos sujeito a um conjunto de regras desigualmente praticadas consoante o próprio grau de escolaridade do redactor do texto; língua morta ou língua viva...) tal juízo não poderá ser impeditivo de que se venham a procurar descrever, como para qualquer outro sistema de língua, os processos organizativos internos. Mas será também fácil de prever que uma prática distribuída por um período de tempo extensivo a uns dez séculos apresente variantes acentuadas e apreciáveis, tanto mais quanto as comunidades humanas se isolaram e perderam contactos linguísticos entre si, ou se fixaram usos analiticamente irreduzíveis a outros (tipos formulares, onde a ordenação sintagmática é inoperante e só a redução a um modelo anterior explícito pode satisfazer). Não seria legítimo que, valorizando uma variante, se pretendesse apontar como incorrecta uma outra. Haveria, pelo menos, que julgar previamente se tal variante era funcional no seu próprio enquadramento (contexto linguístico) e se a sua substituição punha em causa o sistema ou apenas a norma (norma-uso e não norma-regra), se representava apenas a utilização de uma possibilidade, ainda que inovadora em determinadas circunstâncias linguísticas, ou já uma alteração contrária à estrutura.

Pretendemos, em razão de tudo isto, colocar-nos numa perspectiva de análise de tipo descritivo, adequada à delimitação das unidades linguísticas na sua distribuição funcional em sequências, independentemente de um juízo valorativo do grau de gramaticalidade de determinado uso.

O agrupamento de tais unidades em mapas de índices quantitativos permitir-nos-á posteriormente reconhecer quais os dados, unidades e processos, mais representativos de determinado texto e talvez deduzir uma hierarquização dos mesmos, não apenas segundo a ordem numérica das ocorrências, mas de acordo com o grau de interferência na constituição das sequências linguísticas. O juízo não poderá ser imediatamente decisivo, pois depende da natureza do *corpus*, e este teria de ser constituído por amostragens diversificadas extensivas à variedade de usos e textos (com atenção particular aos géneros literários).

2. Consideremos como princípio axiomático o que constitui dado fora de discussão: que a língua é uma estrutura de unidades organizadas em sequências e hierarquizadas segundo um valor de necessidade da sua presença para que essas mesmas sequências existam como meio de comunicação linguística.

Admitindo porém que uma estrutura é um complexo e resulta de interacção da pluralidade dos seus membros, o primeiro dado que, dentro da delimitação dessas unidades, se nos impõe é de ordem negativa: uma unidade é o que a(s) outra(s) não são. Não numa ordem metafísica, evidentemente, mas numa ordem de distribuição funcional. Daí resulta que uma unidade permutável com outra em contexto idêntico é igual a ela, a nível de função. Correlativamente, os elementos não permutáveis desempenham funções distintas, e é da sua correlação ou interacção que se gera a unidade de sequência linguística. Conclui-se igualmente que tal unidade como todo significativo não advém de um somatório de dados mas da interferência de uns sobre os outros, da dependência de uns relativamente aos outros.

Tal dependência supõe marcas que definem o teor da organização e a natureza dos elementos na ordenação sintáctica. Será por isso possível estabelecer hierarquizações de unidades dentro da sequência. Há unidades que apenas coexistem com outras, sem que estas necessariamente coexistam com aquelas. Nem todas elas são indispensáveis à existência da sequência ou não o são de modo idêntico. Algumas há cuja presença ou ausência é indiferente para a mesma, embora seja de admitir um acréscimo de informação pela sua presença, já que dentro de um princípio de economia na relação significativa (independentemente de um valor semântico referencial) não sejam de admitir elementos nulos.

Se, dentro da pluralidade de unidades requeridas pela estrutura, pelo menos algumas (duas no mínimo) são equipolentemente necessárias, e se entre elas se estabelecem laços de dependência, porque inter-relacionadas, a alteração parcial de uma poderá acarretar a alteração parcial da outra, ou esta ser indiferente à modificação sofrida por aquela. Quando a alteração de uma unidade implica a alteração simultânea de outras, há que admitir similaridade entre ambas e reconhecer que a relação mais imediata se estabelece entre elas e só depois com as outras. Mas quando a substituição de uma unidade não envolve alteração sintagmáti-

ca, ela não se processa a nível de organização de estrutura, mas de referência externa.

Se uma unidade linguística se pode identificar assim pelo contexto em que se integra, não será inadmissível que o contexto zero seja também significativo (como será o caso de partículas interjectivas ou do morfema de vocativo, entre outros exemplos).

3. Dentro dos princípios enunciados, poderemos pois estabelecer vários tipos de unidades demarcáveis dentro da sequência linguística:

1) Unidades permutáveis entre si, dentro do mesmo contexto, apresentam valor sintáctico idêntico e formam entre si categorias paradigmáticas.

2) Unidades susceptíveis de receber marca sintagmática idêntica constituem classes determinadas.

3) Unidades identicamente marcadas numa sequência linguística formam um grupo de similaridade em que cada membro é um constituinte imediato, se entre essas mesmas unidades existe também uma marca de dependência que não existe com os demais componentes dessa mesma sequência (caso de adjectivo + substantivo = função de determinativo/qualificativo; adjectivo + verbo = função predicativa). O índice de dependência pode não se encontrar uniformemente representado. Por vezes, a ordem de sucessão no enunciado poderá ser suficiente, embora não exclusivo de tal relação de dependência imediata. Morfemas suprasegmentais (como a entoação) não registados graficamente poderão estar implícitos na sequência como tal. A frequência e a necessidade de ocorrência em determinado contexto poderá igualmente ser tida como pertinente no apuramento de tais grupos. Assim o facto de um copulador-verbo ocorrer em simultâneo com um adjectivo permite associar ao verbo e não ao nome o adjectivo que ocorre em determinada ocorrência.

4) Há unidades obrigatoriamente presentes para que a sequência se realize. A simultaneidade de presença obrigatória pressupõe natureza distribucionalmente e funcionalmente distinta (nome; verbo).

5) Há unidades facultativamente presentes, mas cuja presença está obrigatoriamente associada a outras (adjectivo; advérbio).

6) Há unidades com existência livre (foneticamente delimitada) e independente (cuja inclusão na sequência não supõe a presença de outras); há outras unidades não livres mas amalgamadas (caso, prefixo, sufixo); há unidades livres e dependentes e não amalgamadas (preposição, advérbio, conjunção, conector e marcador — de negação, interrogação...).

Na eventualidade do amálgama haverá que distinguir o núcleo de base e as unidades associadas. Nestas é possível ainda reconhecer uma ordem de sequência e estabelecer alguns grupos de acordo com ela. Assim uma ordem 1 será constituída por todas as unidades secundárias que imediatamente se seguem à unidade de base, não podendo encontrar-se senão nes-

sa posição; a ordem 2 será formada por aquelas que ocupam o lugar após 1 ou seguem o radical na ausência de 1, mas nunca mais longe do que 2; a ordem 3 será constituída pelas unidades que podem ocupar o lugar depois de 1 ou de 2 e eventualmente o lugar deles em caso de ausência, etc.. Teríamos assim estabelecido classes, cujos membros se excluem mutuamente por ocuparem lugares definidos e distintos na sequência de unidades que constituem um todo. Esta prática será particularmente operatória na determinação de morfemas de uma sequência que constituem uma palavra (unidade demarcada foneticamente e sintacticamente).

7) Há unidades cuja interferência na sequência é mais ou menos alargada e marcam a relação sintagmática (preposição; caso; desinência verbal). É possível, todavia, admitir não apenas relações de dependência, mas também de associação (conectores, assíndeto com inflexão suprasegmental). A relação entre unidades complexas (frase) é também possível, quer a nível de associação quer a nível de dependência.

Poder-se-á prever que, para além da organização sintáctica interna, haja índices demarcadores de tipos de enunciados (marcadores suprasegmentais de interrogação, imperativo, optativo...).

8) De um modo global, e tomando a sequência como um todo unitário, será possível estabelecer uma hierarquização de elementos, reconhecendo:

- a) o seu valor sintáctico relativo dentro de uma ordem de sucessão e de construção da sequência;
- b) adequação para receber marca de dependência sintáctica ou de a significar;
- c) comutabilidade das unidades entre si (identificação paradigmática).

4. Dentro do sistema da língua latina (é dentro dela que continuamos a procurar prever ocorrências) é possível descrever:

1) *Unidades livres* (foneticamente e sintacticamente destacáveis):

1 - obrigatoriamente presentes: nome / pronome (substituto de nome); verbo.

Coexistentes sintacticamente e susceptíveis de aceitar marca correspondente e interdependente, embora também distintiva: género, número, caso; pessoa; modo, tempo, voz.

2 - facultativamente presentes, mas obrigatoriamente associados:

a) com marca de associação: adjectivos (género, número e caso, com o nome; por vezes, a oposição dentro destas categorias neutraliza-se: *homo altus / mulier alta; homo / mulier illustris...*)

b) sem marca de associação: advérbio (ainda que marcado por derivação de forma livre) associado a verbo; preposição associada a nome (marcado em caso); conjunção,

associada (coordenador / subordinador) ou não (coordenador) a sequência frásica; marcadores suprasegmentais (injuntor, interrogador...)

3 - unidades isoladas (ocasionalmente ou não): palavras frase (negador; afirmador; interjeição).

2) *Unidades presas* (fonética e sintacticamente amalgamadas):

1 - sem valor sintáctico, e com uma ordem de sucessão obrigatória relativamente à unidade de base (radical): prefixos; sufixos (categorialmente opositivos — grau — ou não); postfinais (interrogativa *-ne*; intensiva *-met*)

2 - com valor sintáctico, e segundo uma ordem de sucessão obrigatória relativamente à unidade de base (radical):

a) nome (e associado —adjectivo; ou substituto —pronome): morfema de tema, género, número, caso, partícula de vocativo;

b) verbo: morfema de tema, aspecto, modo, tempo, número, pessoa, voz;

c) indiferenciado: postfinal copulativa *-que*.

Será de prever que este modelo nem sempre se realiza e que as oposições se neutralizam. A ocorrência de oposições e neutralizações pode verificar-se igualmente a nível de coexistência ou de redução de uma situação a outra.

Nem há que excluir a hipótese de a sequência linguística não apresentar marcas explícitas de ordenação. Independentemente do facto de a ordem sequencial poder representar também essa concatenação ou de um elemento suprasegmental (entoação) o fazer (o registo de escrita não o dá), há que considerar a frequência de associação de unidades sintagmaticamente neutras em relação de dependência, mesmo sem a respectiva marca. Chame-se-lhe associação de tipo semântico ou não, a situação de ambiguidade sintagmática só poderá ser resolvida pelo confronto com unidades semelhantes sintacticamente marcadas. Pelo menos a nível interpretativo, já que a nível de explicação será legítimo o recurso aos processos psicológicos de associação utilizados pelo utente da língua ou à resolução por via transformacional.

5. Numa primeira fase de análise, pretendemos apenas fazer um registo dos factos, isto é, das ocorrências dentro da sequência linguística, organizada como unidade global. Limitamo-nos a recolher as unidades linguísticas, tanto quanto possível dentro de uma ordem de concatenação (excluindo obviamente variantes alomórficas, dependentes de contexto fonético e portanto não opositivas). Por unidades linguísticas, hão-de entender-se tanto as unidades mínimas obtidas por segmentação (morfemas), como, atendendo às possibilidades reais e consentidas pela metodologia de registo, os constituintes imediatos de um grupo (nome e de-

terminante ou determinado; subordinador e tempo de verbo; verbo e regência, etc.). Tentamos assim ultrapassar uma fase de atomização de sequência linguística, que, embora não o fosse na origem por parte dos intervenientes na análise (dado que a parte interpretativa supunha uma consciência do contexto global em que a unidade mínima se incluía) não aparecia como tal no respectivo registo.

Esta precaução metodológica é tanto mais justificada quanto partimos do pressuposto de que os textos que pretendemos analisar (de latim medieval) podem nem sempre estar em conformidade (o que não significa agramaticalidade) com a norma-regra descrita exemplarmente pelos teóricos de determinado período. Se no registo a exclusão do contexto, em que aparece a unidade linguística, pode ser aceite quando é de prever a sujeição a uma norma-uso equivalente a uma norma-regra, já o mesmo não será legítimo quando se pode duvidar dessa equivalência.

Há que ter em conta, no entanto, que as duas perspectivas são divergentes quanto ao objectivo final. Enquanto aquela terá de acabar na verificação da ocorrência ou da não ocorrência, da conformidade ou desvio relativamente a um uso anterior, dando predominância a um índice comparativo, esta pretende entender o processo sintagmático na sua ordenação interna, independentemente de confrontos; aquela compara, esta descreve.

Colocando-nos assim em plano descritivo, não pretendemos desde logo e pelo menos numa primeira fase de abordagem do texto, estabelecer níveis de equivalência que ultrapassem a organização sintagmática. Não nos propomos, por ex., determinar que *modo facile/i* é igual a *facile* (adv.) mas apenas registar que o grupo verbal se encontra determinado por advérbio ou nome marcado por morfema de ablativo. Posteriormente, o confronto do registo dar-nos-á qual a tendência predominante na construção de determinado verbo ou quais são as unidades concorrentes em determinado contexto. Pense-se igualmente na correspondência e/ou não correspondência entre construção de verbo com acusativo e infinitivo ou com conjunção e tempo finito, ou ainda em nome construído com casos diferentes que sim/não se identificam.

Sem julgarmos que a solução de tais equivalências passa necessariamente por uma interpretação que terá de fazer entrar o factor semântico com apelo a uma situação referencial, reconhecemos, no entanto, que a comutação por contextos idênticos, quando se trata da análise de textos cujo sistema linguístico no todo ou em parte há que apurar, porque já não pode haver recurso a intuição da mesma prática linguística criativa ou porque o redactor não tinha necessariamente que ater-se a um esquema normativo regulador da sua actividade, só pode ser feita em fase ulterior.

Não pretendemos por isso prever (ao elaborarmos o esquema que presidirá à recolha de dados sobre o *corpus* que mais imediatamente vai ser constituído por textos latino-medievais, como os já referidos) variantes semânticas inerentes a construções distintas. Caberá isso



a uma fase explicativa e não a um modelo descritivo de levantamento de ocorrências.

6. Poderá a recusa de um critério semântico ser absoluta ? Poderá, por outro lado, ser operatório esse mesmo critério, não obstante toda a margem de ambiguidade ?

Consideremos que o sistema gráfico em que a sequência linguística é registada não é totalmente englobante quanto à inclusão de traços demarcativos. Os traços suprasegmentais só em escala muito reduzida aí figuram, mediante pontuação ou partículas tais como as demarcativas da interrogação (quando esta não é assinalada por pronomes que podem em determinados contextos oferecer variantes específicas). Não poderá negar-se que a inflexão de voz é por vezes suficiente para resolver a ambiguidade sintáctica de uma sequência. Nestas condições, e noutras, só a percepção global da frase (sequência organizada completa, se bem que não passível de delimitação quanto a número de unidades) permite suprir traços necessários para a comunicação linguística e/ou aquilatar da correspondência com outras sequências mais explícitas e determinar os elementos em conformidade com tal correspondência. Tenha-se isso como fenómeno de excepção ou não, na realidade não pode subestimar-se. Nem se pode considerar um sistema teórico como menos operatório pelo facto de apenas mediante aproximação transformacional poder incluir dados que só foram organizados sintacticamente por referência a uma situação que não se integrou plenamente e explicitamente na sequência linguística. Ou rejeitamos tais dados ou imaginamos para eles outro modelo de organização, ou simplesmente os interpretamos por confronto com esquema mais comum. Mas se foi um recurso admitir a interjeição, por ex., como classe paradigmática a nível de unidade constituída pela palavra (foneticamente e sintacticamente isolada) não será tão fácil prever grupos sintácticos incompletos suficientemente homogêneos e delimitados para os reconhecer como classe.

Dentro de um critério estritamente distribucional não haveria que admitir interpretações que ultrapassassem o nível de associação. Assim não se justificaria que uma sequência com marca de acusativo fosse classificada ora como complemento de direcção ora como complemento de objecto directo. Registando o facto da ocorrência, verificaríamos, mais tarde, que a coincidência era apenas aparente, pois o contexto não era formalmente e sempre idêntico, já que a comutação não era viável em todos os casos. Assim, *eo Romam / video Romam* coincidem; mas não ocorreria *eo altitudinem turris* em contraste com *video altitudinem turris*. Teríamos já uma restrição significativa: *eo* apenas entra em situação de regência de acusativo com formas monomorfemáticas como são os nomes de lugares. Mas tal conclusão só resulta de um confronto e mesmo de uma interpretação semântica (também os nomes de pessoas são monomorfemáticos, e só o apuramento de outros contextos onde a permutação não seja possível permitirá uma conclusão pertinente).

Será forçoso deixar para fase posterior a análise de tais distribuições ? A possibilidade de adiantar alguns dados manifestamente interpretativos, logo na fase de recolha, poderá apontar-se como hipótese, pelo menos para factos mais frequentes e constantes. A gramática tradicional utilizava critérios formais e semânticos à mistura. Valerá a pena regressar ou manter essa fase, a título pelo menos de complementaridade ? O problema levanta-se mais de perto quando a marca formal não é plenamente explícita da função, isto é, da relação interna na sequência organizada, e há que decidir acerca dela. Sirva de exemplo o fenómeno do sincretismo ou das neutralizações morfemáticas quando há que definir a relação do grupo nominal com o grupo verbal. A limitação de resultados a que um critério estritamente formal conduzirá poderá ser compensada pelo cruzamento com um critério semântico. Em que medida será legítimo utilizar esta correcção depois das restrições postas anteriormente em nome de um rigor objectivo ? Julgamos que de harmonia com a própria operatoriedade.

Relativamente ao verbo, por exemplo, não seria talvez rentável distinguir entre sujeito animado e inanimado (embora se possa prever que, por ex., *curro* exigirá um animado) ou entre humano e não humano (se bem que de antemão se pudesse estabelecer que *cogito* só admite um humano). O facto é que não há em latim uma distribuição morfemática correspondente a tais oposições. Para julgar de tal rentabilidade haveria que proceder ao levantamento de uma grelha correspondente de análise semântica (o que não está imediatamente em causa). E o que se diz do verbo poderia afirmar-se igualmente do adjectivo relativamente ao substantivo, e correspondentemente do advérbio em relação com o verbo.

Existem todavia outras situações em que o quadro de traços semânticos pode ser mais delimitado e definido, em virtude de uma situação de língua (latina) em que eles eram marcados. Seja o fenómeno da complementarização ou expansão verbal. É obviamente de nível semântico reconhecer se determinado complemento faz parte integrante necessária ou acessória do sintagma verbal da frase (aliás o categorema aduzido é de referência semântica, pois não faria sentido pleno empregarmos aqui uma referência à obrigatoriedade de presença no sintagma). Pode, no entanto, contrapor-se-lhe uma situação em que tal complementarização era ainda marcada (pelo menos para algumas ocorrências). Permite-nos isso verificar até que nível de funções se realizou a neutralização morfemática. Restringir-nos-emos porém a categoremias largas e conhecidos como sejam: sujeito, objecto directo, objecto indirecto, predicativo e complemento circunstancial. Poderá isto servir também para a delimitação de constituintes específicos, entre os quais se procura estabelecer uma certa hierarquização. Considere-se, no entanto, como solução provisória e dentro de um determinado tipo de análise que tem de tratar a palavra como elemento isolado e não pode atingir ainda o enunciado, em razão, nomeadamente, da discontinuidade linear dos constituintes imediatos da sequência linguística. Extrairemos, porém, dados pertinentes, e que, deixados para fase posterior, dupli-

cariam o trabalho: neutralizações casuais e equivalências relativamente à expressão de uma função.

## II – GRELHA DE ANÁLISE

De acordo com os princípios acabados de enunciar, vamos construir e apresentar seguidamente a nossa grelha de análise. Dispensamo-nos de justificação ou comentários, já que, a fazerem-se teriam de ser dilatados para serem compreensivos, e resultariam irrelevantes em bom número de casos. A leitura e interpretação não oferecerão, contudo, dificuldades ou dúvidas de realce, desde que se atentem nos critérios formais de caracterização. Assim não causará estranheza que designemos os substantivos de 1.a, 2.a, 3.a... declinação de acordo com o tema -a, -o, -i, -c... ou o adjectivo de 1.a, 2.a e 3.a classes segundo as variantes 3, 2, 1, por eles apresentadas. Reconhecer-se-á também que omitimos caracterizações cuja informação reputamos diminuta ou irrelevante (género, número e caso para pronomes e adjectivos, por ex.) e que preferimos assinalar outras de cuja importância não se poderá duvidar (no caso do pronome ou do adjectivo a oposição entre forma livre e forma presa permite-nos obter os usos pronominais e substantivados sem termos de aumentar os índices numéricos de distinção no lema), etc..

Uma coluna de análise procura obter dados sobre variantes gráficas. É evidente que a sua utilização apenas será pertinente quando estivermos perante registos de manuscritos e haja razões para supor que a alteração do código gráfico convencional obedece a uma intenção ou reflecte hábitos particulares (descuido de cópia, insegurança de código escrito...).

Não transcrevemos para aqui os símbolos de codificação. Não propriamente por desejo de manter secreta uma chave necessária, mas porque, a juntar à falta de espaço para distribuição gráfica da grelha, consideramos que a técnica de codificação supõe uma metodologia que não pode ser improvisada ou considerada aprendida pela utilização de um símbolo. Não teremos dúvidas, porém, em fornecer o nosso código (aliás simples e que tenta ser o mais transparente possível) a quem estiver em condições de o utilizar e queira correr esse risco.

*AIRES AUGUSTO NASCIMENTO*





