

REFERÊNCIA:

LEFFA, Vilson J. Análise sintática; ensinando o que não se sabe para quem já sabe? In: FORTKAMP, Mailce Borges Mota; TOMITCH, Lêda Maria Braga (Orgs.) *Aspectos da lingüística Aplicada*; estudos em homenagem ao Professor Hilário Inácio Bohn. Florianópolis: Insular, 2000. p. 203-228.

ANÁLISE SINTÁTICA: ENSINANDO O QUE NÃO SE SABE PARA QUEM JÁ SABE?

Vilson J. Leffa, UCPEL

INTRODUÇÃO

O objetivo deste trabalho é investigar o efeito do ensino da análise sintática a longo prazo. Busca-se saber até que ponto a instrução que os alunos possam ter recebido no primeiro e segundo graus permanece quando eles chegam à universidade, levando em consideração não questões de terminologia e metalinguagem, mas conceitos fundamentais na segmentação da frase complexa. Para isso foi delineado um estudo transversal, tendo como tarefa a segmentação de diferentes tipos de coordenação e subordinação por alunos da quarta série do primeiro grau até o nível universitário.

Inicialmente, na parte da revisão teórica, verifica-se que apesar da reconhecida importância da segmentação no processamento da língua, há uma escassez muito grande de estudos empíricos na área, quando se trata da segmentação de orações. O que se encontrou são respingos de abordagens teóricas que podem ter um impacto maior ou menor no processamento das frases complexas, tais como a Gramática de Estados Finitos, ressuscitada por Brazil (1995), a Gramática de Valências (ex.: Borba, 1996) e as propostas das gramáticas pedagógicas, geralmente com ênfase nos indicadores formais de coordenação e subordinação (ex.: Quirk et al., 1985). As limitações e vantagens dessas abordagens é o que se pretende discutir no levantamento dessa literatura, levando à conclusão de que quando se ensina análise sintática aos alunos ensina-se algo que não se sabe explicitar.

Para medir o impacto desse ensino nos alunos, delineou-se um estudo transversal, usando como instrumento uma

tarefa de segmentação de 44 frases complexas, cada uma com duas orações, e usando como sujeitos 271 alunos da quarta série do primeiro grau até o nível universitário, o que levou a um total de 11.924 frases. Os resultados são então discutidos em detalhes, levando em consideração, em primeiro lugar, os elementos que facilitaram e dificultaram a segmentação e, em segundo lugar, o desempenho dos sujeitos nos diferentes níveis de escolaridade, destacando os pontos onde houve e não houve progresso na tarefa de segmentação das frases. O achado mais importante na análise dos resultados foi a aparente e total incapacidade da escola em levar o aluno para além do que ele já sabe quando é iniciado nos estudos da sintaxe.

As implicações pedagógicas são finalmente discutidas, em termos das duas conclusões opostas a que parecem levar os resultados: (1) a necessidade de se abandonar o ensino formal da sintaxe ou (2) a retomada de seu ensino sob uma perspectiva diferente. Uma proposta de como poderia ser essa retomada é discutida com algum detalhamento. A idéia básica é de que não se deve ensinar o que não se sabe mas tentar desvelar o que o aluno já sabe através de um processo de descoberta do conhecimento lingüístico.

A DIFICULDADE EM EXPLICITAR O PROCESSO

A segmentação de uma frase em partes menores pode ocorrer em vários níveis de abrangência. No nível mais alto, no caso de uma frase complexa, temos o segmento da oração, envolvendo um sintagma verbal e seus argumentos (sujeito, objeto, etc.). Num nível intermediário, podemos considerar os segmentos que abrangem os diferentes sintagmas e as palavras isoladas. Num nível bem mais baixo podemos citar, por exemplo, os segmentos que envolvem o fonema.

Todos esses segmentos precisam ser identificados e processados individualmente para que um texto maior possa ser compreendido ou produzido. Essa segmentação, ainda que abaixo do nível da consciência, envolve a capacidade do falante em localizar o início e o fim de cada um desses segmentos, resolvendo problemas de transição entre um

segmento e outro. Para isso, o falante usa sua informação lingüística internalizada, envolvendo conjuntos de traços distintivos — desde pistas acústicas até a classificação gramatical das palavras — e regras de combinação desses traços, de modo a formar uma estrutura onde o segmento a ser processado possa ser encaixado. São as regras de combinação, por exemplo, que vão estabelecer se uma palavra como "ama" é um verbo ou um substantivo, podendo iniciar um sintagma verbal em "Ela ama a gramática", ou terminar um sintagma nominal numa frase como "A ama chegou"

Enquanto que os níveis mais baixos de segmentação têm gerado grande quantidade de pesquisas por parte de muitos estudiosos, os níveis mais altos têm atraído muito pouco interesse. Esse desinteresse já começa a diminuir no nível do sintagma, em que pese a preocupação histórica demonstrada por Saussure, ao alertar que na língua não há apenas diferenças, mas também agrupamentos: "não falamos por signos isolados, mas por grupos de signos, por massas organizadas que são elas próprias signos" (Saussure, 1988, p. 148). Atualmente os estudos sobre as unidades fraseológicas na área da terminologia (ex. Kjaer, 1990; Pavel, 1993) e a abordagem lexical no ensino de línguas estrangeiras (ex.: Nattinger & DeCarrico, 1992; Lewis, 1993) podem também ser citados como exemplos dessa preocupação com o nível do sintagma.

No nível da segmentação de orações, em oposição à abundância terminológica da área (o que para alguns é *oração*, para outros pode ser *cláusula*, ou mesmo *proposição*; o que para alguns é *sentença*, para outros pode ser *frase* ou *período*; e finalmente o que para alguns é *frase complexa*, para outros pode ser *período composto*), há uma escassez muito grande de estudo descritivos da oração em frases complexas, conforme observa Perini (1989, p. 206). Ainda, segundo o próprio Perini, nos raros momentos em que o problema é abordado, a questão é tratada de forma incompleta. Assim enquanto que para Jespersen a oração é o "membro de uma sentença que tem forma semelhante à de uma sentença" (Jespersen, 1933, p.166), para Martinet a oração é composta de "[...] partes do enunciado que

apresentam a mesma forma que uma frase completa [...]" (Martinet, 1979, p., 17).

O que Perini não chega a afirmar é que a definição de Martinet, não é apenas incompleta, mas também incorreta.

Uma oração é mais do que uma frase dentro de uma frase. Um segmento como "Pedro chegar" pode formar uma oração ("Vi Pedro chegar"), mas não pode formar uma frase livre. Teria um grau de "oracionalidade", na expressão de Perini (1989, p. 209), exibindo um "indicador formal de subordinação" na terminologia de Quirk et al. (1985, p. 997).

O que se nota, na verdade, é que a segmentação da frase complexa em orações é uma questão tratada apenas pelas gramáticas pedagógicas, quer tradicionais (ex.: Oliveira, 1960; Almeida, 1992), quer atuais (Ex.: Perini, 1996; Silva e Koch, 1995; Quirk et al., 1985). Todas elas se caracterizam por apresentar uma descrição menos ou mais completas dos diferentes tipos de oração, mas poucas oferecem pistas de como se pode chegar a essa segmentação.

As que oferecem contam com a intuição lingüística do aluno, sem explicitar o processo. Veja-se, por exemplo,

Oliveira (1960, p. 348). Segundo o autor, para determinar as orações é necessário três passos:

- (a) marcar os verbos, tomando cuidado com as locuções verbais;
- (b) marcar as conjunções, relativos e as formas nominais do verbo e
- (c) marcar o fim das orações, "de acordo com o sentido parcial obtido"(p. 348).

Cabe ao aluno decidir quando terá chegado ao "sentido parcial obtido", o que significa, na prática, deixar a solução

do problema para a intuição do falante, pressupondo um conhecimento que se é incapaz de explicitar. Almeida

(1992) também não é muito diferente, quando, para marcar o início da oração, aconselha o aluno a prestar atenção

nas formas conjuntivas (conjunções, pronomes relativos, advérbios relativos), e o alerta de que "A inversão dos

termos não deve perturbar-nos na separação"(p. 532). Fornece, porém, uma pista interessante ao sugerir que

"quando o verbo da oração está na forma nominal, pertencem a essa oração os termos que têm relação com a forma nominal" (p. 532).

OS DIFERENTES PROCESSOS DE SEGMENTAÇÃO

O primeiro problema na segmentação da frase complexa em seus componentes oracionais é definir por onde começar o processo. Há três possibilidades: (1) iniciar pela primeira palavra da frase e avançar da esquerda para a direita palavra por palavra; (2) iniciar pelo marcador formal, geralmente uma conjunção, que identifica o início da oração, coordenada ou subordinada e, a partir daí, buscar, à direita, o fim da oração; ou (3) iniciar pelo sintagma verbal, identificando o tipo de verbo e depois localizar o sujeito e os complementos, quando exigidos pelo verbo.

O processamento da esquerda para a direita é um dos mais antigos na história da lingüística moderna e pode ser atribuído a Hocket (1955), em seu modelo de gramática de estados finitos. Embora descrita em detalhes por Hocket, a gramática é mais conhecida pela descrição que dela faz Chomsky na *Syntactic Structures*:

[o falante ao produzir uma sentença] começa no estado inicial, produz a primeira palavra da sentença, entrando assim num segundo estado que limita a escolha da segunda palavra, etc. Cada estado pelo qual passa representa as restrições gramaticais que limitam a escolha no próximo ponto do enunciado (Chomsky, 1957, p. 19)

Para Chomsky a gramática de estados finitos é limitada e incapaz de gerar todas as frases possíveis da língua, em razão de pressupor um processamento linear, feito palavra a palavra, e portanto incapaz de resolver o problema de orações intercaladas. Na frase *A mulher que beijou o ministro foi presa*, por exemplo, seria impossível processar a oração *que beijou o ministro*, por ela estar contida dentro da oração *A mulher foi presa*.

Apesar das restrições de Chomsky, a gramática de estados finitos tem servido de base para várias abordagens do processamento da língua, incluindo o Augmented Transition Network (ATN) de Woods (1970), na área de inteligência artificial, e a proposta de Brazil (1995), na compreensão da língua oral. Ainda que possa ter um componente *top down*, como ponto de partida, o modelo é essencialmente *bottom up*, em termos de processamento.

O uso de indicadores formais, tanto de coordenação como de subordinação, parte do princípio de que a oração, quando não é absoluta, geralmente inicia por um coordenador ou subordinador. A principal vantagem deste modelo

é poder dispor dos amplos levantamentos já realizados nas diversas línguas (Almeida, 1992; Perini, 1989; Souza e Silva e Koch, 1995; Quirk et al., 1985). Quirk et al. (1985), por exemplo, dedicam à coordenação e à subordinação, 231 páginas de sua gramática. Esses levantamentos mostram que as orações podem iniciar também por outras marcas que não sejam as conjunções. Entre essas marcas, podemos destacar as inversões (*Tivesse ele apoio do congresso, ficaria oito anos*); o uso de formas verbais não-finitas (*Reformar a constituição é a meta de todos os presidentes*), e algumas formas finitas de certos verbos (*Está no governo faz sete anos*).

Os indicadores formais são muito úteis na segmentação da frase complexa quando as orações que eles marcam não estão invertidas e nem intercaladas dentro de outras orações. É mais fácil, por exemplo, segmentar as orações numa frase sem inversão como *José virá se amar Maria* do que *Se amar Maria José virá*. Neste caso o início da segunda oração pode estar em três posições diferentes:

Se amar Maria / José virá.

Se amar / Maria José virá.

Se amar Maria José / virá.

Enquanto que a gramática de estados finitos pressupõe um processamento *bottom up*, da esquerda para a direita, criando estados que vão se modificando gradativamente com o avanço de cada palavra, até chegar ao estado final, o recurso ao uso de marcadores formais de coordenação e subordinação pressupõe um processamento *top down*, da oração aos seus componentes menores. O processamento, no entanto, continua sendo da esquerda para a direita: identificado o marcador, o usuário da língua prepara-se para processar à direita uma oração, que irá até o fim da frase ou início da próxima oração.

Uma abordagem centrada no verbo sugere um processamento não da esquerda para a direita mas de um núcleo central para as partes periféricas, governadas por esse núcleo. A gramática de valências é um exemplo típico dessa abordagem. O verbo não é apenas a razão de ser da oração mas o centro de onde emanam os controles a cada um

de seus argumentos, cujo número varia de acordo com a valência de cada verbo. Segundo Borba (1996), essa valência pode variar de 0 a 4, tendo valência 0 os verbos que não aceitam nem sujeito e nem complemento (*chove*), ficando com valência 4, por exemplo, aqueles que além do sujeito podem aceitar até três complementos (*Ele traduziu o livro do francês para o português*). Neste modelo, para encontrar os limites da oração, tanto à esquerda como à direita, basta identificar a valência do verbo e, a partir daí, localizar os argumentos que estão ligados a esse verbo.

O fato, no entanto, de os verbos serem ambíguos em termos de valência limita a aplicação do modelo. O verbo *chover*, por exemplo, embora tendo valência 0, e portanto sem um lugar para o sujeito ou complemento, pode facilmente aceitar um objeto direto (*Choveu mana no deserto*). É justamente a ambigüidade valencial do verbo *amar*, ora funcionando como transitivo direto, ora como intransitivo, o fator que possibilita as diferentes segmentações, conforme já foi visto, para as duas frases abaixo:

Se amar Maria / José virá.

Se amar / Maria José virá.

O uso real da língua está também muito longe da rigidez de qualquer norma preestabelecida, tradicional ou não, limitando em aplicações práticas as pré-classificações baseadas nas valências. Os exemplos abaixo, tirados de Pontes (1986), mostram como os verbos transitam facilmente de uma subcategoria para outra, mudando seu significado, mas aparentemente, sem dificultar a compreensão da frase pelo usuário da língua:

A Belina deita o banco, sabe? (p. 17) (Belina = camioneta)

A Sarinha tá nascendo dente. (p. 18)

O Carlos André cresceu o nariz (p. 18)

O único trabalho, de meu conhecimento, a tratar especificamente da segmentação da oração é o de Stefanini (1993), na área da inteligência artificial, que apresenta uma proposta de tratamento de língua natural através de um sistema de múltiplos agentes, sendo a segmentação da frase complexa em orações a tarefa executada por um desses agentes

(p. 150 e seguintes). Esse agente segmentador caracteriza-se por conhecer os indicadores formais de subordinação e coordenação e é capaz de detectar as orações principais, coordenadas e subordinadas, construindo árvores para representar a estrutura da frase complexa, incluindo não só as orações mas também possíveis incisos como parênteses, colchetes, travessões, etc. Às vezes consegue também classificar as orações segmentadas em seus diferentes tipos (relativa, adverbial, substantiva).

O critério fundamental para a identificação de uma oração na proposta de Stefanini é a existência de um verbo conjugado: haverá tantas orações quantos forem esses verbos. O início da oração é marcado a partir da conjunção, embora não fique claro como proceder quando a oração estiver invertida ou intercalada. A autora faz referência a algoritmos usados anteriormente, ao que parece baseados na gramática de estados finitos, chama atenção para suas limitações no caso de orações invertidas, mas não chega a explicitar uma solução. Também não fica claro como foram tratadas as orações com verbo no infinitivo (*estudar a gramática*), gerúndio (*estudando a gramática*) ou particípio (*estudada a gramática*). Tem-se também a impressão de que a autora não faz distinção, para fins de processamento, entre construções como *posso trabalhar*, onde temos um sintagma verbal, e *quero trabalhar*, onde temos dois. Essa distinção, no entanto, pode ser muito importante, porque onde há dois sintagmas verbais há duas orações, que, como tais, precisam ser identificadas para o processamento da língua. Como seria possível, por exemplo, processar uma frase como *trabalhar cansa*, onde o verbo *trabalhar*, ainda que no infinitivo, funciona como uma oração substantiva, podendo, portanto, ser o sujeito do verbo *cansar*? A questão não é trivial.

Consideremos, por exemplo, as frases abaixo:

Visitar muitas cidades cansa.

*Visitar muitas cidades cansam

A única maneira de fazer a concordância correta do verbo é tratar todo o segmento anterior como um sujeito singular e não plural. Isso, no entanto, só é possível reunindo as palavras numa oração substantiva, mesmo que o

verbo *visitar* esteja no infinitivo.

O que se nota, portanto, ao se revisar a literatura sobre a segmentação da frase complexa em orações, é a inexistência de uma descrição clara de como ocorre o processo. As propostas de um algoritmo que pudesse ser oferecido ao aprendiz são vagas, baseando-se em intuições que não podem ser explicitadas, ou incompletas, resolvendo apenas um percentual dos problemas encontrados no uso real da língua. A incapacidade de abranger o problema na sua totalidade, para que a solução possa ser eficaz, é percebida nas diferentes abordagens, desde a gramática de estados finitos, até a gramática de valências.

A hipótese principal deste estudo é de que se não há uma descrição clara de como se processa a segmentação, conforme se observa na literatura, a instrução específica em análise sintática externa, por partir de descrições vagas e confusas, não deve produzir qualquer resultado prático para o aluno em termos de capacidade de segmentação da frase. De acordo com essa hipótese, um pós-teste aplicado ao sujeito depois dos anos de instrução que recebe na escola, não terá uma diferença estatisticamente significativa de um pré-teste aplicado antes do período de instrução.

METODOLOGIA

O objetivo desta pesquisa é verificar o impacto da escola na capacidade do aluno em realizar a análise sintática, em geral, e fazer a segmentação da frase complexa, em particular. O estudo é transversal, reunindo sujeitos de nove níveis diferentes de escolaridade, desde a escola fundamental até o nível universitário, e parte do pressuposto de que os sujeitos dos níveis mais altos tenham em algum período de sua escolaridade passado por instrução em específica em análise sintática.

Foram selecionados alunos que freqüentam desde a quarta série do curso fundamental até o nível universitário num total de 271, assim distribuídos: 4ª Série do 1º Grau (35 alunos), 5ª Série do 1º Grau (81 alunos), 6ª Série do 1º

Grau (26 alunos), 7ª Série do 1º Grau (30 alunos), 1º Ano do 2º Grau (16 alunos), 2º Ano do 2º Grau (37 alunos), 3º Ano do 2º Grau (24 alunos) e Nível Universitário (22 alunos). O critério para a seleção foi basicamente a disponibilidade desses sujeitos, o que explica a ausência de uma turma de 8ª série. Essa ausência não deverá influenciar os resultados, considerando que são ao todo oito níveis diferentes de escolaridade. Os alunos são de diferentes escolas da parte Sul do Rio Grande do Sul, na região de Pelotas e Rio Grande. A razão por escolher sujeitos a partir da 4ª série foi a suposição de que antes desse adiantamento deve haver muito pouca instrução em análise sintática.

O instrumento usado foi um teste de segmentação, composto de 44 frases complexas de duas orações cada uma. Procurou-se, nas 44 frases, reunir pelo menos um exemplo de cada tipo de oração, incluindo orações coordenadas (aditivas e adversativas) e orações subordinadas (substantivas, adjetivas e adverbiais). As frases foram todas tiradas e, às vezes, levemente adaptadas, de gramáticas de ensino de língua portuguesa.

O procedimento consistiu basicamente, após a entrega do teste, em solicitar ao aluno que dividisse cada frase em duas partes, ponto um traço vertical onde achasse que a frase pudesse ser separada. Toda metalinguagem foi evitada.

RESULTADOS E DISCUSSÃO

Os testes dos 271 alunos, perfazendo um total de 11.924 frases, foram corrigidos e tabulados. A média de acertos para a segmentação das 44 frases foi de 32,45. Isso significa que todos os sujeitos acertaram em média 73,75% de todos os itens.

Para a análise desses dados, vamos considerar primeiro os resultados produzidos pelo próprio teste, fazendo um levantamento dos fatores que contribuíram para facilitar ou dificultar a segmentação das frases. Num segundo momento, vamos comparar o desempenho dos sujeitos envolvidos, desde a 4ª série até o nível universitário.

Elementos Facilitadores

A Tabela 1 mostra alguns dos fatores que parecem ter contribuído para um maior ou menor número de acertos.

Entre os fatores que mais contribuíram, em termos percentuais, podemos destacar, o uso de vírgula (*Acabada a festa, os músicos partiram*) coordenação (*Sua namorada partiu e não voltou*) e oração adjetiva não-intercalada (*Afasta-te do cão que não ladra*).

Tabela 1 - Fatores que mais e menos contribuíram para a segmentação, em percentuais de acertos

<i>Fatores</i>	<i>Acertos</i>
Orações separadas por vírgula	93%
Coordenadas, sem vírgula	91%
Adjetiva não-intercalada	81%
Subordinadas adverbiais - ordem direta	74%
Encontro de verbos conjugados	71%
Adjetiva intercalada	69%
Substantiva completiva nominal	68%
Substantiva objetiva indireta	67%
Subordinadas adverbiais - ordem indireta	66%
Oração com uma palavra	59%
Substantiva subjetiva	55%
Substantiva objetiva direta	52%
Substantiva predicativa	18%

O maior número de acertos (93%) ocorreu com as orações separadas por vírgula, tanto as coordenadas (*Não só o roubaram, também o feriram*) como as subordinadas (*Se ele entra, eu saio*). O uso da vírgula sobrepujou a ordem das orações na frase, compensando de certa maneira, as dificuldades apresentadas pela ordem indireta. Também anulou o tamanho da oração; embora os sujeitos tivessem uma tendência a desconsiderar orações de uma só palavra, construída a apenas com o verbo, a presença da vírgula levava-os a separar corretamente (*Estudando, a gente aprende*). A presença da vírgula foi fator decisivo, independente do grau de escolaridade dos sujeitos, o que parece

demonstrar que mesmo os alunos da 4ª série do 1º grau já têm noção da pausa que ela representa na frase. Provavelmente não se trata apenas do sinal gráfico, já por si só indicador de separação, mas também de motivações sintáticas. Uma frase como *Quanto chegou, eu estava estudando*, com altíssimo índice de acerto entre os sujeitos, pode indicar uma espécie de acúmulo de marcadores de separação, representados não só pela vírgula mas também por peculiaridades da estrutura sintática.

Retirada a vírgula, o segundo fator mais importante foi a coordenação (91%). Como no caso da vírgula, a coordenação anulou outros fatores que poderiam dificultar a segmentação, como, por exemplo, a combinação de orações coordenadas e subordinadas em 1.

(1) Quero que vás mas peço que volte.

É possível também que as frases complexas selecionadas para o teste apresentassem um elevado índice de independência entre as orações que os compõem, o que facilitaria a segmentação. Como se pode ver no exemplo 1, a conjunção *mas* não chega a competir com as duas conjunções *que*, realmente subordinadas. O *mas* é tão forte que quando aparece na frase tem o poder de anular tudo o que foi dito antes dentro da frase e iniciar uma nova idéia. O instrumento usado, baseado em exemplos que poderíamos considerar prototípicos de coordenação e subordinação, não explora talvez os diferentes graus de independência que pode existir mesmo em orações coordenadas, conforme se pode perceber nos exemplos abaixo:

(2) Quero que vás mas peço que volte.

(3) Quero que vás e peço que volte.

(4) Quero que vás e quero que volte.

(5) Quero que vás e que volte.

(6) Quero que vás e volte.

Uma leitura atenta das frases mostra que há uma diferença muito grande quando se passa da frase 4 para a frase 5.

Em 2, 3 e 4, a conjunção coordenativa une dois segmentos que contêm, cada um, uma oração principal e uma

subordinada substantiva objetiva. Tudo o que está antes da coordenativa é sintaticamente igual ao que está depois, de modo que, de acordo com a tarefa exigida na pesquisa, a conjunção coordenativa seria realmente o ponto correto para fazer a segmentação da frase em duas partes.

Nos exemplos 5 e 6, porém, isso já não acontece. O paralelismo entre o que está antes e depois da conjunção coordenativa não existe mais. A elipse do verbo *quero* deixa a conjunção dentro de segmento subordinado, de modo que a frase teria que ser segmentada antes da conjunção, ou seja, logo depois do primeiro verbo.

Uma possível explicação, portanto, para o elevado índice de acertos quando da ocorrência de orações coordenadas pode estar no uso de construções dos tipos 2 a 4, onde as orações da frase têm autonomia sintática e morfológica.

As orações *Quero que vás* e *Peço que voltes*, podem ser retiradas da frase e formar frases completas sem qualquer alteração.

Conforme se esperava, a ordem das orações na frase foi também um fator que contribuiu para um maior ou menor índice de acertos. Em geral, a ordem direta (*Ela dormia quando ele chegou*) contribuiu para elevar os acertos, em oposição à ordem indireta (*Quando ele dormia ela chegou*) ou às orações intercaladas (*A terra onde nasci é linda*). Note-se, mais uma vez, que no exemplo *ela dormia quando ele chegou*, se a conjunção *quando* for retirada, as duas orações restantes formam duas frases completas, com conteúdo proposicional distinto.

Elementos Complicadores

Fazendo-se uma comparação entre os vários tipos de orações subordinadas (adjetivas, substantivas e adverbiais), nota-se que as substantivas foram as que apresentaram maior dificuldade para os alunos. Uma possível explicação para isso é a duplicidade de função exercida pelas orações substantivas; ao mesmo tempo em que elas formam uma oração completa elas também formam, de modo subalterno, o termo integrante de uma oração maior, na função de

sujeito, objeto ou de complemento nominal. Note-se, por exemplo, a frase *trabalhar cansa*. A palavra *trabalhar*,

sozinha, ao mesmo tempo em que é uma oração substantiva subjetiva, é também o sujeito gramatical de toda a frase, funcionando aí como um sintagma nominal. A dificuldade na segmentação estaria assim no fato de ter que separar o verbo do sujeito gramatical, o que normalmente não acontece. Os sujeitos intuitivamente podem ter sentido essa dificuldade.

Independente de nossa consciência metalingüística e do maior ou menor domínio que possamos ter da terminologia gramatical, fica mais fácil identificar uma oração quando ela se concretiza em termos de sujeito e predicado. Em 7, por exemplo, embora se trate de uma oração principal e uma subordinada, cada oração tem seu sujeito e seu predicado, permitindo perguntas como *Quem dormia? Quem chegou? O que ela fazia?*

(7) Ela dormia quando ele chegou.

No exemplo 8, no entanto isso já não ocorre. Só posso fazer uma pergunta para o sujeito (*O que cansa?*), tendo como resposta toda a oração anterior (*trabalhar*), e outra para o predicado (*O que trabalhar faz?*), tendo então como resposta a oração seguinte. As duas orações estão tão encravadas uma na outra, que não há como separá-las.

(8) Trabalhar cansa.

Há também o aspecto proposicional. Os sujeitos podem ter sentido que a cada oração corresponde uma proposição, um enunciado verbal que pode ser falso ou verdadeiro. Quando as frases são formadas por orações coordenadas, ou mesmo subordinadas adverbiais, parece ser mais fácil identificar as proposições, como no exemplo abaixo:

(9) Fininho fala inglês e Bolão toca viola.

Cada uma das orações forma um enunciado completo que pode ser negado ou afirmado, até mesmo em conjunto (*Fininho não fala inglês e Bolão não toca viola*), o que não acontece com a mesma facilidade nas orações substantivas. Negar as duas orações no exemplo 10, produziria no mínimo uma certa dificuldade de compreensão (*Não importa não viver honestamente*).

(10) Importa viver honestamente,

Dentre as substantivas, as que apresentaram maior dificuldade foram as predicativas, com um nível de acerto de apenas 18%. Pode haver várias explicações para essa dificuldade, mas uma hipótese interessante é a presença do verbo de ligação nos dois exemplos usados no teste (11 e 12).

(11) O certo é que ele não vem.

(12) A verdade é que todos o amam.

A segmentação correta seria depois do primeiro verbo (*O certo é / que ele não vem*) mas os sujeitos preferiram fazer a segmentação antes do verbo (*O certo / é que ele não vem*). Parecem ter sentido, intuitivamente, que o segmento *o certo é* fica incompleto em termos de enunciado, sem possibilidade de afirmar ou negar qualquer coisa, já que a proposição só se completa quando a segunda oração é processada.

Esses resultados são baseados em frases que foram coletadas apenas sob o critério de representatividade quanto aos tipos de oração, incluindo, por exemplo, os diferentes tipos de orações subordinadas substantivas (subjativas, objetivas, completivas nominais, predicativas), etc. Não houve preocupação em controlar outros fatores, tais como a combinação de diferentes tipos de oração na mesma frase (*Quero trabalhar e estudar*) ou o uso de conjunções ambíguas (*Não sei como chegou lá versus Se não tem dinheiro como vai comprar um carro?*), onde a palavra *como* pode ser uma conjunção ou um advérbio. É provável que se esses fatores fossem considerados, os resultados não apenas poderiam ser diferentes mas também analisados num nível maior de detalhamento.

As conjunções coordenativas, por exemplo, têm a propriedade de ligar não apenas orações com a mesma função mas qualquer segmento do texto. Assim, podemos ter, por ordem decrescente, duas ou mais orações independentes (*Sua namorada partiu e não voltou*), subordinadas desenvolvidas (*Quero que vás e que voltes*), subordinadas reduzidas (*Quero ir e voltar*), sintagmas nominais (*A casa de madeira e o castelo de pedra ruíram*), palavras individuais (*A casa e o castelo ruíram*) e até prefixos (*O palestrante considerou aspectos pré- e pós-modernistas*).

É possível que combinações com esses diferentes tipos de segmentos textuais baixasse os elevados escores obtidos

com a segmentação das orações coordenadas. Note-se, por exemplo, a dificuldade em processar sintaticamente uma frase com diferentes níveis de coordenação como *Maria e Ana ajudaram Pedro e João ficou feliz*.

Comparação do Desempenho

A tabela 2 mostra os escores dos sujeitos desde a 4ª série até o nível universitário. O dado que mais chama a atenção é a total falta de progresso entre os alunos da 4ª série e os alunos universitários, ambos exatamente com o mesmo escore, 31 acertos num total possível de 44.

Tabela 2 — Médias de acertos desde a 4ª série até o nível universitário

Nível	4ª série	5ª série	6ª série	7ª série	1º ano	2º ano	3º ano	Univ.
Média	31	31	31	31	37	33	33	31

Essa semelhança de desempenho é encontrada também entre os diferentes itens. O que é difícil para a 4ª série é igualmente difícil para o nível universitário, o que é fácil para a 4ª série é igualmente fácil para o nível universitário. Isto pode ser visto na Tabela 3, onde a segmentação das orações coordenadas foi igualmente fácil para os dois níveis, com índice de acerto superior a 90%. Os dois itens mais difíceis, envolveram as orações predicativas (exemplos 11 e 12), ficando o índice de acertos abaixo de 19% nos dois grupos e muito próximos um do outro (menos de 1%).

Tabela 3 — Percentual de acertos entre orações dificuldade variada.

Nível	4ª série	Univ.
-------	----------	-------

Predicativas	18,6%	18,2%
Coordenadas	90,1%	94,3%
Uma palavra e sem vírgula	50%	72%

A maior diferença encontrada foi com a oração de uma única palavra, sem vírgula (exemplos 13 e 14), que produziu um percentual de 50% de acertos para os iniciantes e um escore bem mais alto (72% de acertos) para os alunos universitários.

- (13) Fico se ele sai.
- (14) Faço como sou mandado.
- (15) Acho que eles estão estudando.
- (16) Saí porque ele entrou.

Uma análise estatística emparelhando os escores de cada um dos 44 itens, entre os dois grupos (4ª série e universitário), mostra um índice de correlação positiva entre ambos ($r = 0,74$), que é estatisticamente significativo. Os resultados mostram, portanto, que os alunos entre a 4ª série do primeiro grau e o nível universitário, não apenas deixaram de progredir de um modo geral, mas também mantiveram praticamente inalteráveis as mesmas dificuldades e facilidades. O que é fácil para as crianças continua sendo fácil para os adultos. O que é difícil continua sendo difícil, sem chegar a uma solução, apesar de todo o trabalho da escola.

O índice de correlação não é mais alto devido ao progresso registrado nas orações de uma só palavra, que foi o único item onde a escolaridade parece ter influenciado. Um levantamento dessas orações mostra na realidade que o aumento significativo de acertos entre a 4ª série e o nível universitário só ocorreu com orações de uma só palavra que apresentassem as seguintes características: (1) a oração é um verbo conjugado; (2) a oração não é substantiva; (3) a oração não está separada por vírgula. A distinção entre ser ou não substantiva foi crucial. Se a oração é substantiva, ela pode não só acarretar falta de progresso mas até retrocesso, conforme mostra o exemplo 17.

- (17) Importa viver honestamente.

Enquanto que na 4ª série 54% dos sujeitos segmentaram corretamente, colocando o travessão depois do verbo

importa, no nível universitário apenas 14% o fizeram. Todos os outros, incluindo 86% no nível universitário, colocaram a separação depois da palavra *viver*. Houve portanto um retrocesso altamente significativo do ponto de vista estatístico; erraram muito mais no nível universitário do que na 4ª série. Parece que não apenas a escola mas a própria vida ensinaram aos sujeitos que *importa viver* forma um segmento mais coeso e inseparável do que *viver honestamente*.

Não se aventava aqui a questão filosófica de que *importa viver* possa ser mais importante do que *viver honestamente*, e por isso menos separável à medida que se vive e se aprende, mas uma questão puramente sintática, implícita na dificuldade maior de separar o sujeito do verbo. Embora os alunos pesquisados fossem provavelmente incapazes de explicitar que *viver honestamente* é o sujeito gramatical de *importa*, podem ter sentido isso intuitivamente. *Viver honestamente* é uma oração que já inclui um verbo nominalizado, portanto a meio caminho entre sintagma verbal e sintagma nominal e sintaticamente funcionando como um nome, podendo, por exemplo, ser substituído na frase *importa viver honestamente* pela expressão *a vida honesta (importa a vida honesta)*. Em outras palavras, não se está mais separando uma oração da outra, mas os elementos da própria oração, operando, portanto, num outro nível de segmentação — o que pode ter desorientado os alunos.

Há também aqui o problema da diferença entre oração desenvolvida e oração reduzida. *Viver honestamente*, além de ser uma oração substantiva, é também uma oração reduzida, que, como sabemos, pode ser de três tipos: (1) reduzida do infinitivo (*viver*); (2) reduzida do gerúndio (*vivendo*); e (3) reduzida do particípio (*vivido*). Embora sejam todas tratadas como orações nas gramáticas tradicionais, na teoria não há unanimidade. Perini (1989), por exemplo, defende que em orações reduzidas do particípio (*Posto o sol, parti*) o particípio não é um verbo mas um adjetivo. Stefanini, como vimos, após ter estudado o problema, resolveu para a implementação de seu sistema de tratamento automático da língua, interpretar como orações apenas os segmentos que tivessem verbos conjugados.

É possível portanto que se ao invés de uma oração reduzida (*Importa viver honestamente*), tivéssemos uma oração desenvolvida (*Importa que se viva honestamente*), os resultados fossem diferentes, aproximando-se daqueles que foram obtidos com os exemplos de 13 a 16.

IMPLICAÇÕES PEDAGÓGICAS

Este estudo, ao verificar especificamente que a escolaridade não produz qualquer efeito na capacidade do aluno em segmentar corretamente as orações de uma frase complexa leva a duas implicações mais ou menos opostas. A primeira é de que o professor não deve se preocupar com a análise sintática das frases. Os alunos, ao entrarem na escola, já desenvolveram a competência sintática da língua, algo que ocorre e se completa em idade pré-escolar, sendo provavelmente incapaz de ser desenvolvida posteriormente. O trabalho do professor portanto é totalmente inútil, tentando ensinar ao aluno algo que ele já sabe e que ele, professor, não sabe explicitar.

A segunda implicação, menos óbvia na análise dos resultados, é a de que a escola precisa tentar desvelar aquilo que está oculto no inconsciente do falante. Uma explicitação do funcionamento da língua ou pelo menos uma sensibilização da complexidade de seu funcionamento pode ser uma maneira de ajudar nesse desvelamento.

O que não se tem ao entrar na escola é consciência do que se sabe sobre a língua, incluindo a idéia de subordinação. Em qualquer segmento lingüístico autêntico há uma parte que é nuclear e outra que é periférica, incluindo a frase, o fonema e todos os segmentos intermediários. Como na física, desde a estrutura do átomo, onde os elétrons giram ao redor do núcleo, até os sistemas solares, onde os planetas giram ao redor de uma estrela, há também na língua esta estrutura de dependência de uns elementos a outros: a oração subordinada está presa à oração principal, o adjetivo ao substantivo, a sílaba à palavra, e assim por diante. Essa dependência de elementos a

um núcleo é importante para localizar os pontos de segmentação, as fronteiras entre um agrupamento e outro. A palavra dentro de uma frase bem formada nunca está solta, e por mais longe que esteja do núcleo nunca passa para o outro lado da fronteira. A língua parece possuir uma estrutura tão complexa quanto a do cérebro que a contém e uma das funções da escola pode ser a de explicitar ao aluno a beleza que subjaz a essa complexidade.

Uma maneira de fazer isso é problematizar a língua para o aluno, mostrando os problemas que surgem quando certas regras não são aplicadas. Alguns programas menos sofisticados de tratamento automático da língua podem ser uma fonte inesgotável desses tipos de problemas. As frases seguintes, produzidas por um tradutor automático ilustram alguns exemplos de problemas originados por segmentação inadequada devido à incapacidade de identificar o núcleo de um segmento:

*Os alunos da escola chegou.

*O carro de Maria é rápida

A ambigüidade da língua, muitas vezes criadas por questões de segmentação é outra fonte inesgotável de problemas, que pode levar o aluno à conscientização de como é feito o processamento de uma frase. Os exemplos seguintes mostram algumas dessas ambigüidades:

Pintou a parede de madeira.

Pintou a parede de verde.

O guarda viu a menina de binóculos.

O guarda viu a menina de vestido verde.

Todos precisam ser identificados.

Todos precisam seres identificados.

Outra questão interessante para debate é o compartilhamento do mesmo sujeito gramatical por diferentes orações coordenadas, o que não parece ser possível com orações subordinadas. Isso pode ser problematizado para os alunos mostrando que das duas orações abaixo uma parece ser mais aceitável que a outra.

Ele chegou e ele arrumou a sala.

Quando ele chegou ele arrumou a sala.

Um projeto maior, possivelmente para alunos mais adiantados, poderia envolver a elaboração de um algoritmo para a identificação das orações a partir da subcategorização do verbo, tendo como base a gramática de valências (Borba, 1996), mas talvez complementada por uma heurística mais local, ad hoc mesmo, para resolver os problemas reais da língua em uso.

Professores e alunos parecem ter o mesmo domínio da língua, e embora os professores provavelmente tenham desenvolvido em grau mais alto a consciência desse domínio, não o parecem ter desenvolvido em grau suficientemente alto para que possam explicitar o processo ao aluno. A proposta, portanto, não é o ensino daquilo que não sabemos ensinar, mas a aprendizagem pela descoberta daquilo que já sabemos. Se já temos o conhecimento tácito da língua, de certo modo recessivo dentro da mente, o que precisamos fazer é torná-lo manifesto. Para isso, é necessário novas propostas de estratégias de descoberta para que esse conhecimento possa ser desvelado.

Por fim, não se pode ignorar a importância da motivação por parte do aluno. Aqui parte-se do pressuposto de que a motivação existe naturalmente. Não existe uma única pessoa que não goste de falar da própria língua. Imitando o que Johnson disse da cidade de Londres, poderíamos dizer que quem não gosta da língua não gosta de viver, pois tudo que está na vida está na língua. A língua, como o sexo, é um instinto forte e somos apaixonados por ela. A sugestão é não matar essa paixão, evitando partir do que não se sabe e que não faz sentido para o aluno, para o que ele sabe ou acha que sabe e, a partir daí, tentar o caminho da descoberta.

A partição do texto em segmentos que se agrupam em torno de um núcleo é um procedimento fundamental no processamento da língua que o aluno deve tentar desvelar. A implicação pedagógica deste trabalho, portanto, não é a de que o professor deva deixar de trabalhar com a análise sintática, mas de que deve trabalhar mais.

CONCLUSÃO

O objetivo deste estudo foi verificar o impacto a longo prazo do ensino da análise sintática em alunos de primeiro e segundo grau. Para isso elaborou-se um instrumento, composto de 44 frases complexas, que foi aplicado em alunos da quarta série do primeiro grau até a universidade. Os resultados mostraram que a escola em nada contribui para a sensibilização gramatical do aluno em termos de análise sintática externa. Confirma-se a hipótese de que o aluno ao chegar à escola já possui o conhecimento sintático da língua e o que a escola ao tentar ensinar o que ele já sabe em nada contribui para desenvolver ou aprimorar esse conhecimento.

A escola também parece fracassar em conscientizar o aluno do que ele já sabe. Uma possível causa desse fracasso é a aparente falta de coerência entre o que se diz ao aluno e a realidade da língua, conforme ele a percebe. O trabalho mostrou que em muitos aspectos onde não houve progresso ou retenção da aprendizagem por parte do aluno também não há unanimidade entre os próprios estudiosos da língua, como é especificamente o caso das orações reduzidas. Quando o que se tentou ensinar entrou em choque com o conhecimento intuitivo do aluno, o ensino foi consciente ou inconscientemente rejeitado.

A principal implicação pedagógica do trabalho é de que a análise da estrutura sintática da língua não seja descartada do currículo, mas enfatizada — dentro de uma perspectiva que parta da percepção lingüística do aluno. A idéia básica é de que a língua deve ser problematizada para que o aluno tenha ao menos a oportunidade de vislumbrar sua complexidade e perceber alguns dos aspectos fundamentais de sua estruturação. A curiosidade inata do falante pela sua língua forneceria a motivação necessária para o seu estudo.

REFERÊNCIAS

- ABNEY, Steven. Chunks and dependencies: bringing processing evidence to bear on syntax. In: COLE, Jennifer, GREEN, Georgia M, MORGAN, Jerry L. (eds.) *Linguistics and computation*. Stanford, California: CSLI Publications, 1995. P. 145-164.
- ALMEIDA, Napoleão Mendes de. *Gramática metódica da língua portuguesa*. São Paulo: Saraiva, 1992.
- BORBA, Francisco S. *Uma gramática de valências para o português*. São Paulo, Ática, 1996.
- BRAZIL, David. *A grammar of speech*. Oxford: University Press, 1995.
- CHOMSKY, Noam. *Syntactic Structures*. The Hague: Mouton, 1957.
- HOCKET, C. F. A manual of phonology. *International Journal of American Linguistics*. v. 21, n. 4: Memoir no. 11, 1955.
- JESPERSEN, Otto. *Analytic syntax*. London: Allen & Unwin, 1933
- KJAER, A. L. Phraseology research; State of the art. *Terminology Science & Research*, vol. 1, n. 1-2, ITTF, p. 3-20, 1990.
- LEWIS, Michael. *The lexical approach*. London: LTP, 1993.
- MARTINET, André. *Grammaire fonctionnelle du français*. Paris: CREDIF-Didier, 1979.
- NATTINGER, James R., DECARRICO, Jeanette S. *Lexical phrases and language teaching*. Oxford: University Press, 1992.
- OLIVEIRA, Cândido de. *Revisão Gramatical*. São Paulo: Luzir, 1960.
- PAVEL, Silvia. La phraséologie en langue de spécialité; Méthodologie de consignation dans les vocabulaires terminologiques. *Terminologies Nouvelles*, n. 10, p. 67-82, 1993.
- PERINI, Mário A. *Gramática descritiva do português*. São Paulo: Ática, 1996.
- PERINI, Mário. *Sintaxe portuguesa: Metodologia e funções*. São Paulo: Ática, 1989.
- PONTES, Eunice Souza Lima. *Sujeito; Da sintaxe ao discurso*. São Paulo: Ática, 1986.
- QUIRK, Randolph; GREENBAUM, Sidney; LEECH Geoffrey; SVARTVIK, Jan. *A comprehensive grammar of English*. London: Longman, 1985.
- SAUSSURE, Ferdinand de. *Curso de lingüística geral*. São Paulo: Cultrix, 1988, 14^a ed.
- SOUZA E SILVA, M. Cecília P. de; KOCH, Ingedore Villaça. *Lingüística aplicada ao português: Sintaxe*. São Paulo: Cortez, 1995. 6^a ed.
- STEFANINI, Marie-Hélène. *Talisman: Una architecture multi-agents pour l'analyse du français écrit*. Grenoble: Université Pierre Mendès-France, Centre de Recherche en Informatique Appliquée aux Sciences Sociales. 1993.
- WOODS, W. A. Transition network grammars for natural language analysis. *Communications of the ACM*, n. 13, 1970, p. 591-606.