

O PROCESSAMENTO DE CONCATENAÇÕES SINTÁTICAS EM TRÊS TIPOS DE ESTRUTURAS FRASAIS AMBÍGUAS EM PORTUGUÊS¹

Marcus Maia - UFRJ
Shelen Alcântara - UFRJ
Simone Buarque - UFRJ
Fernanda Faria - UFRJ

Resumo

Reportam-se neste artigo estudos de questionário e experimentos de leitura auto-monitorada sobre três tipos de concatenações sintáticas ambíguas em português: orações iniciadas por QUE, ambíguas entre aposição como complemento ou como adjunto; formas verbais ambíguas entre a flexão de presente do indicativo ou o particípio passado; sintagmas preposicionais ambíguas entre aposição ao sintagma verbal ou ao sintagma nominal. Analisam-se essas duplas possibilidades de concatenações nos termos de Frazier & Clifton (1996), respectivamente, como relações sintáticas primárias e secundárias, mapeando-as em termos da caracterização apresentada em Chomsky (2001). Os resultados permitem corroborar a atuação no processamento de frases em português do Princípio da Aposição Mínima (Fodor e Frazier, 1978; Frazier, 1979) e apontam no sentido do encapsulamento do parsing sintático.

Palavras-chave

Psicolinguística, Processamento Sintático, Português

1 Este trabalho reúne pesquisas vinculadas ao projeto “O Processamento das Interfaces Sintaxe/Semântica e Sintaxe/Prosódia na Compreensão de Frases em Línguas Naturais”, coordenado por Marcus Maia (CNPq 301365/94-6). Marcus Maia é professor adjunto de Linguística do Museu Nacional e do Programa de Pós-graduação em Linguística da Faculdade de Letras da UFRJ; Simone Buarque, Shelen Alcântara e Fernanda Faria são alunas da Faculdade de Letras da UFRJ, bolsistas de iniciação científica (PI-BIC/CNPq), com projetos orientados pelo primeiro autor da publicação. Os autores agradecem as sugestões dos pareceristas anônimos da Revista Forum e o Auxílio FUIB/UFRJ - Proc. 6.729.

I – Introdução

A subárea da Psicolinguística conhecida como Processamento de Frases (*Sentence Processing*) constitui um campo de pesquisas extremamente produtivo, embora ainda relativamente pouco praticado no Brasil. O objetivo central das teorias de processamento de frases é o de identificar os procedimentos psicologicamente reais que colocamos em jogo ao produzir e compreender frases. A distinção entre um nível de representação gramatical e um nível de acesso a essas representações, consubstanciada na dicotomia competência x desempenho (Chomsky 1965), permitiu um balizamento da área em que perfilam modelos que pressupõem a independência entre a gramática e o processador como um postulado básico. No âmbito do Programa Minimalista (Chomsky, 1995, 1998, 2000, 2001), pode-se indentificar uma aproximação crescente a questões diretamente relacionadas ao processamento, cuja caracterização passou, inclusive, por uma reconceituação fundamental de Chomsky (1995) para Chomsky (1998), abrindo caminho para a identificação progressiva entre a gramática e o processador. Conforme apontamos em Maia (2000), enquanto Chomsky (1995) assume a hipótese de que o processamento, ao contrário do sistema da competência gramatical, é invariável, Chomsky (1998) propõe, à base de evidências psicolinguísticas recentes, que os sistemas de processamento podem variar de língua para língua, sendo modulados pela gramática de cada língua específica. Essa reformulação, que não significa, necessariamente, a anulação completa das diferenças entre o processador e a gramática, permite, no entanto, que se entretenha a hipótese de que os princípios que norteiam o processamento sejam, concebidos como sensíveis aos princípios gramaticais, o que seria menos plausível em um sistema que admitisse variação da competência, mas não do acesso.

No âmbito dos estudos sobre a compreensão de frases, a dicotomia chomskyana (cf. Chomsky, 1965, p. 83-97), ganha contornos mais nítidos, refletindo-se na relação entre a gramática e o processador ou

parser. O termo inglês *parser* ou *parseador* em português, oriundo do Latim *pars orationis*, indica, em sentido estrito, o procedimento de identificação dos constituintes oracionais e de suas relações hierárquicas, no processo de compreensão. Em sentido lato, o termo tem sido tomado até como sinônimo de desempenho, dificultando a determinação dos processos específicos que têm lugar a cada momento em que ouvimos uma frase.

Conforme discutimos em Maia (2001), Gramática e *parser* têm também sido pensados como um mesmo e único objeto, como na instigante proposta de Mike Dillinger (1992), que resolve a dicotomia, tornando-a um monômio: o *parser* seria a própria gramática em ação. Segundo Dillinger, a relação entre a gramática e o processamento seria comparável à relação entre a Anatomia e a Fisiologia, na Medicina, ou entre a Estática e a Cinemática em Física. Esta maneira de colocar o problema equaciona uma teoria da competência com uma teoria estática da faculdade de linguagem e uma teoria do processamento com uma teoria cinemática da faculdade de linguagem. A diferença entre competência e desempenho corresponderia, então, à presença ou ausência do fator tempo².

As diversas teorias psicolinguísticas têm, de uma forma ou de outra, sido balizadas por essa dicotomia, especialmente na subárea do Processamento Sintático. Quando de sua vinda ao Brasil, em 1996, Noam Chomsky, instado a posicionar-se sobre a relevância do binômio, reafirmou sua naturalidade conceitual, lembrando que "... existe uma diferença categorial entre o que você sabe e o que você faz." (Chomsky, 1997), havendo, pois que se distinguir os dois processos, até no que se refere aos aspectos não estritamente linguísticos da cognição. Não obstante, a dicotomia apresenta tensão variável ao longo da história da Psicolin-

2 O fator tempo não seria o único a desempenhar um papel nessa diferença, havendo que se levar em consideração os algoritmos específicos de acesso "top down" aos princípios gramaticais, conforme sugerido por um dos revisores anônimos.

güística, sendo particularmente interessante observar-se, como já aludimos acima, sua reformulação no âmbito do programa minimalista, etapa mais recente da teoria gerativa (cf. Weinberg, 1999).

Em seu mais recente artigo, *Beyond Explanatory Adequacy* (2001), Chomsky aborda especificamente a derivação de estruturas de complementação comparativamente à derivação de estruturas de adjunção. A computação das frases é concebida como um processo em que os itens lexicais são selecionados (*Select*) e, em seguida, concatenados (*Merge*), formando objetos estruturais que serão enviados para a interpretação nas interfaces (*Spellout*). A concatenação mais básica é aquela em que dois itens são integrados em um conjunto (*Set Merge*), por exemplo, a concatenação entre um núcleo e seu complemento. Além desse tipo de concatenação, Chomsky propõe um outro para explicar a junção de adjuntos. Nesse tipo, denominado de concatenação de par (*Pair Merge*), há duas operações de junção: na primeira, um elemento é adicionado ao outro, formando um par, em que um elemento se mantém independente do outro, devendo ser, em um segundo momento, simplificado (*Simplify*) quando da próxima junção com outro item. É interessante notar que esta distinção fornece estatuto teórico ao modelo de processamento conhecido como *Construal*, proposto em Frazier & Clifton (1996). Segundo este modelo, que propõe uma revisão importante da teoria do *Garden Path*³ (Frazier & Fodor, 1978; Frazier, 1979), diferenciam-se relações sintáticas primárias de relações não-primárias, sendo as primeiras exemplificadas como a relação do tipo sujeito-predicado ou aquela que se estabelece entre um núcleo e seu complemento, enquanto que as segundas seriam elaborações de posições argumentais através de adjuntos. Frazier & Clifton

3 Dillinger (1992) traduz o termo "garden-path" como labirinto. No presente artigo, alternaremos livremente o uso dos dois termos para expressar o mesmo fato.

(1996) propõem que o mecanismo de processamento de frases (*parser*) é capaz de distinguir entre esses dois tipos de relações sintáticas, procedendo de maneira específica ao computá-las. No caso das relações primárias, tais como a concatenação de um núcleo a seu complemento, como previsto na teoria do *Garden-Path*, os fatores estritamente sintáticos são prioritários na construção da estrutura sintática pelo processador, invocando-se o princípio da Aposição Mínima (*Minimal Attachment - M.A.*), que leva o processador a decidir pela estrutura com menos nós quando confrontado com ambigüidades sintáticas, ou o princípio da Aposição Mais Recente (*Late Closure - LC*), quando as estruturas ambíguas apresentam o mesmo número de nós. A afirmação fundamental da teoria do *Garden Path* é que: (1) o *parser* usa uma porção do seu conhecimento gramatical isolado do conhecimento de mundo e outras informações para a **identificação inicial** das relações sintagmáticas; (2) o *parser* confronta-se com sintagmas de aposição ambígua e compromete-se com uma **estrutura única**; (3) Pressionado pela arquitetura do sistema de memória de curto prazo, que tem um limite estreito de processamento e armazenamento, o *parser* segue um princípio psicológico na escolha desta estrutura: use o menor número possível de nós (**Minimal Attachment**.) e, se duas aposições mínimas existem, aponha cada nova palavra ao sintagma corrente (**Late Closure**). Os fatores semânticos e pragmáticos não seriam capazes de influenciar a decisão do *parser*, atuando apenas no segundo passe, quando a frase pode ser revista pelo processador temático. Na reformulação dessa versão estrita da teoria do *Garden Path*, desenvolvida no modelo de *Construal*, Frazier & Clifton (1996), propõem que, no caso das relações secundárias, como, por exemplo, a aposição de uma oração relativa a um SN, a decisão estrutural do processador não é tão automática e estritamente sintática quanto no caso das relações primárias, postulando que a oração ambígua seja associada (e não diretamente aposta) ao marcador frasal em cons-

trução através do sistema de construal, permitindo que fatores semânticos e pragmáticos influenciem a interpretação da estrutura, contribuindo para a identificação pelo parser da análise preferencial. No presente artigo, estaremos investigando o processamento de três tipos específicos de estruturas sintáticas do português que instanciam esses dois tipos de relações sintáticas.

II – Três tipos de concatenações sintáticas

Os três estudos reportados a seguir investigam o comportamento do *parser* quando defrontado com dupla possibilidade de concatenação sintática. Embora testando três tipos de ambiguidades sintáticas distintas, os estudos têm como referência a Teoria do *Garden Path* e adotam os mesmos procedimentos experimentais. São conduzidos, em um primeiro momento, estudos *offline*, com base em questionários a serem completados por voluntários, sem medida de tempos em milésimos de segundos. Em um segundo momento, os três estudos reportam experimentos de leitura auto-monitorada, utilizando o programa Psyscope, a fim de medir tempos de leitura em milésimos de segundos. O estudo reportado na seção 2.1. analisa as preferências de aposição sintática no caso de orações subordinadas iniciadas por QUE, que têm como matriz uma oração nucleada por um verbo “dicendi”. A ambigüidade decorre da possibilidade de integrar a subordinada como substantiva objetiva direta ou como adjetiva, formando um SN complexo. Observe-se que, no caso da preferência pela aposição como substantiva, o processador será conduzido a um efeito labirinto (*garden-path*), pois não terá como integrar a oração subsequente, precisando voltar e reanalisar a segunda oração como adjetiva:

(1) A babá explicou para a criança / QUE estava sem sono / que a mamãe iria chegar à noite.

O estudo relatado em 2.2. investiga estruturas em que uma forma verbal ambígua entre presente do indicativo e particípio se segue a um SN, admitindo ser aposta a ele como um verbo principal ou como uma oração adjetiva reduzida, formando um SN complexo. Como se observa no exemplo em (2), a preferência pela aposição como verbo no presente do indicativo poderá levar o leitor ao efeito *garden-path*, pois ao encontrar a oração subsequente não terá como integrá-la no primeiro passe, já que se comprometeu com a estrutura mínima. Deverá, então, reanalisar sintaticamente a estrutura, de forma não mínima, isto é, interpretando a forma verbal como particípio, formando um SN complexo:

(2) A empresária /PAGA com antecedência de um mês/ exige exclusividade.

Finalmente, o estudo apresentado em 2.3. explora a compreensão de estruturas em que um sintagma preposicional pode apresentar ambigüidade de aposição sintática como um adjunto adverbial ou como um adjunto adnominal, formando um SN complexo:

(3) O policial /viu o turista/com o binóculo.

Observe-se que o SP “com o binóculo”, no exemplo acima, pode ser tomado como o instrumento com o qual o policial viu o turista (estratégia de aposição do SP ao SV) ou como um modificador do SN “o turista” (estratégia do SN complexo, não mínima).

Os três estudos têm em comum ainda o fato de haverem manipulado efeitos semânticos e/ou pragmáticos, tanto nos estudos de questionário, quanto nos experimentos de leitura auto-monitorada, com a finalidade de investigar a interferência de fatores não estritamente estruturais no processamento sintático.

2.1. Concatenação de complemento como oração substantiva ou adjetiva

Ao encontrar uma ambigüidade estrutural, o parser parece decidir-se rapidamente por uma das análises possíveis, havendo-se postulado princípios, tais como a “aposição mínima” (*minimal attachment*) e a “aposição baixa” (*late closure*), à base, principalmente, de dados da língua inglesa (Frazier, 1979; Frazier & Rayner, 1982). Com base nestes estudos sobre a compreensão de frases na língua inglesa, tem-se procurado determinar a natureza das informações usadas pelo *parser* para guiar suas decisões de processamento, investigando-se se o parser é estritamente sintático em sua análise inicial ou se já acessa rapidamente informações de natureza semântica e pragmática a tempo de influenciar a análise sintática.

O presente estudo pretendeu, então, investigar a ambigüidade de de aposição sintática em Português do tipo Oração Substantiva x Oração Adjetiva. Frases tais como “O patrão disse ao funcionário que... nunca chegava atrasado que lhe daria aumento”, constituem frases local e temporariamente ambíguas, uma vez que a palavra “que” pode introduzir tanto uma oração subordinada substantiva como uma oração subordinada adjetiva. A ambigüidade é local e temporária porque, mais adiante, na frase, o leitor deverá decidir por apenas uma das interpretações possíveis. Como resenhado na introdução do artigo, a estratégia preferencial, neste caso, deveria ser a concatenação mínima, que instancia uma relação sintática primária, a relação núcleo-complemento. Esta decisão constrói uma estrutura com um número de nós menor do que a concatenação não-mínima que instancia uma relação sintática secundária, a relação de adjunção, caso em que deveria se montar um sintagma nominal complexo, computando-se uma estrutura com maior número de nós. Foram realizados dois estudos: questionário e leitura auto-monitorada, reportados abaixo.

2.1.1. Estudo de questionário

Foram elaborados dois questionários para o experimento *off line*, que foram aplicados em 20 alunos de graduação da UFRJ, a fim de testar a preferência de produção de frases em estruturas ambíguas que permitiam continuações tanto com oração substantiva quanto com oração adjetiva.

2.1.1.1. Procedimentos e Materiais

O experimento *offline* reuniu 60 frases: 40 frases distratoras e 20 frases experimentais. A primeira frase sempre era a experimental, seguida de duas distratoras, e assim sucessivamente. Por frases distratoras, entendam-se aquelas presentes nos questionários com o intuito de “distrair” a atenção dos entrevistados, para que estes não percebam o objetivo do estudo ou a questão teórica implícita nas frases respondidas. As experimentais, por oposição, eram aquelas realmente focalizadas no estudo, contendo as estruturas relevantes para se testar a hipótese, no caso a preferência pela aposição de oração substantiva ou adjetiva. O fato de havermos distribuído, nesse questionário, as frases experimentais em alternância regular com as distratoras, ao invés do procedimento randomizado, garantiu o não seqüenciamento das frases experimentais sem, no entanto, facilitar a identificação das estruturas testadas por parte dos sujeitos que, de modo geral, não souberam precisar de que se tratava o questionário, conforme verificado em entrevistas posteriores à sua aplicação.

Os questionários pediam que se lessem períodos, completando-os da forma que achassem mais adequada. Os dois questionários propunham contextos distintos: um incluiu um contexto chamado (+) plausível e um outro um contexto denominado (-) plausível. O questionário que privilegiou o contexto (+) plausível continha 20 frases experimentais (além das 40 distratoras) que foram compostas, cada uma, por dois períodos, sendo que o primeiro visava a

manipular a plausibilidade de ocorrência da oração adjetiva no segundo período. Vejamos um exemplo:

(4) O patrão convocou dois funcionários. O patrão prometeu ao funcionário que ...

Observe-se, com base no exemplo (4), que incluiu-se no contexto anterior mais de um referente para o objeto indireto da oração principal do período seguinte, visando a verificar se esta inclusão facilitaria a opção pela continuação do segundo período com oração adjetiva, que serviria para identificar mais precisamente o referente do SN objeto. O entrevistado deveria decidir, após a leitura do primeiro período (O patrão convocou dois funcionários), entre continuar a frase como estando diante de um QUE pronomes relativo ou como estando diante de um QUE conjunção integrante. Nosso objetivo, através da aplicação deste primeiro piloto, foi verificar, então, qual seria a preferência de produção dos sujeitos.

Ainda a respeito dos questionários, faz-se necessário observar que todos os verbos neles presentes – com exceção dos verbos das frases distratoras – se enquadram no grupo de verbos denominados “dicendi”, verbos estes que são introdutores de opinião. A presença desses verbos nos pilotos não foi aleatória. Trata-se de verbos com dupla transitividade, ou seja, aqueles que necessitam de um complemento direto, bem como de um indireto, sendo eles: prometer, reclamar, explicar, informar, declarar, negar, afirmar, dizer, ensinar, comunicar, assentir, perguntar, indagar, responder, retrucar, contestar, pedir, solicitar, ordenar, aconselhar.

O questionário que privilegiou o contexto (-) plausível privilegiou a neutralidade no contexto, contendo também 20 frases experimentais (além das 40 distratoras), que eram compostas, cada uma, por também dois períodos, sendo que o primeiro incluía um

critério denominado não-plausibilidade que, em oposição ao contexto plausibilidade, não deveria favorecer a produção de orações adjetivas. Vejamos um exemplo:

(5) O patrão convocou uma reunião. O patrão prometeu ao funcionário que ...

No exemplo (5), o entrevistado não esteve diante de um contexto que favorecesse a produção de orações adjetivas e, portanto, após a leitura do primeiro período (O patrão convocou uma reunião.) mais provavelmente ele continuaria a frase como estando diante de um “que” conjunção integrante. Cada um dos questionários foi aplicado a vinte sujeitos diferentes, em sua maioria alunos de graduação da UFRJ, passando-se, em seguida, à análise das frases experimentais – aquelas realmente relevantes para o projeto – e à computação dos dados.

2.1.1.2. Resultados

Foram preenchidos dois pilotos, totalizando 40 questionários respondidos, o que representou 400 observações preenchidas (20 frases experimentais preenchidas por 20 sujeitos). Houve 20 respostas ambíguas (frases que ainda permitissem dupla interpretação ou outra interpretação que não aquela em estudo) e 4 respostas nulas (frases que não foram preenchidas), que foram excluídas da análise, todas elas referentes ao contexto menos plausível.

A Tabela 1 contém um resumo das informações obtidas pelos questionários. Nesta tabela estão as contagens das preferências pelo pronome relativo e pela conjunção integrante nos contextos mais plausível (+P), menos plausível (-P) e no contexto geral, incluindo ambos os fatores de plausibilidade, bem como os percentuais relativos à escolha pelo pronome relativo em cada um desses contextos.

	+P	-P	GERAL
Conjunção Integrante	275	353	628
Pronome Relativo	125	23	148
% Pronome Relativo	0.31	0.06	0.19

Tabela 1 – Dados do Questionário

Um teste sobre a proporção geral (19%) mostrou que esta é significativamente menor do que 50% ($N = -21.93$, $p \cong 0$). Mesmo no contexto mais plausível, não há preferência maior do que 50% pelo pronome relativo ($N = -8.0904$, $p \cong 0$). De toda forma, o contexto teve forte influência sobre as respostas: no contexto mais plausível houve um aumento significativo da proporção de preferências pelo pronome relativo ($t(397) = 9.56$; $p \cong 0$).

Os resultados demonstram, claramente, a preferência dos sujeitos pela produção com conjunção integrante, o que corresponde a 88,25 % do total de dados analisados. Interessante se faz contrastar os resultados dos dois contextos, o (-) e o (+) plausível, em relação à preferência pela produção com pronome relativo. No contexto (+) plausível, essa produção corresponde a 31,25 % dos dados analisados, ao passo que, no contexto (-) plausível, temos uma diminuição significativa dessa preferência, correspondendo apenas a 5,75 % do total de dados. Observe-se que em ambos os casos, o princípio da aposição mínima parece estar funcionando, levando à produção preferencial de estruturas com menor complexidade estrutural. Entretanto, o contexto + plausível fez com que a preferência pela aposição mínima diminuísse sensivelmente, ainda que se mantendo como a opção preferencial. Vejamos a Tabela 2:

	CONJUNÇÃO INTEGRANTE	PRONOME RELATIVO
(+) Plausível	68,75 %	31,25 %
(-) Plausível	88,25 %	5,75 %

Tabela 2 – Percentuais comparativos

Os dados expostos acima nos permitem observar que, apesar de a plausibilidade buscar facilitar a produção de orações adjetivas (*pair merge*), isto na prática não se observou. Os resultados evidenciaram o contrário: a preferência dos entrevistados se dá pela produção com conjunção integrante (*set merge*), o que corresponde a 275 dados, 68.75 % em valor percentual.

Os resultados contrastantes obtidos no experimento *offline* parecem corroborar a hipótese freqüentemente entretida na literatura psicolinguística de que informações não estritamente estruturais, tal como a plausibilidade pragmática, poderiam estar influenciando as decisões do processador na construção da estrutura (cf. MacDonald et alii (1994) e Sedivy, J. & Spivey-Knowlton, M. (1994), para uma revisão das teorias que atribuem papel relevante aos fatores não sintáticos na resolução da ambiguidade). Entretanto, uma afirmação sobre o processo de *parsing* só poderia ser feita a partir de dados *on line*, que capturassem decisões rápidas, mensuráveis em milésimos de segundos. Incapazes de capturar esses processos, os questionários poderiam estar refletindo apenas a etapa interpretativa, pós-sintática, da compreensão, quando fatores semânticos e pragmáticos seriam considerados, completando e/ou corrigindo as decisões estritamente sintáticas do *parser*. Assim, terminada esta etapa de aplicação e análise dos experimentos *offline*, projetamos a aplicação de um instrumento *on line* que nos possibilitasse identificar com precisão se as diferenças encontradas no estudo de questionário poderiam ser atribuídas à fase de *parsing* da compreensão de frases ou se deveriam ser melhor caracterizadas como resultado da interpretação pós-sintática.

2.1.2. O experimento *on line*

A etapa anterior – a *offline* – fora uma etapa de produção, não-reflexa, incapaz de aferir o processamento sintático que se daria de modo reflexo, automático. O experimento *on line* utilizou o programa computacional para sistema operacional *Macintosh* “*Psyscope*”, desen-

volvido por pesquisadores da *Carnegie Mellon University* para a investigação dos processos cognitivos em ação na compreensão de frases. Nesse sentido, foi realizado um experimento de leitura auto-monitorada, a fim de testar a preferência rápida do processador e a influência de fatores semânticos e pragmáticos no *parsing* sintático.

2.1.2.1. Procedimentos e Materiais

O experimento consistiu de um estudo de leitura auto-monitorada em que os sujeitos, apertando um botão amarelo na “*button-box*”, caixa de botões especialmente projetada para funcionar acoplada a computador *Macintosh*, rodando o programa *Psyscope*, faziam diferentes períodos e segmentos oracionais aparecerem na tela do computador, de maneira não cumulativa, isto é, cada segmento novo substituía o anterior. Após a leitura do último segmento de cada frase, uma pergunta sobre o conteúdo da frase aparecia na tela, só desaparecendo quando o sujeito apertasse a tecla verde da caixa de botões, para responder afirmativamente, ou a tecla vermelha, para responder negativamente. Esse controle, comum em experimentos desse tipo, visava a possibilitar que se retirasse da base de tempos de leitura, computados pelo *Psyscope*, aqueles referentes às frases cujas respostas às questões interpretativas estivessem incorretas, procurando evitar um procedimento de leitura puramente mecânico. Pedia-se aos sujeitos do experimento que lessem cada frase para compreensão, da forma mais rápida e natural possível e respondessem, em seguida a cada frase, a questão interpretativa. Após este evento, o sujeito era instruído a pressionar o botão amarelo de modo a fazer iniciar a próxima frase a ser lida.

24 sujeitos, em sua maioria alunos de graduação da UFRJ, com visão normal ou corrigida, apresentaram-se como voluntários para o estudo. Cada um foi testado em sala isolada da Faculdade de Letras da UFRJ, onde funciona o LAPEX – laboratório de psicolinguística experimental. Um conjunto de 16 frases experimentais e

32 distratoras foi apresentado de forma randômica a cada participante, em sessões de leitura auto-monitorada com duração média de quinze minutos. O projeto do experimento previa a apresentação das frases experimentais em um desenho do tipo “entre sujeitos” (*between subjects*), de modo que cada sujeito fosse exposto a todas as condições experimentais, mas a apenas uma versão de cada conjunto de frases, caracterizando um desenho conhecido como *Latin Square*. Logicamente, como cada sujeito foi exposto a todas as condições experimentais, as condições constituíam um fator *within subject*. As condições experimentais +PL, -PL, +PC, -PC, são exemplificadas em (6), abaixo, sendo as barras inclinadas indicativas dos segmentos apresentados:

(6) Conjunto de dados ilustrativos das condições experimentais

- +PL(+Plausível longa): Havia duas crianças na sala. / A babá explicou para a criança / que estava sem sono / que a mamãe só iria chegar à noite.
- PL (-Plausível longa): Havia uma criança na sala. / A babá explicou para a criança / que estava sem sono / que a mamãe só iria chegar à noite.
- +PC (+Plausível curta): Havia duas crianças na sala. / A babá explicou para a criança / que a mamãe só iria chegar à noite.
- PC (-Plausível curta): Havia uma criança na sala. / A babá explicou para a criança / que a mamãe só iria chegar à noite.

Observe-se que o segmento crítico, cujos tempos médios de leitura foram computados, era a última oração de cada período. Assim, no exemplo acima, a oração “que a mamãe só iria chegar à noite” teve seus tempos de leitura computados.

2.1.2.2. A lógica do experimento

A lógica do experimento era a de que os tempos de leitura do segmento crítico deveriam variar significativamente entre as condições longas e curtas se o Princípio da Aposição Mínima for de fato aplicável no processamento das frases. Note-se que, nas condições longas, há a inclusão, no segundo período, logo após a oração principal, de uma oração iniciada por QUE. No exemplo acima, esta oração é “que estava sem sono”, ambígua entre uma aposição como argumento (*set merge*) ou como adjunto (*pair merge*). Caso o Princípio da Aposição Mínima seja operante em português, de maneira semelhante ao que se demonstrou para o inglês (Frazier, 1978; Frazier & Rayner, 1982), espera-se que a opção pelo QUE integrante seja preferida, levando o leitor a construir incrementacionalmente o período “A babá explicou para a criança que estava sem sono”, em que a segunda oração é apostada à primeira como uma substantiva objetiva direta, satisfazendo, de modo imediato, a grade de subcategorização do verbo *dicendi* “explicar”, que exige um argumento. Neste caso, ao encontrar o segmento crítico, o leitor experienciaria um efeito labirinto (*garden-path*), pois não teria como anexá-lo à estrutura anteriormente construída, devendo reanalisar essa estrutura, reinterpretando a oração anterior como adjetiva e permitindo o encaixe substantivo da última oração ao verbo *dicendi*. Assim, no exemplo em exame, o segmento “que a mamãe só iria chegar à noite”, ocorrendo depois do período longo (PL) “A babá explicou para a criança que estava sem sono”, deveria registrar maior tempo médio de leitura do que quando ocorre nas condições PC, em que vem imediatamente após a oração principal, sendo diretamente integrada como complemento oracional e não provocando o efeito labirinto.

Além dessa hipótese, o experimento *on line* visava ainda a testar a interação entre o fator plausibilidade e a decisão de concatenação, sugerida pelo resultado do experimento *offline*, como descrito acima. Caso essa interação existisse, o último segmento da condi-

ção +PL deveria ser lido mais rapidamente do que o segmento correspondente da condição -PL, uma vez que a condição +PL deveria aumentar a plausibilidade de aposição da oração imediatamente após à principal como adjetiva, evitando o efeito labirinto ao chegar-se na última oração, que poderia, então, ser anexada à estrutura como substantiva, sem necessidade de revisão. Como vimos acima, essa diferença dos tempos de leitura, medidos na unidade de milésimos de segundos, poderia, ao contrário do estudo baseado no preenchimento de questionários, indicar, de fato, se o fator plausibilidade seria acessado rapidamente no processo de leitura, a tempo de influenciar a decisão do parser.

2.1.2.3. Resultados

Os resultados dos tempos médios de leitura dos segmentos críticos das frases experimentais estão representados no Gráfico (1) e resumidos na Tabela (3).

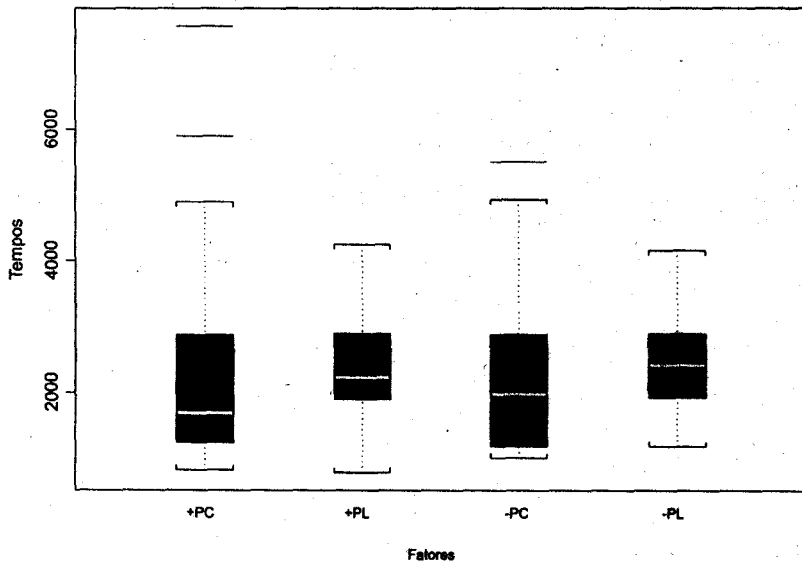


Gráfico 1 – Tempos médios de leitura

No Gráfico 1, os tempos médios de leitura dos segmentos críticos para as frases com a condição curta parecem ser menores e mais dispersos do que aqueles para a condição longa. Na Tabela (3), o efeito da dispersão pode ser notado na coluna desvio padrão e na distância interquartil (3º quartil – 1º quartil). O efeito da locação pode ser notado nos valores das medianas mas são muito sutis nos valores das médias.

CONDIÇÃO	MÍNIMO	1º QUARTIL	MEDIANA	MÉDIA	3º QUARTIL	MÁXIMO	DESVIO PADRÃO
Geral	784	1560	2130	2350	2870	7570	1065,9
+PC	824	1280	1690	2290	2850	7570	1439,3
-PC	1010	1190	1980	2270	2870	5500	1215,9
+PL	784	1910	2220	2370	2880	4240	740,23
-PL	1190	1930	2400	2450	2880	4150	741,82

Tabela 3 – Resultados dos tempos médios de leitura dos segmentos críticos

O gráfico e as estatísticas indicam a presença de outliers na amostras. Estas observações muito grandes camuflam o efeito das condições, como fica claro na Tabela 4, que contém os p-valores da ANOVA para os dados originais e para os dados com *cutoff* das maiores observações.

	NÚMERO DE OBSERVAÇÕES RETIRADAS									
	0	1	2	3	4	5	6	7	8	9
P-Valores	0,85	0,56	0,29	0,23	0,15	0,06	0,07	0,02	0,02	0,01

Tabela 4 – p-valores da ANOVA sem e com *cutoff* de até nove observações

O Gráfico 2 abaixo repete o Gráfico 1 sem as 9 maiores observações.

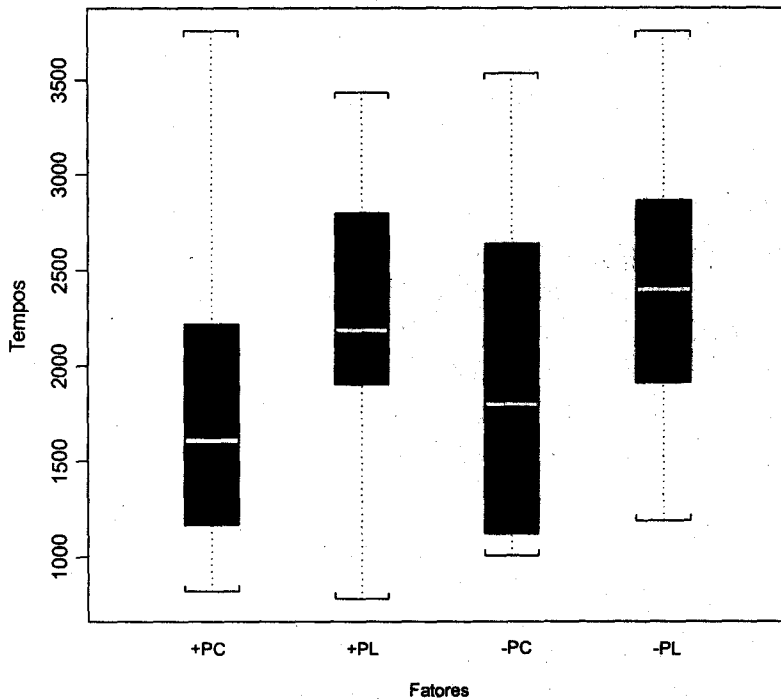


Gráfico 2 – Tempos médios de leitura dos segmentos críticos com cutoff

Para essa amostra podada uma ANOVA com um fator (cujos níveis eram os tipos de frases: +PC, -PC, +PL, - PL) mostrou um resultado significativo ($F(3,147) = 4,17, p = 0,01$). O que significa que há diferenças nos tempos de leituras para cada tipo de frase. Dando prosseguimento a análise fizemos comparações “pairwise” com testes t bilaterais. A tabela (5), abaixo, contém essas comparações

CRUZAMENTO	TESTE
+PC x -PC	$t(72.98) = 0.06$; $p = 0.95$
+PC x -PL	$t(71.30) = -2.61$; $p = 0.01$
+PC x +PL	$t(70.70) = -2.11$; $p = 0.04$
-PC x -PL	$t(69.19) = -2.78$; $p = 0.01$
-PC x +PL	$t(68.51) = -2.26$; $p = 0.03$
+PL x -PL	$t(83.94) = -0.60$; $p = 0.55$

Tabela 5 – Comparações *pairwise* entre as condições experimentais

A tabela acima indica que o efeito que realmente importa é o de curta versus longo. Um teste-t revela que esse efeito é bem significativo ($t(141.7) = -3.47$; $p = 0,00$). Já o fator plausibilidade sozinho não se mostrou significativo ($t(158.93) = -0.44$; $p = 0.66$).

Essas diferenças significativas entre curtas e longas atestam que houve um efeito *garden-path* na leitura do último segmento, evidenciado pelos tempos médios elevados tanto em +PL quanto em -PL (cf. Tabela 3). O fato de que o cruzamento de +PL e -PL não produziu diferenças significativas demonstra que o fator plausibilidade não foi relevante no processamento *on line*.

2.2. Concatenação de SN sujeito a verbo principal ou a oração adjetiva reduzida

O tipo de estrutura examinado nesse estudo é equivalente ao famoso exemplo de Tom Bever “*The horse raced past the barn fell*”. Em inglês, a ambigüidade ocorre porque o passado simples e a forma de particípio passado são idênticas em muitos verbos, permitindo que se interprete a forma “*raced*” como verbo principal da oração, o que levará a um efeito labirinto, quando o leitor encontrar mais adiante a forma “*fell*”, devendo, então, reanalisar a forma “*raced*” como particípio.

Utilizando frases deste tipo, Ferreira & Clifton (1986) manipularam o caráter [\pm animado] do primeiro SN a fim de determinar se a informação semântica ou temática poderia ser usada pelo sistema de processamento para eliminar o *garden-path* em relação com controles

com relativas não reduzidas. Seu raciocínio era o de que SN's inanimados não poderiam ser agentes plausíveis e assim poderiam, em princípio, fornecer feedback temático para o *parser*, que bloquearia uma análise em termos de oração principal com verbo no passado. Usando técnica de monitoramento da fixação ocular (*eye-tracking*), Ferreira & Clifton encontraram evidência contra o uso imediato da informação temática. Observem-se, a título de exemplo, um dos conjuntos de frases estudadas por eles:

(7) The defendant examined by the lawyer was unreliable.

“O réu examinou/examinado pelo advogado não era confiável”

(8) The evidence examined by the lawyer was unreliable

“A prova examinou/examinada pelo advogado não era confiável”

Os resultados reportados por Ferreira & Clifton apóiam a posição de que a inconsistência de significado de frases como (8), em que o SN “*the evidence*”, sendo inanimado não poderia tomar a forma verbal “*examined*” no sentido agentivo, não seria suficiente para bloquear o efeito labirinto. Os autores encontraram o mesmo padrão de leitura para os dois tipos de frase, concluindo que os leitores sempre perseguem a análise em termos de oração principal, em conformidade com o Princípio da Aposição Mínima, sem levar em conta no processamento inicial a informação de natureza semântica.

Como vimos, na língua inglesa, a forma nominal de particípio passado e a forma de flexão modo-temporal de passado simples são homônimas em verbos regulares. Em função dessa homonímia, há ambigüidade na compreensão de frases cujos verbos apresentem essas formas, podendo tais frases ser analisadas tanto como oração principal quanto como oração subordinada adjetiva reduzida de particípio. O Princípio da Aposição Mínima (Minimal Attachment, Frazier, 1978) prevê que, diante de casos assim, haja uma

preferência por considerar a oração como principal, computando-se um menor número de nós sintáticos do que na aposição como oração adjetiva, que formaria um SN complexo. Frazier propôs esse princípio como universal. O presente estudo visou a observar se a preferência pela aposição mínima acontece também na língua portuguesa, posto que homonímia equivalente também ocorre em português com alguns verbos irregulares e/ou abundantes, coincidindo a forma de particípio irregular com a forma de flexão modo-temporal de presente do indicativo. Assim, frases como (9) seriam casos em que o leitor/ouvinte sofreria um efeito labirinto (*garden-path*) se preferisse interpretar a forma verbal “suspeita” como verbo principal e não como um particípio.

(9) Mãe suspeita de assassinato do filho foge

O presente estudo manipulou ainda os traços [\pm humano] do SN sujeito da oração, com a finalidade de verificar se o processador poderia utilizar informações de natureza semântica a tempo de evitar o efeito labirinto. Foram realizados um estudo “offline” e um estudo “online”.

2.2.1. Estudo de questionário

2.2.1.1. Participantes e Materiais

Foram preparados dois questionários contendo cada um 16 frases experimentais e 32 frases distratoras, a serem completadas por 17 voluntários, alunos de graduação da UFRJ. Um dos questionários utilizava SN's do tipo [+humano] e o outro SN's do tipo [-humano], como exemplificado em (10) e (11), a seguir:

(10) A repórter oculta...

(11) A rocha oculta...

2.2.1.2. Resultados

Os resultados do questionário são apresentados na Tabela 6, abaixo, em que se computam as preferências mínimas e não mínimas para cada verbo, no questionário com SN's [+humanos] e no questionário com SN's [-humanos]. A tabela também registra os casos em que a continuação da frase nos questionários não permitiu que se estabelecesse inequivocamente a preferência pela aposição mínima ou não mínima, computadas como ambíguas.

	+HM	+HN	+H AMBÍGUOS	-HM	-HN	-H AMBÍGUOS
suspeita	14	3	-	2	-	15
expulsa	13	4	-	17	-	-
paga	16	-	1	16	-	1
isenta	4	10	3	13	1	3
oculta	14	3	-	11	6	-
aceita	17	-	-	17	-	-
limpa	17	-	-	2	15	-
suja	3	14	-	3	14	-
salva	17	-	-	17	-	-
pega	17	-	-	17	-	-
alerta	11	-	6	5	-	12
desperta	15	-	2	17	-	-
descalça	-	17	-	1	16	-
inquieta	1	16	-	7	10	-
confessa	17	-	-	16	1	-
TOTAL	190	70	12	170	71	31
		272			272	
%	69,9	25,7	4,4	62,5	26,1	11,4
		100			100	

Tabela 6 – Resultados dos questionários offline

Um teste sobre a proporção geral (71%) mostrou que esta é significativamente maior do que 50% ($N = 10.88$, $p \cong 0$). Vemos, na Tabela 7, que na condição +H o percentual de mínimas é maior do que na condição -H, porém esta diferença não é significativa ($t(248) = 0.63$; $p > 0,5$).

	+H	-H	GERAL
Mínima	190	170	360
Não-mínima	70	71	141
% Mínima	0.73077	0.70539	0.71856

Tabela 7 – Dados do Questionário

Observe-se que das 272 observações obtidas para cada questionário (16 frases experimentais completadas por 17 voluntários), registram-se preferências significativas pela estrutura mínima, tanto no questionário em que os SN's tinham o traço [+humano], quanto no questionário em que os SN's tinham o traço [-humano]. Note-se, entretanto, que algumas formas verbais apresentaram maior sensibilidade ao caráter [+humano]/[-humano] do SN, como as formas “isenta”, “limpa” e “fixa”.

2.2.2. Estudo “on line”

O estudo teve por objetivo monitorar a compreensão das frases em milésimos de segundos, de modo a verificar se o processamento preferencial da forma é por posição mínima, testando a teoria do *Garden Path* com dados da língua portuguesa.

2.2.2.1. Procedimentos e Materiais

O estudo utilizou a técnica de leitura auto-monitorada, implementando-se o protocolo no programa *Psyscope*, em computador *Macintosh*, equipado com uma “*button-box*”, caixa de botões especialmente projetada para ser utilizada em conjunto com o programa *Psyscope*. As frases foram apresentadas em quatro segmentos, cujo tempo de permanência na tela era monitorado pelo leitor, que era instruído para ler de modo mais natural e rápido possível. As frases eram monitoradas, apertando-se a tecla amarela na caixa de botões. O quadro (8) abaixo exemplifica a segmentação das frases experimentais.

SEGMENTO 1	SEGMENTO 2	SEGMENTO 3	SEGMENTO 4.	Pergunta
SN sujeito	V	complemento	(conjunção) V + complemento.	
A moça	isenta	de pagamento de mensalidades	(e) divulga o curso.	A moça divulga o curso?

Tabela 8 – Estrutura das frases experimentais

Além das quatro condições experimentais, incluíram-se ainda quatro condições de controle, compostas por formas verbais não-ambíguas, em que as formas de presente do indicativo não são homônimas com as formas de particípio (e.g. fala/falada). A estrutura experimental seguiu o padrão “entre-sujeitos” (*between-subjects*), em que cada sujeito foi exposto a todas as condições experimentais e de controle em teste, mas não às diferentes versões da mesma frase, que eram comparada entre os sujeitos. Cada uma das oito condições foi exposta a cada sujeito quatro vezes, em um total de 32 frases experimentais e de controle. Estas frases foram apresentadas em distribuição randômica com outras 32 frases distratoras (*fillers*), totalizando uma apresentação total de 64 frases por sujeito. Após o monitoramento da leitura de cada frase, uma pergunta interpretativa aparecia automaticamente na tela, devendo ser respondida apertando-se a tecla verde (SIM) ou a tecla vermelha (NÃO) na caixa de botões. 24 sujeitos, alunos de graduação da UFRJ, com visão normal ou corrigida, apresentaram-se como voluntários para o experimento, havendo reportado, posteriormente, ter sido a tarefa relativamente simples. Os quadros (9) e (10), a seguir, ilustram as condições experimentais e de controle e exemplificam um conjunto de frases:

	DE CONTROLE		EXPERIMENTAIS	
	VERBO (V)	PARTICÍPIO (P)	MÍNIMO (M)	NÃO-MÍNIMO (N)
Humanos (H)	HV	HP	HM	HN
Inanimados (I)	IV	IP	IM	IN

Tabela 9 – Condições experimentais e de controle

Exemplo:

	Paga
HV	A empresária recompensa com prêmio em reais mas exige confiança.
HP	A empresária recompensada com prêmio em reais exige confiança.
HM	A empresária paga com antecedência de um mês mas exige confiança.
HN	A empresária paga com antecedência de um mês exige confiança.
IV	A empresa recompensa com prêmio em reais mas exige confiança.
IP	A empresa recompensada com prêmio em reais exige confiança.
IM	A empresa paga com antecedência de um mês mas exige confiança.
IN	A empresa paga com antecedência de um mês exige confiança.

Tabela 10 – Exemplos de cada condição

2.2.2.2. A lógica do experimento

O experimento previa que o segmento crítico (segmento 4) seria lido mais rapidamente nas frases mínimas do que nas frases não mínimas, indicando a preferência pela concatenação da forma verbal ambígua como verbo principal e não como adjetiva reduzida, conforme previsto pelo Princípio da Aposição Mínima. Os tempos de leitura do segmento 4 deveriam ser mais altos nas frases não-mínimas, pois refletiriam o efeito surpresa do *garden-path* e a reanálise subsequente. O experimento pretendia também verificar se haveria diferenças nos tempos de leitura do segmento 4 decorrentes da manipulação dos traços semânticos [+humano] e [-animado] do SN inicial. Caso o processador fizesse uso rápido desse tipo de informação, a tempo de influenciar a decisão do parser na construção da estrutura, os tempos de leitura das condições IM e IN não deveriam apresentar a mesma distribuição dos tempos de leitura das condições HM e HN. Os controles visavam a comparar os tempos de leitura do segmento crítico nas frases com verbos ambíguos com os tempos de leitura do segmento crítico nas frases com verbos não ambíguos, por exemplo, HV x HM ou IV x IM.

2.2.2.3. Resultados

Foi detectada claramente uma preferência geral pela interpretação como verbo principal, comprovando o Princípio da Aposição Mínima. A manipulação dos traços semânticos de animacidade e humanidade não influenciou os tempos de leitura médios do segmento crítico. Observe-se que os tempos abaixo registrados na Tabela 11 e no Gráfico 3 referem-se aos tempos médios de leitura dos segmentos 4, em cada condição experimental. Assim, por exemplo, na frase ilustrada acima, o segmento “mas exige confiança” foi lido mais rapidamente do que o segmento “exige confiança” tanto nas condições em que o SN era [+humano] (HM x HN), quanto nas condições em que o SN era [-animado] (IM x IN). Infere-se que os tempos de leitura significativamente mais elevados dos segmentos críticos das frases não-mínimas decorreram da decisão inicialmente tomada na leitura do segundo segmento de integrá-lo à estrutura em construção como verbo principal e não como partícipio. Havendo tomado esta decisão mínima, conforme previsto pelo Princípio da Aposição Mínima, o parser não terá como integrar o segmento crítico à estrutura precedente, sem reanalisá-la em termos não-mínimos, isto é, adotando a análise adjetival. Embora a metodologia empregada no estudo *on line*, a leitura auto-monitorada, não permita o estudo direto dos padrões regressivos, infere-se, como é prática corrente na literatura, que os tempos médios de auto-monitoração mais elevados do segmento crítico, sejam indicativos do efeito surpresa do *garden-path*, quando o leitor se apercebe que o comprometimento inicial deve ser revisto a fim de “salvar” a frase.

Note-se, por outro lado, que não há diferenças significativas quando se cruzam os tempos médios de leitura dos segmentos críticos das frases mínimas com sujeito [+humano] (HM), com os tempos médios de leitura dos segmentos críticos das frases mínimas com sujeito [-animado] (IM); ou quando se cruzam os tempos médios de leitura dos segmentos críticos das frases não-mínimas com sujeito [+humano] (HN) com os tempos médios de leitura dos segmentos críticos das

frases não mínimas com sujeito [-animado] (IN), indicando não ter havido efeitos decorrentes dos traços semânticos manipulados.

Os resultados dos tempos médios de leitura dos segmentos críticos das frases experimentais estão representados no Gráfico 3 e resumidos no quadro 11.

CONDIÇÃO	MÍNIMO	1º QUARTIL	MEDIANA	MÉDIA	3º QUARTIL	MÁXIMO	DESVIO PADRÃO
Geral	292	1060	1460	1670	1990	7380	903.81
HM	593	1020	1080	1230	1440	2490	865.6
HN	881	1710	1890	2240	2710	4680	692.67
HP	292	1190	1570	1610	1940	3550	425.03
HV	530	1050	1460	1650	2030	3960	819.86
IM	595	1000	1080	1240	1240	3780	785.67
IV	644	1150	1390	1640	1860	3930	732.93
IN	892	1650	1910	2350	2500	7380	560.6
IP	523	935	1220	1510	1860	3300	1365.2

Tabela 11 – Tempos médios de leitura do segmento crítico

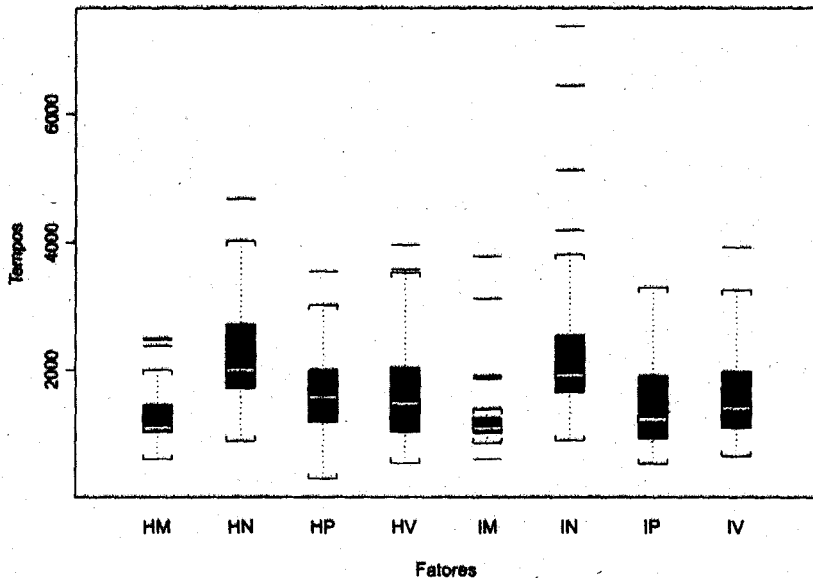


Gráfico 3 – Tempos médios de leitura do segmento crítico e cruzamentos das condições

O Gráfico 3 sugere que algumas condições têm tempos médios de leitura maiores. Uma ANOVA com um fator (cujos níveis eram os tipos de frases) mostrou um resultado significativo ($F(7,294) = 10,16$, $p \cong 0$). Isto significa que há diferenças nos tempos de leituras para cada tipo de frase. Uma ANOVA 2X2 indica que o fator H (humano) ou I (inanimado) não é significativo ($F(1,297) = 0,124$, $p = 0,73$) enquanto que o fator M, N, P, ou V é significativo ($F(3,297) = 23,63$, $p = 0,00$). Dando prosseguimento à análise, fizemos comparações “pairwise” com testes t bilaterais. A tabela (12), abaixo, contém essas comparações.

CRUZAMENTO	TESTE
HM x HN	$t(49,84) = -6,73$; $p = 0,00$
IM x IN	$t(51,53) = -4,78$; $p = 0,00$
HV x HM	$t(48,37) = 2,70$; $p = 0,01$
IV x IM	$t(80,30) = 2,48$; $p = 0,016$
HP x HN	$t(67,70) = -3,49$; $p = 0,00$
IP x IN	$t(61,57) = -3,39$; $p < 0,00$
HM x IM	$t(74,34) = -0,13$; $p = 0,90$
HN x IN	$t(64,91) = -0,44$; $p = 0,66$

Tabela 12 – Comparações *pairwise*

Observe-se que as comparações que contrastam H (humano) vs I (inanimado) não produzem resultados significativos, enquanto que, por exemplo, as comparações que contrastam M (mínimo) com N (não-mínimo) apresentam diferenças significativas.

2.3. Concatenação de SP a SV ou a SN

Esse estudo teve por objetivo investigar a ambigüidade de aposição sintática do Sintagma Preposicional (SP) ao Sintagma Verbal (SV) ou ao Sintagma Nominal (SN) no português brasileiro. Assim, como nos dois outros estudos reportados nas seções 2.1. e 2.2., foram feitos um estudo de questionário e um experimento on line em que se monitoraram tempos de leitura. A exemplo dos experimentos acima,

além de testar a Teoria do *Garden-Path*, este também procurou manipular efeitos semânticos a fim de examinar a possível influência de fatores não estruturais no parsing. Trata-se de determinar se o parser é estritamente sintático em sua análise inicial ou se já acessa rapidamente informações de natureza semântica e pragmática a tempo de influenciar a análise sintática, assim como examinado em diversos estudos sobre estruturas equivalentes na língua inglesa (cf. Sedivy & Spivey-Knowlton, 1994). O princípio específico testado nos estudos sobre a compreensão de estruturas com sintagmas preposicionais com dupla possibilidade de aposição sintática é também o Princípio da Aposição Mínima, que prediz a preferência da aposição ao SV, evitando a criação de um SN complexo.

2.3.1. Estudo de questionário

2.3.1.1. Procedimentos e Materiais

Foram preparados dois questionários contendo cada um 20 frases experimentais e 40 frases distratoras, a serem completadas por 40 voluntários, alunos de graduação da UFRJ. Em ambos os questionários havia uma frase inicial que pretendia funcionar como um contexto para a frase seguinte, onde ocorria a ambigüidade de aposição do SP. No primeiro grupo, havia um contexto que afirmava a existência de um único referente, enquanto que, no segundo grupo, o contexto introduzia mais de um referente possíveis, com vistas a favorecer a aposição baixa do SP, como exemplificado em (12) e (13), a seguir:

(12) Contexto – Plausível: Havia um turista no parque. O policial viu o turista com o binóculo.

(13) Contexto + Plausível: Havia dois turistas no parque. O policial viu o turista com o binóculo.

O contexto –Plausível menciona um referente único, enquanto que o contexto + Plausível menciona dois referentes, pretendendo aumentar

a plausibilidade de se restringir pela aposição do SP ao SN, o universo dos referentes possíveis sobre o qual se faz a afirmação contida na frase. As frases foram apresentadas em blocos de papel, estando no verso da página contendo as frases experimentais uma pergunta focalizando a ambigüidade em teste. Nas frases distratoras a pergunta focalizava outros itens das frases. Por exemplo, para as frases acima, a pergunta era: “Quem estava com o binóculo?”. A disposição dos itens nos blocos de papel visava também a evitar que os informantes consultassem as respostas dadas a itens anteriores nas respostas as perguntas, com vistas a evitar resultados decorrentes de artefato experimental.

2.3.1. 2. Resultados

Os resultados do questionário são apresentados na tabela (13) abaixo, em que se computam as respostas dadas às perguntas, quantificando-se as preferências interpretativas pela aposição do SP ao SV (aposição alta) ou ao SN (aposição baixa), cruzando-as com o contexto +/- plausível. Ao lado de cada indicador, registram-se, também, os valores percentuais:

PLAUSIB./LIGAÇÃO	SP DE SV	SP DE SN	TOTAL
- Plausível	259 /64,9%	141 / 35,1%	400 / 100%
+ Plausível	219 / 54,9%	181 / 45,1%	400 / 100%

Tabela 13 – Resultados e percentuais por condição experimental

Um teste sobre a proporção geral (40%) mostrou que esta é significativamente menor do que 50% ($N = -5.62, p \cong 0$). Mesmo no contexto mais plausível, não há preferência maior do que 50% pelo SN ($N = -1.91, p < 0,05$). De toda forma, o contexto teve forte influência sobre as respostas: no contexto mais plausível houve um aumento significativo da proporção de SN ($t(399) = 2.90; p < 0,01$).

Observe-se, portanto, que, em ambas as condições de plausibilidade, registra-se uma preferência significativa pela aposição alta, isto é, a

interpretação que analisa o SP como ligado ao SV, indicando que o princípio da Aposição Mínima é operativo também neste tipo de estrutura em português. Note-se, no entanto, que na condição mais plausível há diferenças menos acentuadas em favor da aposição alta, computando-se número significativamente maior de aposições baixas, indicando uma interação entre a manipulação do número de referentes no contexto anterior e a escolha da aposição baixa do SP, especificando o SN. De fato, o teste estatístico (teste-t), confirma esses efeitos: As proporções de +P e -P são, respectivamente, 0.6475 e -P é 0.5475 sendo significativamente diferentes ao nível de 1%. O valor de t (estatística do teste) = 2.895267358; v (graus de liberdade) = 399 e o p-valor = 0.003996594.

Embora haja se registrado uma interação significativa entre a informação sobre o número de referentes no contexto discursivo e a decisão de aposição sintática do sintagma preposicional, esta diferença na interpretação do SP poderia estar refletindo uma decisão de reanálise do processador, já que não se monitorou o processamento *on line* da frase. Em outras palavras, o processador poderia haver se comprometido rapidamente com a aposição mínima, apondo o SP ao SV, evitando, assim o SN complexo. Em um segundo passe, no entanto, antes de escrever a sua resposta, o leitor reanalisaria a decisão inicial, considerando, então, o fator pragmático, ou seja, a existência de mais de um referente possível no discurso tornaria mais plausível restringir-se o referente do SN, especificando-o pela aposição do SP. Com o objetivo de se verificar se este efeito registrado no estudo de questionário poderia também ser identificado já na fase inicial do *parsing* sintático, realizou-se o experimento *on line* reportado a seguir.

2.3.2. Estudo *on line*

O estudo teve por objetivo procurar determinar, através de um protocolo de leitura auto-monitorada, se há uma preferência rápida na leitura pela aposição alta do SP ambíguo, confirmando o Princípio da Aposi-

ção Mínima, e se condições pressuposicionais relativas à plausibilidade pragmática são acessadas na fase inicial do processo de *parsing*.

2.3.2.1. Procedimentos e Materiais

O procedimento foi semelhante ao utilizado nos dois estudos *on line* reportados acima. Empregando-se o programa Psyscope, projetou-se um experimento de leitura auto-monitorada, em que 24 sujeitos foram solicitados a ler, da forma mais natural e rápida possível, duas orações segmentadas, conforme indicado pelas barras oblíquas, nos exemplos abaixo. Cada sujeito foi exposto a 16 frases experimentais, apresentadas randomicamente com 32 distratoras, totalizando 48 frases. As frases experimentais distribuíam-se em quatro condições, a saber, +PB (+ plausível baixa), +PA (+plausível alta), -PB (-plausível baixa) e -PA (-plausível alta). As frases experimentais foram comparadas “entre-sujeitos”, evitando-se que um mesmo sujeito fosse exposto a mais de uma versão da mesma frase. Um conjunto experimental completo é exemplificado a seguir:

(12) conjunto experimental

+PB Havia dois turistas no parque./ O policial / viu o turista /com a ferida aberta.

+PA Havia dois turistas no parque./ O policial / viu o turista /com o binóculo preto.

-PB Havia um turista no parque./ O policial / viu o turista /com a ferida aberta.

-PA Havia um turista no parque./ O policial / viu o turista/com o binóculo preto.

Após a leitura de cada frase, cujos segmentos eram apresentados de maneira não-cumulativa, pressionando-se a tecla amarela na caixa de botões acoplada ao computador Macintosh, seguia-se uma pergunta

interpretativa sobre a frase, que devia ser respondida pressionando-se a tecla verde (sim) ou a tecla vermelha (não) na caixa de botões.

2.3.2.2. A lógica do experimento

O experimento pretendeu testar duas hipóteses: (a) o Princípio da Aposição Mínima, que prediz a preferência imediata pela aposição alta do SP; (b) a interferência de fatores não estruturais, tal como a pressuposição pragmática, no processamento *on line*. A hipótese (a) seria confirmada caso a leitura do SP, apresentado no segmento 4 de cada frase experimental, fosse mais lenta nas condições em que se forçava uma leitura baixa (+PB e -PB) do que nas condições onde a aposição alta (+PA e -PA) era possível. Tal deveria ocorrer, pois, sendo operativo em português, o Princípio da Aposição Mínima determinaria que a aposição do SP fosse feito inicialmente, como *default*, ao SV, computando-se o SN simples. Comprometendo-se inicialmente com essa análise, o leitor entraria em um *garden-path* nas frases em que se forçava a leitura baixa, precisando, então rever a sua análise inicial. Tanto o efeito surpresa ao encontrar-se no *garden-path*, quanto a revisão da decisão de aposição sintática inicial requereriam mais operações mentais do que as que seriam necessárias para processar as frases em que a aposição alta do SP era admitida, determinando o maior tempo de leitura dos SP's nas condições baixas. A hipótese (b) seria confirmada caso se detectasse diferenças significativas nos tempos de leitura dos SP's entre as condições +P e -P. Por exemplo, os SP's na condição -PB deveriam registrar tempos médios de leitura mais elevados do que os SP's na condição +PB, se a pista sobre a pressuposição referencial fosse levada em conta rapidamente, assim como parece ter ocorrido na tarefa *offline*.

2.3.2.3. Resultados

Como se pode comprovar, observando as tabelas e o gráfico abaixo, o experimento demonstrou que o Princípio da Aposição Mínima é

operativo também nesse tipo de estrutura ambígua em português. O experimento demonstrou ainda não haver interação significativa entre a informação relativa à pressuposição pragmática e a aposição sintática preferencial do SP. Na tabela (14) aparecem os tempos médios de leitura dos SP's em cada uma das condições experimentais, também representados no Gráfico 3. Abaixo, na tabela Y, estão os resultados da Análise de Variância (ANOVA) que cruzou as condições experimentais relevantes, testando se as diferenças entre as médias revelariam interações significativas. Note-se que as condições em que se forçou a aposição baixa apresentam os tempos de leitura mais altos, conforme predito. Os cruzamentos $-PA \times -PB$ apresentaram diferenças significativas, assim como os cruzamentos $+PA \times +PB$. Também o cruzamento de $PA (+PA \text{ e } -PA)$ e $PB (+PB \text{ e } -PB)$ revelou-se significativo, indicando um efeito principal da interação entre a aposição alta ou baixa do SP e os tempos de leitura, indicativo do efeito *garden-path* e da reanálise, conforme predito. Já os cruzamentos $-P \times P$ não são significativos, nem com relação ao cruzamento de $-PA \times +PA$ (p -valor = 0,8505), nem com relação ao cruzamento de $-PB \times +PB$ (p -valor = 0,1244). Esses resultados indicam que o efeito de pressuposição pragmática obtido no experimento offline não ocorreu no parsing sintático, mas na fase interpretativa da compreensão, sendo o parsing impermeável a esse tipo de informação não-estrutural.

Os resultados dos tempos médios de leitura dos segmentos críticos das frases experimentais estão resumidos no quadro 14 e representados no Gráfico 4.

CONDIÇÃO	MÍNIMO	1º QUARTIL	MEDIANA	MÉDIA	3º QUARTIL	MÁXIMO	DESVIO PADRÃO
Geral	508	936	1170	1350	1660	4470	592.7
+PA	512	886	1080	1160	1330	3180	429.47
+PB	530	989	1430	1480	1840	4390	638.52
-PA	508	858	1060	1140	1280	2890	410.69
-PB	554	1100	1640	1640	1930	4470	698.25

Tabela 14 – Tempos médios de leitura dos segmentos críticos

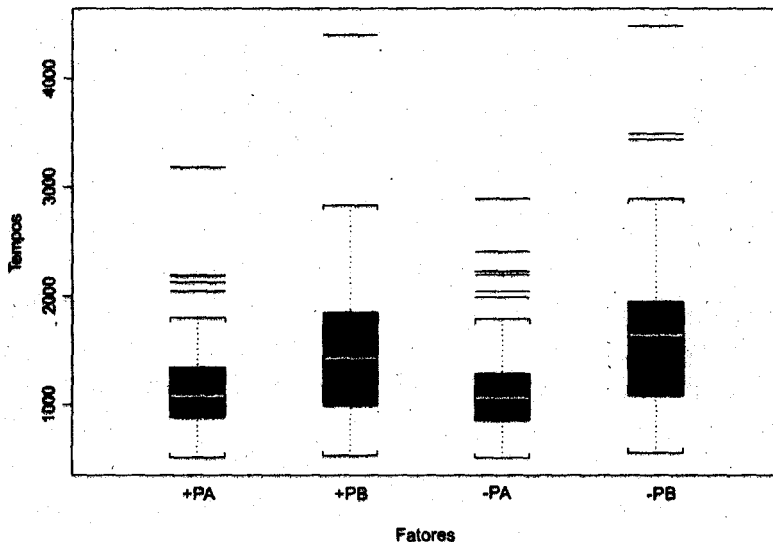


Gráfico 4 – Tempos de leitura do segmento crítico para cada condição

O Gráfico (4) sugere que as frases com B têm tempos médios de leitura maiores. Uma ANOVA com um fator (cujos níveis eram os tipos de frases) mostrou um resultado significativo ($F(3,355) = 18,03, p \cong 0$). Isto significa que há diferenças nos tempos de leituras para cada tipo de frase.

Como as frases com B (aposição baixa) apresentam um tempo de leitura maior do que as frases com A (aposição alta), examinou-se o tempo de leitura considerando o fator com os níveis A e B utilizando uma 2X2 ANOVA. Os resultados da análise com A versus B e plausibilidade como fatores “within-subject” mostraram um “main” efeito de A versus B de ($F(1,356) = 49,62, p \cong 0$) com as frases A (1148,6ms) apresentando um tempo de leitura menor do que as frases B (1563,7ms). O “main” effect de plausibilidade não apresentou um resultado significativo ($F(1,356) = 1,61, p = 0,21$). Dando prosseguimento à análise, fizemos comparações “pairwise” com testes t bilaterais. A tabela abaixo contém essas comparações, confirmando a preferência pela aposição mínima dos sintagmas preposicionais no

processamento de frases em português, independentemente do fator plausibilidade.

CRUZAMENTO	TESTE
+PA X +PB	t(137.28) = -3.7538; p = 0,00
+PA X -PB	t(149.28) = -5.5935; p = 0,00
-PA X +PB	t(132.37) = -4.122; p = 0,00
-PA X -PB	t(144.35) = -5.96; p = 0,00

Tabela 15 – Comparações *pairwise*

III – Conclusões

Reportamos neste artigo experimentos *offline* e *on line* investigando o processamento de três tipos de concatenações sintáticas ambíguas em português: orações iniciadas por QUE, ambíguas entre uma aposição como complemento ou como adjunto; formas verbais ambíguas entre a flexão de terceira pessoa do presente do indicativo ou o particípio passado; sintagmas preposicionais ambíguos entre uma aposição ao sintagma verbal ou ao sintagma nominal. Conforme resenhamos na introdução, o primeiro tipo instancia uma escolha de concatenação entre uma relação sintática primária e uma relação sintática secundária (cf. Clifton & Frazier, 1996), recentemente denominadas em Chomsky (2001), respectivamente, como concatenação de conjunto (*set merge*) e concatenação de par (*pair merge*). Os dados experimentais sobre o português permitiram estabelecer que o primeiro tipo de concatenação é o *default* no processamento: as relações do tipo primário são mais prontamente estabelecidas no parsing do que as relações secundárias que, inclusive, na análise de Chomsky (2001), requerem uma operação a mais do que as primeiras, a simplificação. Nesse sentido, o artigo pretendeu argumentar em favor da realidade psicológica dos princípios gramaticais, sugerindo que o desempenho do parser não é apenas determinado por fatores extra-lingüísticos, como os limites de memória de trabalho, mas que seu funcionamento é, fundamental-

mente, decorrente dos princípios que norteiam a gramática. Assim, procuramos aduzir evidências experimentais em favor da tese de que a faculdade da linguagem é também ótima no sentido de que os mesmos princípios que regem a competência gramatical também podem contribuir para uma teoria do processamento.

O segundo tipo de ambigüidade de concatenação sintática testado por nós também contrasta uma relação sintática primária (SN sujeito concatenado a SV) com uma secundária (SN concatenado a adjunto). Embora não analisada em Chomsky (2001) como constituindo exemplo de concatenação de conjunto, operação restrita naquele estudo à concatenação núcleo-complemento, a concatenação entre o SN sujeito e o composto verbal exibe comportamento paralelo à concatenação de conjunto, comparativamente à opção de concatenação como adjunto, não preferencial.

O terceiro tipo de ambigüidade de concatenação sintática examinado experimentalmente no artigo focaliza o processamento de dois tipos de concatenação de adjunto, residindo a ambigüidade na escolha de adjunção ao verbo ou ao nome. Nesse caso, capturamos a preferência do processador pela adjunção adverbial, evitando a formação de sintagma nominal complexo.

Os três estudos permitiram demonstrar a existência de efeitos *garden-path* em português, provocados pela atuação do Princípio da Aposição Mínima (Frazier e Fodor, 1978; Frazier, 1979), que orienta o parser a decidir pela computação menos complexa sintaticamente. Os três estudos apresentaram também uma simetria interessante no que diz respeito aos resultados dos questionários *offline*, comparativamente aos estudos de leitura auto-monitorada *online*. De modo geral, os estudos *offline* apresentaram maior sensibilidade a efeitos semânticos e pragmáticos, enquanto que os resultados dos estudos *online* parecem sugerir que estes fatores estariam menos disponíveis ao parser do que as informações estritamente estruturais. A natureza das informações acessadas pelo parser no processamento *on line*, bem como o seu curso temporal, são, no entanto, questões

ainda extremamente controversas na literatura, havendo-se mesmo registrado resultados conflitantes decorrentes das técnicas experimentais e metodologias usadas (cf. Clifton & Ferreira, 1989; Steedman & Altmann, 1989). Nesse sentido, acreditamos que a realização de futuros estudos que pretendemos desenvolver sobre a compreensão de frases ambíguas em língua portuguesa, utilizando equipamento de monitoramento da fixação ocular (*eye-tracker*), além de desenho específico, com maior controle dos estímulos, poderá vir a contribuir para determinar de modo mais preciso do que foi sugerido no presente estudo se, de fato, na fase inicial da análise sintática, ocorre o encapsulamento do parser, que funcionaria de modo reflexo, impermeável a informações não estritamente estruturais.

Referências Bibliográficas

- CHOMSKY, N. (1965). *Aspects of a Theory of Syntax*. Cambridge, Mass: MIT Press.
- _____. (1981). *Lectures on government and binding*. Dordrecht: Foris.
- _____. (1986). *Knowledge of language, its nature, origin, and use*. New York: Praeger.
- Chomsky, N. (1991). Some Notes on Economy of Derivation and Representation, in R. Freidin (org) *Principles and Parameters in Comparative Grammar*, Cambridge, MIT Press.
- _____. (1992). *A Minimalist Program for Linguistic Theory*, MIT Occasional Papers in Linguistics, MIT, Cambridge.
- _____. (1995). *The Minimalist Program*. Cambridge, MIT Press.
- _____, N (1997). *A Linguística como uma Ciência Natural*. Entrevista à Revista Mana, vol. 3, nº 2, Ed. Contra-capá, Rio de Janeiro.
- _____. (1998). *Minimalist Inquiries: the Framework*. MIT, manuscript.
- _____. (2000). *Derivation by Phase*. MIT, manuscript.
- _____. (2001). *Beyond Explanatory Adequacy*. MIT, manuscript.

- CHOMSKY, N. & LASNIK, H. (1993). The Theory of Principles and Parameters. In Jacobs, J. et alii (org), *Syntax, an International Handbook of Contemporary Research*. Walter de Gruyter, Berlin, New York.
- CLIFTON, C. & FERREIRA, F. (1989). Ambiguity in Context. *Language and Cognitive Processes*, 4, 77-103.
- DILLINGER, M. (1992). Parsing Sintático. *Boletim da ABRALIN*, 13
- FERREIRA, F. & CLIFTON, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25, 348-368.
- FRAZIER, L. (1979). On comprehending sentences: Syntactic parsing strategies. Doctoral dissertation, University of Connecticut. Distributed by Indiana Linguistics Club.
- FRAZIER, L. & C. CLIFTON Jr. (1996). *Construal*. Cambridge: The MIT Press.
- FRAZIER, L. & J.D. FODOR. (1978). The sausage machine: A new two stage parsing model. *Cognition*, 6, 291-325.
- FRAZIER, L. & K. RAYNER. (1982). Making and correcting errors during sentence comprehension: eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14.
- MACDONALD, M. C., PEARLMUTTER, N.J., & SEIDENBERG, M.S. (1994). The Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676-703.
- MAIA, M. (2000). "A Compreensão de Relações Espaciais em Karajá". *Revista PaLavra*, vol. 6, p. 154-165, Ed. Trarepa, Petrópolis, ISBN 1413-7763.
- _____. (2001). "Gramática e Parser". *Boletim da Abralín*, vol. I, número 26, p. 288-291.
- SEDIVY, J. & SPIVEY-KNOWLTON, M. (1994). The use of structural, lexical and pragmatic information in parsing attachment ambiguities. Hillsdale: Lawrence-Erlbaum.
- STEEDMAN, M. & ALTMANN, G. (1989). Ambiguity in Context:

A Reply. *Language And Cognitive Processes*, 4, 105-122.

WEINBERG, A. (1999) A minimalist theory of human sentence processing. In *Working Minimalism* (eds.) Epstein, D. & Hornstein, N., MIT press, p.283-315