

RESOLUÇÃO PRONOMINAL COM ANTECEDENTES SOBRECOMUNS E COMUNS DE DOIS GÊNEROS EM PORTUGUÊS BRASILEIRO COMO LÍNGUA MATERNA, LÍNGUA DE HERANÇA E COMO SEGUNDA LÍNGUA

Raquel Fellet Lawall (UFRJ)

Marcus Maia (UFRJ/CNPq)

Luiz Amaral (UMass)

RESUMO

Este trabalho investiga as diferenças na computação da resolução pronominal por falantes de português brasileiro nativos, de herança e aprendizes avançados, cuja primeira língua é o inglês. Construções em que um pronome masculino ou feminino retoma um antecedente sobrecomum masculino ou feminino são estudadas através de um experimento de leitura automonitorada com pergunta interpretativa final, a fim de observar efeitos da marcação de gênero. As propriedades formais destas construções são analisadas usando a teoria *Head-driven Phrase Structure Grammar*. Os resultados sugerem que os participantes dos três grupos apresentam diferenças significativas em relação ao processamento da informação de gênero nos níveis sintático e semântico, em medidas on-line e off-line.

PALAVRAS-CHAVE: resolução pronominal, nomes sobrecomuns e comuns de dois gêneros, *Head-driven Phrase Structure Grammar*, processamento sentencial bilíngue, aquisição de segunda língua, português

1. INTRODUÇÃO

Os nomes ditos sobrecomuns em português apresentam traço de gênero gramatical fixo e arbitrário, masculino ou feminino, como os nomes comuns, embora, ao contrário desses, possam referir-se a entidades de ambos os sexos, como nos exemplos em (1).

1. a. A criança (ele/ela)
- b. O ídolo (ele/ela)

Por outro lado, os nomes ditos comuns de dois gêneros são os substantivos com formas masculina e feminina intrinsecamente idênticas, mas diferenciáveis pela alternância de um determinante, como em (2).

- 2.a. O visitante
- b. A visitante

Em inglês, tal classificação não é pertinente, uma vez que os nomes não são diferenciáveis arbitrariamente por um gênero gramatical fixo, como no caso dos sobrecomuns em português, assim como não possuem formas intrinsecamente idênticas, cuja diferenciação de gênero se dê através de um determinante, como é o caso dos comuns de dois gêneros em português. Assim, a tradução para o inglês dos exemplos em (1) segue como em (3) e do exemplo (2) segue como em (4):

- 3. a. The child (he/she)
‘A criança (ele/ela)’
 - b. The idol (he/she)
‘O ídolo (ele/ela)’
4. The visitor
‘O/a visitante’

Os nomes “child” e “idol” podem referir-se a entidades de ambos os sexos “he/she”. De forma similar, o nome “visitor” pode referir-se a “he/she”, sem que haja especificação de gênero no determinante “the”, o qual possui apenas o traço [+determinado]¹.

No presente artigo, investiga-se, através de um experimento de leitura automonitorada, se a diferença entre as propriedades de gênero de antecedentes sobrecomuns em português afeta a computação da correferência pronominal em falantes nativos, de herança, e em aprendizes avançados de português cuja L1 é o inglês.

O experimento testa a correferência intrassentencial em condições em que há concordância ou discordância de gênero gramatical masculino e feminino entre o antecedente sobrecomum e o pronome, e entre antecedentes comuns de dois gêneros em condições em que há apenas concordância de gênero gramatical entre o antecedente e o pronome. Entretém-se análise baseada na *Head-driven Phrase Structure Grammar (HPSG)* (Pollard and Sag, 1994; Sag et al., 2003), que será melhor detalhada na próxima seção.

1. Há, porém, em inglês, evidências de alguns estudos com anáfora conceitual que demonstraram haver uma tendência de certos nomes carregarem uma carga prototípica, tal como os de profissão, tendo sua retomada favorecida através de um pronome masculino ou feminino, como exemplificado a seguir

- a. The carpenter (he vs. She)
‘O/a carpinteiro/a (ele vs. Ela)’
- b. The secretary (she vs. he)
‘O/a secretário/a (ela vs. Ele)’

Nos exemplos acima, as condições com nomes retomados por pronomes que favoreciam o gênero prototípico, ou seja carpenter (carpinteiro(a)) retomado por um pronome masculino (he - ele) e secretary (secretário(a)) retomado por um pronome feminino (she - ela), foram lidas mais rápido do que quando havia discordância entre o gênero prototípico e o pronome - carpenter sendo retomado por she, por exemplo (cf. Carreiras et al.1997, Cacciari & Padovani, 2007; Duffy & Keir, 2004; dentre outros). Kenninson & Trofe (2003) também encontraram convergências em seus resultados com os estudos anteriormente citados. Analisando nomes com gênero prototípico, observaram, em um experimento de leitura automonitorada que, em inglês, os tempos de leitura da região imediatamente após o pronome foram significativamente maiores quando esse desfavorecia a leitura prototípica associada com o antecedente, sugerindo que a resolução pronominal é sensível à discordância de gênero.

O artigo se divide da seguinte maneira: na próxima seção, é apresentada a HPSG. Na seção 3, caracterizam-se os grupos de sujeitos testados no estudo, com foco nas diferenças na competência linguística entre falantes de herança e aprendizes de L2, sendo revistos trabalhos anteriores relevantes, que também compararam esses grupos. Na seção 4, são discutidos estudos prévios sobre o processamento da anáfora pronominal, foco do presente artigo. Na seção 5, é apresentado o experimento de leitura automonitorada que foi aplicado aos três grupos de sujeitos. Na seção 6, apresenta-se uma análise dos resultados obtidos, baseada no quadro teórico da HPSG. Finalmente, a seção 7 é dedicada à conclusão e a apontar os possíveis desdobramentos deste trabalho.

2. DESCRIÇÃO FORMAL

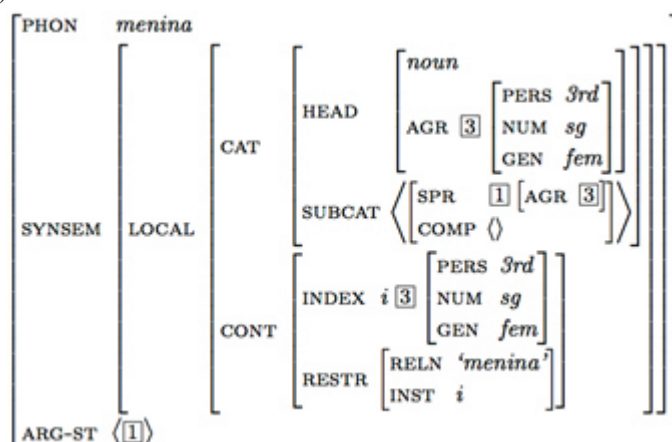
A *Head-driven Phrase Structure Grammar* (HPSG) é um modelo formal de base lexical originalmente proposto por Pollard e Sag (1987) e baseado em modelos sintáticos de base restritiva, como a *Generalized Phrase Structure Grammar* (Gazdar et al, 1985) e a *Lexical Functional Grammar* (Bresnan, 2001). Pollard e Sag (1994) apresentam as bases teóricas da HPSG, assim como a sua aplicação na descrição dos mais variados fenômenos linguísticos².

O mecanismo descritivo básico da HPSG é composto por matrizes de atributo-valor (AVM – do inglês *attribute value matrix*), onde os atributos aparecem em letras maiúsculas e os valores podem ser tanto outras AVMs quanto valores atômicos descritos em letras minúsculas em itálico. Todos os atributos e seus respectivos (possíveis) valores são previstos em uma ontologia hierarquicamente construída, responsável por representar todos os objetos existentes em uma língua.

Neste trabalho usaremos a proposta apresentada por Kathol (1999) para descrever formalmente a diferença entre substantivos sobrecomuns e comuns de dois gêneros. Kathol (1999) propõe um mecanismo de concordância baseado em uma descrição nominal onde as informações sobre gênero e número aparecem tanto na representação sintática, quanto na representação semântica de substantivos. Nós expandimos sua abordagem para explicitar os traços linguísticos que podem influenciar o processamento de enunciados onde exista a correferência entre pronomes e os substantivos descritos acima.

A base da proposta feita por Kathol (1999) pode ser explicada a partir da entrada lexical da palavra “menina” no exemplo (5).

(5) Entrada lexical de “menina”.



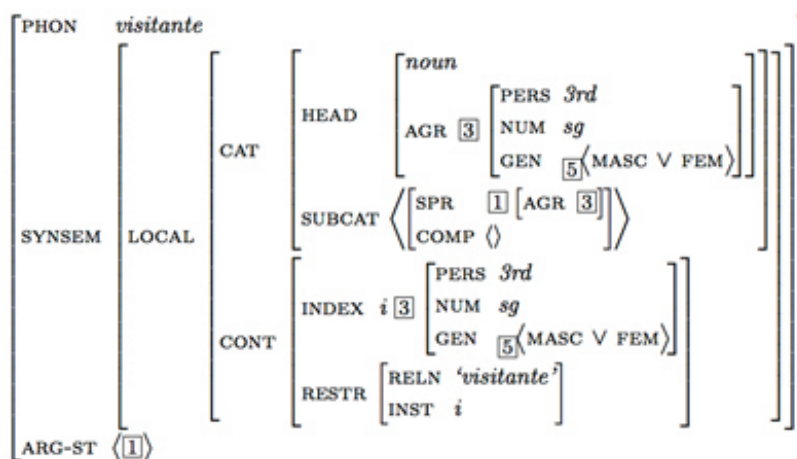
2. Para uma breve introdução à HPSG em português, ver Amaral (2001).

Como vemos no exemplo acima, os três atributos básicos de uma entrada lexical são PHON, SYNSEM e ARG-ST. Nos concentraremos na AVM que serve como valor de SYNSEM | LOCAL. Os atributos CAT e CONT descrevem, respectivamente, as propriedades sintáticas e semânticas da entrada em questão. Repare que o atributo AGR faz parte das características sintáticas do substantivo e que a concordância com o seu especificador é garantida pelo compartilhamento de valores (*structure sharing*) entre AGR e os atributos que descrevem a valência nominal SUBCAT | SPR³. É importante notarmos também que o compartilhamento de valores nesse caso também se dá entre AGR e as propriedades semânticas que descrevem o objeto ao qual o substantivo se refere, e que aparecem como os valores de INDEX. Ou seja, a entrada lexical acima, além de garantir que o substantivo “menina” tenha que concordar sintaticamente com um determinante feminino, também nos indica que esse mesmo substantivo tem que ter como referente um elemento feminino.

2.1. Substantivos sobrecomuns e comuns de dois gêneros:

Baseando-se na descrição dos substantivos vista acima, apresentamos agora as entradas lexicais para os substantivos sobrecomuns e os comuns de dois gêneros.

(6) Comum de dois gêneros:



Como vimos na introdução, os substantivos comuns de dois gêneros aceitam tanto a concordância com determinantes masculinos como femininos. É importante ressaltar que, uma vez que a concordância sintática tenha sido estabelecida, a interpretação semântica forçosamente segue na mesma direção. Ou seja, se o artigo for masculino, o referente descrito pelo substantivo também deverá ser masculino, o mesmo acontecendo com artigos femininos. A entrada lexical em (6) ilustra esses pontos através da introdução de uma disjunção a descrições de gênero (GEN) sintático e semântico. O valor do atributo GEN é uma lista com uma disjunção (exclusiva⁴), onde um dos valores deve ser selecionado. Observe que os valores de AGR e INDEX continuam estruturalmente compartilhados⁵, o que garante que

3. Como veremos mais adiante, seguimos aqui a abordagem proposta por Pollard e Sag (1994), onde o substantivo é o núcleo da projeção máxima das construções nominais – diferentemente de outras abordagens que pressupõem que o determinante seja considerado como núcleo da projeção máxima com função nominal.

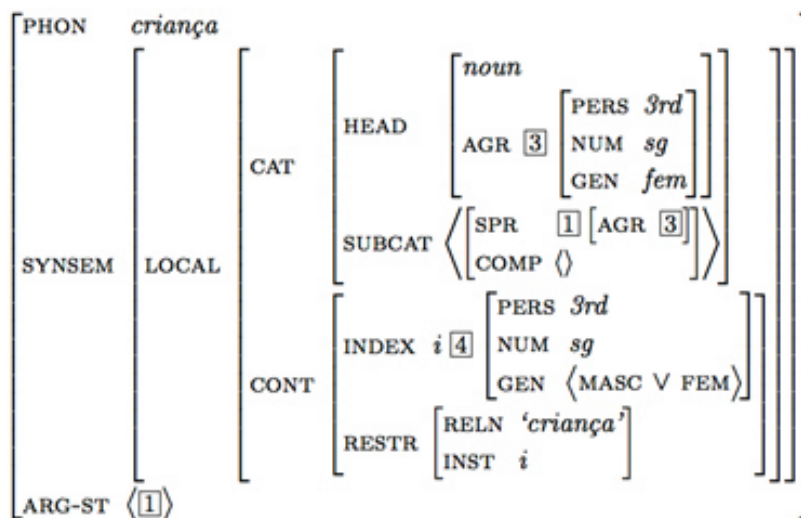
4. Usamos o símbolo v para representar a disjunção exclusiva uma vez que o símbolo \oplus é usado em HPSG para descrever a operação de unificação.

5. Na verdade o termo “estruturalmente compartilhados” (*structurally shared*) representa uma operação onde um único objeto é descrito em dois diferentes lugares. Ou seja, não existe a necessidade de se apresentar duas AVMs com números iguais, como estamos fazendo, já que o objeto descrito pela etiqueta que traz o número é exatamente o mesmo.

eles sejam iguais mesmo depois que a disjunção em GEN seja resolvida através da unificação com o determinante, como veremos na próxima seção.

No caso dos substantivos sobrecomuns, devemos fazer algumas modificações para acomodar suas características sintático-semânticas.

(7) Sobrecomuns



Como vemos no exemplo (7), o ajuste mais importante se dá na etiqueta responsável pelo compartilhamento estrutural entre os valores de AGR e os de INDEX. Nesse caso, fica evidente que o valor sintático de GEN (em AGR) deve obrigatoriamente ser feminino, forçando uma unificação com um determinante também feminino. Porém, no caso dos sobrecomuns, não existe uma obrigatoriedade de que o elemento referido pelo substantivo tenha o mesmo gênero. Ou seja, a palavra “criança”, que sintaticamente é feminina, pode ser usada tanto para se referir a um menino (INDEX | GEN *masc*) como a uma menina (INDEX | GEN *fem*).

Uma descrição linguística que captura as diferenças entre as descrições sintáticas e semânticas dos núcleos de construções nominais nos permite descrever com maior clareza as estruturas envolvidas no processamento dessas construções, em contextos onde exista uma discrepância entre as informações contidas nesses dois níveis descritivos. Porém, antes de partirmos para a descrição do experimento e a análise dos dados, é importante compreendermos como esses itens lexicais influenciam a projeção máxima de suas respectivas construções.

2.2. Descrição sintática de construções nominais

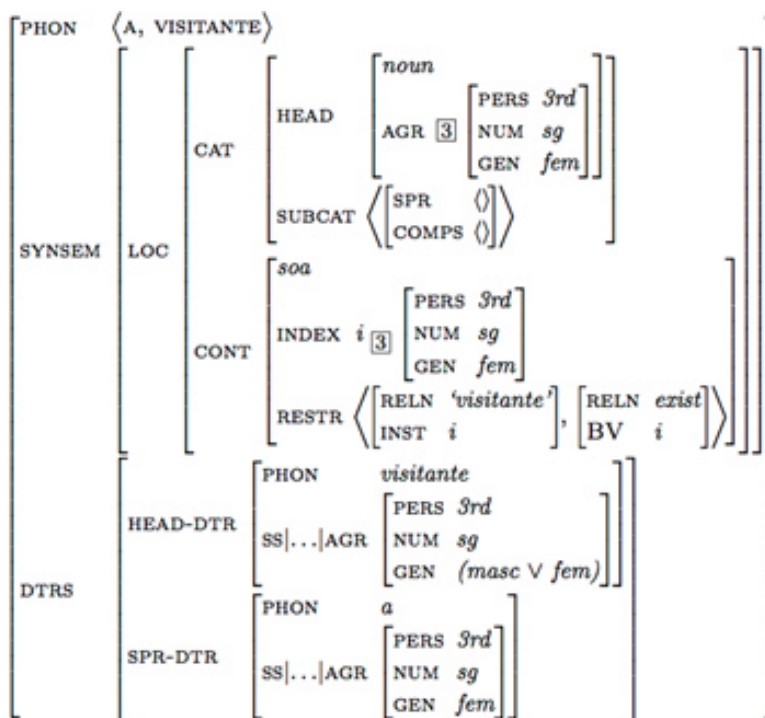
Em HPSG, o *Princípio de Projeção do Núcleo*⁶ garante que as informações apresentadas no atributo HEAD do núcleo de qualquer construção sejam preservadas até a sua projeção máxima. Paralelamente a isso, o *Princípio de Subcategorização* estipula que a projeção máxima de um elemento só se dá quando as listas compostas pelos atributos que compõem SUBCAT estejam vazias; e a *Regra de*

6. Um princípio claramente inspirado pela teoria X-barra.

Núcleo-Especificador força a concordância entre o substantivo e o artigo. Um outro princípio importante, em nosso caso, é o da *Herança Semântica*, que estipula que o valor de INDEX da filha-núcleo (*head-daughter*) será igual ao das suas projeções.

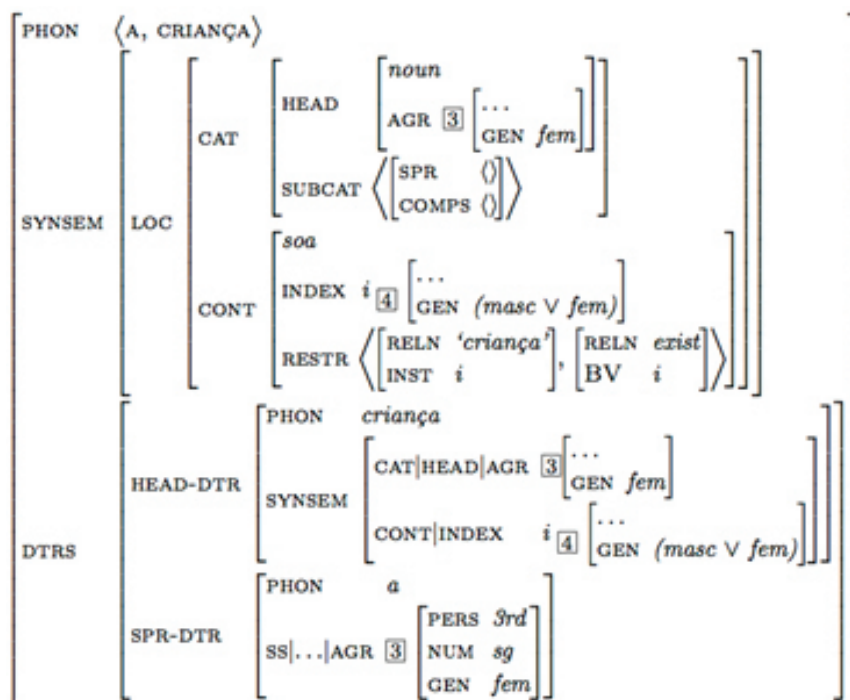
No caso dos substantivos comuns de dois gêneros, como no exemplo (8), vemos que o valor de GEN na filha-núcleo (HEAD-DTR), “visitante”, é uma lista com uma disjunção exclusiva, já o valor de GEN do especificador (SPR-DTR), “a”, está marcado como feminino (*fem*). Assim, quando ocorre a unificação dos valores na formação do sintagma “a estudante”, o valor de GEN é estabelecido como feminino, devido à *Regra de Núcleo-Especificador*. Repare que, nesse caso, por causa do compartilhamento estrutural entre os valores de AGR e INDEX presente na entrada lexical de “visitante” (exemplo 7) e do *Princípio da Herança Semântica*, o valor de INDEX | GEN no sintagma “a menina” também é feminino. Fazendo com que o falante saiba obrigatoriamente o gênero semântico do referente da expressão nominal.

(8) Construção Nominal - *A visitante*



Já no caso de construções nominais com substantivos sobrecomuns a mesma informação sobre o gênero da interpretação semântica da projeção máxima não estará presente.

(9) Construção nominal – *A criança*



Os mesmos princípios responsáveis pela unificação dos elementos lexicais descritos anteriormente para os comuns de dois gêneros se aplicam ao exemplo (9). É importante ressaltar que, assim como ocorre com os comuns de dois gêneros, o valor de AGR | GEN em projeções nominais encabeçadas por substantivos sobrecomuns está completamente resolvido. Nesse caso, não devido ao valor de AGR | GEN do especificador (SPR-DTR), mas devido ao valor de AGR | GEN do núcleo (HEAD-DTR). Contudo, como não existe um compartilhamento estrutural entre as descrições sintática (AGR) e semântica (INDEX), não existe uma maneira de resolver o gênero do referente da expressão a nível local. Ou seja, somente com mais informações contextuais vindas do discurso seria possível se estabelecer se “a criança” se refere a “um menino” (INDEX | GEN *masc*) ou a “uma menina” (INDEX | GEN *fem*). Essa indefinição do valor semântico de gênero vai ser fundamental para explicar as diferenças nos tempos de reação encontrados em nosso experimento, como veremos na seção 5.

3. DEFINIÇÃO DOS GRUPOS DE PARTICIPANTES

O presente artigo explora, através de uma tarefa de leitura automonitorada, a computação da correferência pronominal em três grupos distintos: (i) falantes nativos de português, que atuam como grupo controle, (ii) aprendizes avançados do português como segunda língua, que têm o inglês como L1 e (iii) falantes de português adquirido como língua de herança nos Estados Unidos (EUA). A seguir, caracterizamos mais precisamente esses grupos.

Uma das questões centrais dos estudos linguísticos é a de tentar identificar o que, de fato, sabemos quando sabemos uma língua. A resposta usual a essa pergunta é a de que conhecemos um sistema de sons (ou gestos/sinais) que, organizados de modo sistemático, formam unidades linguísticas de significado, as quais podem ser manipuladas e combinadas para formar unidades mais complexas, tais como sintagmas, sentenças, textos e discursos. Nesse contexto, um falante nativo (que não

possui déficit linguístico e foi exposto a sua língua materna desde criança) compartilha um sistema linguístico com outros membros de sua comunidade, que lhe permite se comunicar, processando os estímulos enviados por seu interlocutor. Esse falante prototípico possui uma pronúncia nativa e um vocabulário compreensível. Ou seja, um falante nativo em condições normais de aquisição apresenta um conhecimento estável e, de certo modo, uniforme de sua língua materna.

Já os falantes de segunda língua podem apresentar certas variações em suas gramáticas que diferenciam sua performance daquela observada em falantes nativos. Alguns autores classificam essas alterações no conhecimento do sistema linguístico da L2 como aquisição incompleta, principalmente em áreas como a fonologia, morfologia flexional, sintaxe e pragmática (Benmamoun, Montrul & Polinsky, 2010). Todos os falantes de segunda língua que participaram de nosso estudo começaram a aprender português após o chamado *período crítico* (ver Birdsong, 1999).

O terceiro grupo considerado em nosso estudo é de falantes de herança. Seguindo a definição dada por Valdes (2000), falantes de herança (*heritage speakers*) são indivíduos criados em lares onde a língua dominante/nacional, no nosso caso o inglês, não é usada ou não é exclusivamente utilizada. Para Polinsky (2011), um falante de herança é um bilíngue que cresceu ouvindo e, possivelmente falando, uma língua minoritária de sua família de imigrantes, em casa, desde a primeira infância. No caso específico do presente estudo, seriam os filhos de pais brasileiros que nasceram e que vivem nos Estados Unidos, tendo aprendido o português em casa como sua primeira língua, mas sendo expostos a toda sua instrução formal na língua majoritária (o inglês). Seguindo Benmamoun, Montrul & Polinsky (2010), falantes de herança são, portanto, falantes bilíngues, de línguas etnicamente minoritárias, que adquiriram a língua alvo cedo, possivelmente antes mesmo da língua majoritária, mas mesmo assim apresentam um conhecimento linguístico que difere à língua falada pelos pais. Ainda seguindo as ideias desenvolvidas por esses autores, como falantes nativos adultos, os falantes de herança são expostos à língua dos pais cedo na infância, dentro do período crítico, em casa e por vários anos, dependendo das condições. Por outro lado, diferenciam-se dos nativos em alguns aspectos importantes. Enquanto os falantes nativos são expostos à sua língua materna dentro e fora de casa, os falantes de herança são confinados à exposição da língua de herança somente no ambiente doméstico, causando uma assimetria de exposição entre a língua minoritária e a majoritária o que pode, posteriormente, se refletir na forma como sua gramática adulta será configurada (cf. Benmamoun, Montrul & Polinsky, 2010). Ou seja, por a língua de herança estar confinada a certos contextos específicos, tipicamente à casa ou a uma comunidade minoritária (quando essa existe), tende a associar-se a um conteúdo mais restrito às relações familiares. Por outro lado, a exposição dos falantes de herança à língua dominante é normalmente mais extensa e contextualmente diversa. Esse *input* mais extenso faz com que a língua majoritária passe a ser dominante em relação língua de herança. Polinsky (2011) aponta que a pesquisa em línguas de herança tem mostrado resultados convincentes de que tanto a produção quanto a compreensão dos falantes de herança é diferente da dos grupos controles de falantes nativos, cuja língua constitui o parâmetro *baseline* para comparação (exemplo de estudos Montrul, 2004, 2006; Polinsky, 2006, 2008). Ainda de acordo com Polinsky, a gramática dos falantes de herança adultos seria marcada por uma aquisição incompleta e por constante *atritão* (*attrition*). Montrul (2008) afirma que a aquisição de uma língua de herança apresenta traços similares e distintos tanto da aquisição de uma L1 quanto da de L2. Dessa forma, nota-se que esse falantes nesse grupo podem apresentar diferenças em relação aos nativos em áreas como a fonologia, conhecimento lexical (vocabulário menos extenso, dificuldade no mapeamento de categorias lexicais), morfologia, sintaxe, marcação de caso e troca de códigos. Para Benmamoun, Montrul & Polinsky (2010), a gramática dos falantes de herança está sujeita a uma aquisição incompleta (devido ao *input*

insuficiente na infância); à atrito ou atrição (*attrition*) (quando um falante de herança demonstra dificuldade em uma propriedade linguística que deveria ser adquirida na infância, por volta de 4-5 anos); e à transferência linguística (quando há uma interferência entre a L1 e a língua primária).

O presente trabalho visa contribuir para um entendimento maior do processamento linguístico e, indiretamente, da competência linguística dos falantes de herança, comparando-os com os aprendizes de L2 e com os falantes nativos.

4. PROCESSAMENTO ANAFÓRICO E MORFOSSINTÁTICO

4.1. Estudos prévios comparando falantes de herança e aprendizes de L2

A maior parte dos estudos, tanto no campo da Linguística Teórica, quanto em Linguística Experimental, focados no desenvolvimento de uma língua, costumam considerar como objeto de estudo falantes, criança ou adulto, nativos dessa língua. Como discutido anteriormente, a competência desse falante é o resultado da aquisição linguística em um ambiente monolíngue, com exposição contínua a essa L1. Entretanto, observando-se dois outros cenários: no primeiro, a aquisição de um sistema linguístico se desenvolve dentro do chamado período crítico, mas sob condições reduzidas de acesso ao *input*; no segundo, em que a língua alvo é aprendida após a puberdade, mediante instrução formal, algumas questões emergem. Como será o processamento linguístico da língua alvo desses dois grupos, já que ambos apresentam uma variação na sua competência linguística com relação ao grupo de falantes nativos?

Esse artigo busca um entendimento acerca dessas questões, comparando as rotinas de processamento de falantes de herança e aprendizes de português como L2, no que tange à computação da correferência pronominal com nomes sobrecomuns. Antes de focar no objeto de estudo específico deste trabalho, vale relatar outros estudos que comparam esses dois grupos, já que alguns deles fornecem *insights* importantes acerca da questão.

Na subárea da morfologia, mais especificamente em um estudo sobre aquisição de gênero gramatical em espanhol, Montrul et al. (2008) entretêm que há uma diferença de processamento entre falantes de herança e aprendizes de espanhol L2, sendo que os primeiros apresentam maior automaticidade em seu processamento gramatical em relação ao segundo grupo. No mapeamento do gênero, falantes de herança com baixa proficiência na língua alvo apresentaram melhor desempenho em uma tarefa de produção, enquanto que aprendizes avançados de espanhol como L2 tiveram melhor resultado em um teste de compreensão, que exigia um conhecimento metalinguístico da L2.

Em outro estudo comparando o mapeamento de gênero em espanhol por falantes de herança e aprendizes de L2, Alarcón (2011) replica os experimentos de Montrul et al. (2008) comparando falantes de herança avançados e aprendizes avançados (L2) e conclui que ambos adquiriram a concordância de gênero em espanhol. Os resultados de seu estudo apontam, na tarefa de compreensão, para uma maior precisão, de ambos os grupos, no estabelecimento da concordância, quando havia um Sintagma Determinante com um substantivo masculino. Na tarefa de produção, como no estudo de Montrul et al. (2008), houve maior precisão nas respostas dos falantes avançados de herança e mais erros dos aprendizes de L2 com as formas femininas. Para a autora, “os resultados favorecem robustamente um *default* da forma masculina, que opera tanto no nível lexical quanto no sintático” (cf. p.342). Os resultados de Alarcón (2011) contradizem uma tendência encontrada em falantes de herança, mas não em falantes nativos e aprendizes de L2, de empregar o sufixo feminino plural como o *default*, em um estudo de aquisição de árabe por Ravid & Farah (1999).

Foote (2011) examina, em seu estudo no campo da morfossintaxe, se falantes de herança e aprendizes de espanhol (L2), que tinham como L1 o inglês, mostram sensibilidade a erros de concordância em contextos com sujeito-verbo e nome-adjetivo, em uma tarefa de leitura, partindo da ideia de que os falantes de herança apresentam maior automaticidade no estabelecimento da concordância.

Dada a escassez de trabalhos na literatura que comparem o português como língua de herança e como L2 no domínio da correferência anafórica de nomes sobrecomuns nessa língua, mostrou-se relevante revisar estudos, como os discutidos acima, já que português e espanhol, compartilham o fato de serem duas línguas que apresentam um sistema dual de gênero masculino e feminino, sendo, em ambas, o feminino a forma mais marcada.

4.2. O processamento linguístico de falantes de herança e aprendizes de L2

A pesquisa em psicolinguística tem se focado, ao longo dos anos, em entender como os falantes nativos, crianças ou adultos, processam a linguagem, em tempo real (processamento *on-line*) ou após a integração de todos os subcomponentes linguísticos (processamento *off-line*), quando compreendem ou produzem palavras, sintagmas, enunciados e textos. Um dos focos principais desses trabalhos tem sido o de buscar entender se a análise sintática ou *parsing* é autônoma nesse processo – pelo menos em um primeiro momento –, ou se processar uma sentença exige a integração simultânea de vários subcomponentes linguísticos (sintaxe, semântica, pragmática, etc). Tais questões têm sido, igualmente, entretidas, na subárea do processamento da correferência anafórica, em que a relação entre, por exemplo, um pronome e um antecedente, têm sido estudada com vistas a estabelecer o chamado “curso temporal” do acesso a diferentes informações no processamento. Outra questão que o presente artigo busca responder é como ocorre o processamento em falantes que aprenderam a língua alvo dentro do período crítico – os falantes de herança – mas receberam o *input* na língua de herança junto com a língua dominante, e como a computação se dá em aprendizes tardios de uma L2, que a adquirem pós período crítico e mediante uma L1 já completamente estabilizada. Mais especificamente, busca-se contribuir para a questão do acesso ao traço gramatical de gênero no estabelecimento da correferência, em falantes nativos, de herança e aprendizes avançados do português, explorando, dessa forma, grupos de sujeitos ainda muito pouco investigados, na literatura de processamento.

Na literatura de processamento de L2, um dos trabalhos seminais que busca compreender o processamento gramatical de aprendizes de L2 em tempo real é o de Clahsen & Felser (2006), que comparam distintos grupos de falantes – nativos adultos, crianças aprendizes de uma língua materna, e aprendizes de uma segunda língua – em diferentes domínios linguísticos (morfologia e sintaxe). O principal foco desse estudo é o de tentar estabelecer como o processamento gramatical de aprendizes de uma L2 difere do de falantes nativos e, até mesmo, do de crianças adquirindo sua L1. Para os autores, crianças e falantes nativos adultos de uma L1 utilizam um mesmo mecanismo de *parsing* (*Continuity Hypothesis*), que não se modifica ao longo do tempo, e as diferenças nas rotinas de processamento desses dois grupos pode ser atribuída a outros fatores, como a capacidade limitada da memória de trabalho das crianças, bem como uma reativação de itens lexicais menos eficiente delas em relação a dos adultos. O processamento não nativo, por outro lado, difere no *parsing* da informação de natureza sintática em relação aos falantes nativos adultos, uma vez que as representações sintáticas dos aprendizes de uma L2 seriam mais superficiais e menos detalhadas do que as dos nativos (*Shallow Structure Hypothesis*), apesar de ambos os grupos serem guiados pelas mesmas pistas léxico-semânticas utilizadas durante o processamento.

Outros modelos que se propõem a analisar o processamento em segunda língua veem a diferença entre aprendizes de uma L2 e falantes nativos através de como a gramática é mentalmente representada e processada (Ullman, 2001, 2006; Paradis 1994, 1997, 2004). Ullman (2001, 2006) advoga em seu Modelo Declarativo-Procedural (*DP Model*, em inglês) que a língua depende de dois sistemas de memória para funcionar. O primeiro deles é o sistema de Memória Declarativa, localizado no lobo temporal, que subscreve a aprendizagem, representação e conhecimento linguístico, sobre fatos e eventos, sendo esse sistema, pelo menos em parte, explícito, i.e., mais consciente. O segundo deles é o sistema de Memória Procedimental, localizado na região do gânglio basal, nos lobos frontais, que subjaz a aprendizagem implícita e não consciente de habilidades e hábitos novos e bem estabelecidos, especialmente os que envolvem sequências. De acordo com Ullman, na aquisição de uma primeira língua, a distinção entre memória procedimental e declarativa seria paralela à noção de gramática e léxico mental. Assim, a memória declarativa subjaz ao léxico, que contém todo o conhecimento linguístico idiossincrático, incluindo os sons e os significados das palavras, bem como palavras complexas. Já a memória procedimental engloba aspectos da gramática mental, ou seja, a computação de estruturas linguísticas complexas (hierárquicas e governadas por regras). Os dois sistemas são previstos para interagirem, tanto cooperativamente, quanto de forma competitiva. Por exemplo, uma criança em processo de aquisição de uma L1 aprenderia palavras complexas e idiossincrasias de sua língua, na memória declarativa, enquanto o sistema procedimental gradualmente adquire o conhecimento gramatical subjacente à combinação governada por regras.

O que ocorreria, então, na aquisição tardia de uma segunda língua e de uma primeira língua que passa a ser minoritária, do ponto de vista do *input*, como no caso dos falantes de herança? Segundo o Modelo de Ullman, em uma L2 a aquisição do conhecimento procedimental/gramatical é mais problemática que a do conhecimento declarativo/lexical, comparada à aquisição infantil. Esse fato pode ser explicado pela atenuação da memória procedimental e pelo crescimento da declarativa, um processo biológico, em que atividades motoras associadas ao primeiro sistema estão sujeitas a efeitos de maturação, havendo um ganho da memória declarativa durante a infância, com seu ápice na adolescência. Assim, dada a facilidade da aprendizagem declarativa em detrimento da procedimental, os aprendizes de uma L2 tendem a se apoiar na primeira, memorizando formas linguísticas complexas como blocos (ex. Andou; o gato), as quais geralmente são computadas composicionalmente por aprendizes de uma L1 (and + ou; o + gato).

Sendo o processamento gramatical em falantes nativos dependente de um conhecimento mais implícito, armazenado na memória procedimental, o resultado tende a ser o de um *parsing* rápido, inconsciente e automático. Já para adultos aprendizes de uma L2, como dito anteriormente, Ullman (2001, 2006) sugere que o sistema de memória procedimental estaria menos ativo para processar uma língua não nativa (talvez devido a efeitos de maturação cerebral), de forma que os aprendizes de uma L2 se apoiariam em recursos da memória declarativa para armazenar o seu conhecimento sobre a língua alvo. Assim, o conhecimento sobre a L2 seria dado como explícito (i.e. consciente), mais do que um sistema de procedimentos computacionais internalizados que se aplicam automaticamente (Paradis, 2004). E no caso dos falantes de herança? O fato de terem aprendido a L1, mesmo que essa aquisição tenha sido interrompida, antes do período crítico, lhes daria o respaldo necessário para adquirir sua L1 utilizando recursos de sua memória procedimental, ou seja, apresentariam um conhecimento implícito de sua L1, mesmo que não mostrem a proficiência de um nativo (devido à interrupção do processo de aquisição)⁷. Nossa hipótese, portanto, é a de que tal conhecimento

7. De fato, segundo aponta parecerista anônimo da Revista Lingüística, segundo Ullman (2001), a prática levaria à proceduralização do conhecimento. Mais determinante do que o sistema que subjaz a aquisição desse conhecimento, o uso determinará o sistema neural que será subjacente a ele, ou seja, sistema de memória declarativo ou procedural. Em outras palavras, não seria a suposta interrupção na aquisição o fator determinante para o desempenho não nativo dos indivíduos, mas sim o uso que ele faz da língua.

lhes permitiria executar tarefas em sua L1 com maior precisão e rapidez, quando comparados aos aprendizes de uma L2. Porém, por terem utilizado o *input* de maneira reduzida na L1 não teriam a mesma automaticidade, precisão e rapidez, quando comparados a falantes nativos de uma dada língua, por isso seriam mais lentos em tarefas psicolinguísticas *on-line*.

O presente estudo busca contribuir para o entendimento da competência gramatical dos dois grupos, tentando perceber se há diferença na computação da correferência pronominal entre falantes de herança e aprendizes de português como L2, usando o grupo de falantes nativos de português brasileiro (PB) como controle. O próximo tópico é dedicado a analisar estudos anteriores relevantes, no campo do processamento anafórico em primeira língua.

4.3. Estudos anteriores com anáfora pronominal

O termo ‘anáfora’ geralmente descreve qualquer expressão que faz referência a um antecedente. Nas décadas recentes, a resolução anafórica se tornou uma questão chave no entendimento dos mecanismos e representações envolvidos na compreensão da linguagem (para uma visão geral, ver Nicol & Swinney, 2002; Garnham, 1999). Vários fatores afetam a facilidade e o curso temporal da computação correferencial, dentre eles a representação conceitual do antecedente e suas características lexicais (Cacciari et al. 1997; Nicol & Swinney, 1989, 2002; van Gompel & Majid, 2004). Há ainda muita discussão na literatura sobre como e quando esses dois níveis da informação intervêm no processamento pronominal. O presente estudo fornece alguma evidência sobre essas questões, buscando contribuir para um entendimento do processamento da correferência pronominal em três grupos distintos: falantes nativos, de herança e aprendizes avançados do português.

Um estudo importante sobre o processamento da concordância de gênero em italiano é o de Cacciari et al. (1997), que avaliaram o processamento de formas pronominais (pronomes plenos, clíticos) e nomes próprios, que tiveram como antecedentes nomes [+animado], [+humano] como ‘vítima’. Os autores encontraram menor tempo de leitura para as condições com pronomes e nomes próprios em que há concordância de gênero entre este e o antecedente.

Outro estudo que elabora achados de Cacciari et al. (1997), utilizando dados de falantes de português brasileiro, foi o de Côrrea (2001), que investigou a concordância de gênero na compreensão de formas pronominais em contexto discursivo, utilizando antecedentes nominais acompanhados de adjuntos, que carregavam ou não informação léxico-semântica de gênero, como “a testemunha de meia-calça/de barba/de roupas claras”. Os dados de Cacciari et al. (1997) não apresentavam a informação léxico-semântica que poderia servir como pista na resolução da concordância de gênero. Os resultados de Côrrea (2001) sugerem que o processamento da concordância de gênero faz uso de informação léxico-semântica na resolução da correferência.

Um estudo diretamente relevante para o que desenvolvemos neste artigo, que traz importantes *insights* acerca do papel do gênero gramatical na resolução pronominal de antecedentes nominais sobrecomuns (epícenos), é o de Cacciari et al. (2011). Esses autores investigaram, em um experimento de leitura automonitorada, o papel do gênero gramatical e do gênero contextual na referência de um antecedente a um pronome, com nomes sobrecomuns e comuns de dois gêneros, em italiano. Os resultados indicaram que, no caso dos sobrecomuns, que identificam o gênero morfológicamente, a

resolução pronominal teria sido guiada, inicialmente, pelo gênero gramatical; no caso dos comuns de dois gêneros, sem indicação morfológica, a resolução seria guiada, ainda que em fase não inicial do processamento, pelo contexto.

5. O ESTUDO DE LEITURA AUTOMONITORADA

5.1. Objetivo

Testar o papel do acesso ao traço gramatical de gênero no estabelecimento da correferência, com antecedentes sobrecomuns e comuns de dois gêneros (controle), em três grupos de falantes do PB: nativos, falantes de herança e aprendizes adultos (L2).

5.2. A tarefa

Um experimento de leitura automonitorada não cumulativa, com questão interpretativa, com sondagem para a correferência, após a leitura de cada frase.

5.3. Hipóteses

Este experimento apresenta as seguintes previsões:

Para o grupo de falantes nativos do português, esperam-se tempos de leitura médios mais rápidos, de modo geral, em todas as condições. Quando há assimetria na concordância de gênero entre o pronome e o nome sobrecomum, por exemplo, ‘o ídolo’ sendo retomado por ‘ela’, esperam-se tempos médios de leitura mais lentos dos segmentos críticos – o pronome e o segmento imediatamente subsequente a ele, o *spillover*.

Para os falantes de herança, esperam-se diferenças nos tempos médios de leitura similares às do grupo de nativos.

Para o grupo de aprendizes adultos, não se esperam as mesmas diferenças nos tempos médios de leitura na resolução pronominal com antecedentes sobrecomuns, previstas para os dois primeiros grupos, uma vez que se hipotetiza que o grupo de L2 não demonstraria sensibilidade semelhante ao traço gramatical de gênero.

5.4. Variáveis

5.4.1. Independentes: gênero dos sobrecomuns (masculino/feminino); gênero do pronome (masculino/feminino); grupos (nativos/aprendizes de L2/falantes de herança).

5.4.2 Dependentes: Tempos médios de leitura dos segmentos críticos e pós-críticos (*on-line*) e índices de decisão (*off-line*).

5.5. Material

O experimento apresenta 12 frases experimentais e 24 distratoras, em um total de 36 sentenças.

5.6. Condições

O experimento apresentou 6 condições: SFF (nome sobrecomum feminino retomado por um pronome feminino); SFM (sobrecomum feminino retomado por um pronome masculino); SMM (sobrecomum masculino retomado por um pronome masculino); SMF (sobrecomum masculino retomado por um pronome feminino); CFF (nome comum de dois gêneros com determinante feminino retomado por um pronome feminino); CMM (nome comum de dois gêneros com determinante masculino retomado por um pronome masculino). Cada sentença era acompanhada de uma pergunta interpretativa, para verificar a aceitabilidade da correferência, como em (11):

- (11) SFF – Como/ a vítima /usava/ cinto de segurança,/ela/sobreviveu /ao acidente.
SFM – Como/a vítima /usava /cinto de segurança,/ele/sobreviveu/ ao acidente.
SMM – Como/ o bebê/ usava/cinto de segurança,/ele/sobreviveu/ ao acidente.
SMF – Como/ o bebê/ usava/ cinto de segurança,/ ela/ sobreviveu/ ao acidente.
CFF – Como/a assistente/usava/cinto de segurança,/ela/sobreviveu/ao acidente.
CMM – Como/ o assistente/ usava/ cinto de segurança,/ ele/sobreviveu/ ao acidente.

Questão: (A vítima, o bebê, o-a assistente) sobreviveu ao acidente?

5.7. Procedimentos

Os participantes foram individualmente testados, usando-se um computador *Apple MacBook*, por meio do programa *Psyscope* (cf. Cohen et alii, 1993), para coletar os tempos de leitura e os índices de aceitabilidade da correferência das condições testadas. Ao pressionar uma tecla, os sujeitos liam partes de uma frase, que apareciam na tela não cumulativamente até o final da frase, assinalado por um ponto, quando se chegava a uma pergunta de interpretação sobre a frase que eles haviam lido anteriormente, de resposta sim ou não. Se aceitassem a correferência, deveriam apertar uma tecla verde e, caso não a estabelecessem, deveriam apertar uma tecla vermelha no computador.

5.8. Participantes

O grupo controle foi composto por 12 falantes nativos de português brasileiro, e o grupo de falantes de herança, 12 estudantes de cursos de português na Universidade de Massachusetts (UMass, Amherst, EUA), de diversos níveis. O terceiro grupo, de aprendizes avançados de PB, cuja língua nativa era o inglês, também foi composto de 12 estudantes de cursos avançados de português na UMass, cujo conhecimento da língua portuguesa é suficiente para se comunicarem com falantes nativos e lerem textos acadêmicos.

5.9. Resultados

Os tempos médios de leitura (ms) do segmento crítico (pronome) e do segmento pós-crítico (verbo), para todos os três grupos, em todas as condições experimentais, estão apresentados na Figura I, a seguir.

Segmentos	CONTROLE		HERANÇA		L2	
	pronome	verbo		verbo		verbo
SFF	505	487	765	684	781	1033
SFM	587	451	663	714	751	1056
SMM	522	480	696	707	669	1075
SMF	596	630	717	930	802	984
CFE	548	490	765	830	786	1142
CMM	543	519	716	747	693	1032

Figura I: Tempos médios de leitura dos segmentos crítico e *spillover*

As comparações relevantes dos tempos médios de leitura do segmento do pronome não apresentam diferenças significativas entre as condições, internamente a nenhum dos grupos, o que não é incomum em estudos de correferência pronominal. Note-se que as condições com os nomes comuns de dois gêneros (CFE e CMM) também não apresentam diferenças importantes nas comparações dentro de cada grupo, seja na região do pronome, seja na região do verbo⁸. Diferenças significativas entre as condições com os sobrecomuns aparecem, no entanto, na região do verbo (*spillover*), conforme reportado a seguir.

Os gráficos e tabelas que se seguem mostram os tempos de leitura do segmento pós-crítico *spillover* (verbo) nas condições com os nomes sobrecomuns (gráfico 1), medida *on-line*, e os índices de aceitabilidade da correferência (gráfico 2), medida *off-line*. Assim, no exemplo (11), repetido abaixo, mediu-se o tempo de leitura de ‘sobreviveu’, no gráfico (1):

(11) Como/ a vítima-o bebê /usava/ cinto de segurança,/ele-ela/sobreviveu /ao acidente.

O gráfico (1) mostra os tempos médios de leitura dos sobrecomuns (masculino/feminino) com um pronome (masculino/feminino) na região do verbo (*spillover*).

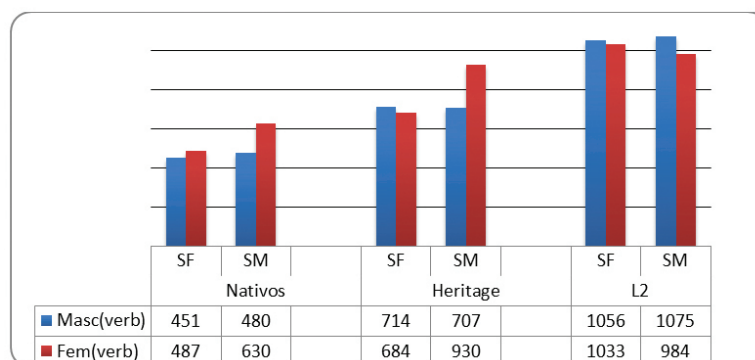


Gráfico 1: Tempos de leitura dos sobrecomuns na região do verbo

8. Ao contrário de estudos como o de Cacciari et alii (2011) para o italiano e o de Faber, Amaral & Maia (2012/a aparecer) para o espanhol, em que a condição com os comuns de dois gêneros pôde ser contrastada de modo interessante com as condições com os sobrecomuns, no presente estudo esta comparação não foi entretida, pois, em português, o gênero desses nomes é necessariamente especificado pelo artigo. Em italiano, o artigo definido pode ser contraído com o nome iniciado por vogal, não sendo obrigatoriamente especificado. Em espanhol, Faber et alii utilizou o pronome possessivo que, ao contrário do que ocorre em português não precisa concordar com o nome que determina, não especificando, portanto, o gênero do sintagma. Os estudos sobre o italiano e o espanhol puderam, assim, contrastar a presença do traço gramatical de gênero, como ocorre nos sobrecomuns, com a sua ausência nos comuns de dois gêneros. Em português, tal contraste seria possível utilizando-se numerais diferentes de um e de dois, que flexionam em gênero, o que será objeto de estudo futuro (e.g. o/a assistente; seu assistente/sua assistente; três assistentes).

O grupo controle apresentou um resultado significativo na comparação entre as duas condições assimétricas (SFM e SMF) como indicado no teste-t: $t(46)=2,85$, $p=0,006^*$.

O grupo de falantes de herança também mostrou significância na área do verbo nas diferenças entre as condições assimétricas: $T(40)= 2,221$, $p=0,032^*$.

Os resultados da comparação das mesmas condições (SFM e SMF) para o grupo de aprendizes de português L2 não foram, no entanto, significativos na medida *on-line*, à diferença do que se obteve para os grupos controle e de herança. O teste-t feito para o grupo L2 não permite, portanto, rejeitar a H_0 para o grupo L2: $T(46)=0,96$, $p=0,33$.

Uma ANOVA bivariada por sujeitos cruzando o fator GRUPO de sujeitos (Nativos, falantes de herança e L2) com os tempos médios de leitura da região do verbo nas condições assimétricas (SFM e SMF) revelou um efeito principal do fator grupo ($F(1, 36)= 16,7$, $p=0,001^{**}$), mas não revelou efeito da assimetria ($F(1,36)=1,18$, $p=0,3ns$) e nem interação entre os dois fatores ($F(1,36)=2,13$, $p=0,13ns$). A assimetria presente no grupo de nativos e no grupo de falantes de herança, bem como os tempos médios de leitura do grupo L2, de modo geral, significativamente mais altos, garante a diferença entre os grupos, mas a inexistência de assimetria SFM x SMF, neste último grupo, oblitera a assimetria presente nos grupos de nativos e de falantes de herança. De fato, quando se realiza a ANOVA considerando no fator GRUPO apenas os nativos e os falantes de herança, ou seja, excluindo-se o grupo L2, observa-se tanto o efeito principal do fator grupo ($F(1,24)=14,6$, $p=0,0004^{***}$), altamente significativo, quanto o efeito principal da assimetria ($F(1,24)=7,5$, $p=0,009^{**}$), sem interação entre os fatores ($F(1,24)=0,61$, $p=0,4 ns$).

O gráfico (2) mostra a aceitabilidade da correferência (respostas sim) para os sobrecomuns (masculino/feminino) com o pronome (masculino/feminino). No exemplo abaixo, há uma sentença experimental seguida de uma pergunta:

(12) Como/ a vítima /usava/ cinto de segurança,/ele-ela/sobreviveu /ao acidente.
Pergunta: A vítima sobreviveu ao acidente?

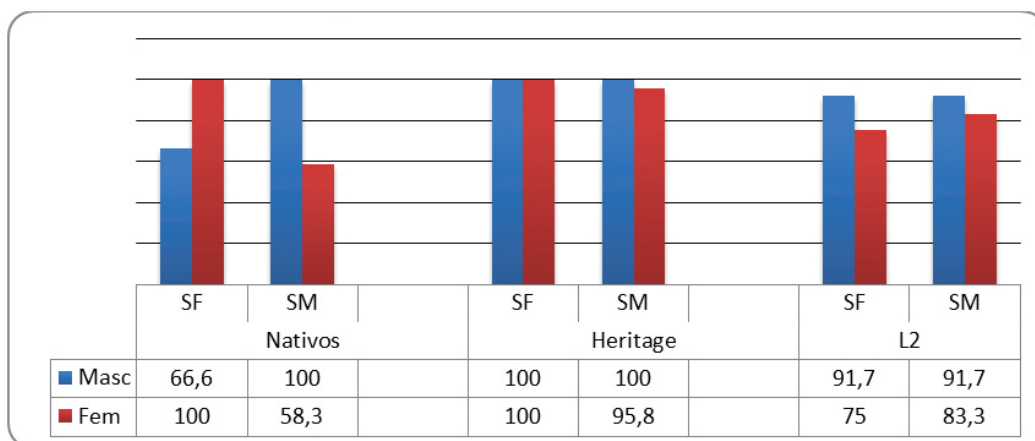


Gráfico 2: Índices de aceitabilidade da correferência para os nomes sobrecomuns

Para o grupo de nativos, cruzam-se as condições a seguir: SMM x SMF ($X^2=22,3$, $p=0,0001^{***}$) e SFF x SFM ($X^2=13,9$, $p=0,0002^{***}$). Ambos os cruzamentos demonstram diferenças significativas entre as condições simétricas e assimétricas nos índices de resposta. Por outro lado, à diferença do que

se obteve nas medidas *on-line*, as assimetrias não revelam diferenças significativas entre si – SMF x SFM ($X^2=1,03$, $p=0.3^{ns}$). Assim, é interessante observar que os tempos médios de leitura mais rápidos obtidos quando o pronome masculino recupera o sobrecomum feminino (efeito *default* do gênero masculino) do que quando o pronome feminino recupera o antecedente sobrecomum masculino, não traz consequências para a fase interpretativa (*off-line*), quando a possibilidade de correferência gramaticalmente legítima em ambos os casos se impõe, indicando recuperação do efeito surpresa *default* capturado *on-line*.

Para os falantes de herança, as comparações demonstram altos índices de aceitabilidade da correferência em todas as condições, não evidenciando o efeito de assimetria obtido para esse grupo na medida *on-line*. Assim, o efeito de estranhamento do pronome feminino recuperando sobrecomum masculino, capturado nos tempos médios de leitura na condição SMF, se esvanece na medida interpretativa, ao contrário do que ocorre com os nativos, o que é interessante. Embora haja um estranhamento reflexo, os falantes de herança diferem significativamente dos nativos na fase reflexiva: SMF nativo x SMF falantes de herança: ($X^2=17,9$, $p=0.0001^{***}$); SFM nativo x SFM falantes de herança: ($X^2=13,9$, $p=0.0002^{***}$)

Já os L2, que não apresentaram diferenças significativas entre as condições na medida *on-line* (ao contrário dos nativos e falantes de herança), continuam sem apresentar diferenças significativas entre as condições na medida *off-line*. Mesmo a maior diferença nos percentuais de resposta não abre significância - SFF x SFM: ($X^2=3,08$, $p=0.08^{ns}$). Entretanto, na comparação com o grupo controle, há diferenças robustas nos percentuais de aceitabilidade da correferência entre o pronome feminino com antecedente sobrecomum masculino: SMF controle x SMF L2 ($X^2=8,8$, $p=0.002^{**}$), confirmando que os L2 não tem sensibilidade para a assimetria também na fase reflexa. O mesmo se dá na outra condição assimétrica: SFM nativos x SFM L2 ($X^2=7,9$, $p=0.004^{**}$).

Assim, comparativamente, pode-se concluir que os nativos tem sensibilidade para a assimetria de gênero na correferência com nomes sobrecomuns, tanto na fase reflexa, quanto na fase reflexiva, embora nessa segunda o efeito *default* do masculino já não se instancie, igualando ambas as condições assimétricas. Os falantes de herança revelam sensibilidade para a assimetria na fase reflexa, mais automática, deixando de apresentá-la na fase da interpretação metalinguística, mais reflexiva. Finalmente, em contraste com os dois grupos anteriores, os L2 não revelam sensibilidade aos traços de gênero nem na fase reflexa e nem na fase reflexiva.

6. ANÁLISE DOS DADOS E REPRESENTAÇÃO GRAMATICAL

Para observarmos as consequências do presente estudo na descrição da representação de projeções nominais nos três grupos de falantes é importante formalizar a restrição imposta à resolução pronominal no discurso.

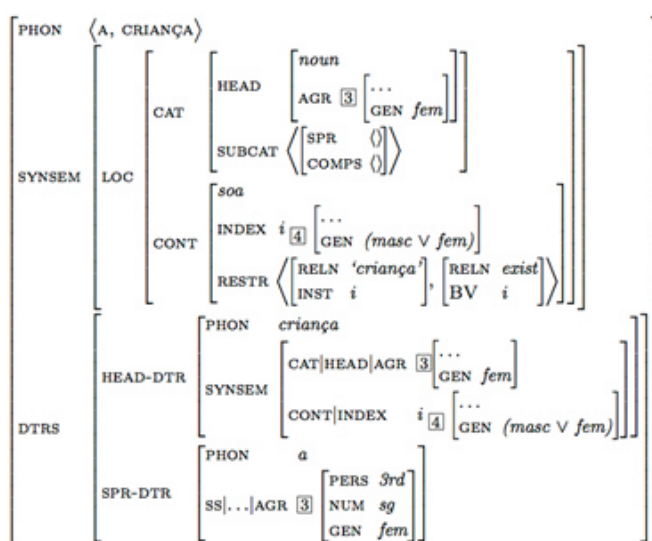
(13) Restrição de resolução pronominal

$$\left\langle \dots \left[\begin{array}{c} \textit{noun} \\ \text{S | L | CONT | INDEX } i \end{array} \begin{array}{c} \text{PERS } 3rd \\ \text{NUM } 1 \\ \text{GEN } 2 \end{array} \right], \dots \dots \left[\begin{array}{c} \textit{pronoun} \\ \text{S | L | CONT | INDEX } i \end{array} \begin{array}{c} \text{PERS } 3rd \\ \text{NUM } 1 \\ \text{GEN } 2 \end{array} \right], \dots \right\rangle$$

Como princípio geral, podemos dizer que a resolução pronominal só é possível se as descrições semânticas do pronome e de seu referente forem compatíveis. Ou seja, como ilustrado no exemplo (13), o valor semântico de NUM e GEN do pronome e da construção nominal deve ser o mesmo⁹. A descrição sintática (AGR), em princípio, não deve interferir na resolução pronominal, pois, como vimos na sessão 2, pode haver uma divergência entre os valores semânticos e sintáticos para gênero no caso dos substantivos sobrecomuns. Reparem que a primeira AVM representa a projeção máxima da construção nominal e não o substantivo que representa o seu núcleo.

Na seção 2, vimos que as construções nominais cujo núcleo é um substantivo sobrecomum não possuem um gênero semântico definido. Como podemos ver no exemplo (9), repetido abaixo para maior comodidade, o valor de INDEX | GEN apresenta uma disjunção, o que permite que ele seja instanciado a partir da co-indexação com um pronome de qualquer gênero.

(9) Construção Nominal – *A criança*.

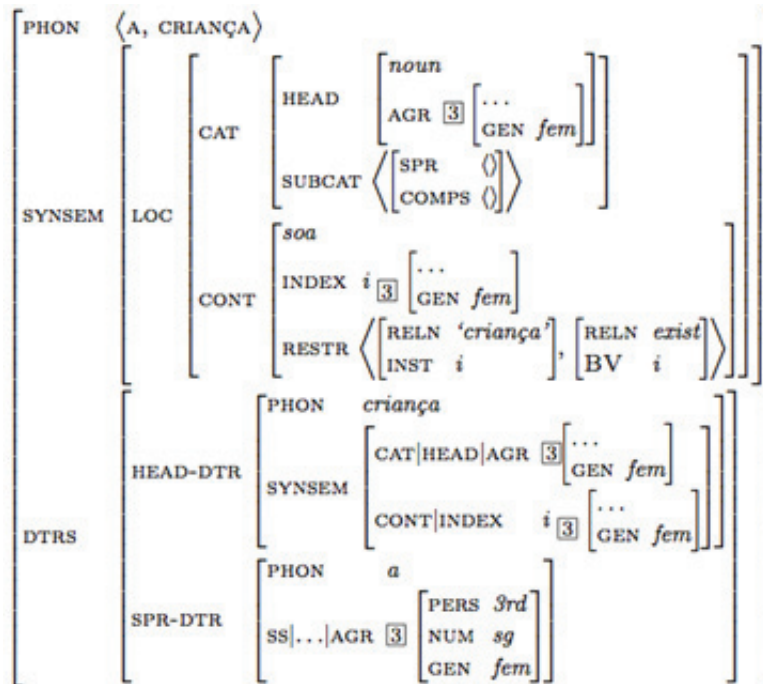


Os dados por nós apresentados comprovam essa possibilidade, uma vez que os falantes nativos no grupo de controle permitem que um sobrecomum masculino seja recuperado por um pronome feminino em 66.6% dos casos, e que um sobrecomum feminino seja recuperado por um pronome masculino 58.3% das vezes. Contudo, é interessante notar que a medida *off-line* mostra que em torno de 40% dos casos os falantes nativos preferem não permitir a resolução pronominal em casos em condições assimétricas. Além disso, os resultados indicam que, para o grupo de falantes nativos, há diferenças significativas entre as condições SFM e SMF em relação aos tempos médios de leitura *on-line* na área do verbo (451 x 655 ms – p= 0,03), sugerindo que o sobrecomum masculino aceita menos a retomada com pronome feminino do que o sobrecomum feminino aceita a retomada com pronome masculino. Em ambas as condições assimétricas (SFM e SMF), a medida *off-line* indica chance no estabelecimento da correferência, o que contrasta significativamente com as condições em que há, respectivamente, simetria entre o antecedente e a retomada pronominal (SFF e SMM), que apresentam alto índice de correferência e tempos médios de decisão significativamente mais baixos. As condições com os comuns de dois gêneros não diferem entre si e contrastam significativamente apenas com a condição SMF.

9. O valor de PERS é irrelevante nesse caso e só está descrito para facilitar a leitura da AVM.

A primeira vista, poderíamos pensar que existe um problema com a maneira como os falantes aplicam a restrição à resolução pronominal apresentada em (13). Porém essa análise não nos parece correta, uma vez que esse princípio é observado sem maiores problemas nos casos de substantivos comuns de dois gêneros. Parece-nos mais plausível supor que os falantes nativos têm uma maior propensão a manter o compartilhamento estrutural dos valores INDEX (semânticos) e AGR (sintáticos), uma vez que, em português, a grande maioria dos substantivos é do tipo comum e, como vimos no exemplo (5), seguem necessariamente a regra que estabelece o compartilhamento dos valores de concordância sintática e semântica. Ou seja, ao invés da descrição em (9), os falantes nativos preferem a representação do sintagma nominal “a criança” descrito em (14) em aproximadamente 40% dos casos.

(14) Variação da Construção Nominal – *A criança*



Repare que o compartilhamento estrutural marcado pela etiqueta [3] faz com que o gênero semântico esteja resolvido na projeção máxima do substantivo. Essa representação força o falante a bloquear a correferência em condições assimétricas.

Essa descrição das projeções nominais é reforçada pelo fato de que os sobrecomuns simétricos não demonstram facilitação da resolução pronominal na medida *on-line*, em relação aos comuns de dois gêneros simétricos, cuja especificação formal do traço de gênero se dá a partir do valor de AGR dos determinantes. Na medida *off-line*, no entanto, os sobrecomuns assimétricos diferem tanto de seus correlatos simétricos, quanto dos comuns de dois gêneros simétricos (comuns de dois gêneros assimétricos não foram testados, por serem claramente agramaticais). Esse padrão aponta para a existência de diferenças significativas na computação do gênero gramatical na sintaxe e na semântica, o que nos parece ser a evidência de que o compartilhamento dos valores de AGR e INDEX imposto aos substantivos comuns e comuns de dois gêneros também é aplicada aos sobrecomuns em alguns casos.

É importante notar também que ambas as assimetrias, de fato, se igualam na medida *off-line*, mas a medida *on-line* demonstra que o pareamento de traços é imediatamente recusado quando o antecedente é masculino e o pronome feminino, contrastando com a condição em que o antecedente é feminino

e o pronome masculino, significativamente melhor aceita *on-line*. Avalia-se que a explicação para o efeito de gênero encontrado na computação de traços no processamento *on-line* está relacionada ao caráter *default* do gênero masculino, que pode incluir o feminino, ao contrário do gênero feminino, que é mais marcado e não inclui o masculino. Ao encontrar o pronome masculino **ele**, o processador busca pelo antecedente correferencial possível. Encontrando tanto um sobrecomum masculino (e.g. o bebê) quanto um feminino (e.g. a visita), o pareamento é feito automaticamente. Por outro lado, ao buscar o antecedente para o pronome feminino **ela**, o pareamento com o antecedente masculino é dificultado por não poder ser incluído no pronome feminino. Posteriormente, na fase interpretativa, menos reflexa, a possibilidade de interpretação não correferencial do pronome é entretida em ambos os casos, impondo-se e obliterando o efeito de assimetria de gênero estabelecido na fase mais automática da computação de traços.

Os resultados do grupo de aprendizes avançados de PB apontam que não há diferença significativa na medida *on-line* em todas as condições, sugerindo menos sensibilidade ao gênero na fase computacional. Na medida *off-line*, os índices de correferência foram mais baixos para as condições com o pronome feminino, inclusive com os comuns de dois gêneros, porém, tal diferença não foi significativa, só tendo, de fato, demonstrado uma tendência de maior correferência com as condições com pronome masculino. Do ponto de vista representacional, é curioso notar que a regra de compartilhamento estrutural entre os valores de AGR e INDEX, que parece tão natural para o falante nativo, não é aplicada com a mesma ênfase no caso de falantes de L2. Ou seja, esse falantes parecem não determinar os valores de INDEX na projeção máxima, o que faz com que eles aceitem a correferência pronominal em condições assimétricas com maior naturalidade. Na verdade, os resultados sugerem que a computação dos valores de INDEX pode estar sendo influenciada por critérios extralinguísticos (ou no mínimo de difícil interpretação).

Os falante de herança, por sua vez, apresentam um caso mais curioso. Os resultados da medida *on-line* do grupo de falantes de herança apontam para a mesma direção que a encontrada no grupo de nativos, com diferenças significativas entre as condições SFM e SMF em relação aos tempos médios de leitura na área do verbo, sugerindo que o sobrecomum masculino aceita menos a retomada com pronome feminino do que o sobrecomum feminino aceita a retomada com o pronome masculino. Por outro lado, na medida *off-line* há um alto índice de correferência em todas as condições, independentemente de haver ou não assimetria de gênero. Com base nesses resultados, sugere-se que, esse grupo, na fase *on-line*, parece utilizar-se de seu conhecimento armazenado na Memória Procedimental (Ullman 2001, 2006), uma vez que consegue captar a diferença entre condições assimétricas e simétricas no estabelecimento da correferência. Tal diferença não foi percebida, na fase *on-line*, pelos aprendizes avançados de PB.

Em termos representacionais, vemos uma discrepância entre o que parece ser dois mecanismo opostos para se estabelecer o valor de INDEX | GEN. Ao analisarmos a medida *on-line* no segmento pós-crítico, temos a impressão de que o falante de herança impõe restrições semelhantes às dos falantes de L1, ou seja, o compartilhamento dos gêneros semântico e sintático. Porém, quando olhamos as medidas *off-line*, vemos que os falantes de herança permitem uma grande flexibilidade na retomada dos pronomes, demonstrando uma variação na atribuição do valor de INDEX | GEN.

A análise do caráter *default* do gênero masculino serve para o grupo de falantes de herança, na medida *on-line*, uma vez que, diante de um pronome masculino, o processador aceita fazer a correferência, tanto com o sobrecomum masculino, quanto com o feminino. Já na busca do antecedente para o

pronome feminino *ela*, o pareamento com o antecedente masculino é dificultado por não poder ser incluído no pronome feminino. Posteriormente, porém, na fase integrativa, há interpretação correferencial do pronome entretida em todas as condições, diluindo-se o efeito *default* do gênero da fase anterior, diferentemente do que foi encontrado para o grupo controle.

Em linhas gerais, o contraste entre as medidas *on-line* e *off-line* dos três grupos nos permite observar como a correlação entre os valores de gênero na sintaxe e na semântica é estabelecida. Talvez a grande dificuldade que os falantes de Português L2 têm com concordância nominal esteja realmente vinculada à falta da restrição sobre o compartilhamento dos valores de AGR e INDEX. Porém, somente com outros estudos específicos, incluindo construções agramaticais com substantivos comuns, poderemos confirmar essa hipótese.

7. CONCLUSÕES

Resumindo:

- Grupo Controle: os resultados indicam que o sobrecomum masculino (o bebê) aceita menos o pronome feminino (ela) do que o sobrecomum feminino (a vítima) aceita o pronome masculino (ele). Na medida *off-line*, os índices de correferência são maiores em condições simétricas (SMM e SFF) do que nas condições assimétricas (SMF and SFM).
- Grupo de falantes de herança: na medida *on-line*, as diferenças nos tempos médios de leitura entre as condições SMF e SFM indicam a presença do efeito *default*, também observado para o grupo controle, com o sobrecomum masculino aceitando menos o pronome feminino do que o sobrecomum feminino aceitando o pronome masculino. Por outro lado, na medida *off-line*, há um alto índice de correferência em todas as condições.
- Grupo de aprendizes avançados de PB: não há diferença significativa na medida *on-line* em todas as condições, sugerindo menos sensibilidade ao gênero na fase computacional. Na medida *off-line*, os índices de correferência foram mais baixos para as condições com o pronome feminino.
- O cruzamento do fator GRUPO de sujeitos (Nativos, falantes de herança e L2) com os tempos médios de leitura da região do verbo nas condições assimétricas (SFM e SMF) revelou um efeito altamente significativo principal do fator grupo, mas não mostrou efeito da assimetria e nem interação entre os dois fatores. A assimetria presente no grupo de nativos e no grupo de falantes de herança, bem como os tempos médios de leitura do grupo L2, de modo geral, significativamente mais altos, garante a diferença entre os grupos, mas a inexistência de assimetria SFM x SMF, neste último grupo, oblitera a assimetria presente nos grupos de nativos e de falantes de herança. De fato, quando se considera apenas os nativos e os falantes de herança no fator GRUPO, ou seja, excluindo-se o grupo L2, observa-se tanto o efeito principal do fator grupo altamente significativo, quanto o efeito principal da assimetria, sem interação entre os fatores.

Esse estudo mostrou um efeito *on-line* do gênero morfológico na resolução pronominal para os nativos e os falantes de herança do português brasileiro. Ambos mostraram um efeito de assimetria de gênero no segmento pós crítico (verbo). Observa-se um efeito *default* do pronome masculino, na fase inicial da computação correferencial, tanto no grupo de nativos quanto no de falantes de herança: um pronome feminino é lido significativamente de forma mais lenta quando o antecedente é um sobrecomum masculino (o bebê-ela). Já os aprendizes adultos de L2, com L1 inglês, não mostraram essa sensibilidade *on-line* de marcação de gênero, mas tendem a estabelecer menos correferência com

pronomes femininos na fase *off-line*. Os resultados dos falantes de herança sugerem, portanto, o uso de um mecanismo mais automático de análise de sua L1 PB, na fase computacional, comparativamente aos aprendizes avançados de PB, que não demonstraram sensibilidade à assimetria de gênero com os sobrecomuns na fase *on-line*.

PRONOUN RESOLUTION WITH DP ANTECEDENTS WHOSE GENDERS REFLECT SEX AND IRREGULAR ONES THAT HAVE GENDER ASSIGNED BY CONVENTION BY SPEAKERS OF BRAZILIAN PORTUGUESE AS A NATIVE LANGUAGE, AND HERITAGE LANGUAGE AND AS A SECOND LANGUAGE

ABSTRACT

This paper explores the differences in the computation of pronominal resolution by native, heritage and second language speakers of Portuguese. We look at constructions where the pronoun's antecedent is an epicene noun to see how gender marking at the syntactic and semantic levels can affect processing. We analyze the formal properties of constructions headed by epicenes using the Head-driven Phrase Structure Grammar framework, and we use this description to interpret the results of a self-paced reading experiment. Our results suggest that participants from the three groups respond differently to information on gender represented at the syntactic and/or semantic levels.

KEY WORDS: pronoun resolution, epicene & bigender nouns, Head-Driven Phrase Structure Grammar, bilingual sentence processing, second language acquisition, portuguese

REFERÊNCIAS BIBLIOGRÁFICAS:

Alarcon, Irma. (2011). Spanish gender agreement under complete and incomplete acquisition: Early and late bilinguals' linguistic behavior within the noun phrase, 332-350. In *Bilingualism: Language and Cognition* 14 (03).

Alarcon, Irma. (2006). *The second language acquisition of Spanish gender agreement: The effects of linguistic variables on accuracy*. Munich: Lincom Europa.

Amaral, Luiz. (2001). *Seleção Semântica de Complementos Verbais em HPSG*. Tese de Mestrado. Rio de Janeiro: Pontifícia Universidade Católica do Rio de Janeiro.

Belikova, Alyona; White, Lydia. (2009). Evidence for the Fundamental Hypothesis or not? *Island Constraints Revisited*. *SSLA*, 31, 199–223.

Benmamoun, Elabbas, Montrul Silvina, Polinsky, Maria. (2010). *White Paper: Prolegomena to Heritage Linguistics*. Harvard University

Bresnan, Joan. (2001). *Lexical-Functional Syntax*. Oxford: Blackwell.

Birdsong, David. (1999). *Second language acquisition and the Critical Period Hypothesis*. Mahwah, NJ: Erlbaum.

Bley-Vroman, Robert. (1989). The logical problem of second language learning, in Gass, S., &

- Schachter, J. (eds.), *Linguistic Perspectives on Second Language Acquisition*, CUP, 41-68.
- Bresnan, Joan (2001). *Lexical-Functional Syntax*. London: Blackwell.
- Cacciari, Cristina; Corradini, Paola; Padovani, Rafael & Carreiras, Manuel (2011). Pronoun resolution in Italian: The role of grammatical gender and context. *Journal of Cognitive Psychology*, 23:4, 416-434
- Cacciari, Cristina; Padovani, Rafael. (2007). Further evidence on gender stereotype priming in language: Semantic facilitation and inhibition on Italian role nouns. *Applied Psycholinguistics*, 28, 277-293.
- Carreiras, Manuel., Garnham, Alan., Oakhill, Jane. V., & Cain, Kate. (1997). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *Quarterly Journal of Experimental Psychology*, 49A(3), 639-663.
- Clahsen, Harald; Felser, Claudia. (2006). Grammatical processing in language learners. *Applied Psycholinguistics* 27 (2006), 3–42.
- Chomsky, Noam (1973). *Conditions on Transformations*, in Anderson and Kiparsky, A Festschrift for Morris Halle, New York: Holt, Rinehart & Winston, pp. 232–286
- Correa, Letícia. (2001). Concordância de gênero no processamento de formas pronominais. *Cad.Est. Ling.*, Campinas, 77-92, Jan/Jun.
- Duffy, S. A., & Keir, J. A. (2004). Violating stereotypes: Eye movements and comprehension processes when text conflicts with word knowledge. *Memory and Cognition*, 32, 551-559.
- Faber, Andrea; Amaral, Luiz & Maia, Marcus. (2012). Processing of anaphoric subjects with epicene and bigender antecedents by Native and English L2 speakers of Spanish. *II Workshop em Processamento Anafórico*. UFC.
- Foote, Rebecca. (2011). Integrated knowledge of agreement in early and late English–Spanish bilinguals. *Applied Psycholinguistics* 32, 187–220.
- Johnson, Jacqueline.S.; Newport, Elissa.L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21, 60-99.
- Garnham, Alan. (1999). Reference and anaphora. InS. Garrod & M. Pickering (Eds.), *Language processing* (pp. 335-362). Hove, UK: Psychology Press.
- Gazdar, Gerald, Ewan Klein, Geoffrey Pullum & Ivan Sag. (1985). *Generalized Phrase Structure Grammar*. Oxford: Basil Blackwell; Cambridge, MA: Harvard University Press.
- Kathol, A. (1999). Agreement and the syntax-morphology interface in HPSG. In Levine, R. and Green, G., editors, *Studies in Contemporary Phrase Structure Grammar*, pages 223–276. Cambridge University Press, Cambridge, UK.
- Kennison, Shelia. M., & Trofe, Jessie. L. (2003). Comprehending pronouns: A role for word-specific gender stereotype information. *Journal of Psycholinguistic Research*, 32, 355-378.

- Montrul, Silvina., Foote, Rebecca., & Perpiñan, Silvia. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language Learning*, 58, 503–553.
- Montrul , Silvina. (2008). *Incomplete acquisition in bilingualism. Re-examining the age factor* . Amsterdam: Benjamins .
- Montrul , Silvina. (2006). On the bilingual competence of Spanish heritage speakers: Syntax, lexical semantics and processing. *International Journal of Bilingualism* , 10 , 37 –69.
- Montrul , Silvina. (2004). Subject and object expression in Spanish heritage speakers: A case of morpho-syntactic convergence. *Bilingualism: Language and Cognition* , 7 , 1 –18.
- Nicol, Janet., & Swinney, David. (2002). The psycholinguistics of anaphora. In A. Barss (Ed.), *Anaphora* (pp. 72- 104). New York, NY: Blackwell Press.
- Nicol, Janet., & Swinney, David. (1989). The role of structure in co-reference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, 18, 5-19.
- Paradis, Michael. (2004). *A neurolinguistic theory of bilingualism*. Amsterdam: John Benjamins.
- Paradis, Michael. (1997). The cognitive neuropsychology of bilingualism. In A. De Groot & J. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 331–354). Mahwah, NJ: Erlbaum.
- Paradis, Michael. (1994). Neurolinguistic aspects of implicit and explicit memory: Implications for bilingualism and SLA. In N. Ellis (Ed.), *Implicit and explicit language learning* (pp. 393–419). London: Academic Press.
- Polinsky, Maria. (2011). Reanalysis in Adult Heritage Language. New Evidence in Support of Attrition. *Studies in Second Language Acquisition*, 33 , 305– 328 .
- Polinsky , Maria. (2008). Gender under incomplete acquisition: Heritage speakers’ knowledge of noun categorization. *The Heritage Language Journal* , 6 , 40 –71. Retrieved September 3, 2009, from <http://www.heritagelanguages.org/>
- Polinsky , Maria. (2006). Incomplete acquisition: American Russian. *Journal of Slavic Linguistics*, 14 , 191 –262.
- Pollard, Carl & Ivan Sag. (1987). *Information-Based Syntax and Semantics*. Vol.1. CSLI Lecture Notes no.13. Stanford: Center for the Study of Language and Information.
- Pollard, Carl and Ivan Sag. (1994). *Head-Driven Phrase Structure Grammar*. Chicago: The University of Chicago Press.
- Ravid, Dorit. & Farah, Rola. (1999). Learning about noun plurals in early Palestinian Arabic. *First Language*, 19, 187-206.

Sag, I., Wasow, T., and Bender, E. (2003). *Syntactic Theory: a formal introduction*. CSLI, Stanford: CA.

Ullman, Michael. (2006). The declarative/procedural model and the shallow structure hypothesis. *Applied Psycholinguistics*, 27, 97–105.

Ullman, Michael. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition*, 4, 105–122.

Valdés, Guadalupe. (2000). *The teaching of heritage languages: an introduction for Slavic-teaching professionals*. The learning and teaching of Slavic languages and cultures, Olga Kagan and Benjamin Rifkin (eds.), 375–403.

Van Gompel, Roger, & Majid, Asifra. (2004). Antecedent frequency effects during the processing of pronouns. *Cognition*, 90, 255-264.