

SIMPÓSIO 29

MORFEMAS GRAMATICAIS EM PORTUGUÊS LÍNGUA NÃO MATERNA

COORDENAÇÃO:

Professora Ana Madeira
Universidade Nova de Lisboa
ana.madeira@fcsh.unl.pt

Professora Maria Francisca Xavier
Universidade Nova de Lisboa
mf.xavier@fcsh.unl.pt

REDES NEURONAIS DE MAPAS AUTO-ORGANIZADOS PARA ANÁLISE DE L2 : UTILIZAÇÃO DE PISTAS MORFOLÓGICAS EM L2

Nuno MARQUES¹Sandro DIAS²Ana MADEIRA³Francisca XAVIER⁴

RESUMO: Apresenta-se um estudo sobre a análise computacional de texto contendo transcrições de entrevistas de aprendentes de L2. Utilizaremos o computador para efetuar a análise do texto dos vários falantes, tendo por base a contagem da frequência relativa de utilização de morfemas gramaticais. As entrevistas conjugam frases apresentadas pelo entrevistador (um falante nativo de Português, cujo texto pode ser tido como referência) e as respostas do entrevistado, as quais - dependendo do seu domínio do Português - podem revelar as dificuldades inerentes à aprendizagem.

A contagem automática da frequência de utilização de morfemas gramaticais é fácil: trata-se de classes gramaticais fechadas, onde basta uma consulta a um dicionário para a sua identificação. Pode-se assim ter acesso a estatísticas sobre unidades de significação linguística importantes como por exemplo os pronomes (sujeito, objeto e possessivos) ou os determinantes. Estas pistas podem depois ser conjugadas com informação verbal através da utilização de um dicionário de flexão verbal do Português (MarquesLopesRocio94). Consegue-se assim obter uma análise de frequência tendo em conta a concordância verbal. Uma vez obtidas as frequências relativas e conjuntas destas pistas morfológicas, será possível a sua caracterização por recurso a algoritmos clássicos de análise de dados. O estudo apresentado baseia-se na utilização de mapas auto-organizados (acrónimo original em Inglês: SOM, Kohonen2002), um algoritmo que tem a capacidade de projetar dados com dimensionalidade elevada num mapa neuronal, tipicamente, a duas dimensões. Cada contagem sobre uma pista morfológica corresponde a uma dimensão distinta, projetada em neurónios no mapa a duas dimensões. Cada neurónio representa um protótipo dos dados, logo, para cada entrevista, um padrão corresponde à frequência detetada de pistas sobre morfemas gramaticais de um indivíduo. O algoritmo SOM efetua aprendizagem competitiva, não supervisionada de forma iterativa: em cada iteração, os neurónios competem para representar o padrão de entrada, sendo selecionado o neurónio mais próximo. Por fim, atualiza-se esse neurónio e todos os seus vizinhos mais próximos. Assim, depois de alguns milhares de iterações, o processo de adaptação é concluído, ficando os padrões semelhantes próximos entre si no mapa neuronal final, i.e. há uma ordenação topológica no mapa final. Será apresentado um estudo de viabilidade (incluindo vantagens e desvantagens) da utilização de diversas técnicas computacionais de análise de

¹ MARQUES, Nuno - UNL, Faculdade de Ciências e Tecnologia, Departamento de Informática, Quinta da Torre, 2829 -516 CAPARICA, PORTUGAL, nmm@di.fct.unl.pt.

² DIAS, Sandro - UNL, Faculdade de Ciências Sociais e Humanas, Centro de Linguística da Universidade Nova de Lisboa, Av. De Berna 26-C, 1069-050 LISBOA, PORTUGAL, sandrodias@fcsh.unl.pt.

³ MADEIRA, Ana - UNL, Faculdade de Ciências Sociais e Humanas, Centro de Linguística da Universidade Nova de Lisboa, Av. De Berna 26-C, 1069-050 LISBOA, PORTUGAL, ana.madeira@fcsh.unl.pt.

⁴ XAVIER, Francisca - UNL, Faculdade de Ciências Sociais e Humanas, Centro de Linguística da Universidade Nova de Lisboa, Av. De Berna 26-C, 1069-050 LISBOA, PORTUGAL, mf.xavier@fcsh.unl.pt.

dados, conjugadas com mapas auto-organizados sobre as frases do corpus de entrevistas de aprendentes de Português L2.

PALAVRAS-CHAVE: Aquisição; L2; pistas morfológicas; redes neuronais; SOM.

Redes Neuronais de Mapas Auto-Organizados para Análise de L2: Utilização de Pistas Morfológicas em L2

Um projeto de investigação sobre a aquisição do português por estrangeiros em desenvolvimento desde 2005 por uma equipa do Centro de Linguística da Universidade Nova de Lisboa tem proporcionado a constituição de diversos corpora textuais, tanto de produções escritas como orais⁵. A organização destes corpora de acordo com a proficiência dos informantes em português L2 tem sido realizada, fundamentalmente, com base na avaliação feita pelos professores de língua. Esta avaliação assenta, contudo, em critérios que visam determinar a competência de comunicação dos aprendentes com o objetivo de os agrupar segundo os níveis de língua estabelecidos para o ensino/aprendizagem de português L2. Assim, a avaliação dos professores excede o estrito conhecimento gramatical dos falantes e não informa suficientemente sobre o que se pretende identificar através de um projeto de investigação cujo objetivo é compreender o processo de desenvolvimento linguístico dos falantes de português L2, através do estudo das propriedades gramaticais que vão sendo adquiridas por grupos de falantes de diferentes línguas maternas, em sucessivos estados de conhecimento da L2.

No âmbito de um projeto desta natureza torna-se necessário desenvolver trabalho em colaboração com investigadores da área do tratamento automático das línguas naturais no sentido de encontrar ferramentas que auxiliem a realização das diversas tarefas de recolha, organização, armazenamento, classificação e extração de dados tanto do perfil dos informantes como de tarefas de elicitación ou dos textos por eles produzidos.

Sendo um dos nossos objetivos de natureza prática determinar automaticamente níveis de conhecimento gramatical de falantes de português L2, a partir de atividades de compreensão e de produção linguísticas, tendo por base determinadas propriedades morfosintáticas da língua em estudo, realizámos algumas experiências de análise computacional de texto com uma ferramenta de redes neuronais, a ferramenta SOM (Vesanto et al. 2000).

Para esta experiência, foi constituído um corpus textual composto por produções orais em português L2, a partir de entrevistas transcritas que foram realizadas a falantes de diferentes línguas. Estes informantes eram estudantes estrangeiros em ensino formal português, nomeadamente dos 3 ciclos do ensino básico e, também, do ensino superior (do programa Erasmus). Foi com a intenção de testar as potencialidades do sistema que, deliberadamente, se constituiu um corpus textual produzido por aqueles estudantes que constituem um grupo muito heterogéneo no que respeita a algumas variáveis que são habitualmente controladas em estudos sobre aquisição de línguas, em particular, a idade, a língua materna e, eventualmente, outras línguas, o nível de escolaridade e o tempo de aprendizagem e de imersão na L2. Estas são informações que efetivamente constam das fichas individuais dos informantes e que podem ser consideradas numa interpretação fina dos resultados da análise computacional dos dados linguísticos, resultante da aplicação da ferramenta àquele corpus. Nas respetivas fichas encontra-se ainda a autoavaliação do conhecimento do português L2, numa escala de A a D, registada pelos informantes.

As entrevistas são em geral constituídas por duas partes, uma em que se pretende que o entrevistado fale livremente de si próprio, dos seus interesses e perspetivas a curto e médio prazo, e outra em que deverá fazer a descrição de uma história em banda desenhada. As entrevistas são transcritas por falantes nativos de português.

O corpus digitalizado e os dicionários já existentes que lhe foram associados permitem a contagem da frequência relativa de utilização de morfemas gramaticais. A contagem automática da frequência de utilização deste tipo de morfemas é fácil: trata-se de classes gramaticais fechadas, onde basta uma consulta a um dicionário para a sua identificação. Pode-

⁵ Disponíveis em <http://l2.clunl.edu.pt/>

se assim ter acesso a estatísticas sobre unidades de significação linguística importantes como por exemplo os pronomes (sujeito, objeto e possessivos) ou os determinantes. Estas pistas podem depois ser conjugadas com informação verbal através da utilização de um dicionário de flexão verbal do Português (MarquesLopesRocio94). Consegue-se assim obter uma análise de frequência tendo em conta a concordância verbal. Uma vez obtidas as frequências relativas e conjuntas destas pistas morfológicas, será possível a sua caracterização por recurso a algoritmos clássicos de análise de dados.

Após várias experiências que implicaram conjuntos de pistas morfológicas diferentes qualitativa e quantitativamente, verificou-se que se obtinha melhores resultados se o número de pistas fosse fortemente restringido. Assim, as pistas implicadas nesta aplicação do sistema têm por base:

- estruturas envolvendo verbo (V) e clítico (C);
- estruturas com vários verbos (VV);
- estruturas com interjeições e verbo(s) (I IV V),

que se exemplificam a seguir:

VCV

“porque o chão tava molhado, depois foi, **queria-se ir** embora e escorregou de novo” (entrevista 1053)

“E depois a menina pensou que ele **ia-se casar**” (entrevista 1065)

VV

“**la pintar** a parede outra vez.”

(entrevista 1267)

IVV

“depois do curso, haa, haa, vou, **haa, vou viajar** para, para o norte de Portugal” (entrevista 1584)

Uma vez obtidas as frequências relativas e conjuntas destas pistas morfológicas, foi possível a sua caracterização por recurso a algoritmos clássicos de análise de dados.

O estudo realizado baseia-se na utilização de mapas auto-organizados (acrónimo original em Inglês: SOM, Kohonen2002), um algoritmo que tem a capacidade de projetar dados com dimensionalidade elevada num mapa neuronal, tipicamente, a duas dimensões. Cada contagem sobre uma pista morfológica corresponde a uma dimensão distinta, projetada em neurónios no mapa a duas dimensões. Cada neurónio representa um protótipo dos dados, logo, para cada entrevista, um padrão corresponde à frequência detetada de pistas sobre morfemas gramaticais de um indivíduo. O algoritmo SOM efetua aprendizagem competitiva, não supervisionada de forma iterativa: em cada iteração, os neurónios competem para representar o padrão de entrada, sendo selecionado o neurónio mais próximo. Por fim, atualiza-se esse neurónio e todos os seus vizinhos mais próximos.

O Algoritmo Som

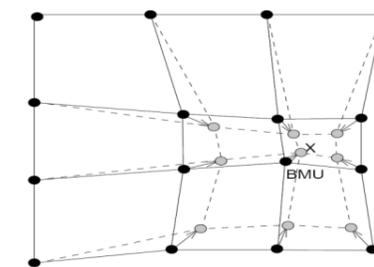


Figura 1- **BMU** (Best matching unit): neurónio que melhor representa **X** (entrevista)

Podemos resumir o algoritmo SOM como o cálculo de uma regressão descrevendo o conjunto de vetores observados $x \in \mathbb{R}^n$ através de um conjunto ordenado de vetores $m_i \in \mathbb{R}^n$. O conjunto de todos os vetores m_i está ordenado topologicamente no mapa de saída e representa o próprio modelo SOM. Cada um dos vetores m_i representa um protótipo (também designado por neurónio).

Na experiência efetuada, cada vetor x representa o conjunto de valores das características calculadas para uma dada entrevista. Esse valor é obtido após normalização da variância para 1.0, quando se tem em conta as contagens na totalidade das entrevistas. O número total de características consideradas foi de 27 (i.e. $n=27$).

O processo iterativo do algoritmo SOM é baseado na seguinte atualização geral:

$$m_i(t+1) = m_i(t) + hc(x(t), i) \cdot (x(t) - m_i(t))$$

onde t é o número de cada iteração do algoritmo. Como dispomos de todos os dados, foi ainda efetuada uma adaptação para épocas do algoritmo geral. Considera-se cada época como a apresentação ao algoritmo de todos os exemplos x . Na variante utilizada, a atualização dos neurónios é apenas efetuada utilizando vetores médios obtidos após cada época (método Batch SOM, Kohonen 2001). Para as experiências apresentadas foi realizado um máximo de 200 épocas. Trata-se de um valor por excesso, pois os mapas obtidos apresentavam estabilidade próximo das 15 épocas, i.e. como há 354 entrevistas, perto das 5000 iterações.

A organização topológica do mapa SOM é obtida pela utilização da função de vizinhança $hc(x, i)$. Esta função decresce consoante a distância entre os neurónios x e i no mapa de projeção do SOM. Utilizou-se uma função de curva gaussiana, onde a vizinhança decresce ao longo das iterações, t . Quando $t=0$ o efeito de vizinhança foi pré-definido para um quarto do mapa, descendo esse valores linearmente até uma vizinhança de 1 (última iteração). Finalmente, para cada iteração t , $c(x(t), i)$ representa o neurónio que melhor se ajusta à observação $x(t)$ (utilizou-se o critério usual da menor distância euclidiana).

Foram efetuados diversos ensaios e testes sobre os parâmetros do SOM até se obter uma configuração estável. Os mapas apresentados de seguida, utilizam um valor inicial proposto por Kohonen 2001: o algoritmo tem início atribuindo aos vetores $m_i(0)$ valores linearmente distribuídos relativos à distância entre os dois maiores vetor próprios (ou autovetor) das duas componentes principais de x .

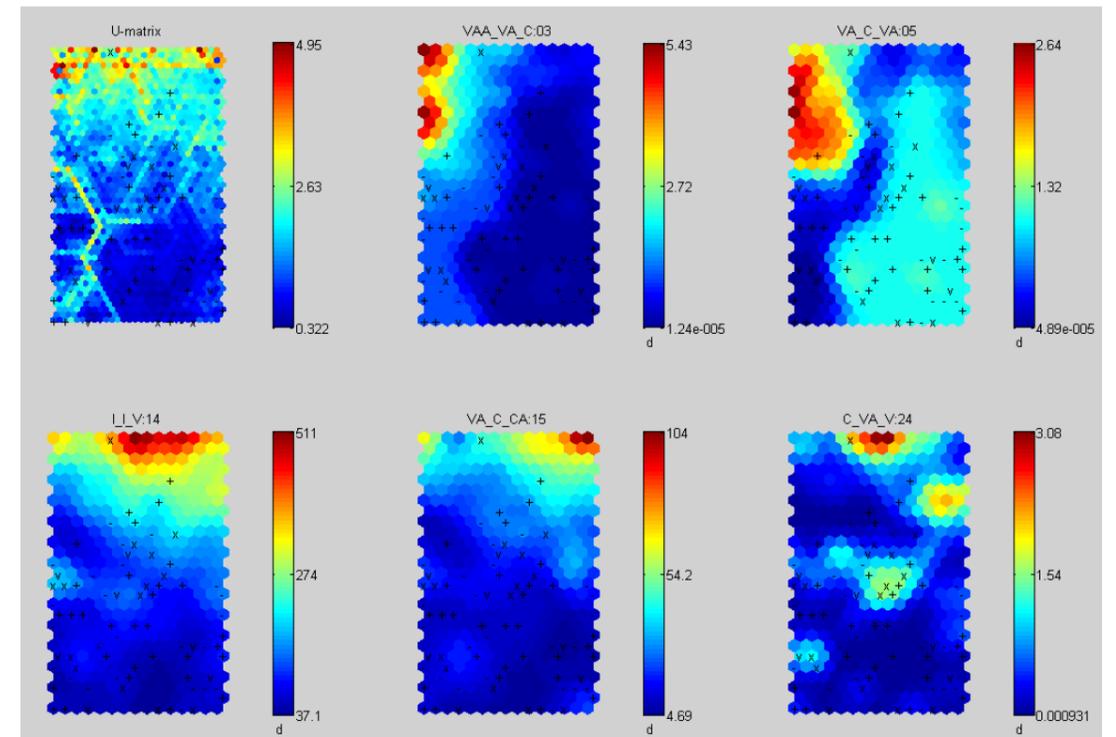


Figura 2 - Mapas neuronais

Verificou-se que é possível relacionar as variáveis morfológicas (calculadas automaticamente) com o mapa conceptual. O primeiro quadro da figura 2 representa a distância de uma unidade (um neurónio) no mapa relativamente às suas unidades (neurónios) vizinhas. Note-se que representando cada neurónio um conjunto de entrevistas, distâncias mais elevadas (a vermelho) representam maior heterogeneidade nessa zona do mapa, enquanto distâncias mais baixas (azul escuro) representam zonas no mapa mais homogéneas. Nos restantes quadros é efetuada uma projeção por componente: como todas as variáveis estão representadas em cada neurónio será possível construir um mapa para cada variável. Assim, na figura estão ainda representados respetivamente (da esquerda para a direita e de baixo para cima) os mapas [I VA C], [VA C VA], [I IV], [VA C CA] e [C VA V]. Neste caso, maiores valores no mapa representam maior predominância daquela variável naquela zona do mapa. Apesar de algum ruído nas pistas utilizadas, os agrupamentos são consistentes porque o SOM agrupou no mesmo neurónio entrevistas com as mesmas características morfológicas. Obtiveram-se bons resultados pois a avaliação obtida através do método utilizado (avaliação essa que é corroborada pelos resultados da análise manual dos corpora) aproxima-se mais do conhecimento gramatical revelado pelos dados dos informantes relativamente às propriedades selecionadas para pistas do que a avaliação mais geral realizada pelos professores de língua ou, ainda, do que a autoavaliação feita pelos informantes, que é subjetiva e distante da avaliação do método utilizado.

CONCLUSÃO

As redes neuronais com mapas organizados são uma ferramenta útil para a caracterização do nível de conhecimento morfosintático dos informantes, expresso no mapa conceptual.

É possível relacionar as variáveis morfológicas (calculadas automaticamente) com o mapa conceptual.

Trabalho Futuro

O próximo passo desta investigação em curso passa pela necessidade de encontrar pistas mais fiáveis, como, por exemplo, a desambiguação morfossintática e o recurso/construção de bons dicionários. Paralelamente, deve-se tentar evitar formas ambíguas (por exemplo : *e – como pista de é - não é um bom indicador de verbo) e procurar evitar o número excessivo de palavras entre pistas (ex: Estraga-o / Por causa disto / casar...).

Referências bibliográficas

J. Lopes, J. G.; Marques, N. C.; Rocio, V. J. (1994) "POLARIS: Portuguese Lexicon Acquisition and Retrieval Interactive System", In The Practical Applications of Prolog, Royal Society of Arts, London, UK.

Teuvo Kohonen. (2001) "Self-Organizing Maps". Springer-Verlag New York, Inc., USA. Vesanto et al. 2000. SOM Toolbox for Matlab 5. Helsinki University of Technology.

SUJEITOS EM CONSTRUÇÕES DE CONJUNTIVO EM PORTUGUÊS L2

Fátima MARTINS⁶

Ana MADEIRA

RESUMO: Tanto o português como o espanhol caracterizam-se por ser línguas de sujeito nulo. É também conhecido que, nestas duas línguas, as orações conjuntivas se distinguem das indicativas quanto às propriedades do sujeito encaixado. Assim, em subordinadas completivas com indicativo, o sujeito da completiva é preferencialmente disjunto do sujeito matriz quando realizado lexicalmente; caso contrário, é interpretado como correferente. Já o conjuntivo, com predicados desiderativos e factivos, por exemplo, apresenta efeitos de obviação i.e., o sujeito é obrigatoriamente disjunto do da matriz, seja ou não realizado lexicalmente. No entanto, o sujeito nulo cria leituras ambíguas, no que respeita à indexação do sujeito, em contextos marcados, tais como na presença de modais, estativos, etc. Este estudo investiga a aquisição das propriedades interpretativas do sujeito em orações conjuntivas, em contraste com contextos indicativos, por aprendentes de português L2, tendo como língua materna o espanhol. Assim, pretende-se estudar o modo como se desenvolve o conhecimento de propriedades que não estão directamente acessíveis no input linguístico nem são ensinadas formalmente. Aplicou-se um teste de compreensão a um grupo de falantes nativos de espanhol (nível intermédio e avançado). O estudo incluiu dois grupos de controlo, um constituído por falantes nativos de português e o outro por falantes de romeno, já que esta língua se caracteriza por ser uma língua de sujeito nulo que não apresenta efeitos de obviação em contextos conjuntivos. Os resultados preliminares indicam um desenvolvimento gradual das propriedades relevantes, evidenciando, os aprendentes, maiores dificuldades na interpretação de sujeitos lexicais em contextos marcados de conjuntivo.

PALAVRAS-CHAVE: aquisição de L2; obviação, sujeitos nulos; modo conjuntivo; modo indicativo.

1. Introdução

Diz-se que numa estrutura se verificam efeitos de obviação sempre que o sujeito de uma oração finita encaixada não pode ser ligado pelo sujeito da oração matriz. Estes efeitos caracterizam as subordinadas completivas com conjuntivo, em línguas como o português e o espanhol (cf. (1) - (3)).

- (1) [... Sujeitoi ... [Sujeitoi T-ind ...]]
- (2) * [... Sujeitoi ... [Sujeitoi T-sub ...]]
- (3)a. *[Eui quero [que (eu)i parta esta noite]]
- b. *[Yoi quiero [que (yo)i parta esta noche]]

Em subordinadas completivas com indicativo, em línguas de sujeito nulo, como é o caso do português, do espanhol e do romeno, o sujeito pronominal da completiva é preferencialmente disjunto do sujeito matriz quando realizado lexicalmente (4a); caso contrário, é preferencialmente interpretado como correferente do sujeito matriz (4b).

- (4)a. O Joãoi acha que ele(i)/j comprou o livro errado
- b. O Joãoi acha que proi /(j) comprou o livro errado

No que respeita à indexação do sujeito, determinados contextos de conjuntivo, que se caracterizam pela presença, por exemplo, de verbos implicativos (doravante designados 'contextos marcados de conjuntivo'), apresentam as mesmas possibilidades interpretativas do que os contextos de indicativo, como se ilustra em (5).

⁶ UNL, Centro de Linguística da Universidade Nova de Lisboa fatima.martins@fch.unl.pt. Boleira de Doutoramento da FCT com a ref. SFRH/BD/48376/2008

- (5)a. O João espera que ele(i)/j consiga despachar-se a horas
 b. O João espera que proi /(j) consiga despachar-se a horas

Este estudo investiga a aquisição das propriedades interpretativas do sujeito em orações conjuntivas, em contraste com contextos indicativos, por aprendentes de português L2, tendo como língua materna o espanhol. Em particular, pretende-se (i) verificar a existência de efeitos de obviação em construções completivas com predicados desiderativos que seleccionam o modo conjuntivo; e (ii) verificar a existência de uma interpretação preferencialmente correferencial do sujeito nulo e preferencialmente disjunta do sujeito lexical em contextos conjuntivos marcados. Com este trabalho procurar-se-á, assim, contribuir para uma melhor compreensão do modo como se desenvolve o conhecimento de propriedades interpretativas que não estão directamente acessíveis no input linguístico nem são ensinadas formalmente. Na secção seguinte, caracteriza-se as propriedades interpretativas dos sujeitos em orações completivas de conjuntivo e de indicativo. Na secção 3, apresenta-se a metodologia adoptada no âmbito deste estudo, que se baseou na aplicação de uma tarefa de compreensão a um grupo de aprendentes de português L2, falantes nativos de espanhol. Os resultados desta tarefa são descritos na secção 4 e, na secção 5, apresenta-se a discussão e as conclusões.

2. A Interpretação dos Sujeitos em Contextos Encaixados

Os efeitos de obviação em orações de conjuntivo são, geralmente, explicados em termos da transparência sintáctica que caracteriza estes domínios oracionais. Por exemplo, Picallo (1985) deriva estes efeitos da inexistência de um tempo de conjuntivo independente – a relação de dependência temporal que se estabelece, necessariamente, entre o tempo encaixado e o tempo matriz produz uma extensão do domínio de ligação da oração de conjuntivo, com a consequência que o sujeito da encaixada não pode estar co-indexado com o sujeito da matriz. Kempchinsky (1985), por outro lado, relaciona a obviação com a presença de um operador nulo, que provocaria a subida de T para C na oração encaixada, logo provocando a extensão do domínio de ligação (ver também San Martin, 2007).

Porém, à semelhança de outras análises propostas na literatura (ver, por exemplo, Ruwet, 1984/1991; Bouchard, 1982, 1983; Farkas, 1992; para uma revisão da literatura sobre obviação, ver Constantini, 2005; Quer, 2006), nem a análise de Picallo nem a de Kempchinsky explicam as assimetrias observadas entre diferentes verbos que seleccionam o conjuntivo, em português. Por exemplo, ao contrário dos complementos seleccionados por verbos volitivos como querer (cf. (3) acima), os complementos de verbos dubitativos e de negação não apresentam estes efeitos (6).

- (6) a. Elei duvida que proi parta amanhã
 b. O João negou que proi chegasse tão tarde.

Além disso, como se observou acima, em determinados contextos, mesmo os verbos que habitualmente espoletam efeitos de obviação permitem o cancelamento desses efeitos. Tal ocorre na presença de determinados elementos na oração encaixada: para além de verbos implicativos (cf. (5) acima), também modais (7), auxiliares (8), estativos (9), passivas (10) e sujeitos focalizados (11) podem produzir este cancelamento.

- (7) José espera que proi possa terminar o trabalho a tempo.
 (8) O José espera que proi tenha ganhado o euromilhões
 (9) A Anai lamenta que proi esteja cansada
 (10) A Anai espera que proi seja eleita para ministra.
 (11) A ministra espera que ELAi/ela própria presidida à sessão

Independentemente da análise que se adopte para explicar a obviação, e apesar dos problemas que possam estar inerentes às diferentes análises que têm sido propostas na literatura, é indubitável que a obviação, em contextos de

conjuntivo, é um fenómeno que é determinado por factores de natureza semântica. Tal distingue as restrições que governam a interpretação de sujeitos encaixados em contextos de conjuntivo daquelas que regem a interpretação dos sujeitos em contextos de indicativo e em contextos marcados de conjuntivo, em línguas de sujeito nulo. De facto, embora estes sujeitos sejam preferencialmente interpretados como sendo referencialmente idênticos ou distintos do sujeito matriz, consoante sejam nulos ou lexicais, respectivamente (cf. (4) acima), estas preferências interpretativas podem ser canceladas em determinados contextos discursivos, como se ilustra em (12). Tal demonstra que a interpretação dos sujeitos nestes contextos é determinada por factores de natureza discursiva.

- (12) a. O João e a Maria foram ontem comprar os manuais de Matemática. O João acha que elei/*j comprou o livro errado
 b. [O pai do João] foi ontem comprar-lhe o manual de Matemática para este ano. Porém, o João acha que pro*i /j comprou o livro errado.

3. O Estudo

3.1. O Problema

Sabendo que as propriedades associadas aos efeitos de obviação e à interpretação de sujeitos lexicais e nulos em contextos de indicativo e em contextos marcados de conjuntivo não estão directamente acessíveis no input nem são ensinadas formalmente, coloca-se as seguintes questões:

- A. Como se desenvolve o conhecimento deste tipo de propriedades?
 B. O desenvolvimento de propriedades interpretativas é influenciado pela L1 dos aprendentes?
 C. Será que se observam assimetrias na aquisição dos diferentes tipos de propriedades interpretativas, ou seja, semânticas (que governam a interpretação em contextos de conjuntivo) e discursivas (que determinam a interpretação em contextos de indicativo e em contextos marcados de conjuntivo)?

A existência de assimetrias entre semântica e discurso é sugerida por ampla evidência empírica, que indica que as propriedades discursivas (Valenzuela 2006, e.o.) poderão ser mais vulneráveis e susceptíveis de atrasos no desenvolvimento que as propriedades semânticas (Slabakova 2008, e.o.). De igual modo, existe alguma evidência que certos tipos de propriedades estão, do ponto de vista do desenvolvimento, mais susceptíveis a efeitos de influência da L1 do que outros. Em particular, alguns estudos parecem indicar que é, sobretudo, no domínio das propriedades discursivas que se observam efeitos persistentes de influência da L1 nos estádios mais avançados.

3.2. Metodologia

3.2.1. Participantes

Aplicou-se um teste de compreensão a um grupo de 10 adultos, falantes nativos de espanhol, com idades compreendidas entre os 30 e os 60 anos. Todos eles começaram a aprender português e vivem em Portugal há, pelo menos, 10 anos, e estão posicionados entre o nível intermédio e o nível avançado de proficiência.

O estudo incluiu dois grupos de controlo: um grupo era constituído por 10 falantes nativos de português, com idades compreendidas entre os 30 e os 60 anos, e com um nível de escolaridade igual ou superior ao 12º ano; o outro grupo era constituído por 9 aprendentes de português, falantes nativos de romeno, com idades compreendidas entre os 20 e os 50 anos, e a aprender português e a viver em Portugal há, pelo menos, 5 anos. Este segundo grupo funcionou como

controlo, uma vez que o romeno se caracteriza por ser, tal como o português e o espanhol, uma língua de sujeito nulo, que, contudo, não apresenta efeitos de obviação em contextos de conjuntivo.

O conjuntivo romeno é caracterizado pela possibilidade de verbos como querer e esperar, que seleccionam o conjuntivo, permitirem tanto leituras de controlo (em que se estabelece correferência entre o sujeito encaixado e o sujeito da oração matriz) como leituras obviativas (em que o sujeito encaixado é interpretado como referencialmente disjunto do sujeito da matriz). Compare-se (13) e (14):

- (13) Vrea să plece.
 quer-3sg conj ir-3sg
 ‘Ela quer ir.’
- (14) (Eai) vrea (eak/el) să plece.
 (elai) quer-3sg (elak/ele) conj ir-3sg
 ‘(Ela) quer que (ela/ele) vá.’
 (Geber & Tonciulescu, 2007: 3)

Assim, ao contrário do português e do espanhol, a língua romena não exhibe efeitos obrigatórios de obviação em estruturas de complementação de conjuntivo.

3.2.2. Condições Testadas

Foram testadas as seguintes condições:

i. Construções conjuntivas (com predicados desiderativos) em contexto de obviação com sujeito nulo (5 itens):

- (15) O João espera que pro ganhe a corrida (pro ≠ o João)

ii. Construções conjuntivas (com predicados desiderativos) em contexto de obviação com sujeito lexical (5 itens):

- (16) O João espera que ele ganhe a corrida (ele ≠ o João)

iii. Construções conjuntivas (predicados desiderativos) sem obviação com sujeito nulo (5 itens):

- (17) O João espera que pro possa ganhar a corrida (pro ≠ o João ou pro = o João)

iv. Construções conjuntivas (predicados desiderativos) sem obviação com sujeito lexical (5 itens):

- (18) O João espera que ele possa ganhar a corrida (ele ≠ o João [leitura preferencial])

v. Construções indicativas (predicados epistémicos) com sujeito nulo (5 itens):

- (19) O João pensa que pro vai ganhar a corrida (pro = o João [leitura preferencial])

vi. Construções indicativas (predicados epistémicos) com sujeito lexical (5 itens):

- (20) O João pensa que ele vai ganhar a corrida (ele ≠ o João [leitura preferencial])

O teste é constituído por 40 itens, 10 dos quais são distractores.

3.2.3. Procedimento

Cada item é precedido por um parágrafo, que tem a função de estabelecer o contexto. Contudo, este contexto é neutro, na medida em que deixa em aberto a interpretação do sujeito da oração subordinada contida na frase de teste. A tarefa do participante consiste em seleccionar a opção que lhe parece estar mais de acordo com a sua interpretação do sujeito da subordinada, de entre as três opções que lhe são apresentadas. Em (21) mostra-se um exemplo de um item de teste:

(21) Todos os anos a Paula e o João divertem-se muito na neve. Este ano, o João acha que vai para a neve ainda mais cedo.

Quem é que o João acha que vai para a neve este ano?

- O João
- A Paula
- Uma outra pessoa

4. Resultados

Nesta secção, apresentam-se os resultados do teste, por condição.

4.1. Contextos de Obviação com Sujeito Nulo

Os resultados para esta condição (ver gráfico 1) mostram uma preferência clara, por parte do grupo de espanhol L1, pela interpretação de referência disjunta, tal como acontece nas respostas obtidas do grupo de controlo constituído por falantes nativos de português. Por outro lado, o grupo de romeno L1 manifesta uma preferência clara pela interpretação de correferência. A assimetria observada entre os resultados do grupo de espanhol L1 e os do grupo de romeno L1 parece indicar a existência de efeitos de influência da L1 – este contextos exibem efeitos de obviação para o grupo de espanhol L1, à semelhança do que ocorre na gramática da sua L1, mas não para o grupo de romeno L1.

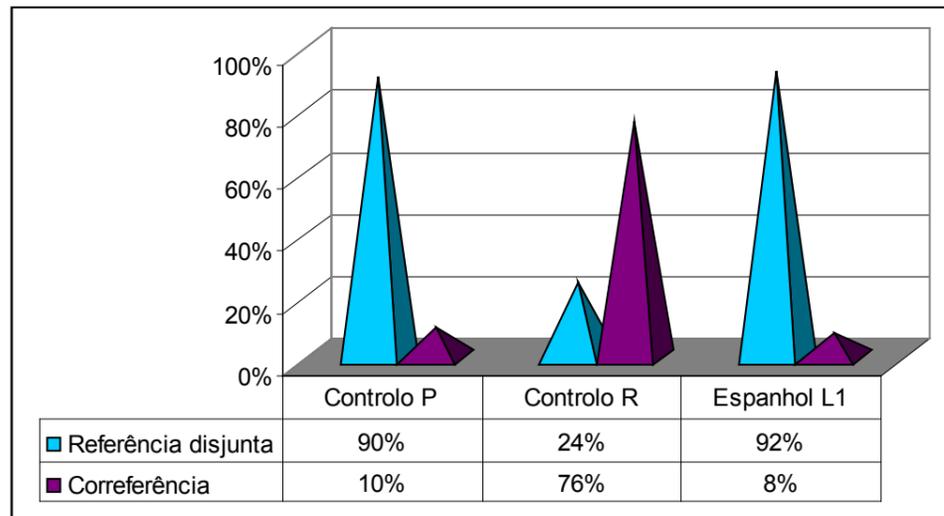


Gráfico 1: Contextos de obviação com sujeito nulo

4.2. Contextos de Obviação com Sujeito Lexical

Nestes contextos, enquanto o grupo de espanhol L1, à semelhança do grupo de controlo de falantes nativos, continua a apresentar efeitos de obviação, com valores próximos de preferência pela interpretação disjunta aos observados com sujeitos nulos, no grupo de romeno L1 verifica-se uma inversão da tendência observada com sujeitos nulos, exibindo aqui uma preferência nítida pela leitura disjunta (ver gráfico 2).

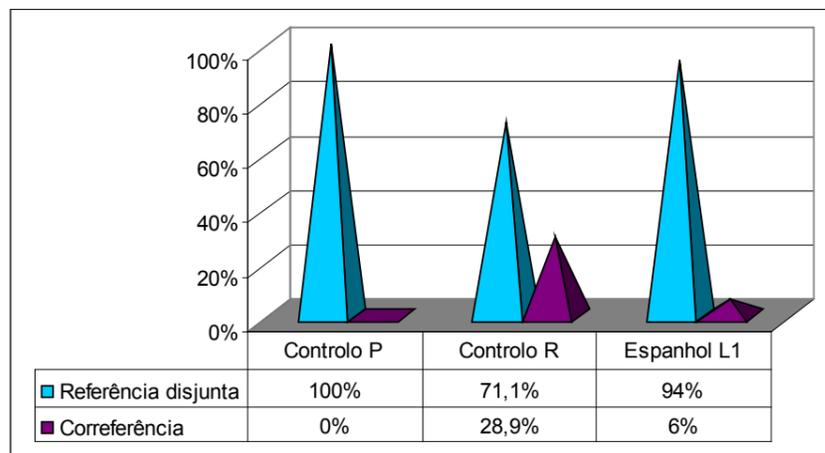


Gráfico 2: Contextos de obviação com sujeito lexical

4.3. Contextos Marcados de Conjuntivo com Sujeito Nulo

Verifica-se que, para o grupo de espanhol L1 (tal como acontece para o grupo nativo), a presença de um verbo implicativo na encaixada provoca um cancelamento dos efeitos de obviação, passando a ser possível tanto a interpretação de

correferência como a de referência disjunta dos sujeitos (ver gráfico 3). Para o grupo de romeno L1, no entanto, a presença de contextos que cancelam a obviação não produz efeito nos resultados, mantendo-se o mesmo padrão de preferência pela interpretação de correferência observado com sujeitos nulos em contextos de obviação (cf. gráfico 1 acima).

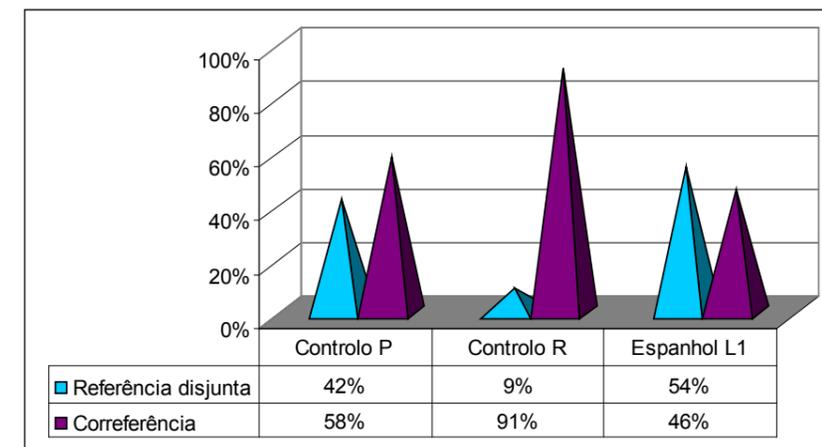


Gráfico 3: Contextos marcados de conjuntivo com sujeito nulo

4.4. Contextos Marcados de Conjuntivo com Sujeito Lexical

Nestes contextos, a preferência pela interpretação disjunta entre o sujeito pronominal lexical da subordinada e o sujeito matriz é evidente nos resultados do grupo de controlo nativo. O mesmo não acontece, porém, quer no caso dos aprendentes espanhóis quer no dos controlos romenos, já que nenhum dos dois demonstra uma clara preferência por qualquer uma das leituras, oscilando entre a correferência e a referência disjunta (ver gráfico 4).

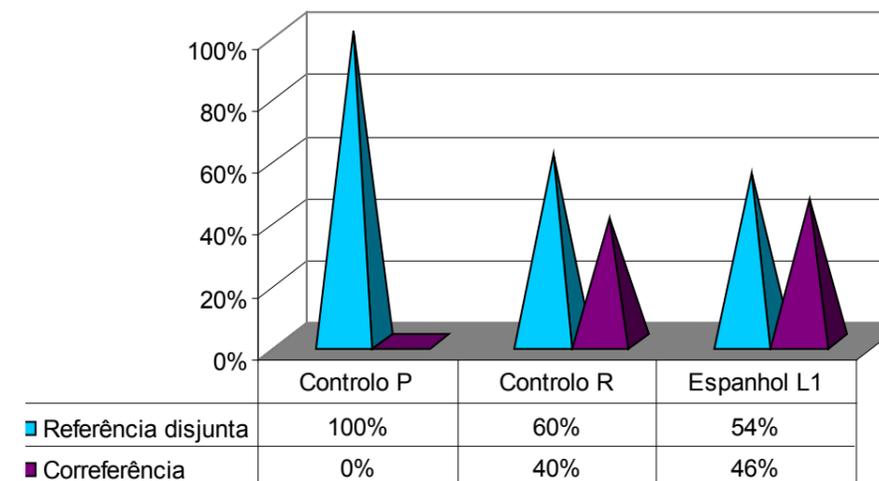


Gráfico 4: Contextos marcados de conjuntivo com sujeito lexical

4.5. Contextos de Indicativo com Sujeito Nulo

Em contextos de indicativo com sujeito nulo, observa-se, em todos os grupos, uma preferência significativa pela interpretação de correferência com o sujeito da matriz. Assim, podemos afirmar que a interpretação de sujeitos nulos

não é problemática em contextos subordinados de indicativo (cf. gráfico 5).

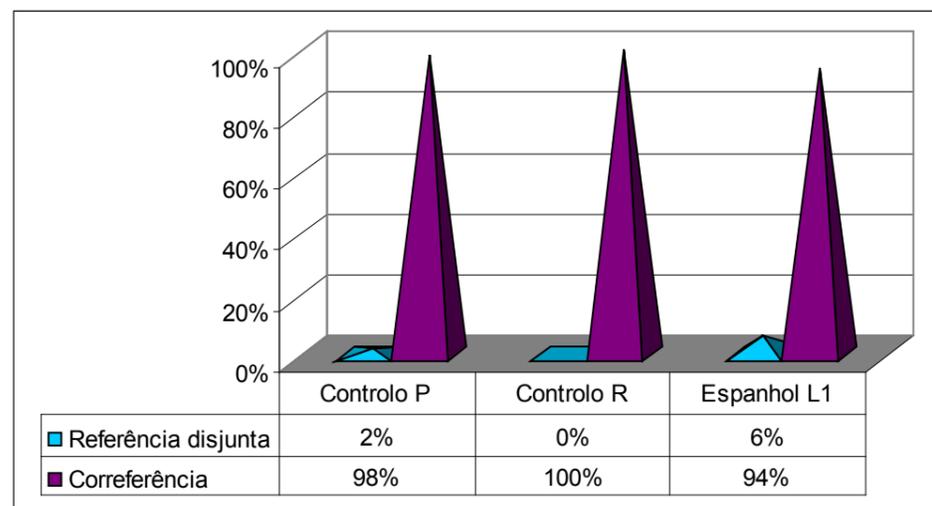


Gráfico 5: Contextos de indicativo com sujeito nulo

4.6. Contextos de Indicativo com Sujeito Lexical

Como é visível no gráfico 6, a presença do sujeito lexical leva a uma preferência pela interpretação de referência disjunta para ambos os grupos de aprendentes, embora não tão marcada como a exibida pelo grupo de controlo nativo.

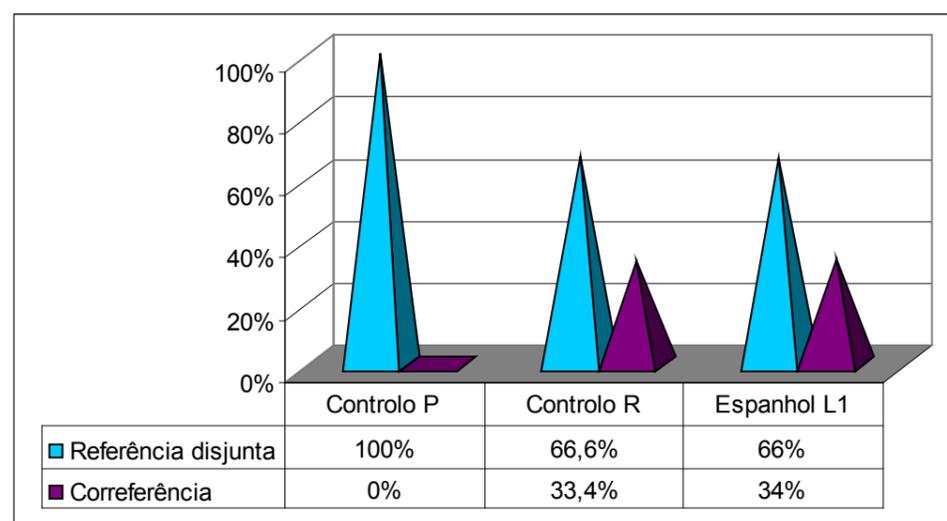


Gráfico 6: Contextos de indicativo com sujeito lexical

5. Discussão e conclusões

Os resultados indicam que os falantes nativos de espanhol não apresentam dificuldades relativamente às restrições

semânticas que caracterizam os contextos com obviação. O desempenho exibido nestes contextos é semelhante ao do grupo de falantes nativos, revelando uma preferência absoluta por uma leitura obviativa do sujeito da oração de conjuntivo, independentemente do seu estatuto.

Este desempenho poderá ser atribuído à influência da L1, uma vez que, como foi referido acima, o espanhol se caracteriza, tal como o português, pela existência de efeitos de obviação nestes contextos. Em alternativa, os resultados observados poderão indicar que o conhecimento de propriedades de natureza semântica, como é o caso das propriedades que determinam as restrições de interpretação dos sujeitos nestes contextos de conjuntivo, se desenvolve cedo, e de forma não problemática, no processo de aquisição da L2.

Uma análise dos resultados obtidos pelo grupo de romeno L1 permite-nos decidir entre estas duas opções. Ao contrário do grupo de espanhol L1, verificou-se que este grupo não revela conhecimento dos efeitos de obviação, interpretando os sujeitos nestes contextos do mesmo modo como interpreta os sujeitos em contextos de indicativo, com leitura preferencial de correferência para os sujeitos nulos e leitura preferencial de referência disjunta para os sujeitos lexicais. Assim, a comparação entre os dois grupos de L2 sugere, muito convincentemente, que a L1 dos aprendentes desempenha um papel importante na aquisição destas propriedades.

Quanto aos contextos marcados de conjuntivo, observou-se que, com sujeitos nulos, o grupo de espanhol L1 não manifesta uma preferência clara pela interpretação de correferência, como seria de esperar, permitindo as duas leituras. Porém, este é também o comportamento exibido pelo grupo de controlo, o que sugere que as condições que determinam a interpretação dos sujeitos nestes contextos não são exactamente idênticas àquelas que caracterizam os contextos de indicativo, em que ambos os grupos mostram uma preferência clara pela leitura de correferência. O grupo de romeno L1, porém, não estabelece uma distinção entre estes dois tipos de contextos, preferindo uma interpretação de correferência em ambos.

Com sujeitos lexicais, por outro lado, o grupo de espanhol L1 apresenta resultados semelhantes aos do grupo de romeno L1: embora se observe preferência por uma leitura de referência disjunta, esta preferência é muito menos marcada do que a demonstrada pelo grupo de falantes nativos. Esta diferença é reveladora de algumas dificuldades na interpretação de sujeitos lexicais, já observada em muitos estudos anteriores (Sorace & Filiaci, e.o.). Neste domínio, a L1 não parece, pois, desempenhar qualquer efeito significativo.

Observam-se, assim, assimetrias na aquisição dos diferentes tipos de propriedades interpretativas: as propriedades semânticas, que governam a interpretação dos sujeitos em contextos de conjuntivo, são adquiridas mais cedo e são mais permeáveis à influência da L1 do que as propriedades discursivas, que determinam a interpretação dos sujeitos em contextos marcados de conjuntivo e em contextos de indicativo. Os resultados do nosso estudo corroboram, pois, as conclusões de estudos anteriores, demonstrando que alguns aspectos da interpretação que são condicionados pelo contexto discursivo não são facilmente transferidos da L1 e se caracterizam por um desenvolvimento mais tardio.

Referências Bibliográficas

Bouchard, Denis. 1982. *On the Content of Empty Categories*. Dissertação de Doutoramento, MIT.

Bouchard, Denis. 1983. The Avoid Pronoun Principle and the Elsewhere Principle. In Peter Sells & Charles Jones (orgs.) *Proceedings of ALNE 13/NELS 13*. Amherst (Mass.): GLSA, pp. 29-36.

Constantini, Francesco. 2005. On obviation in subjunctive clauses: The state of the art. *Annali di Ca' Foscari. Serie occidentale 2005* (vol. 44, 1-2). Università Ca' Foscari di Venezia, pp. 97-132.

Farkas, Donka. 1992. On Obviation. In Ivan A. Sag & Anna Szabolcsi (orgs.) *Lexical Matters*. Stanford University, CSLI, pp. 85-109.

- Geber, Dana & KerenC. Tonciulescu. 2007. Moody Subjunctive in Romanian. In Milica Radišić (org.) *Proceedings of the 2007 Annual Conference of the Canadian Linguistic Association*, 12 pp.
- Kempchinsky, Paula. 1985. The Subjunctive Disjoint Reference Effect. In Carol Neidle & Rafael Nuñez Cedeño (orgs.) *Studies in Romance Linguistics*. Dordrecht: Foris, pp. 123-140.
- Picallo, Carme. 1985. *Opaque domains*. Dissertação de Doutorado, CUNY.
- Quer, Josep. 2006. Subjunctives. In Martin Everaert & Henk van Riemsdijk (orgs.) *The Blackwell Companion to Syntax*, vol. 4. Oxford: Blackwell Publishing, pp. 660-684.
- Ruwet, Nicolas. 1984. *Je veux partir/*Je veux que je parte*: On the distribution of finite complements and infinitival complements in French. *Cahiers de grammaire* 7, pp. 75-138.
- Ruwet, Nicola. 1991. *Je veux partir/*Je veux que je parte*: On the distribution of finite complements and infinitival complements in French. In Nicolas Ruwet (org.) *Syntax and Human Experience*. Chicago-London: The University of Chicago Press.
- San Martin, Itziar. 2007. Beyond the infinitive vs. subjunctive rivalry: surviving changes in Mood. In Luis Eguren & O. Fernández Soriano (orgs.) *Coreference, Modality, and Focus: Studies on the Syntax–Semantics Interface*. Amsterdam: John Benjamins, pp. 171-190.
- Slabakova, Roumyana. (2008) *Meaning in the Second Language*. Studies in Language Acquisition Series, Berlin: Mouton de Gruyter.
- Sorace, Antonella & Francesca Filiaci. 2006. Anaphora resolution in near-native speakers of Italian. *Second Language Research* 22 (3), pp. 339–368.
- Valenzuela, Elena. 2006. L2 end state grammars and incomplete acquisition of Spanish CLLD constructions. In Roumyana Slabakova, Silvina Montrul & Philippe Prévost (orgs.) *Inquiries in Linguistic Development: In Honor of Lydia White*. Amsterdam: Johns Benjamins, pp. 283-304.