

A EXTENSÃO SINTÁTICA NA ESCRITA DE TEXTOS ESCOLARES

Mário MARTINS¹
Rui MARQUES²

Resumo: Neste artigo, apresenta-se um estudo de correlação entre extensão sintática e progressão escolar em textos escritos por crianças e adolescentes em idade escolar monolíngues de português europeu. Para a caracterização da extensão, utilizam-se como medidas a frequência média de palavras por oração (BERMAN, 2007), a frequência média de palavras por unidade-t e a frequência média de orações subordinadas por unidade-t (HUNT, 1965, 1970). Com recurso à plataforma UAM CorpusTool (O'DONNELL, 2008), essas medidas foram aplicadas a um *corpus quasi-longitudinal* composto por 244 textos de registros narrativos (n=122) e argumentativos (n=122), escritos por alunos do 5.º (n=26), do 7.º (n=46) e do 10.º (n=50) ano do sistema escolar português. Os resultados mostram que, por um lado, a extensão da oração, medida pela frequência média de palavras, não mantém correlação com a progressão escolar em nenhum dos dois registros avaliados e, por outro, a unidade-t, quer aferida pela frequência média de palavras, quer pela frequência média de orações subordinadas, só tem representação significativa nos textos do registro narrativo, não sendo detectado o mesmo para os textos argumentativos. Com este trabalho, pretende-se contribuir para uma compreensão mais pormenorizada dos movimentos configuradores do desenvolvimento sintático na língua escrita de crianças e jovens em idade escolar.

Palavras-chave: Extensão sintática. Progressão escolar. Textos escolares.

Introdução

Desde a década de 60, a extensão sintática, enquanto um dos componentes de avaliação da complexidade sintática, tem sido considerada em muitas investigações sobre o desenvolvimento linguístico da língua materna de crianças e jovens. Essas investigações revelam-se expressivas particularmente nas décadas de 60, 70 e 80 (HARPIN, 1973, 1976; LOBAN, 1963, 1976; O'DONNELL *et al.* 1967; PERERA, 1984) e têm tido impacto não apenas no âmbito da investigação, mas também nas práticas pedagógicas. Em seu trabalho seminal, Perera (1984, p. 4) sugere que se reconheça a importância da maturidade (ou da

¹ Universidade Federal do Amapá. *E-mail:* mgcmartins@gmail.com

² Universidade de Lisboa. *E-mail:* rmarques@letras.ulisboa.pt

imaturidade) linguística através da complexidade sintática, o que pode permitir aos professores de línguas compreender melhor os textos produzidos por seus alunos. Segundo Schleppegrell (2004, p. 14), as recomendações curriculares de Walter Loban sobre a complexidade sintática identificada em textos escolares são até hoje referenciadas na elaboração de matrizes curriculares nos Estados Unidos.

Nos anos 90, porém, há uma considerável diminuição do número de estudos, momento em que a preocupação sobre os padrões de realização sintática em língua materna, de acordo com Biber *et al.* (2011, p. 6), cede espaço para os estudos focalizados nos processos cognitivos da escrita, no pensamento crítico e nos contextos socioculturais de textos produzidos em ambiente escolar. A extensão sintática passa a ser tópico recorrente apenas em pesquisas sobre a aquisição de línguas adicionais. O abandono da centralidade da sintaxe, refletido no silêncio de vários relatórios publicados por comissões oficiais dos Estados Unidos, teve um custo explícito, que se vê no enfraquecimento do letramento nacional, tanto na leitura como na escrita (SCOTT, 2004). Com o advento de recursos informáticos mais eficientes para o tratamento de grandes quantidades de texto, renova-se a curiosidade sobre o desenvolvimento linguístico em idade escolar a partir da sintaxe, com McNamara *et al.* (2010), Beers e Nagy (2011), Berninger *et al.* (2011), Biber *et al.* (2011), entre outros, à frente dessa renovação.

Dos estudos passados e recentes, pode-se concluir que a presença de certos comportamentos gramaticais, como se pode descrever pela extensão, implica um controle crescente da língua. Hunt (1965, p. 5), examinando a escrita de crianças e adolescentes, associa esse tipo de comportamento à complexidade sintática, que se refere a um conjunto de características mensuráveis que são observadas com maior incidência em textos de escritores mais velhos e com fraca ou nenhuma ocorrência nos textos de escritores mais novos. Foster e Skehan (1996, p. 303) apontam para uma situação semelhante e salientam que o desenvolvimento gramatical do uso da língua consiste num processo crescente de elaboração, representado por uma grande variedade de padrões sintáticos.

Domínios da extensão

Uma das peculiaridades envolvidas no tratamento da extensão sintática diz respeito às unidades e às estruturas que fundamentam as técnicas de medição. Wolfe-Quintero *et al.* (1998, p. 274) observam que os indicadores considerados na generalidade dos estudos sobre o desenvolvimento linguístico dividem as suas preferências entre a oração e a unidade-t, as

quais tanto podem figurar isoladamente, sendo o foco a sua constituição interna, como em combinação uma com a outra. A oração, entendida como unidade sintática, é definida por Raposo *et al.* (2013, p. 314) como “uma sequência de palavras gramatical, que tem um conteúdo proposicional e como elemento nuclear um verbo”, definição que aqui se adota. A consideração da unidade *oração* tem como principal vantagem, enquanto unidade de análise, a fácil identificação na superfície textual escrita, além de ser uma unidade fiável tanto para uma análise de base sintática, como para uma análise de base semântica ou discursiva.

A unidade-t, proposta por Hunt (1965, p. 21), é a unidade constituída por uma oração independente e, quando é o caso, pela(s) oração(ões) que a ela se subordina(m). Se a unidade-t se constitui unicamente por uma oração independente, trata-se de uma unidade-t simples; ao contrário, se há uma ou mais orações subordinadas, é o caso de uma unidade-t multioracional. Portanto, a unidade-t é uma unidade autônoma das unidades que a antecedem ou se lhe seguem na superfície textual. Biber *et al.* (2007, p. 1069-1070), sobre a unidade-c (*c-unit*), proposta por Loban (1976), afirma que se configura como a menor unidade discursiva para fins de segmentação textual. Por analogia, pode-se dizer o mesmo da unidade-t, já que essas duas unidades se distinguem apenas pelo objeto a que se aplicam: aquela para a fala e esta para a escrita. A escolha pela unidade-t é particularmente recorrente em estudos sobre a escrita escolar porque, segundo Hunt (1965, p. 20-21), permite a um analista contornar os inconvenientes que a pontuação de aprendizes lhe impõe.

Wolfe-Quintero *et al.* (1998, p. 274) referem, em paralelo com a unidade-t e a oração, a frase como uma unidade possível para analisar a extensão sintática. No entanto, acreditamos que incluí-la como unidade de análise de textos escritos escolares pode não ser coerente com a unidade-t, que, afinal, lhe é uma alternativa. Além disso, como lembra Linell (2005, p. 169), o termo *frase* é regularmente usado de modos distintos, por vezes com um sentido vago ou ambivalente, confundindo-se com o termo *oração*.

Técnicas de medição

Os procedimentos de medição da extensão fundamentam-se na frequência de uma dada unidade dentro de outra, que lhe é superior, portanto. Tipicamente, obtém-se a extensão pela frequência média de palavras por enunciado (frase), o que se conhece pelas siglas EME-p (em português, extensão média do enunciado em palavras) ou MLU-w (em inglês *mean length of utterance in words*). Essa é uma adaptação de uma medida proposta

por Brown (1973), com que, originalmente, se pode mensurar a língua falada de crianças pela extensão média de enunciados em morfemas. À frequência média de palavras por frase (EME-p) há alternativas de cálculos, como, por exemplo, a frequência média de palavras por unidade-t e a frequência média de orações subordinadas por unidade-t, como se vê proposto em Hunt (1965). Pode ainda ser uma medida a frequência média de palavras por oração, como defendido por Berman (2007). Também, Berman e Ravid (2009) referem que a complexidade sintática se pode obter alternativamente pela frequência média de orações num texto. McNamara *et al.* (2010), com base na plataforma Coh-Metrix, avaliam a complexidade sintática de um texto com base num conjunto de propriedades acumuláveis (e extraíveis automaticamente), sendo a mais relevante delas a média de palavras à esquerda do verbo principal de uma oração. Ainda, Sun e Nippold (2012) dividem a somatória de todas as orações (principais e subordinadas) pelo total de unidades-t num texto, medida que chamam de densidade oracional.

Por serem medidas amplamente referidas na literatura como fiáveis para a avaliação da extensão do desenvolvimento sintático (BIBER *et al.*, 2011), neste estudo, seguimos os passos de Hunt (1965, 1970) e Berman (2007) e utilizamos a frequência média de palavras por oração, a frequência média de palavras por unidade-t e a frequência média de orações subordinadas por unidade-t. O seu emprego justifica-se também pela possibilidade de se poder comparar os resultados aqui encontrados com resultados de estudos anteriores, mesmo que baseados noutras línguas, de que se pode citar Verhoeven *et al.* (2001), Berman e Verhoeven (2002), Nippold *et al.* (2005) ou Beers e Nagy (2011).

Abaixo, encontra-se um fragmento de um texto argumentativo pertencente ao *corpus* em estudo, segmentado em unidades-t (numeradas) e em orações subordinadas (ladeadas por parênteses retos), para ilustrar a aplicação das medidas acima referidas:

- (1) Eu acho [que hoje em dia é importante [ter computador mas não as redes sociais]].³
- (2) As redes sociais são boas [para conhecer pessoas].
- (3) Mas também começou a haver mais raptos e mortes.

³ É importante referir que, neste estudo, não estão a ser consideradas as estruturas de elipse, razão por que, na análise da unidade-t 1, a estrutura “mas não as redes sociais”, em que parece haver a elipse de *são importantes* (“mas não ~~são importantes~~ as redes sociais”), está incluída na oração anterior. Sem desmerecer a importância do recurso à elipse, a sua identificação (manual) numa quantidade considerável de textos, em paralelo com todos os outros itens de investigação que aqui se utilizam, implica um esforço que ultrapassa em muito as possibilidades definidas aqui, para além da subjetividade envolvida neste tipo de análise.

(4) Eu sou a favor e contra [porque há cada vez mais pessoas a [optarem por [sair da escola e de casa]]].

(5) Mas é bom ter computador [porque os professores mandam trabalhos [para fazer no computador]].

(6) E ficam melhores e apresentáveis.

Como está explícito no fragmento acima, trata-se de um texto com seis unidades-t, das quais duas são simples (3 e 6) – se (como aqui fazemos, por simplicidade) não assumirmos uma análise bioracional das estruturas com verbos auxiliares aspectuais, como as de (3) (*começou a haver...*) – e quatro são multioracionais (1, 2, 4 e 5), com catorze orações (seis independentes e oito subordinadas) e com 71 palavras. Extrai-se disto que há, em média, 5,07 palavras por oração e 11,83 palavras por unidade-t e que a frequência média de orações subordinadas por unidade-t é de 1,33.

É importante ressaltar que o procedimento com base na extensão não deve ser entendido como um mero contar de unidades. Como afirmam Beers e Nagy (2009, p. 186), adotar esse tipo de procedimento para a análise de um texto ajuda a revelar o aspeto qualitativo da complexidade sintático-lexical. Esses autores fundamentam-se na lógica de que uma proporção elevada de palavras por oração, e por analogia por unidade-t, implica um maior encapsulamento de informações. Asseguram ainda que essas frequências refletem a interface entre estruturação sintática e letramento linguístico, representada no emprego de nominalizações, adjetivos atributivos ou sintagmas preposicionais. A interface ‘léxico-sintaxe’ explicita-se bem nas palavras de Ravid (2004, p. 343):

Longer clauses contain more words in general, and more content words in particular, and so they are lexically denser than shorter clauses. Besides, more words per clause translate into larger phrases (NPs, VPs, and PPs) in the clause, and this in turn is indicative of a more complex and diverse clause-internal syntactic architecture.

Hipóteses de trabalho

Estabelecemos como primeira hipótese de trabalho que a extensão sintática se correlaciona positivamente com a progressão nos anos escolares em ambos os registros, ou seja, conforme o aluno progride na escolaridade, as orações e as unidades-t que produzem tornam-se mais robustas em termos de constituição vocabular e as unidades-t tornam-se mais robustas em termos de constituição oracional.

Dada a natureza dos textos do registro argumentativo, que são tipicamente interpretados como mais exigentes ao escritor, tanto cognitiva como linguisticamente (NIPPOLD, 2004, p. 12), tem-se por hipótese secundária que, quanto à extensão das orações e das unidades-t, se notam mudanças mais significativas nos textos de registro narrativo do que nos textos de registro argumentativo. Esta hipótese contraria as assunções de Ravid (2004, p. 351), segundo a qual os textos argumentativos são mais ricos e mais complexos na sua expressão linguística do que os textos narrativos, sendo uma consequência dessa complexidade, como defendem Beers e Nagy (2011, p. 188), a identificação das diferenças ao nível sintático como mais saliente naquele tipo de textos do que neste. Trata-se de uma hipótese que se justifica numa percepção empírica de que a prática escolar, confirmada pelo Programa de Português do Ensino Básico (PORTUGAL, 2009), conduz os alunos a tornarem-se, ano a ano, mais familiarizados com os textos do registro narrativo, como se pode ver pela inserção desse tipo de textos desde os anos iniciais do primeiro ciclo da escolaridade e pela sua manutenção como uma necessidade de aprendizagem até, pelo menos, o 7.º ano, onde tais textos figuram para corroborar as “práticas de relato e reconto de experiências, de acontecimentos, de filmes vistos ou de livros lidos” (PORTUGAL, 2009, p. 173), enquanto conteúdo formalmente proposto para o ensino das habilidades de escrita, aparecendo os textos argumentativos como uma preocupação do ensino da escrita apenas a partir do segundo ciclo da escolaridade.

As hipóteses acima colocadas têm a sua origem em orientações e práticas pedagógicas que comumente defendem que a aprendizagem do português, no plano sintático, deve ser organizada pela noção de complexidade crescente, como se pode ver no Programa de Português do Ensino Básico, em que se considera, por exemplo, a mudança da “frase simples à frase complexa” por meio da expansão do grupo nominal (2009, p. 33) ou a “articulação de frases complexas” (2009, p. 95) como estratégias para assegurar o domínio quer da expressão oral, quer do conhecimento explícito da língua. Assume-se como expectável que o uso da língua em produções textuais efetivas espelhe tais aquisições, sendo incorporadas gradualmente conforme a criança progride nos anos escolares, o que se confirma por estudos que asseguram que crianças e adolescentes nos anos finais da escolaridade produzem textos escritos com mais palavras e, interligadamente, com mais orações, e mais variadas quanto à funcionalidade semântica, do que os alunos dos anos iniciais (MALVERN *et al.*, 2004; RAVID, 2004; NIPPOLD *et al.*, 2005; NIPPOLD *et al.*, 2007; BEERS; NAGY, 2011).

Corpus e métodos

Para o exame da correlação entre a extensão sintática e a progressão escolar, utilizou-se um *corpus* de 244 textos produzidos por alunos monolíngues de português europeu regularmente matriculados no 5.º, no 7.º e no 10.º ano escolar de quatro escolas públicas localizadas no distrito de Lisboa. Cada aluno, a partir dos estímulos propostos⁴, produziu um texto narrativo e um argumentativo, resultando num *corpus* com 244 textos (122 narrativos e 122 argumentativos), com 43.685 palavras, tendo sido todos os textos manualmente segmentados em unidades-t e em orações no programa informático UAMCorpusTool (O'DONNELL, 2008), de onde se extraiu automaticamente a informação sobre os quantitativos de palavras.

Para identificar os graus de relação entre variáveis independentes e dependentes e determinar as diferenças e semelhanças entre os grupos de escolaridade e os registos, foram aplicadas, com recurso ao programa SPSS, técnicas da estatística descritiva e inferencial. Em virtude de os dados obtidos a partir da segmentação e anotação não obedecerem a uma distribuição normal, nem serem passíveis de normalização, pela grande quantidade de valores aberrantes, tornou-se necessária a utilização de testes não paramétricos. Assim, para a identificação da magnitude de correlação entre as variáveis, aplicou-se o coeficiente de correlação de Spearman, sendo corroborado, para detectar a existência de variância significativa na amostra, pelo teste de Kruskal-Wallis (análogo ao teste F da ANOVA a um fator), de que também se extraem a força do tamanho do efeito e os valores de probabilidade de diferenças entre os grupos.

Resultados e discussão

Na tabela a seguir, apresentam-se, diferenciados por registo (narrativo e argumentativo) e por ano escolar (5.º, 7.º e 10.º), os números médios de ocorrência de orações totais (*Orc*) – incluindo-se orações simples, subordinantes e subordinadas –, de orações subordinadas (*Sub*), de unidades-t (*Unt*) e de palavras (*Pal*), com os respectivos

⁴ Estímulo 1 (tarefa de escrita de um texto narrativo): Narra um facto marcante (real ou imaginado) que tu e teu(tua) melhor amigo(a) viveram durante o último verão.

Estímulo 2 (tarefa de escrita de um texto argumentativo): Achas que as redes sociais (Facebook, Twitter, Google+, Windows Live Space, etc.) são importantes hoje em dia? Escreve um texto para ser publicado no blogue da tua escola em que exponhas a tua opinião sobre as redes sociais. Neste texto deves dizer se és a favor ou contra a existência das redes sociais. Não te esqueças de justificar a tua opinião!

desvios-padrão, identificados nos textos sob estudo. São essas as unidades centrais para a medição da extensão da complexidade sintática.

Tabela 1 – Valores médios de ocorrência, por texto, de palavras (Pal), de orações (Orc), de orações subordinadas (Sub) e de unidades-t (Unt) identificados nos registros narrativo e argumentativo

| Registro | Ano | N | Pal (DP) | | Orc (DP) | | Sub (DP) | | Unt (DP) | |
|---------------|------|----|----------|-------|----------|-------|----------|------|----------|------|
| Narrativo | 5.º | 26 | 179,12 | 45,14 | 31,88 | 11,70 | 9,88 | 5,02 | 21,54 | 8,94 |
| | 7.º | 46 | 180,54 | 37,32 | 32,26 | 8,71 | 10,96 | 5,55 | 20,91 | 5,18 |
| | 10.º | 50 | 207,62 | 48,40 | 34,68 | 10,25 | 15,36 | 6,34 | 18,56 | 5,55 |
| Argumentativo | 5.º | 26 | 154,46 | 39,26 | 27,27 | 8,62 | 14,19 | 5,91 | 12,62 | 4,39 |
| | 7.º | 46 | 153,24 | 38,26 | 23,98 | 5,75 | 12,89 | 4,47 | 11,28 | 4,35 |
| | 10.º | 50 | 185,54 | 32,11 | 34,68 | 10,25 | 13,98 | 4,94 | 12,78 | 3,42 |

No registro narrativo, a quantidade média de palavras por texto é crescente de ano a ano. Também o são o número médio de ocorrências de orações e o número médio de ocorrências de orações subordinadas. Num movimento contrário, o número médio de unidades-t é decrescente do 5.º ao 7.º, e deste ao 10.º. Nos textos do registro argumentativo, o 7.º ano escolar é o ponto que anula a linearidade das ocorrências das quatro unidades, já que o número médio de ocorrências é sempre mais baixo neste ano do que nos outros, isto é, do 5.º ao 7.º, há um decréscimo nos valores médios, mas deste ao 10.º ano vê-se uma elevação.

Os valores acima descritos servem de base ao cálculo das três medidas de avaliação da extensão da complexidade sintática escolhidas para este estudo, que são a frequência média de palavras por oração (*Pal/Orc*), a frequência média de palavras por unidade-t (*Pal/Unt*) e a frequência média de orações subordinadas por unidade-t (*Sub/Unt*). O Gráfico 1 e o Gráfico 2, abaixo, apresentam o comportamento pareado destas frequências ao longo dos anos escolares no registro narrativo e no registro argumentativo, respectivamente:

Gráfico 1 – Frequência média de palavras por oração (*Pal/Orc*), de palavras por unidade-t (*Pal/Unt*) e de orações subordinadas por unidade-t (*Sub/Unt*) no registro narrativo ao longo dos anos escolares

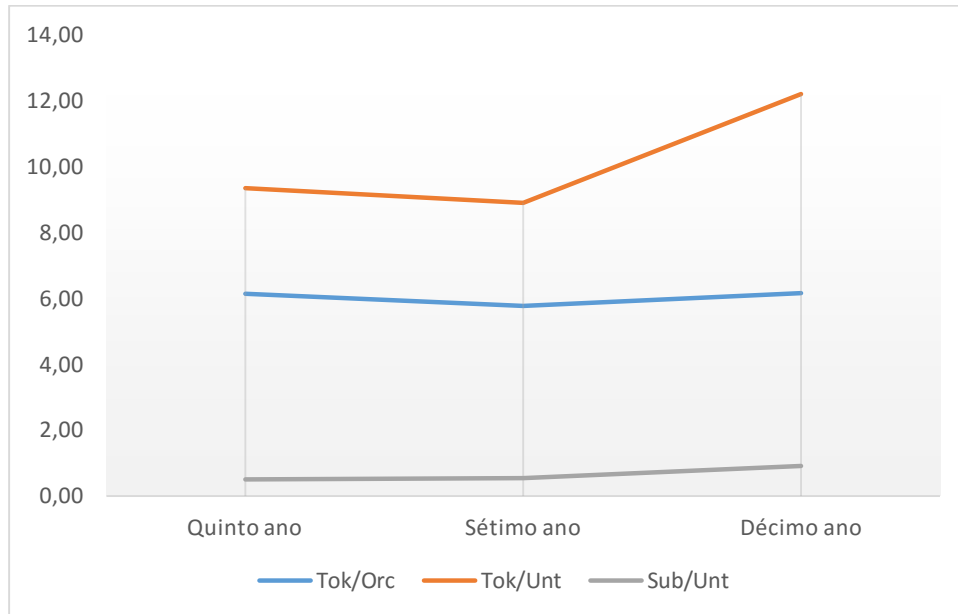
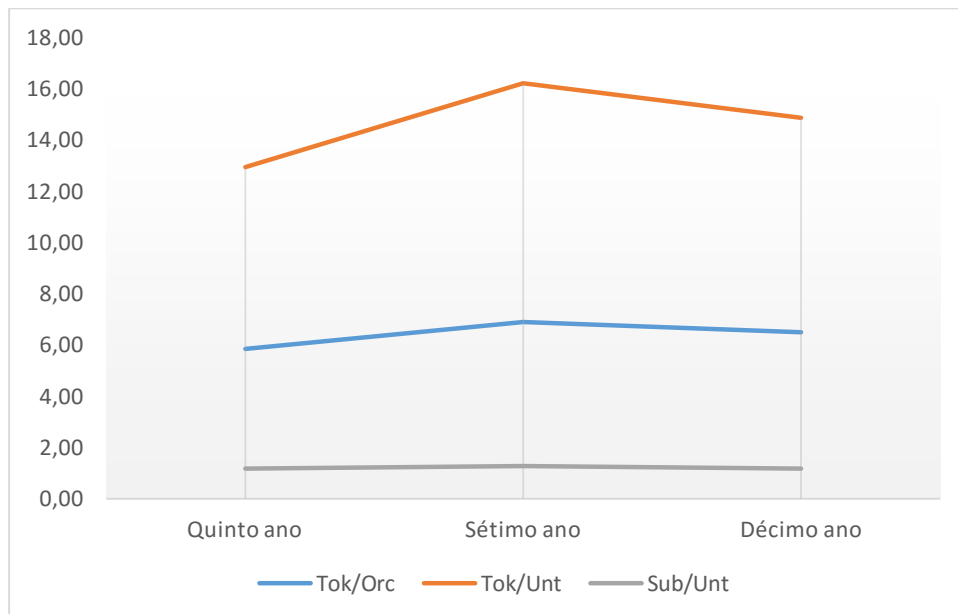


Gráfico 2 – Frequência média de palavras por oração (*Pal/Orc*), de palavras por unidade-t (*Pal/Unt*) e de orações subordinadas por unidade-t (*Sub/Unt*) no registro argumentativo ao longo dos anos escolares



Quanto à frequência média de palavras por oração, percebe-se que os valores médios, no registro narrativo (Gráfico 1), diminuem do 5.º para o 7.º ano, mas aumentam deste ano para o 10.º. No 5.º ano, a frequência média de palavras por oração é de 6,15 (2,05⁵); no 7.º ano, o valor médio é de 5,77 (0,94); no 10.º ano, o valor médio é de 6,16 (1,00). Quanto ao registro argumentativo (Gráfico 2), o movimento é precisamente o oposto, havendo uma ligeira subida do 5.º ao 7.º ano. Deste ano para o 10.º, pelo contrário, ocorre uma diminuição, também ligeira. No 5.º ano, a frequência média de palavras por oração é 5,84 (1,10); no 7.º ano, o valor médio é de 6,89 (1,29); no 10.º ano, o valor médio é de 6,54 (1,20).

A frequência média de palavras por unidade-t reflete um cenário bastante diferenciado da frequência anterior. Quanto ao registro narrativo, o Gráfico 1 sugere que há alguma diminuição nos valores do 5.º para o 7.º ano; porém, verifica-se do 7.º para o 10.º ano uma subida bastante acentuada no uso. A frequência média de palavras por unidade-t, no 5.º ano, é de 9,09 (2,98); no 7.º ano, a frequência média é de 9,09 (2,05); no 10.º ano, o valor médio é de 12,12 (5,96). No registro argumentativo (Gráfico 2), a frequência média de palavras por unidade-t também se mostra diferenciada da frequência de palavras por orações no mesmo registro. Do 5.º ao 7.º ano, há um aumento acentuado na frequência de palavras por unidade-t, seguido de uma diminuição também acentuada deste para o 10.º ano. No 5.º ano, a frequência média é de 12,94 (3,08); no 7.º ano, é de 16,21 (6,29); e, no 10.º ano, é de 14,86 (5,27).

O Gráfico 1 mostra, quanto à frequência média de orações subordinadas por unidade-t no registro narrativo, que há um aumento gradativo nos valores do 5.º ao 10.º ano, mesmo que aparentemente tênue. No 5.º ano, a frequência média de orações subordinadas por unidade-t é de 0,51 (0,27); no 7.º ano, o valor médio passa a 0,54 (0,28); no 10.º ano, o valor médio é de 0,92 (0,61). No registro argumentativo, a partir do Gráfico 2, infere-se que os valores se mantêm quase inalterados de ano a ano, havendo uma subida pouco expressiva dos valores no 7.º ano, quando comparados com os dos outros anos escolares. No 5.º ano, o valor médio de orações subordinadas por unidade-t é de 1,18 (0,46); no 7.º ano, a frequência sobe para 1,27 (0,68); e, no 10.º ano, cai para 1,16 (0,45).

Os testes de correlação aplicados às três medidas corroboram a leitura que se pode fazer dos valores médios apresentados no Gráfico 1 e no Gráfico 2. Para a relação entre a frequência de palavras por oração e a progressão escolar não se identificou correlação

⁵ Desvio-padrão.

matematicamente significativa nem no registro narrativo ($r=0,176^6$; $p=0,052^7$), nem no registro argumentativo ($r=0,154$; $p=0,090^8$). Mesmo sem correlação, é importante referir que, pelo teste de variância, foram notadas, nas médias do registro narrativo, diferenças significativas entre o 7.º e o 10.º ano ($p=0,029$), não havendo diferenças nem entre o 5.º e o 7.º ano ($p=0,930$), nem entre o 5.º e o 10.º ano ($p=0,156$). Quanto às médias dos textos argumentativos, as diferenças localizam-se entre o 5.º e o 7.º ano ($p=0,000$) e entre o 5.º e o 10.º ano ($p=0,004$), não havendo diferenças entre o 7.º e o 10.º ano ($p=0,080$).

Para a frequência de palavras por unidade-t no registro narrativo, identificou-se uma correlação positiva moderada ($r=0,433$; $p=0,000$). Quanto a isto, na análise da variância pelo teste de Kruskal-Wallis, foram detectadas diferenças significativas entre as médias dos três grupos tomados conjuntamente ($H(2)=26,968$; $p=0,000^9$), com os seguintes valores de *rank* – no 5.º, 47,46; no 7.º, 47,75; no 10.º, 81,45 – e com um tamanho do efeito de 0,22. No entanto, confrontados os valores de *rank* de cada grupo, não se detectou diferença entre o 5.º e o 7.º ano ($p=0,699$), diferença que se manifesta apenas entre o 5.º e o 10.º ano ($p=0,000$) e entre o 7.º e o 10.º ano ($p=0,000$). Quanto à frequência de palavras por unidade-t no registro argumentativo e à progressão escolar, não foi identificada nenhuma correlação significativa ($r=0,077$; $p=0,402$).

Por fim, os testes estatísticos apontam para uma correlação positiva moderada entre a frequência média de orações subordinadas por unidade-t nos textos do registro narrativo e a progressão escolar ($r=0,446$; $p=0,000$), sendo tal correlação confirmada pela existência de diferenças significativas entre as médias dos anos escolares ($H(2)=26,918$; $p=0,000$), com os seguintes valores de *rank* – no 5.º ano, 45,35; no 7.º ano, 49,04; no 10.º ano, 81,36 – e com um tamanho do efeito de 0,22. É de referir, no entanto, que entre o 5.º e o 7.º ano, não foram identificadas diferenças significativas ($p=0,606$), que se verificam apenas entre o 7.º e o 10.º ano ($p=0,000$) e entre o 5.º e o 10.º ano ($p=0,000$). Quanto aos textos argumentativos, não foi identificada nenhuma correlação entre a frequência média de orações subordinadas por unidade-t e a progressão nos anos escolares ($r=-0,009$; $p=0,920$).

Os resultados dos testes estatísticos indicam, de um modo geral, que a extensão das unidades linguísticas selecionadas – orações e unidades-t – não se conforma a um mesmo

⁶ Para a interpretação da magnitude das correlações em Spearman, utiliza-se a classificação dos coeficientes de correlação proposta por Hulley *et al.* (2003): $< 0,4$ (correlação fraca), $\geq 0,4$ a $\leq 0,5$ (correlação moderada) e $> 0,5$ (correlação forte).

⁷ Diferença significativa no nível $< 0,01$.

⁸ Correlação significativa no nível $< 0,01$ (2 extremidades).

⁹ Diferença significativa no nível $< 0,05$.

processo uniforme e linear de desenvolvimento ao longo dos anos escolares. Por um lado, a extensão da oração, medida pela frequência média de palavras, não mantém correlação com a progressão escolar em nenhum dos dois registros avaliados e, por outro, a unidade-t, quer aferida pela frequência média de palavras, quer pela frequência média de orações subordinadas, só tem representação significativa nos textos do registro narrativo, não sendo detectado o mesmo para os textos argumentativos. Esses resultados confirmam parcialmente a primeira hipótese de trabalho, segundo a qual a extensão sintática se correlaciona positivamente com a progressão nos anos escolares em ambos os registros. No entanto, confirmam plenamente a segunda hipótese de trabalho, segundo a qual os textos narrativos apresentam mudanças na sua caracterização sintática de modo mais expressivo do que as identificadas nos textos de registro argumentativo.

Os resultados aqui encontrados contrariam uma convicção bastante difundida na literatura (e corroborada também pelos resultados de alguns estudos) sobre o desenvolvimento linguístico em idade escolar. Essa convicção pode-se sintetizar no que afirmam Berman e Ravid (2009, p. 96), segundo as quais, em diferentes línguas e tipos de textos, o número médio de palavras por oração cresce em função da idade/escolaridade e do registro, em particular do registro argumentativo. Corrobora-se tal afirmação no estudo de Johansson (2009), cuja pesquisa, centrada no desenvolvimento linguístico de crianças e adolescentes suecos do 4.º ano de escolaridade até ao primeiro ano da universidade, revela que a extensão média da oração contabilizada em palavras se correlaciona positivamente com o avanço nos anos escolares nos dois registros (narrativos e argumentativos), notando-se no registro argumentativo a mudança mais expressiva. Schleifer (2003), a partir da frequência média de palavras por oração em textos de crianças e adolescentes inscritos no 4.º, 7.º e 11.º anos de escolas israelenses, encontra diferenças significativas no desenvolvimento da extensão de orações apenas no registro argumentativo, não identificando o mesmo nos textos narrativos. Ravid (2004), ao comparar textos expositivos e narrativos escritos por alunos monolíngues falantes de hebreu inscritos também no 4.º, 7.º e 11.º anos de escolaridade, também detecta que as orações dos textos argumentativos, em termos de quantidade de palavras, tendem a ser mais longas do que as dos textos narrativos, mantendo-se uma frequência crescente ao longo dos anos escolares.

Embora não se tenha identificado nenhum tipo de correlação entre a frequência média de palavras por oração e a progressão escolar, não quer isto dizer que a extensão da oração produzida pelos alunos falantes nativos de português, no registro de texto argumentativo, se mantenha inalterada desde o ano inicial até o ano final. Como ficou

ilustrado (Gráfico 2), o 7.º ano apresenta uma média de palavras por oração superior à dos outros anos, o que também foi confirmado pelo teste de Kruskal-Wallis, que apontou para diferenças substanciais entre as médias identificadas no 5.º ano quando comparadas com as do 7.º ano, e também entre as médias comparadas do 5.º ano e do 10.º ano, não havendo, contudo, diferenças entre as médias identificadas no 7.º e no 10.º ano. Disto se pode concluir, portanto, que é no 7.º ano que há um ponto de ruptura da esperada linearidade ascendente quanto a esta medida, o que satisfaria a convicção do crescimento das unidades em função da escolaridade. Dito doutro modo, a ausência de correlação entre a frequência média de palavras por oração e a progressão escolar, nos textos estudados, parece ser amplamente afetada pela ausência de linearidade que se nota entre os grupos dos alunos estudados.

Quanto à extensão da unidade-t, os resultados aqui identificados apontam para mudanças significativas apenas no registro narrativo e, mais uma vez, opõem-se aos resultados de alguns estudos, de que se pode citar, por exemplo, Crowhurst e Piche (1979), que, analisando a qualidade da escrita de crianças norte-americanas do 4.º, 6.º e 9.º ano, concluem que, no registro argumentativo, há uma produção mais significativa de palavras por unidades-t do que no registro narrativo. Também Nippold *et al.* (2005), num estudo que analisa a produtividade sintática de crianças, adolescentes e adultos (respectivamente, 11, 17 e 24 anos), reportam um crescimento constante na extensão da unidade-t em textos argumentativos. Em Beers e Nagy (2009), que estudam textos narrativos e argumentativos de crianças no 7.º e no 10.º ano da escolaridade americana, identifica-se um crescimento contínuo de palavras por unidade-t, em particular mais forte nos textos argumentativos. Johansson (2009) reporta resultados semelhantes aos de Crowhurst e Piche (1979), Nippold *et al.* (2005) e Beers e Nagy (2005). Ainda quanto à extensão da unidade-t em palavras, é importante lembrar que os valores aqui identificados para a medição da extensão são matematicamente diferentes apenas entre o 5.º e o 10.º ano e entre o 7.º e o 10.º, não havendo diferenças significativas entre o 5.º e o 7.º ano.

Sobre a extensão da oração ou da unidade-t em palavras, não foram identificadas investigações focadas na escrita escolar em português de populações semelhantes à que aqui se analisa. Mais recorrentes são os trabalhos sobre a produção linguística de crianças na primeira infância, como fazem Pereira (2008), Rodrigues (2008) ou Mendes (2015). No entanto, duas investigações são notáveis. Carvalho (1989), num estudo transversal com setenta crianças entre os 7/8 e os 13/14 anos de idade, avalia a evolução da escrita nos registros narrativo e descritivo, a partir de uma perspectiva psicolinguística. Um dos

resultados aponta para “uma relativa estabilidade” da extensão do período e da frase nos textos narrativos (CARVALHO, 1989 *apud* CARVALHO, 2003, p. 35), sendo possível detectar mudanças apenas nos textos descritivos. Costa (2010), a fim de examinar o uso de conectores de valor concessivo em produções textuais dos alunos do 4.º, 6.º e 9.º anos da escolaridade portuguesa, emprega, como estratégia complementar, a medição da extensão com base na frequência média de palavras por enunciado (frase), observando que é possível notar um “crescimento entre os escritores mais jovens e os escritores adultos”, mas, semelhantemente ao que se identifica aqui quanto à extensão da oração e da unidade-t, há “uma clivagem entre os alunos mais novos, do 4.º ano e do 6.º ano, cujos textos não revelam aumento da extensão média de enunciados, e os alunos do 9.º ano” (COSTA, 2010, p. 136).

Por fim, como se comprovou com a análise estatística, é a terceira medida selecionada para avaliar a extensão da complexidade sintática – a frequência média de orações subordinadas por unidade-t – que se mostra um indicador de fortes mudanças ao longo da progressão escolar, porém, e de novo, apenas nos textos narrativos, contrariando a expectativa comum de que nos textos argumentativos o índice de ocorrência de orações por unidade-t seria mais expressivo nos textos argumentativos, como se pode ver, por exemplo, em Johansson (2009) ou em Beers e Nagy (2009). É de notar, porém, que a extensão, em termos de uso de orações subordinadas por unidades-t, não se confirma entre o 5.º e o 7.º ano da escolaridade dos grupos investigados.

Quanto ao registro como função da progressão escolar, os estudos acima referidos para efeitos de comparação ecoam, direta ou indiretamente, a seguinte afirmação de Crowhurst e Piche (1979 *apud* BEERS; NAGY, 2011, p. 188) para justificar os resultados mais expressivos centrados no registro argumentativo: “*narration places fewest demands and argument greatest demands on writers to make use of their syntactic resources*”. Por uma lógica reversa, especulamos, no entanto, que o fato de os textos narrativos serem pouco exigentes aos escritores pode servir como um possível argumento para justificar os resultados aqui identificados, que, como se viu, não seguem na mesma direção dos resultados dos estudos citados. Desse modo, é possível especular que, dada a contínua e sistemática exposição aos textos narrativos que se vê ao longo da progressão escolar, por um lado e, por outro, uma inclusão do registro argumentativo que se pode chamar de mais cautelosa nas práticas de sala de aula, corroborada pelas proposições do Programa de Português do Ensino Básico (PORTUGAL, 2009), a instrumentalização dos recursos sintáticos para a produção de orações e unidades-t que as crianças e os adolescentes

possuem parece ser bem menos acionada nos textos argumentativos do que nos textos narrativos.

Considerações finais

Reportou-se, neste artigo, a um estudo de correlação estatística entre a extensão sintática e a progressão nos anos escolares em dois registros (narrativo e argumentativo) produzidos por crianças e adolescentes monolíngues de português europeu. A extensão descreveu-se pela frequência média de palavras por oração, pela frequência média de palavras por unidade-t e pela frequência média de orações subordinadas por unidade-t. Quanto à frequência média de palavras por oração, a aplicação dos testes estatísticos revelou que esse indicador de extensão sintática não mantém correlação com a progressão ao longo dos anos escolares em nenhum dos registros analisados. Ainda quanto à frequência média de palavras, mas desta vez limitada pelos domínios da unidade-t, verificou-se que essa medida de complexidade da extensão sintática mantém uma correlação positiva com a progressão escolar apenas no registro narrativo, não sendo identificado nenhum tipo de correlação no registro argumentativo. Por fim, a aplicação de testes estatísticos revelou que a terceira medida de avaliação da extensão sintática, que é a frequência média de orações subordinadas por unidade-t, mantém uma correlação positiva com a progressão escolar, mas unicamente nos textos de registro narrativo, não havendo correlação com os textos de registro argumentativo. Ao comparar-se esse resultado com o da medida anterior, infere-se que, nos textos narrativos, a quantidade de orações é um indicador de mudanças da extensão sintática realizada nos textos dos alunos estudados.

Como lembra Berman (2007, p. 353), apesar de os mecanismos de combinação de orações serem percebidos desde cedo na escrita de crianças, é apenas um aprendiz em estágios mais avançados, detentor de outros conhecimentos linguísticos, como, por exemplo, um léxico mais expandido, que é capaz de utilizá-los com mais propriedade. Ainda segundo essa autora (2007, p. 353), tais construções são a marca registrada de um estilo mais acadêmico, mais letrado, que só tem lugar a partir do ensino secundário. Essa capacidade de empregar mais orações subordinadas – aqui delimitadas pelos domínios da unidade-t –, revela-se, portanto, uma característica-chave do desenvolvimento linguístico em idade escolar, e em particular do desenvolvimento sintático. Tal capacidade deve ser entendida, na defesa que Slobin faz (1985 *apud* BERMAN, 2007, p. 354), como parte de um processo de reorganização funcional dos usos linguísticos, que se circunscreve pela

atribuição de novas funções a estruturas já adquiridas e, ao mesmo tempo, pela utilização de novas formas para satisfazer funções já dominadas. Essa concepção permite explicar o emprego crescente da subordinação que se nota na construção de complexos oracionais impressos nos textos dos alunos em estágios mais avançados.

MARTINS, Mário; MARQUES, Rui. Syntactic length in school written texts. **Revista do GEL**, v. 13, n. 1, p. 176-194, 2016.

Abstract: This article presents a correlational study between syntactic length and school progression in texts written by school age children and adolescents, monolingual speakers of European Portuguese. The measures used to assess syntactic length are the average frequency of words per clause (BERMAN, 2007), the average frequency of words per t-unit and the average frequency of subordinate clauses per t-unit (HUNT, 1965, 1970). In UAM CorpusTool platform (O'DONNELL, 2008), these measures were applied to a quasi-longitudinal corpus consisting of 244 texts of narrative (n = 122) and argumentative (n = 122) texts written by students in the 5th (n = 26), the 7th (n = 46) and 10th (n = 50) year of the Portuguese basic schooling system. The results show that, on one hand, clause length, measured by average frequency words, does not maintain correlation with school progression in both registers assessed and, on the other hand, the t-unit length, whether measured by the average frequency words or by the average frequency of subordinate clauses, is statistically significant only in narrative texts, what does not happen to argumentative texts. This work aims to contribute to a more detailed understanding of the settings of the syntactic development by school age children and adolescents' written language.

Keywords: Syntactic length. School progression. School texts.

Referências

BEERS, S. F.; NAGY, W. E. Syntactic complexity as a predictor of adolescent writing quality: Which measures? Which genre? **Reading and writing**, v. 22, n. 1, p. 185-200, 2009.

_____. Writing development in four genres from grades three to seven: Syntactic complexity and genre differentiation. **Reading and writing**, v. 24, n. 1, p. 183-202, 2011.

BERMAN, R. A. Developing linguistic knowledge and language use across adolescence. In: HOFF, E.; SHATZ, M. (Ed.). **Blackwell handbook of language development**. Malden/Oxford/Victoria: Blackwell, 2007. p. 347-367.

BERMAN, R. A.; RAVID, D. Becoming a literate language user: Oral and written text construction across adolescence. In: OLSON, D. O.; TORRANCE, N. (Ed.) **Cambridge handbook of literacy**. Cambridge: Cambridge University Press, 2009. p. 92-111.

BERMAN, R. A.; VERHOEVEN, L. Crosslinguistic perspectives on the development of text-production abilities: Speech and writing. **Written Language and Literacy**, v. 5, n. 1, p. 1-44, 2002.

BERNINGER, V. W.; NAGY, W.; BEERS, S. Child writers' construction and reconstruction of single sentences and construction of multi-sentence texts: Contributions of syntax and transcription to translation. **Reading and writing**, v. 24, n. 1, p. 151-182, 2011.

BIBER, D.; GRAY, B.; POONPON, K. Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? **TESOL Quarterly**, v. 45, n. 1, p. 5-35, 2011.

BIBER, D. *et al.* **Longman grammar of spoken and written english**. Londres: Longman, 2007.

BROWN, R. **A first language: The early stages**. Cambridge: Harvard University Press, 1973.

CARVALHO, J. A. B. S. **Escrita: Percursos de investigação**. Minho: DME/UM, 2003.

COSTA, A. L. **Estruturas contrastivas: Desenvolvimento do conhecimento explícito e da competência de escrita**. 2010. 306 f. Tese (Doutorado em Linguística Educacional) – Universidade de Lisboa, Lisboa, 2010.

CROWHURST, M.; PICHE, G. L. Audience and mode of discourse effects on syntactic complexity in writing at two grade levels. **Research in the Teaching of English**, v. 13, n. 2, p. 101-109, 1979.

FOSTER, P.; SKEHAN, P. The influence of planning and task type on second language performance. **Studies in second language acquisition**, v. 18, n. 1, p. 299-323, 1996.

HARPIN, W. S. **Social and educational influences on children's acquisition of grammar**. Swindon: Social Science Research Council, 1973.

_____. **The second 'r': Writing development in the junior school**. Londres: Allen and Unwin, 1976.

HULLEY, S. B. *et al.* **Delineando a pesquisa clínica: Uma abordagem epidemiológica**. Porto Alegre: Artmed, 2003.

HUNT, K. W. **Grammatical structures written at three grade levels**. Champaign: National Council of Teachers of English, 1965.

_____. **Syntactic maturity in schoolchildren and adults**. Chicago: Chicago University Press, 1970.

JOHANSSON, V. **Developmental aspects of text production in writing and speech**. Lund: Lund University, 2009.

LINELL, P. **The written language bias in linguistics: Its nature, origins, and transformation**. Londres: Routledge, 2005.

LOBAN, W. D. **The language of elementary school children**: A study of the use and control of language and the relations among speaking, reading, writing, and listening. Champaign: NCTE, 1963.

_____. **Language development**: Kindergarten through grade twelve. Urbana: NCTE, 1976.

MALVERN, D. *et al.* **Lexical diversity and language development**: Quantification and assessment. Nova Iorque: Palgrave Macmillan, 2004.

MCNAMARA, D. S.; CROSSLEY, S. A.; MCCARTHY, P. M. Linguistic features of writing quality. **Written Communication**, v. 27, n. 1, p. 57-86, 2010.

MENDES, M. **Para um modelo ontogenético da escrita infantil no 1.º ciclo do ensino básico**: A emergência do período como unidade de organização textual. 2015. 458 f. Tese (Doutorado em Linguística Aplicada) – Universidade de Lisboa, Lisboa, 2015.

NIPPOLD, M. A. Research on later language development. In: BERMAN, R. A. (Ed.). **Language development across childhood and adolescence**. Amsterdam/Filadélfia: John Benjamins, 2004. p. 1-8.

NIPPOLD, M. A.; HESKETH, L. J.; DUTHIE, J. K. Conversational versus expository discourse: A study of syntactic development in children, adolescents, and adults. **Journal of Speech, Language, and Hearing Research**, v. 48, n. 5, p. 1048-1064, 2005.

NIPPOLD, M. A.; MANSFIELD, T. C.; BILLOW, J. L. Peer conflict explanations in children, adolescents, and adults: Examining the development of complex syntax. **American Journal of Speech-Language Pathology**, v. 16, n. 2, p. 179-188, 2007.

O'DONNELL, M. The uam corpustool: Software for corpus annotation and exploration. In: BRETONES CALLEJAS, C. M.; SALABERRI, R. S.; SANCHES, E. G. (Ed.) **Applied linguistics now: Understanding language and mind / La lingüística aplicada hoy: Comprendiendo el lenguaje y la mente**. Almería: Universidad de Almería, 2008. p. 1433-1447.

O'DONNELL, R. C.; GRIFFIN, W. J.; NORRIS, R. C. **Syntax of kindergarten and elementary school children**: A transformational analysis. Champaign: NCTE, 1967.

PEREIRA, I. S. P. **Para a caracterização do contexto de ensino – aprendizagem da literacia no 1.º ciclo de escolaridade**: Das competências dos alunos às concepções e práticas dos professores. 2008. 514 f. Tese (Doutorado em Linguística Aplicada) – Universidade do Minho, Braga, 2008.

PERERA, K. **Children's writing and reading**: Analysing classroom language. Londres: Blackwell, 1984.

PORTUGAL. **Programa de português do ensino básico**. Lisboa: Ministério da Educação e Ciência. Direcção-Geral da Educação, 2009.

RAPOSO, E. B. P. Estrutura da frase. In: RAPOSO, E. B. P. *et al.* (Ed.). **Gramática do português**. Lisboa: Fundação Calouste Gulbenkian, 2013. p. 303-400.

RAVID, D. Emergence of linguistic complexity in later language development: Evidence from expository text construction. In: RAVID, D. D.; SHYLDKROT, H. B. (Ed.). **Perspectives on language and language development**: Essays in honor of Ruth A. Berman. Dordrecht/Boston/Londres: Kluwer Academic Publishers, 2004.

RODRIGUES, S. B. P. **Escrita espontânea**: Desenvolvimento das capacidades de composição escrita em crianças do 1º ao 4º ano de escolaridade. 2008. 133 f. Dissertação (Mestrado em Linguística Aplicada) – Universidade Fernando Pessoa, Porto, 2008.

SCHLEIFER, M. **Development of written text production of native israeli and ethiopian immigrant school children and adolescent**. 2003. 166 f. Tese (Doutorado em Linguística Aplicada) – Tel-Aviv University, Tel-Aviv, 2003.

SCHLEPPEGRELL, M. J. **The language of schooling**: A functional linguistics perspective. Mahwah/Londres: Lawrence Erlbaum, 2004.

SCOTT, C. M. Syntactic contributions to literacy learning. In: STONE, C. A. *et al.* (Ed.). **Handbook of language and literacy**: Development and disorders. Nova Iorque: The Guilford Press, 2004. p. 340-362.

SUN, L.; NIPPOLD, M. A. Narrative writing in children and adolescents: Examining the literate lexicon. **American Speech-Language-Hearing Association**, v. 43, n. 2, p. 2-13, 2012.

VERHOEVEN, L.; STROMQVIST, S. Development of narrative production in a multilingual context. In: VERHOVEN, L.; STROMQVIST, S. (Ed.). **Narrative development in a multilingual context**. Amsterdam/Filadélfia: John Benjamins, 2001. p. 1-14.

WOLFE-QUINTERO, K.; INAGAKI, S.; KIM, H. Y. **Second language development in writing**: Measures of fluency, accuracy, and complexity. Honolulu: University of Hawai'i at Mānoa, 1998.

Submetido em 30/10/2015

Aceito em 05/05/2016