

DEPENDÊNCIA: O CONCEITO E AS GRAMÁTICAS DEPENDENCE: THE CONCEPT AND GRAMMARS

Isaac Souza de Miranda Junior¹ and Oto Araújo Vale¹

¹Affiliation not available

November 21, 2022

Abstract

Dependency Grammar has recently gained relevance with the increase of works using its methodology to perform automatic syntactic analysis. They are very productive in terms of syntactic disambiguation, considering the binary arrangement of the trees, and they also have been shown to be much more efficient for automatic parsing (NIVRE, 2005, p.1). With this new demand, Dependency Grammar has become a relevant term in Natural Language Processing works (KÜBLER et al., 2009; DE MARNEFFE and NIVRE; 2019). However, studies related to Dependency Grammars were marginalized compared to the traditional methodology of syntactic analysis, the constituents analysis, which made it difficult to access the theoretical foundation that this type of methodology follows. Considering this, this work intends to make an introductory review of the fundamentals that govern most of the syntactic theories that use the dependency methodology, to make the Dependency Grammar models, as well as the concept of dependency, clearer and more accessible.

DEPENDÊNCIA: O CONCEITO E AS GRAMÁTICAS

DEPENDENCE: THE CONCEPT AND GRAMMARS

Tipo de Contribuição

Revisão de Literatura.

Isaac Souza de Miranda Junior*

Oto Araújo Vale[†]

RESUMO: As Gramáticas de Dependência recentemente adquiriram relevância com o aumento de trabalhos utilizando sua metodologia para a realização de análise sintática automática. Elas são muito produtivas com relação a desambiguação sintática, considerando a disposição binária das árvores, como também têm se demonstrado muito mais eficientes para a realização de parsing automático (NIVRE, 2005, p.1). Com essa nova demanda, *Gramática de Dependência* tornou-se um termo muito relevante dentro dos trabalhos de Processamento de Língua Natural (KÜBLER *et al.*, 2009; DE MARNEFFE e NIVRE; 2019). Entretanto, os estudos relacionados às Gramáticas de Dependência foram marginalizados quando comparados com a metodologia tradicional de análise sintática, a análise de constituintes, o que acabou por dificultar o acesso à fundamentação teórica que esse tipo de metodologia segue. Considerando isso, este trabalho pretende realizar revisão introdutória acerca dos fundamentos que regem a maior parte das teorias sintáticas que se utilizam da metodologia de dependências, a fim de tornar os modelos de Gramáticas de Dependência, assim como o conceito de dependência, mais claros e acessíveis.

PALAVRAS-CHAVE: Linguística Descritiva; Sintaxe; Gramática de Dependência.

ABSTRACT: Dependency Grammar has recently gained relevance with the increase of works using its methodology to perform automatic syntactic analysis. They are very productive in terms of syntactic disambiguation, considering the binary arrangement of the trees, and they also have been shown to be much more efficient for automatic parsing (NIVRE, 2005, p.1). With this new demand, Dependency Grammar has become a relevant term in Natural Language Processing works (KÜBLER *et al.*, 2009; DE MARNEFFE and NIVRE; 2019). However, studies related to Dependency Grammars were marginalized compared to the traditional methodology of syntactic analysis, the constituents analysis, which made it difficult to access the theoretical foundation that this type of methodology follows. Considering this, this work intends to make an introductory review of the fundamentals that govern most of the syntactic theories that use the dependency methodology, to make the Dependency Grammar models, as well as the concept of dependency, clearer and more accessible.

KEYWORDS: Descriptive Linguistics; Syntax; Dependency Grammar.

* Mestre em Linguística Pela UFSCar. ORCID: <https://orcid.org/0000-0002-4004-3182>.

[†] Professor do Departamento de Letras da UFSCar. ORCID: <https://orcid.org/0000-0002-0091-8079>.

RESUMO PARA NÃO ESPECIALISTAS: Dentre as diversas teorias que buscam explicar como se organizam os elementos da gramática de uma língua, um modelo que surgiu nos anos 50, a Gramática de Dependência tem se tornado relevante recentemente por sua praticidade em aplicações voltadas a linguística computacional. O presente artigo pretende realizar uma revisão das principais características que envolvem essa teoria. Dentre essas características, tem-se o conceito de grafo conectado, que ajudaria a explicar de maneira visual como as dependências ocorrem dentro das frases.

Introdução

O conceito de *dependência* em língua natural é extremamente produtivo entre as diferentes formas de análise. Se tomados individualmente, cada estrato linguístico pode ser analisado em função de dependências. Fonética, morfologia, sintaxe, semântica, cada um desses níveis exprime alguma relação de *dependência* dentro de sua estrutura (ENFIELD, 2017). Todavia, as formas mais recorrentes de análises fundamentadas em *dependência* são sintáticas, recebendo o título de *Gramática de Dependência*, doravante GD.

Essas formas de estruturação textual podem ser recuperadas em textos antigos, como as gramáticas do Sânscritos de Panini, como também em trabalhos ao longo século XX, como Sgall, Hajičová e Panevová (1986), Hellwig (1986), Mel'čuk (1988) e entre muitos outros. Mais recentemente, com o exponencial aumento de *parsers* utilizando de GD's como metodologia para o processamento de linguagem, elas ganharam um enorme destaque dentro do Processamento de Língua Natural (PLN) (KÜBLER *et al.*, 2009; DE MARNEFFE e NIVRE; 2019).

Como destaque para os trabalhos de análise sintática baseados em *dependências*, pode-se citar o célebre trabalho de Lucien Tesnière *Eléments de Syntaxe Structurale*. Publicado em 1959, o *Eléments de Syntaxe Structurale* é considerado o marco inicial dos estudos contemporâneos das Gramáticas de Dependência. Em seu livro, Tesnière define conceitos como *árvores*, *núcleos* e *nós* que são de extrema importância para as análises sintáticas com base em dependências posteriores aos seus trabalhos.

De maneira bem sucinta, pode-se definir o conceito de *dependência* em função das palavras pertencentes a um enunciado: “em uma sentença, todas, exceto uma palavra, dependem de outra(s) palavra(s)” (DEBUSMANN, 2000, p. 2). Ou seja, existem relações entre as palavras que constituem a sentença, de maneira que cada palavra necessita de outra, com exceção de apenas umas, para que esteja presente na sentença. Todavia, essa definição não diz muito sobre como identificar ou como construir essas relações entre os elementos de um enunciado.

A fim de elucidar o conceito de *dependência* e sua interação com as GD's, este trabalho tem o intuito de realizar uma revisão introdutória acerca dos fundamentos que regem a maior parte das teorias que se utilizam da metodologia de *dependências* para suas análises. Para isso, este artigo está organizado da seguinte forma: na primeira seção será realizada uma definição formal da relação de *dependência* entre os elementos de um enunciado; na segunda seção será apresentando como o conceito de *dependência* é utilizado pelas GD's para a construção de um sistema de análise linguística; na terceira seção será demonstrado, por meio de uma versão simplificada da Gramática de Hays e Gaifman (HAYS, 1964; GAIFMAN, 1965), como utilizar o conceito de *dependência* para realização de análises linguísticas.

1. Dependência

Como trazido na introdução, pela definição de Debusmann (2000, p. 2), as relações entre os elementos de um enunciado¹, recebem o nome de *dependência*, ou seja, entre dois elementos **a** e **b** relacionados em um enunciado, ou **a** dependente de **b** ou **b** depende de **a**. Todavia, saber que essas relações são dadas por *dependência* não é suficiente para definir, ou mesmo como identificar, as relações entre os elementos dentro de um enunciado.

Em De Marneffe e Nivre (2019) é possível encontrar uma lista de critérios para a identificação de dependência com base em Zwicky (1993 [1985]) e Hudson (1990), na qual a identificação é feita seguindo série de critérios em função da necessidade de processamento computacional, como também em função de diversas características linguísticas.

Segundo De Marneffe e Nivre (2019, p.203), a dependência entre a cabeça (**H**), o dependente (**D**) e uma construção² (**C**) pode ser dada em função dos seguintes critérios:

- H determina a categoria sintática de C e normalmente pode substituir C;
- H determina a categoria semântica de C; D fornece a especificação semântica;
- H é obrigatória, D pode ser opcional;
- H seleciona D e determina quanto D é obrigatório ou opcional;
- A forma de D depende de H, em concordância ou governança;
- A posição linear de D é especificada com referência a H.

¹ Para evitar a ambiguidade envolvida com o termo *palavra*, utilizaremos, assim como Debusmann (2000), o conceito de *elemento do enunciado*, ou seja, cada um dos itens linguísticos que constituem o enunciado.

² Como as gramáticas de dependência não utilizam o conceito de sintagma, as partes de um enunciado recebem a nomenclatura de *construção*.

Mesmo com essa lista, ainda é complicado de definir o conceito de *dependência*, uma vez que ela se utiliza de termos próprios das GD's (dependente, cabeça e governança), por meio de características sintáticas e semânticas, para definir a relação de dependência, o que acaba tornando-a endógena, ou seja, ela explica *dependência* em função dos itens que fazem parte do conceito.

Para uma definição não endógena, pode-se utilizar a presente no trabalho de Robinson (1970, p.260), onde ela, postula quatro axiomas para estruturas de dependência bem formadas com base nos trabalhos de Hays (1964) e Gaifman (1965) formalizam a proposta de Tesnière (2015 [1959]). Os axiomas são os seguintes:

- (a). Um, e apenas um, elemento é independente;
- (b). Todos outros elementos dependem diretamente de algum outro elemento;
- (c). Nenhum elemento depende diretamente de mais de um outro elemento;
- (d). Se um elemento p_1 depende diretamente do elemento p_2 e existe um elemento p_3 entre eles (na ordem linear da sequência de elementos), então p_3 depende diretamente de p_1 ou de p_2 ou de algum outro elemento que esteja entre eles.

Segundo Robinson (1970, p. 260), a relação ***a depende de b*** ($\langle a, b \rangle \in R$) é transitiva, ou seja, se ***a depende de b*** e ***b depende de c***, então ***a indiretamente depende de c***; assimétrica, se ***a depende de b***, então ***b não depende de a***; e irreflexiva, não existe nenhum elemento ***n*** de tal maneira que ***n depende de n***.

Pode-se agora, partindo dos axiomas, definir os termos dependente, cabeça e governança que estão presentes na definição de De Marneffe e Nivre (2019).

Assim, como a relação ***a depende de b*** é binária, é necessário que existam dois elementos que a constituam. O elemento que *depende* recebe o nome de *dependente*, enquanto o elemento que *domina* a relação recebe o nome de *governador* ou *cabeça*.

Com essa nomenclatura pode-se definir a relação oposta a *dependência*, a *governança*, ***a governa b*** ($\langle p_2, p_1 \rangle \in G$), de modo que ela é transitiva, se ***a governa b*** e ***b governa c***, então ***a indiretamente governa c***; assimétrica, se ***a governa b***, então ***b não governa a***; e irreflexiva, não existe nenhum elemento ***n***, tal que ***n governa n***. Porém, a *governança* permite que um elemento governe mais de um dependente, ou seja, um elemento que governa pode ter diversos dependentes, o que não ocorre na relação de *dependência*.



O exemplo (1) traz as relações de dependência na sentença: *A Ana abraçou o João*. No exemplo, as setas indicam o dependente, logo, o elemento ao lado oposto do arco representa o governante. Assim, podemos inferir que o conjunto R que descreve a relação de dependência é constituído pelas tuplas:

$$2. \quad R = \{\langle A, Ana \rangle, \langle Ana, abraçou \rangle, \langle João, abraçou \rangle \text{ e } \langle o, João \rangle\}$$

Enquanto o conjunto G que descreve a relação de governança, é constituído pelas tuplas:

$$3. \quad G = \{\langle Ana, A \rangle, \langle abraçou, Ana \rangle, \langle abraçou, João \rangle \text{ e } \langle João, o \rangle\}$$

Uma questão importante que deve ser ressaltada aqui é a relação entre os itens funcionais, elementos que têm significado em função do sistema linguístico, e itens de conteúdo/lexicais, elementos que representam entidades externas ao sistema linguístico. No exemplo (1), os itens funcionais, no caso os artigos *a* e *o*, são dependentes dos itens de conteúdo, respectivamente os nomes *Ana* e *João*. Isso ocorre pois, na definição de De Marneffe e Nivre (2019) o governante sempre define a categoria sintática da construção, logo, os itens de conteúdo sempre acabam por terem os itens funcionais como dependentes, entretanto, diferentes abordagens podem sugerir que, em alguns casos, os itens funcionais governem os itens de conteúdo.

Em (1) há apenas as relações $\langle a, b \rangle \in R$, ou seja, as dependências entre os pares de elementos do enunciado. As relações de dependência entre *a* e *b* não têm nenhuma distinção entre elas, o que não reflete completamente o sistema linguístico. Mesmo com a definição do conceito de *dependência*, não é possível realizar a descrição total do funcionamento de uma língua, tornando o conceito necessário, mas não suficiente para a descrição linguística por meio de *gramáticas de dependência*.

A próxima seção pretende utilizar da definição de dependência aqui elencada para exemplificar como, a maior parte das teorias, se utilizam desse conceito para a construção de descrições linguísticas com *gramáticas de dependência*.

2. Gramáticas de Dependência

Existem muitos trabalhos que se utilizam do conceito de *dependência* para realizar descrições linguísticas, essa diversidade gerou várias de teorias gramáticas, como por exemplo a Gramática de Hays e Gaifman (HAYS, 1964; GAIFMAN, 1965), a Descrição Gerativa

Funcional (Functional Generative Description) (SGALL et al., 1986), a Gramática de Unificação de Dependências (Dependency Unification Grammar) (HELLWIG, 1986), a Gramática de Palavras (Word Grammar) (HUDSON, 1990), a Teoria do Texto-Sentido (Meaning-Text Theory) (MEL'ČUK, 1988), e a Gramática de Dependência Funcional (Functional Dependency Grammar) (JÄRVINEN e TAPANAINEN, 1998). Entretanto, ainda que muito numerosas e diversificadas, todas as teorias compartilham conceitos básicos para além do conceito de *dependência*.

É possível encontrar em Polguère e Mel'čuk (2009, p. XIV) uma lista dos princípios básicos compartilhados pela maior parte, se não por todas, das teorias gramaticais fundamentadas no conceito de dependência, os quais são:

- (i). Conectividade da estrutura sintática;
- (ii). Direção das relações sintáticas;
- (iii). Organização hierárquica estrita da estrutura sintática;
- (iv). “Significado” das relações sintáticas.

A *conectividade da estrutura sintática* refere-se às relações entre os elementos de um enunciado. “A estrutura sintática forma um todo organizado, isto é, um sistema contínuo de relações sintáticas” (POLGUÈRE e MEL'ČUK, 2009, p. XIV), ou seja, toda e qualquer elemento **a**, existente no enunciado, se relaciona com ao menos um outro elemento **b**. Todos os elementos presentes em um enunciado fazem parte da estrutura sintática, de forma que nenhuma fique de fora. Cada tupla $\langle a, b \rangle \in R$ representa um enunciado mínimo, ou seja, um conjunto de dois elementos dotados de sentido.

Por exemplo, em (2), todos os elementos presentes na sentença de (1) fazem parte de ao menos uma das tuplas pertencentes de R. Também é verificável em (2) que nenhum elemento presente no enunciado, em (1), está fora do conjunto R.

Polguère e Mel'čuk (2009, p. XIV) definem que a propriedade que segue desse princípio é que “a estrutura sintática é um grafo conectado”, ou seja, a estrutura sintática, em função de *dependências*, é um conjunto relacionados entre eles, permitindo que todos os elementos constituintes de um enunciado possam ser analisados como parte de um conjunto finito.

Assim como a relação de dependência é assimétrica, as relações sintáticas estabelecidas por ela também serão, fazendo com que as relações sintáticas sejam direcionadas, ou seja, partem de um elemento na direção de um outro. Cada elemento dentro de um enunciado mínimo domina, ou é dominado, por um outro. É possível demonstrar o princípio (ii) pela

capacidade de um enunciado mínimo expressar seu significado em função da categoria gramatical de seu *governante*. Tomemos os exemplos:

4.

- a. Ana tem olhos **muito belos**
- b. Ana tem olhos **castanhos**

5.

- a. Ana corre **rápido**
- b. Ana corre **muito bem**

Mesmo que exista um advérbio dentro do enunciado mínimo *muito belo(s)*, em (4a), ele não exprime comportamento adverbial, o enunciado mínimo se comporta como um adjetivo, o governante do par, *belo(s)*. Em (5b), o enunciado mínimo *muito bem* se comporta como um advérbio, assim como *rápido* em 5(a), uma vez que o governante do par é *bem*, um advérbio.

Polguère e Mel'čuk (2009, p. XIV) definem que a propriedade formal que segue do princípio de direcionalidade das relações sintáticas é que: a estrutura sintática é um grafo *conectado* e *direcionado*. Esse princípio faz com que a estrutura sintática seja um grafo conectado e organizado, existe um sentido e uma direção dentro do grafo, permitindo identificar quais os elementos mais relevantes para a estrutura.

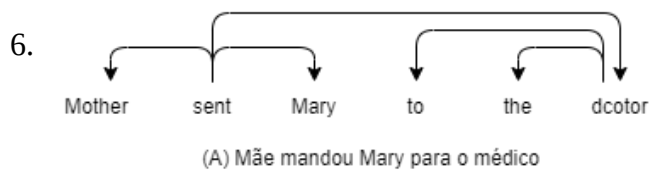
O princípio (iii) define que, assim como proposto pelos axiomas I e II de Robinson (1970, p. 260), cada elemento dentro da estrutura sintática é governado por um outro elemento, com exceção de um único elemento que governa todo enunciado. A organização hierárquica estrita define que cada enunciado mínimo tenha um, e apenas um, governador e que existe um elemento máximo da estrutura de onde todos os outros elementos dependem, de maneira que o elemento que não é governado por nenhum outro é o item mais alto da estrutura de dependência.

O princípio (iii) faz com que não seja possível a existência de ciclos infinitos no grafo da estrutura sintática. A estrutura parte do governador mais alto, passando pelos pares intermediários e termina nos dependentes que não governam nenhum elemento. Dessa forma, nenhum outro elemento irá dominar o governador máximo. Assim, no exemplo (1), o governador máximo do enunciado é o verbo *abraçar*, que não depende de nenhum outro elemento dentro da estrutura, os nomes *Ana* e *João* governam respectivamente os artigos *a* e *o*, que, por sua vez, não têm dependentes.

A propriedade formal que segue do princípio da organização hierárquica restrita do enunciado é que: “a estrutura sintática é um *grafo direcionado, acíclico e conectado*, ou seja,

uma árvore hierarquizada ou, mais brevemente, uma árvore” (POLGUÈRE e MEL'ČUK, 2009, p. XV). Assim, segue-se do princípio (iii) que a estrutura sintática é um sistema télico, ou seja, ele parte de um elemento, o governador máximo, e termina em um, ou mais, outro(s) elemento(s).

Pelos princípios (i), (ii) e (iii), a estrutura sintática ainda é um reflexo das relações de dependência, porém, como já dito, apenas as dependências não são suficientes para exprimir a sintaxe das línguas naturais. Um exemplo disso é a possibilidade de haver duas relações de dependência R_1 e R_2 constituídas pelos mesmos elementos **a** e **b**, mas que exprimem relações sintáticas distintas. Para exemplificar, tomemos as sentenças *Mother Sent Mary to the doctor* e *Mother sent Mary 200\$* trazidas por Polguère e Mel'čuk (2009, p. XV):



Em ambos os exemplos, existe o mesmo par relacional $\langle Mary, to\ send \rangle \in R$, contudo, ele exprime relações sintáticas distintas para cada exemplo. Em (6), a relação entre *Mary* e *to send* é a relação de objeto direto, enquanto em (7), os mesmos elementos estão se relacionando como objeto indireto. Mesmo que o par relacional $\langle Mary, to\ send \rangle \in R$ tenha a mesma distribuição nos enunciados, *Mary* ocorre logo após *to send*, em cada um dos exemplos ele representará uma relação de sentido distintas.

Assim, as relações sintáticas não são apenas distribuições estruturais, mas também são dotadas de significado. Entretanto, deve-se notar que o sentido das relações sintáticas não é pré-definido, ele é passível de ser alterado de acordo com a especificidade do governador do enunciado mínimo, como podemos observar nos seguintes exemplos:

8. A UFSCar contratou novos professores;
9. A UFSCar sofreu cortes do governo;
10. A UFSCar sediará um evento internacional este ano.

Mesmo que a relação $\langle a, b \rangle \in R$ entre *UFSCar* e os verbos das sentenças seja de sujeito, em cada exemplo *UFSCar* desempenha um papel semântico distinto, respectivamente: AGENTE, PACIENTE e LOCATIVO.

Finalizando as propriedades que seguem dos princípios, a propriedade formal que segue do princípio (iv) é que: “a estrutura sintática é uma árvore em que os nós são rotulados e vinculados por arcos nomeados pelas relações sintáticas que representam” (POLGUÈRE e MEL'ČUK, 2009, p. XVI). Assim, cada arco que liga os elementos do grafo da estrutura terá uma nomenclatura específica para representar sua função sintática.

Os princípios (i), (ii), (iii) e (iv) abarcam a maior parte das teorias gramaticais de dependência, mas eles ainda não são suficientes para solucionar problemas cotidianos da linguagem. Um dos problemas mais evidentes é a coordenação. Tomemos os exemplos:

11. Pedro e João visitaram a praia;
12. Pedro gosta de cães e gatos;
13. Pedro comprou e comeu um bolo;

Pelo princípio (iii), as relações sintáticas devem ser estritamente hierarquizadas, todavia isso é contra intuitivo nos exemplos. Tanto *Pedro* quanto *João* são ótimos candidatos para dependente na relação de sujeito do verbo *visitar*, em (11), assim como *cães* e *gatos* são ótimos candidatos para objeto indireto de *gostar*, em (12), e *comprar* e *comer* são ótimos candidatos para governantes em (13). O conceito de dependência, em princípio, não é suficiente para esse tipo de relação.

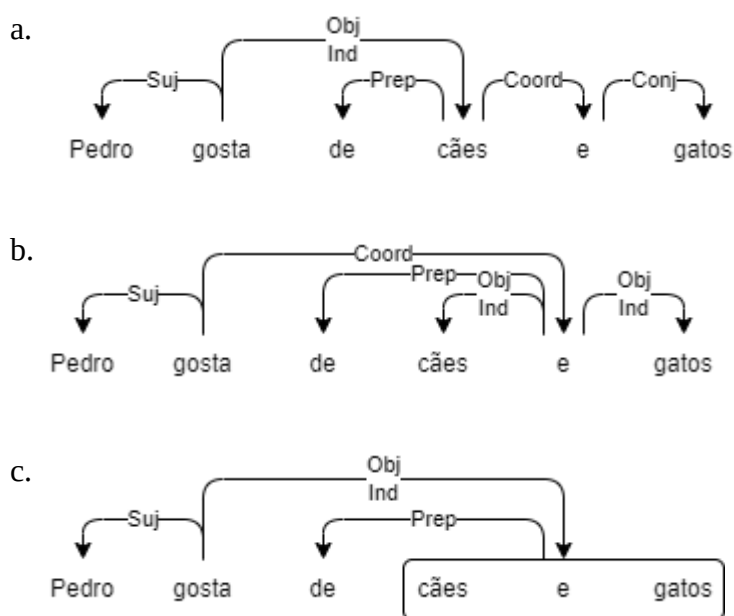
Diversas soluções foram propostas para esse problema. Em Tesnière (2015 [1959]) define-se, além da relação de *dependência*, outras duas relações dentro da estrutura sintática. A primeira é a relação de *junção* (do francês *jonction*), ela é responsável pelas estruturas de coordenação em sentenças como nos exemplos (11), (12) e (13). A segunda relação é a *translação* (do francês *translation*), ela é responsável por alterar a categoria lexical de um elemento. Um exemplo para a *translação* é o que ocorre com a sequência *de frango* em *estrogonofe de frango*, em que o *de* permite que o nome *frango* modifique o nome *estrogonofe*, tal modificação ficaria reservada para a relação de dependência dos adjetivos.

Outras soluções possíveis para o problema da coordenação podem ser observadas no exemplo (14), abaixo. Para (14a), tem-se a proposta de Mel'čuk (1988), em que a coordenação é tratada como uma relação de dependência, onde o primeiro elemento da coordenação é o governante de quem a conjunção depende que, por sua vez, governa o último elemento da coordenação.

Em (14b), tem-se a proposta de Sgall et al. (1986), na qual a relação de dependência *coord* permite que a conjunção *e* seja o governante da estrutura de coordenação, fazendo com que os elementos da coordenação sejam dependentes dela.

E no exemplo (14c), tem-se a proposta de Kahane (1997), nela as relações de dependência não ocorrem apenas entre os elementos da sentença, mas pode ocorrer em um conjunto de elementos (que ele chama de *bolhas*), fazendo com que toda a estrutura *cães e gatos* sejam apenas uma *bolha*.

14.

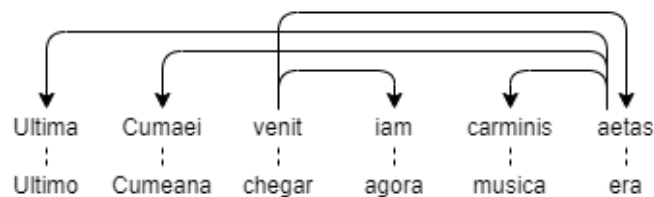


Em todas as propostas a hierarquia estrita dos elementos ainda é mantida, entretanto cada teoria recorra a uma forma distinta para empregar esse conceito.

Por fim, um outro conceito relevante para as teorias gramáticas com base em dependência é a *projetividade* (LECERF, 1960; HAYS, 1964; GAIFMAN, 1965), isto é, para que uma árvore de dependências seja projetiva, os galhos não se intersectam, ou seja, as dependências da árvore não se cruzam entre elas. Ainda que a maior parte das sentenças da maioria das línguas naturais seja projetiva (POLGUÈRE e MEL'ČUK, 2009, p.85), é possível encontrar exemplos em que a *projetividade* é quebrada.

Línguas que marcam caso morfológicamente, como o latim, o russo ou o húngaro, aceitam distribuições livres entre os elementos sintáticos, uma vez que a função sintática é demarcada pela morfologia e não pela ordenação. Essa liberdade com relação a ordenação faz com que existam sentenças que violem a *projetividade*.

15.



“É chegada agora a era da última canção cumeana”, (VIRGIL, Eclogue 4 apud POLGUÈRE e MEL’ČUK, 2009, p.87)

No exemplo (15), em latim, é possível observar que o arco que representa a dependência entre *aetas* e *venit* intersecta (cruza) os arcos que representam as dependências de *Ultima* e *Cumaei* em relação a *aetas*.

As teorias de descrição gramatical tratam a falta de projetividade de diferentes maneiras. Em sua maioria, elas propõem soluções multiestrato (SGALL *et al.*, 1986; MEL’ČUK, 1988; HUDSON, 1990, por exemplo), por essas propostas, a estrutura de dependência não está associada apenas à sintaxe, mas também a outros níveis de granularidade linguística, como morfologia e semântica.

Dentro das teorias multiestrato, algumas, como Sgall *et al.* (1986), assumem que a falta de projetividade só existe nos níveis mais altos, como a sintaxe, e é solucionada nos níveis mais profundos, como a semântica. Outras propostas, como as de Mel’čuk (1988) e Hudson (1990), recorrem a transformações lógicas e interações entre os diferentes níveis para solucionar a falta de projetividade. Entretanto, não há consenso em relação a como a falta de projetividade deve ser tratada.

Finalizada aqui, esta seção trouxe levantamento das características comuns que regem as teorias gramaticais baseadas em dependência. Toda a discussão se estabeleceu de maneira abstrata para que fosse possível relacionar os conceitos com diferentes teorias, mas em nenhum momento foi abordada uma teoria em específico. Para que seja possível observar o funcionamento dos princípios e propriedades levantados ao longo deste artigo, a próxima seção apresentará o funcionamento do modelo proposto por Hays (1964) e Gaifman (1965), um modelo simples, mas que é de grande influência em diversos trabalhos sobre dependência (DEBUSMANN; 2000, p.6).

3. Gramática de Dependência de Hays e Gaifman

Para demonstração do comportamento de um sistema de uma gramática de dependências, examinaremos a proposta da Gramática de Hays e Gaifman (HAYS, 1964; GAIFMAN, 1965), doravante GHG, uma vez que, além de ser a base dos postulados de Robinson (1970) para estruturas de dependência bem formadas, ela influenciou teorias como as de Lombardo e Lesmo (1988) e Lai e Huang (1998) e também pode ser recuperada em trabalhos recentes como Morey et al. (2018).

De fato, Hays (1964) e Gaifman (1965) são os primeiros que visam construir os axiomas necessários e suficientes para a implementação de uma gramática de dependência em algoritmos autônomos de análise textual. Ambos os trabalhos são reestruturações do texto de Tesnière (2015 [1959]), mas partindo de um ponto de vista matemático com foco na relação de *dependência*, deixando de lado as relações de *junção* e *translação*.

Segundo Gaifman (1965, p. 305-306), uma gramática de dependência é constituída por três conjuntos de regras:

- L_I : Regras que tem a forma $X(Y_1, Y_2, Y_3, \dots, Y_m * Y_{m+1}, Y_{m+2}, Y_{m+3}, \dots, Y_n)$, para X na posição de $*$, nas quais m poder ser igual a 0 e/ou igual a n . Essas regras estão relacionadas à capacidade da categoria X ocorrer com as categorias $Y_1, Y_2, Y_3, \dots, Y_n$ como suas dependentes na ordem descrita pela regra.
- L_{II} : Regras que forneçam a lista de elementos pertencentes a cada categoria X , de maneira que um elemento pode ser parte de mais de uma categoria.
- L_{III} : Uma regra que fornece a lista de todas as categorias que possam governar uma sentença, ou seja, categorias que possam ocorrer sem depender de nenhuma outra.

De modo mais claro, L_I se refere a distribuição das categorias, ou seja, como elas podem ocorrer, dentro de uma sentença e quais outras podem ocorrer em seu entorno:

16. O João foi para o escritório

a.	O	João	foi	para	o	escritório
	Det	Nome	Verbo	Prep	Det	Nome

17. Pedro alcançou o ônibus

a.	Pedro	alcançou		o	ônibus
	Nome	Verbo		Det	Nome

L_{II} se refere a um conjunto de funções capaz atribuir a cada elemento de um enunciado uma categoria, como os conjuntos de elementos nos exemplos 19 e 20:

18. Det = {O, A, Os, As, Um, Uma, Uns, Umas, etc}

19. Nome = {Pedro, João, ônibus, bolo, escritório, Maria, etc}

E L_{III} se refere a um conjunto de categorias que possam ocorrer em uma sentença, mas que não dependem de nenhuma outra. Mesmo que na maior parte dos enunciados o *verbo* ocorra como o governador máximo, existem enunciados deverbais, permitindo que existam outras categorias possíveis dentro da regra L_{III} .

Ainda segundo Gaifman (1965, p. 306), uma sentença; o conjunto de elementos $p_1, p_2, p_3, \dots, p_i$; é analisada pelo sistema de dependência em função das categorias $X_1, X_2, X_3, \dots, X_i$ de seus p_i elementos, de maneira que o índice i de p_i seja o mesmo que o de X_i . Entre os elementos da sentença existe a relação binária de dependência $d(p_i, p_j)$, p_i depende de p_j . Para cada relação d pode-se definir a relação d^* , que é sua relação de transitividade.

- I. Para nenhuma p_i , $d^*(p_i, p_i)$.
- II. Para cada p_i existe ao menos um p_j tal que $d(p_i, p_j)$.
- III. Se $d^*(p_i, p_j)$ e existe uma p_k entre p_i e p_j , então $d^*(p_k, p_j)$.
- IV. O conjunto de elementos da sentença é inteiramente conectado por relações d .
- V. Se $p_1, p_2, p_3, \dots, p_m$ são dependentes à esquerda e $p_{m+1}, p_{m+2}, p_{m+3}, \dots, p_n$ são dependentes à direita de algum elemento, e $X_1, X_2, X_3, \dots, X_m, X_{m+1}, X_{m+2}, X_{m+3}, \dots, X_n$ são as categorias de $p_1, p_2, p_3, \dots, p_m, p_{m+1}, p_{m+2}, p_{m+3}, \dots, p_n$, então $X(X_1, X_2, X_3, \dots, X_m * X_{m+1}, X_{m+2}, X_{m+3}, \dots, X_n)$ é uma regra de L_I .
- VI. O elemento g que governa a sentença, ou seja, que não depende de nenhuma outra, é um elemento que corresponde a uma categoria listada em L_{III} .

Em termos mais claros, I afirma que não existe nenhum elemento que possa depender direta ou indiretamente dele mesmo; II e VI afirmam que cada elemento na sentença depende de ao menos um outro, com exceção de g ; III afirma que se um elemento p_i depende, direta ou indiretamente, de um outro p_j e existe um elemento p_k entre eles, então p_k depende direta ou indiretamente de p_j ; e IV afirma que todos os elementos da sentenças estão conectadas por relações dependência.

Assim, se segue de II e VI que existe ao menos um elemento que não depende de nenhum outro e se segue de I que é necessário que seja apenas um elemento que não depende de nenhum outro; III estabelece a ordenação das conexões das relações de dependência; IV estabelece a completude das relações de dependência em relação aos elementos; V estabelece a relação entre a regra L_I e a ordenação dos elementos em uma sentença, uma vez que sentenças

são escritas em função de seus elementos e não das categorias dos elementos; e, por fim, VI afirma que a o elemento que não depende de nenhuma outra precisa ser um elemento possível para governar a sentença.

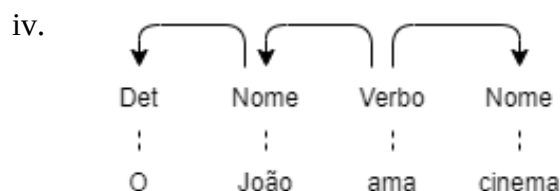
Para a demonstração do sistema, tomemos (D1), uma demonstração simplificada do funcionamento da GHG. Em (D1-i) está presente a sentença analisada, em (D1-ii) o conjunto de regras L_I que ordena a sentença, em (D1-iii) o conjunto de regras L_{II} que vincula os elementos da sentença a suas categorias e em (D1-iv) a árvore de dependências resultantes da análise:

D1.

i. O_1 João₂ ama₃ cinema₄

ii. $Verb_3(Nome_2, * Nome_4)$
 $Nome_2(Det_1, *)$
 $Nome_4(*)$
 $Det_1(*)$

iii. $Verb = \{\text{amar}, \dots\}$
 $Nome = \{\text{João}, \text{cinema}, \dots\}$
 $Det = \{O, \dots\}$



Assim, pode-se definir que, como proposto por Debusmann (2000, p.7), uma gramática de dependência como a GHG é constituída de quatro conjuntos distintos:

Quadro 1 – Gramática de Dependência de Hays e Gaifman

Gramática de Dependência de Hays e Gaifman = $\langle R, L, C, F \rangle$
Em que: R é o conjunto de regras de dependência entre as categorias sintáticas; L é o conjunto de símbolos terminais, ou elementos; C é o conjunto de símbolos auxiliares, ou categorias sintáticas; e F é uma função de atribuição de L para C ($F: L \rightarrow C$), ou seja, uma regra de associação entre os elementos e suas categorias sintáticas.

Fonte – O autor.

O intuito desta seção é demonstrar o comportamento de uma gramática de dependências utilizando os princípios e propriedades da seção anterior, todavia, a GHG não classifica o sentido das relações de dependência, como proposto por Polguère e Mel'čuk (2009), ela apenas indica a dependência entre as categorias. Entretanto, a nomenclatura das relações de

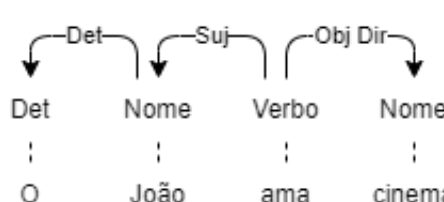
dependência pode ser deduzida a partir distribuição das categorias sintáticas que fazem parte de cada relação de dependência, permitindo que seja possível representar a nomenclatura das categorias sintáticas com o modelo da GHG, apenas com a adição de dois novos conjuntos:

Quadro 2 – Gramática de Dependência de Hays e Gaifman com a nomenclatura das relações sintáticas

Gramática de Dependência de Hays e Gaifman com a nomenclatura das relações sintáticas = $\langle R, N, L, C, F, G \rangle$
Em que: R, L, C e F são análogos ao quadro 1; N é o conjunto de sentidos para cada regra de R; e G é uma função de atribuição de R para N ($G: R \rightarrow N$), ou seja, uma regra de associação entre as regras de dependência e seus sentidos.

Fonte – O Autor.

O que permite a demonstração (D1) ser reescrita, com a adição de mais um passo, (iv), como:
D2.

- i. O_1 João₂ ama₃ cinema₄
- ii. $Verb_3(Nome_2, * Nome_4)$
 $Nome_2(Det_1, *)$
 $Nome_4(*)$
 $Det_1(*)$
- iii. $Verb = \{amar, \dots\}$
 $Nome = \{João, cinema, \dots\}$
 $Det = \{O, \dots\}$
- iv. $Sujeito = Verb(Nome, *)$
 $Det = Nome(Det, *)$
 $Objeto Direto = Verb(*, Nome)$
- v.


Demonstrando assim, um modelo de gramática de dependências que é capaz de exprimir os princípios e as propriedades elencados na seção anterior. Os princípios de *conectividade da estrutura sintática*, de *direção das relações sintáticas*, de *organização hierárquica estrita da estrutura sintática* e de “*significado*” das *relações sintáticas*; e as propriedades que ditam que a estrutura sintática, em função de dependência, é um grafo conectado, ordenado e acíclico em que as conexões entre os elementos recebem nomes representando suas funções sintáticas.

4. Considerações Finais

O presente artigo teve como objetivo realizar uma introdução acerca do conceito de *dependência* e como ele é utilizado por diferentes teorias para realização de análises linguísticas, teorias essas que recebem a nomenclatura de Gramáticas de Dependência. Dessa forma, na primeira seção, por meio dos trabalhos de De Marneffe e Nivre (2019) e Robinson (1970), foi explicitada a forma como os elementos de um enunciado se relacionam entre eles por meio do conceito de *dependência*. As definições da primeira seção foram utilizadas na seção seguinte (segunda seção) para, com base em Polguère e Mel'čuk (2009), demonstrar os princípios básicos compartilhados pela maior parte das Gramáticas de Dependência: *conectividade da estrutura sintática*, de *direção das relações sintáticas*, de *organização hierárquica estrita da estrutura sintática* e de “*significado*” das *relações sintáticas*. Por fim, na seção três, realizou-se a demonstração, por meio de uma versão simplificada do modelo de Hays e Gaifman (HAYS, 1964; GAIFMAN, 1965), de como utilizar o conceito de *dependência*, e os princípios elencados na segunda seção, para a realização de análises em enunciados linguísticos.

Em vista da crescente utilização das GD's como metodologia para a realização de análise sintática automática, acredita-se que a discussão aqui apresentada é de extrema relevância para o debate científico, ainda mais considerando o apagamento sofrido pelas GD's em oposição às teorias “tradicionais” de análise sintática, os modelos em função de constituintes.

Por fim, o presente artigo é apenas um recorte abstrato das características essenciais para as diversas teorias fundamentadas em análises de *dependência*. Para uma leitura mais teórica, recomenda-se ao leitor os trabalhos de Debusmann (2000), Polguère e Mel'čuk (2009) e De Marneffe e Nivre (2019); e para trabalhos com aplicabilidade das GD's, recomenda-se a leitura de Nivre (2006) e De Marneffe *et al.* (2021).

Agradecimentos

Os autores deste trabalho agradecem ao Centro de Inteligência Artificial (C4AI-USP) com apoio da Fundação de Apoio à Pesquisa do Estado de São Paulo (Processo FAPESP 2019/07665-4) e da IBM Corporation.

Conflito de Interesse

Os autores não têm conflitos de interesse a declarar.

Link para *Preprint*

Os manuscritos submetidos a *CadLin* devem ser previamente depositados em um [servidor de *preprint*](#) que suporte comentários públicos. O DOI do *preprint* para deve ser fornecido nesta seção.

Protocolo e Pré-Registro de Pesquisa

Os autores avaliaram os roteiros da Equator Network e não foram capazes de encontrar nenhum modelo relevante para este manuscrito.

Ademais, a pesquisa não foi pré-registrada em um repositório institucional independente.

Declaração de Disponibilidade de Dados

O compartilhamento de dados não é aplicável a este artigo, pois nenhum dado novo foi criado ou analisado neste estudo.

Fontes de financiamento

O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 88887.499210/2020-00.

Referências

DE MARNEFFE, M. C.; MANNING, C. D.; NIVRE, J.; ZEMAN, D. Universal Dependencies. **Computational linguistics**, v. 47, n. 2, p. 255-308, 2021. DOI: https://doi.org/10.1162/coli_a_00402.

DE MARNEFFE, M.C.; NIVRE, J. Dependency grammar. **Annual Review of Linguistics**, v. 5, p. 197-218, 2019. DOI: <https://doi.org/10.1146/annurev-linguistics-011718-011842>.

DEBUSMANN, R. An introduction to dependency grammar. **Hausarbeit fur das Hauptseminar Dependenzgrammatik SoSe**, v. 99, p. 1-16, 2000.

ENFIELD, N. J. Dependencies in language. In: ENFIELD, N. J. **Dependencies in language: On the causal ontology of linguistic systems**. Berlin: Language Science Press, 2017. cap. 1, p. 1-8. ISBN 978-3-946234-88-3. DOI: <https://doi.org/10.5281/zenodo.573773>.

GAIFMAN, H. Dependency systems and phrase-structure systems. **Information and control**, v. 8, n. 3, p. 304-337, 1965.

HAYS, D. G. Dependency theory: A formalism and some observations. **Language**, v. 40, n. 4, p. 511-525, 1964.

HELLWIG, P. Dependency unification grammar. In: **Coling 1986 Volume 1: The 11th International Conference on Computational Linguistics**. 1986.

HELLWIG, P. Dependency unification grammar. In: **Coling 1986 Volume 1: The 11th International Conference on Computational Linguistics**. 1986

HUDSON, R. **English Word Grammar**. Oxford, Uk, 1990.

JÄRVINEN, T.; TAPANAINEN, P. Towards an implementable dependency grammar. In: **Processing of Dependency-Based Grammars**. 1998.

KAHANE, B. Bubble trees and syntactic representations. In: **Proceedings of mathematics of language (mol5) meeting**. DFKI, Saarbrücken, 1997. p. 70-76.

KÜBLER, S.; MCDONALD, R.; NIVRE, J. Dependency parsing. **Synthesis lectures on human language technologies**, v. 1, n. 1, p. 1-127, 2009.

LAI, T. BY; HUANG, Changning. Complements and adjuncts in dependency grammar parsing emulated by a constrained context-free grammar. In: **Processing of Dependency-Based Grammars**. 1998.

LECERF, Y. Programme des conflits, modele des conflits. **Bulletin bimestriel de l'ATALA**, v. 1, n. 4, p. 11-18, 1960.

LOMBARDO, V.; LESMO, L. Unit coordination and gapping in dependency theory. In: **Proceedings of the Workshop on Processing of Dependency-Based Grammars**. 1998. p. 11-20.

MEL'ČUK, I. A. **Dependency syntax: theory and practice**. Albany: State University of New York Press, 1988.

MOREY, M.; MULLER, P.; ASHER, N. A dependency perspective on rst discourse parsing and evaluation. **Computational Linguistics**, v. 44, n. 2, p. 197-235, 2018. DOI: https://doi.org/10.1162/coli_a_00314.

NIVRE, J. Dependency grammar and dependency parsing. **MSI report**, v. 5133, n. 1959, p. 1-32, 2005.

NIVRE, J. Dependency Parsing. *In*: NIVRE, J. **Inductive Dependency Parsing**. Springer Dordrecht, 2006. p. 45-86. ISBN 978-1-4020-4889-0. DOI: https://doi.org/10.1007/1-4020-4889-0_3.

POLGUÈRE, A.; MEL'ČUK, I.A. (Ed.). **Dependency in linguistic description**. John Benjamins Publishing, 2009. ISBN 978-90-272-8958-2.

ROBINSON, J. J. Dependency structures and transformational rules. **Language**, p. 259-285, 1970.

SGALL, P.; HAJIČOVÁ, E.; PANEVOVÁ, J. **The Meaning of the Sentence in Its Semantic and Pragmatic Aspects**. Dordrecht: D. Reidel Publishing Company, 1986.

TESNIÈRE, L. **Elements of structural syntax**. John Benjamins Publishing Company, 2015. [1959].

ZWICKY, A. M. 13 Heads, bases and functors. **Heads in grammatical theory**, p. 292, 1993. [1985].