

A EDIÇÃO DIGITAL DA *HISTÓRIA DO FUTURO*, DE ANTÓNIO VIEIRA: ARQUIVO E FERRAMENTAS¹¹

Ana Paula Banza¹, António Rito Silva² e Irene Rodrigues³

¹ CIDEHUS, Universidade de Évora, ² INESC, ID,

³ Departamento de Informática, Universidade de Évora

anabanza@uevora.pt, Rito.Silva@tecnico.ulisboa.pt, ipr@uevora.pt

Resumo. Aborda-se, no presente texto, a forma como as novas concepções de edição e de texto, resultantes da chegada do Digital ao mundo da Filologia, se reflectiram no projecto de criação de um arquivo digital da obra *História do Futuro*, do Padre António Vieira - “Arquivo digital e análise assistida da *História do Futuro*, de António Vieira” - que, pela sua natureza, incompleta e fragmentária, constitui um desafio filológico inatingível fora do novo paradigma digital. Apresenta-se também brevemente a estrutura e objectivos do projecto, tomando como foco a construção do arquivo e a implementação de determinadas ferramentas de processamento de linguagem natural, seleccionadas e adaptadas em função do tipo de texto, dos desafios que este suscita e das leituras que permite, desenvolvendo-se, a título de exemplo, alguns dos seus possíveis cenários de utilização, para ilustrar as mais valias que o projecto trará, quer do ponto de vista académico, quer na divulgação da obra do Padre António Vieira.

1. O PROJECTO ARQUIVO DIGITAL E ANÁLISE ASSISTIDA DA *HISTÓRIA DO FUTURO*, DE ANTÓNIO VIEIRA

A *História do Futuro* (HdoF) é uma obra muito diferente daquelas pelas quais o autor é geralmente conhecido, que são os sermões e as cartas. Neste caso, estamos perante uma obra de natureza escatológica, sobre o final dos tempos e sobre o papel que o reino de Portugal teria no estabelecimento do Império de Cristo na terra, o Quinto Império, sendo este último aspecto, que é verdadeiramente original e nada ortodoxo, o que causaria ao autor os problemas com a Inquisição que viria a enfrentar.

Apesar de ter trabalhado nesta obra desde 1649 até à sua prisão, em 1665, Vieira deixou-a inacabada e muito longe do plano, previsto para sete volumes. Do que chegou até nós desta obra, pensada para ser monumental, o Livro Antepimeiro da *História do Futuro*, organizado por Vieira, antes de ser preso, como uma introdução à obra, e enviado para a Corte numa espécie de manobra publicitária, teve uma primeira edição em 1718 e várias depois dessa. O resto dos papéis já escritos, de forma fragmentária, para a *História do Futuro* foi apreendido pela Inquisição na altura da prisão de Vieira, em 1665, e ficou apenso ao

¹¹Trabalho desenvolvido no âmbito do projeto UIDB/00057/2020 - FCT - Portugal.

Processo. Lúcio de Azevedo publicou, em 1918, o conjunto mais organizado destes fragmentos (cf. Sérgio e Cidade, 1953: 1-160), permanecendo ainda apenas parcialmente editado um outro conjunto de fragmentos, mais dispersos, cuja identificação e organização levantam vários problemas (cf. Muhana, 1994).

A Representação dos motivos que tive para me parecerem prováveis as proposições de que se trata, escrita em cativo, entre 1665 e 1666, como defesa da obra perante as acusações de heresia da Inquisição, sintetiza e, mais importante, completa de forma resumida a matéria que não tinha chegado a ser escrita.

Assim, do ponto de vista conceptual, da obra *História do Futuro*, fazem parte três obras fisicamente distintas, além dos fragmentos referidos, que é nossa intenção integrar no Arquivo digital da Obra.

No paradigma impresso, seria impossível colocar todos estes textos em diálogo, mas, como bem escreveu Paixão de Sousa (2013: 129):

“a lógica da difusão digital tem [...] como característica central (e [...] mais interessante) a desvinculação entre o texto lógico e seu suporte material. Ela descorporifica o texto”.

É precisamente nesta descorporificação do texto, isto é, na separação entre o que é físico e o que é conceptual, que se baseia este projecto, cujas questões de partida são as seguintes: Pode o meio digital permitir uma leitura nova e mais completa dos textos de António Vieira? Se sim, como?

Acreditamos que sim, porque o arquivo digital integrará toda a matéria escrita para a *História do Futuro*, reconstruindo “arqueologicamente” a Obra, a partir do seu plano, o que vai permitir uma visão nova e mais completa.

Por outro lado, a introdução de ferramentas de análise assistida permitirá identificar automaticamente as relações conceptuais entre os textos, comprovando a relação entre os diferentes textos físicos, além de assessorar a análise e comentário (linguístico, literário, histórico-cultural).

Assim, o projecto organiza-se em quatro tarefas, que se relacionam entre si: o texto, estabelecido na primeira tarefa, é integrado e organizado no arquivo, cuja construção constitui a segunda tarefa; as ferramentas, a implementar na terceira tarefa, usam o conteúdo do arquivo e são também incorporadas nele. Finalmente, a anotação e comentário, previstos para a quarta tarefa, usam o texto e as ferramentas incorporadas no arquivo e são, eles próprios, também incorporados no arquivo.

Relação entre Tarefas

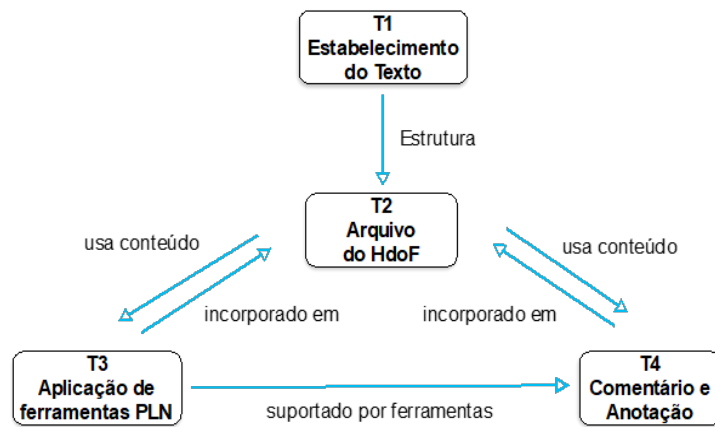


Fig. 1

Relação entre tarefas

2. O ARQUIVO

O objectivo do Arquivo da *História do Futuro* (HdoFArq.) é estruturar os fragmentos da obra do Padre António Vieira de modo a facilitar a sua interacção, promover a integração de ferramentas de Processamento de Linguagem Natural (PLN) e suportar a categorização e os comentários dos conteúdos da HdoF.

No que diz respeito ao primeiro objectivo, a interacção deve ser dirigida pela interpretação teórica associada à HdoF, a desenvolver na primeira fase do projecto, mas para a qual existe já uma base, estabelecida em trabalho anterior (Banza, 2008). Assim, é central fornecer mecanismos de ligação entre as diferentes partes de reescrita da HdoF, facilitando a sua identificação e comparação. Enquanto paradigma global de interacção com o arquivo HdoF, pretende-se adicionar mecanismos de navegação horizontal, entre as partes reescritas da HdoF, ao mecanismo usual de navegação vertical na obra. A interacção deve permitir, em qualquer momento, durante a leitura da HdoF, a possibilidade de ler e analisar todas as reescritas efetuadas pelo Padre António Vieira do fragmento em foco. Esta estrutura combinatória deve, contudo, preservar a ideia de uma obra, uma estrutura vertical, promovendo, não obstante, a combinação de reescritas. O maior desafio que se apresenta à definição da estrutura do arquivo HdoF será o estabelecimento da granularidade dos elementos reescritos, dado que o grão fragmento e sequência de fragmentos, como será inicialmente codificada a HdoF na sua dimensão vertical, pode não corresponder de forma directa às reescritas efetuadas pelo Padre António Vieira. Ou seja, a relação entre partes reescritas pode ser uma relação de muitos para muitos, quando considerado o grão fragmento. Ainda no que diz respeito a este primeiro objectivo, o

arquivo HdoF deverá disponibilizar as funcionalidades usuais de um arquivo associadas à pesquisa do seu conteúdo e ao suporte de meta-dados acerca dos seus fragmentos.

A integração com ferramentas de PLN deverá ter duas vertentes: análise e exploratória. Na vertente exploratória, o investigador, apoiado por um especialista de PLN, deve poder usar os conteúdos do arquivo HdoF para explorar as suas hipóteses de investigação através da aplicação de ferramentas de PLN. A vertente de análise permitirá aos investigadores, mas também a leigos, uma interacção sobre o arquivo HdoF através da utilização de ferramentas de PLN. Evidentemente, a vertente de análise resulta da estabilização dos resultados alcançados pelos investigadores na vertente exploratória, que, uma vez considerados de relevo, serão codificados no arquivo HdoF através da integração das ferramentas de PLN usadas na fase exploratória. Desta forma, poderão utilizadores leigos, não especialistas, usar as ferramentas na sua interacção com o arquivo HdoF. Com vista a suportar a vertente de exploração, o arquivo deverá fornecer um conjunto de interfaces de disponibilização de dados para serem consultados pelas ferramentas de PLN. Note-se que a bancada de experimentação deverá ser externa ao arquivo HdoF, uma vez que cada experimentação terá as suas características e tenderá a evoluir dinamicamente com o próprio processo de exploração. Já a vertente de análise exigirá uma adaptabilidade do arquivo HdoF por forma a permitir a fácil integração de novas ferramentas e dados. Para isso, será necessário que o arquivo HdoF possua uma boa modularidade, que limite o impacto das alterações a efectuar.

Finalmente, para se atingir o objectivo de suporte à categorização e à adição de comentários aos conteúdos da HdoF, será necessário que o arquivo permita, para além da meta-informação estática associada aos seus conteúdos, a adição dinâmica de meta-informação. Pretende-se suportar dois tipos de meta-informação: categorias e comentários. Esta meta-informação poderá ser gerada manualmente por um utilizador ou automaticamente por uma ferramenta de PLN. Tal como para a estruturação do arquivo HdoF, também aqui se levanta a questão de qual o grão do conteúdo sobre o qual se aplica a meta-informação. Sendo provável que a granularidade do fragmento não seja suficiente, serão provavelmente necessárias granularidades maiores, por exemplo, grupos de fragmentos, e menores, por exemplo, parte de um fragmento. Assim, o arquivo HdoF deverá possuir uma estrutura que suporte a introdução dinâmica de novos tipos de meta-informação.

Pretende-se desenvolver o arquivo HdoF estendendo o actual arquivo do *Livro do Desassossego* (LdoD), de Fernando Pessoa (Portela e Rito Silva, 2015; Rito Silva e Portela, 2015) com o objectivo de, através de processos de refactorização da sua base de código, construir um sistema que seja altamente flexível, de forma a poder ser facilmente adaptado para implementar diferentes repositórios de humanidades digitais. Assim, o trabalho terá por base o actual arquivo que se encontra em produção, assim como o trabalho já realizado na sua modularização e que se encontra descrito em Gonçalo et al. (2021).

Um dos resultados finais que se pretende alcançar é um sistema adaptável e reutilizável de módulos que possam ser usados na construção de diferentes arquivos. A Figura 2 apresenta este processo de definição de repositórios por integração de módulos específicos de um repositório com módulos comuns adaptados, exemplificado para os repositórios LdoD e

HdoF. Pretende-se que este processo permita desenvolver novos repositórios de forma robusta e reduzindo o esforço de desenvolvimento.

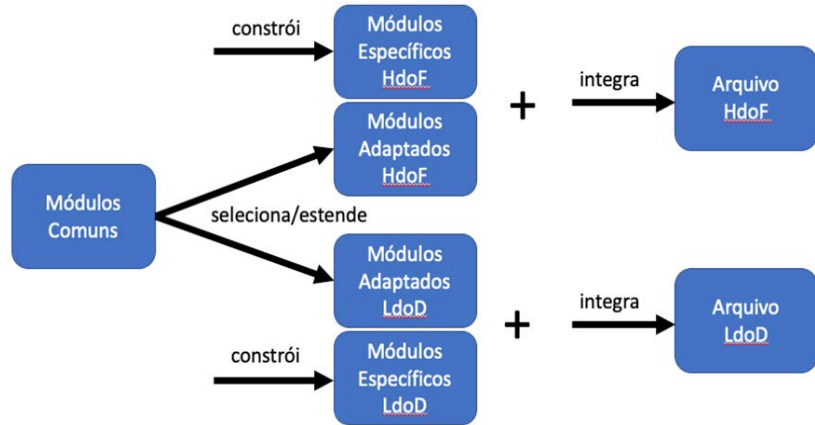


Fig. 2
Processo de Definição de Repositórios

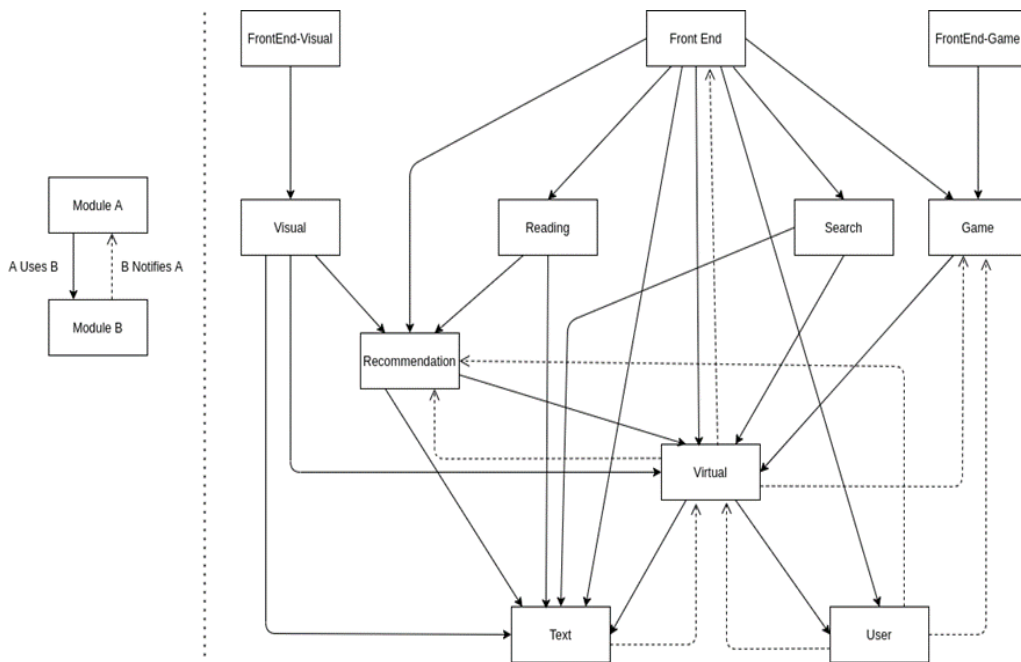


Fig. 3
Arquitetura de Módulos do Arquivo LdoD

Os módulos base que poderão ser adaptados do actual repositório de código são: *Text*, *User*, *Virtual*, *Search* e *Recommendation*. Estes módulos, ainda que requerendo alguma

adaptação e generalização, permitirão suportar alguns dos objectivos definidos para o arquivo HdoF.

O módulo de *Text* permitirá suportar a estrutura do arquivo HdoF, em particular as navegações verticais e horizontais. No arquivo LdoD, a navegação vertical está associada à estrutura de uma edição do *Livro do Desassossego*, enquanto a navegação horizontal está associada às diferentes transcrições de cada um dos fragmentos por parte dos seus editores. No arquivo da HdoF, haverá apenas uma navegação vertical, pelo que a adaptação do repositório existente será trivial. Contudo, a adaptação da navegação horizontal será mais complexa, pois terá de passar a aceitar mais graus de liberdade, devido às relações de muitos para muitos que poderão ocorrer entre diferentes fragmentos da HdoF, como anteriormente descrito.

O módulo de *User* será trivialmente adaptado para suportar a autenticação e autorização de utilizadores do arquivo HdoF, uma vez que estas funcionalidades não são centrais na semântica dos arquivos, sendo apenas logísticas.

Os módulos de *Search* e *Recommendation* podem ser adaptados para suportar as funcionalidades de pesquisa do arquivo; funcionalidades estas que estão associadas ao primeiro objectivo do arquivo.

O módulo *Virtual* implementa as funcionalidades de gestão de edições virtuais do arquivo LdoD. Estas funcionalidades são específicas deste arquivo, contudo incluem um conjunto de sub-funcionalidades que permitem a categorização da totalidade, ou partes, de um fragmento. Adicionalmente, associadas a essa categorização, também existem funcionalidades para comentar. Uma vez que um dos objectivos do arquivo HdoF é o suporte à categorização e comentário dos conteúdos da HdoF, a estratégia que se pretende seguir será a da extracção de um novo módulo de categorização e comentário do actual módulo *virtual*. Uma vez efectuada essa extracção, será possível usar este novo módulo para a categorização e comentário de fragmentos da HdoF.

Dos objectivos definidos para o arquivo HdoF, a integração de ferramentas de PLN não encontra nenhum módulo que possa ser estendido ou adaptado do actual repositório do LdoD. Assim, para se atingir este objectivo, deverá ser definido um novo módulo. Este módulo deverá ser integrado com os restantes módulos, por forma a manter a sua independência, como previamente descrito.

3. AS FERRAMENTAS

No que diz respeito às ferramentas a integrar no Arquivo HdoF, conforme referido anteriormente, elas servem essencialmente dois objectivos: a confirmação da identidade/complementaridade dos textos das diferentes obras físicas e a análise, em diferentes níveis.

Para o primeiro objectivo, serão essenciais as ferramentas de identificação de similaridades. No entanto, verifica-se que as ferramentas actualmente disponíveis não se adequam necessariamente a este tipo de texto e aos objectivos pretendidos, pelo que será necessário, em alguns casos, adaptar os modelos actuais ao caso da HdoF, conseguindo, essencialmente, mais eficiência, em termos de precisão e cobertura.

Nomeadamente, num caso como o dos excertos abaixo, em que as similaridades são evidentes, um dos objectivos será identificar automaticamente, não apenas correspondências exactas, na mesma linha ou em linhas diferentes, como as assinaladas a amarelo, mas também paráfrases, como as assinaladas a rosa, flexões da mesma palavra, como as assinaladas a verde, ou ainda outros tipos de semelhança semântica, como os assinalados a azul.

- **História do Futuro**

- Parecia-te que vias defronte de ti uma **estátua** grande, de estatura alta e sublime e de aspecto terrível e temeroso. **A cabeça** desta **estátua** era de **ouro**, **o peito** e os braços **de prata**, **o ventre** até os joelhos **de bronze**, **os joelhos** de **ferro**, **os pés** de **ferro** e de **barro**. **Estando assim suspenso** no que vias, viste mais que **se arrancava uma pedra de um monte, cortada dele sem mãos**, e que, dando **nos pés da estátua**, a derrubava. Então **se desfizeram** juntamente o barro, o ferro, o bronze, a prata, o ouro, e se converteram **em pó e cinza**, que foi levada dos ventos, e nem aqueles metais apareceram mais, nem o lugar onde tivessem estado; porém **a pedra** que tinha derrubado a **estátua cresceu**, e fazendo-se **um grande monte**, **ocupou e encheu toda a terra**.

- **Representação**

- Tinha a **estátua** a **cabeça de ouro**, **os peitos de prata**, **o ventre de bronze**, **ũa e outra perna de ferro** e **os pés de ferro e barro**. E enquanto Nabucodonosor **estava admirado** do que **via** (porque a **estátua**, como diz o texto, era grande e de aspecto terrível), **viu** mais que **de um montecão ou se arrancava, sem mãos, ãa pedra**, a qual deu um golpe **nos pés da estátua** com tão maravilhoso efeito que a **estátua** e seus metais **se desfizeram** todos em pó e cinza, e **a pedra** **crecendo** se converteu em **um monte** de tão imensa grandeza que **cobriu toda a terra**.

Fig. 4

O episódio da **estátua dos quatro metais**: tipos de similaridades

Para a análise, nos domínios linguístico, histórico-cultural e literário, serão também usadas diferentes ferramentas já disponíveis (como o reconhecimento de entidades nomeadas ou a análise de sentimentos), mas que poderão também ser adaptadas ao tipo específico de texto em causa. Tomando como exemplo as ferramentas de análise linguística, prevê-se a utilização do esquema de Dependências Universais, por exemplo, para a análise morfo-sintáctica.

Neste caso, o desempenho das ferramentas existentes é, no geral, bastante razoável. No entanto, tratando-se de um texto do séc. XVII e de um texto de Vieira, há, apesar de tudo, desafios importantes a resolver, como é o caso da interpretação de grafias antigas (e.g. ãa = uma) ou da interpretação da função sintáctica de constituintes deslocados (Teixeira et al., 2019).

Tomem-se como exemplo as frases assinaladas nos dois excertos já referidos:

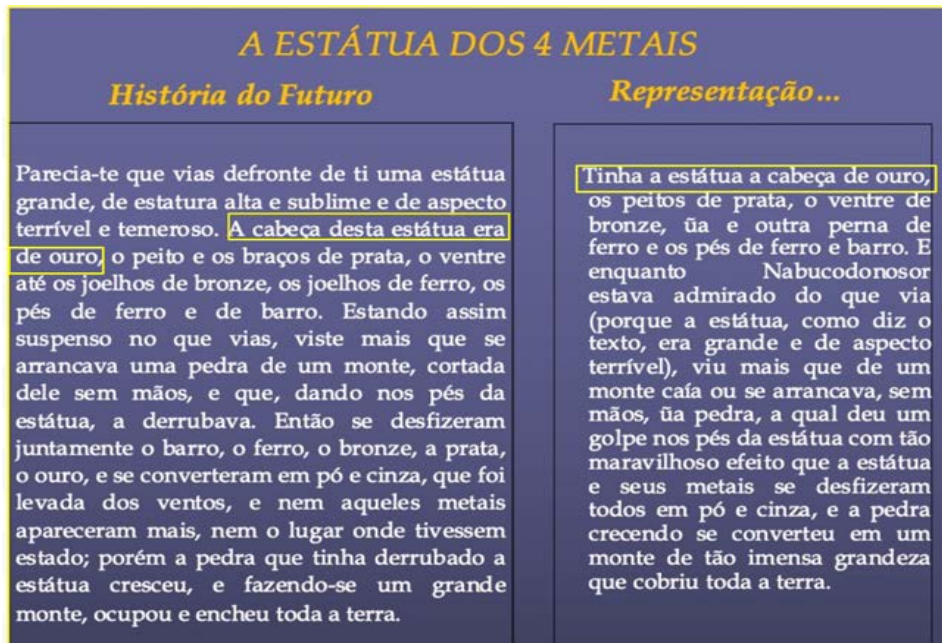


Fig. 5

O episódio da estátua dos quatro metais: ex. de similaridade morfo-sintáctica

Na frase 1, “a cabeça desta estátua era de ouro” (HdoF), a análise de um analisador sintáctico como o stanza.run é a seguinte:

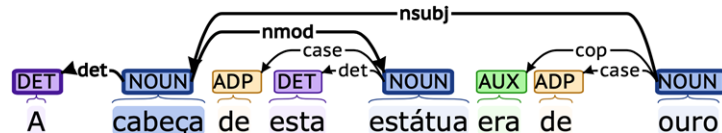


Fig. 6

Análise sintáctica da frase 1

O grafo semântico desta representação sintáctica (Silva et al., 2021) representa as seguintes relações entre os diferentes objectos, permitindo a extracção das seguintes anotações (“parte_de (cabeça,estátua) e parte_de (ouro,cabeça)”).

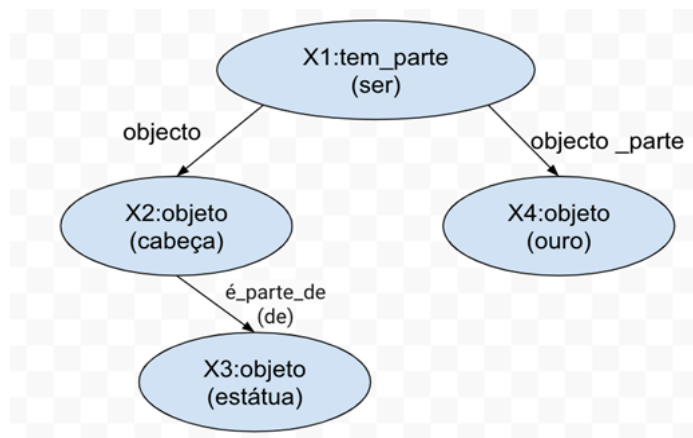


Fig. 7
Grafo semântico da frase 1

Na frase 2, “tinha a estátua a cabeça de ouro” (Representação...), o stanza.run identifica “a estátua” como objecto, por estar depois do verbo, e não como sujeito, como se apresenta na Figura 8:

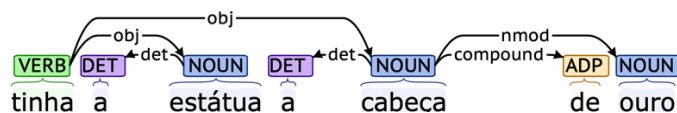


Fig. 8
Análise sintáctica da frase 2

No entanto, apesar de não identificar correctamente o sujeito da frase, esta representação sintáctica pode ser a entrada de uma ferramenta de análise semântica que, usando recursos como dicionários e taxonomias, permite chegar à representação de um termo num grafo semântico como o da Figura 9:

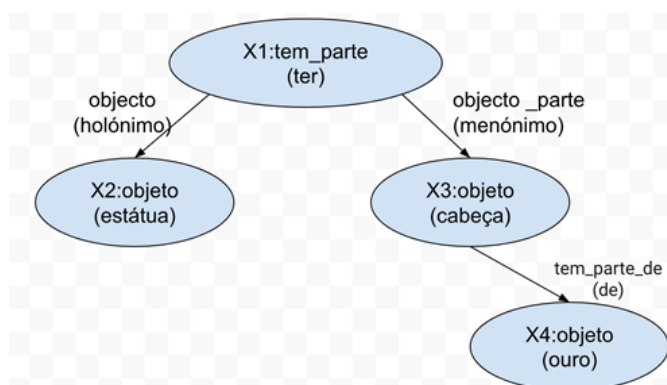


Fig. 9
Grafo semântico da frase 2

A partir deste grafo, podem gerar-se anotações ligadas ao texto na plataforma que auxiliam na detecção, manual ou automática, de similaridades entre textos.

Assim, a partir de ferramentas existentes, como o analisador sintáctico stanza.run, propomos o desenvolvimento de um analisador semântico que use os recursos melhorados no projecto (dicionários e taxonomias adequados ao vocabulário dos textos do Vieira), permitindo extrair e representar alguma informação dos textos e possibilitando a detecção automática de similaridade baseada no seu conteúdo semântico.

Em resumo, a ideia será partir de ferramentas existentes, melhorá-las, tornando-as mais versáteis, de forma a serem úteis neste tipo de texto, e desenvolver novas ferramentas que permitam representar parte do conteúdo semântico dos textos.

4. CONCLUSÕES

O Arquivo HdoF, contendo todos os textos conceptualmente pertencentes à obra, suporta leituras a duas dimensões: vertical (progressão do texto de acordo com o Plano) e horizontal (variantes).

Por outro lado, o arquivo suporta também ferramentas de PLN, para análise comparativa e textual da obra.

Finalmente, o Arquivo HdoF apoia a interacção perito-arquivo, na gestão dos resultados obtidos nas actividades de investigação.

REFERÊNCIAS

1. Banza, A.: P. *Representação perante o Tribunal do Santo Ofício, de Padre António Vieira. Edição crítica e estudo filológico*. Lisboa, Imprensa Nacional – Casa da Moeda (2008).
2. Gonçalves N., Faustino, D., Rito Silva, A., Portela, M.: Monolith Modularization Towards Microservices: Refactoring and Performance Trade-offs. *IEEE 18th International Conference on Software Architecture Companion (ICSA-C)*, pp. 1-8 (2021). Doi: 10.1109/ICSA-C52384.2021.00015.
3. Muhana, A.: *Apologia das coisas profetizadas*. Lisboa, Cotovia (1994).
4. Paixão de Sousa, M. C.: A Filologia Digital em Língua Portuguesa: alguns caminhos. In Gonçalves, M. F., Banza, A. P. (orgs.). *Património Textual e Humanidades Digitais: da Antiga à Nova Filologia*. Évora, CIDEHUS, pp. 113-138 (2013). Disponível em: <https://books.openedition.org/cidehus/1073>.
5. Portela, M., Rito Silva, A.: A model for a virtual LdoD, *Digital Scholarship in the Humanities*, Volume 30, Issue 3, September, 354–370 (2015). <https://doi.org/10.1093/lc/fqu004>.
6. Sérgio, A., Cidade, H.: História do Futuro. In *Obras escolhidas do Padre António Vieira*, IX, pp. 1-160 (1953 [1918]).
7. Rito Silva, A., Portela, M.: TEI4LdoD: Textual Encoding and Social Editing in Web 2.0 Environments. *Journal of the Text Encoding Initiative [Online]*, Issue 8 (2015). <https://doi.org/10.4000/jtei.1171>.
8. Silva, J.Q., Melo, D., Rodrigues, IP., Seco, JC., Ferreira, C., & Parreira, J.: An Ontology based Task Oriented Dialogue. *Proceedings of the 13th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, IC3K 2021, Volume 2: KEOD, Online Streaming, October 25-27, 2021, Book Series: KEOD*, pp. 96-107 (2021).
9. Teixeira, C., Rodrigues, I.: Deciphering Latin sentences using traditional linguistic resources. *Digital Scholarship in the Humanities*, Volume 34, Issue 4, 791–805 (2019).
10. Vieira, A.: *Historia do Futuro. Livro Antepimeiro*. Lisboa Occidental, Oficina de Antonio Pedrozo Galram (1718).