

Complexidade textual em adultos de baixa literacia: análise dos níveis de complexidade e as suas características

Ricardo Monteiro

**Dissertação de Mestrado em Ciências da Linguagem,
Linguística**

Versão corrigida e melhorada após a sua defesa pública

Mai 2024

Dissertação apresentada para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Ciências da Linguagem, área de especialização em Linguística, realizada sob a orientação científica da Professora Doutora Susana Correia e a Professora Doutora Raquel Amaro.

Para o vindouro

AGRADECIMENTOS

Agradeço particularmente às minhas orientadoras, Professora Doutora Susana Correia e a Professora Doutora Raquel Amaro pelos discernimento e sabedoria transmitidos ao longo da execução deste trabalho. O seu senso será o meu siso.

Agradeço também à falecida Professora Doutora Teresa Brocardo por me ter aberto a porta das Ciências da Linguagem. É ela a grande editora de todas as palavras aqui escritas.

Agradeço aos meus colegas do projeto iRead4Skills pelos momentos sérios e leves, pelas aventuras e questões, pela sabedoria partilhada e pelos dilemas escutados. Este trabalho também é deles.

Agradeço aos comparsas do CLUNL e CAN que tanto me deram conversa quando precisava.

Agradeço à Sofia por me lembrar do que consigo fazer.

COMPLEXIDADE TEXTUAL EM ADULTOS DE BAIXA LITERACIA: ANÁLISE DOS NÍVEIS DE COMPLEXIDADE E AS SUAS CARACTERÍSTICAS

RICARDO MONTEIRO

RESUMO

Este trabalho pretende descrever o processo de criação de uma infraestrutura de complexidade textual. Inserida no âmbito do projeto iRead4Skills, esta infraestrutura ambiciona detetar e classificar a complexidade dos textos, nomeadamente aquela que se manifesta para adultos de baixa literacia. Apesar da existência de vários instrumentos que dão orientações sobre a proficiência em leitura (QECR, PISA, etc.), pouca atenção é dada às dificuldades de leitura sentidas por adultos em programas de aprendizagem ao longo da vida e a como estas dificuldades se manifestam linguisticamente nos textos. Baseado em referenciais existentes e em circulação (ANQEP, QECR, PIAAC, ALTE), este trabalho aborda este fenómeno através da análise de níveis e os respetivos descritores dentro destes referenciais, resultando numa proposta original. Composta por três níveis, esta proposta considera, para cada nível, categorias linguísticas como a dimensão léxico-conceptual, estruturas sintáticas e coesão dos textos como indutoras de complexidade. Estas são exibidas e detalhadas através de descritores que dão conta da complexidade textual. Depois disto, estes níveis e descritores foram expostos a um processo de validação, que consistiu na sua avaliação por parte de um *focus group*, composto por professores de centros de aprendizagem para adultos. Paralelamente, estes níveis e descritores de complexidade foram usados na compilação de um *corpus*, através da validação dos textos segundo os níveis e descritores aqui propostos. Por último, uma aplicação da infraestrutura de complexidade foi ensaiada com o objetivo específico de revelar a ligação entre os textos analisados e tipos específicos de complexidade manifestada nestes. Nesta dissertação, são apresentados e discutidos a metodologia de desenvolvimento e os resultados desta infraestrutura de complexidade.

PALAVRAS-CHAVE: Complexidade textual, literacia, proficiência em leitura

ABSTRACT

TEXTUAL COMPLEXITY FOR LOW LITERACY ADULTS: ANALYSIS OF COMPLEXITY LEVELS AND THEIR CHARACTERISTICS

This work aims to describe the process behind the creation of an infrastructure of textual complexity. Framed within the iRead4Skills project, this infrastructure looks to detect and classify the complexity embedded in texts, namely that which manifests for low literacy adults. Although various benchmarks of reading proficiency (CEFR, PISA, etc.) exist, little attention is given to the difficulties felt by adult learners and how these difficulties manifest linguistically in texts. Based on already existing frameworks (ANQEP, CEFR, PIAAC, ALTE), this work approaches this phenomenon through the analysis of levels and their descriptors within these frameworks in order to propose new levels pertinent for the case at hand. Structured in three different levels, our proposal considers linguistic categories like the lexicon-conceptual dimension, syntactic structures and text cohesion, for each level. These categories are explained through descriptors representative of textual complexity. These same levels and descriptors are put through a validation process by a focus group composed of trainers from adult learning centers. Alongside this process, these levels and descriptors served as well in the compilation of a *corpus* through the validation of texts according to the framework of complexity proposed here. Lastly, the application of this complexity infrastructure is put in place with the particular aim of revealing a link between a group of texts and specific types of complexity they manifest. In this work, the results and methodology behind the development of this complexity infrastructure are presented and detailed.

KEYWORDS: Textual complexity, literacy, reading proficiency

ÍNDICE

1. INTRODUÇÃO.....	1
2. ENQUADRAMENTO TEÓRICO.....	3
2.1. Literacia.....	3
2.1.1. Oral vs. Escrito.....	5
2.1.2. Literacia como um conjunto de aptidões cognitivas	6
2.1.3. Literacia como um fenómeno social	9
2.1.4. Literacia no texto.....	11
2.1.5. Adultos de baixa literacia	12
2.1.6. Literacia e a formação para adultos	15
2.2. Complexidade Textual	18
2.2.1. A tarefa do leitor	19
2.2.2. Dimensão quantitativa	20
2.2.3. Dimensão qualitativa.....	22
2.2.4. O uso <i>in tandem</i>	25
3. METODOLOGIA	27
3.1. Níveis de complexidade e proficiência de literacia e as suas características 28	
3.1.1. Levantamento e análise de níveis e descritores de proficiência em leitura.....	28
3.1.2. Proposta de base adequada a adultos de baixa literacia	29
3.1.3. Os níveis e os descritores de complexidade textual	30
3.2. Constituição e validação do <i>corpus</i>	31
3.2.1. Compilação do <i>corpus</i> – critérios de seleção e amostragem.....	31
3.2.2. Recolha dos textos por nível	33
3.2.3. Validação do <i>corpus</i>	35
3.3. Validação dos níveis e descritores de complexidade por formadores de 37	
adultos	37
3.4. Proposta final de complexidade textual para adultos de baixa literacia	38
3.5. Aplicação da infraestrutura de complexidade numa amostra do <i>corpus</i>	38

4. RESULTADOS E DISCUSSÃO	39
4.1. Infraestrutura de complexidade	39
4.1.1. Os níveis de complexidade	40
4.1.2. Os descritores de complexidade textual	41
4.2. A validação do <i>focus group</i>	45
4.3. O <i>Corpus</i> – teste e validação dos níveis e descritores de complexidade	47
4.4. A aplicação da infraestrutura de complexidade	49
5. CONCLUSÃO	58
REFERÊNCIAS	61
Anexo I – Descritores por nível	69
Anexo II – Descrição dos níveis de complexidade	80
Anexo III – Tipologia de textos	84
Anexo IV – Tarefa preparatória para o <i>focus group</i>	86

ÍNDICE DE FIGURAS

Figura 1. Esquema dos componentes da infraestrutura de complexidade textual	28
Figura 2. Exemplo de codificação do documento	34
Figura 3. Documento de validação	36

ÍNDICE DE TABELAS

Tabela 1. Descritores que mudaram de nível	46
Tabela 2. Distribuição dos textos por nível de complexidade	48
Tabela 3. Amostra utilizada na aplicação da infraestrutura de complexidade	49
Tabela 4. Resultados da análise de manifestações de complexidade por descritor	51

LISTA DE ABREVIATURAS

ALTE	Associação de Examinadores de Línguas na Europa
ANQEP	Agência Nacional para a Qualificação e o Ensino Profissional
CA	Cursos de Aprendizagem
CEF	Cursos de Educação e Formação
CP	Cursos Profissionais
EFA	Educação e Formação de Adultos
QECR	Quadro Europeu Comum de Referência para as Línguas
OECD	Organização para a Cooperação e Desenvolvimento Económico
PIAAC	Program for the International Assessment of Adult Competencies
PLN	Processamento de Língua Natural

1. INTRODUÇÃO

Uma boa proficiência de leitura é essencial para os indivíduos prosperarem no mundo atual. A literacia afigura-se, cada vez mais, vital para o desenvolvimento e concretização dos nossos objetivos. Na mesma senda, altos níveis de literacia estão correlacionados com outros benefícios na vida dos indivíduos, desde ocupações profissionais mais bem remuneradas até a melhores hábitos de saúde (OECD, 2013a).

Com efeito, as questões à volta da literacia na sociedade são cada vez mais discutidas e reconhecidas como importantes para uma vida digna. Martins (2010) realça a literacia como uma prática social, na qual os falantes de uma língua utilizam as suas aptidões de leitura e escrita em contextos necessários para o seu desenvolvimento pessoal e profissional. A Organização para a Cooperação e Desenvolvimento Económico (OECD) salienta a relação entre baixos níveis de proficiência em literacia e uma população adulta desempregada (OECD, 2013a).

Para além disto, a noção de literacia tem vindo a complexificar-se desde as suas primeiras menções até hoje, tendo partido duma relação estreita com alfabetização até ao desdobramento em vários conceitos relacionados com âmbitos específicos da sociedade (Leal 2018). De igual forma, também o processo que leva um indivíduo a possuir literacia parece ser igualmente complexo, afigurando-se cada vez mais como um processo do que um estado, isto é, uma meta à qual se tenta chegar.

Esta ideia da literacia como situando-se numa escala de proficiência, variando de indivíduo para indivíduo, tem vindo a abrir o espaço para a discussão do melhoramento de literacia em adultos. À vista disto, vários programas de aprendizagem de adultos, tanto a nível internacional como nacional, têm vindo a ser implementados. Com a meta de providenciar oportunidades educativas, por vezes através do contexto de formações profissionais, os cursos de Educação e Formação de Adultos (EFA) têm o objetivo de melhorar as qualificações de indivíduos com uma escolaridade baixa, adaptando-os às necessidades da sociedade e enquadrando-os no contexto de aprendizagem ao longo da vida. Contudo, encontrar materiais de leitura adequados para este tipo de população – especialmente aqueles que são autênticos e circulam na sociedade – pode ser uma tarefa difícil.

Este contexto está na origem do estabelecimento do projeto iRead4Skills ¹. Reconhecendo a importância de práticas de leitura frequentes, este projeto, iniciado em 2023, tem como meta a promoção do desenvolvimento de práticas de leitura através da criação de um sistema que avalia a complexidade textual e sugere textos adequados a leitores com baixa literacia. O intuito é que este acesso a textos adequados proporcione uma reaproximação dos indivíduos com práticas de leitura, o que, por sua vez, estimulará o desenvolvimento das mesmas.

De modo a alcançar este fim, vários passos são necessários dar. Desde o abalçamento do público-alvo do projeto e as suas necessidades até ao desenvolvimento do sistema de Processamento de Língua Natural (PLN), o projeto subdivide-se em várias metas. Uma destas é a constituição de uma infraestrutura de complexidade textual, a qual funciona como pedra basilar do projeto e dos seus objetivos. Esta dissertação tem como meta apresentar e detalhar a criação desta infraestrutura de complexidade textual, desenvolvida no âmbito do projeto iRead4Skills.

Primeiramente, far-se-á um enquadramento teórico relativo às noções de literacia e complexidade textual, com o propósito de melhor compreender a literacia e como esta pode ser aferida, bem como as noções em redor da complexidade textual e como esta se pode manifestar nos textos.

No terceiro capítulo, expor-se-á a metodologia por detrás da infraestrutura de complexidade. Os objetivos, motivações e procedimentos utilizados serão desenvolvidos nesta secção, bem como a maneira como todos estes se entrosam e complementam.

De seguida, os resultados nos quais a criação desta infraestrutura se alicerça serão apresentados e discutidos. Em primeiro lugar, os dados provenientes de cada passo metodológico serão exibidos, depois analisados e, de seguida, discutidos e relacionados. Por último, uma pequena demonstração da infraestrutura de complexidade e a sua versatilidade de resultados é ensaiada.

¹ The iRead4Skills – Intelligent Reading Improvement System for Fundamental and Transversal Skills Development. Grant number: 1010094837, Topic: HORIZON-CL2-2022-TRANSFORMATIONS-01-07 – Conditions for the successful development of skills matched to needs. <https://iread4skills.com/>

No último capítulo, apresentar-se-ão as conclusões sobre este trabalho, assim como algumas considerações finais sobre a temática.

2. ENQUADRAMENTO TEÓRICO

2.1. Literacia

Nos últimos anos, o conceito de literacia tem verificado um uso cada vez proeminente na sociedade. Seja pela sua natureza desdobrável e produtiva, a qual é responsável por conceitos paralelos como a *literacia digital* ou *financeira*, seja pelo reconhecimento da sua importância no desenvolvimento do indivíduo e da comunidade, a verdade é que o debate sobre a literacia chegou para ficar. A literacia revela-se como fulcral no acesso à informação e ao conhecimento, na possibilidade de aprender ao longo da vida, bem como no exercício de uma análise crítica do mundo (Ávila, 2008). No entanto, do que falamos quando falamos de literacia? Como sabemos se somos detentores de literacia ou não? Como aumentar os nossos níveis de literacia? Não só as perguntas sobre o tema são múltiplas, mas também as suas respetivas respostas.

O termo *literacia* – como o conhecemos hoje – é introduzido no léxico português aquando da divulgação do *Estudo Nacional de Literacia. Relatório Preliminar português*, em 1995² (Pinto, 2002:96). Apropriando a palavra e o conceito de *literacy* – utilizado em países anglo-saxónicos desde final do séc. XIX – a noção é primeiramente associada com o conceito de *alfabetização*. Com efeito, nos primórdios do estudo da literacia, estes dois termos eram frequentemente utilizados com o mesmo significado (Martins, 2010). E ainda que este entendimento seja, hoje em dia, considerado obsoleto, a verdade é que é um excelente ponto de partida para abordar este tema.

A ideia da literacia associada ao conceito de alfabetização sugere-nos dois pontos interessantes: o primeiro, que a distinção entre *saber ler*, por um lado, e *compreender* o que se lê, por outro, não era considerada, ou seja, assumia-se que “juntar letras” e reconhecê-las – isto é, descodificar o princípio alfabético – era sinónimo da sua

² Benavente, A. (coord.); Rosa, A.; Costa, A.; Ávila, P. (1995). Estudo Nacional de Literacia: Relatório Preliminar. Lisboa: Instituto de Ciências Sociais - Universidade de Lisboa.

compreensão por parte do leitor. Como já foi dito, esta ideia foi abandonada em lugar de uma outra que considera a literacia como muito mais do que apenas a capacidade de leitura. Ainda que um consenso sobre uma definição da literacia seja difícil de alcançar, de forma sucinta, um entendimento da literacia pressupõe-na como “a capacidade de pôr em ação conhecimento, aptidões, atitudes e valores, aquando do manuseio de um texto” (Hanemann, 2012:15). Esta definição parece ir de encontro ao entendimento da UNESCO:

Literacy is the ability to identify, understand, interpret, create, communicate and compute, using printed and written materials associated with varying contexts. Literacy involves a continuum of learning in enabling individuals to achieve their goals, to develop their knowledge and potential, and to participate fully in their community and wider society. (UNESCO, 2020, Secção de Definição)

O Program for the International Assessment of Adult Competencies (PIAAC) parece ser do mesmo entendimento, tanto na definição de 2013, como na mais recente, de 2021:

Literacy is understanding, evaluating, using and engaging with written texts to participate in society, to achieve one’s goals, and to develop one’s knowledge and potential. OECD (2013b:19)

Literacy is accessing, understanding, evaluating and reflecting on written texts in order to achieve one’s goals, to develop one’s knowledge and potential and to participate in society. OECD (2021: 42)

Assim, a literacia evidencia-se como mais do que uma aptidão específica, que se tem ou não, mas como uma competência, de natureza contínua, na qual se é mais ou menos proficiente. Para além disso, não parece haver uma correspondência entre a escolaridade de uma população e o seu perfil de literacia (Scribner & Cole, 1981; Rothes & Queirós, 2021). Uma sociedade altamente escolarizada não é, necessária e obrigatoriamente, uma sociedade com mais competências literácitas. Para compreender

esta competência nos indivíduos, é necessário compreender igualmente como estes usam as suas aptidões de leitura e escrita enquanto participantes ativos da sociedade.

Algo igualmente importante a considerar, quando se fala de literacia, é o carácter descritivo, ao invés de normativo, das suas definições (Rothés & Queirós, 2021). Entender a literacia não como uma meta ou estado ideal, mas como um *continuum*, que pode variar em cada indivíduo, é crucial para a sua compreensão e avaliação. Nesta senda, a UNESCO, na sua definição de literacia, considera-a como um *continuum* de proficiência, ou seja, uma competência que está em constante aprimoramento. Esta ideia, reiterada por Hanemann (2012), pressupõe algo bastante importante sobre esta competência que é o facto de estar em constante evolução, tanto no seio de cada indivíduo, tornando possível a ideia de que alguém que possui baixa literacia atualmente se pode tornar em alguém de alta literacia no futuro, como também em relação ao que realmente significa ser letrado. À vista disto, a literacia configura-se cada vez mais como um processo do que um estado, isto é, um “caminho em direção a um objetivo” (Alves, 2005). No entanto, é difícil dizer se este objetivo é alguma vez atingido na íntegra e, consequentemente, se o processo chega a um ponto de culminação, estando a meta deste processo em constante movimento, devido aos avanços e mudanças dentro da sociedade. Esta ideia de *continuum* parece ser, assim, bivalente, no sentido em que não só se deslocam os indivíduos no *continuum*, como também se desloca o próprio *continuum*.

2.1.1. Oral vs. Escrito

Algo que parece também estar contido nestas definições é a ideia de que a literacia se aplica a *textos escritos*. Do ponto de vista pragmático, existem diferenças entre a linguagem oral e escrita, sendo a comunicação oral pautada por uma maior clareza de mensagem (Carvalho, 2003). Isto deve-se ao facto de, na maioria das vezes, o emissor e o recetor das mensagens se encontrarem no mesmo espaço, com acesso ao mesmo contexto e informação prévia. Por outro lado, a linguagem escrita objetiva “pintar uma tela em branco”, isto é, transmitir uma mensagem do modo menos ambíguo possível,

sem ser redundante, considerando contextos distintos entre interlocutores³. Os indivíduos necessitam, assim, de invocar imagens mentais, de estruturar as mesmas, assim como colocá-las dentro de noções espaciotemporais. Em contraste com a linguagem oral, estes processos exigem um esforço cognitivo maior, para além de suscitem, nos indivíduos, os seus conhecimentos linguísticos.

Desta forma, não é estranho notar a correlação positiva entre níveis mais elevados de proficiência em literacia e práticas de leitura frequentes (Roths & Queirós, 2021, baseados nos resultados do *Survey of Adult Skills*- PIAAC (OECD, 2019a). Paralelamente, indivíduos com níveis baixos de literacia apresentam práticas de leitura escassas, muitos deles não lendo no dia a dia. Tal como na maioria das aptidões, o uso e a proficiência reforçam-se mutuamente.

Tudo isto parece apontar para uma relação forte entre os textos escritos e a competência de literacia, a qual é postulada por Fransman (2005), quando este refere a panóplia do que considera como “literacias não textuais”, como literacia financeira, literacia digital, literacia cultural, etc. Assim, apesar da multimodalidade e dinâmica do conceito da literacia nos últimos anos, é possível cingir a noção de literacia às práticas de comunicação que se manifestam em torno de textos escritos (em formato manuscrito, impresso ou digital) (Hanemann, 2012).

2.1.2. Literacia como um conjunto de aptidões cognitivas

A ideia primitiva e obsoleta, que encetou a nossa discussão, de que a literacia era o mesmo que alfabetização, pressupunha que não havia distinção entre descodificação e compreensão. Como vimos acima, essa ideia é hoje recusada. No entanto, uma particularidade inerente a este pensamento é a de que a proficiência em literacia está vinculada ao indivíduo, isto é, a um conjunto de aptidões do indivíduo como leitor (Wagner, 1983). Esta perspetiva de literacia é normalmente apelidada de *perspetiva individual*, e considera a literacia essencialmente como um atributo pessoal,

³ Existem exceções a esta clareza vinculada à linguagem escrita, como é o caso de alguns géneros literários como a poesia, no qual se pode alegar que existe uma subjetividade de conteúdo. Contudo, entende-se que nestes casos, estamos perante manifestações que têm como objetivo contrariar conscientemente a nitidez intrínseca da linguagem escrita.

diretamente ligada às habilidades de cada indivíduo. Esta abordagem à noção de literacia – também designada por modelo *autónomo* (Fransman, 2005) –, considera que o ato interpretativo de qualquer tipo de texto está exclusivamente associado a capacidades cognitivas (Dionísio *et al.*, 2014). Desta forma, a capacidade de decifração de textos escritos, ou seja, a proficiência em literacia de cada leitor, é entendido como algo técnico, resultante de um conjunto de aptidões específicas de cada leitor.

Esta perspetiva encontra no campo da neuropsicologia cognitiva o seu principal sustento. A ideia aqui subjacente é a de que a aprendizagem da leitura e da escrita alteram a forma como o cérebro se organiza (Castro-Caldas & Reis, 2000). As atividades cognitivas desencadeadas durante o processo de alfabetização conduzem ao desenvolvimento de novas estruturas cerebrais e criam uma “ponte” entre o sistema visual e o tratamento de informação. Para além disto, este processo de alfabetização parece responsável por desenvolver componentes linguísticas nos indivíduos – para além da lexical – dotando-os da capacidade de criarem enunciados espontâneos (oralmente e por escrito), sem se alicerçarem numa compreensão semântica ou lexical dos enunciados em si (Castro-Caldas & Reis, 2000).

Neste sentido, leitores iletrados tendem a ter uma grande panóplia de problemas nos processamentos sintáticos e morfológicos, em suma, problemas gramaticais (Perfetti *et al.*, 2001). A literacia parece ser capaz de influenciar o desenvolvimento do cérebro, no que diz respeito à sua capacidade de interagir com o ambiente, assim como de modular o processamento de língua oral (Petersson *et al.*, 2009). Por outro lado, também o processamento fonológico no indivíduo letrado, assim como a sua consciência fonológica, é mais desenvolvido do que no indivíduo iletrado (Huettig, 2015; Perfetti *et al.*, 2001; Petersson *et al.*, 2009). Contudo, o grande problema inerente a estas hipóteses que consideram as aptidões psico-cognitivas como cruciais no âmbito da literacia é o fundamento sobre as quais operam. Todas elas têm como axioma o entendimento da literacia como a capacidade de ler e escrever, considerando assim não a proficiência de leitura, mas a dicotomia entre alfabetismo e analfabetismo. Um entendimento para além da noção “alfabetismo *versus* analfabetismo” é algo ainda subexplorado. No

entanto, uma correlação entre a forte capacidade de processamento linguístico com leitores mais desenvolvidos e maduros parece existir (Huettig, 2015). Neste âmbito, é proposto que um processamento de língua sem esforço está associado a uma boa capacidade de antecipação das manifestações linguísticas seguintes, em forma escrita. Vários estudos foram realizados sobre esta questão (Costa *et al.*, 2009; Fernandez-Duque, 2009; Mishra *et al.*, 2012).

De qualquer forma, um dos aspetos relevantes na perceção da literacia como assentando em características psico-cognitivas é o facto de essa visão assumir que a literacia é mensurável. Porém, encarar a literacia desta forma é uma tarefa delicada, uma vez que existem múltiplas variáveis nestes processos a serem consideradas (Martins, 2010).

Talvez a mais importante destas variáveis a ter em conta, quando adotada esta abordagem à literacia, é a distinção entre escrita e leitura. Embora as duas sejam agrupadas frequentemente, parece haver diferenças fundamentais entre estes dois processos (Soares, 2002).

No processo de leitura, existem competências linguísticas e cognitivas em execução, desde a descodificação da linguagem até a extração de sentido, através da interpretação de sequências no texto, da comparação e contraste das mesmas, da relação entre eventos e entidades do texto. Por outro lado, no processo de escrita, aptidões linguísticas e cognitivas diferentes estão em jogo, nomeadamente competências de codificação de língua, de ortografia, de uso adequado de pontuação, mas também de organização e expressão de ideias (Soares, 2002).

Curiosamente, ainda que se trate de processos distintos, a leitura parece ser o processo pelo qual a própria escrita é amadurecida. O contacto frequente com a leitura fomenta a proficiência da expressão escrita, através do relacionamento dos leitores com os diversos modos como a linguagem escrita se apresenta, conduzindo à consequente interiorização destas manifestações (Ferreiro & Teberosky, 1991). Desta forma, evidencia-se ainda mais a importância da leitura no desenvolvimento da linguagem escrita. Isto posto, o processo da leitura – e as aptidões linguísticas e cognitivas

encerrado nele –, parece ser aquele a ter em consideração aquando da consideração da literacia como algo que se pode manifestar em aptidões individuais.

No entanto, e ainda que esta visão da literacia como uma competência individual seja a abordagem prevalente em Portugal (Alves *et al.*, 2012), bem como na maioria dos países onde esta questão é discutida, muito em razão das vantagens que apresenta para a monitorização dos níveis de literacia (Fransman, 2005), a verdade é que outros entendimentos sobre a literacia se lhe opõem.

2.1.3. Literacia como um fenómeno social

Ainda que fascinante, a perspectiva da literacia como um fenómeno exclusivamente individual comporta um problema – o facto de esta não tomar em consideração todo o contexto no qual a literacia opera (Street, 2005). A noção de literacia não é passível de ser dependente exclusivamente de valências individuais do leitor, dado estar intrinsecamente ligada a fenómenos culturais e ideológicos que a modelam (Street, 2005). Esta visão da literacia é denominada como o *modelo ideológico* ou social.

Examinar a literacia como um fenómeno que vai para lá do indivíduo pressupõe que esta é mediada por textos em práticas sociais, contextualmente e temporalmente situadas (Alves *et al.*, 2012). Por outras palavras, o entendimento da literacia como um fenómeno social admite o uso desta em situações quotidianas, aquando da interação dos indivíduos com os diversos contextos sociais. Estas interações são designadas por *literacy events* (Fransman, 2005). Nestes, a interação do indivíduo com uma prática social é feita através de um texto, e, de forma reflexa, a interação com o texto é feita no âmbito da prática social. Por conseguinte, os usos e práticas dos textos não apresentam sempre as mesmas características, variando conforme as esferas de vida (Dionísio *et al.*, 2014). Desta maneira, torna-se claro como a dimensão social influencia a sua interação com o texto, sendo a proficiência em literacia de cada indivíduo dependente do contexto social no qual é posto em prática.

A ideia da literacia como um fenómeno social nasce no início do milénio e é apelidada como o movimento NLS – *New Literacy Studies* (Street, 1993; Fransman, 2005; Ávila, 2008). Esta escola de pensamento não só é responsável por repensar o conceito de

literacia, mas é também responsável pela sua aferição, já que, ao considerar o contexto na qual a literacia é observada, a apreciação do que é um indivíduo com mais ou menos competência de literacia está inteiramente dependente desse mesmo contexto. Reconsiderando as noções mencionadas atrás de *práticas sociais contextualmente e temporalmente situadas*, não é errado presumir que alguém que viveu há um século atrás, considerado com certo nível de proficiência em literacia então, seria hoje em dia considerado de maneira diferente. O mesmo se estende a culturas e contextos sociais. Para além disto, para apelidar um indivíduo de letrado ou iletrado, várias são as dimensões que podem ser consideradas, para além da sua competência de escrita e leitura (Fransman, 2005), nomeadamente:

- i. os processos que levaram os indivíduos a adquirir as suas competências de leitura e escrita (escola, programas de ensino de adultos, etc.);
- ii. os propósitos que dão a estas competências (argumentação, comunicação de notícias, etc.);
- iii. os contextos nos quais põem em práticas as suas competências (familiares, profissionais, formais).

Todas estas questões acrescentam e condicionam a respetiva proficiência de literacia de cada um. Fatores como a idade do indivíduo, a sua escolaridade, o seu sexo, apresentam resultados diferentes no que diz respeito ao seu peso na proficiência em literacia de cada um. Ainda que, em certos países e culturas, estes mostrem uma tendência para ser responsáveis por níveis de proficiência mais baixos nos indivíduos, noutros estão associados a níveis altos de literacia. Isto parece indicar que não são exatamente estes aspetos que ditam a proficiência de cada um, mas a forma como as respetivas sociedades encaram esses atributos e, por consequência, como estes pesam dentro de cada uma (Roths & Queirós, 2021, baseados nos resultados do *Survey of Adult Skills- PIAAC* (OECD, 2019a).

Com isto em mente, torna-se evidente que a literacia é mais do que um conjunto de competências de escrita e leitura. A literacia é, sim, o uso dessas habilidades para atender a exigências sociais. Esta visão pragmática é construída em cima do conceito da literacia como um conjunto de aptidões do indivíduo. Com efeito, a perspetiva social da

literacia coexiste com a perspectiva individual, embora a primeira atente ao que as pessoas fazem com estas aptidões e como se relacionam com os princípios e práticas sociais. O entendimento da literacia como um fenómeno social destaca-se por rejeitar uma visão da literacia que se baseia exclusivamente em aptidões psico-cognitivas (Dionísio *et al.*, 2014). Assim, o modelo social ou ideológico não procura refutar a conceção da literacia como aptidões cognitivas ligadas à escrita e à leitura, mas sim, acima de tudo, enquadrá-la culturalmente (Street, 1993).

2.1.4. Literacia no texto

Uma ideia que também merece consideração neste trabalho é o do entendimento da literacia como estando ligado simbioticamente aos textos em si (Fransman, 2005). Esta hipótese destaca o suposto poder dos textos em posicionar os indivíduos na sociedade e assenta sobre a ideia abrangente do construtivismo da linguagem, a qual propõe que esta é um produto de constantes construções individuais e sociais (Philpo, 2023). De acordo com este ponto de vista vygotskyano, a linguagem é o produto de construções históricas e sociais, as quais perpetuam o *status quo* embutido nas normas sociais e culturais a ela associadas. Desta forma, quando uma criança aprende a sua língua, esta está, ao mesmo tempo, a aprender todas estas normas e convenções socio históricas embutidas nela. O mesmo acontece com um leitor aquando da sua leitura dos textos. Assim, visto serem o objeto concreto e manifestável do estudo da literacia, os textos contêm o poder de reproduzir, cancelar, legitimar ou condenar as práticas sociais neles patentes, como por exemplo, ideias classistas, racistas ou de desigualdade de género.

Esta ideia pode ser associada à teoria da semiótica social, desenvolvida por Kress (2003), no sentido em que esta refere como a comunicação na sociedade se adapta conforme o contexto e objetivo desta mesma, dando aso à criação de texto *multimodais*. Estes utilizam imagens ou outros elementos semióticos na sua comunicação de significado, os quais, por sua vez, requerem uma compreensão que combine estes elementos não verbais com os elementos verbais, por parte do leitor. Esta aptidão é designada de *multiliteracia* (Leal, 2018). Esta característica da multimodalidade de

certos textos é a prova da sua interligação com a ideia construtivismo social, que, com vista a representar o significado da maneira mais eficaz possível, é responsável por mudar a natureza monomodal dos textos para uma que faz uso de outros elementos não verbais.

A noção basilar a todos estes conceitos é de que os textos carregam uma força escondida. A literacia é aqui equacionada relativamente aos processos de linguagem que são desencadeados aquando da escrita ou leitura de um texto. Com efeito, estes processos textuais carregam consigo relações de poder em ação no mundo em que vivemos (Fransman, 2005). Desta forma, o entendimento do que é a literacia e, por consequência, uma pessoa mais ou menos letrada, estará sempre subordinado ao *status quo*, levando à conclusão de que a literacia não é a *resposta* ao mundo em que vivemos, mas sim o seu acomodamento.

Se é verdade que as abordagens ao que é a literacia e como esta pode ser definida são múltiplas, também é verdade que a sua estrita definição estanque pouco afeta quem realmente é condicionado por estas questões. Neste sentido, uma definição bastante restrita não é possível, nem desejável (Yousif, 2003). Proveitoso será combinar as diferentes visões da literacia, cada uma com o seu fundamento, para a construção de um conceito abrangente capaz de ser utilizado na implementação de programas de literacia.

2.1.5. Adultos de baixa literacia

Uma das ideias mais importantes e que mais revolucionou o conceito de literacia desde a sua criação é o reconhecimento que não existe meta para esta competência. Esta noção da literacia como um *continuum* de proficiência é adicionada à definição de literacia da UNESCO em 2009, quando a ideia de literacia passa também a ser considerada como uma competência indispensável e ao alcance de jovens e adultos, independentemente da sua posição no *continuum* (Wagner, 2011). Para além disso, começa igualmente a ser postulado o seu papel como pré-requisito no desenvolvimento pessoal, social, económico e político de cada um.

Com efeito, a literacia revela uma transversalidade ímpar, uma vez que é o meio pelo qual toda a aprendizagem se dá, considerando que todo o processo de aprendizagem é mediado por práticas da linguagem (Cunha & Dionísio, 2019). Tudo isto parece apontar para a literacia como uma competência que não só permite a compreensão dos textos, da sociedade, do mundo, sem que este último seja barrado ao indivíduo, mas, também, a capacidade de melhoramento de outras competências a partir dela.

Deste modo, o conceito de “aprendizagem ao longo da vida” surge, considerando a não existência de um patamar que impossibilite a um adulto a sua melhoria de competências. Noções como “educação e aprendizagem ao longo da vida”, “educação de adultos” e “aprendizagem de adultos”, parecem todas encerradas sob a definição de *lifelong learning*. Com efeito, a Comissão das Comunidades Europeias (2006) define a questão da educação de adultos como compreendendo todas as formas de aprendizagem postas em prática por adultos após o abandono da educação inicial e da formação, independentemente do nível que tenham atingido (pode, por exemplo, incluir o ensino superior).

Seguindo a prática começada por o “Young Adult Literacy Survey” (1985)⁴ e o “National Adult Literacy Survey” (NALS, 1992)⁵, e Estudo Nacional de Literacia (1995) (a nível nacional) mais recentemente, o *Survey of Adult Skills* (PIAAC) (no qual Portugal também participa) tem tido destaque como a principal ferramenta de avaliação de literacia em adultos (entre outros aspetos). É com os resultados de estudos como este em mente que a literacia tem vindo a ser entendida como algo muito mais consequente e abrangente do que antigamente. Numa análise de vários países, Schleicher (2012, baseado nos resultados do *Survey of Adult Skills-PIAAC* (OECD, 2013a), constata que, em média, indivíduos com níveis de literacia considerados altos têm salários 60% superiores a indivíduos do lado oposto do espectro, isto é, com níveis de literacia considerados baixos. Para além disso, comparando estas duas populações no que diz respeito ao nível profissional, aqueles com proficiência baixa em literacia têm o dobro da probabilidade de se encontrarem desempregados (Schleicher, 2012). Ademais, as consequências dos níveis de literacia noutros aspetos, à primeira vista, sem relação, parecem não ficar por

⁴ O YALS baseou-se em entrevistas a jovens com idades compreendidas entre os 21 e os 25 anos (Kirsch, 2001).

⁵ O NALS ocupava-se de uma população com idade igual ou superior a 16 anos.

aqui: uma população com altos níveis de literacia goza de um maior poder de intervenção política⁶, de uma maior participação em atividades de voluntariado, bem como de uma saúde boa ou excelente (Schleicher, 2012). Com efeito, baixos índices de literacia são frequentemente associados a uma disfuncionalidade profissional e a *deficits* sociais.

O entendimento parece ser consensual: uma forte competência de literacia (assim como numeracia) tem o poder de melhorar a qualidade de vida dos indivíduos, ajudar na concretização dos seus objetivos, assim como proporcionar outras oportunidades de aprendizagem.

Literacy and numeracy are crucial components of a set of essential, foundational or general skills and competencies for the 21st century. These are essential for performing daily tasks, leading healthy lives, finding work, engaging in social and political activities, and independent learning. Strong literacy skills are associated with a range of valuable and desirable development outcomes including access to decent work. (Hanemann, 2012: 14)

A verdade é que, como temos vindo a ver, a importância dada à abrangência da literacia e à vasta gama de consequências que esta tem na vida das pessoas é algo relativamente recente. A noção de que esta competência potencia os indivíduos e, como consequência, as sociedades, era algo completamente rejeitado, por exemplo, no século XVIII, numa Inglaterra em processo de industrialização, quando se considerava que nenhum dos grandiosos avanços da altura se devia à mão de obra letrada e que muito menos a fomentava (Graff, 1987). Nos dias que correm, o impacto de uma sociedade com alta proficiência em literacia é visível ao nível socioeconómico e na competitividade entre países⁷ (Cunha & Dionísio, 2019). Nesta senda, múltiplas campanhas de literacia por parte da UNESCO têm estabelecido a ligação do indicador mensurável da literacia a

⁶ Isto é, uma população com uma competência de literacia alta julga-se mais capaz intelectualmente, assim como mais informada, para participar ativamente e de uma forma não coagida em atos de decisão política.

⁷ A relação entre literacia e desenvolvimento é algo especialmente enraizado nas sociedades ocidentais (Cunha e Dionísio, 2019).

outros indicadores de desenvolvimento, como, por exemplo, de economia ou de modernização (Wagner, 2011).

Em síntese, torna-se aparente que, hoje em dia, não só é a educação vista como um processo em curso ao longo da vida, o qual deve ser fomentado e ambicionado pelos indivíduos, mas também que a ênfase colocada outrora na educação formal escolar e a sua exclusividade como albergadora dos momentos únicos propícios para a educação tem cada vez mais os dias contados.

2.1.6. Literacia e a formação para adultos

O entendimento de que uma boa competência de literacia é acompanhada de outros benefícios que, em geral, elevam a qualidade de vida de um indivíduo é crucial para reconhecer o poder da literacia. No entanto, mais importante ainda é considerar que a falta de literacia leva a consequências nefastas.

Nos países desenvolvidos, os empregos que requerem elevadas qualificações multiplicam-se (Ávila, 2008). Contudo, a mão de obra não acompanha esta mesma evolução, assim se evidenciando um desfasamento entre as duas. Desta forma, indivíduos possuidores de baixos níveis de literacia (e por consequência, baixos níveis de qualificações), estão cada vez mais vulneráveis à exclusão social, revelando-se este fenómeno como um dos principais responsáveis por clivagens sociais, hoje em dia. Em Portugal, os trabalhadores que apresentam níveis baixos de literacia estão especialmente vulneráveis a situações de desemprego (Dionísio *et al.*, 2014). Ademais, esta população parece estar mais alienada de oportunidades profissionais e de educação, as quais serviriam os seus interesses (Schleicher, 2012).

Para além desta vulnerabilidade profissional, também a perceção que esta população tem de si própria parece estar carregada de negatividade. Indivíduos com uma baixa proficiência em literacia tendem a olhar para si próprios como diferentes do resto da população (Dionísio *et al.*, 2014). Desta forma, esta população não considera as suas próprias práticas de leitura e escrita legítimas aos olhos dos “outros” (Alves *et al.*, 2012). Esta autoexclusão não só é prejudicial para a saúde mental destes indivíduos, mas revela-se também como um dos grandes entraves ao desenvolvimento de cada um, já

que fomenta uma situação duplamente adversa, na qual o indivíduo está “refém” da sua falta de literacia (Banda & Mpolomoka, 2012). Curiosamente, a qualidade de “círculo virtuoso” vinculado à literacia (Rothés & Queirós, 2021), na qual uma maior proficiência gera maior engajamento com práticas de literacia e, por consequência, maior acentuação de competências transversais, também pode ser visto pelo prisma oposto, onde uma proficiência deficiente de literacia resulta num afastamento de práticas de literacia, assim como todas os malefícios associados a estas.

Frequentemente, a conjuntura é o “pecado original”. Proveniente de um estrato socioeconómico baixo e/ou com dificuldades de acesso ao ensino e formação, esta população parece estar exposta a tradições culturais que desvalorizam a educação, tradições culturais estas que privilegiam, por exemplo, uma carreira laboral em detrimento de uma educação para lá da básica (Dionísio *et al.*, 2014).

Com vista a contrariar estes “futuros amaldiçoados”, vários programas de incentivo à requalificação pessoal e profissional têm sido postos em prática, tanto a nível internacional como nacional⁸. Estes programas baseiam-se em valores humanísticos, democráticos e emancipatórios e têm como objetivo requalificar e integrar esta população menos letrada (Hanemann, 2012). Através do ensino formal e informal, para todas as faixas etárias e com a maior abrangência de contextos possível, propõem-se aos formandos oportunidades de aprendizagem, seguindo o princípio que tem vindo a ser mencionado aqui: que a competência de literacia de cada um pode ser aprimorada, independentemente de fatores externos.

Os cursos EFA, a nível nacional, são excelentes exemplos destes programas. De acordo com a Portaria nº86/2022 de 4 de fevereiro, estes cursos têm como principais objetivos:

- a) permitir o acesso e a melhoria das qualificações dos adultos, nomeadamente os que não têm o ensino secundário;
- b) constituírem -se como percursos flexíveis e adaptados às necessidades dos

⁸ Destacam-se ao nível nacional: Reconhecimento, Validação e Certificação de Competências Profissionais (RVCC Profissional) proporcionados por Centros Qualifica, (CET) Especialização tecnológica (CET), (PCDI) Qualificação de pessoas com deficiência e incapacidade e PRO_MOV Portugal, este último patroneado pela entidade internacional Reskilling 4 Employment (R4E).

adultos num contexto de aprendizagem ao longo da vida;

Formações como CA (Cursos de Aprendizagem), CP (Cursos Profissionais), CEF (Cursos de Educação e Formação), têm, muitas vezes, como metas tangíveis, a aquisição de certos conhecimentos para o exercício de uma profissão ou para a reinserção no mercado de trabalho, o desenvolvimento de uma identidade cívica e a promoção de um pensamento crítico e comunitário no formando, por vezes, todas ao mesmo tempo. Contudo, para estas metas serem atingidas (ou algumas delas), será necessário, não só a disponibilidade dos formandos, mas também a adaptação dos programas às necessidades dos formandos (Cunha & Dionísio, 2019). Nos dias de hoje, o entendimento parece ser o de que o sucesso destes programas está dependente do grau de adequação dos programas aos seus participantes. Com a meta de englobar a grande diversidade de contextos sociais dos formandos, estes programas apresentam uma grande variedade de atividades, módulos e textos, reconhecendo costumes e comunidades fora daquelas consideradas habituais (Cunha & Dionísio, 2019).

Bastante alicerçada na perspetiva social da literacia, esta abordagem reconhece a ligação entre a literacia e o âmbito social na qual esta se efetiva. Tendo em conta a interação dos indivíduos com diferentes esferas da vida no seu dia a dia, também a sua competência nestas, assim como os textos que nelas circulam, variam (Alves *et al.*, 2012). Desta forma, os textos não apresentam as mesmas características em todos os contextos, variando conforme o âmbito (Dionísio *et al.*, 2014). Por outras palavras, diferentes domínios convocam diferentes textos e interações com os mesmos.

A título de exemplo considera-se um indivíduo cuja atividade profissional é a construção civil. Este indivíduo é considerado como um excelente profissional no seu labor, mas, ao mesmo tempo é igualmente considerado como tendo níveis baixos de literacia. O contacto constante dentro da sua profissão com textos instrutivos, manuais e indicações de segurança, entre outros, certamente fá-lo-á proficiente na compreensão (e talvez até produção) daqueles textos, ou seja, no seu uso. Serão assim os textos dos outros âmbitos sociais (financeiro, mediático, científico), bem como a sua dificuldade com estes, os textos que estarão na base da sua baixa literacia. Por conseguinte, a falta de compreensão da dimensão social da literacia, assim como a uniformização de

programas de requalificação, levará a resultados que beneficiarão alguns, mas que prejudicarão muitos.

O sucesso destes programas revelará igualmente mudanças mais íntimas nos seus participantes. Este novo contacto com os processos de leitura e escrita tem o poder de dar acesso aos traços e valores das comunidades letradas – as quais os formandos viam como não fazendo parte delas –, assim como ao seu capital cultural (Cunha & Dionísio, 2019). Por outras palavras, no sucesso destas formações está embutida a mudança de mentalidade em relação ao valor dos indivíduos, reformulando a maneira como estes se percebem na sociedade.

Estas alterações de perceção do mundo estão interligadas com a noção do *kit de identidade* (Gee, 1996). Nesta ideia, cada indivíduo carrega uma espécie de “bilhete de identidade”, no qual são reconhecidas as suas valências e lacunas não só em relação à sua capacidade de compreensão de certos textos, mas também em relação às atitudes, crenças e valores que estes encapsulam. Desta forma, entende-se que a formação destes indivíduos os doa também de competências extratextuais que serão adicionadas ao seu *kit de identidade*. Uma destas competências extratextuais que é adicionada a cada indivíduo é o entendimento em relação a quem pode ostentar o título de “sujeito letrado” (Dionísio *et al.*, 2014). Esta é uma das grandes mudanças a acontecer nos formandos – aquela que os identifica e legitima como membros funcionais da sociedade e lhes dá um sentimento de pertença –, a qual se refletirá no sucesso das formações, uma vez que lhes asseguram o reconhecimento da sociedade, reconhecimento este que certamente desencadeará mais sucessos futuros.

2.2. Complexidade Textual

O conceito de complexidade numa entidade ou sistema é algo que não é exclusivo do campo da linguística. Também áreas de conhecimento como a biologia, química ou outras se debruçam sobre a questão: “o que torna algo complexo?” (Karlsson, 2014).

Cunningham & Mesmer (2014:256) definem complexidade textual como um conjunto de características inerentes e exclusivas ao texto, as quais podem ser analisadas e estudadas de um modo distinto das dificuldades do leitor. Este

entendimento da complexidade como um fenómeno exclusivo e independente enquadra-se na visão estruturalista do início do século XX, cuja distinção entre “simples vs complexo” é dada através de manifestações linguísticas marcadas e não marcadas (Karlsson, 2014). Nesta ideia, uma manifestação linguística existe num *continuum* e pode-se posicionar tanto no extremo de complexidade como no extremo de simplicidade, dependendo do seu grau de *marcação*. Ao nível da palavra, por exemplo, quanto menos morfemas esta tiver, mais simples será. Assim, categorias como singular, infinitivo, o caso nominativo, são vistos como formas “simples” ou “puras”, ou seja, sem marcação, ao contrário dos seus respetivos opositores: plural, outros tempos verbais e outros casos gramaticais. Esta ideia pode ainda estender-se à sintaxe, com o mesmo contraste a ser feito entre a ordem canónica SVO na língua portuguesa ou a voz ativa, com outras ordens de constituintes e a voz passiva.

Também Dahl (2007) interpreta a complexidade linguística desta forma, considerando a quantidade de informação embutida numa manifestação linguística como sendo a bitola para a medição da sua complexidade: quanto mais informação linguística existir (e.g., a marcação do género gramatical), maior será a complexidade do enunciado em questão.

Por outro lado, o *Common Core State Standards/ English Language Arts* (CCSS/ELA) (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010) define a noção de complexidade como algo subjetivo. Neste entendimento, a complexidade textual consiste na dificuldade inerente em ler e compreender um texto, combinado com as variáveis do leitor em questão. Desta forma, o entendimento do que torna um texto complexo é visto como um fenómeno relativo, o qual varia de leitor para leitor (Kusters, 2003). O CCSS/ELA propõe ainda um sistema tripartido para avaliar a complexidade de um texto: a sua dimensão quantitativa, qualitativa, e a tarefa do leitor para o compreender.

2.2.1. A tarefa do leitor

Os textos nunca são completamente explícitos (Perfetti *et al.*, 2001). À vista disto, a necessidade de inferir significado a partir de um texto é absolutamente necessária para

a sua compreensão. Considerando o fenômeno da complexidade textual como subjetivo e variável, o CCSS/ELA tem em consideração as idiossincrasias de cada leitor e as tarefas cognitivas a executar pelo próprio quando este lê um texto. Por outras palavras, diferentes leitores possuem diferentes proficiências em literacia, assim como diferentes textos requerem diferentes abordagens para serem compreendidos.

Em relação à primeira questão, parece óbvio que hábitos diferentes de leitura nos leitores geram diferentes níveis de proficiência nos mesmos. Por conseguinte, os padrões de leitura de cada um tornam-se mais ricos com a idade e, por isto mesmo, são mais avançados nos adultos do que nas crianças (Flor *et al.*, 2013). Isto sugere a possibilidade de um “amadurecimento” em certas estratégias de leitura, de forma implícita, mas, também, a possibilidade de aperfeiçoamento da leitura de cada indivíduo. Como qualquer aptidão, a sua prática gera o seu progresso e melhoria.

Por outro lado, também diferentes materiais de leitura convocam diferentes conceitos, tópicos, âmbitos e percursos cognitivos para a sua compreensão. O que o texto tenta comunicar e como o faz para concretizar esse objetivo vai também desempenhar um papel na sua compreensão (Goldman & Lee, 2014). Textos literários, os quais tendem a ter um tema narrativo implícito exigem, incontestavelmente, inferências e estratégias cognitivas da parte do leitor mais aguçadas do que a leitura de uma notícia sobre um tópico quotidiano, cujo objetivo é comunicar acontecimentos e os seus detalhes da maneira mais explícita possível. Ainda que o que os textos requerem para serem compreendidos não seja, na maioria dos casos, explícito, adiante notar-se-á como existe uma hierarquia de dificuldade neles.

2.2.2. Dimensão quantitativa

O interesse na maneira como os textos transmitem as suas ideias e como são compreendidos pelos leitores é algo quase tão antigo como a tecnologia da escrita em si. Contudo, foi apenas no século XIX que um estudo formal e constante começou a ser posto em prática (Hiebert & Pearson, 2014). Nesta altura, a ênfase desta primitiva análise de complexidade textual era maioritariamente qualitativa, focando-se em

características textuais específicas e estudando como estas condicionavam a compreensão dos leitores.

À medida que a tecnologia trouxe novos avanços, novas abordagens científicas também surgiram neste campo, com o conceito de *readability*⁹ a ser analisado através de um prisma quantitativo. Estas abordagens consistiam em fórmulas que contemplavam, principalmente, a frequência de palavras e a extensão das frases (Gunning, 1952; Flesch, 1948; Kincaid *et al.*, 1975; McLaughlin, 1969; Dale and Chall, 1948; Spache, 1953; Lorge, 1948). Sustentados em modelos computacionais incipientes, estas fórmulas contavam as ocorrências geradoras de complexidade, as quais tinham sido previamente estabelecidas (e.g., palavras compostas, número de sílabas numa palavra, número de palavras numa frase). Esta contagem resultava num valor que servia de resultado numérico e ditava, assim, a complexidade do texto (Goldman & Lee, 2014).

Estas *fórmulas de leituraabilidade*¹⁰ são ainda os índices mais comumente utilizados para avaliar a complexidade de um texto, operando como a pedra basilar de programas escolares como o CCSS/ELA, assim como tendo sido também adaptados para outras línguas (Hiebert & Pearson, 2014; Rabin, 1988). Mais de 200 fórmulas deste género já foram propostas por investigadores (Hiebert, 2011). No entanto, devido à sua natureza quantitativa, estas fórmulas tendem a determinar a complexidade sintática num texto através da métrica de palavras por frase; e a complexidade semântica através da frequência de ocorrência de uma palavra em comparação com uma lista de palavras preestabelecida, resultando num cálculo da sua familiaridade/raridade (Flor *et al.*, 2013). Esta simplificação de como a complexidade se manifesta e materializa nos textos, juntamente com o facto de, em geral, novos índices de *leiturabilidade* se basearem e se desenrolarem sobre outros índices – em vez de propor novas métricas – levou à pergunta frequente: “Quão precisas são estas fórmulas na atribuição de materiais de leitura a leitores?” (Hiebert & Pearson, 2014).

⁹ *Readability* é a facilidade na leitura criada pela escolha de conteúdo, estilo, design e organização que se adequa ao conhecimento prévio, às habilidades de leitura, ao interesse e à motivação de um público-alvo (DuBay, 2008).

¹⁰ Tradução do autor de *readability*.

Durante os anos 80, as fórmulas de leituraabilidade, ainda que não sendo totalmente inúteis, uma vez que ofereciam resultados que podiam ser interpretados mais tarde, eram vistas como apresentadoras de resultados deficientes (Hiebert & Pearson, 2014). A ideia é a de que certas componentes destas fórmulas são válidas como indicadores iniciais de complexidade textual. Contudo, medidas como a extensão da palavra e da frase podem não servir de medidas para revelarem toda a complexidade embutida num texto (McNamara *et al.*, 2010; Hiebert, 2011). Para além disto, estas fórmulas têm sido também criticadas pelo facto de não detetarem (ou inflacionarem) certas características que induzem complexidade relacionadas com diferentes géneros textuais (Hiebert, 2011). Um exemplo disto é o facto de textos narrativos possuírem grandes porções de texto de diálogos, nos quais figuram, muitas vezes, frases curtas, uma vez que representam a forma como as pessoas falam no quotidiano. Ora, considerando que estes índices de leituraabilidade aferem frases curtas como não sendo complexas, a complexidade presente nestes blocos de diálogo tende a ser subvalorizada. Assim sendo, não só as particularidades de diferentes géneros textuais são negligenciadas por estas fórmulas, mas também certos textos parecem ser condensados a um resultado genérico, quando o texto (e.g., um romance) possui partes mais complexas e mais simples.

Em suma, uma abordagem quantitativa à complexidade textual está sempre dependente de uma interpretação subsequente. A própria Chall, uma das criadoras de uma das maiores e mais usadas fórmulas de leituraabilidade (Chall & Dale, 1948), afirmou que nada existia na sua fórmula que conduzisse à sua má utilização, insinuando, assim, que o problema residia numa má interpretação dos seus resultados (Hiebert, 2011).

2.2.3. Dimensão qualitativa

Nos anos 80 do século XX, com o afastamento da tendência científica dominante do *Behaviorismo* para abordagens mais baseadas na cognição, o debate sobre a complexidade textual mudou igualmente (Hiebert & Pearson, 2014). O enfoque deixou de ser colocado nas características presentes num texto que provocavam complexidade e passou a ser posto nas consequências destas características no leitor, nomeadamente em como estas afetavam a compreensão do texto. Esta mudança na perceção da

complexidade textual levou a uma consequente mudança que colocaria o leitor no centro da questão, ao invés do texto.

As abordagens qualitativas foram assim ressuscitadas. Teorias de esquemas textuais¹¹, de modelos mentais representativos do processo de leitura¹², assim como gramáticas de histórias¹³, foram propostas com o objetivo de melhor compreender como opera a compreensão de um texto (Harley, 2013). Igualmente importante, estas teorias tentavam ir para além das características de complexidade textual propostas por índices de leiturabilidade, as quais se limitavam a dar conta de manifestações de complexidade superficiais. Métodos qualitativos de análise de complexidade textual como *text leveling*, *rubrics* e mapas de texto baseiam-se nestas teorias de compreensão textual e foram postas em prática (Pearson & Hiebert, 2014).

O primeiro método – *text leveling* – consiste na elaboração de níveis de complexidade textual baseados no conhecimento e juízo de peritos nesta questão, para depois relacionar estes níveis com diferentes níveis de proficiência em leitura. A postulação destes níveis de complexidade está dependente do saber empírico de professores de línguas experientes (assim como outros peritos da área), os quais possuem um discernimento proveniente de vários anos de ensino de língua.

Uma versão mais desenvolvida de *text leveling* é um sistema de *rubrics*. Neste método qualitativo de análise de complexidade, um grupo de peritos identificam características nos textos que consideram como indutoras de complexidade textual. Estas características são, depois, colocadas num *continuum*, no qual características responsáveis por uma menor complexidade são colocadas num extremo e as responsáveis por uma maior complexidade são colocadas noutro, acompanhadas de outras características classificadas noutros pontos neste *continuum* (Pearson & Hiebert, 2014). Um exemplo deste tipo de classificação será uma palavra e os seus afixos

¹¹ O conceito de teorias de esquemas textuais baseia-se na organização (esquematisação) dos elementos do texto com base no conhecimento prévio do leitor destes mesmos elementos e a forma como se podem relacionar com outros elementos. Cf. Anderson (2000).

¹² Os modelos mentais de leitura pretendem representar a informação transmitida por um texto à medida que esta informação é dada. O objetivo é reproduzir as imagens mentais criadas pelos leitores na sua memória de curto prazo, e como estas percecionam elementos como o tempo, espaço, causalidade, etc. (Harley, 2013).

¹³ Gramáticas de histórias identificam e analisam a estrutura de uma história, assim como os seus elementos.

derivacionais: uma palavra simples sem afixos seria classificada no extremo de complexidade baixa do *continuum*, enquanto uma com afixos seria marcada no *continuum* como mais complexa, e uma palavra composta como ainda mais complexa. Exemplos destes índices qualitativos são o *Qualitative Assessment of Text Difficulty* (QATD) (Chall *et al.*, 1999) ou o *American College Test* (ACT) (ACT, 2006).

Por último, o método de mapas de texto visa esclarecer uma complexidade mais “invisível” nos textos (Pearson & Hiebert, 2014). Fundamentando-se na noção de gramáticas de história, o propósito de mapas de texto é revelar a estrutura conceptual de um texto e destacar a sua complexidade (ou a falta desta). Como referem Pearson & Hiebert (2014), os mapas textuais permitem identificar o tema, o enredo, o contexto, a estrutura, as personagens e ainda conceitos abstratos aos quais o texto faz alusão, mas que nunca refere explicitamente. Além disso, estas medidas qualitativas do texto potenciam também a associação de certos mapas textuais a certos géneros, evidenciando padrões textuais.

Em suma, uma fórmula qualitativa de análise de complexidade procura simular a inteligência humana e a forma como esta perceciona as manifestações de complexidade. No início dos anos 2000, o desenvolvimento tecnológico abriu o caminho a análises qualitativas de complexidade mais sofisticadas, que se têm vindo a aprimorar a um passo rápido. O índice Coh-metrix é o mais recente (McNamara *et al.*, 2010).

No entanto, todos estes índices qualitativos não estão isentos de críticas. A abordagem holística destes métodos tem sido várias vezes criticada por ter o “fim cego” de classificação de textos e atribuição de níveis a estes, sem a explicação e especificação de como se chegou a uma certa conclusão. Igualmente, o peso dado às diferentes características de complexidade aferidas por estes índices nunca é realmente apresentado (Pearson & Hiebert, 2014). Desta forma, a aplicação em usos práticos destes índices pode falhar, uma vez que não é capaz de informar adequadamente os professores em relação à maneira como os índices podem ajudar os alunos que experimentam dificuldades ou facilidades com os textos propostos (Hiebert & Pearson, 2014). É como se, na tentativa de romper com os resultados numéricos e matemáticos

dos índices quantitativos, os próprios índices qualitativos tivessem subvalorizado a importância da sua própria metodologia.

2.2.4. O uso *in tandem*

Apesar do ressurgimento e aplicação em curso de índices qualitativos de complexidade textual, a verdade é que as fórmulas de leiturabilidade não devem ser desconsideradas. Durante o início do milénio – o mesmo período no qual índices qualitativos se achavam em considerável crescimento –, as fórmulas de leiturabilidade estavam a ser digitalizadas, tornando-se assim disponíveis e convenientes para professores e editoras de livros escolares, influenciando as escolhas dos seus programas escolares (Hiebert & Pearson, 2014). Ao contrário do expectável, os índices qualitativos não estavam a substituir os quantitativos. Na verdade, os índices qualitativos ainda eram vistos como rudimentares nos seus processos de aferição de complexidade e, por isso, não inspiravam a mesma confiança nos seus utilizadores que os índices quantitativos, cujos resultados eram numéricos e “brutos”, mas, de todas as formas, factuais (Pearson & Hiebert, 2014).

Como consequência das deficiências de ambos os tipos de índices, vários são os autores que advogam pelo seu uso *in tandem* (Hiebert, 2011; Hiebert & Pearson, 2014; Cunningham & Mesmer, 2014; Douglas & Miller, 2016; Goldman & Lee, 2014). Considerando que ambos são vistos como tendo as suas limitações, a ideia é de que o uso em conjunto levaria a que os dois se complementassem e minimizassem as suas respetivas falhas (Cunningham & Mesmer, 2014). Na verdade, resultados provenientes de uma abordagem quantitativa não têm significado até serem realizados numa forma qualitativa (Hiebert & Pearson, 2014). Por outras palavras, os resultados de uma abordagem quantitativa são meros números que necessitam de ser interpretados dentro de um contexto específico. Num estudo com o objetivo de aferir a complexidade textual ao nível da frase (Douglas & Miller, 2016), duas medidas de análise de complexidade foram empregues – uma qualitativa (Ai’s Web-based L2 Syntactic Complexity Analyzer) e uma quantitativa (Lexile¹⁴). Esta última medida quantitativa foi considerada como insuficiente, uma vez que considerava unicamente frases mais longas

¹⁴ Apesar do índice Lexile também utilizar medidas qualitativas, o uso dele neste contexto foi exclusivamente como um índice quantitativo.

como sendo mais complexas. No entanto, a medida providenciava informação importante, a qual seria suplementada com uma medida qualitativa, rendendo assim resultados mais apurados.

A combinação de medidas de dimensões diversas não só é benéfica para a geração de resultados mais fidedignos, mas também pode ser empregue de formas inovadoras com o objetivo de identificar e processar certas manifestações de complexidade que ainda não foram analisadas (Hiebert & Pearson, 2014). Efetivamente, em alguns textos (nomeadamente literários e narrativos), tanto índices qualitativos como quantitativos tendem a falhar na deteção de complexidade, classificando-os como tendo uma complexidade baixa, quando, na prática, estes permanecem como de alta complexidade para a maioria dos leitores (Goldman & Lee, 2014).

Em última instância, o fenómeno da complexidade textual pode manifestar-se de diferentes maneiras e dimensões e, portanto, justifica o emprego de diferentes índices de aferição de complexidade para identificar e processar estas diferentes dimensões. Fórmulas de leiturabilidade podem ser utilizadas com o intuito de captar uma ideia geral da complexidade do texto (Hiebert, 2011). Contudo, como o próprio Hiebert (2011:8) afirma, aquando da aferição da complexidade de um texto, o uso de uma única medida – independentemente da sua natureza –, só pode levar ao entendimento do fenómeno da complexidade textual como algo que se manifesta de uma maneira unidimensional.

Este trabalho aborda, assim, a questão da complexidade textual, com os conteúdos referidos no enquadramento teórico em mente. No âmbito do objetivo deste trabalho, três questões se colocam:

- Que características textuais dificultam a compreensão de um texto a adultos de baixa literacia?
- Como se evidenciam estas características nos textos?
- Poderão os textos condicionar a complexidade manifestadas neles?

De modo a responder a estas questões, algumas perspetivas específicas serão adotadas. Primeiramente, a noção de *texto* será cingida à sua modalidade escrita. Depois, a literacia não é entendida como uma noção estanque, apesar de existir um

pendor para a escola de pensamento que a considera um fenómeno social, em constante transformação no espaço e tempo. Com efeito, é essa a ideia que sustenta programas EFA.

No entanto, entende-se que o objetivo não é discorrer sobre a dimensão sociológica da mesma, mas antes explorar como esta se materializa linguisticamente. Assim, a literacia (e os seus níveis de proficiência) estará sempre estreitamente ligada aos textos em si e como estes manifestam complexidade. Esta abordagem será também fundamental para a terceira questão, onde será investigada a relação dos textos com diferentes tipos de complexidade.

No que diz respeito ao conceito de complexidade textual, uma especial atenção é dada aqui à abordagem qualitativa, relegando (mas não refutando) abordagens quantitativas, como índices de leiturabilidade. Desta forma, aborda-se o tema da complexidade com uma abordagem *in tandem*, com uma forte tendência para a sua consideração qualitativa, de modo objetivo e descritível.

3. METODOLOGIA

Para alcançar os objetivos deste trabalho, nomeadamente a criação de uma ferramenta de análise de complexidade textual para adultos de baixa literacia, foram contempladas duas tarefas essenciais, que incluíram uma série de ações:

1. Análise de complexidade textual: definição e enquadramento da complexidade textual, com estabelecimento de níveis de complexidade e descritores relevantes, juntamente com a discussão e validação por professores de centros de aprendizagem de adultos.
2. Compilação do *corpus*: recolha e validação de textos dos níveis de complexidade propostos.

As tarefas que aqui são referidas estão intimamente relacionadas com a infraestrutura de complexidade, basilar para este trabalho. A próxima secção aborda cada uma destas tarefas, as suas respetivas motivações, métodos e procedimentos. O esquema abaixo ilustra os passos discutidos nesta secção.

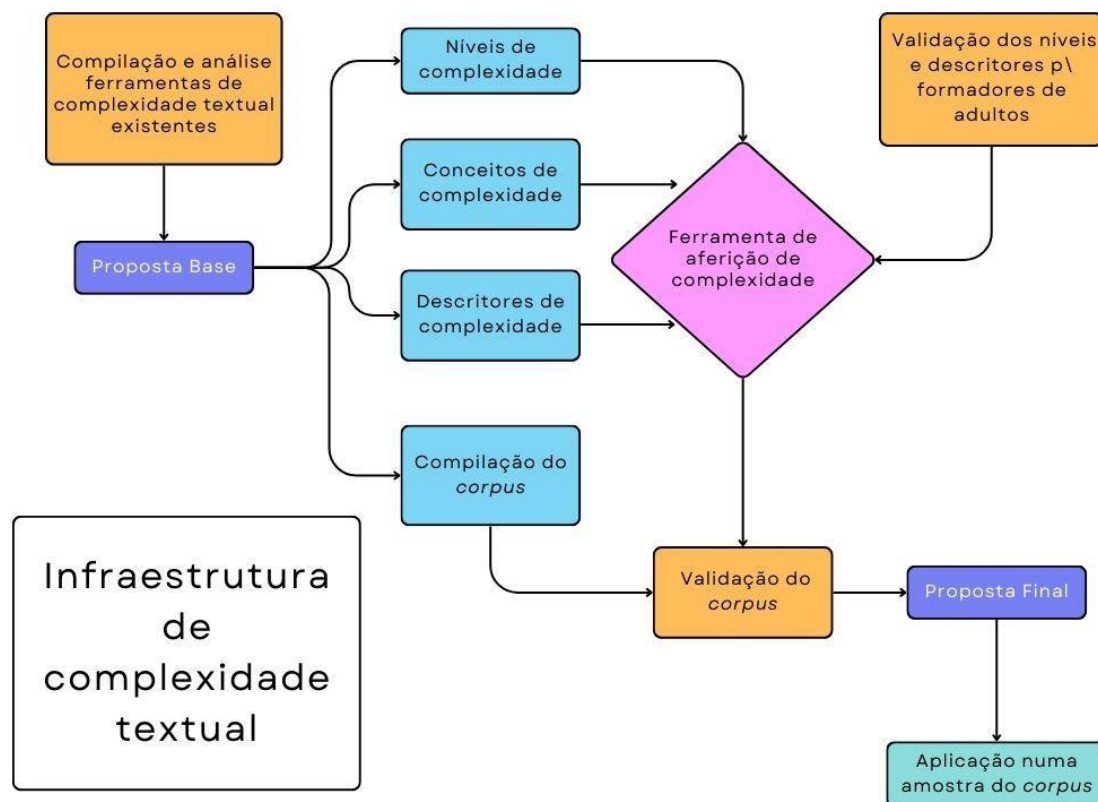


Figura 1. Esquema das componentes da infraestrutura de complexidade textual

3.1. Níveis de complexidade e proficiência de literacia e as suas características

Com o objetivo de estabelecer diferentes níveis de complexidade textual e respetivos descritores, foram dados dois passos:

1. Levantamento e análise dos níveis e descritores presentes em referenciais de proficiência em leitura existentes.
2. Definição de uma proposta de base para os níveis de complexidade adequados a adultos nativos de baixa literacia.

3.1.1. Levantamento e análise de níveis e descritores de proficiência em leitura

O primeiro passo para o estabelecimento de uma infraestrutura de complexidade textual foi dado através da análise de referenciais oficiais. Os referenciais em questão

são: o Quadro Europeu Comum de Referência para as Línguas (QECR), o Inquérito às Competências dos Adultos – PIAAC, o *Can Do Project*, da Associação de Examinadores de Línguas na Europa (ALTE) e o Referencial de Competências-Chave de Educação e Formação de Adultos – Nível Básico da Agência Nacional para a Qualificação e o Ensino Profissional (ANQEP). Estes documentos não têm em consideração o mesmo público-alvo deste trabalho (à exceção do último), abordando a proficiência da linguagem numa ótica mais geral, contudo, as noções e conhecimentos contidos nestes referenciais – nomeadamente o que demarca um nível de complexidade do seguinte e anterior – foram entendidos como um ponto de partida adequado para o desenvolvimento do trabalho.

3.1.2. Proposta de base adequada a adultos de baixa literacia

Depois de analisados os referenciais, foi dada especial atenção aos três níveis iniciais de cada um. No caso dos referenciais propostos pela ALTE, QECR, PIAAC, os três níveis inferiores foram tidos em conta. No caso do referencial da ANQEP, os três níveis propostos já correspondiam aos níveis iniciais.

A razão pela qual os três níveis inferiores de cada infraestrutura de complexidade foram considerados é dupla:

i) Tal como discutido no capítulo 2, o conceito lato de literacia e os diferentes níveis de proficiência tornam difícil a distinção do que é um adulto de baixa literacia. Assim, decidiu-se analisar os níveis de cada referencial e os seus descritores, de modo a perceber o que significava ser um adulto de baixa literacia no dia a dia. Os três níveis inferiores de cada referencial encerram a noção de um adulto de baixa literacia e os grupos de textos em que este é proficiente. Por exemplo, no QECR, o nível imediatamente acima dos que foram considerados pelo projeto – ou seja, o nível B2 – descreve um indivíduo capaz de “ler artigos e reportagens sobre assuntos contemporâneos em relação aos quais os autores adotam determinadas atitudes ou pontos de vista particulares” (CEFR, 2020:177). Uma proficiência de leitura deste género é aqui vista como não sendo associável a um adulto de baixa literacia.

ii) O entendimento de que a literacia não é uma noção estanque conduziu à conclusão de que um adulto de baixa literacia não pode ser associado a apenas um nível de proficiência de leitura, havendo alguma variabilidade e heterogeneidade de competências nesta população. De modo a considerar esta heterogeneidade, foram considerados três níveis para caracterizar esta população.

3.1.3. Os níveis e os descritores de complexidade textual

Depois da tarefa de nivelamento dos referenciais, seguiu-se uma proposta de níveis de complexidade textual, cada um com os seus próprios descritores.

Estes níveis de complexidade correspondem a níveis de proficiência em literacia de adultos. Cada um destes níveis é representado por descritores de complexidade textual (abordados mais à frente), de modo a colocar o foco da complexidade no texto e não no leitor.

É importante notar que, ainda que os níveis de complexidade sejam descritos pela complexidade dos textos, a descrição de cada nível faz alusão a um nível de escolaridade específico, de modo a dar alguma referência objetiva a estas mesmas descrições. Com isto, o objetivo não é caracterizar os leitores no que diz respeito à complexidade textual, mas sim os textos, descrevendo depois o nível de escolaridade que melhor se adapta a textos de dado nível.

Ademais, a proposta de níveis de complexidade textual, com base nos níveis básicos dos referenciais foi feita por se entender que é nos níveis iniciais de proficiência de leitura que se torna relevante a promoção e desenvolvimento de práticas de leitura – um dos objetivos do projeto. Efetivamente, em níveis superiores de proficiência em leitura, estes comportamentos já se encontram consolidados (pelo menos, parcialmente), em virtude dos textos acessíveis e compreensíveis a estes leitores, e em estágio de alfabetização, estes comportamentos ainda se limitam à competência de decodificação da escrita.

É, ainda, nestes níveis iniciais que há mais dificuldade em conseguir materiais de ensino/treino de leitura adequados a adultos com baixa literacia. Considerando esta dificuldade manifestada pelos formadores de programas EFA em recolher e usar textos

adequados para os seus alunos, o estabelecimento destes três níveis de complexidade visa, também, ajudar os formadores a escolher textos adequados ao seu contexto de ensino.

Uma vez estipulada a definição dos níveis de complexidade, a atenção recaiu sobre o que caracterizava cada um deles. Desta forma, descritores de complexidade textual, pertinentes e referentes a cada nível, foram propostos.

3.2. Constituição e validação do *corpus*

Para criar e publicar uma ferramenta de análise de complexidade textual eficaz e suficientemente versátil, um *corpus* foi compilado com a meta de exemplificar e validar os níveis e os descritores de complexidade propostos. O objetivo final deste *corpus* é servir de base para o treino e teste na fase de Processamento de Língua Natural (PLN).

O *corpus* do projeto iRead4Skills teve por base três momentos: a definição de critérios de seleção e amostragem de textos, a sua recolha, e a sua validação.

3.2.1. Compilação do *corpus* – critérios de seleção e amostragem

A primeira etapa da constituição do *corpus* consistiu na definição de critérios para seleção de textos. Foi decidido que os textos selecionados representariam, dentro do possível, o “mundo real” de forma abrangente e levariam em conta os hábitos/necessidades de leitura de uma população de baixa proficiência (patentes nos resultados dos dois questionários iniciais do projeto – *Reading Skills Survey* (Correia *et al.*, 2023) e *Overall skills and gaps survey* (Clément *et al.*, 2023)). Assim, foi criada uma *tipologia textual* para dar conta da execução desta tarefa. O uso desta tipologia tem como principal base a relação entre a dimensão social e a dimensão linguística dos textos (Lee, 2001; Marcuschi, 2002). Tendo em conta a infraestrutura de aferição de complexidade proposta, a organização dos textos segundo esta tipologia é mais adequada.

A criação desta tipologia textual baseou-se nos seguintes critérios:

- **Propósito/Objetivo do texto** – Esta noção relaciona-se com o objetivo comunicativo e conceptual do texto. Assim, entendeu-se que o objetivo do texto estava estritamente associado ao conceito (ou conceitos) explicitado(s) e comunicado(s) no texto. Estes conceitos estão intimamente ligados com esferas ou instâncias de produções discursivas específicas. Por sua vez, esta ideia está intrinsecamente ligada à noção de domínio discursivo, que não é um texto, mas proporciona o surgimento de textos particulares (Marcuschi, 2002).
- **Tipo/Forma de documento** – É uma caracterização das realizações linguísticas de objetivos específicos em situações sociais (Marcuschi, 2002). Neste critério, os textos são entendidos na sua forma social, legitimados sob o prisma de práticas de comunicação humanas. Esta definição encaixa na noção de géneros textuais, isto é, a categorização dos textos aquando da sua manifestação na sociedade (Lee, 2001; Marcuschi, 2002; Miranda, 2012).
- **Área/Conteúdo** – Por último, a Área/Conteúdo refere-se às noções extralinguísticas associadas ao texto, podendo estar ligadas aos contextos nos quais os textos são passíveis de circular, bem como ao teor dos mesmos.

Como exemplo, um panfleto publicitário de um banco poderia ter o Propósito de *Comunicação Comercial*, com o Tipo de documento *Flyer*, pertencendo à área *Área/Conteúdo Financeira/Bancário*.

Desta forma, foram propostas 9 subcategorias de Propósitos e 68 subcategorias de Tipos de documentos¹⁵ (ver anexo III). Alguns aspetos destas propostas devem ser, no entanto, considerados.

Primeiramente, considerou-se que algumas subcategorias incluídas em Propósitos se referiam a textos que, pela sua potencial complexidade estrutural e conceptual, são

¹⁵ Devido às inúmeras especificações de categoria Área/Conteúdo ao qual o texto faz menção, podendo este critério ser entendido de uma forma mais específico ou mais generalista, a contagem de manifestações que constam nesta tipologia não foi considerada relevante.

pouco apropriados ou de pouco interesse a particularmente adultos de baixa literacia – dois em específico: *Comunicação Pessoal* e *Académico*.

A razão da escolha do primeiro (*Comunicação Pessoal*) está relacionada com o facto de uma grande maioria desta população ter afirmado, nos questionários, já ter um extenso contacto com textos do quotidiano, no qual a *Comunicação Pessoal* se insere. Para além disto, no questionário *Reading Skills survey* (Correia *et al.*, 2023), a principal motivação desta população para ler é a de **aprender conteúdos novos**. Torna-se, assim, evidente que o estímulo de leitura para esta população tem de passar por um contacto com textos que a desafia a melhorar não só a sua proficiência em leitura, mas também os seus conhecimentos sobre o mundo. Esta noção, designada de *zone of proximal text difficulty* (Goldman & Lee, 2014), a qual considera benéfico o contacto dos leitores com textos para lá da sua “zona de conforto”, é também aplicável à segunda subcategoria de Propósitos aqui referido – o *Académico*. Considerou-se que devido à complexidade dos assuntos abordados neste tipo de texto, a qual exige um conhecimento prévio aprofundado sobre o conteúdo do texto, os textos com o Propósito Académico se afastavam demasiado de uma zona de possível desafio para um leitor de baixa literacia. Para além disto, textos académicos tendem a refletir práticas de escrita que dificultam a compreensão do texto por serem, muitas vezes, desconhecidas de uma população fora desse mundo (e.g., início do texto com um Resumo, citações no corpo de texto, etc.).

Assim, ainda que as duas categorias acima mencionadas tenham sido entendidas como menos apropriadas para a promoção da leitura, a compilação de textos nelas foi executada, com vista à integridade, diversidade e representatividade do *corpus*.

3.2.2. Recolha dos textos por nível

Com esta tipologia servindo de modelo para a recolha e organização dos textos do *corpus*, este foi compilado através da recolha de vários textos disponíveis *online*. É importante mencionar que, na altura da recolha dos textos, todos os metadados (i.e. autor, editor, data de acesso, etc.) foram recolhidos e armazenados. Uma diversidade de fontes e autores foi tida em conta e foram considerados textos criados por adultos para adultos, escritos na variedade europeia do português.

Esta recolha de textos contou também, necessariamente, com uma primeira inspeção ao seu nível de complexidade. A principal razão para isto prende-se com o facto de a compilação do *corpus* ser simultânea à criação da proposta de descritores de complexidade e, por isso, a sua pré-avaliação teve de ser feita com base na descrição preliminar dos níveis de complexidade.

Cada texto recolhido foi associado a um dos três níveis de complexidade textual (Muito fácil, Fácil, Claro)¹⁶. Foi postulado também um quarto nível – o *Mais Complexo*. A consideração de mais um nível de complexidade acima do nível Claro está relacionada com a necessidade de estabelecimento de um limiar de complexidade superior em relação ao último nível de complexidade (Claro). Como foi referido anteriormente, o mesmo já tinha sucedido em relação ao primeiro nível, com o estabelecimento de uma fronteira inferior, a qual distinguia adultos de baixa literacia de adultos em processo de alfabetização. Estes limiares inferior e superior de complexidade servem para abalizar a aferição de complexidade, a qual tem o público-alvo em consideração. Porém, estes limiares não se enquadram na aferição de complexidade *per se*, uma vez que não representam os textos, mas sim os limites dos níveis de complexidade existentes dentro da infraestrutura de complexidade. Desta forma, o seu objetivo é muito mais funcional do que apurativo, enquadrando a população de baixa literacia. Apesar de estar fora do escopo do projeto, os textos que se achavam neste quarto nível, Mais Complexo, foram compilados de modo a servirem de referência para os outros níveis de complexidade abaixo e, como tal, necessários à análise automática de complexidade a realizar.

Cada texto foi codificado com as seguintes características – *Nível_Propósito do texto_Tipo de documento_número do texto* – e armazenado numa base de dados.

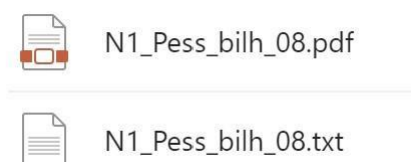


Figura 2. Exemplo de codificação do documento

¹⁶ Cada um deles será caracterizado na secção 4.1.1.

Foi estabelecida a meta de recolha de 2000 textos no total. Para cada Tipo de documento, foram compilados 30 textos, divididos pelos três níveis de complexidade (10 para cada nível). Cada texto era composto por 250 a 500 *tokens*¹⁷. Os textos do nível Mais Complexo, embora compilados e considerados para o efeito acima descrito, não entram nos números totais da recolha.

3.2.3. Validação do *corpus*

O processo de recolha do *corpus* interligou-se necessariamente com uma primeira validação da classificação dos textos. De modo a otimizar este trabalho, foi adotada uma abordagem *in tandem*: a recolha do *corpus* foi realizada por uma equipa, ao mesmo tempo que o processo de validação dos textos era realizado por outra equipa, o que permitia a constante atualização de ambas as tarefas. Desta forma, a recolha e validação dos textos por duas equipas diferentes, de acordo com os descritores de complexidade textual estipulados, permitiu afinar o entendimento do que era um texto do nível Muito Fácil, Fácil e Claro, o que, por sua vez, permitiu uma recolha de textos mais eficiente e precisa. Esta recolha e validação, efetuadas por mais do que uma pessoa, fomentava a constante comunicação e atualização de ambas as tarefas, diminuindo o efeito de enviesamento pessoal.

Para além disto, à medida que a tarefa de validação do *corpus* se desenrolou, verificou-se que certos descritores eram mais adequados a certos tipos de documentos do que noutros. A título de exemplo, a dimensão de coesão textual pareceu ser bastante mais revelante em notícias jornalísticas, que reportavam eventos e que frequentemente os ligavam com outros eventos e outros períodos espaciotemporais, do que, por exemplo, em horóscopos. Embora nenhum dos descritores tenha sido descartado, alguns foram alterados e outros foram considerados como especialmente críticos em certos tipos em certos tipos de documentos¹⁸.

¹⁷ A noção de *tokens* é aqui importada da Linguística Computacional e refere-se às palavras corridas de um texto. É usada como unidade de medida da dimensão dos *corpora*.

¹⁸ Esta ideia é explorada mais a fundo no capítulo que aborda uma análise de uma amostra do *corpus*, adiante.

O processo de validação dos textos foi manual e requereu o julgamento humano, proveniente de mais do que uma pessoa. Este processo de validação consistiu em detetar no texto a presença de descritores do mesmo nível que lhe tinha sido atribuído. Para este fim, os textos recolhidos e armazenados na sua respetiva base de dados eram acompanhados de um documento de validação (ficheiro *Excel*).

FILE NAME	A1 (light/transparent (easy))/ more dependent									
	lexical-conceptual				Verbs				Syntactic structures	
	con (w/ foreign is)	concrete nouns (concrete ideas)	daily concepts (personal life)	rare affixation (-idão, -mente, ...)	simple tenses in Infinitive, Indicative, Gerund	mainly main verbs (no auxiliary)	copulative verbs (ser, estar, ficar, continuar, andar, parecer)	frequent MW verbs (estar/andar a; ir; poder/querer) in Indicative mode + Inf.	simple coordination (e, ou, mas, nem, pois)	simple subordination (subject relative temporal (ante de, depois de, quando, enquanto), causative (pois porque), conditional (se) Indicative)
N1_Did_gloss_01										
N1_Did_gloss_02										
N1_Did_gloss_03										
N1_Did_gloss_04										
N1_Did_gloss_05										
N1_Did_gloss_06										
N1_Did_gloss_07										
N1_Did_gloss_08										

Figura 3. Documento de validação

Neste documento de validação, encontravam-se os descritores de cada nível. Os textos foram analisados e, caso um descritor fosse observado num texto (independentemente do nível), era assinalado no documento, através de um sistema de marcação binária (0 e 1). Os valores provenientes desta marcação eram automaticamente somados de modo contabilizar as ocorrências dos descritores, por nível. No final, o validador¹⁹ marcava o texto como tendo sido mantido no nível proposto ou mudado para outro nível. Desta forma, cada texto foi validado manualmente por mais do que um validador.

Para melhor compreender este processo, considere-se o seguinte exemplo: se um texto tivesse sido pré-avaliado como Muito fácil, na altura da sua recolha, então descritores deste nível como *conceitos concretos*, *tópicos do dia a dia*, *inexistência da voz passiva*, etc., deveriam estar presentes no texto. Se isto acontecesse, a presença de cada um dos descritores seria marcada como positiva (1). Se estes não estivessem presentes, seria marcado (0). De seguida, bastava identificar a que nível (ou níveis) de complexidade pertenciam os descritores manifestados no texto, marcá-los e contar as suas ocorrências. Se existisse uma maioria de ocorrências dos descritores, por exemplo,

¹⁹ Os validadores, integrados no projeto, eram especialistas, com formação em Ciências da Linguagem, capazes de observar a presença dos descritores de complexidade nos textos.

do nível seguinte – Fácil – este texto seria classificado como um texto deste nível e não daquele que tinha sido proposto. Desta forma, os textos seriam reaproveitados e não dispensados.

Ainda que muitos textos tivessem marcas de complexidade exclusivas de um nível, outros textos evidenciaram diversidade, tendo características do primeiro, segundo e terceiro níveis de complexidade, podendo, por exemplo, ter sobretudo estruturas sintáticas previstas como do primeiro nível, muitas formas verbais do segundo e léxico marcadamente do terceiro nível. Este processo de marcação binária permitiu, ao marcar as ocorrências dos descritores, somar e contar as mesmas por nível. Desta forma, foi contabilizado o “resultado” de cada nível em relação ao texto, bem como foi possível perceber, em alguma medida, o peso relativo de determinadas características.

3.3. Validação dos níveis e descritores de complexidade por formadores de adultos

Com o intuito de validar a infraestrutura de complexidade proposta, o julgamento e *input* dos formadores de centros EFA (nomeadamente, Centros Qualifica) foi requisitado, em dois momentos: para a composição e reunião de discussão num *focus group* e para a tarefa de classificação e anotação (Amaro & François, 2023). Devido à ausência de resultados da tarefa de anotação e classificação à altura da escrita deste trabalho, esta não é considerada neste trabalho.

Antes da realização do *focus group*, foi planeada uma tarefa preparatória que tinha como objetivo avaliar os níveis de complexidade do projeto, assim como os descritores, propostos. Esta atividade teve a duração de aproximadamente uma hora e consistiu num questionário online, enviado para vários professores de EFA. Os professores foram convidados a participar no projeto com o seu *know how*.

A tarefa continha uma primeira apresentação dos níveis de complexidade, sem referência às características denotadas pelos descritores para cada nível. De seguida, os descritores de complexidade foram apresentados de forma aleatória e foi solicitado aos professores que os associassem a um determinado nível de complexidade. Dado que

alguns professores não eram professores da língua portuguesa, os descritores de complexidade foram simplificados, encurtados e acompanhados de um exemplo escrito.

A tarefa decorreu durante o período de 10 dias com um total de 12 professores a responderem ao inquérito (ver anexo IV). Os resultados foram compilados e analisados com vista a entender quais os descritores consensuais e quais os problemáticos.

O próximo passo foi o estabelecimento de um *focus group*, constituído pelos mesmos respondentes do questionário, com o objetivo de discutir as suas respostas. Tal como aconteceu no questionário *online*, uma breve apresentação dos níveis de complexidade foi efetuada e, de seguida, todos os descritores foram revistos em conjunto. No total, 11 professores deram o seu contributo. Um dos professores que fez a tarefa preparatória de classificação dos descritores não compareceu na reunião de *focus group*.

3.4. Proposta final de complexidade textual para adultos de baixa literacia

A infraestrutura de complexidade foi revista com base nesta validação por parte dos formadores dos Centros Qualifica e foi feita uma proposta definitiva. Este ajuste final teve o objetivo de aprimorar a infraestrutura de complexidade em relação à proposta inicial, nomeadamente no que concerne o público-alvo.

3.5. Aplicação da infraestrutura de complexidade numa amostra do *corpus*

A infraestrutura de aferição de complexidade desenvolvida no âmbito do projeto iRead4Skills, para além de balizar os níveis contemplados no projeto e validar objetivamente os textos do *corpus* compilado, tem também a capacidade de revelar informação potencialmente relevante para a temática da complexidade textual.

De forma a evidenciar a aplicação da infraestrutura de complexidade textual, bem como os possíveis resultados que se podem obter através dela, foi feito um exercício de estabelecer uma ligação entre manifestações de complexidade e certos *Tipos de documento*. A principal intenção foi a de averiguar até que ponto a complexidade textual varia em função de certas características dos próprios textos. Para este propósito, uma

pequena amostra do *corpus* foi isolada e os resultados referentes às manifestações de complexidade em cada texto foram avaliados.

Especificamente, escolheram-se textos dentro da subcategoria *Comunicação Pessoal* no *Propósito/objetivo do texto*. A motivação para a escolha deste grupo de textos assentou no facto de estar associado a um âmbito da sociedade muitas vezes referido neste trabalho – o quotidiano –, assim como por ser um grupo que integra textos com os quais a população-alvo do projeto diz ter bastante contacto (cf. *Reading Skills survey*) (Correia *et al.*, 2023). Além disso, e ao contrário do que acontece noutros grupos de textos, neste existem textos dos três níveis de complexidade em cada subgrupo.

O capítulo seguinte apresenta e discute os resultados das tarefas descritas acima, congregando a exposição e a interpretação dos resultados.

4. RESULTADOS E DISCUSSÃO

A presente secção apresenta os resultados provenientes de cada componente definida para o objetivo do presente trabalho. Estas componentes são:

- Infraestrutura de complexidade – níveis e descritores
- A validação do *corpus*
- A validação dos níveis e descritores por parte do *focus group*
- A aplicação da Infraestrutura de complexidade numa amostra do *corpus*.

Os resultados de cada componente são interpretados e discutidos avançando hipóteses sobre a sua ocorrência.

4.1. Infraestrutura de complexidade

Após a definição e melhor entendimento do público-alvo do projeto iRead4Skills, assim como as suas relações com a literacia²⁰, bem como a análise e interpretação de níveis de complexidade existentes em referenciais considerados pertinentes – QECR, PIAAC, ALTE, ANQEP –, uma infraestrutura de complexidade foi proposta.

²⁰ cf. *Reading Skills Survey* (Correia *et al.*, 2023) e *Overall skills and gaps survey* (Clément *et al.*, 2023).

4.1.1. Os níveis de complexidade

Os níveis de complexidade propostos correspondem a três níveis de proficiência em literacia de adultos (ver anexo II). Cada um destes níveis é representado por descritores de complexidade textual (apresentados adiante), de modo a colocar o foco da complexidade no texto e não no leitor.

Os três níveis de complexidade textual são os seguintes:

- **Muito fácil:** Textos que são totalmente ou quase totalmente compreendidos por todas as pessoas, incluindo pessoas com escolaridade muito baixa (i.e., que não chegaram ao 6º ano do ensino básico) e que não têm quase nenhuma prática de leitura.
- **Fácil:** Textos que são totalmente ou quase totalmente compreendidos por pessoas com escolaridade baixa (i.e., que completaram o 6º ano do ensino básico, mas que não têm mais que o 9º ano) e que têm pouca prática de leitura.
- **Claro:** Textos que são compreendidos à primeira leitura por pessoas que completaram o 9º ano e que têm uma prática de leitura média.

Novamente, é importante reiterar que, embora se associe um nível de escolaridade a estes níveis, estes refletem a complexidade dos textos e não a proficiência dos leitores. Desta maneira, entendemos que os nomes escolhidos: **Muito Fácil, Fácil e Claro** se enquadram nesta lógica.

Uma questão a considerar nestes níveis é a sua designação. Os nomes de cada nível foram elaborados com base em dois critérios: a sua transparência e a sua versatilidade.

O primeiro aspeto – a transparência –, está relacionado com a simplicidade das designações. Tendo em conta a diversidade dos potenciais utilizadores da ferramenta final de análise de complexidade elaborada neste projeto, foi decidido que nomes transparentes e familiares contribuiriam para o acesso, compreensão e uso fáceis de cada um. Nesta senda, é importante também referir que houve sempre uma tentativa de tornar os nomes escolhidos o menos condescendente possível, de modo a não

ostracizar os leitores e as suas respetivas proficiências em literacia e de focar o texto, e não o leitor.

O segundo aspeto – a versatilidade –, tem que ver com o simples facto de as designações escolhidas serem claras e passíveis de serem transpostas para diferentes línguas e meios e aplicadas nestes (em particular, a designação de "Claro", que corresponde ao nível alvo das iniciativas nacionais e internacionais de *Plain Language*²¹).

4.1.2. Os descritores de complexidade textual

Tendo em conta os níveis de complexidade propostos, a proposta de infraestrutura de complexidade textual neste trabalho considera a análise de complexidade com base nos seguintes aspetos:

- Comprimento do texto (e.g., curto, longo)
- O propósito do texto (e.g., realização de tarefas quotidianas; lazer)
- O conceito (ou conceitos) mencionado no texto (e.g., conceitos do dia a dia; noções especializadas)
- O contexto comunicativo (ou contextos) nos quais o texto circula (e.g., contextos do dia a dia, como transportes públicos; contextos mais específicos, como textos de opiniões políticas)
- O léxico empregue no texto (i.e., relacionado ou não com um domínio específico)
- O tipo de frase utilizadas no texto (com base na sua complexidade sintática) (e.g., frases com ordens não lineares SVO)
- O tipo de formas e construções verbais no texto (e.g., tempos compostos, Modo Conjuntivo)
- A coesão textual (e.g., presença de conetores textuais)

²¹ Exemplos internacionais como a entidade do governo dos E.U.A: <https://www.plainlanguage.gov>; e nacionais como a agência CLARO: <https://claro.pt/about/>

Como resultado, estes aspetos foram desdobrados e detalhados em descritores concretos (ver anexo I). Isto resultou na proposta de 81 descritores contemplando os quatro níveis de complexidade (Monteiro *et al.*, 2023). Tal como será discutido a seguir, a noção de *descritores de complexidade textual* está relacionada com as manifestações concretas de complexidade no texto, manifestações estas subordinadas a problemas textuais. Desta forma, este conjunto de descritores abrange dimensões textuais e permite uma análise da sua complexidade com maior granularidade²².

A elaboração destes descritores baseou-se na noção de que a complexidade textual pode incluir três tipos de problemas de compreensão por parte dos leitores: **problemas superficiais, problemas de conhecimento, problemas de conteúdo** (Beck *et al.*, 1984).

O primeiro – problemas superficiais – relaciona-se com questões estritamente linguísticas, como é o caso de estruturas sintáticas complexas ou um uso inconsistente de conjunções. Por outro lado, problemas de conhecimento é algo potencialmente fora da dimensão linguística, ligado à falta de conhecimento prévio sobre o tópico presente (ou envolvente) no texto. Por último, a noção de problemas de conteúdo tem que ver com a forma como as ideias estão apresentadas e interligadas no texto, ou seja, como o texto está organizado e é comunicado. Nesta última dimensão, ideias e conceitos que se contrariam ou que levam a conclusões contraintuitivas podem levantar problemas.

A partir desta tipologia de problemas textuais, a complexidade textual foi analisada sob o prisma de observação de quatro dimensões textuais distintas: **Léxico-conceitual, Estruturas Sintáticas, Coesão, Formas Verbais**.

A escolha destas dimensões difere das dos referenciais analisados devido ao público-alvo em questão neste trabalho. Por exemplo, falantes não nativos de uma língua tendem a sentir mais complexidade do foro gramatical (Heilman *et al.*, 2007) e crianças tendem a ter menos proficiência em leitura em comparação com adultos, devido simplesmente à sua menor experiência de leitura (Mesmer *et al.*, 2012). Estes exemplos demonstram como existem especificidades inerentes a grupos de leitores, as quais não podem ser desconsideradas. Assim, tendo em conta a literatura presente neste trabalho,

²² É importante lembrar que a lista dos descritores que aqui se apresenta inclui as mudanças provenientes da validação do *focus group*.

estas categorias consideram-se indutoras de complexidade e especialmente relevantes para o público-alvo em questão – adultos de baixa literacia.

Por último, ainda que estas categorias se desdobrem em manifestações mais específicas de complexidade, estas manifestações estão todas subordinadas a estas categorias.

No que diz respeito à categoria Léxico-conceptual, é importante salientar que a infraestrutura de complexidade aqui proposta reconhece a importância desta dimensão como sendo a mais relevante para dificultar a compreensão de um texto aos leitores. Com efeito, basta o leitor desconhecer o significado de cerca de 15% das palavras de um texto para a compreensão ser prejudicada (Viana & Ribeiro, 2020). Os problemas de compreensão lexicais são os primeiros a emergir quando um leitor lê um texto (Harley, 2008).

Em relação à categoria Estruturas Sintáticas, esta infraestrutura tem em conta medidas de *leiturabilidade* tradicionais como a extensão *da frase* como responsáveis por induzir complexidade num texto, assim como manifestações qualitativas de sintaxe complexa, como *estrutura não linear* e ocorrências de *subordinação e coordenação frequente e combinada*.

Para além disto, uma característica como a *voz passiva* pode também revelar-se problemática. As construções na voz passiva tendem a manifestar-se em textos de carácter mais formal (Street & Dąbrowska, 2014). Por esta razão, leitores mais proficientes nestes “textos formais”, isto é, com práticas de leitura mais desenvolvidas e referentes a uma maior diversidade de textos, manifestam mais facilidade em lidar com frases construídas na voz passiva do que indivíduos de baixa literacia. Assim sendo, a voz passiva pode induzir complexidade num texto, para um leitor de baixa literacia²³.

²³ Para além da frequência de contacto dos leitores com esta, as razões por detrás de a voz passiva ser indutora de complexidade podem ser de natureza distintas: Harley (2013) afirma que, para a voz passiva ser processada cognitivamente, esta requer que o leitor cogite, primeiramente, o equivalente na voz ativa, para depois converter a frase de volta para a voz passiva. Por outro lado, Ambridge *et al.* (2016), destacam as propriedades semânticas do verbo, nomeadamente os seus papéis temáticos (*θ-roles*), como facilitadores ou agravantes na compreensão da estrutura passiva, sendo verbos com os papéis temáticos de Agente-Verbo-Tema os mais fáceis em compreender em frases passivas, e verbos com os papéis temáticos envolvendo Tema-Experienciador, os mais difíceis.

No que diz respeito à Coesão, esta infraestrutura mede, nomeadamente, a manifestação de coesão referencial, que se refere à composição textual (Fang & Pace, 2013), através das relações anafóricas e localizações temporais (Glenberg *et al.*, 1987).

Por último, algo que também mereceu realce foi a manifestação de complexidade proveniente das Formas e Construções Verbais (designada como Morfossintaxe nos descritores de complexidade – Anexo I). Apesar de estas poderem ser encaradas como apenas mais uma parte dos itens lexicais num texto, as formas verbais, pela sua capacidade de integrar informações aspetuais, temporais e modais, manifestam complexidade textual de uma forma particular e, por isso, foram considerados independentemente.

Os verbos são, assim, tratados como manifestações linguísticas muito flexíveis e que, por esta razão, podem ser geradores de uma complexidade particular. Enquanto, por exemplo, a presença de formas verbais simples flexionadas em tempos como o Presente, ou conjugadas em modos como o Infinitivo ou Indicativo, são vistas como claras e descomplicadas, o emprego de modos como o Conjuntivo ou tempos verbais compostos revelam-se mais difíceis de compreender, mesmo em verbos frequentes ou conceptualmente comuns. A causa desta complexidade está ligada ao facto de muitas destas formas verbais complexas serem de uso mais raro ou associado a um certo tipo de discurso (e.g., o pretérito mais-que-perfeito associado a uma linguagem literária), mas também ao facto de estas serem responsáveis por criar espaços temporais objetivamente mais complexos e difíceis de processar cognitivamente. Assim, a criação de espaços temporais que aludem a mais do que uma localização temporal, como acontece várias vezes em tempos verbais compostos, conduzem a uma dificuldade na conceptualização da ação em questão (e.g., “O João tinha comido o bolo quando o Rui entrou em casa”).

Nesta senda, também as perífrases verbais foram contempladas. Expressões como “estar a + Verbo no Infinitivo”, “começar a + Verbo no Infinitivo”, etc., são manifestações linguísticas de aspeto gramatical, que, embora possuam uma natureza composta, se revelam bastante “decifráveis” para todos os falantes. A forma como o tempo (ou

tempos) e aspeto são codificados nas manifestações linguísticas, seja por flexão verbal, por meio de advérbios ou através de perífrase verbais, parece revelar uma hierarquia de complexidade, que segue na senda do entendimento da complexidade textual de Dahl (2007), referido na secção 2.2., e que considera a informação linguística embutida na manifestação linguística em questão. Assim, tendo em conta que é através da flexão verbal que mais informação linguística é concentrada numa única palavra, e que as perífrases verbais “diluem” esta informação linguística, podemos considerar esta hierarquia como tendo a primeira como mais complexa e a última como menos complexa.

Com tudo isto em mente, esta infraestrutura efetua a sua aferição de complexidade de uma forma maioritariamente qualitativa, isto é, através de descritores objetivos referentes a manifestações de complexidade. Porém, e embora em menor escala, também são contabilizadas avaliações quantitativas da complexidade, nomeadamente no que diz respeito ao comprimento do texto.

4.2. A validação do *focus group*

O processo de validação contou com o discernimento de um grupo de formadores de Centros Qualifica e teve como objetivo validar os níveis de complexidade propostos, bem como os descritores associados a cada um. Para além da tarefa preparatória que congregou 12 destes formadores – efetuada via inquérito online – foi realizada uma reunião de discussão com este *focus group*. O resultado da tarefa preparatória e desta reunião levou a três alterações de nível de complexidade de três descritores:

Descritor	Nível (pré-focus group)	Nível (pós focus group)
O texto tem palavras compostas cujo sentido não é facilmente compreendido através da combinação das duas partes (e.g., <i>meio-termo, porta-voz, obra prima</i>).	Claro	Mais Complexo
As frases no texto incluem frequentemente orações relativas correspondentes ao sujeito da frase (e.g., « <i>Francisca, a quem os amigos tratam por “Kika”, começou a surfar muito pequena.</i> »).	Claro	Mais Complexo
O texto apresenta sentidos figurativos menos comuns (e.g., « <i>O Barcelona varreu o Real Madrid; O Estoril esmagou o Santa Clara.</i> »).	Claro	Mais Complexo

Tabela 1. Descritores que mudaram de nível

Os três descritores mencionados acima partilham o facto de a sua mudança ter sido efetuada do nível *Claro* para o nível *Mais complexo*. Na prática, esta mudança fez com que estes descritores fossem retirados da infraestrutura de análise de complexidade, uma vez pertencerem a um nível fora do escopo do projeto.

Para além disto, algo que foi revisto e posteriormente alterado foi a frequência relacionada com a ocorrência das marcas linguísticas indutoras de complexidade nos descritores. Certos descritores distinguiram-se de outros no que diz respeito ao seu nível de complexidade por meio da frequência com a qual se manifestavam num texto. Por exemplo, os descritores “*o texto tem poucos estrangeirismos comuns*” e “*O texto faz uso de estrangeirismos comuns frequentemente*”, pertencem a níveis de complexidade distintos, sendo o primeiro do nível *Fácil* e o segundo do nível *Claro*. Esta abordagem considerava que quanto mais alta a frequência de certo descritor num texto, mais complexidade este traria ao mesmo texto. Contudo, depois desta revisão com o *focus group*, este contestou o grau de frequência em alguns descritores, afirmando que, na sua experiência, a frequência com que alguns descritores se manifestavam num texto não era sinónimo de maior ou menor complexidade. Desta forma, concluiu-se que certos descritores induziriam complexidade no texto, independentemente do número de vezes em que ocorriam no texto.

Em todo o caso, foi decidido em conjunto que apenas alguns descritores operariam neste *continuum* de frequência de ocorrências.

Estes descritores são referentes a:

- ✓ A frequência do grau de concretude das palavras (descritores 20, 21)
- ✓ A frequência dos estrangeirismos presentes no texto (descritores 35, 36, 37).
- ✓ A frequência de orações relativas (descritores 53, 54, 55, 56)
- ✓ A frequência da presença da voz passiva (descritores 57, 58, 59)

Estes descritores enquadram-se no campo léxico-conceptual (os dois primeiros) e sintático (os dois últimos).

A continuidade dos dois primeiros descritores na infraestrutura de complexidade segue em linha com a ideia de um limite de 15% das palavras de um texto para a compreensão do mesmo (Viana & Ribeiro, 2020). Ou seja, aquando da consideração de complexidade lexical, existe uma margem – ainda que curta – para a consideração da frequência da sua manifestação.

Por sua vez, os outros dois descritores, sendo de natureza sintática, enquadram-se numa avaliação de leituraabilidade (Dale & Chall, 1948; Flesch, 1948; Gunning, 1952; Kincaid *et al.*, 1975; Lorge, 1948; McLaughlin, 1969; Spache, 1953), cuja medida de aferição de complexidade é o número de ocorrências presentes no texto.

Em suma, tendo em conta que a alteração dos três descritores se reportou a uma mudança do nível mais alto de complexidade (Claro) para o nível fora do escopo (Mais Complexo), a ilação é de que, conforme se avança nos níveis, a distinção entre os níveis de complexidade se torna mais difícil de precisar e assinalar. Desta forma, as fronteiras destes níveis superiores são mais “nebulosas”, havendo uma sobreposição maior entre níveis e uma menor distinção entre os seus descritores.

4.3. O *Corpus* – teste e validação dos níveis e descritores de complexidade

A recolha de um *corpus* permitiu testar a infraestrutura através da sua aplicação nos textos, isto é, a sua adequação prática em textos. Por outro lado, o *corpus* seria classificado e compartimentado com base nos níveis propostos.

O processo de compilação do *corpus* teve duas fases distintas: a recolha e classificação dos textos e a validação final da classificação dos mesmos. Os resultados que aqui se apresentam são referentes ao *corpus* validado. O resultado desta validação

foi um total de 2186 textos e um total de 942 818 *tokens* (Pintard *et al.*, 2023). A repartição deste número total de textos por nível de complexidade foi a seguinte:

Nível de Complexidade			
Mais complexo	Claro	Fácil	Muito fácil
38,3%	28,6%	29,3%	3,8%

Tabela 2. Distribuição dos textos por nível de complexidade

Como se pode ver pela tabela 2, a validação do *corpus* através da infraestrutura de complexidade resultou numa repartição de textos desequilibrada, ficando patente a escassez de textos do primeiro nível de complexidade – Muito Fácil. A maior percentagem de textos validados enquadra-se no nível de complexidade mais alto do projeto – Mais Complexo –, nível este que não é tido em conta para os objetivos do projeto, servindo apenas como nível de abalizamento.

Estas percentagens revelam, claramente, a dificuldade de recolha de textos que se enquadrem em níveis de complexidade mais baixa, nomeadamente no nível Muito Fácil. De modo geral, textos cujo objetivo comunicacional se inseria no de Comunicação Pessoal foram os mais associados ao nível Muito Fácil, enquanto todos os outros revelaram o contrário. Este fenómeno pode ser entendido como estando relacionado com a complexidade conceptual do texto, já aqui referida. Estes resultados evidenciam a ligação que existe entre léxico complexo e conceitos complexos (Fang & Pace, 2013; Harley, 2013; Pearson & Hiebert, 2014). Tendo em conta que textos enquadrados em Comunicação Pessoal têm uma maior probabilidade de aludirem a conceitos do quotidiano (assim como uma maior inclinação para os transmitirem de uma forma discursiva informal), torna-se mais óbvio por que razão textos que não se limitam a conceitos do dia a dia possam não verificar os requisitos para serem classificados como do primeiro nível.

Por outro lado, algo igualmente importante no que diz respeito à escassez de textos do primeiro nível de complexidade é a ilação extralinguística que pode daqui ser extraída. Se estivermos certos e, de facto, o nível Muito fácil representar os textos compreendidos por adultos do mais baixo nível de literacia, então os textos que lhes são

acessíveis são poucos. Isto valida a principal motivação deste trabalho, uma vez que, com a exceção de livros infantis, há poucos textos escritos para pessoas com poucas competências de leitura. Desta maneira, a vivência em sociedade desta parte de população encontra-se obstruída pelo número baixo de textos que este consegue compreender.

4.4. A aplicação da infraestrutura de complexidade

Com vista a ensaiar um outro teste sobre a infraestrutura de complexidade proposta, uma amostra do *corpus* foi isolada e testada.

Esta amostra consiste no grupo *Comunicação Pessoal*. Dentro deste grupo, existem quatro tipos de documento: *Diário*, *Carta*, *Lista*, *Bilhete*. Para esta atividade, pretendeuse considerar 10 textos por cada subgrupo, por nível de complexidade. Este número nem sempre foi atingido, uma vez que certos tipos de documentos correspondentes a certos níveis de complexidade não foram recolhidos. Três subgrupos ficaram incompletos. Assim, no total, esta amostra de *corpus* foi formada por 108 textos. A tabela 3 mostra a distribuição destes textos.

À altura da análise que aqui se apresenta, os textos encontravam-se validados e classificados para um nível específico. Assim, o exercício consistiu em averiguar a ocorrência dos descritores de complexidade pertencentes ao nível no qual um certo texto se inseria, contando a frequência da sua realização. Esta contagem foi possibilitada pela natureza binária da validação que tinha sido feita (cf. Secção 3.2.3. – Validação *corpus*).

<i>Textos\Níveis</i>	<i>Muito Fácil Claro</i>		
	<i>Fácil</i>		
<i>Diário</i>	4	10	10
<i>Carta</i>	10	10	10
<i>Lista</i>	10	10	10
<i>Bilhete</i>	10	6	8
<i>TOTAL:</i>	34	36	38

Tabela 3. Amostra utilizada na aplicação da infraestrutura de complexidade

É importante mencionar que esta contagem reflete a ocorrência dos descritores previstos num nível. Ora, se um grupo de textos estiver classificado como Muito Fácil, por exemplo, e descritores de complexidade desse mesmo nível ocorrerem com uma frequência baixa nesses mesmos textos, isto não quer dizer que estes textos apresentem descritores mais complexos, mas sim que simplesmente não apresentam os descritores previstos neste nível.

Assim sendo, o objetivo foi observar como os descritores de complexidade ocorriam em grupos de textos particulares e comparar os seus resultados, de modo a mostrar a sua relevância. Pressupunha-se, assim, que os Propósitos de comunicação dos textos podem condicionar a complexidade patente nos seus textos, quer possibilitando, quer inibindo manifestações específicas de complexidade, o que, neste caso, se traduzia na ocorrência de certos descritores.

Para além disto, esperava-se também que esta análise respondesse à grande crítica efetuada em relação às avaliações qualitativas de complexidade, que acusava estas avaliações de não atribuir peso ou importância diferentes às características de complexidade contempladas na análise (Pearson & Hiebert, 2014). Através desta aplicação, esperava-se também identificar a categoria (ou categorias) que mais importância revelaria aquando da aferição de complexidade.

A tabela 4 apresenta os resultados desta aplicação. Através da apreciação dos valores percentuais registados em cada dimensão de complexidade textual, atentou-se a dois aspetos: i) a evolução da sua manifestação, de nível para nível; ii) a sua manifestação em diferentes tipos de documento. Com isto em mente, duas hipóteses se consideravam: i) existe uma influência das práticas de comunicação inerentes aos tipos de documento na manifestação de complexidade nestes textos; ii) a configuração do texto condiciona a manifestação da sua complexidade.

Níveis	Descritores		Comunicação Pessoal (ocorrência de descritores)							
			Diário		Carta		Lista		Bilhete	
Muito Fácil	Léxico-conceptual	Léxico básico	3	13/16 (81%)	5	30/40 (75%)	6	31/40 (78%)	10	29/40 (73%)
		Ideias concretas	4		8		10		10	
		Quotidiano	4		9		8		9	
		Afixação rara	2		8		7		0	
	Formas Verbais	Tempos simples	3	7/20 (35%)	4	12/40 (30%)	4	12/40 (30%)	4	9/40 (23%)
		Sem verbos auxiliares	0		4		5		0	
		Verbos copulativos	0		2		2		2	
		Perífrases verbais e verbos modais	4		2		1		3	
	Estruturas sintáticas	Coordenação Simples	3	5/8 (63%)	9	15/20 (75%)	2	6/20 (30%)	5	7/20 (35%)
		Subordinação Simples	2		6		4		2	
	Coesão	Poucas anáforas	3	7/8 (88%)	7	11/20 (55%)	3	8/20 (40%)	3	7/20 (35%)
		Uma localização temporal	4		4		5		4	
Fácil	Léxico-conceptual	Léxico específico	9	27/40 (68%)	8	24/40 (60%)	9	24/40 (60%)	5	13/24 (54%)
		Ideias menos concretas	9		5		2		4	
		Âmbitos para lá do quotidiano	3		6		5		4	
		Afixação recorrente	6		5		8		0	
	Formas Verbais	Modo Conjuntivo	4	12/40 (30%)	4	13/40 (33%)	2	6/40 (15%)	1	2/24 (8%)
		Verbos auxiliares	6		8		4		1	
		Verbos copulativos menos frequentes	0		0		0		0	
		Verbos modais menos frequentes	2		1		0		0	
	Estruturas sintáticas	Coordenação c/ conjunções menos frequentes	2	14/30 (47%)	3	13/30 (43%)	4	18/30 (60%)	3	8/18 (44%)
		Subordinação mais complexa	9		7		9		2	
		Voz Passiva	3		3		5		3	
	Coesão	Anáforas e elipses frequentes	7	12/20 (60%)	7	10/20 (50%)	4	8/20 (40%)	2	3/12 (25%)
		Mais do que uma localização temporal	5		3		4		1	
Claro	Léxico-conceptual	Léxico especializado	9	29/40 (73%)	9	23/40 (58%)	10	26/40 (65%)	6	12/32 (38%)
		Ideias abstratas ou especializadas	7		5		5		5	
		Contextos particulares/técnicos	5		3		5		0	
		Afixação constante	8		6		6		1	
	Formas Verbais	Todos os tempos e Modos	9	11/20 (55%)	9	16/20 (80%)	5	5/20 (25%)	3	3/16 (19%)
		Tempos verbais compostos frequentes	2		7		0		0	
	Estruturas sintáticas	Coordenação e subordinação frequentes e interligadas	10	13/20 (65%)	9	15/20 (75%)	7	9/20 (45%)	1	1/16 (7%)
		Voz Passiva c/ Passiva “-se”	3		6		2		0	
	Coesão	Anáforas (de todos os tipos) frequentes	5	9/20 (45%)	3	9/20 (45%)	6	10/20 (50%)	4	5/16 (31%)
		Localizações temporais variadas e não lineares	4		6		4		1	

Tabela 4. Resultados da análise de manifestações de complexidade por descritor

Léxico-conceptual

Os quatro tipos de documentos apresentam resultados interessantes. É patente a ocorrência dos descritores da categoria Léxico-conceptual, de forma uniforme, em todos os níveis de complexidade. A sua mais patente manifestação dá-se no nível *Muito Fácil*. As percentagens referentes a esta categoria demonstram uma queda dos seus valores nos seguintes níveis de complexidade, em comparação com o nível Muito Fácil. Em regra geral, textos em Diários parecem manter um valor de ocorrências de descritores léxico-conceptuais alto, contudo, os restantes tipos de documento assistem a um decrescimento.

Estes resultados evidenciam dois factos interessantes. Em primeiro lugar, estes resultados revelam que a complexidade proveniente da categoria Léxico-conceptual parece ser a que mais relevância tem nos três níveis de complexidade, visto que a ocorrência dos seus descritores se evidencia frequente em todos os níveis. Apesar de uma redução de manifestações nos níveis mais complexos (Fácil e Claro), este tipo de complexidade está presente em todos os textos, assim como em todos os níveis.

O facto de textos em Diários manterem um valor alto de manifestações dos descritores em cada categoria, contrastando com a descida de valores nos outros tipos de documento é de realçar. Antes de tudo, lembre-se que o objetivo aqui não é afirmar que textos em Diários são mais complexos do que textos nos outros tipos de documento. O objetivo é, sim, destacar como cada tipo de documento manifesta complexidade. Com isto em mente, é necessário ter em conta que cada nível de complexidade contempla certos descritores. A ocorrência de descritores de um certo nível em textos classificados como desse mesmo nível indica a existência de complexidade de uma categoria nesses textos, tal como é prevista pelos descritores. Por outras palavras, a ocorrência dos descritores desta categoria num texto somente valida a sua complexidade léxico-conceptual e não a sua complexidade global.

Assim, no que diz respeito ao caso dos textos em Diários, que revelam uma ocorrência frequente de descritores em todos níveis, parece que, na origem disto, pode estar o facto de estes textos poderem conter marcas de subjetividade regulares, como opiniões, assunções, presunções e idealizações, próprias do teor deste tipo de texto. Esta hipótese em torno da

noção de subjetividade alicerça-se em algo contemplado na infraestrutura de aferição de complexidade – que a concretude de ideias e tópicos nos textos está associada a um nível de complexidade lexical mais baixo e vice-versa. Com efeito, note-se como descritores como *ideias menos concretas* (Fácil) e *ideias abstratas ou especializadas* (Claro), ocorrem frequentemente nos níveis superiores. A juntar a isto, descritores como *léxico específico* e *léxico especializado* também ocorrem frequentemente nos níveis superiores, sendo estes quatro descritores os maiores responsáveis pelo fenómeno aqui analisado.

Textos em Diários possuem, assim, a capacidade de evidenciar complexidade proveniente da categoria em questão de uma forma bastante contundente, contrastando com os outros tipos de documentos, cuja ocorrência de descritores em níveis altos de complexidade está vinculada de modo mais forte a outra categoria.

Por último, algo que também aqui pode ser deduzido é o facto de os descritores desta categoria aferirem adequadamente a complexidade nos três níveis. Relembre-se que este grupo de textos se encontra validado no *corpus* nos seus respetivos níveis. Assim, esta consonância destes descritores com o respetivo nível de complexidade dos textos revela a sua afinação e ajuste a cada nível de complexidade.

Estruturas sintáticas

De seguida, constata-se que um número elevado de textos verifica a ocorrência de descritores da categoria Estruturas Sintáticas. Em todos os níveis, a ocorrência dos descritores de complexidade sintática é a que revela mais constância – à exceção do primeiro, onde os descritores de complexidade lexical se evidenciam mais.

Contudo, uma particularidade se salienta. Em dois dos três níveis (Muito Fácil e Claro), os textos em Listas e Bilhetes apresentam os valores de ocorrência de descritores mais baixos desta categoria (Listas – Nível Muito Fácil – 30%), estando até um deles entre os valores mais baixos de toda a análise em questão (Bilhetes – Nível Claro – 7%). De forma curiosa, o oposto acontece no nível Fácil, no qual a maior manifestação de complexidade sintática se dá no tipo

de documento Listas. Ainda que em minoria, e pelo que se vai propor de seguida, este acontecimento considera-se muito interessante.

Em relação ao primeiro nível de complexidade, é importante notar que esta categoria integra descritores de complexidade como *subordinação* e *coordenação simples*. Tendo em conta que estes pressupõem um nível básico de gramaticalidade, é possível que a razão pela qual esta categoria não possua os seus descritores como os que mais ocorrências verificam neste nível esteja relacionada com a configuração inerente a certos textos. Regra geral, textos em Listas são, na maioria dos casos, enumerações de elementos lexicais, geralmente não relacionados entre si, isto é, que não estão interligados sintaticamente. Desta forma, a necessidade de coordenação e/ou subordinação não existe, provindo daí a ocorrência baixa destes descritores neste primeiro nível.

Nos níveis superiores de complexidade, os descritores desta categoria como *subordinação mais complexa* ou *voz passiva* tendem a apresentar números de ocorrências baixas em textos em Bilhetes e Listas. O facto de este tipo de textos ter, geralmente, mensagens curtas e diretas pode ser a razão para a baixa ocorrência destes descritores nestes níveis.

Esta possibilidade torna-se ainda mais patente quando as ocorrências destes descritores nos textos em Diários e Cartas são comparadas com as ocorrências dos textos em Bilhetes e Listas. Existe uma natureza comunicativa mais direta no teor de um texto em Bilhetes e Listas do que num texto em Diários e Cartas. Desta forma, no primeiro par pode inibir-se o surgimento de descritores de complexidade sintática de níveis mais elevados, ao passo que no segundo par, isto não acontece.

Da mesma forma, o suporte de textos em Bilhetes e Listas, que tende a ser reduzido (e.g., formato de folha em A7, A8 ou A9, *post-it*), pode estar por detrás de um condicionamento da ocorrência dos descritores, dado que um espaço reduzido de escrita implica uma transmissão da mensagem de uma forma mais direta, por vezes telegráfica.

Desta forma, os descritores desta natureza sintática, especialmente nos níveis superiores de complexidade, não se manifestam com frequência nos textos em questão. Isto deve-se, possivelmente, ao teor e formato dos textos que inibem o aparecimento destes descritores.

Coesão

No que concerne esta categoria, podemos afirmar que os seus descritores ocorrem de uma maneira relativamente constante, em todos os níveis. Com valores um pouco abaixo da categoria abordada acima, os descritores desta categoria ocorrem com mais frequência nos textos do primeiro nível de complexidade. Embora seja patente um decrescimento nos níveis de complexidade mais altos (Fácil e Claro), os seus valores mantêm-se proporcionais, muito à semelhança da categoria Léxico-conceptual.

A relação desta categoria com os tipos de documentos em questão é óbvia. Os dois tipos de documento nos quais os descritores expectáveis desta categoria mais ocorrem são os Diários e Cartas. Uma exceção apresenta-se no último nível (Claro), onde mais descritores desta categoria ocorrem em textos de Listas. No entanto, por estarmos perante uma diferença tão mínima – trata-se de apenas menos uma ocorrência do que em Diários e Cartas –, esta não se considera significativa.

Considerando que textos em Listas são, várias vezes, textos em que não há ligação ou estrutura frásica entre as expressões/palavras, e textos em Bilhetes tendem a ser curtos, isto conduz ao entendimento de que existe uma superfluidade de elementos de coesão textual nestes textos, isto é, que elementos de coesão textual tendem a não ocorrer. Descritores relacionados com a presença de anáforas e várias localizações temporais são, geralmente, inibidos de ocorrer devido a serem prescindíveis neste tipo de texto.

Por oposição, devido à natureza fortemente narrativa de textos em Diários e Cartas, os descritores característicos desta categoria tendem a surgir muito mais frequentemente. Contudo, existe um decrescimento no se aparecimento, à medida que aumentam os níveis de complexidade, também nestes tipos de documento. Descritores como *anáforas (de todos os tipos) frequentes* e *localizações temporais variadas e não lineares* (nível Claro) têm uma baixa frequência de ocorrência, não chegando a metade das suas possíveis ocorrências. Esta escassez de ocorrências de descritores previstos nos níveis mais avançados, nomeadamente no último, pode ser interpretada da seguinte forma: pode existir, de facto, uma influência das práticas de comunicação nos tipos de documento que condiciona a manifestação dos

descritores de complexidade desta categoria nestes textos. No entanto, ao contrário do que se propôs em relação aos textos em Listas e Bilhetes, em que a sua natureza não narrativa pode inibir a ocorrência de certos descritores, nesta situação, não se consegue encontrar uma razão sustentada para os descritores não surgirem nos níveis mais avançados.

De qualquer maneira, esta categoria evidencia algo muito interessante. Torna-se novamente patente – após o ter sido na categoria Estruturas Sintáticas – que certos textos nunca irão revelar uma complexidade sintática ou de coesão referentes a níveis de complexidade elevados, dado que não possuem estruturas sintáticas de todo (ou estas são escassas), nem elementos de coesão.

Formas verbais

Em último lugar, é interessante notar como a categoria cujos descritores menos ocorrem é a de *Formas Verbais*. Esta categoria mantém, praticamente, os mesmos valores nos dois primeiros níveis de complexidade aumentando, significativamente, no último.

Apesar da baixa frequência de ocorrência dos descritores desta categoria, os textos em Cartas e Diários são os que registam valores mais elevados. Visto serem textos narrativos, considera-se que existe a possibilidade de uma dispersão da narrativa, ou seja, da especulação de histórias e narrativas paralelas. Esta dispersão de narrativa pode recorrer a tempos (Futuro) e modos verbais (Conjuntivo e Condicional) associados à expressão de valores modais que implicam a referência e a interpretação de pressupostos. Estes valores modais podem ter uma natureza epistémica (e.g., um texto em Diário onde se acha que alguém cometeu um crime), uma natureza deôntica (e.g., um texto em Carta onde se dá um parecer sobre algo), uma natureza apreciativa (e.g., um texto em Diário onde se julga uma ação e as consequências desta sobre alguém), ou até um valor mediativo (Futuro do Pretérito do Indicativo – “ele terá viajado”).

Esta motivação pode estar por detrás da ocorrência baixa de descritores no primeiro e segundo níveis de complexidade. Ademais, tendo em conta que esta escassez de ocorrência

de descritores se dá no primeiro nível de complexidade, isto pode sugerir que existe uma ligação forte entre as estratégias de expressão de valores modais e conjuntivos e os tipos de documento Cartas e Diários.

No que diz respeito a textos que integram Bilhetes e Listas, estes revelam ocorrências residuais dos descritores verbais em todos os níveis. A razão para isto pode estar relacionada com o facto de estes textos não invocarem a mesma dispersão narrativa referida atrás. A presença pobre dos descritores desta categoria está também diretamente relacionada com a ausência de formas verbais. Mais uma vez, textos em Listas, que fazem uso quase exclusivamente de nomes, possuem uma escassez de outras classes de palavras. Porém, quando formas verbais ocorrem, estas tendem a ser, frequentemente, formas simples, como o Imperativo (previsto no nível Muito Fácil), recorrente em textos em Bilhetes.

Deste jeito, as baixas ocorrências dos descritores desta categoria em textos em Diários e Cartas podem ser justificadas pelo facto de os textos simplesmente não fazerem uso dos descritores dos níveis mais baixos, evidenciando uma forte ligação com os descritores do nível Claro; enquanto no caso dos textos em Bilhetes e Listas, a justificação pode passar pela inexistência de formas verbais em geral.

Em suma, a análise realizada permite observar que existem oscilações na ocorrência dos descritores e no peso destes na determinação de nível de complexidade de um texto. Note-se que, ainda que descritores de uma dada categoria não ocorram em determinados textos, estes foram validados como pertencendo a um certo nível, o que significa que outros descritores ocorrem neles.

Primeiramente, salvo a categoria Formas Verbais, todas as categorias evidenciam uma redução na ocorrência dos seus descritores nos textos, ao longo dos níveis de complexidade. Este fenómeno torna-se mais patente nos níveis de complexidade Fácil e Claro. As razões para esta baixa seguem em linha com as hipóteses iniciais. Em todas as categorias, a hipótese que considerava a influência das práticas de comunicação nos textos na manifestação de complexidade nestes textos, verificou-se. A ocorrência de descritores de certas categorias é inibida (Léxico-Conceptual, Estruturas Sintáticas e Coesão) ou fomentada (Formas Verbais)

devido ao teor da mensagem comunicativa do texto em si. Na categoria Estruturas Sintáticas, a hipótese que considerava a configuração do texto, isto é, o seu suporte, como condicionante foi também verificada. O facto de textos terem um suporte reduzido pode estar na origem da inibição de certos descritores de níveis de complexidade mais altos.

Para além disto, os descritores da categoria Léxico-conceptual ocorreram de forma frequente em todos os tipos de documento e todos os níveis. Esta estabilidade mostra a sua relevância para a infraestrutura de complexidade. É este o tipo de complexidade que mais universalmente e frequentemente se manifesta nos tipos de documento analisados aqui. Este fato parece atestar a importância atribuída à categoria Léxico-conceptual, no que concerne a sua capacidade de dificultar a compreensão aos leitores (Harley, 2008), anteriormente advogada neste trabalho.

Finalmente, a manifestação de complexidade está também condicionada pelas particularidades dos textos em si – diferentes géneros textuais manifestam complexidade textual através de dimensões linguísticas diferentes (Hiebert, 2011).

5. CONCLUSÃO

Nesta dissertação, foram apresentados o desenvolvimento, a validação e a aplicação de uma infraestrutura de complexidade textual. Enquadrado no projeto iRead4Skills, o objetivo por detrás deste trabalho era contribuir para a criação de uma infraestrutura de complexidade textual, de modo a determinar níveis de proficiência de leitura adequados para adultos de baixa literacia. Dentro deste objetivo, foram propostos níveis de complexidade e descritores para cada um desses níveis. De modo a afinar o resultado do trabalho, foi recolhido um *corpus* de textos autênticos e conduzida uma validação desses descritores e níveis por formadores de centros EFA.

Em relação à proposta inicial da infraestrutura de complexidade, esta assentou conceptualmente em referenciais de proficiência em linguagem já existentes e em execução, mas muitas vezes relativos a falantes não nativos ou em processo de aprendizagem de uma

língua segunda. Depois de uma análise e nivelamento destas, foram propostos três níveis de complexidade, adequados a adultos de baixa literacia. Caracterizando estes níveis, foram determinadas quatro grandes categorias linguísticas como sendo indutoras de complexidade nos textos: Léxico-conceptual, Estruturas sintáticas, Coesão e Formas Verbais. Abrangendo os quatro níveis de complexidade e relativos a estas categorias, foram propostos 81 descritores.

Estes níveis e descritores foram propostos à validação de formadores de centros de aprendizagem para adultos, o que resultou na alteração de nível de três descritores. Esta alteração, ainda que mínima, evidenciou a natureza fluída tanto dos níveis como dos seus descritores, nomeadamente à medida que os níveis vão aumentando. À vista disto, considerase que o conhecimento empírico do *focus group* participante nesta tarefa foi vital para garantir a aplicabilidade prática da infraestrutura, para lá dos fundamentos teóricos.

Por sua vez, esta infraestrutura foi essencial à compilação e validação de um *corpus* de textos autênticos relativos aos níveis relevantes para o projeto. Assim, entende-se este *corpus* como um conjunto de dados essencial para os futuros passos e avanços do projeto iRead4Skills.

A aplicação da infraestrutura de complexidade também permitiu revelar como alguns tipos de texto manifestam a sua complexidade de uma maneira particular e constante. Ainda que uma associação entre género textual e/ou estrutura de texto estivessem ambas previstos como fatores importantes na forma como um texto manifesta complexidade (Hiebert, 2011; Mesmer *et al.*, 2012), o exercício de aplicação e análise da infraestrutura de complexidade permitiu confirmar que certas categorias linguísticas têm, efetivamente, tendência a ocorrer mais em certos textos do que em outros e que há diferenças de peso relativo entre as categorias de descritores propostas.

A infraestrutura de complexidade e a consequente análise presentes neste trabalho constituem um contributo para a aferição de complexidade em textos destinados a adultos de baixa literacia. Atendendo a que uma grande parte da população adulta dentro das sociedades atuais é, cada vez mais, obrigada a atualizar-se e adaptar-se às constantes mudanças tecnológicas, sociais e do mercado de trabalho, é expectável que a população que busca

desenvolver a sua literacia não diminua. Novas e constantes etapas de aprendizagem se adivinham para os indivíduos do futuro, tornando a aprendizagem ao longo da vida um requisito para uma vida potenciada ao máximo. Desta forma, pretende-se que este trabalho não só contribua para refletir sobre esta questão, como também possa ser a base para uma discussão sobre como contrariar futuros níveis baixos de literacia na população. A ideia de que a “literacia alta de hoje” pode ser “a literacia baixa de amanhã”, mencionada no início deste trabalho, será sempre atual.

REFERÊNCIAS

- ACT. (2006). *Reading between the lines: What the ACT reveals about college readiness in reading*.
- ALTE. (2002). The ALTE CAN DO PROJECT (1992–2002). Consultado a 6 de janeiro de 2023. Disponível em: www.alte.org/can_do/alte_cando.pdf.
- Alves, J. (2005). Literacia e modernização económica. In Helena Moura (Org.). *Diálogos com a Literacia*. Lisboa Editora, 15-22.
- Alves, A. C. & Arqueiro, A. S. & Dionísio, M. (2012) Vidas letradas de adultos em processo de Reconhecimento, Validação e Certificação de Competências. In Pereira, I. & Dias, I. & Pinto, H. & Menino, H. & Cadima, R. (orgs.). *Conferência Internacional — investigação, práticas e contextos em Educação*. Escola Superior de Educação e Ciências Sociais – Instituto Politécnico de Leiria.
- Amaro, R., & François, T. (2023). iRead4Skills - Annotation Schema (V1.0). Zenodo. <https://doi.org/10.5281/zenodo.10069877>
- Ambridge, B., Bidgood, A., Pine, J. M., Rowland, C. F., & Freudenthal, D. (2016). Is Passive Syntax Semantically Constrained? Evidence From Adult Grammaticality Judgment and Comprehension Studies. In *Cognitive Science*, 40(6), 1435–1459. <https://doi.org/10.1111/cogs.12277>
- Anderson, J.R. (2000). *Cognitive psychology and its implications*. W.H. Freeman.
- ANQEP. (2023). *News*. Consultado a 6 de janeiro de 2023. Disponível em: <https://www.anqep.gov.pt/np4/home>
- Ávila, P. (2008). *A Literacia dos Adultos. Competências-Chave na Sociedade do Conhecimento*. Lisboa: Celta Editora.
- Banda, S. & Mpolomoka, D. (2012). If you cannot read, Forget about the other skills. In *Better Skills, Better Jobs, Better Lives*. OECD.
- Batista, A. (2021). Literacia, textos híbridos e práticas educativas. *Cadernos Pedagógicos*, 1, 9–33.
- Beck, I.L, McKeown, M.G, Omanson R., & Pople, M. (1984). Improving the comprehensibility of stories: The effects of revisions that improve coherence. In *Reading Research Quarterly*. 19. 263-277.

- Benavente, A. (coord.); Rosa, A.; Costa, A.; Ávila, P. (1995). Estudo Nacional de Literacia: Relatório Preliminar. Lisboa: Instituto de Ciências Sociais - Universidade de Lisboa.
- Carvalho, J. (2003). *Escrita: percursos de investigação*. Braga: Universidade do Minho, Instituto de Educação e Psicologia, Departamento de Metodologias da Educação.
- Castro-Caldas, A & Reis, A. (2000). Neuropsicologia do analfabetismo. Considerações a propósito de um projecto em desenvolvimento. In *Literacia e Sociedade*. Lisboa. Caminho.
- Chall, J. S., Bissess G., Conrad, S., & Harris-Sharples, S. (1999). *Qualitative assessment of text difficulty*. Brookline Publishers.
- Clément, F., Hauret, L., & Amaro, R. (2023). iRead4Skills - Overall skills and gaps survey (V1.0). Zenodo. <https://doi.org/10.5281/zenodo.10181135>
- Comissão das Comunidades Europeias. (2006, Outubro). *Educação de adultos: nunca é tarde para aprender*. Comunicação da Comissão. Bruxelas.
- Correia, S., Amaro, R., & Gauchola, R. (2023). iRead4Skills - Reading skills survey (V1.0). Zenodo. <https://doi.org/10.5281/zenodo.10179536>
- Costa, J., Lobo, M, Silva, C, & Ferreira, E. (2009). Produção e Compreensão de Orações Relativas em Português Europeu: Dados do Desenvolvimento Típico, de PEDL e do Agramatismo. Alexandra Fiéis & Maria Antónia Coutinho (orgs.): *XXIV Encontro Nacional da Associação Portuguesa de Linguística. Textos Seleccionados*. Lisboa: APL/Edições Colibri.
- Council of Europe (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing, Strasbourg. Available at: www.coe.int/lang-cefr.
- Cunha, J. & Dionísio, M. (2019). O lugar da literacia nas ofertas de educação e formação de jovens em Portugal. In Dantas, T. & Dionísio, M. & Loffin, M. (orgs.). *Educação de jovens e adultos: políticas, direitos, formação e emancipação social*.
- Cunningham, J. W., & Mesmer, H. A. (2014). Quantitative measurement of text difficulty: What's the use? In *Elementary School Journal*. 115, 255-269.
- Dionísio, M. & Silva, A. & Castro, R. (2014). Vidas de literacia – (re)configurações da relação com o escrito entre adultos em processos de RVCC. In Alves, N. &

- Rummert, S. & Marques, M. (org.). *Educação e Formação de Jovens e Adultos em Portugal e no Brasil: Políticas, Práticas e Atores*. Instituto de Educação da Universidade de Lisboa.
- Dahl, Ö. (2007) Definitions of complexity. In *Proceedings of the Colloquium on complexity, accuracy and fluency in second language use, learning & teaching*. 41-46.
- Dale, E. & Chall, J. (1948). A Formula for Predicting Readability. In *Educational Research Bulletin*. 27, 28. 11-20, 37-54.
- Douglas, Y., & Miller, S. (2016). Syntactic and Lexical Complexity of Reading Correlates with Complexity of Writing in Adults. In *International Journal of Business Administration*, 7(4).
- DuBay, W. H. (2008). Unlocking Language: The Classic Readability Studies. *IEEE Transactions on Professional Communication*, 51(4), 416–417.
<https://doi.org/10.1109/tpc.2008.2007872>
- Fang, Z., & Pace, B. G. (2013). Teaching With Challenging Texts in the Disciplines. In *Journal of Adolescent & Adult Literacy*. 57(2). 104–108.
- Fernandez-Duque, D. (2009). Cognitive and neural underpinnings of syntactic complexity. *Typological Studies in Language*, 433–460.
<https://doi.org/10.1075/tsl.85.16cog>
- Ferreiro, E. & Teberosky, A. (1991). *Psicogénese da Língua Escrita*. Porto Alegre: Artes Médicas.
- Flesch, R. (1948). A new readability yardstick. In *Journal of Applied Psychology*. 32. 221-233.
- Flor, M., Klebanov, B., Sheehan, K. (2013). Lexical tightness and text complexity. In *Proceedings of the Workshop on Natural Language Processing for Improving Textual Accessibility*. 29–38. Atlanta, GA.
- Fransman, J. (2005). *Understanding literacy: A concept paper. Background paper for EFA global monitoring report*. Paris: UNESCO.
- Gee, J. P. (1996). *Social linguistics and literacies: Ideology in discourses*. Taylor & Francis.
- Glenberg, A.M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. In *Journal of Memory and Language*,

- Graff, H. (1987). *The Labyrinths of literacy. Reflections on literacy past and present*. The Falmer Press.
- Goldman, S.R., & Lee, C.D. (2014). Text complexity. State of the art and the conundrums it raises. In *The Elementary School Journal*, 115, 290–300.
- Goodman, Y. (1984). Unity in Reading. In C. Purves & O. Niles, (Eds.). *Becoming Readers in a Complex Society*. Chicago: National Society for the Study of Education
- Gunning, R. (1952). *The technique of clear writing*. McGraw-Hill.
- Hanemann, U. (2012). Shaping the literacy agenda from a lifelong learning perspective. In *Better Skills, Better Jobs, Better Lives*. OECD. <https://doi.org/10.1787/9789264177338-en>
- Harley, T. A. (2013). *Psychology of Language: From Data to Theory*. Taylor and Francis.
- Heilman, M., Collins-Thompson, K., Callan, J., Eskenazi, M. (2007). Combining lexical and grammatical features to improve readability measures for first and second language texts. In Proceedings of the NAACL Human Language Technology Conference, 460–467.
- Hiebert, E.H. (2011). Using multiple sources of information in establishing text complexity. In *Reading Research Report*. 11.
- Hiebert, E.H., & Pearson, P.D. (2014). Understanding text complexity. In *The Elementary School Journal*, 115, 153–160.
- Huettig, F. (2015). Literacy influences cognitive abilities far beyond the mastery of written language. In *Adult literacy, second language, and cognition*. Centre for Language Studies.
- Jain, M. (2005) Shikshantar: The Peoples' Institute for Rethinking Education and Development. In *Journal of Futures Studies*. 10(1). 115–118.
- James, C. T. (1975). The role of semantic information in lexical decisions. In *Journal of Experimental Psychology: Human Perception and Performance*. 1(2). 130–136.
- John, P. (2023). Constructivism: Its Implication in Language Teaching and Second Language Acquisition. In *Papers in Education and Development*. 33/34. 217-230.
- Karlsson, F. (2014). Complexity in linguistic theorizing. In *The Mental Lexicon*. 9 (2). 144–169.

- Kincaid, J. P., Fishburne, R. P. J., Rogers, R. L. and Chissom, B. S. (1975). *Derivation of new readability formulas Automated Readability Index, Fog Count and Flesch Reading Ease Formula for Navy enlisted personnel*. Naval Technical Training, U. S.
- Kirsch, I. (2001). The International Adult Literacy Survey (IALS): Understanding What Was Measured. Research Report. Princeton: Educational Testing Service.
- Kress, G. (2003). Multimodality, Multimedia and Genre. In *Literacy in the New Media Age*. (106-121). Routledge.
- Kusters, W. (2003) *Linguistic complexity: the influence of social change on verbal inflection*. LOT.
- Leal, A. (2018) Multimodalidade e multiliteracia: elementos verbais e não verbais nos textos de divulgação científica. In Gonçalves, M. & Jorge, N. (orgs.). *Literacia científica na escola*. NOVA FCSHCLUNL.
- Lee, D. (2001). Genres, registers, text types, domains and styles: Clarifying the concepts and navigating a path through the BNC jungle.
- Lorge, I. (1948). The Lorge and Flesch Readability Formulae: A Correction. In *School and Society*. 67. 141-142.
- Marcuschi, L. (2002) Gêneros textuais: definição e funcionalidade. In Dionisio *et al.* (orgs). *Gêneros textuais e ensino*. 19-36.
- Martins, M. (2010). *Processos discursivos de (re)construção do conceito de literacia: o papel dos media*. [Tese de Mestrado em Ciências da Educação, Universidade do Minho]. Repositório Universidade do Minho.
<https://hdl.handle.net/1822/13870>
- McLaughlin, G. H. (1969). SMOG Grading — a New Readability Formula. In *Journal of Reading* 128. 639-646.
- McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). CohMetrix: Capturing Linguistic Features of Cohesion. In *Discourse Processes*. 47(4), 292–330.
- Mesmer, H. A., Cunningham, J. W., & Hiebert, E. H. (2012). Toward a theoretical model of text complexity for the early grades: learning from the past, anticipating the future. In *Reading Research Quarterly*, 47, 235–258. doi:10.1002/rrq.019.
- Miranda, F. (2012). Os gêneros de texto na dinâmica das práticas de linguagem. In *Cadernos Cenpec*, 2(1). <https://doi.org/10.18676/cadernoscenpec.v2i1.96>

- Mishra, R. K., Singh, N., Pandey, A., & Huettig, F. (2012). Spoken language-mediated anticipatory eye movements are modulated by reading ability: Evidence from Indian low and high literates. In *Journal of Eye Movement Research*, 5 (1), 1-10.
- Monteiro, R., Amaro, R., Correia, S., Pintard, A., Gauchola, R., Moutinho, M., & Blanco Escoda, X. (2023). iRead4Skills - Complexity Levels (V1.0). Zenodo. <https://doi.org/10.5281/zenodo.10459090>
- National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards*.
- OECD (2013a). *Technical report of the survey of adult skills (PIAAC)*. OECD Publishing. Paris.
- OECD. (2013b). *The Survey of Adult Skills Reader's Companion*. OECD Publishing. Paris.
- OECD (2019b). *The Survey of Adult Skills: Reader's Companion (Third Edition)*. OECD Publishing. Paris. <https://doi.org/10.1787/f70238c7-en>
- OECD (2019a). *Skills Matter: Additional Results from the Survey of Adult Skills*. OECD Publishing. Paris. <https://doi.org/10.1787/1f029d8f-en>
- OECD (2021). *The Assessment Frameworks for Cycle 2 of the Programme for the International Assessment of Adult Competencies*. OECD Publishing. Paris. <https://doi.org/10.1787/4bc2342d-en>.
- Pearson, P. D., & Hiebert, E. H. (2014). The State of the Field. In *The Elementary School Journal*. 115(2). 161–183.
- Perfetti, C., Van Dyke, J., & Hart, L. (2001). The psycholinguistics of basic literacy. In *Annual Review of Applied Linguistics*, 21, 127–149.
- Petersson, K.M., Ingvar, M., & Reis, A. (2009). Language and literacy from a cognitive neuroscience perspective. In *Cambridge Handbook of Literacy*. 152–182.
- PIAAC Literacy Expert Group. (2009). *PIAAC Literacy: A Conceptual Framework*. OECD Education Working Papers. <https://doi.org/10.1787/220348414075>
- Pintard, A., François, T., Nagant de Deuxchaisnes, J., Barbosa, S., Reis, M. L., Moutinho, M., Monteiro, R., Amaro, R., Correia, S., Rodríguez Rey, S., García González, M., Mu, K., & Blanco Escoda, X. (2023). iRead4Skills Dataset 1: corpora by complexity level for FR, PT and SP (V1.0) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.10055909>

- Pinto, M. (2002). Da literacia ou de uma narrativa sempre imperfeita de outra identidade pessoal. In *Revista Portuguesa de Educação*. 15(2). 95-123.
- Portaria nº86/2022 de 4 de fevereiro. *Diário da República nº25 – 1ª série*. Ministério da Educação e Ministério do Trabalho e Segurança Social.
- Rabin, A. T. (1988). Determining Difficulty Levels of Text Written in Languages Other than English. In *Readability: It's Past, Present, & Future*.
- Roths, L & Queirós, J. (2021) Práticas de leitura e competências de literacia: resultados e ilações do PIAAC. *Plano Nacional de Leitura 2027* (Ed.). *Entreler*. 1. 32-42.
- Schleicher, A. (2012). Better skills, better jobs, better lives. In *Better Skills, Better Jobs, Better Lives*. OECD.
- Scribner, S., & Cole, M. (1981). *The Psychology of Literacy*. Harvard University Press.
- Soares, M.B. (2002). *Letramento: um tema em três géneros*. Belo Horizonte: Autêntica Editora.
- Spache, G. (1953). A new readability formula for primary-grade reading materials. In *The Elementary School Journal* 537.410–413.
- Street, B. (1993). Introduction: the new literacy studies. In Street, B. (org.). *Cross-cultural Approaches to Literacy*. Cambridge University Press.
- Street, B. (2005). Understanding and defining literacy. Background Paper for the Education for All Global Monitoring Report 2006: *Literacy for Life*.
- Street, J. A., & Dąbrowska, E. (2014). Lexically specific knowledge and individual differences in adult native speakers' processing of the English passive. In *Applied Psycholinguistics*, 35(1), 97–118. <https://doi.org/10.1017/S0142716412000367>
- UNESCO (2010). *Confintea VI. Sixth International Conference of Adult Education. Final Report*. Hamburgo. UNESCO.
- UNESCO. (2020, June 22). *Literacy*. Uis.unesco.org. Consultado em 31/10/2023. Disponível em: <https://uis.unesco.org/node/3079547>
- Valente, A. (coord.) (2021). *Referencial de Competências-chave de Educação e Formação de Adultos – Nível Básico*. ANQEP.
- Viana, F. L., & Ribeiro, I. (2020). Conhecimento lexical. LER: Leitura, escrita, recursos.
- Wagner, D. (1983). Ethnographics: Introduction. In *International Journal of the Sociology of Language*, 42, 5-8.

Wagner, D. (2011). What happened to literacy? Historical and conceptual perspectives on literacy in UNESCO. In *International Journal of Educational Development*, 42(3), 319–323.

Yousif, A. (Junho, 2003). *Literacy: An Overview of definitions and assessment*.

Comunicação apresentada em Expert Meeting on Literacy Assessment, UNESCO, Paris.

Anexo I – Descritores por nível

	Descritores	Muito fácil	Fácil	Claro	+ complexo
<i>Tamanho do texto</i>	1 O texto tem até 50 palavras.				
	2 O texto tem até 150 palavras.				
	3 O texto tem até 250 palavras.				
	4 O texto tem até 500 palavras.				

	Descritores	Muito fácil	Fácil	Claro	+ complexo
<i>propósito</i>	5 O texto serve como apoio para desempenhar tarefas familiares/quotidianas (e.g., comprar bilhetes de autocarro, identificar tipo e preço de produtos)				
	6 O texto apresenta informação nova (e.g., uma introdução sobre a importância da reciclagem como "A reciclagem é quando transformamos o lixo em coisas novas e úteis.")				
	7 O texto tem o propósito de entreter o leitor em momentos de lazer (e.g., desporto, crónica, notícias)				
	8 O texto serve para aprender/realizar tarefas novas (e.g., instruções sobre como utilizar o e-mail da empresa)				
	9 O propósito do texto pode ser sobre vários tópicos de interesse do leitor, quer seja para o informar ou para lazer (e.g., artigo de revista sobre natureza).				

		Descritores	Muito fácil	Fácil	Claro	+ complexo
contexto comunicativo	10	O texto circula em comunicações do dia a dia (e.g., horários de transportes, ementas, instruções, preços de produtos, comunicações pessoais, etc.).				
	11	O texto é utilizado em contextos profissionais ou de comunicação dos <i>media</i> (e.g., comunicação ou instrução profissional, notícias, anúncios, desporto).				
	12	O texto é utilizado em contextos de lazer (e.g., literatura e ficção, reportagens, artigos de opinião).				
	13	O texto circula em contextos artísticos (e.g., literatura, poesia).				
	14	O texto é empregue em contextos especializados (e.g., comunicação profissional formal, fóruns <i>online</i>).				
	15	O texto é utilizado em contextos científicos e académicos (e.g., ensaios, relatórios, artigos científicos).				

		Descritores (PT)	Muito fácil	Fácil	Claro	+ complexo
diversidade conceptual	16	O texto menciona conceitos relacionados com atividades do dia a dia (e.g., comida, lida da casa, roupa).				
	17	O texto menciona conceitos presentes em âmbitos específicos, mas que são de interesse do leitor (e.g., profissional, desporto, hobbies).				
	18	O texto refere conceitos que os leitores precisam de interligar para compreender totalmente (e.g., estrelas, galáxias, vulcões).				
	19	O texto faz referência a conceitos novos e multifacetados (e.g., matéria negra, desenvolvimento do cérebro, estatísticas).				

		Descritores (PT)	Muito fácil	Fácil	Claro	+ complexo
concretude	20	O texto apresenta mais palavras que referem objetos concretos (e.g., <i>cadeira, maçã, martelo, computador</i>).				
	21	O texto contém algumas poucas palavras que referem ideias abstratas (e.g., <i>felicidade, imaginação, amor</i>).				
	22	O texto emprega palavras que remetem para conceitos e ideias abstratas (e.g., <i>envelhecimento, razão, fidelidade</i>).				
	23	O texto usa palavras que referem conceitos e ideias abstratas em contextos multifacetados e/ou âmbitos especializados (e.g., <i>poesia, democracia, fusão nuclear</i>).				

		Descritores (PT)	Muito fácil	Fácil	Claro	+ complexo
vocabulário/ conjunto de léxico	24	O texto não emprega abreviações.				
	25	O texto utiliza abreviações comuns após ter explicado o que significam (e.g., <i>l – litro, m – metro</i>).				
	26	O texto utiliza abreviações comuns (e.g., <i>m, km, g., l, Exmo.</i>).				
	27	O texto não utiliza siglas.				
	28	O texto apenas utiliza siglas e acrónimos conhecidos (e.g., <i>CP, IVA, RTP, SIC, TVI</i>).				
	29	O texto utiliza siglas e acrónimos não conhecidos juntamente com o nome a que se referem (e.g., <i>Organização das Nações Unidas (ONU), Instituto Público (IP)</i>).				
	30	O texto emprega todos os tipos de siglas (e.g., <i>ADN, UNESCO, EU, UNICEF</i>).				
	31	O texto utiliza palavras e expressões simples relacionadas com assuntos do quotidiano, fazendo uso de substantivos frequentes (e.g., <i>batatas, azeite, legumes</i>).				
	32	O texto tem palavras e expressões frequentes que fazem referência a assuntos do dia a dia, mas também a outros âmbitos específicos ou profissionais (e.g., <i>segunda-feira, curso, ano escolar</i>).				
	33	O texto tem palavras e expressões um pouco menos conhecidas e frequentes que fazem referência a âmbitos específicos (e.g., <i>doença cardíaca, artéria, tratamento</i>).				

	34	O texto utiliza palavras e expressões de âmbitos específicos, jargão ou termos arcaicos (e.g., <i>enfarte do miocárdio, artéria coronária, dissecação espontânea, genoma</i>).				
	35	O texto não faz uso de estrangeirismos ou faz uso de estrangeirismos muito frequentes e conhecidos (e.g., <i>internet, donut</i>).				
	36	O texto tem poucos estrangeirismos comuns (e.g., <i>site, stress</i>).				
	37	O texto faz uso de estrangeirismos comuns frequentemente (e.g. <i>link, hobby, marketing, mail, gay</i>).				
	38	O texto contém palavras polissêmicas (palavras que possuem mais do que um sentido) que são ambíguas e/ou pouco frequentes. (e.g., <i>interno, representar, forma, paciente</i>).				
	39	O texto contém palavras polissêmicas (palavras que possuem mais do que um significado) ambíguas e/ou bastante frequentes (e.g., <i>base, experiência, leve, seca, aplicar, banco</i>).				

		Descritores (PT)	Muito fácil	Fácil	Claro	+ complexo
morfologia	40	O texto utiliza palavras compostas comuns (e.g., <i>feijão verde, couve-flor</i>) ou palavras compostas transparentes, ou seja, palavras ou expressões que podem ser facilmente compreendidas através da combinação das duas partes (e.g., <i>tira-caricas</i>).				
	41	O texto utiliza palavras compostas comuns e seu sentido pode ser percebido pela combinação das duas partes (e.g., <i>óculos de sol, mal-educado, meia-idade, maus-tratos</i>).				
	42	O texto tem palavras compostas cujo sentido não é facilmente compreendido através da combinação das duas partes (e.g., <i>meio-termo, porta-voz, obra-prima</i>).				
	43	O texto possui várias palavras compostas com sentidos que não são facilmente compreendidos pela combinação das duas partes (e.g., <i>pós-graduação, mais-valia, pé-de-meia, cavalo de batalha</i>).				
	44	O texto tem palavras derivadas com prefixos e sufixos comuns e familiares em grande número (e.g., <i>infeliz, calmamente, organizar, trabalhador, casinha</i>).				
	45	O texto tem palavras algumas derivadas mais comuns, com prefixos e sufixos frequentes (e.g. <i>pré-cozinhado, desorganizar, embalagem, organização, jardineiro, finíssimo</i>).				
	46	O texto utiliza palavras derivadas com sufixos e prefixos variados (e.g., <i>nervosismo, lavável,</i>				

		curiosidade, altitude, superpoder, preconceituoso).				
	47	O texto utiliza palavras derivadas com sufixos e prefixos menos comuns (e.g., megatonelada, tendinite, geopolítico, neoliberal).				

--	--	--	--	--	--	--

		Descritores (PT)	Muito fácil	Fácil	Claro	+ complexo
sintaxe	49	No texto, há mais frases curtas e diretas ou combinadas através de conjunções familiares (e.g., <i>A criança fugiu ontem. O pai trabalha na fábrica, mas a mãe fica em casa.</i>				
	50	O texto combina frases através de conjunções coordenadas e subordinadas (e.g., <i>Os músicos estavam tocando, <u>mas</u> poucos clientes ouviam. <u>Quando</u> pararam de tocar, apenas um senhor se manifestou.</i>				
	51	As frases no texto são frequentemente combinadas através de conjunções coordenadas e subordinadas e, por vezes, com conjunções menos comuns (e.g., <i><u>Apesar da</u> confusão provocada pelos animais à solta, a situação foi controlada <u>sem</u> os animais serem atropelados.</i>)				
	52	As frases no texto são longas e compostas por diversas outras frases combinadas através de conjunções coordenadas e subordinadas (e.g., <i><u>Dado que</u> grande parte deste livro fala na gestão das partes difíceis do trabalho a solo e analisa modos de as corrigir ou de as transcender, primeiro quero deixar bem claro que trabalhar sozinho é fantástico:</i>				

	<i>é por isso que tantas pessoas o consideram uma opção.)</i>				
53	O texto tem poucas orações relativas referentes ao sujeito da frase (e.g., <i>O senhor que vive na casa ao lado ouviu o barulho</i>).				
54	O texto tem orações relativas que referem o sujeito da frase (e.g., <i>O senhor que vive na casa ao lado ouviu o barulho</i>).				
55	As frases no texto incluem frequentemente orações relativas correspondentes ao sujeito da frase (e.g., <i>Francisca, a quem os amigos tratam por «Kika», começou a surfar muito pequena</i>).				
56	As frases no texto incluem frequentemente orações relativas correspondentes ao sujeito ou ao objeto da frase (e.g., <i>O rapaz que abraçou a mãe dirigiu-se à entrada</i> ; <i>O rapaz que a mãe abraçou dirigiu-se à entrada</i>).				
57	No texto, há mais frases na voz ativa (e.g., <i>O cão comeu o bolo</i> (ativa) vs. <i>O bolo foi comido pelo cão</i> . (passiva)).				
58	O texto tem poucas frases na voz passiva (e.g., <i>A visita foi paga pelo patrão</i>).				
59	O texto apresenta frases na voz passiva (e.g., <i>O almoço foi pago pela empresa</i>), incluindo construções impessoais comuns e passivas de -se (e.g., <i>O trabalho faz-se bem</i>).				
60	O texto tem frases simples apresentadas numa ordem básica e linear (sujeito-verbo-complementos) (e.g., <i>Eu ajudei os meus filhos a comprar casa</i>).				

	61	O texto contém poucas frases simples apresentadas não numa ordem linear (sujeito-verbo-complementos) (e.g., <i>À escola, chegaram os autocarros da visita de estudo</i>).				
	62	As frases do texto não estão necessariamente numa ordem básica e linear (sujeito-verbo-complementos) (e.g., <i>De ar puro, todos precisamos.</i>).				
	63	As frases no texto são apresentadas em diversas ordens não lineares (sujeito-verbo-complementos; verbo-complementos-sujeito, etc.) (e.g., <i>Até final de março, porque antes seria impossível encontrar todos os ingredientes, haverá na ementa vários pratos de caça</i>).				

		Descritores (PT)	Muito fácil	Fácil	Claro	+ complexo
coesão textual: pronomes, anáforas e elipses	64	O texto tem nenhum ou poucos elementos omitidos (e.g., <i>O acidente causou dois feridos leves. Os bombeiros levaram os feridos para o hospital</i>).				
	65	O texto tem elementos repetidos em vez de pronomes (e.g., <i>O João não gosta da <u>Maria</u> porque a <u>Maria</u> não é simpática</i>).				
	66	O texto tem alguns elementos omitidos (e.g., <i>Gostava de fazer artesanato. vs. Eu gostava de fazer artesanato</i>).				
	67	O texto usa pronomes (e.g., <i>O João não gosta da <u>Maria</u> porque <u>ela</u> não é simpática</i>).				

	68	O texto apresenta elementos omitidos (e.g., <i>[] Mostrámos em estudos anteriores que a gordura solidifica quando [] arrefece.</i> vs. <i>Nós mostrámos em estudos anteriores que a gordura solidifica quando <u>ela</u> arrefece</i>).				
	69	O texto apresenta pronomes e palavras relacionadas para indicar o que é referido (e.g., <i>No livro, nós damos uma definição complete de uma tempestade de neve e como <u>esta</u> se forma. Também indicamos onde o <u>fenómeno</u> pode ser visto</i>).				

		Descritores (PT)	Muito fácil	Fácil	Claro	+ complexo
estilo	74	O texto não faz uso de sentidos figurativos (e.g., <i>Este verão os preços <u>subiram</u></i> vs. <i>Este verão os preços <u>dispararam</u></i>).				
	75	O texto apresenta sentidos figurativos comuns (e.g., <i>Este verão os preços <u>dispararam</u></i>).				
	76	O texto apresenta sentidos figurativos menos comuns (e.g., <i>O Barcelona <u>varreu</u> o Real Madrid; O Estoril <u>esmagou</u> o Santa Clara.</i>)				
	77	O texto apresenta sentidos figurativos mais incomuns (e.g., <i>Os primeiros raios de sol <u>aconchegaram-me</u> as faces</i>).				
	78	O texto utiliza apenas comparações, personificações e hipérboles mais comuns (e.g., <i>Ela é <u>linda como</u> uma flor; As minhas costas <u>estão a dar cabo de mim</u>; A tua mala <u>pesa uma tonelada</u></i>).				

79	O texto faz uso da metonímia (e.g., <i>beber um copo</i>), assim como metáforas familiares (e.g., <i>Meu pai tem olhos de águia</i>).				
	80 O texto faz uso de metáforas incomuns (e.g., <i>A sociedade está doente; ele é um lobo em pele de cordeiro</i>).				
	81 O texto apresenta diversas figuras de estilo (e.g., <i>O cão da fábrica de curtumes acendia latidos fosforescentes nas noites de julho, quando o pólen da acácia me cobria as pálpebras, eu, morto de amores pela mulher de Sandokan, descobria-me unicórnio trancado na retrete da escola...</i>).				

Anexo II – Descrição dos níveis de complexidade

NÍVEL 1 – MUITO FÁCIL

Textos curtos e simples até cerca de 50 palavras, com o objetivo de desempenhar tarefas do quotidiano ou de apresentar nova informação (e.g. textos didáticos). **Conceitos simples do dia a dia**, com o tópico da comunicação a ser apresentado inicialmente.

Contextos básicos de comunicação: quotidianos (e.g. horários de transportes/ listas/ menus/ informação de preço de um produto), comunicações pessoais (e.g. cartas, mensagens).

Palavras simples e frequentes, comumente usadas em contextos do dia a dia (transporte, comida, família, trabalho). Uso predominante de palavras concretas que fazem referência a objetos concretos no mundo real (*maçã, cadeira, fogão*). Ausência de abreviações, siglas e estrangeirismos, à exceção daqueles muito frequentes (internet, hamburger).

Palavras compostas comuns (*couve-flor, guarda-chuva*).

Palavras com afixos frequentes (*infeliz, calmamente, trabalhador*).

Frases curtas e simples, utilizando conjunções simples e na **ordem direta (SujeitoVerbo-Complemento)**.

Estruturas de coordenação frequentes (copulativas, disjuntivas e adversativas) estão presentes (e.g. «O Rui e/ou a sua irmã foram à festa, mas o João não foi.»).

Estrutura de subordinação frequentes estão presentes (e.g. oração subordinada adverbial temporal – «*Desde que voltou, O Rui não dormiu.*»).

Uso predominante da voz ativa.

Textos utilizam tempos verbais no modo Infinitivo, Indicativo e Imperativo, nomeadamente o Presente, e Pretéritos Perfeito e Imperfeito. Ocorrência de construções perifrásticas comuns como estar/começar/andar + a + infinitivo ou deixar/acabar + de + infinitivo.

Não figuram tempos compostos, à exceção do Futuro que é transmitido através de ir + infinitivo.

Cadeia referencial completa sem omissão de elementos, com repetição em vez da utilização de pronomes (e.g. «O Rui gosta da Maria, porque a Maria é simpática.»).

Localização temporal linear, dada por advérbios e conjunções simples (e.g. «ontem, o Rui foi à praia. Os bombeiros chegaram depois da polícia.»).

Ausência de figuras de estilo.

NÍVEL 2 – FÁCIL

Textos curtos até 150 palavras, que são interessantes para o leitor **se informar/aprender algo novo ou para momentos de lazer.**

Conceitos mais específicos, mas do interesse do leitor (e.g. profissional, desporto, hobbies). Contextos de comunicação profissionais e dos media (e.g. instrução profissional, notícias, anúncios, desporto).

Palavras e expressões frequentes que fazem referência a assuntos do dia a dia, mas também a outros âmbitos mais específicos (e.g. curso, ano escolar). O texto contém algumas palavras que referem ideias abstratas mais frequentes (*felicidade, imaginação, carinho*).

O texto utiliza abreviações comuns, depois de as ter explicado (e.g. *litro – l/ metro – m*).

O texto apenas utiliza siglas e acrónimos conhecidos (e.g. CP, IVA, RTP, SIC, TVI).

Uso de estrangeirismos menos frequentes (e.g. *site, souvenir*).

O texto utiliza palavras compostas comuns com o seu sentido a poder ser compreendido pela combinação das duas partes (*óculos de sol, mal-educado, meia-idade, mastratos*).

Palavras com afixos menos frequentes (*pré-cozinhado, embalagem, finíssimo*).

Frases simples e complexas através de conjunções coordenadas e subordinadas (e.g. «*os músicos estavam tocando, mas poucos clientes ouviam. Quando pararam de tocar, apenas um senhor se manifestou.*»). Uso de orações relativas de sujeito (e.g. «*O senhor que vive na casa ao lado ouviu o barulho.*»). O texto pode conter frases apresentadas numa ordem não linear (Complemento-Verbo-Sujeito) (e.g. «*À escola, chegaram os autocarros da visita de estudo.*»).

Possível uso da voz passiva.

Uso de modos e tempos verbais variados.

Uso de tempos verbais compostos e de expressões verbais com *ter/poder/dever* (e.g. «*posso/devo/tenho de gostar de bolo.*»). Modos Conjuntivo, Condicional e Gerúndio.

Expressões com os verbos *permanecer, querer, poder, conseguir* em tempos simples (e.g. «*permaneça calmo, queria almoçar, consigo ir.*»).

Utilização do Futuro simples (e.g. «*ele gostará de te ver.*»).

O texto tem alguns elementos omitidos e utiliza pronomes (e.g. «Eu gostava da Maria vs. Gostava da Maria/ Mas ela não gosta dele.»).

Sequência ou localizações temporais lineares, dadas através de advérbios de tempo ou conjunções **mas também por tempos verbais** (e.g. «*ele está a caminho e a festa já começou.*»).

Figuras de estilo frequentes são utilizadas.

O texto faz uso da **metonímia** (e.g. «*beber um copo*»), assim como **metáforas familiares** (e.g. «O meu pai tem olhos de águia.»).

NÍVEL 3 – CLARO

Textos com cerca de 250 palavras sobre vários tópicos, sejam estes de informação ou lazer, com os leitores a conseguirem sair fora da sua zona de conforto.

Conceitos variados que necessitam de ser interligados para a sua compreensão total (e.g. ciência, política).

Contextos comunicativos variados: lazer (literatura, reportagens); **artes** (poesia); **contextos especializados** (*fóruns* online, comunicação profissional formal).

Palavras e expressões menos frequentes que fazem referência a âmbitos específicos (e.g. *doença cardíaca, artéria, tratamento*). As palavras remetem para **conceitos mais abstratos** (e.g. envelhecimento, razão, fidelidade).

Utilização de siglas e acrónimos não conhecidos pelo leitor juntamente com o nome a que se referem (e.g. *Organização das Nações Unidas (ONU), Instituto Público (IP)*).

O texto utiliza abreviações comuns, sem explicação prévia (e.g. *m, km, g., l, Exmo.*). Uso de estrangeirismos comuns frequentemente (e.g. *link, hobby, marketing, mail, gay*).

O texto contém palavras que são ambíguas e/ou pouco frequentes. (e.g. *interno, representar, forma, paciente*).

Palavras com sufixos e prefixos variados (e.g. *nervosismo, lavável, curiosidade, altitude, superpoder, preconceituoso*).

Frases frequentemente combinadas através de conjunções coordenativas e subordinativas e, por vezes, com conjunções menos comuns (e.g. «*apesar da confusão provocada pelos animais à solta, a situação foi controlada sem os animais serem*

atropelados.»). **Presença frequente de ordem não linear** de constituintes (sujeitoverbo-complementos) (e.g. «*de ar puro, todos precisamos*».).

Utilização habitual da voz passiva (e.g. «*o almoço foi pago pela empresa*».), incluindo construções impessoais comuns e passivas de -se (e.g. «*O trabalho faz-se bem*».).

Utilização de todos os modos e tempos verbais compostos e simples, à exceção do Pretérito Mais-que-perfeito.

Elementos omitidos (e.g. «*[] mostrámos em estudos anteriores que a gordura solidifica quando [] arrefece*. vs. *Nós mostrámos em estudos anteriores que a gordura solidifica quando ela arrefece*.»). Utilização de **pronomes e anáforas lexicais** para indicar o que é referido (e.g. «*no livro, nós damos uma definição complete de uma tempestade de neve e como esta se forma. Também indicamos onde o fenómeno pode ser visto*».)

Sequências ou localizações temporais não lineares (e.g. «*após analisarem as respostas dos doentes que tinham sofrido um ataque cardíaco, os médicos confirmaram que a gordura é o alimento mais perigoso*.»)

Uso de **figuras de estilo** menos frequentes.

Metáforas incomuns (e.g. «*A sociedade está doente; ele é um lobo em pele de cordeiro*».).

Anexo III – Tipologia de textos

<i>propósito & meio</i>		<i>documento</i>	<i>área/conteúdo</i>
comunicação pessoal		bilhete	
		Email/carta pessoal	
		SMS/chats online (WhatsApp, ...)	
		lista/agenda (de compras, de tarefas, ...)	
		diário	
comunicação institucional/profissional		carta/mail	
		relatório	
		instruções	
		ata	
		comunicado de imprensa	
		comunicado/newsletter	
		página web (acerca/sobre/missão...)	
comunicação social (jornais, revistas)		editorial	Política Sociedade Desporto Cultura e artes Viver Ciência e ambiente Economia e finanças
		notícia	
		reportagem	
		entrevista	
		artigo de opinião	
		artigo de divulgação científica	
		perfil (breve apresentação de uma personalidade)	
		classificado	
		horóscopo	
		obituário	
		boletim meteorológico	
comunicação comercial/disseminação		anúncios	Saúde Banca Seguros Restauração Automóvel ...
		cartazes/posters/outdoors	
		folhetos	
		ementa	
		rótulo	
		instruções/manuais de instruções	
		bula	
livro	não-ficção	(auto)biografia	
		crónica	
		ensaio	
		diário	
		prefácio	
		dedicatória	
		autoajuda	
		relato de viagem	
		memória/testemunho	
		cartas	

		guia turístico	
	ficção	conto	<i>Policial Thriller Terror Ficção científica Romance Romance histórico Fantástico Erótico Infanto-juvenil ...</i>
		fábula	
		épico	
		romance/novela	
		lírico (poesia), letras de música, ...	
		dramático (peça de teatro)	
	didático	manual	<i>de várias áreas do saber (Arquitetura, Biologia, Ciências Naturais, Geologia.</i>
		enciclopédia/atlas	
		livro de receitas	
		glossário	
	académico	artigo	
		relatório	
		resumo/síntese	
		recensão crítica	
		ensaio	
		dissertação/tese	
		proposta de projeto	
político		discurso	
		moção/relatório	
		programa	
legal		lei	
		contrato	
		ofício/memorando/estatutos/edital	
religioso		oração	Cristianismo Judaísmo Islamismo Budismo ...
		escrituras/salmos/epístolas	
		homilias	
		sermão/folheto/catequismo	

menos relevantes para promoção de leitura

Anexo IV – Tarefa preparatória para o *focus group*

Questões sobre os descritores

Nota 1: As questões devem ser apresentadas de forma aleatória.

Nota 2: O texto de cor verde representa a informação/questões apresentadas ao focus group.

Nota 3: Cada descritor é acompanhado por um texto que serve de exemplo ao mesmo (a sombreado).

Nota 4: A informação sobre os níveis de complexidade que segue os descritores (muito fácil, fácil, claro, +complexo) é para informação interna, não estando acessível aos respondentes.

Níveis de complexidade do texto

Muito fácil: Textos que são totalmente ou quase totalmente compreendidos por todas as pessoas, incluindo pessoas com escolaridade muito baixa (i.e., que não chegaram ao 6º ano do ensino básico) e que não têm quase nenhuma prática de leitura.

Fácil: Textos que são totalmente ou quase totalmente compreendidos por pessoas com escolaridade baixa (i.e., que completaram o 6º ano do ensino básico mas que não têm mais que o 9º ano) e que têm pouca prática de leitura.

Claro: Textos que são compreendidos à primeira leitura por pessoas que completaram o 9º ano e que têm uma prática de leitura média.

1. Qual o nível de complexidade que associa a esta característica?

O texto tem até 50 palavras. *(repetir para cada um dos 80 descritores)*

☐ muito fácil

☐ fácil

☐ claro

☐ não é claro/ mais complexo

Comentários ou sugestões: _____

	<i>descriptor</i>
<i>tamanho do texto</i>	1. O texto tem até 50 palavras. (muito fácil) 2. O texto tem até 150 palavras. (fácil) 3. O texto tem até 250 palavras. (claro) 4. O texto tem até 500 palavras. (+complexo)
<i>propósito</i>	5. O texto tem o propósito de desempenhar tarefas familiares/quotidianas (e.g., instruções para comprar bilhetes de autocarro, instruções de utilização no rótulo de um produto) (muito fácil) 6. O texto apresenta um tópico novo (e.g., um texto introdutório sobre a importância da reciclagem) (muito fácil) 7. O texto tem o propósito do leitor se entreter em momentos de lazer (e.g., The text is interesting for the reader to inform him/herself in moments of leisure (e.g., desporto, crónica, notícias) (fácil) 8. O texto serve para aprender/realizar tarefas novas (e.g., instruções sobre como utilizar o e-mail da empresa) (fácil) 9. O propósito do texto pode ser sobre vários tópicos de interesse do leitor, quer seja para o informar ou para lazer (e.g., artigo de revista sobre natureza). (claro)
<i>contexto comunicativo</i>	10. O texto circula em comunicações do dia a dia (e.g., horários de transportes, ementas, instruções, preços de produtos, comunicações pessoais, etc.). (muito fácil) 11. O texto é utilizado em contextos profissionais ou de comunicação dos <i>media</i> (e.g., comunicação ou instrução profissional, notícias, anúncios, desporto). (fácil) 12. O texto é utilizado em contextos de lazer (e.g., literatura e ficção, reportagens, artigos de opinião). (claro) 13. O texto circula em contextos artísticos (e.g., literatura, poesia). (claro) 14. O texto é empregue em contextos especializados (e.g., comunicação profissional formal, fóruns <i>online</i>). (claro) 15. O texto é utilizado em contextos científicos e académicos (e.g., ensaios, relatórios, artigos científicos). (+ complexo)
<i>diversidade conceptual</i>	16. O texto menciona conceitos relacionados com atividades do dia a dia (e.g., comida, lida da casa, roupa). (muito fácil) 17. O texto menciona conceitos presentes em âmbitos específicos (e.g., profissional, desporto, hobbies). (fácil) 18. O texto refere conceitos que os leitores precisam de interligar para compreender totalmente (e.g., estrelas, galáxias, vulcões). (claro) 19. O texto faz referência a conceitos novos e multifacetados (e.g., matéria negra, desenvolvimento do cérebro, estatísticas) (+complexo)
<i>concretude</i>	20. O texto é maioritariamente composto por palavras que referem objetos concretos (e.g., cadeira, maçã, martelo, computador). (muito fácil) 21. O texto tem uma ligeira percentagem de palavras que referem ideias abstratas (e.g., felicidade, imaginação, amor). (fácil) 22. O texto emprega palavras que remetem para conceitos e ideias abstratas (e.g., envelhecimento, razão, fidelidade). (claro) 23. O texto usa palavras que referem conceitos e ideias abstratas em contextos multifacetados e/ou âmbitos especializados (e.g., poesia, democracia, fusão nuclear). (+complexo)

vocabulário/ conjunto de léxico	24. O texto não emprega abreviações (e.g., km, c., V. Exa), exceto formas frequentes de tratamento (e.g., Sr. Dra.). (muito fácil) 25. O texto utiliza abreviações comuns após ter explicado o que significam (e.g., l – litro, m – metro). (fácil) 26. O texto utiliza abreviações comuns (e.g., m, km, g., l, Exmo.). (claro) 27. O texto não utiliza siglas. (muito fácil) 28. O texto utiliza siglas conhecidas, isto é, que podem ser compreendidas tendo sido utilizadas sem explicação previa (e.g., empresas de transportes – CP, impostos – IRS, IVA, partidos políticos – PCP, PS) (fácil) 29. O texto utiliza siglas depois de as ter explicado (e.g., Organização das Nações Unidas (ONU), Instituto Público (IP)). (claro) 30. O texto emprega todos os tipos de siglas (e.g., ADN, UNESCO, EU, UNICEF). (+complexo) 31. O texto utiliza palavras e expressões simples relacionadas com assuntos do quotidiano, fazendo uso de substantivos frequentes (e.g. batatas, azeite, legumes). (muito fácil) 32. O texto tem palavras e expressões frequentes que fazem referência a assuntos do dia a dia, mas também a outros âmbitos específicos ou profissionais (e.g., segunda-feira, curso, ano escolar). (fácil) 33. O texto tem palavras e expressões menos conhecidas que fazem referência a âmbitos específicos (e.g., doença cardíaca, artéria, tratamento). (claro) 34. O texto utiliza palavras e expressões de âmbitos específicos, jargão ou termos arcaicos (e.g., enfarte do miocárdio, artéria coronária, dissecação espontânea, genoma). (+complexo) 35. O texto não faz uso de estrangeirismos, exceto aqueles que são muito frequentes e conhecidos (e.g., internet, donut). (muito fácil) 36. O texto tem estrangeirismos comuns (e.g., site, stress), mas são uma percentagem reduzida. (fácil) 37. O texto faz uso de estrangeirismos comuns frequentemente (e.g. link, hobby, marketing, mail, gay) (claro) 38. O texto contém palavras polissémicas, ou seja, palavras que possuem mais do que um sentido e podem ser ambíguas, pouco frequentes. (e.g., interno, representar, forma, paciente). (claro) 39. O texto contém palavras polissémicas, ou seja, palavras que possuem mais do que um significado e podem ser ambíguas, bastante frequentes (e.g., base, experiência, leve, seca, aplicar, banco). (+complexo)
--	--

<i>morfologia</i>	40. O texto utiliza palavras compostas comuns (e.g., feijão verde, couve-flor) ou palavras compostas transparentes, ou seja, palavras ou expressões que podem ser facilmente compreendidas através da combinação das duas partes (e.g., tira-caricas). (muito fácil) 41. O texto utiliza palavras compostas comuns, cujo sentido pode ser mais ou menos compreendido através da combinação das duas partes (e.g., óculos de sol, mal-educado, meia-idade, maus-tratos). (fácil) 42. O texto tem palavras compostas cujo sentido não é facilmente compreendido através da combinação das duas partes (e.g., meio-termo, porta-voz, obra prima). (claro) 43. O texto possui várias palavras compostas cujo sentido não é facilmente compreendido através da combinação das duas partes (e.g., pós-graduação, mais-valia, pé de meia, cavalo de batalha). (+complexo) 44. O texto tem palavras derivadas com prefixos e sufixos comuns e familiares (e.g., infeliz, calmamente, organizar, trabalhador, casinha), mas estas estão em maioria no texto. (muito fácil) 45. O texto tem palavras derivadas com prefixos e sufixos frequentes (e.g. precozinhado, desorganizar, embalagem, organização, jardineiro, finíssimo), mas estas não estão em maioria no texto. (fácil) 46. O texto utiliza palavras derivadas com sufixos e prefixos variados (e.g., nervosismo, lavável, curiosidade, altitude, superpoder, preconceituoso). (claro) 47. O texto utiliza palavras derivadas com sufixos e prefixos menos comuns (e.g., megatonelada, tendinite, geopolítico, neoliberal) (+complexo)
-------------------	--

morfosintaxe	48. Por favor, considere as seguintes estruturas e indique o(s) nível(eis) de dificuldade em que acha que estas ocorrem mais frequentemente/são mais típicas.				
		<i>muito fácil</i>	<i>fácil</i>	<i>claro</i>	<i>mais complexo</i>
	Verbos no infinitivo (e.g., <i>gostar, comer, fugir</i>)				
	Verbos no Presente do Indicativo (e.g., <i>gosto, comem, foge</i>)				
	Verbos no Pretérito Imperfeito do Indicativo (e.g., <i>gostavam, comia, fugias</i>)				
	Verbos no Pretérito Perfeito do Indicativo (e.g., <i>gostou, comi, fugiram</i>)				
	Verbos no Pretérito Mais-que-perfeito do Indicativo (e.g., <i>gostara, comeras, fugira</i>)				
	Verbos no Future Simples do Indicativo (e.g., <i>esquecerá, gostará</i>)				
	Verbos no Gerúndio (e.g., <i>gostando, comendo, fugindo</i>)				
	Verbos no Imperativo (e.g., <i>come!, falai!; não comas!; não falem!</i>)				
	Verbos no Condicional (e.g., <i>gostaria, comerias, fugiriam</i>)				
	Verbos no Presente do Conjuntivo (e.g., <i>gosta, coma, fujam</i>)				
	Verbos no Pretérito Imperfeito do Conjuntivo (e.g., <i>gostasses, comessem, fugisse</i>)				
	Verbos no Futuro do Conjuntivo (e.g., <i>se eu gostar, se tu comeres, se eles fugirem</i>)				
	Verbos no Pretérito Perfeito Composto do Indicativo (e.g., <i>tenho gostado, tem comido, têm fugido</i>)				
	Verbos no Pretérito Mais-que-perfeito Composto do Indicativo (e.g., <i>tinhas gostado, tinham comido, tinha fugido</i>)				
	Verbos no Futuro Composto do Indicativo (e.g., <i>terei gostado, terás comido, terão fugido</i>)				
	Verbos no Pretérito Perfeito Composto do Conjuntivo (e.g., <i>tenha gostado, tenha comido, tenham fugido</i>)				
	Verbos no Pretérito Mais-que-perfeito Composto do Conjuntivo (e.g., <i>tivesses gostado, tivessem comido, tivesse fugido</i>)				
	Verbos no Futuro Composto do Conjuntivo (e.g., <i>tiverem gostado, tiver comido, tiveres fugido</i>)				
	Verbos auxiliares <i>poder, ir, estar</i> no Indicativo (e.g., <i>posso gostar, vai comer, estão a fugir</i>)				
	Expressões com os verbos <i>permanecer, querer, poder, conseguir</i> em tempos simples (e.g., <i>permaneça calmo, queria almoçar, consigo ir</i>)				
	Expressões com os verbos <i>estar, ser, permanecer, querer, poder, dever</i> em tempos compostos (e.g., <i>tenham estado ligados; foram feitas; temos de permanecer calmos; devíamos ir</i>)				

sintaxe	49. A maioria das frases no texto são curtas e diretas ou combinadas através de conjunções familiares (e.g., A criança fugiu ontem. O pai trabalha na fábrica <u>mas</u> a mãe fica em casa). (muito fácil) 50. O texto combina frases através de conjunções coordinativas e subordinativas (e.g., Os músicos foram surpreendidos pela polícia <u>e</u> estão concentrados em frente ao centro comercial. <u>Se</u> ainda não pagou a conta, pode pagar aqui.) (fácil) 51. As frases no texto são frequentemente combinadas através de conjunções coordinativas e subordinativas e, por vezes, com conjunções menos comuns (e.g., <u>Apesar da</u> confusão provocada pelos animais à solta, a situação foi controlada <u>sem</u> os animais serem atropelados.) (claro) 52. As frases no texto são longas e compostas por diversas outras frases combinadas através de conjunções coordinativas e subordinativas (e.g., Dado que grande parte deste livro fala na gestão das partes difíceis do trabalho a solo e analisa modos de as corrigir ou de as transcender, primeiro quero deixar bem claro que trabalhar sozinho é fantástico: é por isso que tantas pessoas o consideram uma opção.) (+complexo) 53. O texto tem orações relativas correspondentes ao sujeito da frase (e.g., O senhor <u>que vive na casa ao lado</u> ouviu o barulho), contudo, estas não estão em maioria no texto (muito fácil) 54. O texto tem orações relativas correspondentes ao sujeito da frase (e.g., O senhor <u>que vive na casa ao lado</u> ouviu o barulho.). (fácil) 55. As frases no texto incluem frequentemente orações relativas correspondentes ao sujeito da frase (e.g., Francisca, <u>a quem os amigos tratam por «Kika»</u>, começou a surfar muito pequena.). (claro) 56. As frases no texto incluem frequentemente orações relativas correspondentes ao sujeito ou ao objeto da frase (e.g., O rapaz <u>que abraçou a mãe</u> dirigiu-se à entrada; O rapaz <u>que a mãe abraçou</u> dirigiu-se à entrada.). (+complexo) 57. A maioria das frases no texto estão na voz ativa (e.g., O cão comeu o bolo (ativa) vs. O bolo foi comido pelo cão. (passiva)) (muito fácil) 58. O texto tem frases na voz passiva (e.g., A visita foi paga pelo patrão), no entanto, não estão em maioria. (fácil) 59. O texto apresenta frases na voz passiva (e.g., O almoço foi pago pelo empresa.), incluindo construções impessoais comuns e passivas de -se (e.g., O trabalho faz-se bem.) (claro) 60. O texto tem frases simples apresentadas numa ordem básica e linear (sujeito-verbo-complementos) (e.g., Eu ajudei os meus filhos a comprar casa.). (muito fácil) 61. O texto contém frases simples apresentadas não numa ordem linear (sujeito-verbo-complementos) (e.g., À escola, chegaram os autocarros da visita de estudo), no entanto, estas não estão em maioria no texto. (fácil) 62. As frases do texto não estão necessariamente numa ordem básica e linear (sujeito-verbo-complementos) (e.g., De ar puro, todos precisamos.). (claro) 63. As frases no texto são apresentadas em diversas ordens não lineares (sujeito-verbo-complementos; verbo-complementos-sujeito, etc.) (e.g., Até final de março, porque antes seria impossível encontrar todos os ingredientes, haverá na ementa vários pratos de caça.) (+complexo)

coesão textual: pronomes, anáforas e elipses	64. O texto tem nenhum ou poucos elementos omitidos (e.g., O acidente causou dois feridos leves. Os bombeiros levaram os feridos para o hospital.). (muito fácil) 65. O texto tem elementos repetidos em vez de pronomes (e.g., O João não gosta da <u>Maria</u> porque a <u>Maria</u> não é simpática) (muito fácil) 66. O texto tem alguns elementos omitidos (e.g., [] Reparo desde pequena nos pássaros. vs. <u>Eu</u> reparo desde pequena nos pássaros.). (fácil) 67. O texto usa pronomes (e.g., O João não gosta da <u>Maria</u> porque <u>ela</u> não é simpática). (fácil) 68. O texto apresenta elementos omitidos (e.g., [] Mostrámos em estudos anteriores que a gordura solidifica quando [] arrefece. vs. <u>Nós</u> mostrámos em estudos anteriores que a gordura solidifica quando <u>ela</u> arrefece). (claro) 69. O texto apresenta pronomes e palavras relacionadas para indicar o que é referido (e.g., No livro, nós damos uma definição complete de uma tempestade de neve e como <u>esta</u> se forma. Também indicamos onde o <u>fenómeno</u> pode ser visto.) (claro)
coesão textual: referências temporais	70. O texto apresenta uma sequência ou localizações temporais simples e lineares, dadas através de advérbios de tempo e conjunções como hoje, amanhã, antes de, depois de, quando, enquanto (e.g., Os bombeiros chegaram depois da polícia.)(muito fácil) 71. O texto tem uma sequência ou localizações temporais lineares, dadas através de advérbios de tempo ou conjunções como após, assim que, antes de e também por tempos verbais (e.g., Ele <u>está</u> a caminho e a festa <u>já</u> começou. A banda começou a tocar <u>antes de</u> todos saírem.) (fácil) 72. O texto inclui uma sequência ou localizações temporais não lineares (e.g., <u>Após</u> analisarem as respostas dos doentes que <u>tinham</u> sofrido um ataque cardíaco, os médicos <u>confirmaram</u> que a gordura é o alimento mais perigoso.). (claro) 73. O texto tem diversas sequências ou localizações temporais não lineares (e.g., <u>Será preciso fazer</u> alguma coisa aos cadáveres e, por isso, <u>é</u> bom que <u>haja</u> gente que <u>se debruce</u> sobre o tema. Eu <u>seria</u> incapaz, porque nem sequer <u>consigo</u> ir a funerais, como já <u>fui</u>, onde <u>repousam</u> caixões abertos com os indivíduos em descanso eterno.) (+complexo)
estilo	74. O texto não faz uso de sentidos figurativos (e.g., Este verão os preços subiram vs. Este verão os preços dispararam (fig.)) (muito fácil) 75. O texto apresenta sentidos figurativos comuns (e.g., Este verão os preços dispararam.) (fácil) 76. O texto apresenta sentidos figurativos menos comuns (e.g., O Barcelona <u>varreu</u> o Real Madrid; O Estoril esmagou o Santa Clara.) (claro) 77. O texto apresenta sentidos figurativos menos incomuns (e.g., Os primeiros raios de sol <u>aconchegaram-me</u> as faces.) (+complexo) 78. O texto utiliza apenas comparações, personificações e hipérboles (e.g., Gostar de alguém <u>é como</u> fazer prédios; As doenças escondem-se em todo o lado; Eu estava gelada; A tua mala pesa uma tonelada.)(muito fácil) 79. O texto faz uso da metonímia (e.g. beber um copo;), assim como metáforas familiares (e.g., O meu caminho é outro. Não quero estudar mais. (fácil) 80. O texto faz uso de metáforas incomuns (e.g., Marcelo e Costa, um casamento de convergência. Os dois coabitam desde 9 de março de 2016, quando Marcelo assumiu funções.) (claro) 81. O texto apresenta diversas figuras de estilo (e.g., O cão da fábrica de curtumes acendia latidos fosforescentes nas noites de julho, quando o pólen da acácia me cobria as pálpebras, eu, morto de amores pela mulher de Sandokan, descobria-me unicórnio trancado na retrete da escola...) (+complexo)