



Jan Kochanowski University Press

This is a contribution from *Token: A Journal of English Linguistics*
Volume 15/2022.

Edited by John G. Newman, Marina Dossena.

© 2022 Jan Kochanowski University Press.

Rewriting novels for a young audience: A corpus-assisted comparison between two versions of *The Da Vinci Code* by Dan Brown

Francesca Bianchi and Elena Manca¹

University of Salento

ABSTRACT

This paper explores the differences between the young adult version and the original version of *The Da Vinci Code* by Dan Brown through the lens of corpus linguistics. The study adds to the limited number of linguistic case studies analysing intralingual literary adaptations for children and highlights ‘traditional’ adaptation strategies such as Purification, Language adaptation, Abridgment and Localization. However, the analyses also highlighted other types of adjustments not covered in the above categories and previously observed in another study on intralingual adaptations for young audiences (Bianchi 2018). This suggests the need for creating a classification of adaptation strategies specific to intralingual adaptations targeting young readers.

Keywords: corpus linguistics, POS tagging, semantic tagging, Wmatrix, adaptation, novel, teenagers.

1. Introduction

The Da Vinci Code by Dan Brown was first published in 2003. It soon became a worldwide bestseller – with 80 million copies sold as of 2009.² In 2016 a new version of the book was published by Delacorte Press, an imprint of Random House Children’s Books, a division of Penguin Random House LLC, New

¹ The first author is responsible for sections 1, 2, 3, 4.1.2, 4.2 and 5; the second author is responsible for the analyses in section 4.1.1.

² https://en.wikipedia.org/wiki/The_Da_Vinci_Code, accessed August 2021.

York. Interestingly, this new version is an adaptation for a ‘young adult’ audience written by Dan Brown himself, as the Copyright notice asserts. Press releases announcing the imminent launch of this version of the book defined the target audience as youngsters in their ‘early teens’ (Deahl 2016) or 13+ (BBC 2016).

Despite the enthusiasm of the author and publisher for this new literary endeavour, the general public and the press (e.g., Colyard 2016; Erizanu 2016; Tufayel 2016) did not seem to understand or appreciate the need for it and expressed the view that a dedicated edition was offensive to teenagers and their understanding and reading abilities. Admittedly, the author’s justification of the young adult version is extremely vague, and so is his description of the differences between the two versions. These are offered in a video³ available on YouTube and can be summarized along the following lines: the author had long wanted to do a young adult version of *The Da Vinci Code*; the 2003 book contains r-rated material that he has tried to dilute; the 2003 book was very long, while the new version is shorter, which makes it more appropriate for a young audience; adults and children can read the new version together, and it will make them think and ask questions; *The Da Vinci Code* shows that art, architecture and documents are based in fact and can be studied as history; at the same time, young readers will be captivated by the adventure story. The young version aims to make young readers think and learn. On the other hand, the publishers make it a point to advertize that the young version “[i]ncludes over twenty color photos showing important locations, landmarks, and artwork, taking readers from Paris to London and beyond!”⁴

Rather than settling the matter, Dan Brown’s and his publisher’s explanations suggest new questions:

1. What is Dan Brown’s (or his publisher’s) idea of “more appropriate for readers in their early teens”?
2. Are the changes made by Dan Brown in keeping with traditional adaptation habits?
3. Are the changes made by Dan Brown in keeping with the actual needs of teenagers?

The current study addresses these questions by comparing the young adult edition to the original adult version. In doing so, it contributes to the

³ <https://www.youtube.com/watch?v=vg6zs0L6k3Y>, accessed August 2021.

⁴ <https://www.penguinrandomhouse.com/books/549656/the-da-vinci-code-the-young-adult-adaptation-by-dan-brown/>, accessed August 2021.

very few linguistic case studies analysing intralingual literary adaptations for the young. A theoretical framework for the analysis of the differences between the two texts is provided in Section 2 and draws from theoretical and empirical studies in children's literature, developmental psychology, language acquisition and reading comprehension, as well as linguistic studies on popularization. Section 3 introduces the materials and methods adopted in the current study. Section 4 illustrates the findings. Finally, Section 5 attempts to draw some conclusions.

2. Theoretical framework and literature review

Linguistic studies of the strategies used or needed to adapt literary texts to young audiences are still scarce. To the best of our knowledge, the only detailed linguistic case study comparing literary adaptations for children to the original adult text is Bianchi (2018), summarized in Section 2.1. However, as the same paper points out, theoretical and empirical studies in children's literature, developmental psychology, language acquisition and reading comprehension can be used to build a theoretical framework for interpreting analytical data. This framework is based on three axes and is schematically outlined below.

A. Habits in adaptations for children

Adaptations are generally based on what society believes to be pedagogically and morally appropriate for children (e.g., Klingberg 1986; Shavit 1986). Across time, they have always had a 'didactic intention' (Beckett 2009: 19). To the best of our knowledge, the only classification of typical adjustments is that by Klingberg (1986, in Anderson 2000). His classification, though deriving from an analysis of translations of children's books into foreign languages, can be considered to suitably describe "manipulations [...] performed in the adaptation of adult texts for children" (Anderson 2000: 276)⁵ and includes cultural context adaptation, modernization, purification (i.e., removal of inappropriate content), language adaptation, abridgement, and localization.

⁵ Anderson (2000: 276) explains that "Literature designated as 'children's' is either adapted from works originally intended for adults or written directly for an audience of children. Because it is not generally taken seriously as literature but is filtered through adults for its pedagogical possibilities, the translation norms that apply to it are quite different from those applied to adult literature; in fact, they often resemble the norms used in adapting adult works for children."

B. Needs of pre-adolescent and adolescent readers

Across a person's life, five reader roles can be identified, each one depending on the affective, social, cognitive and experiential needs of the different age groups (Appleyard 1991): the early childhood reader is a Player; the later childhood reader is a Hero or Heroine; the adolescent reader is a Thinker; at college age and beyond, the reader is an Interpreter; in adulthood s/he becomes a Pragmatic Reader. In their late childhood (including pre-adolescence), readers require adventures with characters identifiable as either villains or heroes. They should feature fairly simple sentences and short paragraphs, limited descriptions of characters and settings and ample focus on dialogue and action. Naturally, at both the structural, thematic and linguistic levels, different degrees of complexity are envisaged as the child grows and moves towards adolescence. To meet the needs of an adolescent reader, a literary work should have a narrative structure that is complex enough to hold his/her attention and characters with whom the adolescent can identify. In particular, the characters should "match their readers' newfound sense of complexity" without exceeding it (Appleyard 1991: 106). In terms of content, adolescents appreciate realism and stories that make them think (Appleyard 1991). In particular, novels should treat topics of specific interest to adolescents, such as death and sexuality (Sellinger Trites 2000; James 2009).

C. Linguistic needs of less-skilled readers

Pre-adolescents and adolescents can – perhaps – be considered less-skilled readers than adults, in so far as they are still building their reading abilities. Less-skilled readers have problems understanding specific language features which should thus be avoided. These include: embedded subordinate clauses, hidden negative clauses, and the passive voice (Reid – Donaldson 1977, in Gamble – Yates 2002); cataphoric reference, ellipsis, and conjunctive ties (Chapman 1987, in Gamble – Yates 2002); low-frequency words (e.g., Nation – Snowling 1998); and figurative language (Gamble – Yates 2002). Furthermore, less-skilled readers have difficulty in making causal inferences (Long – Oppy – Seely 1997). Finally, they benefit from the presence of section titles (e.g., Yuill – Joscelyne 1988; Cain – Oakhill 1996).

2.1 A linguistic analysis of adaptations for children

As mentioned, to the best of our knowledge, the only detailed linguistic analysis comparing literary adaptations for children to original adult texts is Bianchi (2018). Using corpus-linguistic methods, she analysed two narrative

versions of Shakespeare's *Romeo and Juliet* written in contemporary English by an experienced author specializing in bringing Shakespeare to young audiences. One of the two versions targets teenagers, while the other is for children aged 8-11. Bianchi's study showed how this particular author operationalized the concept of adaptation into a specific – and systematically used – range of structural, linguistic, and stylistic choices, some of which were applied in both adaptations and some in the teenage version only.

The techniques common to both texts are:

- a. Integrating ample stretches of dialogue into the narrative texture.
- b. Conveying the idea of action and continuous changes of scene, linguistically marked by the presence of the verb *go* (e.g., Let 's go to Capulet's party, Romeo. No-one will mind; Alright then, off we go. [END OF CHAPTER]).
- c. Explicating relations between circumstances and events, possibly to facilitate comprehension of the characters' actions and decisions.
- d. Explicating the character's emotions by conveying them through concrete descriptions of their facial expressions.
- e. Explicating the character's distress by underlining it with exclamations 'oh' or 'oh no'.
- f. Preferring overt negative structures to covert ones.
- g. Using very common verbs and nouns with a frequency that is well above average. (e.g., be; go; stop; turn; do; word; find; go; fellow).
- h. Repeating set expressions.
- i. Emphasizing the young age of the characters.

The techniques found in the teenage version only are:

- j. Using long stretches of metaphorical or lyrical descriptions.
- k. Including technical vocabulary, sometimes even requiring knowledge of medieval habits and institutions.
- l. Making frequent reference to women, love and sex.

The techniques observed in Bianchi (2018) serve specific aims. In particular, a. and b., besides maintaining some elements of the original play in the new narrative text type, place emphasis on action and make the plot more dynamic. According to Appleyard (1991; see Section 2, this paper), action and dynamism are requirements for late-childhood readers, as in that age group, characters are what they do, which involves presenting them primarily through dialogue and action, plus a few distinctive traits. Action

is also useful in holding the reader's attention, a fundamental need for readers in their late childhood. Techniques c., d. and e. could be grouped together as explication strategies. Explication strategies are intended to limit the need for the reader to make causal inferences, a difficult cognitive task for the less skilled (see this paper, Section 2, C.). Techniques f., g., and h. are all forms of linguistic simplification, in favour of less-skilled readers (see this paper, Section 2, C.). In particular, the frequent use of common words implicitly entails avoiding low-frequency words. Technique i. helps readers to identify with characters, a specific need of adolescents (see this paper, Section 2, B.). Features j. and k. do not seem to correspond to any particular needs in our framework. Finally, l. represents a form of content selection possibly aimed at meeting adolescents' specific interest in sexuality (see this paper, Section 2, B.). In fact, constant references to women, love and sex, though already powerfully present in Shakespeare's play, does not have to be a dominant feature in an adapted version (as the many purified adaptations across time testify).

3. Materials and methods

To find an answer to the Research Questions outlined in Section 1, corpus linguistics methods were applied. Using Wmatrix (Rayson 2009), a corpus analysis and comparison tool that performs part-of-speech (POS) as well as semantic tagging, *The Da Vinci Code (The Young Adult Adaptation)* (Brown 2016) was directly compared to the original version of the novel (Brown 2003); furthermore, both books were compared to the BNC Sampler Written Imaginative, the latter providing a reference corpus including prose fiction (77%), poetry (13%) and drama (10%) in British English, and the outcomes were contrasted. Comparisons were performed at the level of keywords, key POS tags, and key semantic categories. Focus was placed on items having log-likelihood (LL) higher than 6.63, a threshold corresponding to a significance value of 0.01. The items under consideration were analysed by reading their concordance lines, thus considering the words and the POS or semantic categories in context.

Before that, the two versions of *The Da Vinci Code* were compared at a general quantitative level, considering the features summarized in Table 1. The corresponding measures were obtained using Wordsmith Tools 4.0, a corpus concordancer that provides measurements that include number of sentences and mean sentence length.

Table 1. General quantitative overview of the two versions

	Adult version	Young version
Tokens	141,905	89,006
Types	11,424	7,984
Type-Token Ratio (TTR)	8.06	8.98
Standard TTR (1000)	46.12	45.06
Sentences	12,896	8,530
Mean sentence length	10.98 (s.d. 8.58)	10.42 (s.d. 8.05)

4. Findings

4.1 Young vs. adult version – a direct comparison

As Table 1 (in Section 3) shows, the young version is much shorter than the original version, with -37% tokens (words), -30.11% types, and -34% sentences. On the other hand, TTR, standard TTR and mean sentence length are similar. These raw data suggest abridgement but little stylistic difference between the two texts.

4.1.1 An analysis of keywords

The two texts were directly compared to each other at word level, which retrieved 45 positive keywords characterizing the young version and 22 positive keywords characterizing the adult version. Of these, only 12 items appear in the young version and 28 in the adult version. To make sense of the extracted keywords, it was necessary to read their concordance lines and scan each novel's word list. This led to the identification of the following types of events: 1. substitutions (words replaced by other words); 2. deletions (words or phrases removed from the text); 3. additions (words or phrases added to the text).

More specifically, in the adaptation process for young adults, the author replaced:

- a. Words derived from Latin with words derived from the Anglo-Saxon or Old English stages of the English language, thus replacing for example *erased* or *obliterated* with *wiped out*, *exited* with *went out*, *trepidation* with *unease*, *contention* with *train of thought*, *decelerated* with *slowed down*.

- b. Technical and specific expressions with more generic or clearer ones, thus passing, for example, from *secretariat* to *secretaries*, *docent* (in the meaning of lecturer in a cathedral) to *verger* (someone who acts as an attendant during ceremonies), from *cryptanalysts* to *code-breaker*, from *high-strength polypropylene lamps* to *high-power lights*, from *blurbs* and *endorsements* to *quotes*, from *Father* to *Reverend*, from *ymbologist* to *expert*, from *palate* to *taste buds*, from *my zodiac iconography* to *my knowledge of the zodiac*, from *the modernist Cubist movement* to *the early twentieth-century Cubist movement*.
- c. (Phrases including) proper names with more generic ones, thus for example replacing *Mount Vesuvius* with *volcano*, and *Gare du Nord train terminal* with *the train station*.
- d. Phrases referring to sex with more general ones, thus for example replacing the phrase *sexual abuse of children* with *terrible misdeeds*.
- e. American English words with British English words, thus passing for example from *restroom* to *toilet*, from *rotary* to *roundabout*, from *gas* to *petrol*, from *nibblies* to *nibbles*, from *movies* to *films*, from *movie theatre* to *cinema*, from *trash can* to *rubbish bin*, from *closet* to *wardrobe*, from *vest* to *waistcoat*, from *vacation* to *holiday*, from *casket* to *coffin*. (It must be noticed however that one case of counterevidence was also found – see example (2) below – where the author replaces UK with US spelling, passing from *Centre* to *Center*).

Cases from a. to c. above clearly represent attempts to simplify the language in the text. Case d) is illustrative of the purification approach. The replacement of American with British words is a case of localization. Its rationale, however, is not easy to explain, especially if we consider that not only the hero in the book but also Dan Brown himself and the publishing house of the young version are all American. A possible hypothesis could be that the publishers had a British audience in mind for the young version. This would seem to be supported by the deletion of the reference to Scotland – unnecessary for a British audience – in example (1) below; however, why should an American publishing house limit its target readers to the UK only? It goes against the logic of commerce and profit. Another possible hypothesis – alarming as it may be from the perspective of modern linguistics – is that either author or publisher (or possibly both) believes that British English is or should be the standard variety for children to learn. This however would suggest that all the efforts made by English linguistics to recognize equal status to its different varieties have not yet filtered out of academia.

Deletions seem to be motivated by a desire to shorten and simplify the text by removing expendable details, as in the following examples:

(1) *Adult version:*

Rosslyn Chapel – often called the Cathedral of Codes – stands seven miles south of Edinburgh, Scotland, on the site of an ancient Mithraic temple.

Young version:

Rosslyn Chapel stands seven miles south of Edinburgh, on the site of an ancient temple built in honor of the god Mithras.

(2) *Adult version:*

London's Opus Dei Centre is a modest brick building at 5 Orme Court, overlooking the North Walk at Kensington Gardens.

Young version:

The Opus Dei Center in London is a modest brick building at 5 Orme Court in Kensington, West London.

Finally, additions seem to fulfil a need for clarification, as in the following examples:

(3) *Adult version:*

London's Opus Dei Centre is a modest brick building at 5 Orme Court, overlooking the North Walk at Kensington Gardens.

Young version:

The Opus Dei Center in London is a modest brick building at 5 Orme Court in Kensington, West London.

(4) *Adult version:*

Langdon looked again at the digits, sensing it would take him hours to extract any symbolic meaning. If Saunière had even intended any. To Langdon, the numbers looked totally random. He was accustomed to symbolic progressions that made some semblance of sense, but everything here – the pentacle, the text, the numbers – seemed disparate at the most fundamental level.

Young version:

Langdon looked again at the digits. The numbers appeared to be totally random. When numbers were used as part of a system of symbols, they usually made some sort of sense – a progression or

pattern, for instance. But nothing here – the pentacle, the text, the numbers – seemed to have a link to each other.

Example (4) also indicates ample reformulations and substitution of metaphorical expressions with more concrete ones (*seemed disparate at the most fundamental level* vs. *seemed to have a link to each other*).

4.1.2 An analysis of key POS tags and key semantic tags

Direct comparison between the two texts at the level of key POS tags retrieved six items with statistically significant higher frequency in the young version and only two items with statistically significant higher frequency in the adult version (Table 2).⁶

Table 2. Key POS tags characterizing each version

Young version	Adult version
PPIS1: 1st person sing. subjective personal pronoun (I)	JJ: general adjective
RL: locative adverb	NN2: plural common noun
PPHS1: 3rd person sing. subjective personal pronoun (he, she)	
PPY: "you"	
VBZ: "is"	
VVD: past tense of lexical verb	

Furthermore, a direct comparison between the two texts at the level of key semantic tags retrieved two semantic fields with a statistically significant higher frequency in the young version and four fields with a statistically significant higher frequency in the adult version (Table 3).

Table 3. Key semantic tags characterizing each version

Young version	Adult version
Z8: Pronouns	S3.2: Relations: Intimacy and sex
Q2.1: Speech: Communicative	Z99: Unmatched
	Z3: Other proper names
	O1: Substances and materials generally

⁶ The table reports the tag followed by the official explanation provided in Wmatrix.

These findings suggest a relatively greater presence of dialogues in the young version, as shown by:

- Presence of pronouns 'I' (PPIS1) and 'you' (PPY) and the simple present tense, third person singular, of verb 'be' (VBZ) among the key POS tags in the young version, and presence of category Z8, i.e., pronouns, among key semantic fields. In particular, as many as 30% of the instances in category Z8 correspond to first- and second-person pronouns and possessive adjectives.
- Presence of key semantic tag Q2.1: Speech: Communicative, represented by verbs that introduce direct or indirect dialogue (e.g., "*I'm sorry, Langdon said*";⁷ "*As I told you, he said, [...] we have touched nothing*") or refer to (ongoing) conversation (e.g., "*What is he saying?*").

Furthermore, the data indicate a relative prevalence of narration of events, comparatively fewer or shorter descriptions and a greater attempt to narrate through 'cinematic' images, respectively conveyed by:

- Lexical verbs in the past tense (VVD).
- A significantly lower frequency of general adjectives (JJ) and plural common nouns (NN2) in the adult version.
- The frequent use of locative adverbs (RL) (e.g., "*Outside, the city was just now winding down*"; "*Saying nothing, he stared dead ahead at the chrome doors*"), but also reference to (in)distinguishable dialogue in the background or quality of voice (Q2.1: Speech: Communicative; e.g., "*Voices echoed down the marble corridor*"; "*When we possess the keystone, the Teacher whispered, [...]*").

Finally, these findings suggest ample reformulation. This is indicated by the presence of key POS tag PPHS1 in the young version, third person pronouns being an easy replacement for more complex noun phrases whenever anaphoric reference is possible, but above all by the total or relative absence of some semantic fields characterizing the adult version. More specifically:

- S3.2 is a positive key semantic tag for the adult version, which suggests the reduction or removal of words indicating intimacy or sex in the young version.
- Semantic fields Z3: Other proper names and Z99: Unmatched also are positive key semantic tags for the adult version. The analysis of their concordances shows that they include names of characters, places

⁷ Here and elsewhere in the text, underlining indicates the node word.

and gods, as well as words in Italian or French and some adjectives that must have been removed or replaced with something else in the young version.

- Semantic field O1: Substances and materials generally – a positive key semantic field in the adult version – contains seven items in all that show other reformulation strategies, including the elimination of metaphorical expressions (“[...], *vertical dignity that seemed more a by-product of noble ancestry than any kind of conscious [...]*”; “*He let the pain of his body fuel his supplications*”; “*The mention of Sophie’s name had been the catalyst*”), the elimination of text that slowed down the advancement of the plot (“*Sophie drank her tea and ate a scone, feeling the welcome effects of caffeine and food*”; “[Langdon’s book] included several sections about Mary Magdalene that were going to raise some eyebrows. Although the material was well documented and had been covered by others, Faulkman had no intention of printing [...]”), the substitution of less familiar terms with more commonly used ones (“*grabbing a grease pen*”, *grease pen* being a less common word for *marker*), and simplification and abridgement (“*As a veteran of la Guerre d’Algérie, the curator had witnessed this horribly drawn-out death before. For fifteen minutes, he would survive as his stomach acids seeped into his chest cavity, slowly poisoning him from within*” is replaced in the young version with “*From his war experiences he knew he had fifteen minutes to live*”).

4.2 Young and adult versions compared to the BNC Sampler Written Imaginative

Comparison of the two corpora with the BNC Sampler Written Imaginative offers an indirect view of differences between the two versions.

Such comparison retrieved 19 key POS tags for the young version and 22 for the adult version; of these, 17 are common to both texts and will thus be ignored in the current analysis.⁸ The different ones are listed in Table 4 and discussed in this section. A comparison of the two versions with the BNC at the level of semantic tags returned 86 key semantic fields for the young version and 102 for the adult version. Of these, 82 are common to both texts and will thus be ignored; the remaining key semantic fields – specific for each version of the novel – are listed in Table 5.

⁸ The common tags would be useful for an analysis of Dan Brown’s style, but provide no information as for the differences between the two versions of the novel.

Table 4. Key POS tags vs. BNC Sampler Written Imaginative

Young version	Adult version
PPHS1: 3rd person sing. subjective personal pronoun (he, she)	MCMC: hyphenated number (40-50, 1770-1827)
RP: prep. adverb, particle (e.g., about, in)	RGR: comparative degree adverb (more, less)
	VBDZ: was
	VCN: been
	DA2: plural after-determiner (e.g., few, several, many)

Table 5. Key semantic tags vs. BNC Sampler Written Imaginative

Young version	Adult version
H1: Architecture, houses and buildings	A1.4: Unlucky
N3.8+: Speed: Fast	A11.1+: Important
S1.1.4+: Deserving	A4.2+: Detailed
X9.1-: Inability/unintelligence	A6.2-: Comparing: Unusual
	I1.3+: Expensive
	N3.7++: Long, tall and wide
	N3.7+++ : Long, tall and wide
	O4.2++: Judgement of appearance: Positive
	S7.3+: Competitive
	X2.1-: Without thinking
	X3: Sensory
	X4.1: Mental object: Conceptual object
	A9: Getting and giving: possession
	S7.1: Power, organizing
	W5: Green issues
	Q4: The Media
	Q4.1: The Media: Books
	Q4.3: The Media: TV, Radio and Cinema
	S1.1.3: Participation
	S1.1.3+++ : Participating

The data are indicative of four specific strategies. These are summarized in the following paragraphs.

First of all, the data show a greater attempt to narrate through 'cinematic' images in the young version. Detailed analysis of POS tag RP and its concordance lines – which characterize the young version – highlighted that 89.7% of the prepositions in this tag were used to describe movement (56.9%), gaze (20.1%), surroundings (4.9%), places (3.0%) or sound (4.7%), thus contributing to a lively, cinematic narrative, while the remaining 10.3% of instances referred to Relationship (1.2%), Time (3.0%) and Other (e.g., phrasal verbs; 6.1%). Similarly, the presence of the semantic field N3.8+ (Speed: Fast) can be interpreted in the light of the young version's greater focus on action.

Secondly, the data testify to the shortening or omission of elements that slow down the narration. This is suggested by the POS and semantic tags that are key in the adult version but not in the young adaptation. POS tag MCMC includes hyphenated numbers: they belong to a rather long list of names and dates of the Grand Masters of the Priory, reported in the adult book but not in the young version. The remaining POS tags characterizing the adult version all point to the elimination, reformulation or shortening of descriptions in the young version in favour of events. In fact, comparative forms (RGR) and quantifiers (DA2) appeared in descriptive sentences; in 74.4% of cases, the word 'been' (VBN) was preceded by 'had', i.e., it belonged to a past perfect verb phrase, which suggests the author largely removed or shortened descriptions of facts preceding the timeline of the story, possibly because they provide only background information on characters and do not advance the story; finally, concordances of the verb 'was' (VBDZ) are all clear cases of descriptions. Similarly, key semantic tags A1.4-, A11.1+, A4.2+, A6.2-, I1.3+, N3.7+, O4.2+, S7.3+, X2.1-, X3, and S7.1 indicate a greater presence of adjectives and adverbs (i.e., descriptive elements of places, peoples and events) in the adult version. Finally, the presence of semantic tags A9 and X4.1 in the adult version corresponds to the elimination of side comments and side details that are not fundamental to the plot. This is particularly evident in key semantic tags Q4, Q4.1, Q4.3, S1.1.3 and S1.1.3+, all referring to the role of the media in the story and on Langdon being a writer.

Thirdly, there is evidence of focus shifts. Compared to the young version, the 2003 version puts greater emphasis (in proportional terms over the entire text) on Nature as a pre-Christian object of worship, but also the need to understand the 'true nature of the Holy Grail' (key semantic tag W5). On the other hand, the young version emphasizes the idea of

worthiness (e.g., being worthy of the Grail; being worthy of unlocking the keystone; only the worthy ones will receive a reward; key semantic tag S1.1.4+). Whether the above are the result of voluntary decisions or rather the outcome of shortening and simplifying is impossible to establish. Furthermore, key semantic tag H1 suggests that in the young version the author, while generally shortening or removing descriptions of people and places, indulges in the description of buildings, the reason for this possibly being his desire to help children to discover art and architecture (also supported by the presence of colour pictures; see Section 1).

Finally, these data suggest ample reformulation in passing from the adult to the young version. This is signalled by key POS tag PPHS1 in the young version, which indicates reformulation by anaphoric reference, as observed in section 4.1.

5. Conclusions

This study has applied basic and advanced methods typical of corpus linguistics to compare and contrast two versions of *The Da Vinci Code* by Dan Brown, in an attempt to shed light on this author's view of adaptation to the needs of a young audience. By approaching the texts from different angles – keywords, key parts of speech, and key semantic domains – and by comparing them to each other but also to an external reference corpus – the BNC Sampler Written Imaginative – the analyses have managed to identify the strategies at play and also illustrate the structural, semantic, linguistic, and stylistic choices used by Dan Brown to operationalize such strategies.

Four out of the six adaptation strategies listed by Klingberg (1986; Section 2, A) seem to be at play in this young version:

- Purification, i.e., adaptation based on what society believes to be morally appropriate for children, observable in the removal or replacement of references to sexual matters (see Sections 4.1.1 and 4.1.2). This was one of the author's stated purposes in creating a new adapted version (see Section 1).
- Language adaptation, achieved through the substitution of: technical and specific expressions with more generic or clearer ones (Section 4.1.1); words derived from Latin with words derived from the Anglo-Saxon or Old English stages of the English language (Section 4.1.1); (phrases including) proper names with more generic ones (Section 4.1.1); less familiar terms with more commonly used ones (Section 4.1.2); complex noun phrases with third person pronouns whenever anaphoric

reference is possible (Sections 4.1.2 and 4.2); metaphorical expressions with more concrete ones (Section 4.1.1). Furthermore, the elimination of metaphorical expressions (Section 4.1.2) can also be considered a form of language adaptation.

- Abridgement, achieved through the deletion of expendable details, i.e., those parts of the original text which are not directly functional to the advancement of the plot, such as descriptions of people and places, side comments and details (Sections 4.1.1, 4.1.2, and 4.2).
- Localization, evidenced by the replacement of American English words with British English words. However, if this strategy is clear, its rationale is dubious, and any possible explanation remains in the realm of hypotheses (Section 4.1.1).

Other phenomena were also observed, i.e., additions for clarification purposes (Section 4.1.1), emphasis on dialogue (Section 4.1.2), the narration of events through 'cinematic' images (Sections 4.1.2 and 4.2), and focus shifts (Section 4.2). Although some focus shifts are difficult to explain, the emphasis placed on the description of buildings in the 2009 version was certainly deliberate, a way to attract the young reader's attention towards art and architecture (Section 1) and, matched to the presence of colour pictures, fulfil an instructive aim. Beckett's (2009) observation of a general pedagogical drive behind adapting literature to young readers, thus, applies also to this new millennium.

Some of the adaptations observed in *The Da Vinci Code – Young Adult Edition* are also in keeping with the needs of pre-adolescent and adolescent readers, who require a focus on dialogue and action and limited descriptions of characters and settings (Section 2, B). At the same time, the purification of sex-related references sharply contrasts with their needs and interests and indicates a rather puritan attitude to education. Furthermore, the analyses suggest that Dan Brown (or the publishers) considers teenagers to be less experienced readers, since at least a couple of the linguistic features that are difficult for less skilled readers (Section 2, C) were rather systematically addressed. In fact, low-frequency words, including words of Latin origin and technical or specialized terms, were replaced with corresponding high-frequency expressions, and figurative language was removed or replaced with more concrete expressions (Sections 4.1.1 and 4.1.2).

Finally, the structural, semantic, linguistic, and stylistic choices observed in this study show some overlapping with those discovered by Bianchi (2018) in her analysis of two *Romeo and Juliet* narrative adaptations for young audiences. In particular, the observation of emphasis on dialogue

and (cinematic) action also to be found in Dan Brown's text for teenagers confirms that this derives from the idea of adaptation and not from the type of source text. Furthermore, despite what could be called 'technical' differences in the linguistic and stylistic choices observed, the approaches adopted by the two authors largely share the same intents, which can be summarized in the following strategies: simplification (of content and/or language); explication and clarification; and content selection/shift in focus.

The results of these two case studies – i.e., Bianchi 2018 and the current study – suggest the need for a revised classification of adaptation strategies, specific for intralingual adaptations for the young. Furthermore, linguistic analysis of a wider variety of adaptations from different types of source texts could help to outline such revised classification and to identify an even wider set of technical linguistic adjustments.

REFERENCES

Primary sources

BNC Sampler *Written Imaginative*

<https://ucrel.lancs.ac.uk/bnc2sampler/sampler.htm>, accessed December 2022.

Brown, D.

2003 *The Da Vinci Code*. New York: Doubleday.

2016 *The Da Vinci Code (The Young Adult Adaptation)*. New York: Delacorte Press.

Secondary sources

Anderson, K. J.

2000 "Children's literature in English translation." In: O. Classe (ed.) *Encyclopedia of Literary Translation Into English: A-L*. London/Chicago: Fitzroy Dearborn Publishers, 276-277.

Appleyard, J. A.

1991 *Becoming a Reader: The Experience of Fiction From Childhood to Adulthood*. Cambridge: Cambridge University Press.

BBC

2016 "Da Vinci Code: Young adult version to be released", 18th May, <https://www.bbc.com/news/entertainment-arts-36320670>, accessed August 2021.

- Beckett, S. L.
2009 *Crossover Fiction. Global and Historical Perspectives*. New York/London: Routledge.
- Bianchi, F.
2018 "Rewriting *Romeo and Juliet* for a young audience. A case study of adaptation techniques", *Lingue e Linguaggi* 27, 43-65.
- Cain, K. – J. V. Oakhill
1996 "The nature of the relationship between comprehension skill and the ability to tell a story", *British Journal of Developmental Psychology* 14, 187-201.
- Chapman, J.
1987 *Reading: From 5-11 Years*. Milton Keynes: Open University Press.
- Colyard, K. W.
2016 "Teens don't want *The Da Vinci Code* watered down", 18th May, <https://www.bustle.com/articles/161612-the-da-vinci-code-young-adult-adaptation-sparks-controversy-among-ya-readers-writers>, accessed August 2021.
- Deahl, R.
2016 "YA edition of Dan Brown's *Da Vinci Code* announced for fall", 18th May, <https://www.publishersweekly.com/pw/by-topic/childrens/childrens-book-news/article/70408-ya-edition-of-dan-brown-s-da-vinci-code-announced-for-fall.html>, accessed August 2021.
- Erizanu, P.
2016 "Do teens really need a special YA version of Dan Brown's *The Da Vinci Code*?", 18th May, <https://www.theguardian.com/books/2016/may/18/do-teens-really-need-a-special-ya-version-of-dan-browns-the-da-vinci-code>, accessed August 2021.
- Gamble, N. – S. Yates
2002 *Exploring Children's Literature. Teaching the Language and Reading of Fiction*. London: Paul Chapman Publishing.
- James, K.
2009 *Death, Gender and Sexuality in Contemporary Adolescent Literature*. New York/London: Routledge.
- Klingberg, G.
1986 *Children's Fiction in the Hand of Translators*. Malmö: CWK Gleerup.
- Long, D. L. – B. J. Oppy – M. R. Seely
1997 "Individual differences in reader's sentence- and text-level representations", *Journal of Memory and Language* 36, 129-145.
- Nation, K. – M. J. Snowling
1998 "Semantic processing and the development of word recognition skills: Evidence from children with reading comprehension difficulties", *Journal of Memory and Language* 39, 85-101.

- Rayson, P.
 2009 *Wmatrix: A Web-Based Corpus Processing Environment*, Computing Department, Lancaster University. <http://ucrel.lancs.ac.uk/wmatrix/>, accessed January 2018.
- Reid, J. – H. Donaldson
 1977 *Reading: Problems and Practices*. London Ward: Lock.
- Sellinger Trites, R.
 2000 *Disturbing the Universe: Power and Repression in Adolescent Literature*. Iowa City: University of Iowa Press.
- Shavit, Z.
 1986 *Poetics of Children's Literature*. Athens/London: The University of Georgia Press.
- Tufayel, A.
 2016 "Dan Brown is releasing a young adult *Da Vinci Code* and no one's sure why", 18th May, <https://www.newsweek.com/dan-brown-releasing-ya-da-vinci-code-and-no-ones-sure-why-461180>, accessed August 2021.
- Yuill, N. – T. Joscelyne
 1988 "Effects of organizational cues and strategies on good and poor comprehenders' story understanding", *Journal of Educational Psychology* 80 (2), 152-158.

Address: FRANCESCA BIANCHI, Università del Salento, Dipartimento di Umanistici, Ed. Buon Pastore, Via Taranto 35, 73100 Lecce, Italy.
 ORCID code: <https://orcid.org/0000-0002-9382-4521>

Address: ELENA MANCA, Università del Salento, Dipartimento di Umanistici, Ed. Buon Pastore, Via Taranto 35, 73100 Lecce, Italy.
 ORCID code: <https://orcid.org/0000-0002-4742-7330>