




Coherence in popularized medico-pharmaceutical reports: The case of EPAR summaries for the public

Aage Hill-Madsen¹ ·  <https://orcid.org/0000-0002-0843-3252>
Aalborg University, Denmark

ABSTRACT

The present article investigates a particular lay-oriented medico-pharmaceutical genre, viz. the *European Public Assessment Report – summary for the public*, which is a summarized version of a much longer specialized report published by the European Medicines Agency. The specialized report sets out the results of the Agency's evaluation of a given pharmaceutical company's application for marketing authorization of a medicinal product in EU countries. The *summary for the public* reports the gist of the specialized source text in a linguistic register suitable for a lay readership. However, while the summaries are thus products of registerial adaptation, it is uncertain whether the rewriting results in texts that are sufficiently *coherent*, given that the key features of summary writing – condensation and selection of information – may be at odds with the production of coherent text. Therefore, based on an analytical methodology originating in Systemic-Functional Linguistics, this article investigates a small corpus of *EPAR summaries* to assess the level of coherence in these rewritten texts. The investigation uncovers the ways and the extent to which the summaries employ structural and non-structural cohesive devices to achieve coherence. It is concluded, however, that while the use of these devices in the texts is such as to create a clear potential for coherence, this potential is not adequately exploited, which means that the level of coherence in the texts leaves something to be desired.

Keywords: coherence, cohesion, health communication genres, rewriting, expert-lay communication

¹ **Corresponding author** – Kroghstræde 3, 2-212, 9220 Aalborg Øst (Denmark).
E-mail: hill@ikl.aau.dk

RESUMEN

El presente artículo investiga un género médico-farmacéutico particular orientado a los profanos, a saber, el Informe Público Europeo de Evaluación - resumen para el público general, que es una versión resumida de un informe especializado mucho más largo publicado por la Agencia Europea de Medicamentos. El informe especializado expone los resultados de la evaluación por la Agencia de la solicitud de una determinada empresa farmacéutica para la autorización de comercialización de un medicamento en los países de la UE. El resumen para el público recoge lo esencial del texto fuente especializado en un registro lingüístico adecuado para un público no especializado. Sin embargo, aunque los resúmenes son, por tanto, producto de la adaptación registral, no está claro si la reescritura da como resultado textos suficientemente coherentes, dado que las características clave de la redacción de resúmenes - condensación y selección de información - pueden estar reñidas con la producción de textos coherentes. Por ello, basándose en una metodología analítica procedente de la Lingüística Sistémico-Funcional, este artículo investiga un pequeño corpus de resúmenes EPAR para evaluar el nivel de coherencia de estos textos reescritos. La investigación desvela las formas y el grado en que los resúmenes emplean dispositivos cohesivos estructurales y no estructurales para lograr la coherencia. Sin embargo, se llega a la conclusión de que, si bien el uso de estos dispositivos en los textos es tal que crea un claro potencial de coherencia, este potencial no se aprovecha adecuadamente, lo que significa que el nivel de coherencia de los textos deja algo que desear.

Palabras clave: coherencia, cohesión, géneros de comunicación sanitaria, reescritura, comunicación experto-no experto

1. Introduction

During the last couple of decades, healthcare systems have become much more patient-oriented than used to be the case. This development means that patients are no longer relegated to the role of passive receivers of treatment, dispensed by uncommunicative health practitioners. Nowadays, *patient engagement* has come to the fore, making healthcare “a partnership among practitioners, patients, and their families (when appropriate) to ensure that decisions respect patients’ wants, needs, and preferences and that patients have the education and support they need to make decisions and participate in their own care” (Carman et al., 2013, pp. 223-224). Within written health communication, the modern trend towards stronger patient engagement manifests itself in a broad range of genres aimed at providing patients with information and (in some cases) instructions in relation to their condition and the care they are receiving or about to receive. One of these genres – a purely informative one – is the object of study for the present article, viz. the *European Public Assessment Report – summary for the public* (henceforth the *EPAR summary* for short), published by the EU’s medicines regulator, the European Medicines Agency

(EMA). Briefly told, texts belonging to the genre summarize the most important points and findings from a much longer, scientific medico-pharmaceutical report in which the EMA evaluates a given pharmaceutical company's application for authorization of a particular medicinal product for marketing in the EU. By rendering the main points from this specialized report, the *EPAR summary* serves as a brief layman's introduction to the drug and to the reasons for its release onto the market (for further details on the contents of the genre, see Section 2).

However, it is a well-known fact that communicating specialized knowledge and information to an audience of non-experts is a thorny undertaking, with a considerable potential for miscommunication (see, e.g., Hill-Madsen, 2022a; 2022b on registerial inconsistencies in knowledge communication genres aimed at lay readerships). Moreover, problems may arise from the fact that some popularized texts are not just 'specialized knowledge couched in lay terms', but are products of textual *derivation*, being summarized versions of specialized sources, as in the present case. Summarizing, as Hidi and Anderson (1986, p. 473) point out, entails both condensation and a selection of information. This, indeed, is characteristic of *EPAR summaries*, which, metaphorically speaking, are 'jig-saw puzzles' of snippets of information derived from their much longer specialized sources (for further details, see Section 2). Given that *EPAR summaries* are products of such derivation, condensation and registerial transformation, the present article investigates to what extent specimens of the genre achieve the qualities associated with properly functioning *texts*. Here, the essential property of 'text-ness' or *texture* (etymologically, 'interwovenness (of threads)') will be equated with *coherence*, in the sense of 'semantic continuity and connectedness'. In other words, the question being asked here is: To what extent do the processes of derivation indicated above produce coherent *texts*? Apart from being a question of interest *per se* to studies of text generation and expert-lay communication, the question is important from a health-communication perspective, since a *lack* of coherence may be a source of confusion for readers (see Sanford, 2006, p. 585) and thus potentially damaging to the *patient engagement* agenda referred to above.

Based on Hasan (1984, p. 184; see also Hill-Madsen, 2019), the present investigation assumes that coherence is a variable phenomenon: Texts may be *more* or *less* coherent, and language users are sensitive to such variation (Hasan, 1984, p. 184). To gauge degrees of coherence, an analytical framework derived from Systemic-Functional Linguistics (SFL) will be applied, of which a certain set of principles known as *Cohesive Harmony Analysis* (to be accounted for in Section 3) enables quantification. Data consist of a small corpus of six texts (to be detailed in Section 2) sampled from the EMA's website.

2. Contextualization of the EPAR summary genre and data selection

The *EPAR summary* genre is relatively young, dating back only around 20 years. The genre was created in 2004 by EU law stipulating that “[t]he European Public Assessment Report (EPAR) shall include a summary written in a manner that is understandable to the public” (European Parliament/European Council, 2004, article 13(3)). In a reflection paper published by the EMA in 2006, the contents of the texts and the intended target audience were defined in the following manner:

It is suggested to keep the scope of the EPAR summaries at a basic level. At the same time, the contents should include an appropriate amount of information, enabling patients and the general public to obtain adequate information of the given product. In other words, the summaries will target the ‘average layperson’ both in terms of readability and contents. (EMA, 2006, p. 3)

Content-wise, *EPAR summaries* are derived from two specialized medico-pharmaceutical texts, viz. the *EPAR – public assessment report* and the *Summary of Product Characteristics*. The former reports the EMA’s full assessment of the drug producer’s application for approval of the drug for marketing in the EU. From this specialized assessment report, the *summary for the public* recontextualizes information about indications (the type of disease/disorder that the drug targets), mechanism of action (how it works), and the gist of the results of the clinical trials preceding authorization as well as the drug’s risk profile. From the *Summary of Product Characteristics* is sourced information about dosage and method of administration (how to take the drug). In accordance with the above quote from the EMA reflection paper, the information from the specialized sources is registerially adapted via a process of lexical and grammatical simplification² so as to be stylistically suitable for a lay target readership. Significant to present purposes is the complexity of these textual origins: Given that the *EPAR summaries* are products of textual derivation and recontextualization, with two different texts as sources, there would seem to be at least potential obstacles to the possibility of creating coherent texts, with concomitant comprehensibility problems for the target audience. Regarding comprehensibility, two investigations of this particular aspect of the

² For the specific grammatical and lexical simplification procedures underlying the textual derivation of the EPAR summaries from their specialized sources, see Hill-Madsen (2022). See also Muñoz-Miquel (2012), Hill-Madsen (2015), Hill-Madsen and Pilegaard (2019), Hill-Madsen (in press) for the specific types of simplification procedures applying to the parallel lay-oriented genre of *Patient Information Leaflets*, also published by the EMA and derived from the very same specialized sources as the *EPAR summaries*.

genre, viz. Askehave and Zethsen (2008)³ and Raynor and Bryant (2013), both conclude that the texts are indeed problematic in this respect. The two studies, however, do not indicate whether the problems with comprehensibility are linked with inadequate textual coherence.

However, the investigations by Askehave and Zethsen (2008) and Raynor and Bryant (2013) date sixteen and eleven years back, respectively, which means that their findings may no longer be valid, especially since the genre underwent certain stylistic changes in 2017 (see EMA, 2017). To compile a corpus representative of the genre in 2024, therefore, the sampling window opted for was a 12-month period from August 2022 to August 2023 (the latter being the point of sampling). For the sake of representativeness, the sampling principle adopted was *maximum diversity* (see Patton, 2002), which was ensured by selecting texts that all belong to different ATC categories.⁴ Since the genre is highly standardized, with a number of formulaic wordings recurring in virtually all texts, a corpus size of six specimens was found to be sufficient to allow generalization to the genre as a whole. In terms of length, the texts range between 1.6 and 3.3 standard pages (see also Subsection 4.5).⁵

3. Coherence and cohesive devices

3.1. The concepts of coherence and cohesion

In text-linguistic literature, there appears to be consensus about informally defining the concept of *coherence* as referring to the textual property of ‘connectedness’ (Sanford, 2006; Hasan, 1984). However, when it comes to a deeper understanding of the particular character of such ‘connectedness’, interpretations diverge. Thus, in the cognitive, or psycholinguistic, approach, coherence is conceived of as a quality to be

³ Askehave and Zethsen’s (2008) study, it should be noted, was concerned with the Danish version of the texts. The present study, on the other hand, focuses on English-language originals.

⁴ ATC refers to the Anatomic Therapeutic Chemical classification system, which is “an international systematic standard for pharmaceutical drugs that groups them according to the organ or tissue on which they act, their mode of action, and their chemical composition.” (Porta and Last, 2018).

⁵ Bibliographic data for all six texts are provided in Appendix A. In connection with in-text citations, the individual texts will be referred to by the name of the medicinal product in question. One exemplary text, concerned with the product *Sunlenca*, is reproduced in full length in Appendix B.

established in the reader's mind (Blum-Kulka, 1986; Carrell, 1982; Givón, 1995; Sanders et al., 1993; Charolles, 1997; Seidlhofer & Widdowson, 1999): A "mental textual world", in the words of Menzel et al. (2017, p. 1), generated from input on the linguistic surface in the form of cohesive markers (Givón, 1995). In this way, the cognitive approach shares with the functionalist/sociosemiotic approach the distinction between coherence and cohesion. In the two approaches alike, the distinction is one of meaning versus linguistic form: Coherence is a semantic property of texts (semantic connectedness), whereas cohesion and cohesive features are linguistic surface phenomena. The two approaches would also concur on the connection between coherence and cohesion: Coherence is something *cued by* and *dependent on* cohesive markers in a text (Parsons, 1991; Thompson, 1994; Hoover, 1997; Tanskanen, 2006, pp. 20-21; Givón, 1995). The discord between the two approaches relates to the very nature and locus of semantics (cf. Halliday & Matthiessen 1999, p. 416): Whether meaning is a language-immanent phenomenon, as held by the functionalist tradition, or a transcendent one residing outside language, e.g. in the minds of interlocutors, as in the cognitive tradition. As in Hill-Madsen (2019) and Hill-Madsen and Dam-Jensen (2022), the former view (according to which meaning is an immanent feature of language) will be adopted, as argued by Hasan:

I believe that the act of meaning is made possible only through the creation and existence of codes which provide the potential for meaning; and that in a very important sense, we are able to mean through language, by virtue of the fact that the signs of language have meanings quite irrespective of what any one individual might contrive them to mean on an individual occasion. (Hasan, 1984, p. 161)

In other words, texts mean what they do because they instantiate (parts of) the linguistic code or system, which is a socially determined phenomenon. It is the social use of signifiers by a community that establishes the relations between linguistic forms (lexis and grammar) and meanings, and not the individual, cognitive act of a given language user. If/when, in the process of being decoded, a text produces a cognitive response in a reader's mind, it is because the meanings realized by the linguistic forms have already been established by the language community, and the readers' cognitive response is dependent on their being already familiar with these meanings (cf. Myers, 1991). Thus, coherence, too, is a language-immanent property of texts:

When I say that coherence in a text is the property of hanging together, I mean that the patterns of language manifest – or realize – the existence of

semantic bonds, because it is in their nature to do so; not simply because someone is making them do so. (Hasan, 1984: 162)

However, it is important to note that the sociosemiotic/functionalist interpretation of coherence (coherence as an immanent feature of texts, as argued by Hasan above, and not a mental phenomenon) should not be taken to mean that coherence is *inherent* in texts. On the contrary, coherence is a variable phenomenon (see above), exactly because it is realized in linguistic forms (cohesive markers), and thus dependent on the *sufficient*, *successful* and *competent* use of such markers. The consequence of too limited or incompetent use of cohesive cues is incoherent text. This is why, as is the aim of the present article, it makes sense to investigate the *degree* of coherence in texts originating in circumstances that are least potentially detrimental to the achievement of a satisfactory degree of semantic connectedness. In Subsection 3.2 below, the various cohesive devices identified by functionalist linguistics will be detailed.

3.2. Cohesive devices: Subtypes and functions

Functionalist literature (e.g., Halliday & Hasan, 1976; Hasan, 1989; Martin, 1992; Taboada, 2004; Halliday & Matthiessen, 2014) recognize the following non-structural phenomena as cohesive devices: *Conjunction* (the use of sentence connectors), *reference*, *lexical cohesion*, *substitution* and *ellipsis*. Further, one structural phenomenon to which cohesive significance is ascribed in a number of studies (e.g., Halliday, 1967; Firbas, 1974; Chafe, 1976; Fries, 1983; Lambrecht, 1994; Gundel & Fretheim, 2004; Büring, 2007; Krifka, 2008; Halliday & Matthiessen, 2014) is the interaction between *thematic structure* and *information structure*. The individual devices will be defined in the following, except for *substitution* and *ellipsis*, which will be ignored henceforth, simply for not being represented in the corpus.

3.2.1. Thematic structure (Theme-Rheme) and information structure (Given-New)

Thematic structure consists of the two constituents *Theme* and *Rheme*, of which Theme is syntactically identical with the initial element of a clause (either the grammatical subject or an adjunct). Semantically, the Theme represents the speaker's point of departure or the topic of their message. The Rheme, then, represents the actual content of, or what is predicated by the speaker about, the Theme (Halliday & Matthiessen, 2014, p. 89). Theme and Rheme are thus sender-oriented constituents, whereas information structure is receiver-oriented. Information structure is composed of the two constituents *Given* and *New*, encoding

information that is ‘familiar’ vs. ‘unknown’ to the listener (Halliday & Matthiessen, 2014, pp. 114-118). Theme-Rheme and Given-New are also very different in terms of realization: Theme-Rheme is realized in grammatical structure with a clear demarcation between the two elements, whereas information structure is realized in intonation, i.e. phonologically, and not grammatically (Halliday & Matthiessen, 2014, p. 117). This means that determining what is Given and what is New in writing can sometimes be problematic. For present purposes, identification of Given vs. New will rely on the semantics of the structure (‘known’ vs. ‘new/unknown’ information). However, Martin (1992) points to a significant grammatical characteristic of ‘given’ information which can be determined without recourse to intonation patterns, and which will accordingly be taken into account in the analyses in Section 4. This feature is the grammatical definiteness/indefiniteness of noun phrases (NPs): “Basically, indefinite nominal groups [*i.e.*, NPs] code the identity of the participant being realised as not recoverable [*i.e.* as *new information*], whereas pronouns, demonstratives, the definite article and proper names signal that the participant’s identity is in some sense known” (Martin, 1992, p. 92).

The significance of thematic structure and information structure is the way the two interact to enable the natural progression as well as the comprehensibility of a text. A ‘proper’ kind of interplay between the two is indeed a prerequisite for the creation of *texture*, in the sense of strings of wordings being ‘woven together’ to form a coherent whole. Thus, the default interaction between the two structures is for the Theme of a sentence to map onto ‘given’ elements (Halliday & Matthiessen 2014, p. 120). This enables speakers to take their point of departure (Theme) in information that is already known to the receiver (‘given’ information, which is the most ‘natural’ kind of starting point), and to proceed to supply new information in delivering the actual content of the message (the Rheme). The natural ‘flow’ of discourse thus takes the shape of a wave-like motion back and forth between ‘given’ and ‘new’ information, sentence per sentence, with each new message starting out from information that the receiver is already familiar with. The analyses (of this aspect of cohesion) in Section 4 will focus on the extent to which this principle is observed in the EPAR summaries.

3.2.2. Conjunction

Conjunction refers to types of logical relations between sentences, i.e. relations such as ‘addition’, ‘contrast’, ‘causality’ and ‘temporality’ (Halliday & Hasan 1976, pp. 226-273). On the linguistic surface, such relations are encoded in sentence connectors (termed *Conjuncts* in Systemic-Functional Grammar) such as *moreover*, *furthermore* (‘addition’), *however*, *nevertheless* (‘contrast’ or ‘opposition’), *therefore*, *consequently*

(‘causality’) and *in the long run, next* etc. (‘temporality’). It should be noted that in many cases, logical relations are not explicitly coded but merely presupposed, as in an example like: *The stallholders at the market started packing up their stalls. Rain was coming*, in which a causal relation between the second and the first sentence is implied. In the present article (as in Hill-Madsen & Dam-Jensen, 2022), however, Halliday and Matthiessen’s (2014, p. 622) principle of “not reading too much into text” will be adhered to, meaning that only relations that are explicitly realized by means of Conjunctions will be registered.

3.2.3. Reference

As defined by Halliday and Hasan (1976, p. 31), *reference* is the phenomenon where certain items on the linguistic surface (personal and demonstrative pronouns primarily) “make reference to something else for their interpretation”. A cohesive tie of reference is thus established between the two items *the stallholders* and *they* in an example like: *The stallholders started packing up the stalls. They had heard that rain was coming*. What is significant to the purposes of this article is the fact that reference items typically form *chains* (Hasan, 1989), whereby the same referent (or *participant* (Martin, 1992) is tracked down through a text, e.g., *the stallholders ... they ... they ... their ... they* (for the significance of cohesive chains, see Subsection 3.3 below).

3.2.4. Lexical cohesion

Lexical cohesion is concerned with the way connections between lexical items contribute to the semantic unity of a text (Halliday & Hasan, 1976, ch. 6; Halliday & Matthiessen, 2014, pp. 642-650). The following short text (an excerpt from a style guide on writing in ‘plain language’ aimed at government officials in the US) may serve to illustrate the principle:

(1) [1] Examples help you clarify complex concepts, even in regulations. [2] They are an ideal way to help your readers. [3] In spoken English, when you ask for clarification of something, people often respond by giving you an example. [4] Good examples can substitute for long explanations. [5] The more complex the concept you are writing about, the more you should consider using an example. [6] By giving your audience an example that’s relevant to their situation, you help them relate to your document. (Plainlanguage.gov, 2011, p. 70)

Apart from the repetition of the exophoric reference to the reader (*you*), and the reference item *they* in sentence [2], the cohesiveness of the excerpt in Example (1) is

largely achieved through the connectedness of lexical items. In texts of sufficient length, such connectedness generates *chains* of multiple lexical items, as in Example 1, where no less than ten such chains run through the short text:

- A chain concerned with ‘exemplification’⁶ (instantiated through the item *example* in sentences [1], [4], [5] and [6])
- ‘concept’ (the item *concept* in [1] and [5])
- ‘complexity’ (*complex* in [1] and [5])
- ‘question and reply’ (*ask for* [3], *respond* [3])
- ‘helping’ (*help* [1] and [6])
- ‘clarification’ (*clarify* [1], *clarification* [2], *explanation* [4])
- ‘providing’ (*giving* [3], *using* [5], *giving* [6])
- ‘readership’ (*readers* [2], *audience* [6])
- ‘texts’ (*regulations* [1], *document* [6])
- ‘linguistic media’ (*spoken* [3], *writing* [5])

It should be noted that the formation of lexical chains may be based on a number of different types of ties between tokens (cf. Martin, 1992, pp. 332-338). As in Example (1) above, the most frequent type is probably *repetition* of the same lexeme, though sometimes with morphological variation (e.g. *clarify* – *clarification*). Other types of ties are the well-known lexical sense relation types (see, e.g., Lyons, 1977), such as synonymy (e.g., *giving* and *using*, which must be considered synonymous in the way they are used in Example 1) and hypo-/hyperonymy, as in *regulations* – *document* (with *regulations* as a hyponym of *document*) and *clarification* – *explanation* (assuming that *explanation* may be regarded as a subtype of *clarification*). Yet other types of semantic ties are antonymy, as in *ask* – *respond*, and meronymy/holonymy (not instantiated in the Example (1)) but with an invented example being, e.g., *document* – *section*). What should also be noted is the fact that one and the same chain may feature several different types of ties. This was the case in what was dubbed the ‘clarification’ chain in Example (1), in which the relation between the first and the second item in the chain was repetition (of the lexeme *clarify*), but hyponymy between the second and the third item (*clarification* – *explanation*).

⁶ One-word names (in very few cases multi-word) in single inverted commas will be used to refer to individual lexical chains. In each case, the name is intended to capture the semantic content or theme of the chain. Individual tokens belonging to the chain are italicized.

3.3. The principles of Cohesive Harmony

It is a central point in Hasan (1989) that, at least in texts of some lengths, the presence of chains – referential and lexical ones – is a prerequisite for cohesiveness, in that such chains are semantic carriers of thematic (in the sense of ‘topic-based’) relations between sentences. In fact, Hasan (1989) posits that the higher the proportion of items (lexical or referential) that are members of chains, the stronger the cohesiveness of a text. Thus, in Example (1) above, a majority of lexical and referential tokens (35 out of 53 altogether, i.e. around two thirds) are chain members. However, while a necessary condition for cohesiveness, a high proportion of lexical and referential items entering into chains is not a sufficient condition. The presence of chains running ‘vertically’ through a text merely ensures thematic (‘topic-based’) connectivity between sentences but does not *ipso facto* guarantee any connectedness or interrelation *between* such textual topics. For a text to be truly coherent, then, some degree of integration between chains is necessary. On the linguistic surface, such integration – termed *chain interaction* by Hasan (1989, p. 91) – will manifest itself ‘horizontally’ *within* sentences, viz. as syntagmatic relations between tokens from different chains. Thus, to mention an instance from Example (1) above, in the very first sentence the two chains ‘clarify’ and ‘concept’ are linked via the clausal relation between verb and direct object, or, in terms of what is termed *experiential clause structure* in Systemic-Functional Grammar (see Halliday & Matthiessen, 2014, ch. 5), between Process and Verbiage.⁷ A second example is the chains ‘complexity’ and ‘concept’, which are linked at phrase level through a Premodifier–Head relation.

However, according to Hasan (1989, p. 91), a single point of syntagmatic interlinkage between two chains (i.e. between just one token from chain A and one token from chain B) is insufficient for chain interaction to be ‘truly’ present. The reason is that most or all tokens in a chain will necessarily form syntagmatic relations with other message components (lexical or referential items), which will in many or most cases be tokens belonging to other chains: Tokens of chains occur in sentences and thus as part of syntactic configurations, and hence one-point interlinkages with tokens from other chains will be more or less unavoidable. Instead, it is only when a particular type of syntagmatic chain interrelation is *repeated* that integration between two chains can be said to ‘truly’ obtain. Hasan (1989, p. 91) thus posits a minimum of *two* occurrences of the same type of

⁷ In the analyses in Section 4, the clause-level syntagmatic relations to be identified between chains will be those from the experiential clause grammar of Systemic-Functional Grammar. A brief overview of types of configurations will be provided in Appendix D.

syntagmatic relation between tokens from different chains. Once again, Example (1) exhibits this feature also, with the same type of syntagmatic interaction occurring no less than three times between the ‘providing’ and the ‘exemplification’ chains, which are interlinked via the relation between Process and Goal in sentences [3], [5] and [6].

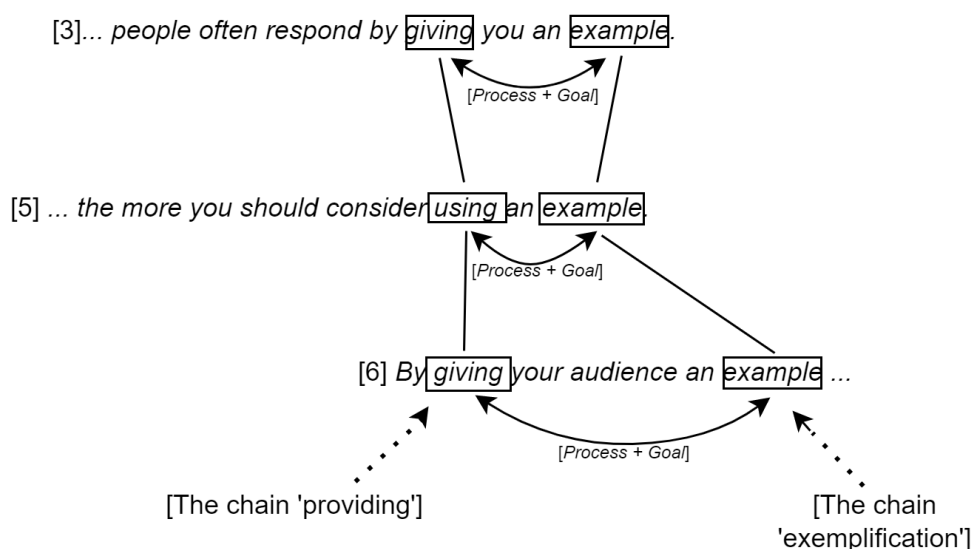


Figure 1. A Chain Interaction Series from Example (1).

In more informal terms, the reason behind the repetition requirement is, in Hasan’s words, that “when speakers are engaged in the process of creating a coherent text, they stay with the same and similar things long enough to show how similar the states of affairs are in which these same and similar things are implicated” (1989, p. 94). In accordance with Hill-Madsen (2019) and Hill-Madsen and Dam-Jensen (2022), recurrent chain interactions will be referred to as *chain interaction series* (CIS). In Fig. 1 above, six tokens enter into the CIS illustrated, viz. *giving* – *using* – *giving* and the three instantiations of *example*.

From the notion of chain interaction as a necessary condition for coherence, it follows that all individual chains should preferably be interlinked with at least one other chain, thus forming a web or ‘mesh’ of interconnections: Chain A in a text may be related to chain B, B to C and D, C to E and F, D to G, G to H and I, etc., ideally forming an unbroken web without interruptions, or at least as few as possible. An uninterrupted web like that will be the linguistic reflex of a fluid and seamless textual progression from topic to topic, for which Hasan’s technical term is *cohesive harmony*. The principle of a tight-knit ‘web’ of interacting chains is illustrated generically in Fig. 2 below, where the vertical rows symbolize imaginary chains of

lexical or referential items (with tokens represented by three letters each), and the curved, double-pointed arrows stand for syntagmatic relations connecting the chains 'horizontally'.

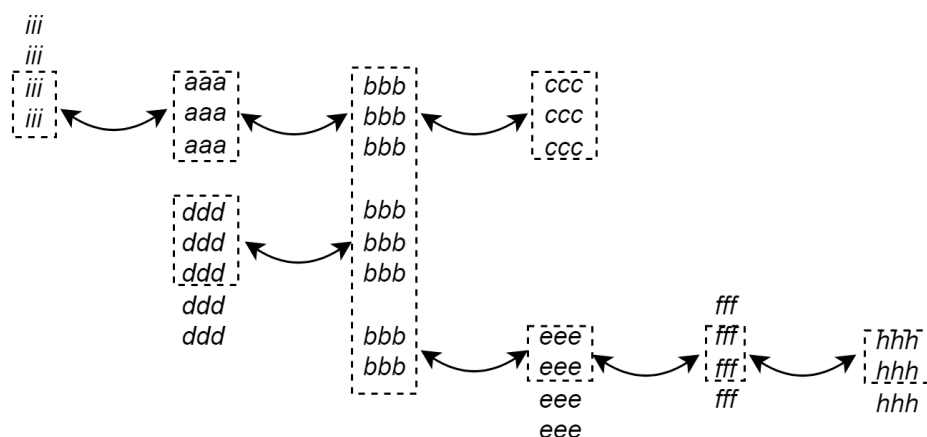


Figure 2. An imaginary web of interacting chains.

3.4. Quantitative operationalization of the principles of Cohesive Harmony

From the account of the principles of *Cohesive Harmony* outlined in Subsection 3.3, four key concepts emerge which allow for quantitative operationalization. The concepts are *relevant tokens*, *peripheral tokens*, *central tokens* and *non-central tokens*:⁸

- *Relevant tokens* are ones that are members of chains.
- *Peripheral tokens* do not enter into any chains.
- *Central tokens* are a subset of *relevant tokens*: *Central tokens* are members of Chain Interaction Series, i.e. tokens that interact syntagmatically (in the way defined in Subsection 3.3) with tokens from other chains.
- *Non-central tokens* constitute the other subset of *relevant tokens*, i.e. those that are not members of CISs.

The semantic connectivity of a text is thus expressible in terms of the following ratios (Hasan, 1989, pp. 93-94):

1. The ratio of relevant (R) to peripheral (P) tokens. The higher the ratio of R to P, the higher the number of tokens that will be related, via chains, to other tokens. The ratio thus measures a text's degree of integration in terms of subject matter

⁸ Definitions in the following are taken from Hasan (1989, p. 93).

(see also Hill-Madsen & Dam-Jensen, 2022): The higher the ratio, the stronger the textual focus on the set of topics being treated.

2. The ratio of central (C) to non-central (NC) tokens. The ratio measures the degree of connectedness *between* subthemes: The higher the ratio of C to NC, the higher the number of tokens that participate in 'horizontal' (syntagmatic) interaction with other chains.
3. The number of interruptions in the web of interrelated chains. The lower the number, the stronger the connectivity of the text and the more fluid its thematic evolution.

To calculate the number of central, non-central and peripheral tokens, Hasan recommends a procedure that she terms *lexical rendering* (Hasan, 1989, p. 87). The procedure consists in interpreting and registering each individual referential item as the corresponding lexical item that it replaces or refers back to. Thus, in Example 1, the referential item *they* in sentence (2) would be lexically rendered as *example*, because this is the lexical item (occurring in sentence 1) that it refers back to. This procedure of lexical rendering is followed in the present investigation.

4. Qualitative and quantitative analyses

The present article follows the analytical procedure outlined in Hill-Madsen and Dam-Jensen (2022), which comprises six steps, with five qualitative and one quantitative (step no. 5). Steps 1-2 focus on the continuity from sentence to sentence, whereas steps 3-5 are concerned with the connectivity produced by cohesive chains.

- 1 and 2: Sentence by sentence, all texts were analyzed for conjunctive relations and for the interaction between thematic structure and information structure. Results will be reported in Subsection 4.1.
- 3: All texts were analyzed for referential and lexical relations throughout, with the purpose of identifying the cohesive chains running through each text. Results will be reported in Subsection 4.2.
- 4: Chain Interaction Series (CISs) were identified through syntagmatic analysis of all clauses and phrases containing adjacent tokens belonging to different chains. The analyses will be illustrated in 4.3.
- 5: Tokens in chains and in CISs were quantified, enabling the calculation of the relevant-peripheral ratio and the central-non-central ratio (see Subsection 3.4). Quantitative results are presented in Subsection 4.4.

4.1. Analytical steps 1 and 2: Connectivity between adjacent sentences

The qualitative analyses reveal what must be considered a satisfactory degree of connectedness between adjacent sentences in all six texts. This is especially because Theme-Rheme and Given-New structures in most sentences interact in the way presented as optimal for sentence-to-sentence connectivity in Subsection 3.2.1, i.e. with the Theme of a sentence intersecting with, or containing, 'given' elements, as in Example 2 below:

(2) [6] Sunlenca is available as tablets to be taken by mouth and as a solution for injection. [7] Sunlenca tablets are taken at the start of the treatment, on days 1, 2 and 8. [8] One week after that, patients are given Sunlenca injections every 26 weeks as maintenance treatment. [9] Injections are given under the skin by a doctor or nurse. (*Sunlenca*)

In Example (2), the Theme of each sentence is underlined, showing that sentences (7), (8) and (9) all link back to a message component introduced or mentioned previously, i.e. to 'given' or 'known' elements: In (7), both *Sunlenca* and *tablets* are mentioned in sentence (6); in (8), the referential item *that* refers back to *the start of the treatment* in (7), and in (9) *Injections* tie in with *injections* mentioned in (8). In all three sentences (7–9), the element(s) 'picked up' by the Theme in all cases occur as 'new' elements in the preceding sentence, which means that the development from sentence to sentence is thematic *progression*, i.e. with the Theme changing from sentence to sentence. It should be noted, however, that across the texts, certain parts tend to be dominated by a lack of Theme variation. This especially applies to the first section, headlined *What is X and what is it used for?*. In these cases, the Theme is typically realized by the name of the medicinal product (e.g., *Sunlenca is used ... Sunlenca is given ... Sunlenca contains ...*). The lack of Theme variation reflects a rather narrow topical focus (on the characteristics of the product) in these sections. Nevertheless, since the product name is a 'given' element in all such sentences, the recurrence of the name in thematic position still ensures intersentential connectivity.

Conjunction, on the other hand, is very rare as a cohesive device in the corpus, with individual texts typically containing only one or two such items. One instance is the item *therefore* in Example 3 below, which is used in all six texts to introduce the central conclusion to the EMA's assessment of the producer's application for approval of the drug:

(3) [33] The side effects of Sunlenca are considered manageable. [34] Therefore, the European Medicines Agency decided that Sunlenca's benefits

are greater than its risks and it can be authorised for use in the EU.
(*Sunlenca*).

The findings give rise to the question whether the general absence of explicit conjunctive relations in the texts may be interpreted to indicate that no such relations are present, or whether they are semantically implicit and up to the reader to infer. Possibly, in all those cases where no conjunctive adverb is present, a relation of 'addition' between two adjacent sentences may be inferred. In this case, the second sentence of two adjacent ones should simply be understood as adding further information in relation to the preceding sentence. However, as previously mentioned, Halliday and Matthiessen (2014, p. 622) caution against assuming such implicit relations, and the question must therefore be left unanswered.

4.2. Analytical step 3: Identification of cohesive chains

While the exact nature of the cohesive chains identified in the corpus is not in itself important to the main purposes of this paper, their main characteristics should nevertheless be summarized: First of all, many of the same chains recur in all texts, which is hardly surprising, given the high degree of standardization that characterizes the genre: All texts rigidly adhere to the very same generic structure, and a number of formulaic wordings recur across the texts. As regards the nature of the chains, it turns out that the majority are lexical ones and only very few referential ones. As for lexical relations, the vast majority consist in mere repetition of the same lexeme, albeit with certain exceptions: All texts feature a chain centered around the name of the medicinal product in question (*Hemgenix*, *Omvoh*, etc.), which is frequently repeated down through the individual text. Nevertheless, in all texts, the repetition sometimes alternates with hyperonymy, manifested in the superordinate lexeme *medicine*. Similarly, in all texts the term for the particular disorder (such as *haemophilia B* in the text about *Hemgenix*, *ulcerative colitis* in *Omvoh*, etc.) alternates with superordinate terms such as *disorder* and/or *disease*. Antonymy also features in connection with certain chains that are common to all texts: One such is concerned with the administration of the drug, instantiated in tokens such as *giving* as well *receiving*, while another is concerned with the effects of the drug, manifested in *benefits* as well as its opposite, *side effects*. Certain texts also feature chains of co-meronymy between 'body parts', such as a *large intestine*, *vein*, *skin*, *anus*, *nose* and *throat*.

4.3. Analytical step 4: Identification of Chain Interaction Series

Below, an example from the text *Sunlenca* has been selected to illustrate syntagmatic interaction between chains:

(4) [a]⁹ **Sunlenca** [b] **is** a [c] new [d] type of treatment for [e] **controlling** [f] **HIV-1** [g] **infection**, [h] which [i] **was** [j] **effective** at [k] reducing [l] **viral** [m] **load** in [n] patients with [o] infections [p] that [q] **are** [r] **resistant** to other [s] **treatments**. (*Sunlenca*)

In Example (4), all tokens that are members of chains are underlined and CIS members are bolded as well. The CISs represented in the example are listed below, with tokens italicized and followed by the semantic designation chosen for the chain in question (except in those cases where the lexical item is identical with the name chosen for the chain). The syntagmatic configuration underlying the chain interaction is indicated next:

- [a] *Sunlenca* ('medicine') and [b] *is* ('being'): Carrier and Process (clause level)
- [e] *controlling* + [g] *infection*: Process and Goal (clause level)
- [f] *HIV-1* ('virus') and [g] *infection*: Premodifier and Head (phrase level)
- [i] *was* ('being') and [j] *effective* ('effects'): Process and Attribute (clause level)
- [l] *viral* ('virus') and [m] *load* ('amount'): Premodifier and Head (phrase level)
- [q] *are* ('being') and [r] *resistant*: Process and Attribute (clause level)
- [r] *resistant* and [s] *treatments*: Head and Postmodifier (phrase level)

The sentence thus features syntagmatic configurations from altogether seven different CISs, with the token from one particular chain (*infection*) participating in two different CISs. In Figure 3 below, the names given to all chains in the *Sunlenca* text are diagrammed,¹⁰ with connecting lines between names indicating chain interactions:

⁹ Letters in square brackets indicate message components (lexical or referential).

¹⁰ Appendix C lists all chains identified in the *Sunlenca* text, with the tokens of each chain indicated.

Text	Hem- genix	Omvo h	Opzel- urah	Sun- lenca	Tezspire	Ximluci	Mean	St. dev.
6. Interacting chains (% of total chains)	62.5	70.9	66	65.1	69.4	75.6	68.3	4.5

Table 1. Quantitative CHA results.

The first horizontal row of the table states the length of each text in terms of standard pages, with one standard page being defined as 1800 characters including spaces. The second row states the total number of message components (MCs) per text. The third provides the number of peripheral tokens as a percentage of total MCs, and the fourth indicates relevant (R) tokens as a percentage of total MCs. In the fifth row, two values are stated: The first indicates central (C) tokens as a percentage of R tokens, and the second indicates C as a percentage of total MCs. Finally, the sixth row states the number of interacting chains as a percentage of the total number of chains in the text. Vertically, the penultimate column provides the average for each type of value (except total MCs, which is irrelevant), and the final column indicates standard deviation.

From the results, the following patterns are discernible:

- The number of tokens entering into chains (relevant tokens) is relatively high for all texts, averaging 80% of total MCs, and the peripheral percentage number is correspondingly low. The very low standard deviation (σ) value (2.7) reflects that all texts are highly consistent in this regard.
- The number of central tokens (i.e. ones participating in chain interaction), however, is not very high, with an average C-to-R ratio of only 48.7 percent. The relatively low σ value indicates that this is a consistent characteristic across the texts also.

Nevertheless, since the σ value for the C-to-R ratios (4.0%) does indicate some variation between the texts in this regard, it needs to be considered whether the variation is linked with differences in text length. At first glance, this might appear to be the case, given that the longest text (*Hemgenix*) represents one of the highest C-to-R ratios (50.6), while at the same time one of the shorter texts (*Sunlenca*) has the lowest (41.4). A possibility would be that a longer text simply provides more and better opportunities for chain interaction. Yet, some of the other texts appear to refute the conjecture: *Omvo*, which is a short text, represents the highest C-to-R ratio, and the very shortest text, *Tezspire*, has a ratio that is almost the same as the very longest text (*Ximluci*). Of course, a much larger sample would be required for a reliable testing of the conjecture to be possible.

With regard to chain statistics (the bottom row in Table 1), the six texts also evidence a relatively clear pattern: In all texts, only around two thirds of all chains interact. As illustrated in Figure 3 above (diagramming the *Sunlenca* text as a representative case in this regard), the interacting chains do tend to form an uninterrupted 'web' of connected chains, but a third of chains are isolated ones.

5. Discussion and conclusion

The analyses in Section 4 have revealed that the *EPAR summaries* evidence clear sentence-to-sentence progression throughout the texts. This, however, was shown to be mainly due to the way Theme-Rheme structures are made to interact with Given-New patterns, in that the Theme of most sentences was seen to contain previously introduced elements, thus ensuring proper connectedness between adjacent sentences. The use of conjunctive cohesion, on the other hand, was shown to play a much more insignificant role in the corpus, with individual texts featuring very few explicit conjunctive relations.

Regarding the non-structural connectedness represented by referential and lexical links, the quantitative results revealed that the chain-dependent cohesiveness of the texts is not particularly strong. Granted, a relatively high percentage of message components do participate in chains, which indicates the texts to be focused in terms of subject matter(s), manifested in (separate) cohesive chains. However, the quantitative results revealed that the syntagmatic interlinkage between chains that is required for a text to be truly coherent is moderate, evidenced in the not-very-high proportion of central tokens. According to Hasan (1984, p. 218), for a text to be "unquestionably coherent", central tokens must amount to at least 50% of all tokens (total MCs). This is a requirement that none of the texts fulfil, with percentage values in this regard being considerably lower (39% on average). What the results reveal is thus that the *integration* of topics in the individual texts is relatively weak, which is also reflected in the fact that around one third of chains in all texts are not linked with other chains. Overall, therefore, it must be concluded that the *EPAR summaries* are not coherent to a satisfactory degree.

Strikingly, much the same type of results was arrived at in Hill-Madsen and Dam-Jensen's (2022) investigation of coherence in a very different summary genre, viz. summaries of economic reports issued by the European Central Bank. In this genre, too, a less than satisfactory degree of coherence was uncovered, with the core of the problem residing, not in the proportion of relevant tokens, which was high, but in limited chain interaction. Thus, the possibility emerges that the less-than-satisfactory degree of coherence may be linked with the fact that the texts in both cases are

summaries. The explanation suggested in Hill-Madsen and Dam-Jensen (2022) was that “these are texts where much more extensive content and a large number of sub-themes are condensed into a very short space” (p. 94). The consequence of such condensation is that cohesive chains typically comprise rather few tokens each, as evidenced in the *Sunlenca* text, in which around half the chains consist of only two or three tokens (see Appendix B). The point is that longer texts with longer chains (i.e. comprising higher numbers of tokens) afford greater *potential* for chain interaction. Whether this hypothesis is sound, and whether it is *inherently* difficult to achieve a satisfactory degree of coherence in summaries (because they are usually of limited length) is a question for future research.

About the author

Aage Hill-Madsen is Associate Professor of modern English language and linguistics at Aalborg University, Denmark. His main research interests are translation, knowledge communication, and register and genre studies. He has published in international journals such as *Meta – Journal des Traducteurs*, *Fachsprache – Journal of Specialized Communication* and *Semiotica*, and he has contributed to *The Routledge Handbook of Intralingual Translation*.

Authorship and contribution policy

Authorship

Conceptualization: Aage Hill-Madsen
Methodology: Aage Hill-Madsen
Software: Not applicable.
Validation: Aage Hill-Madsen
Writing – original draft: Aage Hill-Madsen
Writing – review & editing: Aage Hill-Madsen

Funding

Full name of the funding agency: Not applicable.
Project reference number (if applicable): Not applicable.
Project or program title (if applicable): Not applicable.
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Image use consent

Not applicable.



Conflict of interest

The author declares no conflict of interest.

Data availability statement

The data of this study were in the public domain (open-access website of the European Medicines Agency).

License

This article is published under the CC BY 4.0 License.

Editorial history

Received: 23 February 2024

Accepted: 7 April 2024

Revised: 21 April 2024

Published: 30 June 2025

References

- Askehave, I., & Zethsen, K. K. (2008). Mandatory genres: The case of European Public Assessment Report (EPAR) summaries. *Text & Talk*, 28(2), 167-191. <https://doi.org/10.1515/TEXT.2008.008>
- Blum-Kulka, S. (1986). Shifts of cohesion and coherence in translation. In J. House & S. Blum-Kulka (Eds.), *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies* (pp. 17-35). Narr.
- Büring, D. (2007). Intonation, semantics and information structure. In G. Ramchand & C. Reiss (Eds.), *The Oxford Handbook of Linguistic Interfaces*, pp. 445-474. Oxford University Press.
- Carman, K. L., Dardess P., Maurer, M., Sofaer, S., Adams, K., Bechtel, C. & Sweeney, S. (2013). Patient and family engagement: A framework for understanding the elements and developing interventions and policies. *Health Affairs*, 32(2), 223-231. <https://doi.org/10.1377/hlthaff.2012.1133>
- Carrell, P. (1982). Cohesion is not coherence. *TESOL Quarterly*, 16(4), 479-488. <https://doi.org/10.2307/3586466>.
- Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics and point of view. In C. N. Li (Ed.), *Subject and Topic* (pp. 25-55). Associated Press.
- Charolles, M. (1997). L'encadrement du discours: univers, champs, domaines et espaces. *Cahier de Recherche Linguistique*, 6, pp. 1-73. Université de Nancy 2.
- European Medicines Agency. (2006). Reflection paper – EPAR summary for the public. <<https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/>

- reflection-paper-european-public-assessment-report-summary-public_en.pdf> [10/02/2025].
- European Medicines Agency. (2017). EPAR summary template. <https://www.ema.europa.eu/en/documents/presentation/presentation-epar-summary-template-e-scanlan_en.pdf> [10/02/2025].
- European Parliament/European Council. (2004). Council Regulation (EC) No. 726/2004. <<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32004R0726>> [12/12/2023].
- Firbas, J. (1974). Some aspects of the Czechoslovak approach to problems of functional sentence perspective. In F. Danes (Ed.), *Papers on functional sentence perspective* (pp. 11-37). Academia.
- Fries, P. H. (1983). On the status of Theme in English: Arguments from discourse. In J. S. Petöfi & E. Sözer (Eds.), *Micro and macro connexity of texts* (pp. 116-152). Helmut Buske.
- Givón, T. (1995). Coherence in text vs. coherence in mind. In M. A. Gernsbacher & T. Givón (Eds.), *Coherence in spontaneous text* (pp. 59-116). Benjamins.
- Gundel, J., & Fretheim, T. (2004). Topic and focus. In L. R. Horn & G. Ward (Eds.), *The Handbook of Pragmatics* (pp. 175-196). Blackwell Publishing.
- Halliday, M. A. K. (1967). Notes on transitivity and theme in English: Part 2. *Journal of Linguistics*, 3(2), 199-244. <https://doi.org/10.1017/S0022226700016613>
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Longman.
- Halliday, M. A. K., & Matthiessen, C. M. I. M. (1999). *Construing experience through meaning: A language-based approach to cognition*. Cassell.
- Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional grammar*. Routledge.
- Hasan, R. (1984). Coherence and cohesive harmony. In J. Flood (Ed.), *Understanding reading comprehension: Cognition, language and the structure of prose* (pp. 181-219). International Reading Association.
- Hasan, R. (1989). The texture of a text. In M. A. K. Halliday & R. Hasan (Eds.), *Language, context and text: Aspects of language in a social-semiotic perspective* (pp. 70-96). Oxford University Press.
- Hidi, S., & Anderson, V. (1986). Producing written summaries: Task demands, cognitive operations, and implications for instruction. *Review of Educational Research*, 56(4), 473-493. <https://doi.org/10.3102/00346543056004473>
- Hill-Madsen, A. (2015). Lexical Strategies in Intralingual Translation between Registers. *HERMES – Journal of Language and Communication in Business*, 54, 85-105. <https://doi.org/10.7146/hjlc.v27i54.22949>

- Hill-Madsen, A. (2019). Coherence in hybrid texts: The case of the Patient Information Leaflet genre. *HERMES – Journal of Language and Communication in Business*, 59, 153-187. <https://doi.org/10.7146/hjlc.v59i1.117038>
- Hill-Madsen, A. (2022b). Transformational strategies in diaphasic translation: three case studies. *Perspectives – Studies in Translation Theory and Practice*, 30(4), 643-661. <https://doi.org/10.1080/0907676X.2021.1953085>
- Hill-Madsen, A. (2022a). Semantic density and gravity in lay-oriented medical knowledge communication: The case of the EPAR Summary for the public. *Nordic Journal of English Studies*, 21(2), 195-225. <https://doi.org/10.35360/njes.781>
- Hill-Madsen, A. (2024). Semantic density and gravity in lay-oriented medical knowledge communication II: The case of Patient Information Leaflets. *Nordic Journal of English Studies*, 24(1), 41-74. <https://doi.org/10.35360/njes.v24i1.41116>
- Hill-Madsen, A. (in press). 'Simplification' and 'accessibility' in diaphasic intralingual translation: A semantic interpretation. In H. Karas & H. B. Zeev Shyldkrot (Eds.), *Intralingual translation: Language, text and beyond*. John Benjamins.
- Hill-Madsen, A. & Dam-Jensen, H. (2022). Coherence in non-technical summaries of European Central Bank financial reports. *Fachsprache – Journal of Professional and Scientific Communication*, 44(1-2), 80-101. <https://doi.org/10.24989/fs.v44i1-2.1856>
- Hill-Madsen, A., & Pilegaard, M. (2019). Variable scope for popularization of specialized terminology: The case of medico-pharmaceutical terms. *Fachsprache – Journal of Specialized Communication*, 41(1-2), 22-40. <https://doi.org/10.24989/fs.v41i1-2.1623>
- Hoover, M. L. (1997). Effects of textual and cohesive structure on discourse processing. *Discourse Processes*, 23(2), 193-220. <https://doi.org/10.1080/01638537709544989>
- Kleiber, G. (1999). Associative anaphora and part-whole relationships: The condition of alienation and the principle of ontological congruence. *Journal of Pragmatics*, 31(3), 339-362. [https://doi.org/10.1016/S0378-2166\(98\)00072-1](https://doi.org/10.1016/S0378-2166(98)00072-1)
- Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica*, 55(3-4), 243-276. <https://doi.org/10.1556/aling.55.2008.3-4.2>
- Lambrecht, K. (1994). *Information structure and sentence form: topic, focus, and the mental representations of discourse referents*. Cambridge University Press.
- Lyons, J. (1977). *Semantics*. Cambridge University Press.
- Martin, J. R. (1992). *English text: System and structure*. Benjamins.
- Menzel, K., Lapshinova-Koltunski, E., & Kunz, K. (2017). Cohesion and coherence in multilingual contexts. In E. Lapshinova-Koltunski & K. Menzel (Eds.), *New perspectives on cohesion and coherence: Implications for translation* (pp. 1-11). Language Science Press.
- Myers, G. (1991). Lexical cohesion and specialized knowledge in science and popular

- science texts. *Discourse Processes* 14(1), 1-26.
<https://doi.org/10.1080/01638539109544772>
- Parsons, G. (1991). Cohesion coherence: scientific texts. In E. Ventola (Ed.), *Functional and systemic linguistics: Approaches and uses* (pp. 415-430). De Gruyter.
- Patton, M. Q. (2002). *Qualitative methods and research evaluation*, 3rd ed. SAGE.
- Plainlanguage.gov (2011). Federal Plain Language Guidelines.
<https://plainlanguage.gov/media/FederalPLGuidelines.pdf> [12/12/2023].
- Porta, M. and Last, J. M. (2018). Anatomical therapeutic chemical classification (ATC). In A Dictionary of public health. 2nd ed.
<https://www.oxfordreference.com/view/10.1093/acref/9780191844386.001.0001/acref-9780191844386-e-187?rskey=edQ4wk&result=2> [12/12/2023].
- Raynor, D. K., & Bryant, D. (2013). European Public Assessment Report (EPAR) Summaries for the Public: Are they Fit for Purpose? A User-Testing Study. *British Medical Journal Open*, 3(9), 1-17. <https://doi.org/10.1136/bmjopen-2013-003185>
- Sanders, T., & Sanders, J. (2006). Text and Text analysis. In K. Brown (Ed.), *Encyclopedia of Language and Linguistics* (pp. 597-607). Elsevier.
- Sanders, T. J. M., Spooren, W. T. M., & Noordman, L. G. M. (1993). Coherence relations in a cognitive theory of discourse representation. *Cognitive Linguistics*, 4, 93-133.
<https://doi.org/10.1515/cogl.1993.4.2.93>
- Sanford, A. (2006). Coherence: Psycholinguistic approach. In K. Brown (Ed.), *Encyclopedia of Language and Linguistics* (pp. 585-591). Elsevier.
- Seidlhofer, B., & Widdowson, H. (1999). Coherence in summary: The contexts of appropriate discourse. In W. Bublitz et al. (Eds.), *Coherence in spoken and written discourse: How to create it and how to describe it. Selected papers from the International Workshop on Coherence, Augsburg, 24-27 April 1997* (pp. 205-219). Benjamins.
- Taboada, M. T. (2004). *Building coherence and cohesion: Task-oriented dialogue in English and Spanish*. Benjamins.
- Tanskanen, S.-K. (2006). *Collaborating towards coherence: Lexical cohesion in English*. Benjamins.
- Thompson, S. E. (1994). Aspects of cohesion in monologue. *Applied Linguistics*, 15(1), 58-75. <https://doi.org/10.1093/applin/15.1.58>

Appendix A: Bibliography of corpus texts

- European Medicines Agency (2022). Sunlenca – EPAR: Medicine overview.
<https://www.ema.europa.eu/en/documents/overview/sunlenca-epar-medicine->

overview_en.pdf.> [19/01/2024].

European Medicines Agency (2022). Tezspire – EPAR: Medicine overview. <https://www.ema.europa.eu/en/documents/overview/tezspire-epar-medicine-overview_en.pdf> [19/01/2024].

European Medicines Agency (2022). Ximluci – EPAR: Medicine overview. <https://www.ema.europa.eu/en/documents/overview/ximluci-epar-medicine-overview_en.pdf> [19/01/2024].

European Medicines Agency (2023): Hemgenix – EPAR: Medicine overview. <https://www.ema.europa.eu/en/documents/overview/hemgenix-epar-medicine-overview_en.pdf> [19/01/2024].

European Medicines Agency (2023). Omvoh – EPAR: Medicine overview. <https://www.ema.europa.eu/en/documents/overview/omvoh-epar-medicine-overview_en.pdf> [19/01/2024].

European Medicines Agency (2023). Opzelura – EPAR: Medicine overview. <https://www.ema.europa.eu/en/documents/overview/opzelura-epar-medicine-overview_en.pdf> [19/01/2024].

Appendix B: The *Sunlenca* EPAR summary

[1] What is Sunlenca and what is it used for?

[2] Sunlenca is used, together with other medicines, to treat adults infected with human immunodeficiency virus type 1 (HIV-1), a virus that causes acquired immune deficiency syndrome (AIDS). [3] Sunlenca is given when the virus is resistant to other treatments.

[4] Sunlenca contains the active substance lenacapavir.

[5] How is Sunlenca used?

[6] Sunlenca is available as tablets to be taken by mouth and as a solution for injection. [7] Sunlenca tablets are taken at the start of the treatment, on days 1, 2 and 8. [8] One week after that, patients are given Sunlenca injections every 26 weeks as maintenance treatment. [9] Injections are given under the skin by a doctor or nurse.

[10] Before starting treatment, the doctor must ensure that the patient agrees to keep to the schedule of injections and should explain why this is important. [11] The treatment schedule helps keep the virus under control. [12] If a patient misses

treatment doses, virus levels may increase, or the virus may become resistant to treatment. [13] If treatment with Sunlenca is stopped, another treatment to suppress the virus must be started.

[14] Sunlenca can only be obtained with a prescription and should be prescribed by a doctor who has experience in the management of HIV infection.

[15] For more information about using Sunlenca, including the schedule for the injections, see the package leaflet or contact your doctor or pharmacist.

[16] How does Sunlenca work?

[17] The active substance in Sunlenca, lenacapavir, is a substance that binds to the proteins that make up the outer layer of the HIV-1 virus (the capsid). [18] By binding to these proteins, Sunlenca interferes with different steps that are necessary for the virus to multiply. [19] This reduces the amount of HIV in the blood and keeps it at a low level. [20] Sunlenca does not cure HIV infection or AIDS, but it can hold off damage to the immune system and the development of infections and diseases associated with AIDS.

[21] What benefits of Sunlenca have been shown in studies?

[22] Sunlenca, taken together with other treatments to control HIV-1 infection, was effective at reducing the amount of HIV-1 virus in the blood (viral load) in one main study involving adults who had already tried other treatments and who did not respond or were no longer responding to most of the medicines used to control HIV-1 infection. [23] In the first two weeks of the study, patients were given Sunlenca or placebo (a dummy treatment) in addition to their usual HIV medicines. [24] After this time, 87.5% (21 out of 24) of the participants who were given Sunlenca showed a meaningful decrease in viral load, compared with 16.7% (2 out of 12) of the participants who were given a placebo. [25] The 12 patients first given placebo then also received Sunlenca, and all 36 patients were given maintenance injections every 26 weeks. [26] Viral load was under 50 copies of the virus per mL (which is a threshold considered indicative of durable clinical and immunological benefits) in 80.6% (29 out of 36) of patients after 26 weeks, and in 83.3% (30 out of 36) of patients after 52 weeks of treatment.

[27] What are the risks associated with Sunlenca?

[28] The most common side effects with Sunlenca (which may affect more than 1 in 100 people) are reactions at the injection site and nausea.

[29] For the full list of side effects of Sunlenca, see the package leaflet.

[30] Why is Sunlenca authorised in the EU?

[31] Sunlenca is a new type of treatment for controlling HIV-1 infection, which was effective at reducing viral load in patients with infections that are resistant to other treatments. [32] These patients often lack options to manage their infection, and Sunlenca is considered to address an unmet medical need for this population. [33] The side effects of Sunlenca are considered manageable.

[34] Therefore, the European Medicines Agency decided that Sunlenca's benefits are greater than its risks and it can be authorised for use in the EU.

[35] What measures are being taken to ensure the safe and effective use of Sunlenca?

[36] Recommendations and precautions to be followed by healthcare professionals and patients for the safe and effective use of Sunlenca have been included in the summary of product characteristics and the package leaflet.

[37] As for all medicines, data on the use of Sunlenca are continuously monitored. [38] Suspected side effects reported with Sunlenca are carefully evaluated and any necessary action taken to protect patients.

[39] Other information about Sunlenca

[40] Further information on Sunlenca can be found on the Agency's website: ema.europa.eu/medicines/human/EPAR/Sunlenca.

Appendix C: Chains with tokens in the *Sunlenca* text

'Action': *measures, action*

'Active substance': *active substance, lenacapavir, active substance, lenacapavir*

'Administration': *given, taken, taken, given, given, taken, given, given, given, given, received, given,*

'Adults': *adults, adults, who, who* [lexified as *adults*]

'Amount': *amount, amount, load, load, load*

'Association': *associated, associated*

'Authorisation': *authorised, authorised*

'Being': *is, is, is, is, is, are, was, was, are, are, is, is, was, are, are*

'Binding': *binds, binding*

'Body parts': *mouth, skin, blood*

'Considering': *considered, considered, considered*

'Control': *control, management, control, control, controlling, manage*

'Disease': *acquired immune deficiency syndrome (AIDS), AIDS, diseases, AIDS, nausea*

'Documents': *package leaflet, package leaflet, summary of product characteristics, package leaflet*

'Effects': *benefits, effective, benefits, risks, side effects, which [lexified as side effects], side effects, effective, side effects, benefits, risks, effective, effective, side effects,*

'EMA': *European Medicines Agency, Agency's*

'EU': *EU, EU,*

'HCP [= health care professionals]': *doctor, nurse, doctor, doctor, who [lexified as doctor], doctor, pharmacist, healthcare professionals*

'Increase/decrease': *increase, reduces, reducing, decrease, reducing*

'Infection': *infected, infection, infection, infections, infection, infection, infection, infections, that [lexified as infection], infection*

'Information': *information, information, information*

'Injection': *injection, injections, injections, injections, injections, injections, injection*

'Keeping': *keep, keep, keeps*

'Level': *levels, level*

'Layer': *layer, capsid*

'Maintenance': *maintenance, maintenance*

'Medicine': *Sunlenca, it [lexified as Sunlenca], medicines, Sunlenca, Sunlenca, Sunlenca, Sunlenca, Sunlenca, doses, Sunlenca, Sunlenca, Sunlenca, Sunlenca, Sunlenca, Sunlenca, it [lexified as Sunlenca], Sunlenca, medicines, Sunlenca, medicines, Sunlenca, medicines, Sunlenca, Sunlenca, Sunlenca, Sunlenca, Sunlenca, Sunlenca, Sunlenca, Sunlenca, Sunlenca,*

Sunlenca, medical, Sunlenca, Sunlenca, its [lexified as Sunlenca], it [lexified as Sunlenca], Sunlenca, Sunlenca, medicines, Sunlenca, Sunlenca, Sunlenca, Sunlenca

'Patients': patients, patient, patient, patients, participants, who [lexified as participants], participants, who [lexified as participants], patients, patients, patients, patients, people, patients, patients, their [lexified as patients], population, patients, patients

'Placebo': placebo, placebo, placebo

'Prescription': prescription, prescribed

'Proteins': proteins, that [lexified as proteins], proteins

'Responding': respond, responding

'Resistant': resistant, resistant, resistant

'Safety': safe, safe

'Schedule': schedule, schedule, schedule

'Seeing': see, see

'Starting/stopping': start, starting, stopped, started,

'Study': studies, study, study

'Taking': taken, taken

'Time periods': days, week, weeks, weeks, time, weeks, weeks, weeks

'Treatment': treat, treatments, treatment, treatment, treatment, treatment, treatment, treatment, treatment, treatment, cure, treatments, treatments, treatment, treatment, treatment, treatments

'Use': used, used, used, using, used, use, use, use

'Virus': virus, virus, virus, virus, virus, virus, virus, HIV, HIV-1 virus, virus, HIV, it [lexified as HIV], HIV, HIV-1 virus, viral, HIV-1, HIV, viral, virus, viral, virus, HIV-1, viral

'Tablets': tablets, tablets

Appendix D: Experiential clause configurations according to

Systemic-Functional Grammar

Below, the so-called experiential clause grammar according to SFG (see Halliday & Matthiessen, 2014, ch. 5) is outlined in a somewhat simplified version. This part of the grammar consists of six different clause types, each representing a certain configuration of process (realized in the verb) and participants (corresponding to *arguments* or *semantic roles* in other conceptualizations). The six clause types are:

- **Material clauses** (clauses representing 'doings and happenings'). Primary participants: Actor and Goal. Example: *The dog* (Actor) *chased* (Process) *the postman* (Goal).
- **Behavioural clauses** (clauses representing 'behaving'). Participant: Behaver. Example: *The old man* (Behaver) *sneezed* (Process).
- **Mental clauses** (clauses representing processes of perception and cognition). Primary participants: Senser and Phenomenon. Example: *We* (Senser) *couldn't see* (Process) *the sun* (Phenomenon).
- **Verbal clauses** (clauses representing processes of saying). Primary participants: Sayer, Verbiage and Receiver. Example: *The old man* (Sayer) *told* (Process) *his grandchild* (Receiver) *a story* (Verbiage).
- **Relational clauses** (processes of being and having). Primary participants: Carrier and Attribute. [Example: *The old man* (Carrier) *was* (Process) *a good story-teller* (Attribute).]
- **Existential clauses**. Participant: Existant. [Example: *There is* (Process) *a lot of snow* (Existant) *outside*.