Fragmented but coherent: Lexical cohesion on a YouTube channel

Sanna-Kaisa Tanskanen

Department of Languages, University of Helsinki, P.O. Box 24, 00014 Helsinki, Finland

ARTICLE INFO

ABSTRACT

This paper investigates the use and functions of lexical cohesion within and across modes as well as levels of interaction on a YouTube channel, arguing that lexical cohesion contributes to coherence by establishing links between the video, the comments and sources elsewhere on the internet in a cross-modal and intertextual manner. The analysis focuses on one video and its comments section, investigating the use of cohesion in maintaining coherence and unity across a lengthy YouTube interaction initiated by the video and continued in the comments. Cohesive ties are examined with regard to their functions in the process of meaning-making in this mediated and multimodal context. The results reveal that the use of cohesive ties is extensive within and across modes and levels of interaction. Chains of cohesion are created within the comments, with cross-modal links to the video. Despite the complex multimodality of YouTube and the fragmentation of the comments as many participants post short comments one after another, the YouTube interaction is shown to be coherent.

1. Introduction

While the use of cohesion has been analysed in other, more text-based forms of digital discourse (e.g. in weblogs by Hoffmann 2012 and in mailing-list discourse by (Tanskanen, 2006)), studies of cohesion on YouTube have been scarce. The few existing studies include Bou-Franch et al. (2012), Liebschner (2021) and Schubert (2017). With its combination of a video and a discussion made up of comments posted by users who have viewed the video, YouTube interaction lends itself well to an analysis of cohesion from multi-modal and cross-modal perspectives, as the interaction takes place on different levels in any modal form (Benson 2016; Schmidt & Marx 2019). Cohesion is sensitive to context, i.e. its use varies according to the parameters of the communicative context, and investigating cohesion in the context of a YouTube channel can shed light on how it works across different modes and levels of interaction, and how it helps maintain coherence and unity in this specific context.

This paper investigates the use and functions of lexical cohesion within and across modes on a YouTube channel, arguing that lexical cohesion contributes to coherence by establishing links between the video and the comments as well as between the comments in a cross-modal and intertextual manner. Cohesive ties are therefore not analysed as purely linguistic features but are

examined with regard to their functions in the process of meaning-making in this mediated and multimodal context (Van Leeuwen 2005). Furthermore, the paper answers the call by Androutsopoulos and Tereick (2015) to investigate the use of cohesion in maintaining coherence and unity across a lengthy YouTube interaction.

The YouTube channel analysed is *The Hydraulic Press Channel*, which publishes videos of various objects being crushed in a hydraulic press, with a commentary by the YouTuber, Lauri Vuo-hensilta. One video with comments posted by viewers was selected for close analysis. In Adami's (2014: 240; 2015: 235) words, the video is "a prompt" to which viewers react with their comments (see also Johansson 2017). It is therefore not surprising that cohe-sive ties should exist between the video and the individual comments; at the same time, however, links are created between the comments. A particular challenge in the study of YouTube cohesion and coherence is the fact that the structure as well as the partici-pation framework of a YouTube discussion can appear fragmented, as many participants post short contributions one after another (Boyd 2014; Dynel 2014; Marcoccia 2004). The analysis shows how ties formed by lexical cohesion help offset the fragmentation and create a coherent discussion. The research questions set for the analysis are the following:

How does lexical cohesion contribute to the unity and coher-ence of a YouTube channel?

E-mail address: sanna-kaisa.tanskanen@helsinki.fi

More precisely, how do lexical cohesive devices establish links within and across the separate but interrelated levels of interaction of a YouTube channel, i.e. the video clip and the comments; and how do the devices operate across the entire lengthy interaction?

I begin the paper by first briefly reviewing previous research on the social media platform selected for the present study, i.e. YouTube, and continue with a discussion on lexical cohesion in both offline and online contexts. In section 3, I introduce the material and methods used for the analysis. I then discuss, in section 4, the results of the analysis, which throw light on how lexical cohesion creates coherence in the multimodal context of a YouTube channel. Conclusions are offered in section 5.

2. Background

2.1. YouTube

YouTube is an online video-sharing platform launched in 2005. It allows its users to upload, view, share and comment on videos. Users can subscribe to the channels of other users and receive notifications of new video uploads, the number of subscriptions being the main indication of popularity. The combination of user-generated videos and the viewers' discussion of them through their comments creates a multimodal context: the video is audio-visual and the comments are text-based. Furthermore, the multimodal layers of YouTube encompass social media operators, commercials, and suggestions for further YouTube videos (Johansson 2017; Schmidt & Marx 2019: 117). In Benson's words, on its launch "YouTube represented a high point of multimodality in web-based computer-mediated communication that is, arguably, yet to be surpassed" (Benson 2016: Ch. 3).

Previous linguistic studies on YouTube have examined, for instance, how YouTubers construct their identities and engage in identity play (Senft 2008; Leppänen & Häkkinen 2012). The content of YouTube comments has been the subject of research as well. For instance, Jones and Schieffelin (2009) draw attention to the redundancy of YouTube comments: in their material, similar queries and opinions recur, which leads them to conclude that commenters tend to read only the most recent comments. Bou-Franch and Garcés-Conejos Blitvich (2014) show how conflict in YouTube multi-party interaction (or polylogue) evolves and is resolved, most often by withdrawal.

The participation framework of YouTube interaction has also been investigated (Boyd 2014; Dynel 2014; Schmidt & Marx 2019). Participation frameworks refer to the relations which participants establish in and through their communication (Goffman 1981), and they are affected by situational and technological parameters. According to Dynel (2014), there are three basic levels of interaction in YouTube communication: (1) the video interaction, (2) the sender-recipient interaction, and (3) the comments, each with their own forms of interaction and participant roles. Drawing on Dynel's participation framework, Herring and Chae (2021) analyse types of addressee and content of messages in three YouTube comment threads, showing that there is an interaction between topic and addressee type in the threads.

Finally, as for the interaction between YouTubers, Burgess and Green (2009) note that YouTubers frequently ask for feedback of any kind on their videos. Frobenius (2014: 70) argues that although there is no immediate audience participation in YouTube videos and the speakers/senders have to address imagined viewers/recipients, they actively try to involve the audience. The viewers may then respond in written form in their comments, and these exchanges bear resemblance to face-to-face conversation (see also Schmidt & Marx 2019: 125–126).

2.2. Lexical cohesion

Halliday and Hasan introduced the concept of cohesion in 1976. They use it to refer to relations of meaning that exist within a text and that define it as a text (Halliday & Hasan 1976: 4). The definition is semantic, and like all components of the semantic system, cohesion is realised through grammar and vocabulary. Accordingly, cohesive devices are traditionally divided into grammatical and lexical ones. Grammatical cohesion includes devices such as reference, substitution, ellipsis and conjunction, while lexical cohesion includes relations between lexical items, through either reiteration or collocation (habitual co-occurrence). In some of the more recent models of cohesion, the strict division into grammatical and lexical devices has been abandoned in favour of a view highlighting the similar functions of, for instance, repetitions of pronouns and nouns (see e.g. Hoey 1991; Tanskanen 2006; Hoffmann 2012). This is the view adopted in the present study as well, but since not all grammatical cohesive devices can be subsumed under lexical cohesion on the basis of their functions in text (e.g. conjunction), it seems appropriate to refer to "lexical cohesive devices" when referring to the items studied here. It could be mentioned at this point that even in those studies where the division is retained, the number and role of lexical cohesive devices tend to be clearly more significant than those of grammatical cohesive devices.

The view of cohesion in this paper is that while they are neither sufficient nor necessary for the perceived coherence of discourse, cohesive devices contribute to coherence, and communicators can use them to signal the relatedness within and between their contributions (Tanskanen 2006: 22–27). Previous research has established that lexical cohesion is sensitive to context, as there are differences in the use of lexical cohesion between texts produced in different contexts, both offline and online. Face-to-face conversations, for instance, show high frequencies of reiteration, while academic writing shows a relatively low frequency; collocation is especially rare in multi-party conversations (Tanskanen 2006: 166). In online contexts, mailing list discussions showed medium to high frequencies for reiteration and high frequencies for collocation (Tanskanen 2006: 166), whereas blog comments showed high frequencies for both reiteration and collocation (Hoffmann 2012: 171–172).

How cohesive devices function across different semiotic modes in multimodal contexts has also been investigated. Building on Halliday and Hasan's work, Royce (2007) suggests that the sense relations of reiteration and collocation also work intersemiotically (i.e. in a cross-modal manner), creating links between verbal and visual modes. The sense relations range from repetition (identical experiental meaning represented in verbal and visual modes) to collocation (an expectancy or high probability to co-occur) (Royce 2007: Table 2.1). Liu and O'Halloran (2009) further develop the framework of intersemiotic cohesive devices operating between language and images, arguing that the devices "generate semantic ties between linguistic and pictorial components, and thus integrate different modalities together into a coherent product" (Liu & O'Halloran 2009: 385).

Of the studies which have so far explored the use of cohesion on YouTube, Schubert 2017 and Liebschner 2021 take into account cross-modal cohesive links between the video and the comments. In his pilot study, Schubert (2017) discusses how cohesion creates links within and across the video, the comments and other thematically related videos. Liebschner's analysis of a Russian YouTube channel operating in English is informed by Royce (2007) and thus also looks at relations between verbal and visual modes. The results show that reiteration is overwhelmingly more frequent than collocation. Most of the cohesive links occur between comments, but there are also links which connect the visual or spoken content of the video with the comments (Liebschner 2021: 59–60).

In Bou-Franch et al.'s study on Spanish YouTube interaction, the focus of the analysis of cohesion is on cross-turn cohesion, i.e. cohesion between text-based comments. The analysis reveals an overall reliance on lexical cohesion in the comments (the study does not report the numbers for reiteration and collocation separately). The researchers conclude that despite massive and fluid participation, YouTubers manage to create collaborative and coherent interaction (Bou-Franch et al. 2012: 515).

For a while, i.e. until 2013, YouTubers had the option of responding to a video with a video, and relatedness and cohesion between these has been studied by Adami. Interestingly, she found that the responses can display visual repetition with the prompting video, so that, for instance, the blinking of the eyes by the sender of a video is repeated by the sender of a response video (Adami 2014: 244–246). All in all, the video responses studied by Adami ranged from fully related and cohesive to incoherent and unrelated (Adami 2014; 2015).

To conclude, a final point about cohesion in the online context. Eisenlauer (2013) provides a taxonomy for cohesion across the different levels of online interaction, which he calls nodes. In his model, *intranodal* cohesion operates within one node (e.g. YouTube comments), *internodal* within one "hypertextual database" (e.g. between the video and the comments on a YouTube channel) and *extranodal* across "different hypertextual databases" (e.g. between the comments and material elsewhere on the internet). This taxonomy offers a useful additional basis for the discussion of cohesive ties across the levels of YouTube interaction.

3. Material and methods

3.1. Material

The material for the present study comes from the *Hydraulic Press Channel* (HPC) on YouTube, which publishes videos of various objects being crushed in a hydraulic press. A Finnish couple run the channel, but the language used in the videos as well as in most of the comments is English. The videos include commentary by the YouTuber, Lauri Vuohensilta, as well as reactions to the crushing by his wife, Anni. The channel was created in 2015 and became popular in 2016 after one of the videos was shared on Reddit. The channel currently has more than 2.8 million subscribers and well over 400 million views. The viewers actively write comments on the videos and suggest new items to be crushed. The combination of action and commentary by the YouTuber in the videos as well as the active commenting by viewers made the channel seem befitting for an analysis from multimodal and cross-modal perspectives.

According to the Wikipedia page on the channel, the YouTuber himself thinks that the success of the channel is due to a combination of the unexpected results of the crushing and "the humour value of everything, my accent and stupid jokes" (en.wikipedia.org/wiki/Hydraulic_Press_Channel). Vuohensilta comments on the crushing in a very matter-of-fact way in his Finnish accent, which sometimes becomes the topic in the viewers' comments. Anni Vuohensilta reacts to the action, most often with laughter. She also makes small clay figures, which are crushed as bonus clips.

The video chosen for analysis in this paper is among the earlier ones on the channel, published in 2016, and also among the twenty most popular on the channel. The video shows the crushing of a legendary Nokia mobile phone model, namely the 3310 (Fig. 1). The video was selected because it is old enough so that a good number of comments had accumulated, but the number of comments was not overwhelming for the purposes of the analysis.

The screenshot in Fig. 1 shows the opening frame of the video, which has so far been viewed over five million times. Of these

viewers, 39,000 have indicated that they like the video, whereas a thousand viewers have wanted to record their dislike. All viewers are members of the video's audience, regardless of whether they actively record their like/dislike, comment on the video, or just view it and/or the comments. Below this information is the logo and name of the channel, together with the current number of subscribers and a subscribe button. At the bottom of the screenshot, we see the number of comments (6,506 at the time of the screenshot), which can be sorted either chronologically starting with the newest or according to the top-comments algorithm (which takes into account, among other factors, whether the comment was posted by a subscriber and what its like/dislike ratio is). Of all the viewers, the commenters take a more active role, as they comment on the video, sometimes addressing the sender or other commenters directly.

The sorting and ordering of comments determines how commenters see the other comments, and this is relevant especially if we take into account the point made by Jones and Schieffelin (2009: 1063) that commenters tend to read the most recent comments only. To illustrate how different or similar the result is depending on whether the top comments or the newest comments are selected, Appendices D and E provide screenshots of the first HPC Nokia 3310 comments organised according to these two sorting orders. The top comments have received more replies (quite naturally since they are all older comments), but the contents in the top and newest comments are very similar. If they act as prompts to new comments, it is hard to see what the differences caused by this could be.

Collecting material from a dynamic platform such as YouTube raises the methodological question of "freezing" the material (Androutsopoulos & Tereick 2015: 357; Schmidt & Marx 2019: 138), and this applies to the HPC Nokia video as well. We will be able to analyse the comments contributed up to the point of data collection, but because the video can still be viewed and commented on at any time, the number of comments is bound to be accumulating further.

The comments were collected and ordered chronologically for the analysis. The collecting was carried out with a web-scraping script written in the Python programming language for this purpose.[1] Using the YouTube API, the script retrieved all the comments, ordered them chronologically and separated sequences of comments (i.e. the initiating comment and replies to it) from individual comments so that it was possible to see which comments belong to the same sequence.

The majority of the commenters follow the decision by Lauri Vuohensilta to use English, but there are several comments in other languages as well, which suggests an international following. It is hardly surprising that some comments are in Finnish, although most viewers with Finnish-sounding usernames post in English. There are also comments in Russian, and occasional ones in Arabic, Chinese, French and Japanese, among others. Portuguese is an intriguing case: in 2018, i.e. two years after the publication of the video, there is a sudden influx of comments in Portuguese. Many of these mention the name Felipe Neto, who is an extremely popular Brazilian YouTuber. Visiting Neto's channel and searching for Nokia 3310 reveals that he has embedded a part of the HPC video into his own video and comments on it. Apparently inspired by the video, some of his viewers decide to come and view the original and also leave comments. Such circulation of videos and their promotion by other YouTubers provides yet another level of interaction in YouTube's multimodal network. Sometimes a shared

---

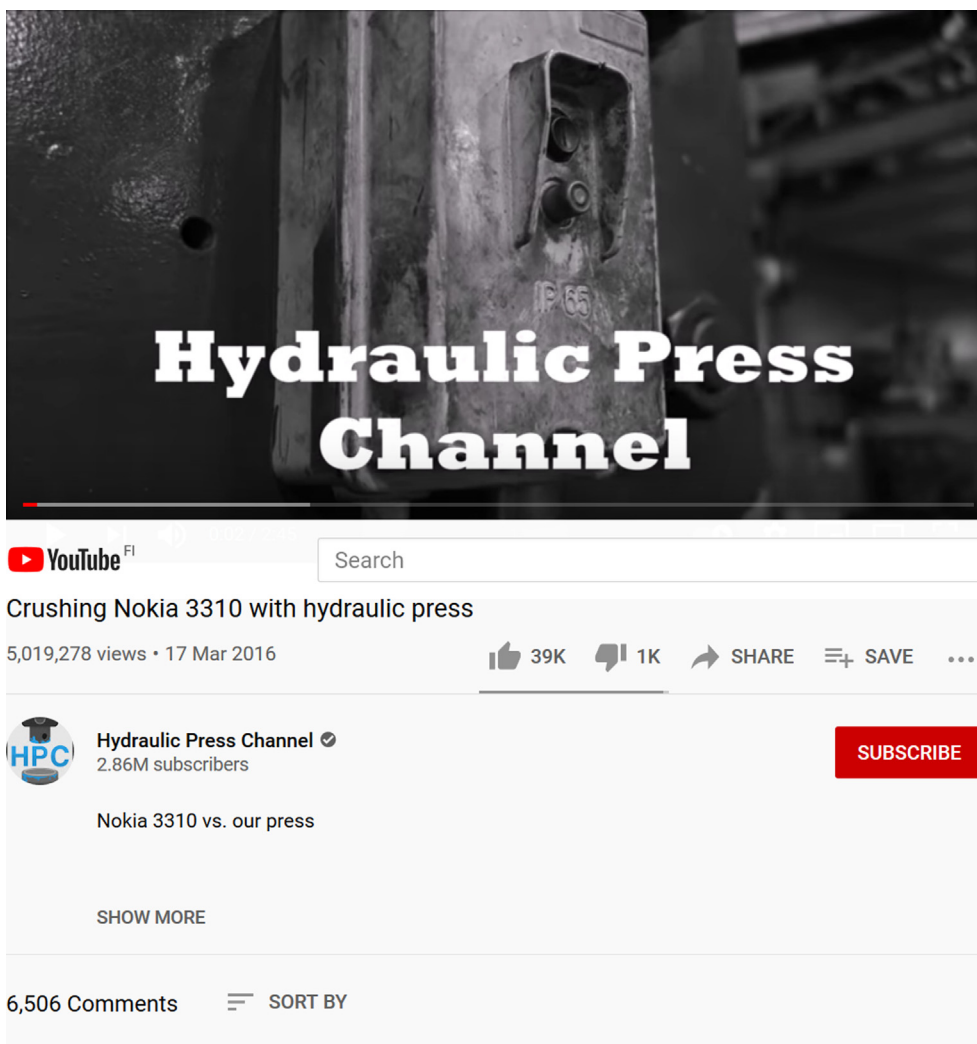[1] I thank my research assistant Jonas Haverinen for his invaluable help with the script.

Fig. 1. Screenshot of the beginning of the video (link to the video). Permission to include the link has been granted by Lauri Vuohensilta.

video can even go viral and enjoy enormous popularity, as more and more people share it and talk about it.

3.2. Methods

The objective of the study is to investigate lexical cohesion in the multimodal context of a YouTube channel. The analysis therefore focuses on the cohesive relations that participants create with their lexical choices, but it also recognises cross-modal (or inter-semiotic) links between verbal and visual modes. The classification of lexical cohesive devices follows that of Tanskanen (2006). Lexical cohesion is divided into reiteration and collocation (Table 1):

The video commentary was transcribed, and cohesive ties within the commentary as well as between the verbal commentary and the visual content of the video were analysed. The comments file was exported to AntConc (Anthony 2019) and MSExcel. The various functions of AntConc (word list, concordance, concordance plot and N-grams)[2] were used to explore the material, while the analytical coding was carried out in Excel. For the detailed analysis of lexical cohesion in the comments, the decision was made to study

Table 1
Classification of lexical cohesive devices.

Reiteration	Collocation
simple repetition (*phone – phone*)	ordered set (*today – tomorrow*)
complex repetition (*press (N) – press (V)*)	activity-related collocation (*crush – hydraulic press*)
substitution (*phone – it*)	elaborative collocation (*phone – battery*)
equivalence (*crush – smash*)	
generalisation (*Nokia 3310 – phone*)	
specification (*phone – Sony*)	
co-specification (*Nokia – Sony*)	
contrast (*real – fake*)	

the first 100 and the last 100 comments as well as 100 comments from the middle of the discussion (i.e. 300 comments in total). Cohesive ties in the comments, together with cross-modal ties with the video, were analysed. Additional sequences where a comment receives replies were analysed for their cohesion in order to find out if cohesion in such sequences differs from cohesion in strings of individual comments. Finally, chains of lexical cohesion were investigated, because they (especially long chains) have been found to be important for the overall coherence of a text (Hasan 1984; Hoey 1991; Morris & Hirst 1991; Parsons 1991; Tanskanen 2006).

---

[2] A concordance line shows the target word in its context (i.e. with the words surrounding it); a concordance plot shows the distribution of the target word in the file in a barcode format; and an N-gram is a sequence of words.

4. Results and discussion

4.1. Lexical cohesion in the video interaction

The results are presented according to the levels of interaction on YouTube, starting with the video interaction (which is simultaneously the sender-recipient level, since there are no other senders). The Nokia 3310 video is 2 min 46 s long. Vuohensilta welcomes the viewers and then proceeds to crush the phone. Below is a full transcription of his commentary, together with descriptions of the action in the video:

> Welcome to the Hydraulic Press Channel. Today we have legendary Nokia 3310 for our press. [places phone in press] It's quite tough phone but I'm quite confident that our press can handle it. [crushing; real life and slow motion] I think that the press wins and the phone lose. Maybe we will try one more time for the bottom part of the phone. The back cover is somehow glued to the bottom [of the press]. I think I just press there one more time. [crushing] There. It was full power against Nokia and for Nokia's glory I must say it still looks like a phone. [takes phone and turns it around in his hand] It's quite thin and quite warm, it's heated up while pressing. Yeah, that's all for today. Thank you all for watching and have a nice day.

The video then continues with text asking viewers for ideas: "If you have good ideas about stuff to crush, please write them to comments!". At the end of the video, there is a bonus clip of the crushing of a clay figure, commented on by Vuohensilta:

> "And for the extra present for our new viewers my wife has made this magnificent clay fox and now we will see what does the fox say under the press. Eeeergh [crushes the fox and pretends to screech on its behalf]".

We never actually see Vuohensilta or his wife on this video, only his hand as he places the phone and the clay fox under the press (see Fig. 2). Anni's reactions, mainly laughter, can be heard in the background at various points. The video thus focuses the attention of the viewer on the phone (and the clay fox), the press, the action of crushing and the commentary.

The multimodal video interaction reveals a tight patterning of lexical cohesion within the commentary and between the verbal commentary and the visual content of the video. Reiteration dominates: in the spoken commentary, there are simple repetitions of *phone* and *it* as well as substitutions, e.g. *phone – it*, and equivalence, *Nokia 3310 – phone* and *warm – heated up*. Similarly, *press* is repeated several times (both simple and complex repetition occurs, as *press* is used both as a noun and a verb), as is *fox* in the bonus part. All the while, multimodal cohesive ties are created between the mentions in the commentary and the actual items seen in the video: the press, the phone, and the clay fox in the bonus part. Fig. 3 illustrates how the mentions relating to the phone form a chain of lexical cohesion, which is multimodally linked with the visual.

If we consider the clip from the perspective of audience design, we see that the YouTuber uses several features aimed at involving the viewers. He speaks to the imagined audience, first welcoming them to view the clip and at the end thanking them for viewing. New viewers are addressed in the bonus clip. Like most YouTubers, Vuohensilta also asks his viewers to add comments (Burgess & Green 2009): he addresses them directly (*you*) and encourages them to suggest ideas for items that could be crushed in future videos. This request is delivered as written text, which shows multimodal cohesive links to the commentary (*crush* in an equivalence relation with *press*) and cross-modal links to the comments (more precisely to the title of the level 'comments'). Using Eisenlauer's (2013) terminology, the former is an instance of intranodal cohe-

sion, linking within the video, while the latter is an example of internodal cohesion, creating a link to another node or level in the YouTube interaction.

4.2. Lexical cohesion in the comments

4.2.1. The cohesive profile of the comments

This section presents the results of the analysis at the level of the comments. Out of the 300 comments studied in detail, only 38 show no lexical cohesive links; all the rest have at least one, and many have several, as we shall see in the examples. The comments lacking cohesion are mainly very short reactions to the video (e.g. *cool, haha, lol, omg, my life is a lie*). Lexical cohesion in the comments is therefore pervasive. As in the video interaction, reiteration dominates the cohesive profile in the comments as well. Collocation is infrequent (for occurrences, see example 4). As might have been expected, just as in the video interaction, *Nokia*, *phone* and *press* are frequently repeated:

---

(1) João Correia: Obviously *fake*. The *nokia* activated its self destruct mechanism to avoid damaging your *press*
Rekkone: *FAKE NOKIA 3310*.
→ ɹǝʞoſ: +Rekkone YEAH[3]
Jordy: A *real Nokia 3310* would've *crushed* that *hydraulic press*. This is obviously a *fake model*.
ShiroiKage009: The *phone* got *crushed*? Must be a *fake*.

The *phone/Nokia* and the *press* mentioned in (1) appear in both the visual and verbal (spoken) modes in the video. *Nokia*, for instance, thus creates a cross-modal link between the written comments by the viewers, the spoken commentary by the YouTuber and the phone seen in the video. At the level of the comments, the mentions create a chain of lexical cohesion (cf. commentary in Fig. 3). Unlike in the commentary, *phone* is less frequent in the comments, as commenters more often refer to it by the brand name. *Phone* has 274 hits, while *Nokia* has 1078, which means that it occurs in every sixth message (*Nokia* is at #5 in the word frequency list, *press* (N + V) at #9; see Appendix A). *Crushed* is also repeated, as is *fake* (both will be discussed in detail below).

What is interesting about example (1) is that although the comments are very similar and clearly related at the level of cohesion, most of them have been posted as individual comments.[4] Joker's comment is the only reply, corroborating Rekkone's comment. When we remember how replies are presented on YouTube, i.e. that in order to read replies, the viewer has to click the "view replies" prompt (see Appendix E for examples), it is tempting to postulate that the commenters may actually be strategically posting their comments as individual ones to guarantee their visibility in the string of comments. Although the related comments do not appear as a sequence, they are still adjacent in the chronological order of the comments.

---

[3] The arrow at the beginning of the line indicates that the comment is a reply to the previous comment, or, in the case of multiple replies, to the initiating comment of the sequence. Comments without an arrow are individual comments.

[4] The relatively high number of individual comments differentiates the HPC Nokia 3310 discussion from the discussions studied by Bou-Franch et al. (2012) and Liebschner (2021), in which participants reply to other comments more frequently. The comments also appear to be longer in the two previous studies, if compared with the fairly short HPC comments. The differences between the initiating videos may partly explain the variation in comment length. The purpose of the HPC Nokia 3310 video is mainly entertainment, whereas the two initiating videos in Bou-Franch et al's study deal with abortion and domestic violence, and the one in Liebschner's study with Russian life style. These videos encourage viewers to exchange opinions, while the HPC video induces viewers to record their reactions.

Fig. 2. Screenshot of the video showing the phone, the (bottom part of the) press and the YouTuber's hand.

We saw above how the YouTuber involves his audience by addressing them directly, and the viewers reciprocate, creating exchanges resembling face-to-face conversation (Frobenius 2014: 70). *You* occurs 1,044 times in the comments, and about half of them refer to the YouTuber (see example 2; only the pronoun is highlighted in this example). As pointed out above, pronoun repetition is included in the model of analysis alongside other simple repetitions, and repetitions of *you* referring to the sender thus create cross-modal and internodal links between the comments and the sender as well as intranodal links between the comments.
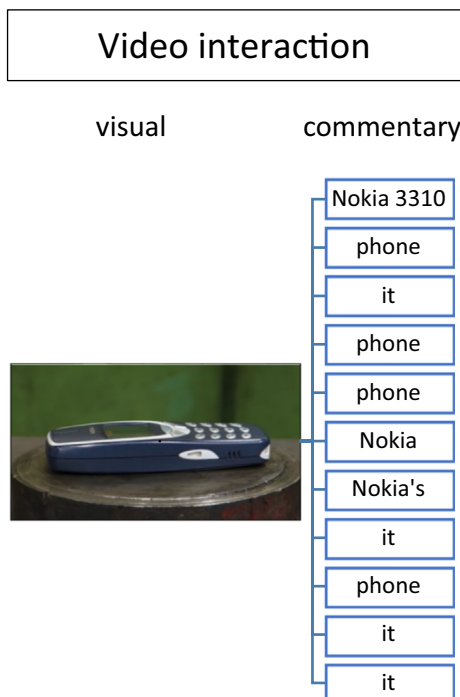

Fig. 3. Chain of lexical cohesion in the commentary, multimodally linked with the visual in the video.

(2) Claudius Chileanus: can *You* crush what's left of my soul next? :c
OssQ: *You* are breaking the laws of physics here mate. Or is that press made of 3310′s?
Windows 98: Suggestion: *You* should have individual thumbnails for all of *your* videos, it'll make it easier to decide what to watch and increase viewership.

Interestingly, there are only two replies from the YouTuber (using the channel name as username), both towards the beginning of the comments (for one of them, see example 4 below). Perhaps as Vuohensilta uploads new videos, he no longer follows the comments posted on older videos so closely. Regardless of the lack of interaction from the sender at the level of comments, the viewers keep addressing him when they post their ideas for items to crush or offer other comments on the video or the channel.

In addition to single lexical items, cohesion is created by repetitions of structures or sequences of words (see Berglund 2009). One of the frequent clusters in the material is *still works*, which is at #7 in the cluster frequency list with 157 hits (see Appendix A). Example (3) also shows reiteration relations between *phone*, *it* and *nokia* as well as *crushed* and *crush* (substitution and simple repetition), and *fake* is repeated:

(3) Mad Dirty: I'm quite confident that *phone still works*.
Alex: may be *crushed*, but *it* probably *still works*.
Feedmeee: *crush* a dead mouse
Finlay Craig: a god has been killed.
Fabian Barrios: *Fake* :v
wild jester: Not possible! Well, I'm sure *it still works*, otherwise *it* was a *fake nokia* :))

Here, too, as in example (1) above, the comments have been posted as individual ones. The relatedness created by the repetitions of *still works* is enhanced by the references to the *phone*. Alex refers to the *phone* introduced by Mad Dirty with the substitution item *it*, which wild jester then repeats, although there are three

intervening comments between the comments. As in example (1), all the mentions of the phone create cross-modal links to the phone seen and talked about in the video.

All but one of the comments discussed so far have been posted as individual comments. Regardless of this, they have been shown to be connected by cohesive ties between them, sometimes over intervening comments. To illustrate cohesion in a sequence of comments, which have been posted as replies to the initiating comment, let us consider example (4):

---

(4) TwitchFast: I assume the *liquid* is from the *battery*?
    → Sanna XO: +TwitchFast *It* is
    → Hydraulic Press Channel: +TwitchFast This time I took *battery* out of the *phone* before *crushing*. The *Sony* smelled so toxic that I thought that it is maybe better to take *it* out first.
    → Timinimification: +TwitchFast *Liquid LCD* from the *screen* perhaps?
    → Leo: +TwitchFast I'm sorry to tell you this… but that's *Nokia blood*.
    → Cwazywazy1240: +Hydraulic Press Channel Might be the *screen* or something. Anyway, you should try *crushing* potato chips, see if any *oil* will be squeezed out or something.
    → David Braucht: +TwitchFast That *liquid* will later reassemble itself into a new *nokia*……..

TwitchFast's comment is the first one to mention *liquid* and *battery*; the liquid referred to can be seen in the video towards the end of the crushing of the Nokia. In other words, the mention of the liquid creates a cross-modal tie with the video. The *battery* is intriguing from this perspective, because in fact it cannot be seen in the video; the YouTuber reveals a little later that the battery has been removed. However, because a phone and a battery are related through their habitual co-occurrence, a cross-modal relation by elaborative collocation is created with the first mention of the battery. Other ideas then follow for the origin of the liquid. In this short extract, we have two short chains of cohesion operating. *Liquid – liquid LCD – Nokia blood – oil – liquid* is the first, and the phone-related *battery – phone – Sony – screen – Nokia* the second. In addition, there is repetition of *crushing*. Cohesion in this sequence (and other sequences) is undeniably tight, but the same is true with related but individual comments, as we have seen. In terms of lexical cohesion, then, whether comments are posted as replies or individual comments does not appear to make a difference in this online discussion.

4.2.2. Chains of cohesion: Fake and crush

To answer the research question of how lexical cohesion operates over a lengthy YouTube interaction, we now turn our attention to cohesive chains. Looking at examples (1) and (3), the frequently re-occurring word *fake* stands out. Without a doubt, *fake* creates cohesion between all the comments in example (1), and between the two last comments in example (3). The word frequency list in Appendix A reveals that *fake* is the first content word in the list, at #4 with 1143 hits. Furthermore, there are three clusters with *fake* among the 15 most frequent clusters (*fake Nokia, a fake, is fake*). Such high frequency of occurrence reveals that for some reason many of the commenters allude to 'fakeness' in their comments.

Appendix B presents the concordance plot created by AntConc for *fake*. Each vertical line in the plot represents an instance of *fake* among all the words occurring in the comments. The darker the plot, the higher the frequency of instances of *fake* in the string of

comments. The plot tells us that *fake* occurs throughout the comments, at times peaking in frequency and never actually fading (if anything, the occurrences seem to get more frequent in the newer messages). A selection of other words related to 'fakeness' also occur in the comments, such as *copy, replica, knockoff, rip-off* and *bootleg*, although they are far less frequent than *fake* (with altogether a couple of dozen hits). They are related with *fake* through equivalence and thus join the 'fakeness' chain.

On the basis of the frequency information and the concordance plot, it seems safe to suggest that the 'fakeness' chain creates unity between the comments throughout the by now over-four-year span of the online discussion. For an explanation as to why the idea of 'fakeness' is so central, we can turn to the commenters themselves (5):

---

(5) graham grasdal: That was probably a Chinese knockoff of a real Nokia 3310 because we all know they are indestructible.
    Joseph H: Um…….…did I miss something? Why are people saying it's fake when it…it just happened. xD Is it really that hard to believe that something broke that?
    Matthew Bolan: its a meme. in meme culture, nokia phones are indestructible. All the fake comments are just memesters memeing

---

In addition, knowyourmeme.com informs us that

Indestructible Nokia 3310 is the nickname given to Nokia's 3000 series mobile phones and customer review parodies poking fun at their durability and heavy weight. The joke typically manifests itself in image macros, in which the Nokia phone is shown to be an incredibly powerful or destructive force.

Indeed, a search for the 'Nokia 3310 meme' on the internet produces hundreds of images where the phone is presented as indestructible; see Appendix C for examples of these. The comments about 'fakeness' directed towards the phone or the action in the video should thus be interpreted with reference to a meme making fun of the qualities of the phone, while comments mentioning the indestructibility of the phone or the death of a meme directly reference the meme. In this way it creates cohesion in both a cross-modal and intertextual manner.

Fig. 4 shows how cross-modal and intertextual cohesion is created between the comments as well as between the comments and the video through the meme. Starting from the left, the video itself is audio-visual. Although the commentary does not mention the meme, only that the phone is "tough", it can be regarded as a reference to the meme: the video is especially interesting because it shows the crushing of an allegedly indestructible object. In the centre, the meme relies on text and image to convey the intended message, and as Dancygier and Vandelanotte (2017) note, the combination of text and often iconic images makes the memes themselves multimodal (see also Yus 2021). The undoubtedly iconic images in the Nokia 3310 memes in Appendix C also make the memes intertextual; knowledge about *The Lord of the Rings* and *Star Wars* is required for a full interpretation (see Wiggins 2019: 34).

On the right-hand side of Fig. 4 are the text-based comments, with the one in the middle in bold typeface, because 'fakeness' is the most frequent theme. Comments referring to 'fakeness' and those referring to the indestructible quality of the phone or the death of the meme are related through their reference to the meme. When the commenters refer to the indestructibility of Nokia 3310, the death of the meme or the fakeness of either the video or the phone seen in the video, the cohesive links formed by these mentions are not merely items on the surface of the
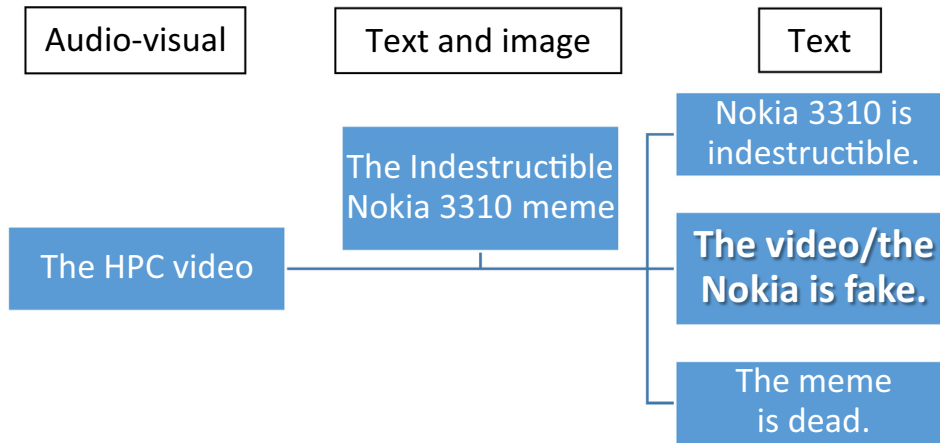
Fig. 4. Cross-modal and intertextual cohesion between the video, the meme and the comments.

comments: their relatedness is the result of a cross-modal, multimodal and intertextual network of meaning-making.

The Nokia 3310 meme is not the only meme in this video and discussion; a second one is mentioned in the bonus part of the video, namely the "What does the fox say?" meme. Originally an electronic dance song, The Fox became a meme, most often combining an image, the question "What does the fox say?" and suggestions for what it might be saying. The mention of the clay fox in Vuohensilta's commentary is a direct reference to the meme, which many of the commenters recognise:

(6) RiotDemon: What does the fox say? "EHHHugh"
Confuseh: "what does the fox say… under the press" - *"EEEEEEEEEEEEEEEEEEEEEHHHHHHH"* im ded
Jon Baxter: Crushing a nokia… what does the fox say… the memes are off the charts…

In example (6), RiotDemon and Confuseh first repeat the question, as is done in the meme, with the repeated structure helping to create cohesion (Berglund 2009). They then spell out the screech of the clay fox (or Vuohensilta) in the video as suggestions for its reaction, with Confuseh adding "im ded". Jon Baxter takes the discussion to another level, commenting on the use of memes in the video more generally. The comments referencing the two memes produce a web of cross-modal cohesive links through the online discussion, built on multimodality and a reliance on intertextual knowledge.

Having considered the longest chain, we can now discuss the other long chain in the comments. This chain is created by the viewers' replies to the text in the video prompting them to suggest items for crushing. Accordingly, *crush* is at #14 in the word frequency list with 660 hits, and 'crush a' tops the cluster frequency list (Appendix A).

(7) Sorrells: *Smash* a burger!
Karrut: *Crush* your hand.
Reptoid: *crush* a can of beans or some food like that
Ptolemaic Taweret: *Crush* a coconut please
Nick Breen: Yeah I think some sort of soft metal like brass or copper would have good results, old brass ornaments.
You Tube: a soda can!!!

The fact that viewers respond to a prompt by the sender creates cross-modal cohesion in itself, but it is even clearer with the repe-

tition of *crush*; also *smash* and *press* occur in the suggestions, which reiterate *crush* with an equivalent item. The items suggested range from the more concrete *baseball* and *coconut* to the more abstract *your hopes and dreams* and *Trump's dreams of presidency*. The last two suggestions in example (7) are interesting: Nick Breen suggests old brass ornaments and You Tube a soda can, but without repeating the verb. Although Halliday and Hasan's (1976) category of ellipsis or substitution by zero is not included in our model, we can use it to explain why the last two are related to the other suggestions: the verb has been omitted because it is not an essential element in question–answer (or prompt-response) sequences (cf. Hoey 1991: 74; Tanskanen 2006: 53–54).

Fig. 5 illustrates how *a soda can!!!* and other suggestions like it are cohesive with the other suggestions for items to crush even when there is no explicit mention of a related verb. Altogether the suggestions form both intranodal links within the comments, and internodal and cross-modal links between the comments and the video.

To conclude the discussion of chains, let us consider the differences between the two chains. The *fake* chain has more hits than the *crush* chain, but more interesting than the numbers is how the hits are situated in the string of comments. This is revealed by the concordance plots in Appendix B. Compared to the plot for *fake*, where the hits are quite evenly distributed throughout the comments, the plot for *crush* shows a different pattern. There are clearly more occurrences of *crush* towards the beginning of the discussion (i.e. the plot is darker), and the occurrences seem to fade as we come to the newer comments. Viewers thus seem to keep commenting on the perceived fakeness of the video all
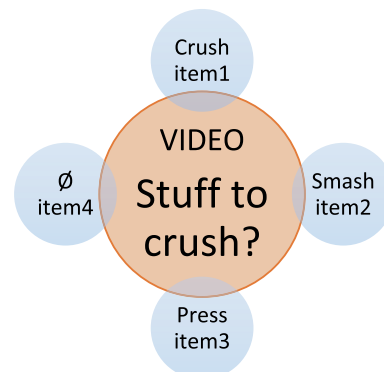

Fig. 5. Cohesion between suggestions for "stuff to crush", and between the video and the suggestions.

through the comments, but they are not as eager to suggest items for crushing after the first third of the comments. At least two possible reasons appear likely: either the viewers recognise that the YouTuber may not be actively following the commenting on older videos, or the suggestions may already have been taken up and carried out by the YouTuber in the newer videos.

5. Concluding remarks

Millions of users have viewed and several thousand have so far posted comments on the HPC Nokia 3310 video in an online interaction that has now lasted for over four years. The analysis has shown how lexical cohesion helps maintain coherence over shorter segments of the interaction (e.g. the spoken commentary in the video), and, quite remarkably, throughout the entire lengthy interaction in the comments with a long cohesive chain focusing on a central theme, i.e. fakeness. The use of cohesive ties is extensive within the video interaction and the comments as well as between these levels of interaction. Cohesion in the comments is dense regardless of the fact that only a minority of the comments form sequences, i.e. are posted as replies to an initiating comment; most comments have been posted as individual items (cf. Bou-Franch et al. 2012; Liebschner 2021). The cohesive items form intranodal connections within a level of interaction (in the video interaction and in the comments), internodal connections between the comments and the video, and extranodal ones with material elsewhere on the internet, the latter two built on cross-modality and intertextuality.

Of the two main categories of lexical cohesion, reiteration clearly dominates the cohesive profile, while collocation is rare. This finding is in line with previous research on YouTube cohesion (Liebschner 2021; Schubert 2017). Comparisons with mailing list discussions (Tanskanen 2006: 166) and blog comments (Hoffmann 2012: 171–172) reveal interesting similarities and differences. Reiteration is frequent in all these online contexts and dominates the cohesive profiles. Collocation, on the other hand, is clearly more frequent in mailing list discussions and blog comments than in the material of the present study. If comparisons are extended to offline contexts, the cohesive profile of the YouTube interaction bears resemblance to two-party conversations in terms of reiteration and multi-party conversations in terms of collocation (Tanskanen 2006: 166).

The analysis showed how even very short comments such as *Fake!* or *a soda can!!!*, which may initially appear unrelated and disconnected, are in fact cohesive both with other comments and cross-modally with the video. *Fake* makes perfect sense as a member of the long chain, building upon a cross-modal, multimodal and intertextual relation with an internet meme, while *a soda can* takes its place among other suggestions offered as a written response to a written prompt shown in the video. The fragmentation of a YouTube interaction, brought about by the fact that a large number of participants post relatively short comments on the video they have viewed, is a feature on the surface only. The above and other instances discussed in this paper show that a cross-modal, multimodal and intertextual perspective is useful for a study of YouTube cohesion, as cohesive ties exist between all the levels of the interaction as well as with extranodal sources. Limiting the study to the level of comments would have revealed only a partial picture of the network of cohesive ties in the interaction. In terms of cohesion, the video and the comment interaction are not separate entities.

Although the present study focused on one video clip and the comments section on one YouTube channel only, it adds to existing knowledge of cohesion on YouTube, with the analysis revealing a tight web of cohesion within and across all the levels of the interaction. Future research could probe the use of cohesion in YouTube interaction which is even more complex in its multimodality. For instance, how cohesion functions across a video and an embedded video, or in situations involving an even greater number of participants producing the video or commenting on it would be interesting themes for future studies, further adding to our understanding of the role of cohesion in a context of complex multimodality and fragmentation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. The most frequent words (1) and two-word clusters/N-grams (2) in the comments.

(1)

Rank	Frequency	Word
1	2150	the
2	2115	a
3	1253	it
4	1143	fake
5	1078	nokia
6	1044	you
7	1031	is
8	965	I
9	937	press
10	770	this
11	765	of
12	751	to
13	744	that
14	660	crush
15	597	and

(2)

Rank	Frequency	Cluster
1	389	crush a
2	363	hydraulic press
3	282	the press
4	257	it s
5	246	this is
6	195	the nokia
7	157	still works
8	152	a nokia
9	135	fake nokia
10	115	it still
11	113	a fake
12	108	the hydraulic
13	101	is fake
14	91	don t
15	91	press is

Appendix B:. AntConc concordance plots for *fake* (1) and *crush* (2). Each vertical line is a hit of the target word in the file (normalised to the width of the bar).

(1) *fake*

(2) *crush*

Appendix C. Three examples of the indestructible Nokia 3310 meme from KnowYourMeme (knowyourmeme.com).

Appendix D. Top comments for the HPC Nokia 3310 video.

Edho Kuddah 4 years ago

obviously reversed, this is how it's made

👍 4.9K 👎   REPLY

▼ View 24 replies

Forum Cheeki 4 years ago

these are really good special effects! Obviously a nokia 3310 cannot be destroy

👍 1.4K 👎   REPLY

▼ View 15 replies

edwin 4 years ago

the Nokia just temporary changed the shape to not damage his press..

👍 414 👎   REPLY

▼ View 2 replies

Michael Wen 4 years ago

the 3310 has a self destruct sequence so it doesn't break your press

👍 121 👎   REPLY

Cassandra F 4 years ago

Spolier Alert....the nokia phone breaks his press. You're welcome.

Appendix E. Newest comments for the HPC Nokia 3310 video.

Ringcisten 1 week ago

That nokia was an imposter.

👍 👎 REPLY

margareta magicalowska 1 week ago

FAKE. nokia 3310 rozkurwiła by tą prase w chuj

👍 👎 REPLY

Chirag Daga 1 week ago

And then we realised that this was just the demo piece..
The real has broken his 2 machines

👍 1 👎 REPLY

The Biggest hit 2 weeks ago

hot to break okie phone
how to break hamer

👍 👎 REPLY

Fabiofac a 2 weeks ago

Fake

👍 👎 REPLY

References

Adami, E., 2014. "Why did dinosaurs evolve from water?": (in)coherent relatedness in YouTube video-interaction. Text & Talk 34 (3), 239–260.

Adami, E., 2015. What I can (re)make out of it: Incoherence, non-cohesion, and re-interpretation in YouTube video responses. In: Chovanec, J., Dynel, M. (Eds.), Participation in public and social media interactions. John Benjamins, Amsterdam, pp. 233–257.

Androutsopoulos, Jannis and Jana Tereick. 2015. YouTube. The Routledge handbook of language and digital communication. Abingdon & New York: Routledge.

Anthony, Laurence. 2019. AntConc (Version 3.5.8) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

Benson, P., 2016. The discourse of YouTube: Multimodal text in a global context. Taylor & Francis. [VitalSource Bookshelf], London.

Berglund, T.Ö., 2009. Disrupted turn adjacency and coherence maintenance in instant messaging conversations. Language@Internet 6 (2), 1860–2029.

Bou-Franch, P., Garcés-Conejos Blitvich, P., 2014. Conflict management in massive polylogues: a case study from YouTube. Journal of Pragmatics 73, 19–36.

Bou-Franch, P., Lorenzo-Dus, N., Blitvich, P.-C., 2012. Social interaction in YouTube text-based polylogues: A study of coherence. Journal of Computer-Mediated Communication 17 (4), 501–521.

Boyd, M.S., 2014. (New) participatory framework on YouTube? Commenter interaction in US political speeches. Journal of Pragmatics 72 (1), 46–58.

Burgess, J., Green, J., 2009. The entrepreneurial vlogger: Participatory culture beyond the professional/amateur divide. In: Snickars, P., Vonderau, P. (Eds.), The YouTube reader. National Library of Sweden, Stockholm, pp. 89–107.

Dancygier, B., Vandelanotte, L., 2017. Internet memes as multimodal constructions. Cognitive Linguistics 28 (3), 565–598.

Dynel, M., 2014. Participation framework underlying YouTube interaction. Journal of Pragmatics 73, 37–52.

Eisenlauer, V., 2013. A critical hypertext analysis of social media: The true colours of Facebook. Bloomsbury, London & New York.

Frobenius, M., 2014. Audience design in monologues: How vloggers involve their viewers. Journal of Pragmatics 72 (1), 59–72.

Goffman, E., 1981. Forms of talk. University of Pennsylvania Press, Philadelphia.

Halliday, M.A.K., Hasan, R., 1976. Cohesion in English. Longman, London.

Hasan, R., 1984. Coherence and cohesive harmony. In: Flood, J. (Ed.), Understanding reading comprehension: Cognition, language, and the structure of prose. International Reading Association, Newark, DE, pp. 181–219.

Herring, S., Chae, S.W., 2021. Prompt-rich CMC on YouTube: To what or to whom do comments respond? In: Proceedings of the 54th Hawaii International Conference on System Sciences, pp. 2906–2915.

Hoey, M., 1991. Patterns of lexis in text. Oxford University Press, Oxford.

Hoffmann, C.R., 2012. Cohesive profiling: Meaning and interaction in personal weblogs. John Benjamins, Amsterdam & Philadelphia.

Johansson, Marjut. 2017. YouTube. In Hoffmann, Christian R. & Wolfram Bublitz (Eds.), Handbooks of pragmatics (HOPS), Vol. 11: Pragmatics of social media, 173–200. Berlin: Mouton de Gruyter.

Jones, G.M., Schieffelin, B.B., 2009. Talking text and talking back: "My Bff Jill" from Boob Tube to YouTube. Journal of Computer-Mediated Communication 14 (4), 1050–1079.

Leppänen, S., Häkkinen, A., 2012. Buffalaxed superdiversity: representations of the other on YouTube. Diversities 14 (2), 17–33.

Liebschner, A., 2021. Language evolution online? The use of multimodal cohesion on YouTube. Service Plus 15 (2), 51–61.

Liu, Y.u., O'Halloran, K.L., 2009. Intersemiotic texture: analysing cohesive devices between language and images. Social Semiotics 19 (4), 367–388.

Marcoccia, M., 2004. On-line polylogues: conversation structure and participation framework in internet newsgroups. Journal of Pragmatics 36 (1), 115–145.

Morris, J., Hirst, G., 1991. Lexical cohesion computed by thesaural relations as an indicator of the structure of text. Computational Linguistics 17 (1), 21–48.

Parsons, G., 1991. Cohesion coherence: scientific texts. In: Ventola, E. (Ed.), Functional and systemic linguistics: Approaches and uses. Mouton de Gruyter, Berlin & New York, pp. 415–429.

Royce, T., 2007. Intersemiotic complementarity: A framework for multimodal discourse analysis. In: Royce, T., Bowcher, W. (Eds.), New directions in the analysis of multimodal discourse. Lawrence Erlbaum, Mahwah, NJ and London, pp. 63–109.

Schmidt, A., Marx, K., 2019. Multimodality as challenge: YouTube data in linguistic corpora. In: Pflaeging, J., Bateman, J., Seizov, O., Tseng, C.-I. (Eds.), Wildfeuer, Janina. De Gruyter, Multimodality. Berlin & Boston.

Schubert, Christoph. 2017. Discourse and cohesion. In Hoffmann, Christian R. & Wolfram Bublitz (eds.), Handbooks of pragmatics (HOPS), Vol. 11: Pragmatics of social media, 317–343. Berlin: Mouton de Gruyter.

Senft, T.M., 2008. Camgirls: Celebrity and community in the age of social networks. Peter Lang, New York.

Tanskanen, Sanna-Kaisa, 2006. Collaborating towards coherence: Lexical cohesion in English discourse. John Benjamins, Amsterdam & Philadelphia.

Van Leeuwen, Theo. 2005. Introducing social semiotics. London: Routledge.

Wiggins, Bradley E., 2019. The discursive power of memes in digital culture: Ideology, semiotics, and intertextuality. Routledge, Abingdon & New York.

Yus, Francisco, 2021. Incongruity-resolution humorous strategies in image macro memes. Internet Pragmatics 4 (1), 131–149.