*June 15, 2023*

Louis Cotgrove

Using YouTube comments as teaching material: The benefits of the NottDeuYTSch corpus

Developments within the field of Second Language Acquisition (SLA) have meant that scholars are increasingly engaging with corpora and corpus-based resources, providing a source of "'authentic' language" to learners and educators (Mitchell 2020: 254), and contributing to "state-of-the-art research methodologies" (Deshors and Gries 2023: 164). However, there are areas in which progress can still be made, particularly in the area of metadata, such as information about the speaker and contexts of the language use, as well as increased variety in the text types and genres of corpora used to develop SLA materials (Paquot 2022: 36). This post discusses one such possibility for increasing the variety of text types and providing a rich source of authentic language that can be used to create engaging SLA materials, particularly for young people learning German, namely the use of the *NottDeuYTSch* corpus (to download the corpus in a variety of formats, see Cotgrove 2018).

The *NottDeuYTSch* corpus is a collection of over 3 million YouTube comments extracted from German-language videos aimed at a mainstream youth audience, and published between 2008 and 2018 (Cotgrove 2023). The corpus provides the opportunity for a wide range of analyses of digitally-mediated communication (DMC), specifically written by young people. These include studies of language change over time, the comparison of language based on YouTube video channel and category, as well as more traditional sociolinguistic analyses of the lexis, sentence structure, and interaction between young people online. Additionally, the corpus can be used as a basis for developing learning resources for the teaching of German as an additional language (referred to in German as *Deutsch als Fremdsprache/Deutsche als Zweitsprache* or DaF/DaZ), which is analysed in more depth here.
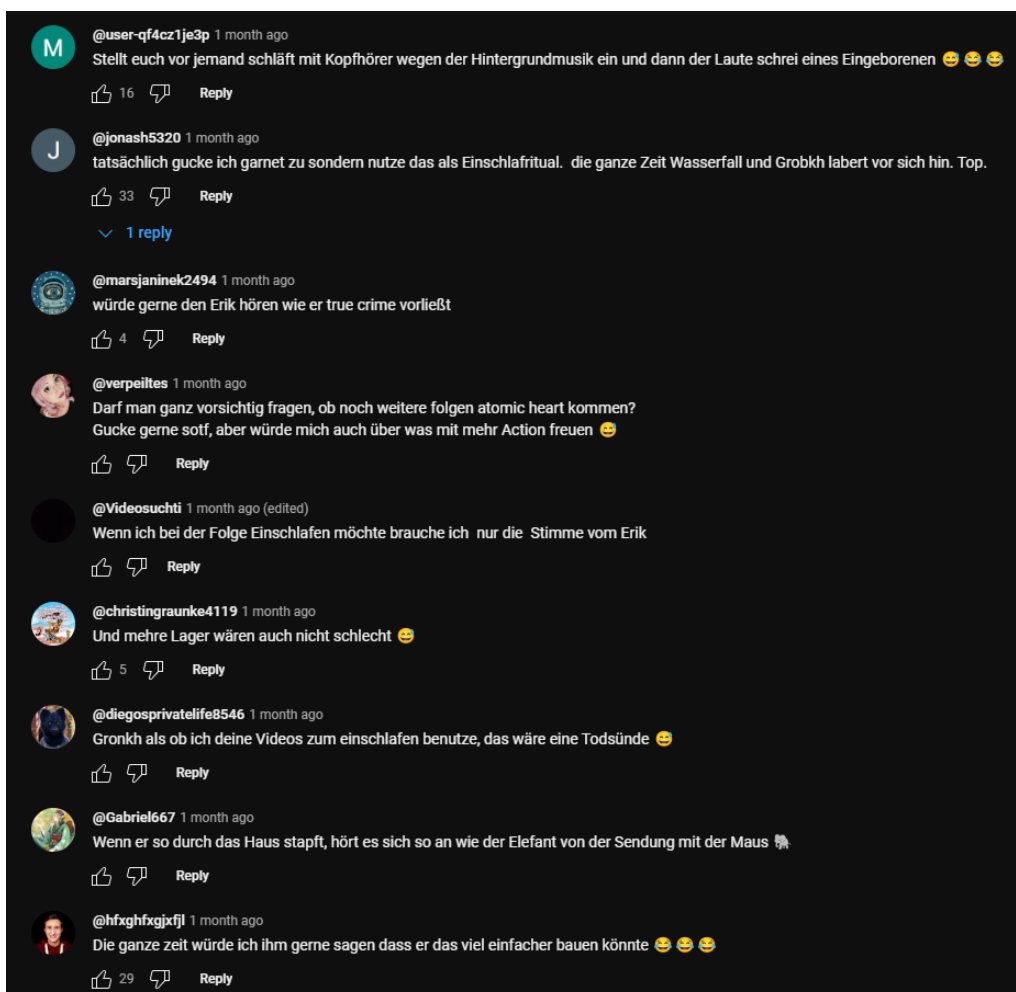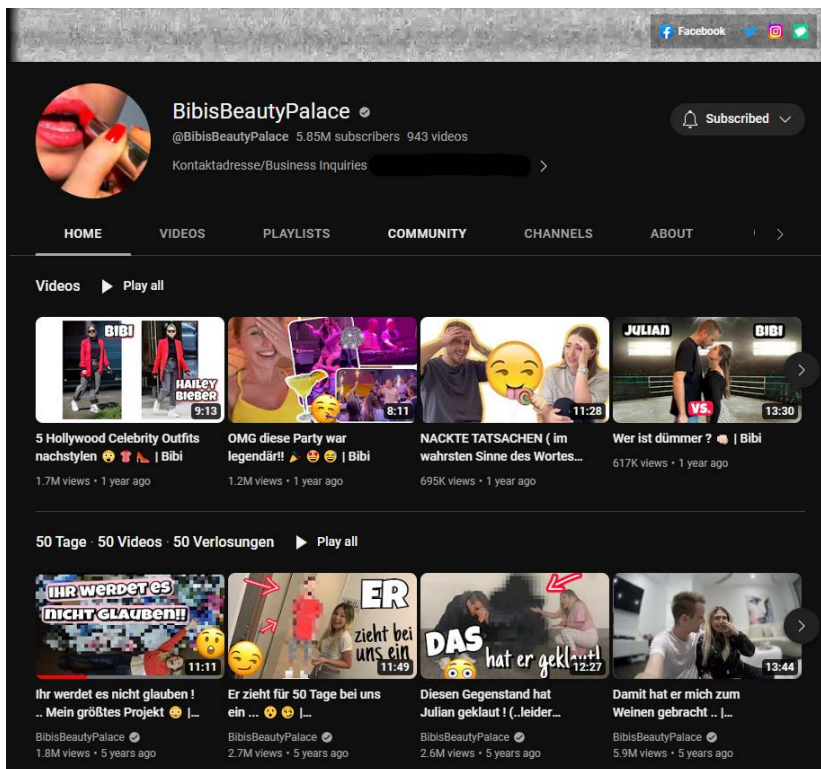
Considerable research, especially in the field of psycholinguistics, has demonstrated that the learning environment is crucial for SLA in young and teenage learners (MacIntyre et al. 2003). There are a number of factors that influence the learning environment; three are particularly relevant to young learners of languages: motivation, foreign language anxiety (FLA), and willingness to communicate (WTC).

Motivation is defined here as the "desire to achieve the goal, positive attitudes, and effort" (MacIntyre 2002: 46) and has long been positively linked with achievement in SLA (Krashen 1982: 30). It is important to create a learner-centered environment to increase motivation in SLA, e.g. adapting materials that are relevant and accessible to learners. FLA is "the subjective feeling of tension, apprehension, nervousness, and worry associated with an arousal of the autonomic nervous system […] limited to the language learning situation" (Horwitz et al. 1986: 125), and is particularly prevalent in teenage learners (MacIntyre et al. 2003; Russell 2020). This is partly due to psychological apprehensions experienced by teenagers related to the display of vulnerability or ridicule associated with adolescence (Harklau 2007: 639-640). One way in which FLA can be combatted is by creating a learning environment that contains authentic materials that are relevant to learners. WTC is an "individual's intention to initiate or participate in communication [in] the target language at a particular moment and situation" (Reinders and Wattana 2014: 105). Therefore, the use of digitally-mediated content and focus on digitally-mediated communication provided by the *NottDeuYTSch* corpus can help provide a familiar environment and produce accessible, enjoyable learning materials for young learners, increasing their engagement with the material, increasing motivation and WTC, and combatting FLA, which ultimately leads to enhanced SLA.

Researchers have advocated for the use of authentic language for SLA since the 1970s (see, e.g., Craik and Lockhart 1972), but available corpora were too small and unsuitable for the task, meaning that educators have only recently been able to utilise these resources. Exposure to authentic language has been demonstrated to have the following benefits for learners: an

increase in quality of the target language output (Li 2017), the development of pattern recognition (Gilquin and Granger 2022), and the awareness of pragmatic aspects of language, such as the awareness of the appropriateness of certain linguistic stuctures in certain communicative situations (Gilmore 2007: 100; Staples and Fernández 2019; Wain et al. 2019).

Contrived examples, i.e. those thought up by educators, often do not present learners with realistic communicative situations. However, research has continually demonstrated the value of using authentic materials as they "provide the best source of rich and varied input for language learners" and "impact on affective factors essential to learning" (e.g. motivation, FLA, and WTC) (Mishan 2005: 41–42), thus improving comprehension and language production. It is clearly useful, therefore, to design learning activities using authentic language. For example, YouTube comments, such as those in the *NottDeuYTSch* corpus, could be used for tasks involving register, informal language, and sociolinguistic aspects of language. The corpus could also be used to illustrate non-standard features of language, such as lexis and grammar, for example, the online-specific borrowings, *Abo* (subscription) or *klickgeil* (being desperate for clicks). YouTube comments can also be used as a springboard for learners to produce their own language, using the corpus as reference material to post comments under other videos or interact with other commenters in the target language (see Ritchie and Black 2012). Corpus-based resources also enable educators to devise resources that train particular aspects of language learning (see Li 2017), due to the wide range of features and situations that can be identified, including collocations, idioms, and discourse markers.

The use of the *NottDeuYTSch* corpus can therefore allow educators to improve motivation and WTC and reduce FLA in learners: in particular by allowing educators to extract subcorpora that focus on a specific feature or structure, thereby 'scaffolding' learning (see Wood et al. 1976). The corpus harmonises with other approaches to SLA, such as the use of textbooks, by providing extra material to reinforce the pedagogical goal of a particular exercise. Finally, learners can adapt the examples in the corpus to produce their own authentic communication, which can fulfill the "ultimate goal of language learning", namely "authentic communication between persons of different languages and cultural backgrounds" (MacIntyre et al. 1998: 559).

References

Cotgrove, Louis Alexander. 2018. 'Das Nottinghamer Korpus Deutscher YouTube-Sprache (the NottDeuYTSch Corpus)' (LINDAT/CLARIAH-CZ) <http://hdl.handle.net/11372/LRT-4806>

———. 2023. 'New Opportunities for Researching Digital Youth Language: The NottDeuYTSch Corpus', in Neue Entwicklungen in Der Korpuslandschaft Der Germanistik, ed. by Marc Kupietz and Thomas Schmidt (Tübingen: Narr)

Craik, Fergus I.M., and Robert S. Lockhart. 1972. 'Levels of Processing: A Framework for Memory Research', Journal of Verbal Learning and Verbal Behavior, 11.6: 671–84 <https://doi.org/10.1016/S0022-5371(72)80001-X>

Deshors, Sandra C, and Stefan Th Gries. 2023. 'Using corpora research on second language psycholinguistics', in The Routledge handbook of second language acquisition and psycholinguistics, Routledge handbooks in second language acquisition, ed. by Aline Godfroid and Holger Hopp (London: Routledge), pp. 164–77

Gilmore, Alex. 2007. 'Authentic Materials and Authenticity in Foreign Language Learning', Language Teaching, 40.2: 97–118 <https://doi.org/10.1017/S0261444807004144>

Gilquin, Gaëtanelle, and Sylviane Granger. 2022. 'Using Data-Driven Learning in Language Teaching', in The Routledge Handbook of Corpus Linguistics, 2nd Edition, ed. by Anne O'Keefe and Michael McCarthy (Abingdon: Routledge), pp. 430–42

Harklau, Linda. 2007. 'The Adolescent English Language Learner', in International Handbook of English Language Teaching, Springer International Handbooks of Education, ed. by Jim Cummins and Chris Davison (Boston: Springer), pp. 639–53 <https://doi.org/10.1007/978-0-387-46301-8_41> [accessed 14 April 2023]

Horwitz, Elaine K., Michael B. Horwitz, and Joann Cope. 1986. 'Foreign Language Classroom Anxiety', The Modern Language Journal, 70.2 (JSTOR): 125–32 <https://doi.org/10.1111/j.1540-4781.1986.tb05256.x>

Krashen, Stephen. 1982. Principles and Practice in Second Language Acquisition (Oxford: Pergamon Press)

Li, Li. 2017. New Technologies and Language Learning (London: Palgrave Macmillan)

MacIntyre, Peter D. 2002. 'Motivation, Anxiety and Emotion in Second Language Acquisition', in Individual Differences in Second Language Acquisition, ed. by P. Robinson (Amsterdam: John Benjamins), pp. 45–68

MacIntyre, Peter D., Susan C. Baker, Richard Clément, and Leslie A. Donovan. 2003. 'Sex and Age Effects on Willingness to Communicate, Anxiety, Perceived Competence, and L2 Motivation Among Junior High School French Immersion Students', Language Learning, 53.S1: 137–66 <https://doi.org/10.1111/1467-9922.00226>

MacIntyre, Peter D., Richard Clément, Zoltán Dörnyei, and Kimberly A. Noels. 1998. 'Conceptualizing Willingness to Communicate in a L2: A Situational Model of L2 Confidence and Affiliation', The Modern Language Journal, 82.4: 545–62 <https://doi.org/10.1111/j.1540-4781.1998.tb05543.x>

Mishan, Freda. 2005. Designing Authenticity into Language Learning Materials (Bristol: Intellect)

Mitchell, Rosamund. 2020. 'Corpora and Instructed Second Language Acquisition', in The Routledge Handbook of Second Language Acquisition and Corpora, Handbooks in SLA, ed. by Magali Paquot and Nicole Tracy-Ventura (London: Routledge), pp. 252–64

Paquot, Magali. 2022. 'Corpora and Second Language Acquisition', in The Routledge Handbook of Corpora and English Language Teaching and Learning, Routledge Handbooks in Applied Linguistics, ed. by Reka R. Jablonkai and Eniko Csomay (London: Routledge), pp. 26–40 [accessed 28 February 2023]

Reinders, Hayo, and Sorada Wattana. 2014. 'Can I Say Something? The Effects of Digital Gameplay on Willingness to Communicate', Language Learning & Technology, 18.2: 101–23 <https://www.lltjournal.org/item/10125-44372/>

Ritchie, Mathy, and Catherine Black. 2012. 'Public Internet Forums: Can They Enhance Argumentative Writing Skills of Second Language Learners?', Foreign Language Annals, 45.3: 349–61 <https://doi.org/10.1111/j.1944-9720.2012.01203.x>

Russell, Victoria. 2020. 'Language Anxiety and the Online Learner', Foreign Language Annals, 53.2: 338–52 <https://doi.org/10.1111/flan.12461>

Staples, Shelley, and Julieta Fernández. 2019. 'Corpus Linguistics Approaches to L2 Pragmatics Research', in The Routledge Handbook of Second Language Acquisition and Pragmatics, ed. by Naoko Taguchi (New York: Routledge), pp. 241–54 <https://doi.org/10.4324/9781351164085>

Wain, Jennifer, Veronika Timpe-Laughlin, and Saerhim Oh. 2019. 'Pedagogic Principles in Digital Pragmatics Learning Materials: Learner Experiences and Perceptions', ETS Research Report Series, 2019.1: 1–21 <https://doi.org/10.1002/ets2.12270>

Wood, David, Jerome S. Bruner, and Gail Ross. 1976. 'The Role of Tutoring in Problem Solving', *Journal of Child Psychology and Psychiatry*, 17.2: 89–100 <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>

Louis Cotgrove is a researcher at the Leibniz Institute for the German Language in Mannheim, and specialises in corpus linguistic methods to analyse youth and online language. He can also be found programming APIs for lexical resources as part of the Germany-wide Text+ project (https://www.text-plus.org/).