Producing Standard German Subtitles for Swiss German TV Content

Johanna Gerlach, Jonathan Mutal, Pierrette Bouillon
FTI, University of Geneva, Switzerland
johanna.gerlach, jonathan.mutal, pierrette.bouillon@unige.ch

Abstract

In this study we compare two approaches (neural machine translation and edit-based) and the use of synthetic data for the task of translating normalised Swiss German ASR output into correct written Standard German for subtitles, with a special focus on syntactic divergences. Results suggest that NMT is better suited to this task and that relatively simple rule-based generation of synthetic data could be a valuable approach for cases where little training data is available and transformations are simple.

1 Introduction

In Switzerland, two thirds of the population speak Swiss German, which is primarily a spoken language with many regional dialects. Swiss German has no standardised written form (Honnet et al., 2018), thus written communication relies on Standard German. Swiss German is widely used on Swiss TV, for example in news reports, interviews or talk shows. In order to make these contents accessible to people who cannot understand spoken Swiss German, either due to hearing impairments, or because they only understand Standard German, these TV programs need to be subtitled in Standard German. For daily TV content, where large amounts of subtitles need to be produced within a short time frame and in a cost-effective manner, being able to automate the subtitling process would be advantageous. The PASSAGE project, which is the context of the present study, focuses on this task.

Subtitling can be automated by combining automatic speech recognition (ASR) with intralingual machine translation to improve the output to achieve compliance with subtitling standards (Buet and Yvon, 2021). In the PASSAGE project a first ASR step is used to produce a normalised transcription of spoken Swiss German, keeping the original syntax and expressions but only using Standard German words. In a second step, a neural machine translation (NMT) and an edit-based approach are explored to transform this normalised transcription into correct written Standard German (see Figure 1 for an example). To achieve this, multiple issues must be dealt with: ASR errors, incorrect detection of sentence boundaries, features related to spontaneous spoken language, such as dysfluencies or informal language, and finally the syntactic divergences between Swiss German and Standard German (Scherrer, 2011; Arabskyy et al., 2021).

Spoken Swiss German:
und d'Regierig hät no wiiteri Idee zum d'Stüüre abetue.

Normalised transcription (ideal ASR result):
Und die Regierung hat noch weitere Ideen zum die Steuern senken

Standard German:
Und die Regierung hat noch weitere Ideen, um die Steuern zu senken.

Figure 1: Example of the subtitling steps

In the present study, we focus on the second step, and more specifically on the systems' ability to transform Swiss German syntactic phenomena into their Standard German counterparts. We compare different approaches and investigate whether additional synthetic training data targeting these phenomena can improve the models. To evaluate the systems' performance on this task, we perform human evaluations of several test suites.

The paper is structured as follows, Section 2 introduces the syntactic phenomena we have focused on, Section 3 presents the data and architectures used, followed by Section 4 which describes the evaluation approach. Results are given in Section 5. Section 6 presents our conclusions and directions for future work.

Ninth Workshop on Speech and Language Processing for Assistive Technologies (SLPAT-2022), pages 37 - 43

Corpus	Segments	Words
GSW_NORM	98,126	2,630,824
DE (original subtitles)	101,150	1,414,744
DE_PE	20,634	347,232
GSW_NORM-DE	70,374	1,265,846 - 871,435
sDE_PE	4,418	94,194 - 94,065
sDE	13,896	223,146 - 221,944

Table 1: Overview of the data sets. GSW_NORM-DE was automatically aligned

2 Syntactic divergences between Swiss German and Standard German

The syntactic differences between Swiss German and Standard German can be classified into two main types: features related to the mainly spoken usage of Swiss German on one hand and dialect-specific features on the other (Scherrer, 2011). The latter are language phenomena involving among others the positioning of verbal forms, the construction of clauses or the use of cases and pronouns. These phenomena also differ from region to region (Glaser and Bart, 2021), thus the TV content, which includes transcripts of speakers from all regions of German speaking Switzerland, covers a large number of variations. For this study, we have focused on a subset of phenomena that occur in our corpora and that require different transformations:

- Adjective phrases with intensity adverbs often present different determiner usage than in Standard German, with the determiner placed after the adverb, or doubled. (advArtAdj)
- The verb *tun* 'do' used as an auxiliary with a trailing infinitive, referred to as *tun-periphrase*, is very common in many dialects and in spoken German, but is considered informal, and therefore is not used in subtitles. (tun)
- The particles *für* or *zum* are used to introduce final clauses instead of the Standard German complementiser *um ... zu* 'in order to'. (umZu)
- Reversed verb order compared to Standard German, often referred to as *verb raising* (for an overview, see Wurmbrand, 2017) occurs in different cases, e.g. in subordinate clauses the modal verb is placed before the infinitive, or the auxiliary precedes the participle. (verbsAuxPP and verbsModalInf)
- The uninflected particle *wo* is often used instead of nominative and accusative relative pronouns. (wo)

See Table 5 in the Appendix for examples.

3 Data and systems

In this section we describe the initial data that were provided to build the systems, the aligned and synthetic corpora that were derived from these data, and the different architectures that we have used.

3.1 Data

Table 1 summarises the corpora with the number of segments and words. Initially SRF (Schweizer Radio und Fernsehen) provided the following data for several TV shows:

GSW_NORM: normalised human transcriptions of Swiss German speech, keeping the original syntax and expressions but using German words. These data were created to train the Swiss German speech recogniser and correspond to an ideal ASR result.

DE: the original Standard German subtitles of the TV shows, not aligned with the transcriptions.

Based on these data, we created three aligned corpora used for system training:

GSW_NORM-DE_PE: this corpus was produced by manual post-editing of GSW_NORM into Standard German.

GSW_NORM-DE: this corpus was aligned automatically using (Plüss et al., 2021) modified to take as input GSW_NORM instead of speech. The alignment finds similar word chunks between GSW_NORM and DE which are then post processed to reconstruct sentences based on punctuation. The result has not been validated manually and therefore could contain errors.

sDE_PE and sDE: Since the training data for this task is scarce, we have chosen to generate synthetic parallel data specifically for the syntactic phenomena described in Section 2 (Lee and Seneff, 2008; Hassan et al., 2017; Lample et al., 2018). To this end, we have used the SpaCy toolkit's Matcher[1] to create transformation rules that identify syntactic patterns in Standard German text based on sequences of tokens, POS or morphological features, and transform these into the corresponding Swiss German patterns, e.g. by changing word order or verbs forms. We have applied these rules to the two available Standard German corpora, DE_PE and DE. Table 2 provides an overview of the synthetic data.

Finally, our project partner recapp [2] provided ASR output for a subset of the TV shows. This was used for the evaluations described in section 4.

3.2 Systems

In this study we compare the performance of four systems based on two approaches: NMT and edit-based.

NMT: Transformer architecture with copy attention that is usually used in tasks where small changes are needed (Gehrmann et al., 2018). We trained the system with GSW_NORM-DE and specialised it with GSW_NORM-DE_PE (as suggested in Sennrich and Zhang, 2019). The purpose of this approach is to use a larger corpus with low quality segments for training to increase vocabulary coverage (Poncelas and Way, 2019) and then to specialise with high quality segments to eliminate noise.

Ed: Edit-based system that predicts types of edits instead of words (see more, Berard et al., 2017). We trained the system using GSW_NORM-DE and GSW_NORM-DE_PE, but since we did not achieve an optimal loss, the final version was trained using only GSW_NORM-DE_PE.

sNMT and sEd: Same architectures as NMT and Ed respectively, with addition of the synthetic data after the post-edited data (DE_PE) used for system specialisation. (see similar approach for grammar error correction, Wang et al., 2021).

[1]https://spacy.io/api/matcher
[2]https://recapp.ch/

	DE_PE	DE
orig. segments	20,634	101,196
advAdjArt	15	676
tun	26	2,167
umZu	21	1,088
verbsAuxPP	148	5,373
verbsModalInf	1,083	4,525
wo	187	5,204
transformed	4,418	13,896

Table 2: Synthetic training data: number of segments in the original corpora used for extraction, number of occurrences of each phenomenon in the synthetic data, final number of segments transformed by the rules and included in the synthetic training data

4 Evaluation methodology

The objective of our systems is to convert as many Swiss German syntactic phenomena as possible into Standard German, while not introducing any additional errors into the ASR output. To assess the systems' performance, we have therefore performed two human evaluations, as described in the following sections.

4.1 Syntactic divergences

To evaluate the systems' ability to transform the syntactic phenomena described in Section 2 into their Standard German counterparts, we have created a set of test suites. Starting with a corpus of 5,000 segments of unseen real ASR output, we have extracted sets of examples for each phenomenon. The extraction was performed semi-automatically in a two step process. In the first step, we extended the work by (Haberkorn, 2022) using the SpaCy toolkit's Matcher. Hand-crafted rules describing simple patterns are used to extract candidate sentences for each phenomenon. This extraction is not entirely accurate since the ASR output contains recognition errors as well as features of spontaneous speech (e.g. repetitions or incomplete phrases) that cannot be taken into account by simple rules. Therefore, in a second step, the extracted candidates were manually validated by a native German speaker to build test suites for each phenomenon, keeping up to 50 segments per phenomenon.

After processing with the four systems (NMT, Ed, sNMT and sEd), these test suites were annotated by two native German speakers, to determine whether the phenomena had been transformed cor-

test suite (N)	NMT	sNMT	Ed	sEd
advArtAdj (50)	38 (76%)	44 (90%)	1 (2%)	39 (78%)
tun (50)	14 (28%)	12 (24%)	2 (4%)	2 (4%)
umZu (31)	8 (26%)	10 (32%)	0 (0%)	3 (10%)
verbsAuxPP (31)	23 (74%)	31 (100%)	1 (3%)	19 (63%)
verbsModalInf (50)	45 (90%)	47 (94%)	9 (18%)	10 (20%)
wo (50)	43 (86%)	44 (88%)	35 (70%)	30 (60%)

Table 3: Results of the human evaluation of the test suites: number and fraction of segments where the selected phenomenon was transformed correctly

rectly or not. In this evaluation, only the phenomenon of interest was considered, disregarding the remainder of the segment. Disagreements between the two judges were reevaluated in order to reach a final common judgement.

4.2 Relevance of the systems' modifications

To evaluate the models' ability to make only relevant modifications, we have created a test corpus by randomly selecting a subset of 54 segments from the unseen ASR data. These were processed with the four systems, then word-level edits made by the systems (deletions and insertions) were highlighted automatically and annotated manually by two native German speakers. Edits that improved the output or performed a change that did not adversely affect the output, e.g. by replacing a word by a synonym, were marked as correct; edits that degraded the output were marked as incorrect. When improvement of the output requires replacement of one word by another, e.g. when the particle *wo* should be replaced by a pronoun, a deletion must be paired with a correct insertion to be of use. In these cases we have counted the deletion as correct only if the corresponding insertion was present and correct. Based on the edit counts, we calculated a precision score as the fraction of correct edits among all edits performed by each system.

5 Results

5.1 Syntactic divergences

Results of the evaluation of the test suites are reported in Table 3. We observe large differences between the test suites, which strongly suggests that some phenomena are easier to identify and correct than others. The percentage of correct transformations is substantially higher for the phenomena that only require reordering (such as advArtAdj and the two verb phenomena) than for those that require transformation of individual words (tun). For the more complex transformations, e.g. the replacement of the *tun-periphrase*, we observe partially correct transformations, with changed word order but unchanged verb forms.

Overall the NMT systems outperform the edit-based systems, without and with the synthetic training data.

5.2 Relevance of the systems' modifications

Results of the evaluation of precision are reported in Table 4. Overall we observe that the two NMT systems make more than twice as many edits as the Ed systems. In terms of precision, the NMT systems outperform the Ed systems. Agreement between the two annotators is moderate (Cohen's Kappa 0.566), suggesting that annotation is difficult and possibly ambiguous. Often segments include multiple overlapping issues such as ASR errors and dysfluencies which make sentences difficult to understand and edits difficult to assess.

For both approaches, NMT and edit-based, the addition of targeted synthetic data reduces the total number of edits. For NMT, the percentage of correct edits is slightly increased, while for the edit-based approach it is about the same, showing that the addition of synthetic data does not degrade overall precision. Further analysis is required to see if this reduced number of edits is related to the order in which the corpora are used for specialisation.

6 Conclusion

In this study we have compared two architectures and the use of synthetic data for the task of translating normalised Swiss German ASR output into correct written Standard German, with a special focus on syntactic differences. In terms of syntactic transformations, the NMT systems outperform the edit-based systems. We observe large differences between the studied phenomena, some being transformed more successfully than others. For NMT,

	NMT	sNMT	Ed	sEd
Total edits	201	145	69	45
Correct	173 / 153	127 / 122	52 / 52	34 / 25
Precision	0.861 / 0.761	0.876 / 0.841	0.754 / 0.754	0.756/ 0.556
#Edits/#Words	15.9%	10.6%	6.3%	4.0%

Table 4: Word-level edits performed by the systems on the corpus of 54 segments (1214 words) with correct edits and precision for the two annotators

the addition of targeted synthetic training data improves the results, producing a larger number of transformed phenomena while also having a slight positive impact on precision. These results suggest that the relatively simple rule-based generation of training data could be a valuable approach for cases where little training data is available and transformations are simple (e.g. inversion, insertion or replacement).

While results are promising, this study presents several limitations. We have only studied a subset of the syntactic phenomena that distinguish Swiss German from Standard German. Additionally, due to the constraints of human evaluation, only a limited set of data could be included. In terms of synthetic training data, we have only aimed to reproduce the syntactic phenomena, but not the orality markers which are very frequent in the ASR output the systems need to deal with. Finally, the evaluation in this study was focused on system performance in terms of performed edits. An ongoing evaluation with different target groups will show whether these syntactic changes have an impact on understandability, accessibility and general satisfaction.

Future work includes extending to other phenomena and specialising with different settings.

Acknowledgements

This project has received funding from the Initiative for Media Innovation based at Media Center, EPFL, Lausanne, Switzerland. We would also like to thank Melanie Arnold for their contribution to data annotation.

References

Yuriy Arabskyy, Aashish Agarwal, Subhadeep Dey, and Oscar Koller. 2021. Dialectal speech recognition and translation of swiss german speech to standard german text: Microsoft's submission to swisstext 2021 (short paper). In *Proceedings of the Swiss Text Analytics Conference 2021, Winterthur, Switzerland, June 14-16, 2021 (held online due to COVID19 pandemic)*, volume 2957 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Alexandre Berard, Laurent Besacier, and Olivier Pietquin. 2017. Lig-cristal submission for the wmt 2017 automatic post-editing task. In *Proceedings of the Second Conference on Machine Translation*, page 623–629. Association for Computational Linguistics.

François Buet and François Yvon. 2021. Vers la production automatique de sous-titres adaptés à l'affichage. In *Traitement Automatique des Langues Naturelles*, pages 91–104, Lille, France. ATALA.

Sebastian Gehrmann, Yuntian Deng, and Alexander Rush. 2018. Bottom-up abstractive summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, page 4098–4109. Association for Computational Linguistics.

Elvira Glaser and Gabriela Bart. 2021. *Syntaktischer Atlas der deutschen Schweiz (SADS)*. A. Francke Verlag.

Veronika Christine Haberkorn. 2022. Automatic post-editing of subtitles - rule-based post-editing of subtitles from Swiss German to Standard German. Master's thesis, Faculty of translation and interpreting, University of Geneva.

Hany Hassan, Mostafa Elaraby, and Ahmed Tawfik. 2017. Synthetic data for neural machine translation of spoken-dialects. In *Proceedings of the 14th International Workshop on Spoken Language Translation*, pages 82–89, Tokyo, Japan.

Pierre-Edouard Honnet, Andrei Popescu-Belis, Claudiu Musat, and Michael Baeriswyl. 2018. Machine translation of low-resource spoken dialects: Strategies for normalizing swiss german. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. European Language Resources Association (ELRA).

Guillaume Lample, Alexis Conneau, Ludovic Denoyer, and Marc'Aurelio Ranzato. 2018. Unsupervised machine translation using monolingual corpora only. In *International Conference on Learning Representations*.

John Lee and Stephanie Seneff. 2008. Correcting misuse of verb forms. In *Proceedings of ACL-08: HLT*,

pages 174–182, Columbus, Ohio. Association for Computational Linguistics.

Michel Plüss, Lukas Neukom, Christian Scheller, and Manfred Vogel. 2021. Swiss parliaments corpus, an automatically aligned swiss german speech to standard german text corpus. In *Proceedings of the Swiss Text Analytics Conference 2021, Winterthur, Switzerland, June 14-16, 2021 (held online due to COVID19 pandemic)*, volume 2957 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Alberto Poncelas and Andy Way. 2019. Selecting artificially-generated sentences for fine-tuning neural machine translation. In *Proceedings of the 12th International Conference on Natural Language Generation*, page 219–228. Association for Computational Linguistics.

Yves Scherrer. 2011. Syntactic transformations for Swiss German dialects. In *Proceedings of the First Workshop on Algorithms and Resources for Modelling of Dialects and Language Varieties*, pages 30–38, Edinburgh, Scotland. Association for Computational Linguistics.

Rico Sennrich and Biao Zhang. 2019. Revisiting low-resource neural machine translation: A case study. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, page 211–221. Association for Computational Linguistics.

Yu Wang, Yuelin Wang, Kai Dang, Jie Liu, and Zhuo Liu. 2021. A comprehensive survey of grammatical error correction. *ACM Transactions on Intelligent Systems and Technology*, 12(5):1–51.

Susi Wurmbrand. 2017. Verb clusters, verb raising, and restructuring. In *The Wiley Blackwell Companion to Syntax, Second Edition*, pages 1–109. John Wiley & Sons, Ltd.

A Appendix

Phenomenon	Example	Conversion
advArtAdj	ein Land, wo **sehr einen** hohen Standard hat Punkto Sicherheit [...] → ein Land, wo **einen sehr** hohen Standard hat Punkto Sicherheit [...]	Reverse the order of adverb *sehr* and article *einen*
	Diese Mitarbeiter haben **einen sehr einen** hohen Ausbildungsstand → Diese Mitarbeiter haben **einen sehr** hohen Ausbildungsstand	Remove the doubled article *einen*
tun	Man **tut** sich solchen Fragen sicher nicht **verschliessen** → Man **verschliesst** sich solchen Fragen sicher nicht	Replace *tun* by the finite verb form *verschliesst* of the infinitive *verschliessen*
umZu	Man braucht eine Ausbildung **zum** sich können ablösen und von der Sozialhilfe wegkommen. → Man braucht eine Ausbildung, **um** sich ablösen **zu** können und von der Sozialhilfe wegkommen.	Replace the particle *zum* by the complementiser *um ... zu*
verbsAuxPP	Freunde wo in der Intensivpflegestationen **sind gewesen** [...] → Freunde wo in der Intensivpflegestationen **gewesen sind** [...]	Reverse order of auxiliary *sind* and participle *gewesen*
verbsModalInf	Wir haben da das Gefühl gehabt, man muss den Leuten sagen, was man **kann machen** [...] → Wir haben da das Gefühl gehabt, man muss den Leuten sagen, was man **machen kann** [...]	Reverse order of modal *kann* and infinitive *machen*
wo	zum Beispiel diese Leute **wo** gelitten haben bei dem an Bergsturz → zum Beispiel diese Leute **die** gelitten haben bei dem an Bergstur	Replace uninflected particle *wo* by relative pronoun *die* that agrees in number and gender with the noun *Leute* to which it refers

Table 5: Examples of conversions from Swiss German patterns to Standard German patterns for the syntactic divergences included in the study. Examples are extracted from the test-suites. Only the sequences in bold have been edited, errors may subsist in the remainder of the segments.