Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 5063–5070
Marseille, 20-25 June 2022

Standard German Subtitling of Swiss German TV content: the PASSAGE Project

Jonathan Mutal, Pierrette Bouillon, Johanna Gerlach, Veronika Haberkorn
Faculty of translation and interpreting, University of Geneva
40, bd du Pont d'Arve, 1211 Geneva, Switzerland
{jonathan.mutal, pierrette.bouillon, johanna.gerlach}@unige.ch, veronika.haberkorn@etu.unige.ch

Abstract

In Switzerland, two thirds of the population speak Swiss German, a primarily spoken language with no standardised written form. It is widely used on Swiss TV, for example in news reports, interviews or talk shows, and subtitles are required for people who cannot understand this spoken language. This paper focuses on the task of automatic Standard German subtitling of spoken Swiss German, and more specifically on the translation of a normalised Swiss German speech recognition result into Standard German suitable for subtitles. We compared different statistical and deep learning machine translation systems for this task. We also produced an aligned corpus of normalised Swiss German and Standard German subtitles. Results of two evaluations, automatic and human, show that the systems succeed in improving the content, but are currently not capable of producing entirely correct Standard German.

Keywords: automatic subtitling, Swiss German, machine translation, automatic post-editing

1. Introduction

In Switzerland, two thirds of the population speak Swiss German, thus this language is widely used on Swiss TV, for example in news reports, interviews or talk shows. Swiss German is primarily a spoken language, with many regional dialects and no standardised written form (Honnet et al., 2018). In order to make these Swiss German 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 "Sous-titrage automatique du suisse allemand en allemand standard"[1], a Swiss project financed by IMI ("Initiative for Media Innovation"), focuses on this task.

One way to automate subtitling is to combine a speech recognition system with an intralingual machine translation (MT) system. In this process, MT can be used to different ends, for example correcting speech recognition issues or transforming content to achieve compliance with subtitling standards (Buet and Yvon, 2021). In this study, we specifically focus on the translation of normalised Swiss German speech recognition output into Standard German and explore different MT architectures to deal with the divergences between these two languages (Scherrer, 2011). Figure 1 illustrates the complete subtitling pipeline. The first step is automatic speech recognition (ASR) of Swiss German (GSW) to produce a normalised transcription (GSW_REC), keeping the original syntax and expressions, but using German words (Arabskyy et al., 2021) since there is no standardised written form for Swiss German. This is followed in a second step by MT into Standard German (DE). Our contributions to this pipeline concern the second step and are therefore the following:

- a comparison of different statistical and deep learning approaches to translate normalised Swiss German into Standard German subtitles
- an aligned corpus of normalised human Swiss German transcripts and Standard German subtitles

This article is structured as follows: we begin by describing the data used for this study (Section 2) and presenting the different MT architectures and models (Section 3). We continue with two evaluations, automatic and human, to compare the different architectures' performance on a normalised Swiss German human transcriptions (Sections 4.1 and 4.2). This is followed by a section presenting the results of a preliminary evaluation using real ASR output (Section 5). Section 6 concludes and outlines future work.

2. Data

The data were provided by SRF (Schweizer Radio und Fernsehen) and consist of:

- GSW_NORM: normalised human transcriptions of TV shows. These data were created to train the Swiss German recogniser and correspond to an ideal ASR result.
- DE: the unaligned original Standard German subtitles of the TV shows.

Based on these data, we produced two aligned corpora:

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

[1] https://www.media-initiative.ch/project/subtitling-of-swiss-german-into-standard-german-automatic-post-editing/

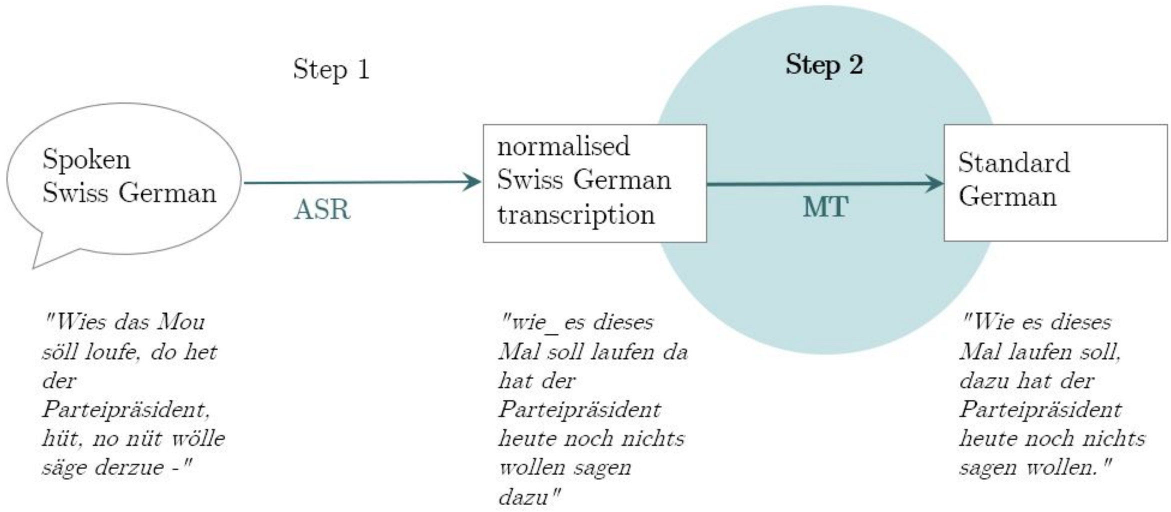

Figure 1: Overview of the subtitling pipeline

Transcription (GSW_NORM)	**Original Standard German Subtitle (DE)**
Weil	
dort steht eigentlich, was man mit dem, also was	Dort steht, was man tun muss.
man eigentlich muss machen. Also zum Beispiel, dass man ähm die neue Pensionskasse muss angeben.	Man muss z.B. seine neue Pensionskasse melden.
Oder	
wenn man jetzt zum Beispiel nicht gerade wieder geht gehen arbeiten, ähm in was für eine Freizügigkeitseinrichtung das Geld soll hin.	Wenn man nicht sofort wieder arbeiten geht, muss man angeben,an welche Freizügigkeitseinrichtung das Geld ausbezahlt werden soll.
Das ist ein so ein riesiges Volumen.	Das ist ein riesiges Volumen.

Table 1: Examples of transcriptions automatically aligned with the original subtitles

- 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 then finds similar chunks of words between GSW_NORM and DE. The results of the alignment is shown in Table 1. The alignment has not been manually validated and therefore could contain errors.

These data allow us to focus on the divergences between spoken Swiss German and written Standard German. To translate the human transcriptions (GSW_NORM) into Standard German (DE_PE), the post-editors performed different transformations. A number of these were related to Swiss German word order, which differs from Standard German, for example for the position of modal verbs. Other frequent divergences lie in the combination of prepositions with cases, or the use of some subordinating conjunctions. Beyond the correction of phenomena specific to Swiss German, the post-editors also corrected issues related to spoken language, such as interjections or disfluencies, as well as grammatical errors. Table 2 shows examples of some transformations and table 3 summarises the data with the number of segments and words.

3. Architecture

In this section, we describe the different approaches used to automatically translate GSW_NORM into Standard German. We trained a baseline, three SMT systems and two NMT systems.

3.1. Baseline

systemSMT_baseline: Phrase-based machine translation (Koehn et al., 2003, PBMT) system, trained with GSW_NORM-DE data.

3.2. SMT systems

systemSMT_bigLM: trained with the GSW_NORM-DE data, but fine-tuned with the post-edited data (GSW_NORM-DE_PE). We used the German OpenSubtitles2018 corpus (Lison et al., 2018) to train the language model.

systemSMT_backTranslation: Same system as systemSMT_baseline, but we added back-translated data to the GSW_NORM-DE data. The post-edited (DE_PE) segments were back-translated (Feng et al., 2021) using English as a pivot language, a method

Transformation	GSW_NORM	DE_PE	
Place modal after infinitive	Also, der einzige Ort, wo ich würde gehen ist Spanien.	Also, der einzige Ort, wo ich hingehen würde, ist Spanien.	'The only place I would go to is Spain'
Change temporal subordinating conjunction	Wir haben es auch gesehen das letzte Jahr, wo ein Putschversuch [...]	Wir haben es auch im letzten Jahr gesehen, als ein Putschversuch [...]	'We also observed it last year, when a coup attempt [...]
Disfluencies	die inländischen produ- ähm Produzenten geschützt sind	die inländischen Produzenten geschützt sind	'the domestic producers are protected'

Table 2: Examples of transformations performed by the post-editors on the transcriptions

Data	Segments	Words
DE_PE	21,097	347,232
DE (original subtitles)	119,150	1,414,744
GSW_NORM	115,126	2,630,824
GSW_NORM-DE	87,923	1,265,846 (source) -871,435 (target)

Table 3: Number of segments and words of the data sets. GSW_NORM-DE was automatically aligned

commonly used to generate unstructured data (Hedderich et al., 2021).

systemSMT_filter: Same system as systemSMT_baseline, but we filtered the translation model by automatically removing misaligned segments (15,000 segments removed), using the word-normalised Levenshtein metric (Johnson et al., 2007).

3.3. Neural Machine Translation systems

systemNMT: Neural Machine Translation (NMT) architecture, usually used for automatic summarising tasks (Gehrmann et al., 2018, Transformer with copy attention). We trained the system with GSW_NORM-DE and specialised with GSW_NORM-DE_PE. The idea is to train with more vocabulary and then specialise with the corrections made by the post-editors.

systemAPE: Model with a task-specific attention mechanism, which is particularly recommended in a scenario with little data. The system predicts the type of edits instead of the word (insertion or deletion of a word, substitutions or keeping the source word), see more (Berard et al., 2017). At first, we trained a system using GSW_NORM-DE and specialised it with GSW_NORM-DE_PE without reaching neither an optimal loss nor a pertinent accuracy during the training step. We then decided to only use GSW_NORM-DE_PE for training. A possible explanation is that there were too many differences between the normalised transcriptions and the original subtitles for the model to learn appropriate edits.

In the following sections, we describe how the systems were evaluated, first using the normalised Swiss German transcriptions provided by SRF, then using ASR output.

4. Evaluation on normalised transcriptions

For our first evaluation, we use the human normalised transcriptions (GSW_NORM), which simulate a perfect speech recognition output. This enables us to estimate the performance of the models on ideal data. In order to compare the different architectures, we carried out an automatic and human evaluation. The automatic evaluation aims at giving an overview of the quality of the systems and the number of modifications made by each system. The human evaluation aims at understanding whether the changes were useful.

In the following sections, we present the automatic and human evaluations, with the results.

4.1. Automatic Evaluation

4.1.1. Design

For the automatic evaluation, we built two test sets using 2,000 consecutive segments extracted from GSW_NORM. For the first test set (PE_test), we use the post-edited version (PE_DE) as reference, to measure the systems' ability to post-edit GSW_NORM to produce Standard German. For the second (DE_test), we used the corresponding real subtitles (DE) as a reference and aimed at quantifying the systems' ability to produce sentences that are close to the official subtitles. These two test-corpora contain the same sentences segmented differently (see Table 4), since the segmentation is not the same in the GSW_NORM and DE corpora.

We used the HTER, TER (Snover et al., 2006) and BLEU (Papineni et al., 2001) metrics. The HTER score allows us to quantify the post-editing effort, in this case the number of edits carried out by the systems; the TER and BLEU scores quantify the similarity with the reference text. We also calculated the proportion of exact matches on the sentence level between system output

Test set	Segments	Data	Evaluation
PE_test	2,000	GSW_NORM - PE_DE	automatic
DE_test	2,000	GSW_NORM - DE	automatic

Table 4: Overview of the test sets. GSW_NORM - DE was automatically aligned

System	BLEU	TER	HTER	Exact match
systemSMT_baseline	46.79	33.43	31.20	6.4%
systemSMT_bigLM	50.80	32.73	15.98	10.0%
systemSMT_backTranslation	44.02	35.02	18.92	6.0%
systemSMT_filter	58.88	25.62	10.41	14.2%
systemNMT	64.91	23.30	22.59	16.9%
systemAPE	61.49	24.37	12.69	15.0%

Table 5: Results for the PE_test test set with manually post-edited transcriptions as reference

and reference.

4.1.2. Results

Table 5 shows the results for the first test set (PE_test). We observe that the neural systems (NMT and APE) achieve the best BLEU and TER scores and outperform the best statistical system (systemSMT_filter). They also produce the highest proportion of exact matches. As the baseline statistical system (systemSMT_baseline) was trained with the automatically aligned corpus (GSW_NORM-DE), it makes the most changes (highest HTER score). The systems specialised with GSW_NORM-DE_PE however produce less changes, since the post-edited corpora are the result of a minimal post-edition. Although these systems make less changes, systemSMT_filter makes the least changes. This can be explained by the removal of the misaligned segments from the training data. The APE system achieves the lowest HTER score of the neural architectures, since it does not perform a real translation, but rather focuses on specific edits.

Table 6 shows the results for the second test set (DE_test). Overall, scores are worse than for the first evaluation, indicating that the system output is not close to the original subtitles. The statistical system SMT_bigLM produces the most exact matches. systemSMT_filter achieves the best BLEU and TER scores. The HTER scores cannot be compared with those of the first evaluation, since the segments are not the same, but they follow almost the same trend, with the APE system making the least modifications among the neural systems, and systemSMT_filter making the least modifications among the SMT systems. The baseline statistical system still obtains the highest score on HTER.

4.2. Human Evaluation

The human evaluation was designed to answer two questions (Mutal et al., 2019), namely 1) whether the modifications performed by our systems would be considered as improvements by German native speakers, and 2) whether these modifications are sufficient to produce correct Standard German, or if further changes are required.

4.2.1. Design

For these evaluations, we only used results from the two best MT systems, namely sytemNMT and systemAPE, with DE_PE as reference. Both evaluations were carried out on segment level.

To evaluate whether individual transformations produced by the systems were an improvement, we presented the normalised transcription (GSW_NORM) side by side with the system output, with differences highlighted in colour. The sentences were shown with their context, i.e. preceding and following sentences. For each sentence pair, participants were asked to indicate whether the modification performed by the system was necessary and correct. Figure 2 shows an example of a segment given to the evaluators.

To evaluate whether the transformations performed by the systems were sufficient to produce correct Standard German and to see if the post-editors performed overcorrection (do Carmo et al., 2021), we presented the system output side by side with the post-edited equivalent (DE_PE) to serve as reference, again with differences highlighted in colour. Participants were asked to indicate whether the modification in DE_PE was necessary.

The test sets for these evaluations were built by randomly selecting 49 segments with modifications from the test set used in the automatic evaluation. Sentences with multiple modifications were duplicated in order to evaluate one modification at a time. In total, the test set used for the first evaluation contained 60 segments with one modification each and the second 69. All the segments were extracted with their context (previous and following sentences) to allow evaluation in context.

Both evaluations were done by the same participants (7 for the evaluation with GSW_NORM and 5 for the evaluation with DE_PE). They were all native speakers of Standard German with no familiarity with Swiss German.

System	BLEU	TER	HTER	Exact match
systemSMT_baseline	28.24	59.25	40.26	6.4%
systemSMT_bigLM	31.21	56.00	29.68	14.4%
systemSMT_backTranslation	28.24	59.25	32.26	6.2%
systemSMT_filter	33.71	56.67	22.05	9.7%
systemNMT	30.81	59.19	31.50	8.2%
systemAPE	28.25	61.73	19.62	7.0%

Table 6: Results for the DE_test test-set, with original subtitles as reference

Figure 2: An example of a segment given to the evaluators

4.2.2. Evaluation of improvements

Table 7 presents the results of the human evaluation comparing GSW_NORM with the system output.

System	Modification necessary	Modification correct
MT	50/60 (83%)	45/60 (75%)
APE	59/60 (98%)	54/60 (90%)

Table 7: Human evaluation of modifications

Considering majority judgements (4 or more of the 7 evaluators agree), we observe that nearly all changes performed by the APE are considered as necessary by the evaluators (98.3%), while 16.7% of those produced by the MT approach were rejected. When considering the results for the two approaches combined, agreement between annotators is moderate for this task (Light's Kappa 0.571). However, calculation of distinct Kappa scores for each of the approaches reveals that evaluators agree more often for the MT approach than for the APE approach (0.63 vs 0.289). This could be explained by the fact that APE sometimes makes changes that improve the sentence but do not entirely fix the issue, resulting in items that are difficult to evaluate systematically. For both systems, five segments included modifications that were judged as necessary but incorrect, i.e. the word or phrase modified by the system was indeed incorrect in GSW_NORM, but the modification performed was not entirely correct.

4.2.3. Evaluation of final quality

System	Segments requiring additional modifications
MT	67/69 (97%)
APE	67/69 (97%)

Table 8: Human evaluation of final quality

Results of the evaluation comparing the system output with the post-edited transcriptions (DE_PE) reported in Table 8 strongly suggest that the output of our systems requires further editing to become fully correct Standard German. For most of the segments, the majority of evaluators considered the differences between system output and DE_PE to be necessary modifications. However, the presentation of the two versions side by side may have influenced the evaluation results (Läubli et al., 2020), as some of the modifications judged necessary may not have been obvious if the evaluators had only been presented with the system output, without any reference. Agreement on this task was lower than for the first task (Light's Kappa 0.337).

5. Evaluation on ASR data

In order to see if the models are able to generate subtitles from the speech recognition results, we carried out a small automatic evaluation using the two best systems, i.e. systemNMT and systemAPE. To do so, we manually aligned 4'926 sentences from ASR output to DE to serve as reference. We then calculated BLEU on the ASR and the system outputs. Both system outputs achieve a higher BLEU score than the raw ASR output (22.06 and 18.56 for systemNMT and systemAPE against 17.73 for ASR). This shows that the post-processing with MT brings the content closer to the reference.

6. Conclusion

The aim of this paper was to see if we could build a useful machine translation system to improve the quality of normalised Swiss German subtitles, using a corpus of human normalised Swiss German transcriptions aligned with the post-edited version in Standard German for training or specialisation. The systemAPE and systemNMT systems obtain the best BLEU and TER scores, with the DE_PE corpus as reference. However, human evaluation shows that the APE is the most precise. NMT systems make more changes, but not all are necessary and/or correct.

Although our corpora are not large enough to train a neural architecture that produces entirely correct Standard German, we demonstrated that our models improved both human transcription and ASR data and were able to learn some of the main divergences between the two languages (word order, lexical differences, etc.).

All these results suggest that the APE system is already good enough to help human post-editors in the task of post-editing normalised Swiss German speech recognition output and producing new parallel data. Future work in this project will focus on new types of edits such as compression for simplified subtitles and correction of speech recognition errors. Finally, it would also be interesting to measure the impact of the alignment (manually aligned data vs automatic alignment) and training data (ASR data vs normalised human transcriptions).

7. Acknowledgements

This project has received funding from the Initiative for Media Innovation based at Media Center, EPFL, Lausanne, Switzerland.

Bibliographical References

Arabskyy, Y., Agarwal, A., Dey, S., and Koller, O. (2021). Dialectal Speech Recognition and Translation of Swiss German Speech to Standard German Text: Microsoft's Submission to SwissText 2021. *arXiv:2106.08126 [cs, eess]*, Jul. arXiv: 2106.08126.

Berard, A., Besacier, L., and Pietquin, O. (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.

Buet, F. and Yvon, F. (2021). Vers la production automatique de sous-titres adaptés à l'affichage. In Pascal Denis, et al., editors, *Traitement Automatique des Langues Naturelles*, pages 91–104, Lille, France. ATALA.

do Carmo, F., Shterionov, D., Moorkens, J., Wagner, J., Hossari, M., Paquin, E., Schmidtke, D., Groves, D., and Way, A. (2021). A review of the state-of-the-art in automatic post-editing. *Machine Translation*, 35(2):101–143, Jun.

Feng, S., Gangal, V., Wei, J., Chandar, S., Vosoughi, S., Mitamura, T., and Hovy, E. (2021). A Survey of Data Augmentation Approaches for NLP. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, page 968–988. Association for Computational Linguistics.

Gehrmann, S., Deng, Y., and Rush, A. (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.

Hedderich, M. A., Lange, L., Adel, H., Strötgen, J., and Klakow, D. (2021). A Survey on Recent Approaches for Natural Language Processing in Low-Resource Scenarios. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, page 2545–2568. Association for Computational Linguistics.

Honnet, P.-E., Popescu-Belis, A., Musat, C., and Baeriswyl, M. (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), May.

Johnson, H., Martin, J., Foster, G., and Kuhn, R. (2007). Improving Translation Quality by Discarding Most of the Phrasetable. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, page 967–975. Association for Computational Linguistics, Jun.

Koehn, P., Och, F. J., and Marcu, D. (2003). Statistical phrase-based translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - NAACL '03*, volume 1, page 48–54. Association for Computational Linguistics.

Läubli, S., Castilho, S., Neubig, G., Sennrich, R., Shen, Q., and Toral, A. (2020). A Set of Recommendations for Assessing Human–Machine Parity in Language Translation. *Journal of artificial intelligence research*, 67:653–672, March. Publisher Copyright: © 2020 AI Access Foundation. All rights reserved.

Lison, P., Tiedemann, J., and Kouylekov, M. (2018).

OpenSubtitles2018: Statistical Rescoring of Sentence Alignments in Large, Noisy Parallel Corpora. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. European Language Resources Association (ELRA), May.

Mutal, J., Volkart, L., Bouillon, P., Girletti, S., and Estrella, P. (2019). Differences between SMT and NMT Output - a Translators' Point of View. In *Proceedings of the Human-Informed Translation and Interpreting Technology Workshop (HiT-IT 2019)*, pages 75–81, Varna, Bulgaria, September. Incoma Ltd., Shoumen, Bulgaria.

Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2001). BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics - ACL '02*, page 311. Association for Computational Linguistics.

Plüss, M., Neukom, L., Scheller, C., and Vogel, M. (2021). Swiss Parliaments Corpus, an Automatically Aligned Swiss German Speech to Standard German Text Corpus. *arXiv:2010.02810 [cs]*, Jun. arXiv: 2010.02810.

Scherrer, Y. (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, July. Association for Computational Linguistics.

Snover, M., Dorr, B., Schwartz, R., Micciulla, L., and Makhoul, J. (2006). A Study of Translation Edit Rate with Targeted Human Annotation. In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers*, page 223–231. Association for Machine Translation in the Americas, Aug.

Appendix

Normalised Swiss German transcription	systemNMT	systemAPE	Reference (DE_PE)
also ähm müsste man sie jetzt einfach nur noch zwingen, s- auch es tatsächlich zu praktizieren.	also müsste man sie jetzt einfach nur noch zwingen, auch es tatsächlich zu praktizieren.	Also müsste man sie jetzt einfach nur noch zwingen, auch es tatsächlich zu praktizieren.	Also müsste man sie jetzt einfach nur noch zwingen, es auch tatsächlich zu praktizieren.
[...] das heisst der kunde nimmt das geld bar	[...] das heisst der Kunde nimmt das Geld bar	[...] das heisst, der Kunde nimmt das Geld bar.	[...] das heisst, der Kunde nimmt das Geld in bar,
Also zum Beispiel, dass man ähm die neue Pensionskasse muss angeben.	Also zum Beispiel, dass man die neue Pensionskasse angeben muss.	Also zum Beispiel, dass man die neue Pensionskasse angeben muss.	Also zum Beispiel, dass man die neue Pensionskasse angeben muss.
Und da muss ich sagen das Nachtfahrverbot das ist diskussionslos.	Und da muss ich sagen das Nachtfahrverbot ist diskussionslos.	Und da muss ich sagen, das Nachtfahrverbot ist diskussionslos.	Und da muss ich sagen, dass das Nachtfahrverbot diskussionslos ist.
Heute ist glaube ich die Zeit vom Mitdiskutieren, Mitreden ähm , zusammen Lösungen finden.	Heute ist glaube ich die Zeit des Mitdiskutieren, Mitreden, zusammen Lösungen finden.	Heute ist glaube ich die Zeit vom Mitdiskutieren, Mitreden, zusammen Lösungen finden.	Ich glaube, heute ist die Zeit des Mitdiskutierens, Mitredens, der gemeinsamen Findung von Lösungen.
Es hat 6 Jugendliche gehabt . Aktuell sind wir am reinigen von der Wohngruppe.	Es hat 6 Jugendliche gehabt. Aktuell sind wir am reinigen von der Wohngruppe.	Es hat 6 Jugendliche gehabt. Aktuell sind wir am reinigen der Wohngruppe.	Es hatte 6 Jugendliche. Aktuell reinigen wir die Wohngruppe.
Die Räumlichkeiten die werden nicht genutzt, wie es vorgesehen ist	Die Räumlichkeiten werden nicht genutzt, wie es vorgesehen ist.	Die Räumlichkeiten werden nicht genutzt, wie es vorgesehen ist,	Die Räumlichkeiten werden nicht genutzt, wie es vorgesehen ist.
und ähm auch sogar Leute darunter, wo sagen nein ich weigre, mich Abfall zu produzieren.	und auch sogar Leute darunter, die sagen nein ich weigre, mich Abfall zu produzieren.	und auch sogar Leute darunter, die sagen: "Ich ich weigre, mich Abfall zu produzieren.	Und es gibt sogar auch Leute, die sagen, nein ich weigere mich, Abfall zu produzieren.

Table 9: Examples of transformations performed by the systems on the normalised Swiss German transcriptions.