

# Speech Resources in the Tamasheq Language

Marcely Zanon Boito<sup>1</sup>, Fethi Bougares<sup>2</sup>, Florentin Barbier<sup>3</sup>, Souhir Gahbiche<sup>3</sup>,  
Loïc Barrault<sup>2</sup>, Mickael Rouvier<sup>1</sup>, Yannick Estève<sup>1</sup>

<sup>1</sup>LIA - Avignon University, France

<sup>2</sup>LIUM - Le Mans University, France

<sup>3</sup>Airbus - France

**contact:** {marcely.zanon-boito, yannick.esteve}@univ-avignon.fr

## Abstract

In this paper we present two datasets for Tamasheq, a developing language mainly spoken in Mali and Niger. These two datasets were made available for the IWSLT 2022 low-resource speech translation track, and they consist of collections of radio recordings from daily broadcast news in Niger (Studio Kalangou) and Mali (Studio Tamani). We share (i) a massive amount of unlabeled audio data (671 hours) in five languages: French from Niger, Fulfulde, Hausa, Tamasheq and Zarma, and (ii) a smaller 17 hours parallel corpus of audio recordings in Tamasheq, with utterance-level translations in the French language. All this data is shared under the Creative Commons BY-NC-ND 3.0 license. We hope these resources will inspire the speech community to develop and benchmark models using the Tamasheq language.

**Keywords:** speech corpus, speech translation, tamasheq, zarma, hausa, fulfulde, french

## 1. Introduction

The vast majority of speech pipelines are developed for and in *high-resource* languages, a small percentage of languages for which there is a large amount of annotated data freely available (Joshi et al., 2020). This not only limits the investigation of language impact in current pipelines, as the applied languages are usually from the same subset, but it also fails to reflect the real-world performance these approaches will have in diverse and smaller datasets.

In recent years, the IWSLT evaluation campaign<sup>1</sup> introduced a low-resource speech translation track focused on developing and benchmarking translation tools for under-resourced languages. While for a vast majority of these languages, there is not enough parallel data for training large translation models, in these cases we might still have access to limited disparate resources, such as word-level translations, small parallel textual data, monolingual texts and recordings. This track of the IWSLT campaign thus focuses on leveraging these different kinds of data for building effective translation systems under realistic settings.

In this paper we present the resources in the Tamasheq language we share in the context of the IWSLT 2022: low-resource speech translation track. Tamasheq is a variety of Tuareg, a Berber macro-language spoken by nomadic tribes across North Africa in Algeria, Mali, Niger and Burkina Faso (Heath, 2006). It accounts for approximately 500,000 native speakers, being mostly spoken in Mali and Niger (Ethnologue: Languages of the World, 2021).

We share a large audio corpus, made of 224 hours of Tamasheq, together with 417 hours in other four languages of Niger (French, Fulfulde, Hausa and Zarma). We also share a smaller corpus of 17 hours

of Tamasheq utterances aligned with French translations. We hope that these resources will represent an interesting use-case for the speech community, allowing them to not only develop low-resource speech systems in Tamasheq, but also to investigate the leveraging of unannotated audio data in diverse languages that co-exist in the same geographic region.

This paper is organized as follows. Section 2 presents the source content of the data shared: thanks to the *Fondation Hironnelle Initiative* and local partners, we are able to collect broadcast news in diverse African languages. Section 3 presents the small Tamasheq-French parallel corpus, and Section 4 presents the collection of unannotated audio data in French from Niger, Fulfulde, Hausa, Tamasheq and Zarma. Finally, Section 5 presents a speech translation baseline model for the IWSLT 2022 campaign, and Section 6 concludes this work.

## 2. The source content: The Fondation Hironnelle Initiative

The Fondation Hironnelle<sup>2</sup> is a Swiss non-profit organization founded in 1995 by journalists, with the goal of supporting local independent media in areas of social unrest. They produce and broadcast information and talk shows in different countries, providing local partners with editorial, managerial and structural support and training to function in a sustainable manner.

In this work we focus on their daily radio broadcasts episodes, produced and broadcast by local partners in different languages. These allow the local communities to get informed in their own dialects, in contrast to mainstream media that tends to cover only the countries' official languages. For the Tamasheq lan-

<sup>1</sup><https://iwslt.org/2022/low-resource>

<sup>2</sup><https://www.hironnelle.org/en/>

guage, we find these episodes being produced daily in Mali (Studio Tamani<sup>3</sup>) and Niger (Studio Kalangou<sup>4</sup>).

**Speech Style and Quality.** The radio episodes are recorded in local studios: for each episode, one or two hosts present the news, and often interviews and advertisements are included. Most of the speech is of good quality, with rare instances of background music during advertisements. For interviews, we notice some cases of overlapping speech, mainly when simultaneous translation is performed, and background noise such as outdoor sounds.

**Audio Web-crawling.** With the authorization of the Fondation Hironnelle and partners, we downloaded the .mp3 episodes by generating URLs from the local partners' broadcast webpages.<sup>5</sup> The corpora presented in Section 3 and Section 4 use these audio files as source content.

### 3. The Tamasheq-French Parallel Corpus

This corpus corresponds to 17 hours of *controlled* speech utterances, with manual translations to the French language. We also share a longer version of this corpus, including 2 additional hours of potentially noisy speech segments. We detail below the steps for creating this corpus, and present general statistics.

**1. Data Downloading.** 100 episodes were downloaded from the Studio Kalangou website in February 2019: 23 episodes from 2016, 36 episodes from 2017 and 2018, and 5 episodes from 2019. This results in 25 hours of raw speech, with an average episode duration of 15 minutes.

**2. Translation Process.** We commissioned ELDA (Evaluations and Language resources Distribution Agency)<sup>6</sup> for translating the 25 hours of Tamasheq into French text. No transcriptions were commissioned. The translations were produced by at least two native Tamasheq speakers,<sup>7</sup> with posterior text correction by proficient French speakers. The translators had access to 5 pages of guidelines, including segmentation guidelines for slicing the episodes into utterances. Annotation used the *Transcriber* open-source tool.<sup>8</sup> Lastly, some utterances contain gender annotation and speaker identification. Unfortunately, this annotation was not standardized across the different translators, and therefore some speakers are referred by different speaker ids in different files, with difficult disambiguation. We thus caution to the use of this information, as the current speaker ids

might represent an upper-bound over the real number of speakers in this dataset.

**3. Translation Post-processing.** From the original *Transcriber* annotation files, we filtered out segments corresponding to pauses, noise and music, and removed segments flagged by the annotators as corresponding to foreign languages, such as Arabic and French. We then applied *sacremoses*, the python port of the Moses toolkit (Koehn et al., 2007), for punctuation normalization and tokenization in French. During post-processing, we noticed that some segments (roughly two hours) were flagged as being of poor source audio quality. For these, the translation was produced nevertheless, so we decided to include them in a larger *less controlled* version of the shared corpus.

**4. Audio Post-processing.** We use the resulting collection of segments described in 3. to split the episodes into utterance-level audio files. For posterior use in standard speech processing libraries, we also convert the original .mp3 files into .wav, 16bits, 16KHz, single channel. We then remove all utterances shorter than 1s and longer than 30s. This is the same audio pre-processing from Baevski et al. (2020).

**5. Statistics.** Table 1 presents the statistics for the two versions of the Tamasheq-French parallel corpus we share with the community. The difference between *clean* (17h) and *full* (19h) is that the latter includes potentially noisy segments. Both are available through GitHub: <https://github.com/mzboito/IWSLT2022.Tamasheq-data>.

Regarding the gender distribution, we notice that almost all labeled utterances correspond to male speech. We also observe that more than a half of the utterances are unlabeled (*unknown*). For having a better idea of the gender distribution for this dataset, we performed gender labeling using the *LIUM\_SpkDiarization* tool (Meignier and Merlin, 2010). The results should be interpreted as an estimation, but we observed that all the unlabeled utterances seemed to belong to the male category. We thus believe that this dataset is unfortunately very gender unbalanced.

### 4. The Niger-Mali Audio Collection

This unannotated audio collection corresponds to 671 hours of episodes in five languages: French from Niger, Fulfulde, Hausa, Tamasheq and Zarma. We automatically segmented this audio data, generating 641 hours of content ready for deployment in speech processing pipelines. We detail below the creation of this audio collection, and present some general statistics.

**1. Data Downloading.** Similarly to Section 3, we downloaded 606 episodes in Tamasheq from Studio Tamani,<sup>9</sup> and 2,160 episodes in all the avail-

<sup>3</sup><https://www.studiotamani.org/>

<sup>4</sup><https://www.studiokalangou.org/>

<sup>5</sup><http://<studio-name>.org/journaux/>

<sup>6</sup><http://www.elra.info/en/about/elda/>

<sup>7</sup>The number and identity of the translators was not disclosed to us.

<sup>8</sup><http://trans.sourceforge.net/>

<sup>9</sup>For Studio Tamani news are broadcast twice a day. These correspond to *matin* and *soir* segments in the source files, respectively morning and evening shows.

	clean (17 h)				full (19 h)			
	male	female	unknown	total	male	female	unknown	total
# utterances	2,313	10	3,506	<b>5,829</b>	2,643	11	3,625	<b>6,279</b>
duration	07:37:54	0:00:48	10:04:49	<b>17:43:33</b>	08:49:11	0:00:51	10:28:42	<b>19:18:45</b>

Table 1: Statistics for the clean (left) and full (right) Tamasheq-French parallel corpus.

	# episodes	duration	# utterances	duration (male)	duration (female)	duration (total)
<b>French</b>	464	116:22:09	38,332	52:15:07	58:46:00	111:01:07
<b>Fulfulde</b>	459	114:23:40	39,255	73:31:36	35:54:47	109:26:23
<b>Hausa</b>	424	105:32:48	35,684	75:05:12	25:39:40	100:44:52
<b>Tamasheq</b>	1,014	234:36:29	75,995	134:11:32	90:33:44	224:45:16
<b>Zarma</b>	405	100:42:34	34,198	57:03:37	38:55:33	95:59:10
<b>Total</b>	<b>2,766</b>	<b>671:37:43</b>	<b>223,464</b>	<b>392:07:04</b>	<b>249:49:44</b>	<b>641:56:50</b>

Table 2: Statistics for the Niger-Mali audio collection raw content (left) and automatically segmented version (right), produced by the use of a speech segmentation system with gender labeling.

able languages from Studio Kalangou: French (464), Fulfulde (459), Hausa (424), Tamasheq (408) and Zarma (405). These episodes correspond to the content we managed to retrieve with our web-crawler.<sup>10</sup> It explored URLs ranging from November 2019 to September 2021 for Studio Kalangou, and from January 2020 to September 2021 for Studio Tamani.<sup>11</sup> The left portion of Table 2 presents the statistics for the downloaded .mp3 files: 671 h of audio, being 116 h in French, 114 h in Fulfulde, 105 h in Hausa, 234 h in Tamasheq, and 100 h in Zarma. This corresponds to a total of 2,766 episodes, with an average episode duration of 15 minutes for files from Studio Kalangou, and 13 minutes for Studio Tamani. Finally, we highlight that the choice of having more audio data in Tamasheq was deliberated, since in this paper we focus on building resources for the Tamasheq language.

**2. Segmenting Episodes into Breath Turns.** The episodes downloaded from the websites are used as input for the LIUM\_SpkDiarization tool. The goal of this step is (i) to produce a format compatible with current speech processing models, that cannot deal with very long speech turns, and (ii) to remove silence, music and other non speech events. The LIUM\_SpkDiarization performs speech diarization, separating turns of speech. This allows us to slice the episodes into smaller audio chunks (breath turns).<sup>12</sup> It also has the advantage of producing gender annotation, which allow us to estimate the gender distribution for each language. After applying this diarization tool, we remove the first 12 seconds of each episode, as these often corresponded to intro jingles. The right portion of Table 2 presents the obtained result: 641 h of audio, being 111 h of French, 109 h of Fulfulde, 100 h of Hausa, 224 h of Tamasheq, and 95 h of Zarma. There are 392 h

estimated to be from male speakers, and 249 h from female speakers.

**3. Resulting Corpus.** We make both versions of this corpus (Table 2) available to the community: the 671 h corpus based on episodes, and the 641 h version based on breath turns. This is because, even though we believe our segmentation process to be of good quality, it is still supported by an automatic diarization tool. By providing the source content, we allow the community to choose their own segmentation approach. The audio collection is made available through a dedicated website: <https://demo-lia.univ-avignon.fr/studios-tamani-kalangou/>. In the next section, we briefly elaborate on the five languages available.

#### 4.1. The Languages

The speech resources we collect and share in this paper correspond to five languages spoken in Niger: French, Fulfulde, Hausa, Tamasheq and Zarma. We now provide a brief description of these languages.

- **French (FRA):** French is the official language of the Niger, and a high-resource romance language from the indo-european family. At first, we intended to include only the other four languages listed in this section in the audio collection. However, we noticed some french segments in the Tamasheq annotation from Section 3, and hypothesized that some lexical borrowing might happen due to the coexistence of these languages in the same region.<sup>13</sup>
- **Fulfulde (FUV):** Fulfulde, also known as *Fula*, *Peul* or *Fulani*, is a Senegambian branch of the Niger-Congo language family. Unlike most Niger-Congo languages, it does not have

<sup>10</sup> Accessing and downloading date: 07/10/2021.

<sup>11</sup> Since the sites vary in their file indexing, not all episodes from the indicated periods are successfully retrieved.

<sup>12</sup> By default, the maximum turn length is set to 20 milliseconds.

<sup>13</sup> The same could also be true for the Arabic language, as annotators identified some instances of Arabic terms in the Tamasheq speech from Section 3.

tones (Williamson, 1989). The number of speakers is estimated to be above 40 million (Hammarström, 2015). The native speakers of this language, the Fula people, are one of the largest ethnic groups in the Sahel and West Africa (Hughes, 2009).<sup>14</sup>

- **Hausa (HAU):** Hausa is a Chadic language, member of the Afro-Asiatic language family. It is spoken mainly within the northern half of Nigeria and the southern half of Niger, with Wolff et al. (1991) and Newman (2009) estimating the number of speakers between 20 and 50 million. Early studies in Hausa showcased a remarkable number of loanwords from Arabic, Kanuri, and Tamasheq (Schön, 1862).
- **Tamasheq (TAQ):** Tamasheq is a variety of Tuareg, a Berber macro-language spoken by nomadic tribes across North Africa in Algeria, Mali, Niger and Burkina Faso (Heath, 2006). It accounts for approximately 500,000 native speakers, being mostly spoken in Mali and Niger (Hammarström, 2015). The livelihood of the Tuareg people has been under threat in the last century, due to climate change and a series of political conflicts (Decalo, 1997). This reduced considerably the number of speakers of Tamasheq however, partially due to the Malian government’s active promotion of the language in recent years, Tamasheq is now classified as a *developing language* (Hammarström, 2015).
- **Zarma (DJE):** Zarma, also spelled *Djerma*, is a leading indigenous Songhay language of the southwest lobe of the west African nation of Niger, spoken by over 2 million speakers. This tonal language is also spoken in Nigeria, Burkina Faso, Mali, Sudan, Benin and Ghana (Britannica, The Editors of Encyclopedia, 2015).<sup>15</sup>

## 5. Use case: Speech Translation Baseline

In this paper we present speech resources for the Tamasheq language, and in four other geographically close languages. They are shared in the context of the IWSLT 2022 low-resource speech translation track. In this section we present as use case our end-to-end speech translation baseline that uses the Tamasheq-French Parallel Corpus from Section 3.

**Dataset.** We run this baseline experiment using both versions of the dataset from Section 3, with data splits detailed in Table 3. We extract 80-dimensional mel filterbank features from the Tamasheq utterances. For

<sup>14</sup>Lexical resources can be found at: <http://www.language-archives.org/language/ful>

<sup>15</sup>Lexical resources can be found at: <http://www.language-archives.org/language/dje>

	train	valid	test
clean (17 h)	4,444 / 13h50	581 / 1h53	804 / 1h59
full (19 h)	4,886 / 15h24	-	-

Table 3: The (Number of utterances / duration) per set. Both *clean* and *full* share the same validation and test sets.

	valid	test
clean (17 h)	2.22 (20.6/3.6/1.1/0.4)	1.80 (18.8/2.9/0.8/0.3)
full (19 h)	2.31 (18.5/3.3/1.0/0.4)	1.90 (15.9/2.6/0.9/0.4)

Table 4: End-to-end speech translation BLEU4 results for the baselines, with detailed scores between parentheses.

the French text, we build a 1k unigram vocabulary using *Sentencepiece* (Kudo and Richardson, 2018) without pre-tokenization.

**Architecture.** We use the `fairseq s2t` toolkit (Wang et al., 2020), training end-to-end speech translation Transformer models (Vaswani et al., 2017), preceded by two convolutional layers for dimensionality reduction.<sup>16</sup> These models are trained for 500 epochs using the Adam optimizer (Kingma and Ba, 2014) with 10k warm-up steps. For decoding, we use beam search with a beam size of 5, and we evaluate the models using the best checkpoint with respect to the loss in the validation set.

**Results and Discussion.** Table 4 presents detokenized case-sensitive BLEU scores computed using *sacreBLEU* (Post, 2018). Looking at these results, we notice that the *full* version of the dataset improves slightly over the *clean* version. The former contains roughly two extra hours in its training set, and thus this could hint that having more data in data scarcity scenarios is beneficial, even when this data is of questionable quality. Nevertheless, the performance of both baselines is *very* low. They highlight the challenge of low-resource end-to-end speech translation when the only data used is of parallel nature. We believe results can be further improved by using auxiliary monolingual tools and models. The next paragraphs elaborate on this.

For the text, and since French is a high-resource language, one could incorporate pre-trained embeddings to the translation decoder. For the decoding procedure, language models – such as CAMEMBERT (Martin et al., 2020) and FLAUBERT (Le et al., 2020) – can be used. Pre-trained decoders like MBART (Liu et al., 2020) could also be incorporated.

For the speech, the self-supervised speech representation produced by models such as HUBERT (Hsu et al., 2021) and WAV2VEC 2.0 (Baevski et al., 2020) can replace mel filterbanks features for the speech translation encoder. One can use freely available pretrained models in high-resource languages, or train these mod-

<sup>16</sup>Settings are detailed at their `s2t_transformer_xs` recipe.

els from scratch. For the latter option, the resources from Section 4 can be used. In both cases, self-supervised (also called *task-agnostic*) fine-tuning in the target language can increase results, but the best option seems to be to fine-tune in the target task directly (Evain et al., 2021; Babu et al., 2021).

Lastly, an interesting research direction is the leveraging of multilingual data in self-supervised models for speech. There are massive multilingual models that produce speech representations from many unrelated languages seen during training (Conneau et al., 2020; Babu et al., 2021). However, we currently do not know if these models are in fact better than having *dedicated models* trained with a smaller set of languages that are closely related (i.e. in speech style, geography, phonology, linguistic family). Thus, it might be interesting to compare the speech representations produced by a multilingual model based on the languages from Section 4, against current multilingual baselines, such as XLSR-53 (Conneau et al., 2020) and XLS-R (Babu et al., 2021).

## 6. Conclusion

In this paper we presented two resources that focus on the Tamasheq language. The **Niger-Mali audio collection** contains 641 hours of speech in French from Niger, Fulfulde, Hausa, Tamasheq and Zarma. The presence in the data of audio recording from languages spoken in the same geographical area is particularly interesting for research related to transfer learning (Wang and Zheng, 2015) and self-supervision (Baevski et al., 2020; Hsu et al., 2021; Conneau et al., 2020). This resource is publicly available on our website: <https://demo-lia.univ-avignon.fr/studios-tamani-kalangou/>.

The second resource we share with the research community, the **Tamasheq-French Parallel Corpus**, focuses on speech translation. It contains 17h of speech in Tamasheq aligned at the utterance-level to French translations. We believe this dataset is an interesting resource for those interested by low-resource speech translation. It is publicly available on this GitHub page: [https://github.com/mzboito/IWSLT2022\\_Tamasheq\\_data](https://github.com/mzboito/IWSLT2022_Tamasheq_data).

Lastly, we also presented a baseline model for IWSLT 2022 low-resource track using the Tamasheq-French parallel corpus. The obtained scores highlight the great challenge of developing effective approaches in such low-resource settings. We believe that by leveraging monolingual tools and data in the translation model, notably through the use of the audio collection presented in this paper, one might be able to develop more effective models for Tamasheq.

## 7. Acknowledgements

We are very thankful to Fondation Hironnelle, Studios Kalangou from Niger, and Studios Tamani from Mali,

for allowing us to download, use and distribute their audio data under the Creative Commons BY-NC-ND 3.0 license for non-commercial use.

This work was funded by the French Research Agency (ANR) through the ON-TRAC project under contract number ANR-18-CE23-0021. This paper was also partially funded by the European Commission through the SELMA project under grant number 957017. We would like to thank Antoine Caubrière from LIA for all the help with the data packaging.

## 8. Bibliographical References

- Babu, A., Wang, C., Tjandra, A., Lakhota, K., Xu, Q., Goyal, N., Singh, K., von Platen, P., Saraf, Y., Pino, J., et al. (2021). Xls-r: Self-supervised cross-lingual speech representation learning at scale. *arXiv preprint arXiv:2111.09296*.
- Baevski, A., Zhou, H., Mohamed, A., and Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. *arXiv preprint arXiv:2006.11477*.
- Britannica, The Editors of Encyclopedia. (2015). Songhai languages. [Online; accessed 3-01-2022].
- Conneau, A., Baevski, A., Collobert, R., Mohamed, A., and Auli, M. (2020). Unsupervised cross-lingual representation learning for speech recognition. *arXiv preprint arXiv:2006.13979*.
- Decalo, S. (1997). Historical dictionary of niger.
- Ethnologue: Languages of the World. (2021). Tamasheq. [Online; accessed 07-12-2021].
- Evain, S., Nguyen, H., Le, H., Boito, M. Z., Mdhaffar, S., Alisamir, S., Tong, Z., Tomashenko, N., Dinarelli, M., Parcollet, T., et al. (2021). Task agnostic and task specific self-supervised learning from speech with lebenchmark. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*.
- Hammarström, H. (2015). "ethnologue" 16/17/18th editions: A comprehensive review. *Language*, pages 723–737.
- Hsu, W.-N., Bolte, B., Tsai, Y.-H. H., Lakhota, K., Salakhutdinov, R., and Mohamed, A. (2021). Hubert: Self-supervised speech representation learning by masked prediction of hidden units. *arXiv preprint arXiv:2106.07447*.
- Hughes, M. E. (2009). Africa and the americas: Culture, politics, and history. *Reference & User Services Quarterly*, 48(4):402–403.
- Joshi, P., Santy, S., Budhiraja, A., Bali, K., and Choudhury, M. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*.
- Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W.,

- Moran, C., Zens, R., et al. (2007). Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th annual meeting of the association for computational linguistics companion volume proceedings of the demo and poster sessions*, pages 177–180.
- Kudo, T. and Richardson, J. (2018). Sentencepiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. *arXiv preprint arXiv:1808.06226*.
- Le, H., Vial, L., Frej, J., Segonne, V., Coavoux, M., Lecouteux, B., Allauzen, A., Crabbé, B., Besacier, L., and Schwab, D. (2020). FlauBERT: Un-supervised language model pre-training for French. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 2479–2490, Marseille, France, May. European Language Resources Association.
- Liu, Y., Gu, J., Goyal, N., Li, X., Edunov, S., Ghazvininejad, M., Lewis, M., and Zettlemoyer, L. (2020). Multilingual denoising pre-training for neural machine translation. *Transactions of the Association for Computational Linguistics*, 8:726–742.
- Martin, L., Muller, B., Ortiz Suárez, P. J., Dupont, Y., Romary, L., de la Clergerie, É., Seddah, D., and Sagot, B. (2020). CamemBERT: a tasty French language model. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7203–7219, Online, July. Association for Computational Linguistics.
- Meignier, S. and Merlin, T. (2010). Lium spkdiarization: an open source toolkit for diarization. In *CMU SPUD Workshop*.
- Newman, P. (2009). *Hausa and the chadic languages*. Routledge.
- Post, M. (2018). A call for clarity in reporting bleu scores. *arXiv preprint arXiv:1804.08771*.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.
- Wang, D. and Zheng, T. F. (2015). Transfer learning for speech and language processing. In *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, pages 1225–1237. IEEE.
- Wang, C., Tang, Y., Ma, X., Wu, A., Okhonko, D., and Pino, J. (2020). Fairseq s2t: Fast speech-to-text modeling with fairseq. *arXiv preprint arXiv:2010.05171*.
- Williamson, K. (1989). *Benue–Congo Overview*. In *The Niger–Congo Languages*. J. Bendor-Samuel ed. Lanham: University Press of America.
- Wolff, E. H., von Gleich, U., and Wolff, E. (1991). Standardization and varieties of written hausa (west africa). *Gleich, Utta von and Wolff Ekkehard (eds.)*, pages 21–33.

## 9. Language Resource References

- Heath, J. (2006). *Dictionnaire touareg du Mali: tamachek-anglais-français*. KARTHALA Editions.
- Schön, J. F. (1862). *Grammar of the Hausa language*. Church missionary house.