Adding Audio to Wordnets

Francis Bond
Dept. of Asian Studies and Sinofon Project
Faculty of Arts
Palacký University
bond@ieee.org

Abstract

This paper explores the integration of sound files into wordnets, transforming them from static lexical databases into multimodal tools for linguistics, language learning and maintenance. Traditionally, wordnets focus on textual representations. Adding sound improves usability for language learners and linguists, especially in less-documented or endangered languages.

We extracted sound data for basic vocabulary in 24 languages from the TUFS Basic Vocabulary Modules, link them to senses and make them available as small wordnets. We also discuss the issues involved with merging the data into an existing wordnet, looking at the Open English Wordnet. In addition, this paper outlines the process of integrating audio, discusses potential use cases, and evaluates the technical challenges involved. Finally we suggest an extension to the wordnet formats to allow sound for examples and definitions as well.

1 Introduction

Wordnets are powerful lexical databases that organize words into synsets (sets of synonyms), which are linked by various semantic relations. Originally developed for English, wordnets have been created cover many languages, providing a valuable resource for natural language processing (NLP) tasks and lexicographic applications (Miller, 1995; Fellbaum, 1998; Vossen, 1998; Vossen et al., 1999; Bond and Foster, 2013). Despite their utility, wordnets, like most lexicographical resources, traditionally focus on textual representations of lexical knowledge, lacking multimodal components such as sound, which limits their usability as comprehensive dictionaries.

The addition of sound files to wordnets has the potential to transform them from static lexical databases into more dynamic, user-friendly resources. By associating sound files with specific synsets or lexical entries, wordnets can provide users with the ability to hear pronunciations, which is especially valuable for language learners, phoneticians, and researchers focusing on dialectology or sociolinguistics. This is particularly important for less well-known or endangered languages, where wordnets might serve as the primary lexicon, if not the only resource available. Several projects involved with language documentation and maintenance have noted that they would like to add audio files to their wordnets (Sio and Costa, 2019; Morgado Da Costa et al., 2023). The African Wordnet project (AWN) noted that pronunciation (in this case information about tones) is essential for disambiguating some words, although it is not shown in the standard orthography (Bosch and Griesel, 2018). In these cases, sound files help preserve and make available phonetic information, supporting both pronunciation accuracy and the documentation of spoken forms where orthographies may be underdeveloped or nonstandardized. Sinha et al. (2020) went as far as synthesizing audio for the Hindi WordNet, which underscores the value of sound. And of course, sound has been incorporated into many online dictionaries beyond wordnets with most online dictionaries of English having audio recordings of entry words, and a few even adding audio for example sentences (Fuertes-Olivera and Bergenholtz, 2011, p254). Some lexicons even include *sound effects* to present an ostensive definition of sounds.

While audio files are still not widely integrated into wordnets, there have been related efforts to enhance wordnets with pronunciation data. The 2021 release of Open English WordNet included pronunciation information for nearly 35,000 entries.[1] Efforts have also been made to extract pronunciation data from Wiktionary for use in word-

[1] https://github.com/globalwordnet/english-wordnet/releases/tag/2021-edition

nets (Declerck et al., 2020; Declerck and Bajčetić, 2021) but to the best of our knowledge this has not yet been incorporated into any actual wordnets. For general users, particularly non-linguists, sound files offer a straightforward way to learn correct pronunciations without requiring knowledge of phonetic transcriptions like the International Phonetic Alphabet (IPA). This makes wordnets more accessible to the average person, removing potential barriers created by complex orthographies or regional phonetic variations. Additionally, for languages with significant dialectal diversity, sound files can capture and preserve different regional variations, providing a richer and more comprehensive resource that reflects the full linguistic diversity of the language community.

In this paper, we explore the technical and linguistic considerations involved in adding sound files to wordnets, focusing on the challenges of linking audio data to synsets, managing multilingual and dialectal variation, and ensuring the quality of the recordings. We present a case study of a wordnet that has incorporated sound files, showcasing the process and reflecting on user feedback. By doing so, we aim to demonstrate that sound-enriched wordnets offer a significant improvement in both educational and research contexts, expanding their role beyond traditional text-based lexical databases.

2 Resources

In this section describe the main resources we use.

2.1 TUFS

The TUFS Basic Vocabulary Modules (Kawaguchi et al., 2007) are a resource developed for online language learning, containing vocabulary words and example sentences across 24 languages. This dataset, particularly focused on Asian languages, serves as a foundational tool for learners by providing vocabulary, grammar sketches, dialogues and example sentences in a formal conversational register. The vocabulary is designed to cover commonly used words, using words selected according to the Japanese Language Proficiency Test (N5). Additionally, the dataset provides translations across different languages, some of which introduce nuanced differences, such as gender distinctions in languages like German and Vietnamese.

Bond et al. (2020b) linked the TUFS data with the Open Multilingual Wordnet (OMW). This gave new resources for evaluating and enriching existing wordnets, created new wordnets for languages such as Khmer, Korean, Lao, Mongolian, Tagalog, Urdu, and Vietnamese. The linking process identifies multilingual connections between vocabulary items, improving cross-linguistic understanding and data integration across languages. However, they did not take advantage of the fact that the TUFS modules include good quality audio data for all of the vocabulary and example sentences.

2.2 Cantonese Wordnet

Recent work on the Cantonese wordnet (Sio and Costa, 2019) has focused on adding pronunciation, in the form of both transliterations and audio. The most recent release (Sio et al., 2025, this volume) includes over 2,000 pronunciations. Each pronunciation has a phonemic transcription using jyutping (the Linguistic Society of Hong Kong Cantonese Romanization Scheme), and an audio file, either reused from wikimedia commons or newly recorded.

2.3 Open Multilingual Wordnet

We use (and extend) the software for reading and displaying wordnets for the Open Multilingual Wordnet 2.0 (Bond et al., 2020a). This open-source software handles global wordnet association WordNet Lexical Markup Framework (WN-LMF) a standard format for representing wordnet-style lexical databases (Vossen et al., 2016; Bond et al., 2016; McCrae et al., 2021).

3 Creating and displaying

Since 2021 (WN-LMF 1.1) has had the capability to represent the pronunciation of lemmas (McCrae et al., 2021). This is in the <Pronunciation> element, which gives the IPA text. It has the following attributes:

- **variety** encodes the language variety, for example by using the IETF language tags to indicate dialect, where British English in IPA would be `en-GB-fonipa`. We do not have a general standard for how these are labelled, each wordnet can decide on its own.
- **notation** can encode further information such as indicating the speaker particular dialect (this was **notes** in McCrae et al. (2021))
- **phonemic** indicates whether the transcription

is phonemic, **true**, or phonetic, **false**, defaulting to **true** (phonemic)

  - Phonemic transcription represents the phonemes (distinctive sounds) of a language. It is more abstract and focuses on the sounds that change the meaning of words. Typically it is presented with sounds enclosed in slashes, e.g., /kæt/.
  - Phonetic transcription represents the exact pronunciation of speech, including fine details of how each sound is produced. It is more detailed and is typically shown in brackets, e.g., [kʰæt̚].

  Phonemic transcription is generally more suitable for dictionaries, because dictionaries aim to represent how words are typically pronounced in the language without overwhelming users with excessive phonetic detail. Phonemic transcription balances simplicity and accuracy by focusing on sounds that change meaning, which is important for learners to distinguish between words.

  For example, in a dictionary, users usually need to know whether a word starts with a /b/ or /p/ to distinguish *bat* from *pat*, but they don't necessarily need to know that in some dialects the t in *cat* might be unreleased.

- **audio** gives the URL of an audio file of the pronunciation

An example of encoding is given in Figure 1. Here we show the pronunciation under the lemma, it is also possible to store the pronunciation under the variant forms.

We have extended the OMW format (Bond and Foster, 2013) to read the pronunciation and store it in the database. The schema for the table for pronunciation is shown in 2. This is similar to how it is stored in the Python WN module (Goodman and Bond, 2021), which can also read and access the pronunciation.

3.1 Display

We display the pronunciation when we look at the sense of a word. If there is audio, then we show a loudspeaker (🔊), clicking on which plays the audio file. If there is a `value` then we show this, either between // if phonemic is true, or between [] if it is false.

For the TUFS wordnets, most of them have only audio. For the Open English Wordnet, entries with pronunciation generally only have the `value`, we have created a hybrid example here to show both (Figure 3). The Cantonese Wordnet has both pronunciation and audio, but has chosen to add them to the variants. Figure 4 shows two variants with different pronunciations, one showing the 'lazy' pronunciation (Chen, 2018; Cheng et al., 2022).

4 Discussion

We have shown that the current GWA format can usefully represent audio for wordnet words. In addition we have extracted audio from TUFS and linked it to senses for 24 languages, for a total of 10,945 pronunciations. These are all common words from a beginners vocabulary and thus useful for early learners. A release of this data will be made available at `https://github.com/fcbond/tufs`.

They can be loaded as wordnets, but are not full wordnets — they have no internal links, or definitions (except for Japanese). Our hope is that the audio links (and example sentences and new senses) will be taken up by existing projects and merged. This can be done fully automatically for TUFS nodes that map to only one ILI link, if there is not existing information about the pronunciation. If there is already pronunciation (as there is in the Open English WordNet), and there are multiple variants, then they must be disambiguated. For example, *fall* "autumn" has three pronunciations in the OEWN: /fɔːl/ (GB) and /fɔl/ (US) and /fɑl/ (unmarked). The pronunciation in TUFS is closest to /fɔːl/.[2] To be safe, even words with only one pronunciation shown should be checked, in case the pronunciation is not the same.

We compared the pronunciations in OEWN (2024 release candidate) and TUFS (English). Most words in OEWN have no pronunciation (0), for those that do, over a third have two or more pronunciations (Table 1).

When we looked at TUFS, 454 English words have associated audio. Of these, 381 have a pronunciation in OEWN (the other 63 do not). 126 have only one pronunciation. We plan to cooperate with the OEWN to merge this data.

To merge the pronunciations requires a competent speaker of the language in question. We do not

[2] `https://www.coelang.tufs.ac.jp/mt/en/vmod/sound/word/word_1142.mp3`

```
<LexicalEntry id="ex-rabbit-n">
    <Lemma writtenForm="rabbit" partOfSpeech="n"/>
        <Pronunciation variety="en-GB-fonxsamp en-US-fonxsamp"
            audio ="https://path/rabbit.flac">'r\{bIt</Pronunciation>
    <Pronunciation variety="en-AU-fonxsamp" notation="weak vowel merger"
            audio ="https://path/rabbit1.flac">'r\{b@t</Pronunciation>
     </Lemma>
 </LexicalEntry>
```

Figure 1: WN-LMF representation of pronunciation

```
CREATE TABLE pronunciations
  (id INTEGER PRIMARY KEY ASC,
   w_id INTEGER NOT NULL,    -- id of the linked word
   value   TEXT,             -- phonemic or phonetic realization
   variety TEXT,             -- encoding of the realization
   phonemic BOOLEAN NOT NULL CHECK (mycolumn IN (0, 1)) DEFAULT 1,
   notation TEXT,            -- any comments
   src_id INTEGER NOT NULL, -- which wordnet it came from
   t TIMESTAMP DEFAULT CURRENT_TIMESTAMP,
   FOREIGN KEY(w_id) REFERENCES w(id));
```

Figure 2: SQL representation of the pronunciation

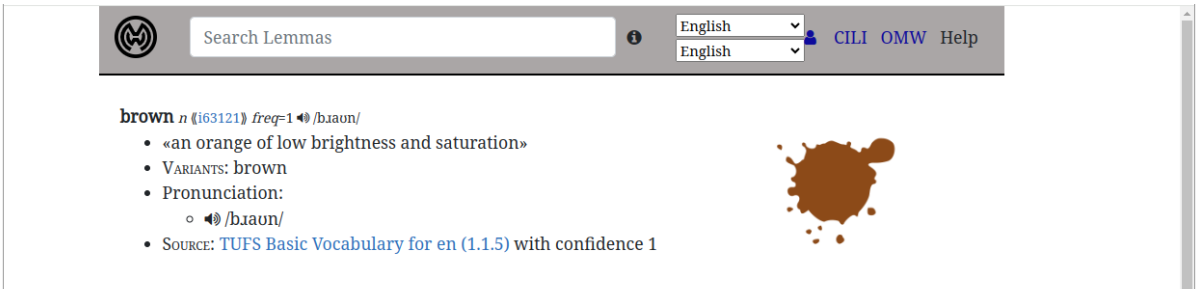

Figure 3: Pronunciation for *brown*, showing the pronunciation and an image from ARASAAC

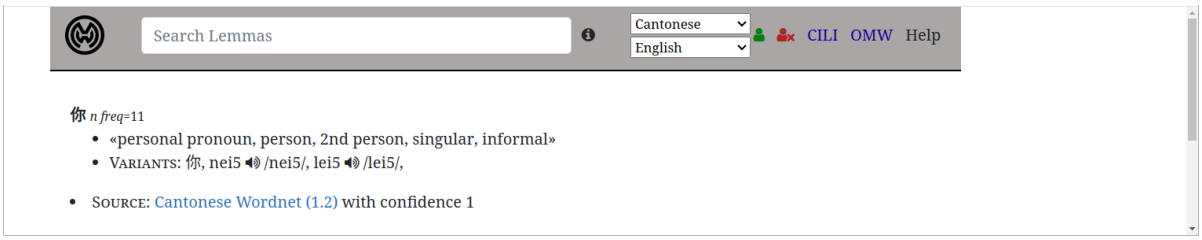

Figure 4: Pronunciation for 你 *nei5* showing a variant with the lazy l

Pronunciations	Number	Example
0	128,313	heel
1	22,606	brown
2	10,202	dog
3	479	fall
4	45	wolf
5	6	croissant
6	2	scallop

Table 1: Distribution of Pronunciations in OEWN (2024)

speak all 24 languages in the TUFS collection, but are happy to work with existing wordnet projects who wish to use his data.

4.1 Extending the GWA formats

Currently `Pronunciation` is only defined for lemmas. However, it would be easy to also add it to `Example` and possibly even `Definition`. Having pronunciation, especially audio, available for example sentences provides significant advantages for language learners. One of the most important reasons is that pronunciation in isolation can differ from pronunciation in context. Words often change due to natural speech phenomena such as connected speech, assimilation, or elision. By hearing how words sound in full sentences, learners can better understand how they flow together and how stress and rhythm work in natural speech, improving their fluency and pronunciation (Lew, 2011, pp255–266).

Additionally, example sentences demonstrate important aspects of prosody, such as intonation and sentence stress, which affect meaning and convey emotion. In many languages, the intonation pattern of a sentence can alter its meaning, such as rising intonation in questions or stress on certain words to indicate emphasis. Learners benefit from hearing these subtleties, which helps them sound more natural and better interpret the nuances of spoken language in real conversations.

For language documentation, where the orthography may not be fully standardized, having audio recordings of example sentences is even more crucial. The audio preserves the authentic pronunciation and prosody of the language, especially in cases where writing systems might not fully capture the phonetic details or where different writing systems coexist. This ensures that even if the orthography changes or develops over time, the spoken language, as documented in audio form, remains accessible and accurately represented for future researchers and learners.

Lastly, providing pronunciation for sentences aids listening comprehension. In connected speech, words may be pronounced differently than in isolation, and learners need exposure to such variations to understand real-world speech. Moreover, hearing sentences in context allows learners to grasp common collocations, idiomatic expressions, and how prosody affects meaning, which enriches their vocabulary and overall understanding. By including sentence-level pronunciation, learners can bridge the gap between theoretical knowledge and everyday spoken language. For these reasons, the TUFS modules includes the pronunciation for example sentences, and it would be good to take advantage of this.

4.2 Images

In addition, researchers have been trying to associate images with wordnet since at least Bond et al. (2008). We are now using illustrations from the Aragonese Center of Augmentative and Alternative Communication (Arasaac) to illustrate the OMW. The pictographic symbols are the property of the Government of Aragón and have been created by Sergio Palao, under a Creative Commons License BY-NC-SA.[3] These illustrations have several advantages for illustrating lexicons. First, they are designed for communicative purposes and widely used. Secondly, the symbols have been augmented with many localised versions for different cultures, such as Arabic, Bulgarian, SEA and Urdu.[4] Thirdly, the symbols are line drawings, which have been shown to be more effective for dictionary users (Dziemianko, 2022). Finally, they have been linked to Princeton WordNet 3.1 by Schwab et al. (2020), and so can be linked to CILI.

8,402 ili concepts are linked to illustrations. We further extend the coverage by illustrating a concept with an illustration of its direct hypernym if it exists. We show an example in Figure 3.

5 Conclusion

Incorporating audio into wordnets is the next step towards creating more comprehensive and user-friendly linguistic tools, enhancing both educational and research potential. This paper has outlined practical methods for linking sound files with

[3] https://arasaac.org/
[4] Global Symbols CIC

synsets, unlocking new possibilities for both linguistic research and language education. While challenges remain in ensuring broad accessibility and accurate audio representation, the benefits for language preservation and phonetic research are clear. Future research should focus on improving audio integration and scaling its application in collaborative, multilingual wordnets, further enhancing their utility. In particular, we would like to make the interface better suited for use on a smaller device like a telephone.

5.1 Acknowledgments

ChatGPT (4o) was used to copyedit the paper and help with the LaTeX formatting. ChatGPT (4o) and Claude (3.5 Sonnet) were used to debug the code for extracting the data from TUFS and formatting it for OMW.

References

Francis Bond and Ryan Foster. 2013. Linking and extending an open multilingual wordnet. In *51st Annual Meeting of the Association for Computational Linguistics: ACL-2013*, pages 1352–1362, Sofia.

Francis Bond, Hitoshi Isahara, Kyoko Kanzaki, and Kiyotaka Uchimoto. 2008. Boot-strapping a WordNet using multiple existing WordNets. In *Sixth International Conference on Language Resources and Evaluation (LREC 2008)*, Marrakech.

Francis Bond, Luis Morgado da Costa, Michael Wayne Goodman, John Philip McCrae, and Ahti Lohk. 2020a. Some issues with building a multilingual wordnet. In *Twelth International Conference on Language Resources and Evaluation (LREC 2020)*, Marseilles.

Francis Bond, Hiroki Nomoto, Luís Morgado da Costa, and Arthur Bond. 2020b. Linking the TUFS basic vocabulary to the open multilingual wordnet. In *Twelth International Conference on Language Resources and Evaluation (LREC 2020)*, Marseilles.

Francis Bond, Piek Vossen, John McCrae, and Christiane Fellbaum. 2016. CILI: The collaborative interlingual index. In *Proceedings of the 8th Global Wordnet Conference (GWC 2016)*, pages 50–57.

Sonja Bosch and Marissa Griesel. 2018. African Wordnet: facilitating language learning in African languages. In *Proceedings of the 9th Global Wordnet Conference*, pages 306–313, Nanyang Technological University (NTU), Singapore. Global Wordnet Association.

Katherine Hoi Ying Chen. 2018. Ideologies of language standardization: The case of Cantonese in Hong Kong. In *The Oxford Handbook of Language Policy and Planning*. OUP.

Lauretta Cheng, Molly Babel, and Yao Yao. 2022. Production and perception across three Hong Kong Cantonese consonant mergers: Community- and individual-level perspectives. *Laboratory Phonology*, 13.

Thierry Declerck and Lenka Bajčetić. 2021. Towards the addition of pronunciation information to lexical semantic resources. In *Proceedings of the 11th Global Wordnet Conference*, pages 284–291, University of South Africa (UNISA). Global Wordnet Association.

Thierry Declerck, Lenka Bajcetic, and Melanie Siegel. 2020. Adding pronunciation information to wordnets. In *Proceedings of the LREC 2020 Workshop on Multimodal Wordnets (MMW2020)*, pages 39–44, Marseille, France. The European Language Resources Association (ELRA).

Anna Dziemianko. 2022. The usefulness of graphic illustrations in online dictionaries. *ReCALL*, 34(2):218–234.

Christine Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press.

Pedro A. Fuertes-Olivera and Henning Bergenholtz, editors. 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. A&C Black.

Michael Wayne Goodman and Francis Bond. 2021. Intrinsically interlingual: The Wn Python library for wordnets. In *11th International Global Wordnet Conference (GWC2021)*.

Yuji Kawaguchi, Toshihiro Takagaki, Nobuo Tomimori, and Yoichiro Tsuruga, editors. 2007. *Corpus-Based Perspectives in Linguistics*, volume 6 of *Usage-Based Linguistic Informatics*. John Benjamins Publishing Company, Amsterdam.

Robert Lew. 2011. Multimodal lexicography: The representation of meaning in electronic dictionaries. *Lexicos*, 20.

John P. McCrae, Michael Wayne Goodman, Francis Bond, Alexandre Rademaker, Ewa Rudnicka, and Luís Morgado da Costa. 2021. The global wordnet formats: Updates for 2020. In *11th International Global Wordnet Conference (GWC2021)*.

George A. Miller. 1995. WordNet: A lexical database for English. *Commun. ACM*, 38(11):39–41.

Luís Morgado Da Costa, František Kratochvíl, George Saad, Benidiktus Delpada, Daniel Simon Lanma, Francis Bond, Natálie Wolfová, and A.l. Blake. 2023. Linking SIL semantic domains to wordnet and expanding the Abui wordnet through rapid words collection methodology. In *12th International Global Wordnet Conference (GWC2023)*.

Roberto Navigli and Simone Paolo Ponzetto. 2012. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250.

Didier Schwab, Pauline Trial, Céline Vaschalde, Loïc Vial, Emmanuelle Esperanca-Rodier, and Benjamin Lecouteux. 2020. Providing semantic knowledge to a set of pictograms for people with disabilities: a set of links between WordNet and arasaac: Arasaac-WN. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 166–171, Marseille, France. European Language Resources Association.

Shikha Sinha, Kalika Bali Narayan, and Prabhakar Singh. 2020. Synthesizing audio for hindi wordnet. In *Proceedings of the 10th Global WordNet Conference*, pages 230–235.

Joanna Ut-Seong Sio and Luis Morgado Da Costa. 2019. Building the Cantonese Wordnet. In *Proceedings of the 10th Global Wordnet Conference*, pages 206–215, Wroclaw, Poland. Global Wordnet Association.

Piek Vossen, editor. 1998. *Euro WordNet*. Kluwer.

Piek Vossen, Francis Bond, and John McCrae. 2016. Toward a truly multilingual global wordnet grid. In *Proceedings of the 8th Global Wordnet Conference (GWC 2016)*. 419–426.

Piek Vossen, Wim Peters, and Julio Gonzalo. 1999. Towards a universal index of meaning. In *Proceedings of ACL-99 Workshop, Siglex-99, Standardizing Lexical Resources*, pages 81–90, Maryland.