Usage

- **Practical Prerequisites**
- **Expectations**
- **First Run**
- **Subsequent Runs (CV)**
- **Subsequent Runs (Tone)**
- **Miscellaneous**

Practical Prerequisites

LIFT Database to Check

A→Z + T requires a LIFT database to check. Fortunately these are not difficult to generate; LIFT is an open XML specification for storing lexical data. LIFT databases can be created by a number of routes:

- Collect words in WeSay (Download here).
  - WeSay uses LIFT natively, so the same repository can be used for WeSay and A→Z + T (though it is not recommended to use them both at the same time).
  - WeSay is Chorus enabled, which allows easy tracking of changes and off-site archiving, including the changes to your database and reports made by A→Z + T.
  - N.B.: I highly recommend the excellent library of images that works with WeSay, the Art of Reading
- Build a dictionary in FLEx, and export to LIFT/WeSay.
  - assuming you want to get the checked database *back* into FLEx, you want to use send/receive for WeSay. Import/Export (a different process) can work, but there is no protection against overwriting your data, without doing backups yourself.
  - there is an active list of users to help with problems doing this.
- Store your data in some other form (text, spreadsheet, database) and convert it to LIFT (*PLEASE* don't do this unless you *really* know what you're doing, and have a *good* reason to; the other options above are much easier, and much less likely to result in data corruption)

LIFT Database Requirements

LIFT databases can be minimal or very complex. For the purposes of running A→Z + T, you just need the following:

- citation (or lexeme) forms (tagged with your language code, of course)
  - forms with spaces or other non-wordforming characters are ignored.
- glosses (or definitions) in at least one language (again coded for gloss language)

- **N.B.:** Long definitions cause enough problems with the UI, that they are now truncated to the first three words. If you don't like this (I wouldn't!), set up proper gloss fields, and A→Z + T will use them.
- `Grammatical Category/Part of Speech (ps)` indication:
  - stored in `sense/grammatical-info@[value]`
  - whatever ps names (e.g., 'Noun', 'Nom', 'Njina', 'noun', 'n', etc) are in your database is what you will select from to study, so name them in a way that will help your work
  - entries with no ps value will be left out of the A→Z + T analysis
- No markup. If you have data from another tool that added any kind of markup (formatting, references, etc), that will likely cause you problems with A→Z + T. If you have and want to preserve markup inside any of the above fields, I would recommend copying it into another field before stripping it for use with A→Z + T.

Collaboration and Archival

I **strongly** recommend using a version controlled repository (e.g., mercurial, git), as is normally done in WeSay and in recommended FLEx collaboration schemes. Even if you are the only one to ever see this data (why would that be?), the advantages in history and preservation of your data are already there. But if you will be sharing changes with others, you really **must** have an easy way to do this, or you will get bogged down in the logistics of sharing data changes. I use Language Depot, though there are certainly other ways to meet this need. In any case, setting this up early is always easier than trying to merge divergent data later.

Error Tracking

Once A→Z + T is running normally, it will create a log with more information that you probably want), in a file called `log_<date>.txt`. If something unexpected happens, that file should contain information that will help understand what happened and why. That is recreated on each A→Z + T startup, though, so add a meaningful name and send it to me before running A→Z + T again.

If A→Z + T has an exception, it should create another file, `log_date_time.xz`. It is already named for time, so you should find one of these for each such error. Any time you get one of these, please send it to me with a brief description of what you observed.

Expectations

Changes to Expect in Your LIFT Database

A→Z + T will place links to `citation/lexeme`, `plural`, and `imperative` recordings in the appropriate fields, coded as a voice writing system (e.g., xyz-Zxxx-x-audio).

A→Z + T will place sorting information in the `entry/sense` node, under an example node for each context/frame:

- form —of word in frame, as presented to user during sorting
- translation —of word in frame, as presented to user during sorting
- frame name
- subgrouping name
- link to sound file

Results of the analysis of multiple frame groupings (e.g., from the Tone Report) is placed in a separate `entry/sense/field[@type='tone']` (i.e., not in an example field), as this is a summary/analysis of the values contained in the example nodes.

Changes to Expect near Your LIFT Database

A→Z + T assumes your LIFT file is in a directory set apart for its analysis. As a result, expect to find generated in that directory:

- A syllable profile analysis file (`ProfileData`), on first run, and if CV analysis parameters change. This can take a while to run, so it is stored and loaded on startup, rather than running it on each startup.
- A settings file (`CheckDefaults`), any time a check is run, to preserve settings, including current `ps/profile` selection. This allows you to continue where you left off.
- A tone frame definitions file (`ToneFrames`), once at least one has been defined. I strongly recommend not modifying this file outside of A→Z + T.
- A verification status file (`VerificationStatus`), where your progress in verification is stored.
- A Sound card settings file (`SoundSettings`), where your sound card settings are stored.
- An `audio` folder, where sound files are saved, once at least one has been recorded.
- A `report` folder, where report files (txt and xml) are stored whenever run by the user. These are set apart from the rest of the repository to reduce clutter. N.B.: relative links to audio work in this hierarchy, if you move a report from this directory, be sure to update the links accordingly.
- An ad hoc groups file (`AdHocGroups`), if you have defined any ad hoc groups.

First Run: Be Patient and Orient Yourself

Tell A→Z + T Where to Find Your Database

The first time you run A→Z + T, you will need to select your LIFT database location. A→Z + T stores this location in `lift_url.py`, so you only have to do this once. But if you need to check a different database, delete `lift_url.py`, and A→Z + T will ask again where your database is.

Syllable Profile Analysis

If you open A→Z + T without a saved syllable profile analysis file (e.g., for the first time), it will open your database and go through the entries there, and sort them by syllable profile and part of speech (CVC v CVCV for each of Nouns and Verbs, for example). This can take a couple of minutes. If you are running A→Z + T in a terminal, you should see its progress.

If you use symbols that A→Z + T doesn't recognize as word-forming (including a space), entries using those symbols will be excluded from the syllable profile analysis, and thus from any work in A→Z + T. For your convenience, a window appears on boot if this applies to you, including which characters are found, and how many entries they impact. If you feel entries are being excluded inappropriately (e.g., you have an orthographic symbol I haven't already accounted for), please write me with the details, and include a log.

There is a Segment Interpretations Settings (Advanced) menu item and window, to fine tune how A→Z + T does this analysis. For instance, some people will want to sort CV? data differently than CVC data, depending on the language situation. This can be done repeatedly, with two caveats:

1. Each time a change is made on this page (and you hit "Use these settings"), a reanalysis will happen, and this takes some time.
2. Each time you reanalyze your syllable profiles, you will be lumping or splitting groups. If you lump two or more groups together, this will impact your sorting done up to that point. E.g., if you had CV? and CVC separate, then join them, your new CVC group will have sort values from the old CV? **and** CVC sortings, though there is no necessary correspondence between those group name/numbers. Practically, this means that any time you lump groups, you should verify each of those groups again.

Settings File

If you open A→Z + T without a saved settings file, the program makes assumptions for you (based on your LIFT file), so you can get started right away, if you want to, e.g.:

- the largest syllable profile group, and its ps
- vowels first, before doing consonants or tone
- the segment with the largest representation in the database is chosen first.

You can change any setting by clicking on where that setting is displayed on the main screen, or through the menus.

Settings are saved to file each time you run a check, and loaded again on the next boot. If you change a setting that leaves another setting invalid (e.g., V1=g, or C1=C2 for VCV nouns), the invalid setting is removed and replaced with an assumption (as above).

The Main Window

On the upper left of the main window, each of these settings are indicated:

- interface language
- analysis language
- gloss language(s) —first (and possibly second)
- part of speech and syllable profile (with count)
- type of check (i.e., C, V, or T) and current stage/frame

There are by default **no menus**; if you see something you want to change (including the way the language name appears), click on it, and a menu will appear for you to change it.

If you prefer menus, you can *show menus* through a context menu via right click anywhere in the main window –these menus allow you to change all these settings, and more.

If the **font size** is too big for you (e.g., things are getting pushed off the screen), the context menu also allows you to select a theme with smaller fonts.

Tone Frames

If you don't have any tone frames set up yet, you will be asked to do so when you try to sort on tone. You can also do that in the Advanced menu, for as many frames as you want to define. Note the *name* is important, as this is how you will refer to this frame in A→Z + T, and how it will be identified in your database in the future (unless and until you change that). So if you're testing the plural form, something like "Pl" or "Plural", or "Pluriel" might be appropriate –but this is just a name, so make it distinct but useful to your workflow.

I hope the `Add Tone Frame` window is otherwise clear, though two points are in order:

1. I have seen at least one MS Windows system, where keyboarding that takes multiple keystrokes to produce a character (like 'n' + '>' → 'ŋ') showed up as '?' in the entry field. If this happens, *do not* ignore it, as this will be on every window, and added to your database examples. Rather, type the correct characters into another program (e.g., a text editor), then cut and paste into this field, and it should appear correctly.

2. The frame calculator is not particularly smart; **it only puts content before and after the form and gloss(es)** for each word, so you need to give it that information. If that information (in the form or gloss) alternates in agreement or harmony with the lexicon word forms, you should think through how you want to resolve potential clashes. The two main options are

- including all options in the frame:
    - `pl form: '__s/z/ɪz'` (with all forms given for each word in that frame)
- or allowing for ungrammaticality:
    - `pl form: '__s'` (knowing that dʒədʒ + pl will come out 'dʒədʒs', not 'dʒədʒɪz')

The user will have the option to skip a word that doesn't work in a frame. This was put there for syntactic or semantic clashes, but could be used to exclude phonologically weird frame combinations, too.

Once you have the form and gloss content in the appropriate boxes, click on 'see it on a word from the dictionary', and you will get the frame as you have defined it applied to some word (in the filter you have currently set). You can try this on a number of words; **continue to click that button**, to see how it will look on different entries. There is no easy way to change this frame after you define it (other than deleting them all and starting from scratch), so I encourage you to take your time before moving on. When you are happy with the frame, click on "use this tone frame".

If you absolutely regret a tone frame you have set up, all your frames are stored in `<lift filename>_ToneFrames.py` next to your LIFT file. Be careful editing this, though; you may need to redefine all your frames if you corrupt this file (This would be a great time to ask for help if you don't absolutely certainly know what you're doing).
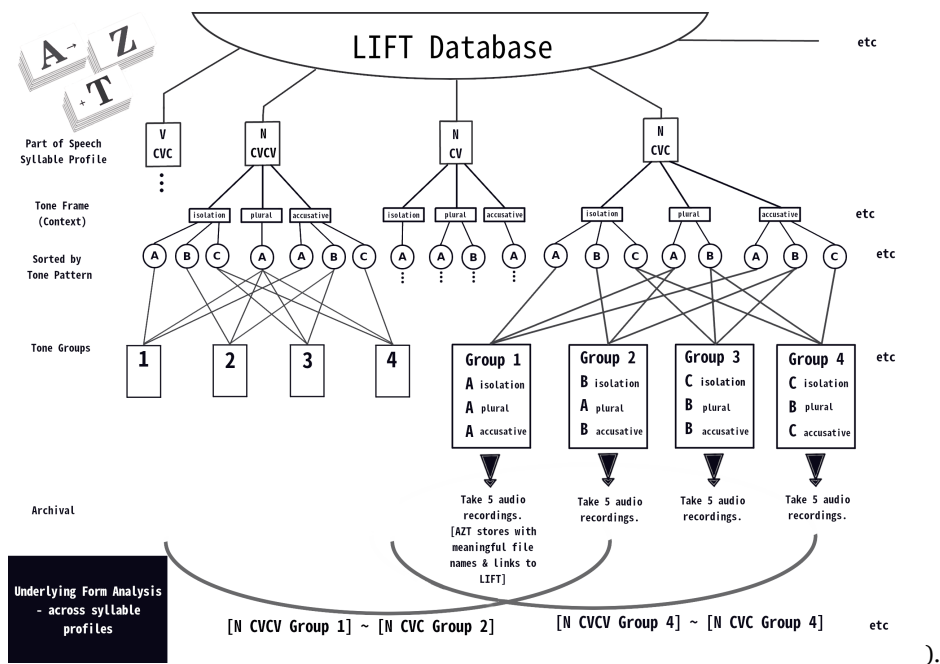

Tone Frame Groups

A→Z+T by default labels the groups into which you sort your data by frame with numbers. This is because, at least initially, the fact that it is its own thing (all one thing, and unlike the other groups) is more important than any description of the group, however objective it may be.

That said, there may come a time where you want to give one or more of these groups a particular name, like HL or [˥˩ ˩˩], perhaps because you want to remember how you thought of the surface form at the time, or because you don't want the group names in your database to be just numbers. To do this, right click on the verification window *while verifying the group you want to rename*, and click on Show Menu. This will provide a menu which will allow you to change the name.

Regarding name changes, please be cognizant of the fact that these names are for a *part of speech* and *syllable profile*, **in a given frame**. These should describe the *surface* tone only; hypotheses regarding underlying tone come later (see Joining and Renaming Draft Tone Groups, below; see also the following process flow chart:

LIFT Database

A→Z +T

etc

Part of Speech
Syllable Profile

| V | N | N | N |
| CVC | CVCV | CV | CVC |

Tone Frame
(Context)

isolation plural accusative   isolation plural accusative   isolation   plural   accusative   etc

Sorted by
Tone Pattern

A B C   A A B C   A A B A   A B C   A B   A B C   etc

Tone Groups

| 1 | 2 | 3 | 4 |

Group 1	**Group 2**	**Group 3**	**Group 4**	etc
A isolation	B isolation	C isolation	C isolation	
A plural	A plural	B plural	B plural	
A accusative	B accusative	B accusative	C accusative	

Archival

Take 5 audio recordings. [AZT stores with meaningful file names & links to LIFT]

Take 5 audio recordings.

Take 5 audio recordings.

Take 5 audio recordings.

Underlying Form Analysis
– across syllable
profiles

[N CVCV Group 1] ~ [N CVC Group 2]   [N CVCV Group 4] ~ [N CVC Group 4]   etc

).

Recording

The first time you try to record, you will be asked to tell A→Z + T what sound card parameters you want. You can set frequency, bit depth, and sound card number (to select between multiple cards, for your microphone and for your speakers. This window is designed with test `play` and `record` buttons, so you can set parameters and test them, before moving on. I suggest you budget some time to play with the settings there, until you are satisfied with them.

N.B.: There are some combinations of settings (likely those beyond the physical limits of your sound card) which **play fine in AZT, but do not produce good sound files**. This is simple enough to discover, by playing the sound file in another tool, such as praat. Please check for this before recording language data.

The settings window will you select the highest quality that it thinks your card can do; test to see if it records and plays back OK, then confirm by playing the sound file in anoter player. I have found several computers with cards that can record at 96khz, somewhat to my surprise —though be sure to think about your microphone and environment, etc, too! If you are making recordings for easy sharing over low bandwidth (as opposed to linguistic study), consider the implications of these setting on the size of your files.

I recommend not storing your sound card settings in your repository. They will be saved in a file, so you won't have to keep setting them, once you discover what works best in your context. However, because they are specific to a given computer, sharing them across computers with different sound cards is probably not what you want. The sound card settings dialog is also available through a menu item.

Regarding which sound card to choose (as A→Z + T will most likely give you multiple options):

- There may be multiple ways of telling A→Z + T to use a given interface, and they may not each give you the same mileage. For instance, you may get more consistency from setting the interface as the `default` sound card in your OS settings, then selecting `OS default` (or `System default` whatever) in A→Z + T —or the reverse; test and find out what works best for you.

Regarding external (e.g., USB) sound interfaces, there is an issue worth pointing out:

- In some contexts (perhaps particularly in places with lower quality power, or running off a generator?), I have noticed inconsistencies building up over time in recordings. They may start with pops or other problems in the recordings, and ultimately not record at all. In any case, there is an easy fix: shut down and restart A→Z + T. But this is a good reminder to listen to your recordings in real time, as you make them.

Subsequent Runs: CV analysis

(View data and run reports)

A→Z + T doesn't do CV sorting and verification (Yet!), but you can make recordings and filter your data and look at it through a number of checks (e.g., by C1, or by V1 = V2, etc.).

Recording Citation and Secondary Forms

When Consonants or Vowels are selected, you can click on "Record Dictionary Words", which will give you a page of `Record` buttons next to words filtered by ps-profile combinations, largest first. To skip to the next slice of data, just click "Next Group". For each entry in the data slice, this page provides a button to record a sound file for citation or lexeme fields, but also plural and imperative fields, if in the database. These recordings should then appear in reports, FLEx, and other uses of the LIFT database (e.g., the Dictionary App Builder).

Consonant and Vowel Reports

To look at filtered data, set up the desired parameters so they appear on the main screen, then click "Report!". This results in a window with data to scroll through, and an XLing-Paper XML document with the same output, ready for printing to PDF or HTML through the XMLmind XML Editor (XXE).

For a more thorough report, use the Basic Report menu item. This will select your top ps-profile combinations, and present each of those slices of your data in order, with the relevant checks for consonants and vowels for each. So the CVCV profile will start with $C1 = C2$, before moving to other data sorted by C1, then by C2 —then it will do the same for vowels. Longer syllable profiles will thus include more checks, though maybe not containing much data (e.g., if you don't have much $C1 = C2 = C3$ data).

This tool will ultimately be able to help with the sorting and correction of consonants and vowels, but until those functions are implemented, these reports should be helpful.

Subsequent Runs: Tone (Sort, and Follow Directions)

Sorting progression: The Status Table

Once you have done any sorting for the selected part of speech, to the right on the main window you will see a status pane, with groupings by syllable profile and check stage

(for one part of speech and check type at a time). To see progress for another check type or part of speech, switch to that check type or part of speech.

The program is designed to step through the process relatively automatically; once things are set up, you should be able to just open the program, and click Sort. If you need a break, click quit on whatever window you're in, and your progress should be there when you return.

You will, of course from time to time want to move to another part of speech or syllable profile, or check type. Click on any cell in the status table to select that combination of profile and check/frame (this can also be done on the main window menus) and the next time you click Sort or Record, the appropriate data will be selected, and those changes saved to the preferences file.

At some point you will likely want to give your sorting groups (within a ps-profile-frame) more meaningful names than the default (numbers). Until then, the status table gives you a count of the names, plus a count of the verified groups (w/+, in parentheses). Once you name a group, all groups for that ps-profile-frame will be listed separately, in the appropriate cell. If you don't like that, there is a context menu to "Hide group names," which will go back to just showing counts. In any case, the status table also indicates the presence of unsorted data in a group, via a preceding '!'.

Recording Data Sorted in Frames

Recording data in frames can be done at any point once sorting has begun. However, when a word is presented for recording, a button for each example (sorting context) in the sense appears. So if you sort one field, then record, then sort the next, you will see your earlier recordings again. Recordings can move rather quickly, so I recommend putting them last in your workflow, and doing them all at once —at least once you've tested that they're working correctly with your recording equipment, etc.

Tone Reports

Once you have done some sortings, it makes sense to run a report. The tone report will show your groupings in just one frame, if that's all you have done, but its real value lies in comparing values across multiple frames, so you'll want to check a couple tone frames before doing much with the tone reports.

The report by default has two phases: analysis and reporting. When analyzing your lexicon (from 'Do'/'Reports'), it compares sets of frame-value correspondences across the ps-profile slice being analyzed. That is, it separates the data until each group has exactly the same sorting values for each tone frame.

Once separated into these groups, the data is presented either by sense (including each sorted tone frame under each sense) or by tone frame/location (including each sense which has been sorted in each tone frame), depending on the report requested by the user.

In addition to the window that appears when the report is finished, each report is also exported to text and XLingPaper XML files. XLingPaper reports have similar organization, but more detail, as to what you will see in the report window.

Joining and Renaming Draft Tone Groups

The tone groupings provided by default in A→Z+T are definitely of the 'splitting' kind, as we want to make judgments ourselves as to which are the same underlying group, rather than have the computer do that. This will certainly lead to groups in the output which you would like to join/merge into a single group.

To join groups, use the 'Advanced' menu item 'Tone Reports':'Join/Rename…' This will present a window where you can tell A→Z+T which groups should be presented together in the report. These groupings are also written to the LIFT file, overwriting the place draft tone UF groupings for each sense impacted by the change.

Once you have made these groupings, it is important to **only run the tone report from the 'Advanced' menu**, if you want it to be based on your manual groupings. Otherwise, the default report will reanalyze your data and present draft groups from scratch.

To undo any grouping, simply run a tone report from the 'Do' (not 'Advanced') menu. This is a good idea to do whenever you have added sorting data to your database, so you can be sure that the groups you join/merge are based on the most recent 'splitting' draft groups. Or if you accidentally joined the wrong groups, or if you want to restart joining groups for any other reason.

Ad Hoc Sorting Groups

After you have done the analysis of noun and verb roots, you may want to move to a smaller grammatical category, and you may find that the slicing of data into strict ps-profile groups makes it difficult to see what's going on in your data. If you want to manually select a set of morphemes to sort, either a subset of a CV profile group, or including data from multiple CV profiles (or both), click on "Add/Modify Ad Hoc Sorting Groups" (In the Advanced menu). This provides a window where you can give a (unique!) name to your *ad hoc* group, and select the senses from your dictionary that you want to compare in it.

This window will leave you with that group selected, so you are immediately ready to sort the senses included in it. To avoid a conflict, A→Z+T will also switch your check type to tone.

Your *ad hoc* group is treated by A→Z + T as another CV profile group, so you can select it from the "Change/syllable profile" menu item, or else move back to a standard CV profile group. If you select a non-tone check while you have an ad hoc group selected (which is only valid for tone), A→Z + T will pick a non-tone profile for you.

Once you have an *ad hoc* group defined, you can modify it by selecting this menu item with that group selected on the main screen. To create another *ad hoc* group, select a standard CV profile on the main screen before selecting this menu item.

Miscellaneous

If any of the directions are unclear or inappropriate in any way, please let me know so I can fix them.

The UI is currently in French and English; if you want to translate it into another language, please let's talk!