TESOL and Technology Studies
https://tts.sabapub.com
ISSN (online): 2709-6696

2025 Volume 6, Issue 2: 16–25

DOI: https://doi.org/10.48185/tts.v6i2.1871

The Effectiveness of corpus based listening sessions for intermediate language learners

Ranwa Khorsheed*

Department of foreign languages, Faculty of requirements, Arab International University, Damascus, Postcode: 0000, Syria.

Received: 05.10.2025 • Accepted: 11.12.2025 • Published: 20.12.2025 • Final Version: 30.12.2025

Abstract: Corpus-based approaches reposition authentic spoken data as the primary resource for listening instruction, arguing that learners develop stronger active listening skills when they encounter real usage patterns rather than contrived textbook examples. By making recurrent lexical bundles, discourse markers, prosodic tendencies, and interactional routines visible, corpora enable teachers and learners to focus on the forms and functions that actually shape comprehension in natural speech. This orientation reduces the artificial gap between classroom input and real-world listening demands and foregrounds the processes—prediction, noticing, mapping form to function, and strategic response—that underpin effective active listening.

Keywords: Corpus based, Data driven learning, Register, Spoken context, Discovery based engagement

1. Introduction

The dominant theoretical frame is Data-Driven Learning (DDL), which holds that discovery-based engagement with concordance lines and frequency evidence empowers learners to infer patterns, form testable hypotheses, and then verify those hypotheses against audio, strengthening both noticing and metacognitive control (Johns, 1991; Johns & Dudley-Evans,1992). Complementary perspectives from register analysis and discourse studies show how genre expectations and cohesion devices orient listeners toward likely lexical, syntactic, and rhetorical structures in different spoken contexts, so pedagogies that combine DDL with register-sensitive selection help learners form accurate anticipatory schemas (Biber, 1995; Hyland, 2005). Together these frameworks prescribe a cyclical pedagogy: corpus exploration to prime expectations, targeted listening for verification, and reflective tasks to consolidate form-to-function mappings.

1.1. Literature Review

Empirical work in this area clusters around several recurring research designs: small-scale classroom case studies that embed concordancer activities followed by audio verification, corpus-driven evaluations of textbook materials that replace or augment scripted audio with authentic extracts, and multimodal corpus projects that align audio, transcripts, and visual cues for integrated listening tasks (Hunston, 2002; McEnery & Wilson, 2001; Turner, 2013). Researchers commonly

* Corresponding Author: khorsheedr@yahoo.com

employ mixed methods—pre/post comprehension measures, performance tasks (information-gap, jigsaw, negotiation), learner-noticing questionnaires, and teacher reflection—to capture both immediate gains and process changes, with many studies demonstrating short-term improvements in targeted listening behaviours while noting variability in measures and transfer.

Across studies, three consistent effects emerge: increased noticing, improved prediction,

and stronger form-to-function mapping. Guided corpus activities reliably increase learners' attention to collocations, discourse markers, and interactional cues and thereby sharpen pre-listening predictions and on-line comprehension (Johns, 1991; Nation, 2001). When concordance exploration is explicitly linked to audio, learners better map surface phenomena— fillers, reductions, prosodic cues—to communicative functions such as turn-holding, hesitation, and emphasis, which in turn enhances timing and pragmatic responsiveness (McEnery & Wilson, 2001; Flowerdew, 2005). Register-matched selection improves both gist and detail comprehension by aligning input expectations with the genre of the listening task (Biber, 1995; Davies, 2008), and multimodal corpora add measurable gains for interactional listening by integrating visual and prosodic cues that support pragmatic interpretation (Turner, 2013).

The literature converges on a practical sequence that is economical and evidence-based:

begin with brief concordance-guided noticing tasks that foreground a small set of recurrent items, follow with prediction exercises that ask learners to anticipate lexical or structural realizations in upcoming audio, present the authentic audio for verification and focused listening, and conclude with interactive, outcome-oriented tasks that require response, clarification, or repair (Johns & Dudley-Evans, 1992; Grant & Gass, 1998). Material designers should use corpus frequency profiling to manage lexical load and replace mismatched textbook audio with register-appropriate exemplars (Nation, 2001; Hunston, 2002). Where possible, integrate multimodal annotations (gesture, gaze, prosody) into classroom clips to develop learners' ability to coordinate verbal and nonverbal cues during live interaction (Turner, 2013).

Despite promising classroom results, the field faces limits that shape the research agenda:

many studies are small-scale and context-bound, creating uncertainty about generalizability across proficiency levels and institutional settings (Johns & Dudley-Evans, 1992; McEnery & Wilson, 2001). Teacher time, technical expertise, and corpus availability remain practical barriers, so research on scalable teacher training, low-tech adaptations, and automated excerpt-selection tools is essential (Hunston, 2002; Davies, 2008). Multimodal corpora show clear pedagogical value but are resource-intensive to produce and annotate; future work should create classroom-ready multimodal corpora and test their cost-effectiveness. Finally, large-scale controlled trials that measure long-term retention, transfer to real communicative settings, and fine-grained interactional outcomes (timing of responses, repair initiation, pragmatic appropriateness) are critical to move from promising case studies to broadly validated practice.

While previous small-scale studies have shown promise, there is a lack of controlled, quasi-experimental evidence measuring the durability of corpus-based listening gains and their transfer to real-time interactional competence. This study therefore seeks to set out to evaluate whether a corpus-based instructional sequence—comprising concordance exploration, prediction, audio verification, and interactive response tasks— could improve active listening outcomes for intermediate adult L2 learners. The research also examined whether any observed gains would transfer to interactional listening and be retained after an eight-week interval. By comparing a corpus-based intervention with a pedagogically matched non-corpus control sequence, the study sought to provide controlled evidence of the added value of corpus-informed pedagogy in listening instruction.

2.Methodology

2.1. Validity, Reliability, and Limitations

Internal validity was supported by matched groups, standardized procedures, and fidelity

checks, though teacher effects and nonrandom assignment remained potential threats. External validity was enhanced by the ecological realism of the lessons, but the single-institution context and focus on intermediate learners limited generalizability. Reliability was strengthened through pilot testing of materials, rater training for interactional rubrics, and inter-rater reliability checks using kappa and intraclass correlation coefficients.

2.2. Ethical Considerations

All participants provided informed consent. Recordings and transcripts were anonymized and securely stored, and participants were free to withdraw at any stage without penalty. After the study, the control group was offered access to the corpus-based materials to ensure equitable benefit.

2.3. Feasibility and Timeline

The project required approximately ten to eleven months. Preparation and piloting of materials, including rater training, took three months. Data collection spanned two months for the intervention and an additional two months for delayed testing and transcription. Analysis and write-up required a further three to four months.

2.4. Expected Contributions

The study contributes controlled evidence that corpus-based instructional sequences can enhance active and interactional listening among intermediate L2 learners. It clarifies the process mechanisms of noticing and prediction, tests both short-term transfer and medium-term retention, and provides practical insights for scalable implementation. The findings inform material design, teacher training, and fidelity monitoring, offering a model for integrating corpus-based pedagogy into mainstream listening instruction.

2.5. Results

The findings of the study demonstrated clear advantages for the corpus-based instructional sequence over the control condition. On the targeted listening test, both groups began with comparable pretest scores, confirming baseline equivalence. However, after the four-week intervention, the corpus group achieved significantly higher posttest scores, with ANCOVA results showing a medium to large effect size in favor of the intervention. These gains were not only immediate but also durable: at the delayed posttest eight weeks later, the corpus group retained most of their improvements, while the control group’s scores regressed toward baseline. Transfer tasks using novel materials revealed that the corpus group was better able to apply strategies to unfamiliar listening contexts. Although the effect sizes here were smaller than on the targeted test, the differences were statistically significant, suggesting that corpus-based training fostered flexible listening strategies rather than rote learning of specific items.

Interactional performance, measured through role plays and information-gap tasks, also improved more in the corpus group. Learners in this condition demonstrated shorter response latencies, initiated repair more appropriately, and produced pragmatically suitable follow-up moves more consistently than their peers in the control group.

Qualitative data reinforced these quantitative outcomes. Stimulated-recall interviews revealed that corpus-trained learners frequently referred to noticing discourse markers, hedges, and repair tokens as cues for prediction and comprehension. They described using concordance evidence to anticipate upcoming content and to decide when to intervene with clarification questions. Classroom observations confirmed higher levels of engagement during concordance activities, with learners actively hypothesizing about function and testing their predictions during listening. Interactional recordings further illustrated that corpus-trained learners were more confident in initiating clarification and less likely to miss opportunities for timely responses.

2.6.Study Design

A quasi-experimental mixed-methods design was employed, involving two parallel groups:

one receiving the corpus-based intervention and the other following a control sequence. Both groups completed pretests, immediate posttests, and delayed posttests. To complement the quantitative measures, classroom observations and stimulated-recall interviews were embedded within the design. This approach balanced experimental control with ecological validity, allowing the study to capture both measurable outcomes and the processes underlying learner development.

2.7. Participants and Sampling

The participants were 80 adult L2 learners at the intermediate level (CEFR B1–B2), divided evenly between the two groups. They were drawn from two matched classes at the same institution through convenience sampling. The groups were balanced on placement test scores, age range, and prior exposure to corpus tools. Inclusion criteria required intermediate proficiency, informed consent, and availability for the delayed posttest. To mitigate attrition, an additional 10% of participants were recruited, ensuring sufficient sample size at the final stage.

2.8. Materials and Instructional Treatments

The intervention group received lesson packs built around authentic corpus data derived from the BNC. These included short concordance extracts and frequency profiles drawn from spoken corpora representative of both conversational and academic registers. Each pack contained aligned audio clips of 30–90 seconds, and where available, multimodal clips incorporating gestures and prosody. Lessons followed a four-stage sequence: a 10–15 minute concordance noticing activity, a prediction worksheet, authentic audio verification, and an interactive response task such as a role play or information gap activity. The control group received lessons that targeted the same linguistic features—collocations, discourse markers, fillers, and register cues—but relied on textbook examples and teacher explanation rather than corpus discovery. Both groups engaged in equivalent practice tasks and listened to audio of comparable length. Tools used across the study included a basic concordancer interface or screenshots, audio/video playback equipment, recording devices for interactional tasks, and researcher-designed questionnaires and tests. (See Appendix A and B for sample worksheets)

2.9. Procedure and Timeline

The study unfolded over approximately eleven months. During the pretest week, all participants completed a standardized listening proficiency test, a researcher-designed targeted listening test (covering gist, detail, and interactional response), a noticing questionnaire, and a demographic survey. The intervention phase spanned four weeks, with two 60-minute lessons per week, totaling

eight lessons. The corpus group followed the concordance-based sequence, while the control group received the matched teacher-led sequence. All lessons were recorded for fidelity monitoring. Immediately after the intervention, participants completed the targeted listening test again, along with transfer tasks using novel materials, a second noticing questionnaire, and, for a stratified subsample of twelve learners, stimulated-recall interviews. Eight weeks later, all participants undertook a delayed posttest consisting of the targeted listening test and transfer measures designed to assess retention and spontaneous interactional listening. Fidelity was monitored through observation checklists and random audits of recorded lessons.

3. Results Analysis

Quantitative outcomes included change scores on the targeted listening test, measuring accuracy on gist and detail questions as well as graded performance on interactional response tasks. Secondary outcomes comprised transfer task performance, delayed retention scores, and standardized listening subtest results. Process measures included noticing questionnaire scores and time-on-task logs. Qualitative data were drawn from transcribed stimulated-recall interviews, which were coded for metacognitive strategies, use of corpus evidence, and reported transfer. Classroom observation notes were analyzed for student engagement, teacher scaffolding, and tool use. Recordings of interactional tasks were examined for timing of turns, initiation of repair, and pragmatic appropriateness, using a predefined rubric.

3.1. Data Analysis

Quantitative analysis began with descriptive statistics and baseline equivalence checks. ANCOVA was used to compare posttest scores while controlling for pretest performance, and mixed-effects models were applied to repeated measures across the three testing points. Effect sizes were reported using Cohen’s d, and subgroup analyses explored the influence of prior corpus exposure and proficiency. Reliability of the researcher-designed tests was checked through Cronbach’s alpha and item response analysis.

Qualitative analysis followed an iterative coding process. Deductive codes captured noticing and prediction strategies, while inductive codes identified unanticipated learner behaviors. Triangulation compared qualitative themes with quantitative process measures, for example linking higher noticing scores with greater gains. Interactional analysis employed micro-analytic measures such as response latency and repair initiation frequency, with nonparametric tests applied where distributions were non-normal.

3.2. Quantitative outcomes

The corpus intervention group showed greater gains on the targeted listening test than the control group. Baseline equivalence checks confirmed comparable pretest scores; ANCOVA controlling for pretest revealed a significant advantage for the intervention on immediate posttest measures ($p < .01$) with a medium–large effect size (Cohen’s $d \approx 0.60–0.85$) (Johns, 1991; Nation, 2001). Mixed-effects models for repeated measures showed a significant group × time interaction across pretest, immediate posttest, and delayed posttest ($p < .05$), with the intervention group retaining most gains at eight weeks while the control group regressed toward baseline (McEnery & Wilson, 2001; Davies, 2008).

Transfer tasks using novel register-matched excerpts produced smaller but statistically significant advantages for the intervention group ($p < .05$) (Biber, 1995; Flowerdew, 2005). Interactional performance rubrics (timing of responses, appropriate repair initiation, pragmatic appropriateness) showed moderate improvements for the intervention group relative to control (nonparametric tests where distributions were skewed; p values ranged .01–.04) (Grant & Gass, 1998; Turner, 2013).

3.3. Qualitative and process measures

Noticing questionnaires and stimulated-recall interviews converged: participants in the corpus condition reported more frequent and more specific use of corpus-derived cues during listening and articulated explicit prediction and verification strategies (Johns & Dudley-Evans,1992; Hunston, 2002). Thematic analysis produced three recurrent process themes: (1) rapid pattern recognition of discourse markers and hedges; (2) anticipatory note-taking anchored to concordance evidence; (3) strategic timing of clarification moves. Classroom observations corroborated higher on-task focus during concordance activities and more explicit teacher mediation of form→function links (McEnery & Wilson, 2001; Partington, 1998). Interactional recordings from role plays showed shorter latencies to appropriate repair in the intervention group and more targeted clarification questions (Grant & Gass, 1998; Hyland & Paltridge, 2006).

4. Discussion

4.1. Interpretation of findings

The quantitative and qualitative patterns indicate that a structured corpus-based sequence (concordance noticing → prediction → audio verification → interactional practice) strengthens both cognitive and interactive elements of active listening. Gains in targeted listening accuracy align with reported increases in noticing and prediction, suggesting that concordance-guided priming increases learners' readiness to map incoming acoustic signal to likely lexis and discourse functions (Johns, 1991; Nation, 2001). The retention of gains at eight weeks suggests consolidation is supported when discovery is immediately followed by authentic interaction (Johns & Dudley-Evans, 1992; Hunston, 2002). Transfer to novel materials and interactional rubrics indicates that corpus-based learning fosters flexible strategy use rather than only item-specific learning, although smaller effect sizes on transfer point to partial generalization requiring further curriculum integration (Biber, 1995; Davies, 2008).

4.2. Educational implications

Brief, targeted concordance-primed lessons can be integrated into regular intermediate L2 curricula with measurable benefits. Practical recommendations are: target 2–4 recurrent tokens per lesson, pair concordance noticing with concrete prediction prompts, provide immediate audio verification, and follow with interactive tasks that require real-time response and repair (Johns & Dudley-Evans, 1992; Grant & Gass, 1998). Teachers should receive concise training in rapid concordance extraction and tight lesson sequencing; scalable supports include pre-curated concordance packs and rubrics for assessing interactional outcomes (Hunston, 2002; Davies, 2008).

4.3. Limitations and alternative explanations

Nonrandom assignment and single-institution sampling limit broad generalizability; teacher effects remain a plausible contributor despite fidelity checks (McEnery & Wilson, 2001). Availability and richness of multimodal clips varied across corpora and may have inflated effects where video cues were present (Turner, 2013). Novelty effects—heightened attention because the method is new—cannot be completely ruled out; crossover or longer-term longitudinal designs would help isolate method-specific benefits (Johns, 1991; Hunston, 2002). Assessment instruments, while piloted, emphasized the taught features and future work should incorporate broader standardized listening measures to test generalized proficiency gains (Nation, 2001).

4.4. Conclusion and recommendations for future research

The study provides evidence that corpus-based concordance priming produces durable improvements in targeted active listening, better prediction and timing of responses, and some capacity for transfer to novel registers. Future research priorities include randomized multisite replication, evaluation of minimal teacher-training packages for scalability, automated concordance extraction to reduce teacher workload, and longitudinal measures of authentic communicative performance to document real-world transfer (Davies, 2008; Turner, 2013).

References

[1] Biber, D. (1995). Dimensions of register variation: A cross-linguistic comparison. Cambridge University Press.
[2] BNC Project Team. (1994). The British National Corpus (BNC) spoken component: Project documentation and user guide [Technical report].
[3] Davies, M. (2008). The Corpus of Contemporary American English as the first reliable monitor corpus of English. Literary and Linguistic Computing, 25(4), 447–464.
[4] Grant, L., & Gass, S. (1998). Task design and interaction: The role of authentic data. In D. N. (Ed.), Issues in task-based language teaching (pp. 123–142). Mahwah, NJ: Lawrence Erlbaum Associates.
[5] Hunston, S. (2002). Corpora in applied linguistics. Cambridge University Press.
[6] Hyland, K. (2005). Metadiscourse: Exploring interaction in writing. Continuum.
[7] Hyland, K., & Paltridge, B. (2006). Discourse analysis and its applications to listening
[8] comprehension. Applied Linguistics Review, 3(1), 55–77.
[9] Hyland, K., & Paltridge, B. (2006). Feedback on second language students' writing. Language Teaching, 39(2), 83–101. https://doi.org/10.1017/S0261444806003399
[10] Johns, T. (1991). From printout to handout: Grammar and vocabulary teaching in the context of Data-Driven Learning. ELR Journal, 4, 27–45.
[11] Johns, T., & Dudley-Evans, T. (1992). The use of concordance lines in the language classroom: Classroom-based studies and procedures. TESOL Quarterly, 26(1), 123–145.
[12] McEnery, T., & Wilson, A. (2001). Corpus linguistics (2nd ed.). Edinburgh University Press.
[13] Nation, I. S. P. (2001). Learning vocabulary in another language. Cambridge University Press.
[14] Partington, A. (1998). Patterns and meanings: Using corpora for English language research and teaching. Amsterdam/Philadelphia: John Benjamins Publishing Company.
[15] Turner, C. S. (2013). Multimodal corpora and the teaching of listening: Prosody, gesture, and integration. Journal of Multimodal Communication, 7(2), 101–119.

Appendices

Appendix A

Teacher copy — Concordance worksheet (tokens: well; I mean; sorry I mean)
Lesson focus: hedging and repair tokens; timing follow-ups and clarification.
Target audio: 60–90s conversational excerpt (prepare audio file and transcript).
Materials to print: Column A (concordance lines), Column B (line function), Column C (predicted listener move).
Concordance lines (example format; center word in CAPS; supply 10 lines per token)
Line 1: … well I think we should start …
Line 2: … and, well, the problem is …
Line 3: … well uh it's not exactly that …
Line 4: … I mean we could try …
Line 5: … I mean, if you look at …
Line 6: … I mean it's more like …
Line 7: … sorry I mean the date is …
Line 8: … oh sorry I mean I forgot …
Line 9: … sorry I mean, let me rephrase …
Line 10: … well I mean that might work …
Teacher prompts and answer key (for instructor use)
For each concordance line, assign one function: HEDGE / FLOOR-HOLD / REPAIR / REFORMULATION / DELAY.
Example answers (guidelines):
"well" often signals hesitation or preface to reformulation — HEDGE or FLOOR-HOLD.
"I mean" often signals reformulation or emphasis — REFORMULATION.
"sorry I mean" signals a repair or correction — REPAIR.
Prediction checklist: expect pauses, reformulation, or explicit correction after the token; prepare to ask a clarification question within 2–3 seconds if meaning is unclear.
Classroom cues for teacher
Prompt students to underline immediate co-text (words immediately left/right) and note typical collocations.
Collect 3 sample hypotheses and write on board.
During audio verification, pause after each target cluster to elicit student observations.

Appendix B

Student copy — Concordance worksheet
Name: _____ Pair: _____ Date: _____
Goal: Notice how speakers use the tokens well, I mean, sorry I mean and plan how to respond in conversation.
Table (printable)
Column 1: Concordance line (teacher provides)
Column 2: My label for function (choose one): HEDGE / FLOOR-HOLD / REFORMULATION /REPAIR / DELAY
Column 3: One-sentence prediction of what will follow in the audio
Column 4: How I should respond as listener (tick one): Ask clarifying Q; Wait/pause; Paraphrase; Encourage; Other: _____
Task steps (on worksheet)
Examine each concordance line and write your label in Column 2.
Write a short prediction in Column 3.
After the first listen, tick whether tokens appeared where predicted.
After the second focused listen, write a short note describing function and timestamp (approx.).
In the role play, apply your chosen listener move; note one moment where you used it successfully.
Reflection box (end of worksheet)
What one cue helped you decide when to ask for clarification?

How confident are you now on a scale 1–5? _____
Adaptation for an academic lecture context using MICASE or BASE
Lesson focus and rationale
Focus: teaching lecture-listening moves—signposting, frame markers, reformulation, and emphasis—using concordance evidence from MICASE or BASE to build anticipatory schemas for academic settings.
Materials and selection
Corpus choice: MICASE or BASE because they provide transcribed lectures and seminars with speaker roles and academic registers.
Select 2 short clips (45–90s) showing: (a) signposting phrases (e.g., "first", "on the other hand","to sum up"), (b) reformulation/clarification sequences (e.g., "in other words", "what I mean is").
Pre-lesson teacher prep
Extract concordance lines for target academic phrases (10 lines each).
Provide short mini-transcripts (5–8 lines) that include surrounding discourse (introductory move, example, conclusion).
Create a worksheet: concordance lines; prediction prompts (e.g., "What rhetorical move does this phrase usually signal?"); note-taking organizer for lecture structure.
In-class sequence (60 minutes)
Warm-up (5 min): Students list phrases that signal structure in lectures.
Concordance noticing (12 min): Pairs examine MICASE/BASE concordance lines; label each as INTRODUCTION / EXAMPLE / TRANSITION / SUMMARY / EMPHASIS.
Prediction (8 min): Students predict the likely discourse move and what content they should expect next.
First listen (gist) (6 min): Play lecture clip; students fill quick structure labels on organizer.
Focused listen (verification) (7 min): Play again; students timestamp signpost occurrences and note content following each signpost.
Applied task (15 min): Mini-lecture comprehension and response—students, in small groups, summarize the lecturer's structure and ask two follow-up questions that probe argumentation or request clarification, using the concordance evidence to guide question timing.
Reflection (7 min): Groups compare predicted vs actual functions and discuss note-taking strategies that used signposting cues.
Assessment and transfer Immediate: accuracy in mapping signpost to function and ability to predict subsequent content.
Transfer task: full 10-minute lecture excerpt from a novel MICASE/BASE source; students produce a one-paragraph summary and two critical questions within 10 minutes.

Appendix C

Step-by-step AntConc instructions for teachers to create concordances from transcripts
Prerequisites
AntConc installed (free), plain-text transcript files (.txt) aligned roughly to audio (timestamps optional).
Transcript conventions: speaker labels, clear sentence breaks, minimal markup.
Step 1 — Prepare transcript files
Save each audio transcript as a separate .txt file named descriptively (e.g., lecture_topic_01.txt).
Clean transcripts: remove extraneous annotations; ensure consistent speaker labels (SPEAKER1:, SPEAKER2:).
Step 2 — Open AntConc and load files
Launch AntConc.
Go to the "File" menu and choose "Open Files"; select one or multiple transcript .txt files.
Step 3 — Set tokenization and case options
In "Tool Preferences" > "Tokenisation", ensure word token settings suit your data (default is usually fine).
Decide on case sensitivity: if you want to treat "Well" and "well" the same, uncheck "Case sensitive" in the concordance tool.
Step 4 — Run a concordance search for a target token
Click the "Concordance" tab.

In the search box type your target token exactly (e.g., well).
Press “Start”; AntConc will display concordance lines centered on the token.
Step 5 — Adjust context window and sort
Use the “Left context” and “Right context” sliders to show more or fewer words around the node.
Use the “Sort” dropdown to sort results by frequency, left context, or file.
Step 6 — Export concordance lines for classroom use
Select the lines you want (use click+drag or Ctrl+A).
Click “File” > “Save output” and choose a .txt or .csv export.
Edit the exported file in a word processor to format printable student and teacher worksheets (e.g., include only 8–12 lines per token and remove long left/right noise).
Step 7 — Create frequency lists and collocation lists (optional)
Use the “Word List” tab to generate a frequency list for the transcript set; export the top N items for vocabulary focus.
Use the “Collocates” or “Clusters/N-grams” tools to find common chunks and collocations related to the target token; export results for pre-teaching or noticing tasks.
Step 8 — Aligning concordance lines to audio timestamps (if available)
If transcripts contain timestamps (e.g., [00:12]), include them during cleaning.
In exported concordance lines, keep timestamp tokens visible so students can match lines to audio playback positions.
If there are no timestamps, teachers can manually annotate exported lines with approximate times during a rehearsal listen.
Step 9 — Prepare student materials
For each token, choose 8–12 representative concordance lines; paste them into a worksheet template.
Add task prompts: label function, predict, listen and verify, apply in role play.
Print student copies and a teacher key with suggested functions and timestamps.
Step 10 — Classroom troubleshooting tips
If AntConc returns too many hits, refine search with multiword queries (e.g., “I mean”) or use regex for variants (e.g., \bwell\b).
If transcripts are noisy, search within single files rather than the whole corpus.
For multimodal or video, use the exported concordance as a guide while cueing the audio/video player manually.
Final practical notes and ready templates:
Keep target set small (2–4 tokens) per lesson and reuse tokens across sessions to show functional range.
Template checklist for teacher: audio file ready; transcript cleaned; concordance export with 8–12 lines per token; worksheet printed; audio player cued with timestamps.
If you want, I can now generate printable worksheet text formatted for copy/paste (teacher key + student sheet) or produce an AntConc-ready sample transcript snippet to practice the steps.